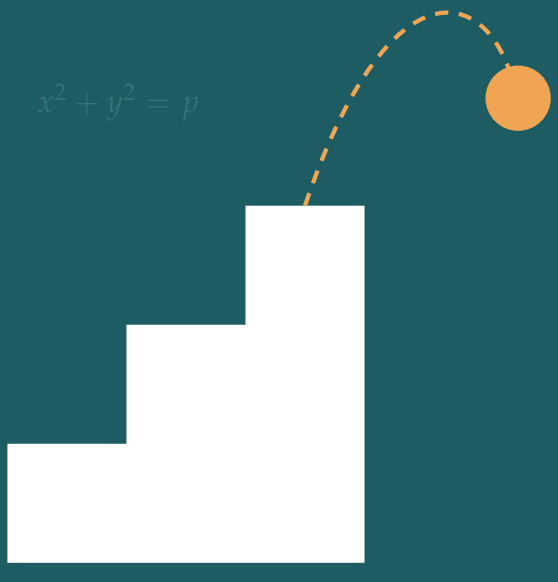
STEP 备试 · 大学数学

# Beyond STEP

## STEP I, II 及 III

1997 年详细解答



KEITH McNULTY

# Beyond STEP

以历届 STEP 试题为门
走进大学数学

STEP I, II 及 III

1997 年详细解答

Keith McNulty

**Beyond STEP：以历届 STEP 试题为门，走进大学数学**

**1997 年详细解答**

1997 年 STEP 试卷（STEP I, II 及 III）全部 42 题的详细解答。自样本试卷以来每一道 STEP 题目的详细解答、*Beyond STEP* 全书以及各章的单行本，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

阅读解答之前，请先认真尝试解题。每个答案均已用数值方法核对，但解答数量如此之多，难免会有疏漏：如果发现有可疑之处，请相信自己的演算，并仔细检查。

由机器从英文翻译而成；尚未经审阅。

2026 年 11 月首次于网上发表。

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

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

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

# 目录

## STEP I


第 1 题 硬币组合的计数 纯数学 3

第 2 题 反三角函数的导数 纯数学 6

第 3 题 幂与个位数字 纯数学 10

第 4 题 绝对值方程 纯数学 13

第 5 题 几何中的向量 纯数学 16

第 6 题 以积分估计 $\pi$ 的界限 纯数学 19

第 7 题 高斯求积法 纯数学 23

第 8 题 指数方程的根 纯数学 26

第 9 题 相对速度 力学 29

第 10 题 圆周运动 力学 33

第 11 题 有阻力的运动 力学 37

第 12 题 中心极限定理 概率与统计 41

第 13 题 条件概率 概率与统计 44

第 14 题 期望成本的最小化 概率与统计 47


## STEP II


第 1 题 数字和与计数 纯数学 50

第 2 题 递推关系 纯数学 53

第 3 题 部分分式积分法 纯数学 56

第 4 题 余式定理 纯数学 59

第 5 题 复指数映射 纯数学 62

第 6 题 三角方程 纯数学 66

第 7 题 曲线描绘 纯数学 70

第 8 题 以积分证明不等式 纯数学 74

第 9 题 摩擦力与刚体平衡 力学 77
第 10 题 爆炸与弹性碰撞 力学 81
第 11 题 抛体运动 力学 85
第 12 题 项分布的奇偶概率 概率与统计 89
第 13 题 蒲丰投针问题 概率与统计 93
第 14 题 放松过程 概率与统计 97

## STEP III

第 1 题 指数幂级数 纯数学 101
第 2 题 曲线描绘 纯数学 105
第 3 题 单位根 纯数学 109
第 4 题 算术几何平均不等式 纯数学 113
第 5 题 摆线的弧长 纯数学 116
第 6 题 切比雪夫微分方程 纯数学 119
第 7 题 $e$ 的无理性 纯数学 123
第 8 题 二次型的对角化 纯数学 126
第 9 题 涉及弹簧的静力学 力学 130
第 10 题 有滑动的滚动 力学 134
第 11 题 刚体转动 力学 138
第 12 题 赠券收集问题 概率与统计 142
第 13 题 正态分布的矩生成函数 概率与统计 146
第 14 题 随机变量的乘积 概率与统计 149

详细解答

# STEP I 1997，第 1 题

纯数学 · 硬币组合的计数

> **STEP I 1997，第 1 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s1-q1

本题在 *Beyond STEP* 第 9.5 节*凑零钱*中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/making-change.html

## 关键思路

要数“方法”的数目，必须有一套系统，把每一个组合恰好列出一次。最自然的做法，是先定下面额最大的硬币的数目，然后是次大的，如此类推。一旦只剩下 2p 和 1p 硬币，数目便立即可得：2p 硬币的数目决定了一切。这样做还有一个好处：10p 的答案可以直接用于 20p 的答案。

## 解答

次序并不重要：一种“方法”就是选定每种硬币各用多少个。因此我们要数的是非负整数（non-negative integer）解的数目。

### 基本组件：2p 和 1p 硬币

要用 2p 和 1p 硬币凑成 $r$ 便士，我们选定 2p 硬币的数目 $b$，它可以是 0 至 $\lfloor r/2 \rfloor$ 之间的任何整数。余下的 $r - 2b \geq 0$ 便士便由 1p 硬币凑成，而且恰好只有一种方法。所以方法的数目是

$$\left\lfloor \frac{r}{2} \right\rfloor + 1.$$

### 凑成 10 便士

- 使用一个 *10p* 硬币。它独自凑成整个金额：1 种方法。
- 不用 *10p* 硬币。我们按 5p 硬币的数目分类，即 0、1 或 2 个：

| 5p 硬币数目 | 留给 2p 和 1p 的金额 | 方法数目 $\lfloor r/2 \rfloor + 1$ |
|---|---|---|
| 2 | 0 | 1 |
| 1 | 5 | 3 |
| 0 | 10 | 6 |

这样共有 $1 + 3 + 6 = 10$ 种方法。

每一个组合都有确定的 10p 硬币数目和 5p 硬币数目，所以它恰好被数了一次，不多不少。总数是 $1 + 10 = 11$ 种方法，正如题目所求。

### 凑成 20 便士

以 $F(n)$ 表示用 5p、2p 和 1p 硬币凑成 $n$ 便士的方法数目。我们刚才已得出 $F(10) = 10$，而显然 $F(0) = 1$（不用任何硬币）。至于 $F(20)$，我们按 5p 硬币的数目分类，由 0 至 4 个：

| 5p 硬币数目 | 留给 2p 和 1p 的金额 | 方法数目 |
|---|---|---|
| 4 | 0 | 1 |
| 3 | 5 | 3 |
| 2 | 10 | 6 |
| 1 | 15 | 8 |
| 0 | 20 | 11 |

所以 $F(20) = 1 + 3 + 6 + 8 + 11 = 29$。

? **想一想** 怎样把 20p 和 10p 硬币纳入计算，而又不会重复计数?

再按大面额硬币分类。要么我们使用 20p 硬币，那么它必定是唯一的硬币：1 种方法。要么不用，这时我们按 10p 硬币的数目分类：

- 两个 10p 硬币：没有余额，$F(0) = 1$ 种方法；
- 一个 10p 硬币：余下 10p，$F(10) = 10$ 种方法；
- 没有 10p 硬币：余下 20p，$F(20) = 29$ 种方法。

这些情况互不相交（disjoint），并且涵盖了所有可能，所以凑成 20 便士的方法数目是

$$1 + 1 + 10 + 29 = 41.$$

## 验算

- 亲手列出 10p 的情况：10；5+5；5+2+2+1、5+2+1+1+1、5+1 × 5；然后是 2 × 5、2 × 4+1 × 2、…、1 × 10（这类共六个）。合共 11 个。
- 用电脑简单地数一次（对每种硬币的数目逐一循环），得出 11 和 41。这同时确认了中间值 $F(10) = 10$ 和 $F(20) = 29$，以及不用 20p 硬币时共有 40 种方法。
- 小的情况很容易亲手验算：用 5p、2p 和 1p 硬币凑成 5p 共有 $F(5) = 4$ 种方法，与 $1 + 3$ 吻合。

## 常见错误

- 把次序也计算在内，以致 5+2+2+1 和 2+5+1+2 变成不同的“方法”。它们其实是同一组硬币。
- 遗漏只用一个硬币的方法：单独一个 10p 硬币，以及单独一个 20p 硬币。
- 只给出一个总数，却看不出任何系统。题目已提醒，没有解释的答案不会得分，所以分类讨论本身才是重点。

## STEP 之外

这些其实是把 $n$ 拆成取自某个固定集合的部分的分拆（partition）。处理它们的巧妙方法是使用生成函数（generating function）：方法的数目就是

$$\frac{1}{(1-x)(1-x^2)(1-x^5)(1-x^{10})(1-x^{20})},$$

中 $x^n$ 的系数，而我们的分类讨论，不过是把这些因式逐个乘开而已。生成函数把计数问题化为代数，并由此通往欧拉（Euler）和拉马努金（Ramanujan）关于分拆的研究。想进一步探索，可参考 Siklos 的 *Advanced Problems in Mathematics*；至于数论方面，则可参考 Davenport 的 *The Higher Arithmetic*。

详细解答

# STEP I 1997，第 2 题

纯数学 · 反三角函数的导数

> **STEP I 1997，第 2 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s1-q2

## 关键思路

在第 (i) 部分，f 的导数（derivative）会化简为零，所以 f 是常数，但只限于它有定义的每一个区间（interval）之内：点 $x = -1$ 把实数线分成两段，而两边的常数并不相同。在第 (ii) 部分，先求 $x$ 对 $y$ 的导数，再把结果倒转过来会较为容易；功夫在于把 $\sin y^2$ 和 $\cos y^2$ 改写成以 $x$ 和 $y$ 表示的式子，而 $y$ 的取值范围正正在这里起作用。

## 解答

### 第 (i) 部分

这个函数对所有 $x \neq -1$ 都有定义。设 $u = \dfrac{1-x}{1+x}$。由商法则（quotient rule），

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{-(1+x)-(1-x)}{(1+x)^2} = \frac{-2}{(1+x)^2}, \qquad 1+u^2 = \frac{(1+x)^2+(1-x)^2}{(1+x)^2} = \frac{2(1+x^2)}{(1+x)^2}.$$

利用 $\dfrac{\mathrm{d}}{\mathrm{d}x}\tan^{-1} u = \dfrac{1}{1+u^2}\dfrac{\mathrm{d}u}{\mathrm{d}x}$（链式法则（chain rule）），

$$\begin{aligned} \mathrm{f}'(x) &= \frac{1}{1+x^2} + \frac{(1+x)^2}{2(1+x^2)} \cdot \frac{-2}{(1+x)^2} \\ &= \frac{1}{1+x^2} - \frac{1}{1+x^2} = 0 \qquad (x \neq -1). \end{aligned}$$

> ? **想一想** $\mathrm{f}'(x) = 0$ 是否表示 f 在所有地方都等于同一个常数?

不是。导数为零只能保证函数在一个区间上是常数（这是由中值定理（mean value theorem）得出的）。f 的定义域（domain）是 $(-\infty, -1) \cup (-1, \infty)$，即两个区间，所以 f 在

每一段上都是常数，而这两个常数可以不同。要求出每个常数，可以代入一个方便计算的点，或者取极限（limit）。

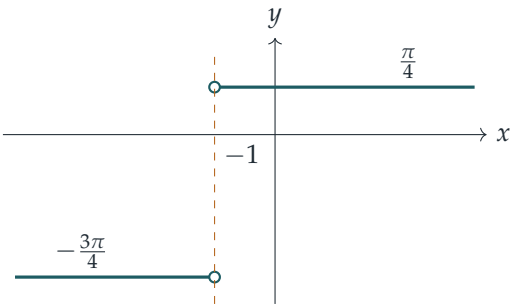
**当 $x > -1$：** 取 $x = 0$。于是 $\mathrm{f}(0) = \tan^{-1} 0 + \tan^{-1} 1 = \frac{\pi}{4}$。

**当 $x < -1$：** 令 $x \to -\infty$。这时 $\tan^{-1} x \to -\frac{\pi}{2}$，而 $\dfrac{1-x}{1+x} \to -1$，所以 $\tan^{-1}\left(\dfrac{1-x}{1+x}\right) \to -\frac{\pi}{4}$。由于 f 在这个区间上是常数，这个常数就等于该极限，即 $-\frac{3\pi}{4}$。（另一个做法：取 $x = -\sqrt{3}$，得 $\dfrac{1+\sqrt{3}}{1-\sqrt{3}} = -(2+\sqrt{3}) = -\tan\frac{5\pi}{12}$，所以 $\mathrm{f}(-\sqrt{3}) = -\frac{\pi}{3} - \frac{5\pi}{12} = -\frac{3\pi}{4}$。）

因此

$$\mathrm{f}(x) = \begin{cases} \dfrac{\pi}{4}, & x > -1, \\ -\dfrac{3\pi}{4}, & x < -1. \end{cases}$$

函数在 $x = -1$ 处出现大小为 $\pi$ 的跳跃，是因为 $\dfrac{1-x}{1+x}$ 在该处由 $-\infty$ 直冲到 $+\infty$，于是它的反正切（inverse tangent）便由 $-\frac{\pi}{2}$ 跃升至 $\frac{\pi}{2}$。




**第 (ii) 部分**

利用积法则（product rule）和链式法则，求 $x = y\sin y^2$ 对 $y$ 的导数：

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \sin y^2 + 2y^2 \cos y^2.$$

现在把两个三角函数项都用 $x$ 和 $y$ 表示。

- 由于 $y > 0$，可以除以 $y$：$\sin y^2 = \dfrac{x}{y}$。
- 由于 $0 < y^2 < \frac{\pi}{2}$，有 $\cos y^2 > 0$，所以取正平方根：

$$\cos y^2 = \sqrt{1 - \sin^2 y^2} = \sqrt{1 - \frac{x^2}{y^2}} = \frac{\sqrt{y^2 - x^2}}{y},$$

其中最后一步用到 $\sqrt{y^2} = y$，这是因为 $y > 0$ 才成立。

所以

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{x}{y} + 2y\sqrt{y^2 - x^2} = \frac{x + 2y^2\sqrt{y^2 - x^2}}{y}.$$

这个式子严格为正，因为 $x > 0$、$y > 0$，而平方根是非负的。因此可以把它倒转（反函数法则（inverse function rule）$\frac{\mathrm{d}y}{\mathrm{d}x} = 1/\frac{\mathrm{d}x}{\mathrm{d}y}$，在 $\frac{\mathrm{d}x}{\mathrm{d}y} \neq 0$ 时成立）：

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{y}{x + 2y^2\sqrt{y^2 - x^2}},$$

这正是所要证明的结果。

> ? **想一想** 为甚么题目坚持要 $0 < y < (\pi/2)^{1/2}$?

没有这个条件，$\cos y^2$ 就有可能是负数，而正确的公式在平方根前面便要有一个负号。这个限制同时令 $y \sin y^2$ 在这个范围内严格递增（strictly increasing）（上面求得的导数是正数），由 0 上升至 $(\pi/2)^{1/2}$，所以在 $0 < x < (\pi/2)^{1/2}$ 内的每一个 $x$ 都恰好对应一个 $y$，$y$ 才真正是 $x$ 的函数。

## 验算

- 第 (i) 部分：$\mathrm{f}(1) = \frac{\pi}{4} + \tan^{-1} 0 = \frac{\pi}{4}$，而 $\mathrm{f}(-3) = \tan^{-1}(-3) + \tan^{-1}(-2) \approx -1.2490 - 1.1071 = -2.3562 \approx -\frac{3\pi}{4}$。我们在 10,000 个随机点上验证过这两个常数。
- 第 (ii) 部分：当 $x = 0.5$，以数值方法解 $y \sin y^2 = 0.5$ 得 $y \approx 0.81355$，而公式给出 $\frac{\mathrm{d}y}{\mathrm{d}x} \approx 0.60284$，与用数值方法求得的反函数的有限差分（finite difference）导数吻合。
- 在 $x = 0$ 附近，$y \sin y^2 \approx y^3$，所以 $y \approx x^{1/3}$，且 $\frac{\mathrm{d}y}{\mathrm{d}x} \approx \frac{1}{3y^2}$；公式给出 $y/(y^3 + 2y^3) = \frac{1}{3y^2}$，两者一致。

## 常见错误

- 断定对所有 $x$ 都有 $\mathrm{f}(x) = \frac{\pi}{4}$，忘记了 $x = -1$ 把定义域分成两部分。
- 写出 $\cos y^2 = \pm\sqrt{\cdots}$，却没有说明为甚么应取正根。

## STEP 之外

$\mathrm{f}' = 0$ 只能在定义域的各个连通（connected）部分上得出常数，这是中值定理的推论，而中值定理是大学分析（analysis）课程中最早遇到的重要结果之一。第 (ii) 部分是反函数定理（inverse function theorem）的一个入门例子：导数不为零的函数可以在局部求反

函数，而反函数的导数就是原来导数的倒数。Abbott 的 *Understanding Analysis*（分析学入门教材）对这两个课题都有仔细的论述。

详细解答

# STEP I 1997，第 3 题

纯数学 · 幂与个位数字

> **STEP I 1997，第 3 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s1-q3

## 关键思路

动笔之前，先把 $a_7$ 写成 3 的单一个幂（power）：把 3 的幂取立方，只是把指数（exponent）乘以三。之后各部分其实是就同一个庞大的数提出不同的问题。个位数字（units digit）只取决于指数模（modulo）4 的值，因为 3 的幂的个位数字以周期（period）4 循环出现。大小的界只需要与 10 的幂作一个粗略的比较。最后一部分则建基于这个界：$a_7$ 极大，$1/a_7$ 便小得可以忽略。

## 解答

首先求出封闭形式（closed form）。若 $a_n = 3^k$，则 $a_{n+1} = (3^k)^3 = 3^{3k}$，所以每一步指数都乘以 3。由 $a_1 = 3^1$ 出发，用归纳法（induction）可得

$$a_n = 3^{3^{n-1}}, \qquad \text{所以} \qquad a_7 = 3^{3^6} = 3^{729}.$$

### 第 (i) 部分

乘积的个位数字只取决于各因数的个位数字。因此要追踪 $3^k$ 的个位数字，只需把上一个个位数字乘以 3，然后只保留个位：

| $k$ | 1 | 2 | 3 | 4 | 5 | ⋯ |
|---|---|---|---|---|---|---|
| 个位数字：$3^k$ | 3 | 9 | 7 | 1 | 3 | ⋯ |

到 $k = 4$ 时个位数字回到 1，此后 $3, 9, 7, 1$ 这个规律便以周期 4 重复出现。（换句话说，$3^4 = 81 \equiv 1 \pmod{10}$，所以 $3^{4m+r} \equiv 3^r \pmod{10}$。）

? **想一想** 怎样不用长除法求出 729 模 4 的值?

由于 $3 \equiv -1 \pmod 4$, 我们有 $3^6 \equiv (-1)^6 = 1 \pmod 4$。(直接计算: $729 = 4 \times 182 + 1$。) 因此 $3^{729} = 3^{4\cdot 182+1}$ 的个位数字与 $3^1$ 的相同，即

故$a_7$ 的个位数字是3.

### 第 (ii) 部分

把指数每三个分成一组：$729 = 3 \times 243$，所以

$$a_7 = 3^{729} = \left(3^3\right)^{243} = 27^{243} > 10^{243} \geq 10^{100},$$

当中用到 $27 > 10$（以及正数取正数次幂后大小次序不变这个事实)。因此 $a_7 \geq 10^{100}$，正如所求。这个界远远谈不上紧：事实上 $a_7$ 有 348 位数字。

### 第 (iii) 部分

把分数拆开：

$$\frac{a_7 + 1}{2a_7} = \frac{1}{2} + \frac{1}{2a_7}.$$

由于 $a_7 > 0$，第二项是正数。根据第 (ii) 部分，它同时极小：

$$0 < \frac{1}{2a_7} \leq \frac{1}{2 \times 10^{100}} < 10^{-100}.$$

因此

$$0.5 < \frac{a_7 + 1}{2a_7} < 0.5 + 10^{-100} < 0.505.$$

凡是严格介乎 0.5 与 0.505 之间的数，取至小数点后两位都是 0.50：小数点后第三位是 0，远远未到需要进位的 5。所以

$$\frac{a_7 + 1}{2a_7} = 0.50 \text{ （准确至小数点后两位)。}$$

事实上，它的小数展开式以 $0.5000\ldots$ 开始，在 5 之后至少有 100 个零。

## 验算

电脑可以精确地处理 $3^{729}$。把 3 连续取立方六次，可确认 $a_7 = 3^{729}$，这个数有 348 位数字，个位数字是 3，并且 $27^{243} = 3^{729} > 10^{243}$。用精确的分数运算，可确认 $\frac{a_7+1}{2a_7}$ 严格介乎 $\frac{1}{2}$ 与 $\frac{1}{2} + 10^{-100}$ 之间。你也可以用较小的例子检验第 (i) 部分的方法：$a_2 = 3^3 = 27$ 的指数是 $3 \equiv 3 \pmod 4$，由此预测个位数字是 7，结果确实如此。

## 常见错误

- 把 $a_7$ 看成 $3^{3\cdot 7}$ 或 $3^{18}$，而不是 $3^{3^6}$：每一步是把指数乘以 3，而不是加 3。

- 把指数对模 10 化简，而不是对模 4 化简。个位数字以周期 4 循环，所以起作用的是指数模 4 的值。
- 在第 (iii) 部分只写“0.5”而不加任何理据。“准确至两位小数”指的是 0.50，而且你需要用上下界说明这个值不可能达到 0.505。

## STEP 之外

第 (i) 部分让你初尝模算术（modular arithmetic）和欧拉定理（Euler's theorem）的滋味：$3^4 \equiv 1 \pmod{10}$，是因为 $\varphi(10) = 4$。数论学家要找出像 $3^{3^3}$ 这类巨大幂塔（power tower）的末几位数字，靠的正是逐层重复运用这种化简。Davenport 的 *The Higher Arithmetic* 从零开始建立同余（congruence）和元素的阶（order of an element）的理论。

详细解答

# STEP I 1997，第 4 题

纯数学 · 绝对值方程

**STEP I 1997，第 4 题**

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

step.maths.org/questions/97-s1-q4

## 关键思路

每个绝对值（modulus）项只会在其内部为零之处改变公式，因此四个“临界点”（critical points）$x = -1, 0, 1, 2$ 把实数线分成五个区间（interval）。在每个区间上，每个绝对值都可以换成符号确定的普通括号，左边于是变成线性函数（linear function），方程便容易处理。唯一要留意的是：线性方程的根，必须落在你正在处理的那个区间内才算数。

## 解答

设 $L(x) = |x+1| - |x| + 3|x-1| - 2|x-2|$。回想当 $t \geq 0$ 时 $|t| = t$，而当 $t < 0$ 时 $|t| = -t$。我们由右至左逐个区间处理。

$x \geq 2$。四个绝对值的内部全部非负，所以

$$\begin{aligned} L(x) &= (x+1) - x + 3(x-1) - 2(x-2) \\ &= x + 1 - x + 3x - 3 - 2x + 4 = x + 2. \end{aligned}$$

方程变成 $x + 2 = x + 2$，对每一个 $x \geq 2$ 都成立。所以整个区间 $x \geq 2$ 都是解。

? **想一想** 方程变成了恒等式（identity）。我们是否在代数运算上出了错?

没有。当 $x \to \infty$ 时，各绝对值中 $x$ 的系数加起来是 $1 - 1 + 3 - 2 = 1$，常数项加起来是 $1 - 0 - 3 + 4 = 2$，恰好与右边的 $x + 2$ 吻合。含绝对值的方程完全可以有一整个区间的解，而题目说“求所有解”，正是暗示可能会出现这种情况。快速检验一下：在 $x = 3$ 时，$L(3) = 4 - 3 + 6 - 2 = 5 = 3 + 2$。

$1 \le x < 2$。 此时 $x-2<0$，所以 $|x-2| = -(x-2)$，而

$$L(x) = (x+1) - x + 3(x-1) + 2(x-2) = 5x - 6.$$

于是 $5x-6 = x+2$ 给出 $x=2$，它不在 $[1,2)$ 内。（它确是一个解，但我们已经把它计算在内。）没有新的解。

$0 \le x < 1$。 此时 $x-1<0$ 亦成立，所以

$$L(x) = (x+1) - x - 3(x-1) + 2(x-2) = -x.$$

于是 $-x = x+2$ 给出 $x=-1$，不在 $[0,1)$ 内。没有解。

$-1 \le x < 0$。 此时还有 $|x| = -x$，所以

$$L(x) = (x+1) + x - 3(x-1) + 2(x-2) = x.$$

方程 $x = x+2$ 根本没有解。

$x < -1$。 四个绝对值的内部全部为负，所以

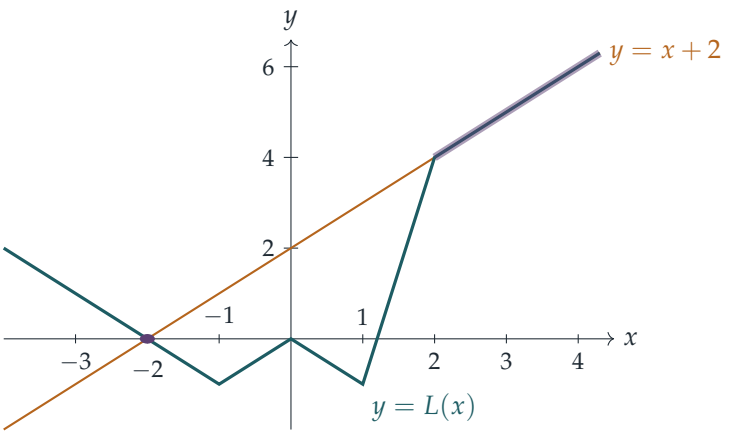
$$L(x) = -(x+1) + x - 3(x-1) + 2(x-2) = -x - 2.$$

于是 $-x-2 = x+2$ 给出 $x=-2$, 而 $-2<-1$, 所以这是一个真正的解。验算：$L(-2) = 1-2+9-8 = 0 = -2+2$。

**结论**。 方程的解是

$$x = -2 \qquad \text{以及所有} x \ge 2.$$

图像显示了原因：$y = L(x)$（蓝绿色）是由线段组成的折线，它与直线 $y = x+2$（琥珀色）在 $x=-2$ 处相交一次，然后由 $x=2$ 起与该直线重合。



## 验算

代入几个数值：$L(-2) = 0 = -2 + 2$，$L(5) = 6 - 5 + 12 - 6 = 7 = 5 + 2$，而在 $x = 0$ 时得 $L(0) = 1 + 3 - 4 = 0 \neq 2$。以电脑用精确分数运算，在 100,000 个随机点上确认了 $L$ 的五条线性公式；并发现在 $[-20, 20]$ 上步长为 0.01 的格点中，方程恰好在 $x = -2$ 以及所有 $x \geq 2$ 的格点上成立，其余各处均不成立。

## 常见错误

- 在 $x \geq 2$ 时得到 $x + 2 = x + 2$，却写“无解”或“有无穷多个解，所以一定是哪里出了错”，而没有看出这是一整个区间的解。
- 接受了像 $x = -1$（来自 $0 \leq x < 1$）这类落在所考虑区间之外的根。
- 遗漏了最外面的区域 $x < -1$，而孤立解 $x = -2$ 正正就在那里。

## STEP 之外

像 $L$ 这样的函数是*分段线性*（piecewise linear）的：它连续（continuous），并且在有限多个“角点”之间是线性的。这类函数是线性规划（linear programming）和最优化（optimisation）的核心，在那里，求绝对值之和 $\sum |x - a_i|$ 的最小值会引出中位数（median）而非平均数（mean）。它们也是“连续但在孤立点上不可微（not differentiable）”的函数的第一个例子，这个主题在 Abbott 的 *Understanding Analysis*（分析学入门教材）中有仔细的阐述。

详细解答

# STEP I 1997，第 5 题

纯数学 · 几何中的向量

**STEP I 1997，第 5 题**

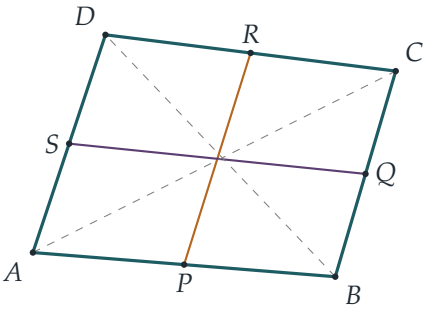
继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/97-s1-q5

## 关键思路

长度的平方就是点积（dot product），中点（mid-point）就是平均，所以用位置向量（position vector）便可把整条恒等式化为代数运算。有用的窍门是看出等式两边其实都暗藏着对角线（diagonal）$AC$ 和 $BD$：只要把每个差都用 $\overrightarrow{AC}\cdot\overrightarrow{BD}$ 表示出来，恒等式便水到渠成。至于其后的推论，只需用到四根杆的长度永远不变这一事实。

## 解答




设 $A, B, C, D$ 相对于任意原点的位置向量为 $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$，并记 $|\mathbf{x}|^2 = \mathbf{x}\cdot\mathbf{x}$。各中点为

$$\mathbf{p} = \tfrac{1}{2}(\mathbf{a}+\mathbf{b}), \quad \mathbf{q} = \tfrac{1}{2}(\mathbf{b}+\mathbf{c}), \quad \mathbf{r} = \tfrac{1}{2}(\mathbf{c}+\mathbf{d}), \quad \mathbf{s} = \tfrac{1}{2}(\mathbf{d}+\mathbf{a}).$$

### 四条边

把每一项按 $|\mathbf{y}-\mathbf{x}|^2 = |\mathbf{x}|^2 + |\mathbf{y}|^2 - 2\,\mathbf{x}\cdot\mathbf{y}$ 展开，平方项 $|\mathbf{a}|^2, \ldots, |\mathbf{d}|^2$ 在两边各出现一次，互相抵消，余下

$$\begin{aligned}|AB|^2 + |CD|^2 - |AD|^2 - |BC|^2 &= -2\,\mathbf{a}\cdot\mathbf{b} - 2\,\mathbf{c}\cdot\mathbf{d} + 2\,\mathbf{a}\cdot\mathbf{d} + 2\,\mathbf{b}\cdot\mathbf{c}\\ &= 2(\mathbf{a}-\mathbf{c})\cdot(\mathbf{d}-\mathbf{b}).\end{aligned}$$

（把 $(\mathbf{a}-\mathbf{c})\cdot(\mathbf{d}-\mathbf{b})$ 乘开，即可核实这个因式分解。）用文字来说，这就是 $-2\,\overrightarrow{AC}\cdot\overrightarrow{BD}$。

### 两条中点连线

我们有

$$\overrightarrow{PR} = \tfrac{1}{2}(\mathbf{c}+\mathbf{d}-\mathbf{a}-\mathbf{b}) = \tfrac{1}{2}\mathbf{u}, \qquad \overrightarrow{QS} = \tfrac{1}{2}(\mathbf{d}+\mathbf{a}-\mathbf{b}-\mathbf{c}) = \tfrac{1}{2}\mathbf{v},$$

姑且这样记。

> ? **想一想** 怎样才可以求出 $|\mathbf{u}|^2 - |\mathbf{v}|^2$，而不必展开两个四项式的平方？

利用向量的平方差公式（difference of two squares）：$|\mathbf{u}|^2 - |\mathbf{v}|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})$，此式成立是因为 $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$。这里

$$\mathbf{u}+\mathbf{v} = 2(\mathbf{d}-\mathbf{b}), \qquad \mathbf{u}-\mathbf{v} = 2(\mathbf{c}-\mathbf{a}),$$

所以

$$|PR|^2 - |QS|^2 = \tfrac{1}{4}\left(|\mathbf{u}|^2 - |\mathbf{v}|^2\right) = (\mathbf{c}-\mathbf{a})\cdot(\mathbf{d}-\mathbf{b}) = \overrightarrow{AC}\cdot\overrightarrow{BD}.$$

### 合并结果

把四条边的结果加上中点连线结果的两倍，

$$\begin{aligned}&|AB|^2 + |CD|^2 - |AD|^2 - |BC|^2 + 2\left(|PR|^2 - |QS|^2\right)\\ &\quad = 2(\mathbf{a}-\mathbf{c})\cdot(\mathbf{d}-\mathbf{b}) + 2(\mathbf{c}-\mathbf{a})\cdot(\mathbf{d}-\mathbf{b}) = 0,\end{aligned}$$

移项后得

$$|AB|^2 + |CD|^2 + 2|PR|^2 = |AD|^2 + |BC|^2 + 2|QS|^2,$$

这正是所要证明的。我们从没有用过凸性（convexity），甚至没有用过各点共面这一条件，所以恒等式对任意四点都成立，包括折叠过来、以致其中两根杆相交的四边形。

### 推论

把恒等式移项，

$$|PR|^2 - |QS|^2 = \tfrac{1}{2}\left(|AD|^2 + |BC|^2 - |AB|^2 - |CD|^2\right).$$

各杆都是刚性的,所以无论接合处怎样屈曲,右边的四个长度永远不变。因此 $|PR|^2 - |QS|^2$ 在整个运动过程中是常数。换句话说，两条对角线的纯量积（scalar product）$\overrightarrow{AC} \cdot \overrightarrow{BD}$ 保持不变，尽管对角线本身的长度和方向都在改变。

## 验算

- 对于边长为 $a$ 和 $b$ 的长方形，$|AB| = |CD| = a$ 而 $|AD| = |BC| = b$，而 $PR$ 和 $QS$ 连接对边的中点，所以 $|PR| = b$ 及 $|QS| = a$。恒等式两边都等于 $2a^2 + 2b^2$。
- 对于菱形（rhombus），四条边全部相等，所以推论指出在任何位置都有 $|PR| = |QS|$。事实上，菱形的对角线互相垂直，所以 $\overrightarrow{AC} \cdot \overrightarrow{BD} = 0$。
- 数值上，在 $10^5$ 个随机四边形上，恒等式的误差都在 $10^{-12}$ 之内。对于一个 $|AB| = 3$、$|BC| = 4$、$|CD| = 5$、$|DA| = 6$ 的连杆机构（linkage），屈曲经过约 200 个位置，$|PR|^2 - |QS|^2$ 始终保持在 $\frac{1}{2}(36 + 16 - 9 - 25) = 9$。

## 常见错误

- 搞错哪些边与哪条连线配对：$PR$ 连接 $AB$ 和 $CD$ 的中点，而在恒等式中与 $PR$ 放在一起的正是这两条边。
- 采用坐标设置，例如把 $A$ 放在原点、$B$ 放在坐标轴上，却忘记了一般的四边形需要其余各点完全自由。用向量便可免去这些繁琐的处理。

## STEP 之外

$P, Q, R, S$ 四点总是构成一个平行四边形（瓦里尼翁定理，Varignon's theorem），而 $PR$、$QS$ 就是它的对角线。这里的恒等式与平行四边形恒等式（parallelogram law）$|\mathbf{u} + \mathbf{v}|^2 + |\mathbf{u} - \mathbf{v}|^2 = 2|\mathbf{u}|^2 + 2|\mathbf{v}|^2$ 是近亲；在大学一年级的线性代数课程中，后者恰好刻划了那些由内积（inner product）导出的范数（norm）。机构运动时某个量保持不变，这种量称为不变量（invariant），这个想法在数学中反复出现。Strang 的 *Introduction to Linear Algebra*（线性代数入门）是一个好去处，可以看到点积和内积如何得到严谨而完整的阐述。

详细解答

# STEP I 1997，第 6 题

纯数学 · 以积分估计 $\pi$ 的界限

> **STEP I 1997，第 6 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s1-q6

本题在 *Beyond STEP* 第 4.3 节22/7 大于 $\pi$ 中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch04/twenty-two-sevenths.html

## 关键思路

被积函数（integrand）是一个有理函数（rational function），其分子的次数高于分母，所以第一步是以 $x^2+1$ 作多项式除法（polynomial division）：得出一个多项式商式，积分起来毫不费力，再加上一个常数余数除以 $1+x^2$，积分后得到反正切（arctangent）。$\pi$ 正是由这个反正切而来。至于最后的不等式，则利用在 $[0,1]$ 上 $1 \le 1+x^2$，把被积函数与 $x^4(1-x)^4$ 本身作比较。

## 解答

### 除法

以二项式定理（binomial theorem）展开，

$$x^4(1-x)^4 = x^4(1-4x+6x^2-4x^3+x^4) = x^8-4x^7+6x^6-4x^5+x^4.$$

我们以 $x^2+1$ 作除式，每次消去首项：

$$\begin{aligned}
x^8-4x^7+6x^6-4x^5+x^4 &= x^6(x^2+1)+\left(-4x^7+5x^6-4x^5+x^4\right)\\
-4x^7+5x^6-4x^5+x^4 &= -4x^5(x^2+1)+\left(5x^6+x^4\right)\\
5x^6+x^4 &= 5x^4(x^2+1)+\left(-4x^4\right)\\
-4x^4 &= -4x^2(x^2+1)+4x^2\\
4x^2 &= 4(x^2+1)-4.
\end{aligned}$$

把 $x^2+1$ 的各个倍数合并，

$$x^4(1-x)^4=\left(x^6-4x^5+5x^4-4x^2+4\right)(x^2+1)-4,$$

所以

$$a_6=1,\quad a_5=-4,\quad a_4=5,\quad a_3=0,\quad a_2=-4,\quad a_1=0,\quad a_0=4,\quad b=-4.$$

快速验算一下：当 $x=0$ 时，右方是 $4\cdot1-4=0$；当 $x=1$ 时，右方是 $(1-4+5-4+4)\cdot2-4=0$，两次都与左方相符。

### 等于 $\frac{22}{7}-\pi$ 的积分

把恒等式两边除以 $1+x^2$（它永不为零），

$$\frac{x^4(1-x)^4}{1+x^2}=x^6-4x^5+5x^4-4x^2+4-\frac{4}{1+x^2}.$$

在 $[0,1]$ 上逐项积分，并利用 $\int\frac{1}{1+x^2}\,\mathrm{d}x=\arctan x$，

$$\begin{aligned}
\int_0^1\frac{x^4(1-x)^4}{1+x^2}\,\mathrm{d}x &= \left[\tfrac17x^7-\tfrac23x^6+x^5-\tfrac43x^3+4x\right]_0^1-4\Big[\arctan x\Big]_0^1\\
&= \tfrac17-\tfrac23+1-\tfrac43+4\;-\;4\cdot\tfrac{\pi}{4}\\
&= \tfrac17+3-\pi\\
&= \frac{22}{7}-\pi,
\end{aligned}$$

因为 $-\frac23-\frac43=-2$ 及 $\arctan1=\frac{\pi}{4}$。这就是所求的结果。

### $x^4(1-x)^4$ 的积分

利用开首的展开式，

$$\int_0^1x^4(1-x)^4\,\mathrm{d}x=\tfrac15-\tfrac46+\tfrac67-\tfrac48+\tfrac19.$$

通分至公分母 $2520=5\cdot7\cdot8\cdot9$，

$$\frac{504-1680+2160-1260+280}{2520}=\frac{4}{2520}=\frac{1}{630}.$$

### $\pi$ 的上下界

设 $f(x) = x^4(1-x)^4$，则当 $0 < x < 1$ 时 $f(x) > 0$。

上界（*upper bound*）。在 $(0,1)$ 上，被积函数 $f(x)/(1+x^2)$ 严格为正且连续（continuous），所以它在 $[0,1]$ 上的积分严格为正。因此 $\frac{22}{7} - \pi > 0$，即 $\pi < \frac{22}{7}$。

下界（*lower bound*）。当 $0 < x < 1$ 时，$1 + x^2 > 1$，而由于 $f(x) > 0$，

$$\frac{f(x)}{1+x^2} < f(x).$$

**想一想** 为甚么两个积分之间的不等式是严格的，而不仅是 $\leq$?

两者之差 $f(x) - \dfrac{f(x)}{1+x^2} = \dfrac{x^2 f(x)}{1+x^2}$ 在整个开区间 $(0,1)$ 上连续且严格为正，所以它在 $[0,1]$ 上的积分严格为正。（一个非负的连续函数，只要在某处为正，其积分便为正。）因此

$$\frac{22}{7} - \pi = \int_0^1 \frac{f(x)}{1+x^2}\,\mathrm{d}x < \int_0^1 f(x)\,\mathrm{d}x = \frac{1}{630},$$

移项后得 $\pi > \frac{22}{7} - \frac{1}{630}$。结合两个界，

$$\frac{22}{7} > \pi > \frac{22}{7} - \frac{1}{630},$$

这正是所要证明的。

## 验算

- 以数值计算，$\frac{22}{7} - \pi \approx 0.0012645$，而对第一个积分使用辛普森法则（Simpson's rule），所得的值准确至小数点后十二位都相同。
- $\frac{1}{630} \approx 0.0015873$，明显大于 $0.0012645$；上下界表示 $3.141270 < \pi < 3.142857$，而事实上 $\pi \approx 3.141593$。
- $\frac{1}{630}$ 这个值亦可由贝塔积分（beta integral）公式 $\int_0^1 x^m(1-x)^n\,\mathrm{d}x = \dfrac{m!\,n!}{(m+n+1)!}$ 得出：它给出 $\dfrac{4!\,4!}{9!} = \dfrac{576}{362880} = \dfrac{1}{630}$。

## 常见错误

- 在长除法中漏掉 $a_3$ 和 $a_1$ 这两个零系数：每一个幂次都要保留在原位，否则余数便会算错。
- 写成 $\int_0^1 \frac{4}{1+x^2}\,\mathrm{d}x = 4\ln 2$（与 $\int \frac{2x}{1+x^2}\,\mathrm{d}x$ 混淆）；正确答案是 $4\arctan 1 = \pi$。

- 在最后一步只证明了 $\leq$。题目要求的是严格不等式（strict inequality），所以必须说明为何两个积分不可能相等。

## STEP 之外

这个优美的积分在二十世纪才被发现，它一举说明了课本上的近似值 $\frac{22}{7}$ 高估了 $\pi$。同样的想法——把 $x^n(1-x)^n$ 乘以简单的函数后积分，再证明结果虽小但为正——正是证明 $\pi$ 和 $e$ 是无理数（irrational）的经典方法的核心：介乎 0 与 1 之间的整数并不存在。形如 $\int_0^1 x^m(1-x)^n \,\mathrm{d}x$ 的积分是贝塔函数（beta function）的值，在大学一年级的分析和概率课程中便会遇到。至于如何严谨地处理积分之间的严格不等式，可参阅 Abbott 的 *Understanding Analysis*。

详细解答

# STEP I 1997，第 7 题

纯数学 · 高斯求积法

**STEP I 1997，第 7 题**

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

step.maths.org/questions/97-s1-q7

## 关键思路

等式两边对 P 都是线性（linearly）的，所以只要公式对四个基本的三次多项式（cubic）1、$t$、$t^2$ 和 $t^3$ 成立，它就对所有三次多项式成立。这样一来，问题便化为含四个未知数的四条方程。这些方程虽然是非线性的，但奇次幂与偶次幂可以干净利落地分开，余下的工作就交给关于 $t = 0$ 的对称性（symmetry）。

## 解答

设 $\mathrm{P}(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3$。则

$$\int_{-1}^{1} \mathrm{P}(t)\,\mathrm{d}t = c_0 \int_{-1}^{1} 1\,\mathrm{d}t + c_1 \int_{-1}^{1} t\,\mathrm{d}t + c_2 \int_{-1}^{1} t^2\,\mathrm{d}t + c_3 \int_{-1}^{1} t^3\,\mathrm{d}t,$$

而 $a_1\mathrm{P}(u_1) + a_2\mathrm{P}(u_2)$ 也可以用完全相同的方式拆开。因此，公式对所有三次多项式成立，当且仅当它对 $\mathrm{P}(t) = t^k$，$k = 0, 1, 2, 3$ 成立。（“仅当”是因为每个 $t^k$ 本身都是广义的三次多项式；“当”则只需把各项加起来。）由于 $\int_{-1}^{1} t^k\,\mathrm{d}t$ 在 $k = 0, 1, 2, 3$ 时分别为 2, 0, $\frac{2}{3}$, 0，我们需要

$$a_1 + a_2 = 2, \tag{A}$$

$$a_1 u_1 + a_2 u_2 = 0, \tag{B}$$

$$a_1 u_1^2 + a_2 u_2^2 = \tfrac{2}{3}, \tag{C}$$

$$a_1 u_1^3 + a_2 u_2^3 = 0. \tag{D}$$

? **想一想** 这些方程是非线性的。有没有哪一个组合可以因式分解（factorise）？

取 (D) 减去 $u_2^2$ 乘以 (B)：含 $a_2$ 的项互相抵消，得

$$a_1u_1^3 - a_1u_1u_2^2 = a_1u_1(u_1-u_2)(u_1+u_2) = 0.$$

我们逐一排除前三个因式为零的情况。

- 若 $a_1 = 0$：由 (A) 得 $a_2 = 2$，由 (B) 得 $u_2 = 0$，于是 (C) 变成 $0 = \frac{2}{3}$。不可能。
- 若 $u_1 = 0$：由 (B) 得 $a_2u_2 = 0$，所以 $a_2u_2^2 = 0$，(C) 同样变成 $0 = \frac{2}{3}$。不可能。
- 若 $u_1 = u_2$：两点重合，(B) 变成 $(a_1+a_2)u_1 = 2u_1 = 0$，所以 $u_1 = u_2 = 0$，(C) 再次不成立。

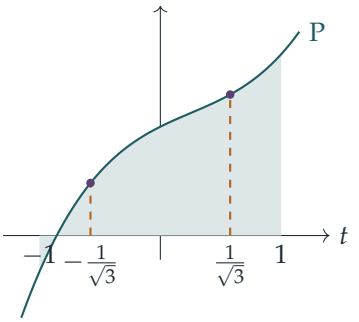
因此 $u_2 = -u_1$，且 $u_1 \neq 0$。这时 (B) 变成 $(a_1 - a_2)u_1 = 0$，所以 $a_1 = a_2$，再由 (A) 得 $a_1 = a_2 = 1$。最后由 (C) 得 $2u_1^2 = \frac{2}{3}$，即 $u_1^2 = \frac{1}{3}$；(D) 随之自动成立。所以

$$\boxed{a_1 = a_2 = 1, \qquad u_1 = -\tfrac{1}{\sqrt{3}}, \quad u_2 = \tfrac{1}{\sqrt{3}}}$$

（或把 $u_1$ 与 $u_2$ 对调），也就是说，

$$\int_{-1}^{1} \mathrm{P}(t)\,\mathrm{d}t = \mathrm{P}\Big(-\tfrac{1}{\sqrt{3}}\Big) + \mathrm{P}\Big(\tfrac{1}{\sqrt{3}}\Big) \quad \text{对每一个三次多项式P.}$$

以上论证同时说明，除了重新标记之外，这是唯一的选择。




无论画的是哪一个三次多项式，阴影部分的面积都等于两条虚线高度之和。

## 验算

- $\mathrm{P}(t) = t^2$：积分为 $\frac{2}{3}$，而公式给出 $\frac{1}{3} + \frac{1}{3}$。
- $\mathrm{P}(t) = 1 + t + t^2 + t^3$：积分为 $2 + \frac{2}{3} = \frac{8}{3}$；公式给出 $2 + 2\cdot\frac{1}{3} = \frac{8}{3}$，因为奇次项互相抵消。
- 这条公式对 $t^4$ 并不准确：它给出 $2\cdot\frac{1}{9} = \frac{2}{9}$，而积分却是 $\frac{2}{5}$。可见三次多项式确实是极限。
- 数值上，以 $10^4$ 个随机三次多项式测试，两边相差不超过 $10^{-14}$。

## 常见错误

- 一开始就假设对称（$u_2 = -u_1$，$a_1 = a_2$）。这样做可以很快得到正确答案，却没有证明其他选择都行不通；上面的因式分解则做到了这一点。
- 在除以某个量时，忘记考虑 $a_1 = 0$ 或 $u_1 = u_2$ 的情况。
- 只用 $\mathrm{P}(t) = 1$ 和 $t^2$ 来检验：迫使对称性出现的，正是奇次条件 (B) 和 (D)。

## STEP 之外

这就是两点高斯求积法（Gaussian quadrature）。只要适当选取 $n$ 个点和相应的权（weights），就可以准确地求出所有次数不超过 $2n-1$ 的多项式的积分，是 $n$ 个等距点所能做到的两倍。最佳的点是勒让德多项式（Legendre polynomials）的零点（这里是 $\frac{1}{2}(3t^2-1)$，其零点为 $\pm\frac{1}{\sqrt{3}}$），而这些多项式在 $[-1,1]$ 上是正交（orthogonal）的。正交多项式会在大学一年级的数值分析（numerical analysis）和求解微分方程（differential equations）时出现；Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教材）在级数解一章中会讨论勒让德方程。

详细解答

# STEP I 1997，第 8 题

纯数学 · 指数方程的根

> **STEP I 1997，第 8 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s1-q8

## 关键思路

方程 $x = a^x$ 的根必定是正数，因为 $a^x > 0$。因此我们可以取对数（logarithm），把问题转化为研究 $f(x) = \ln x - x \ln a$ 在 $x > 0$ 上的零点（zero）。当 $a > 1$ 时，这个函数先升后降，所以只有在其极大值（maximum）不小于 0 时才有零点。当 $0 < a < 1$ 时，它严格递增（strictly increasing），所以最多只能穿过零一次，问题就在于它究竟有没有穿过。

## 解答

### 化为 $f(x) = 0$

全题中 $a > 0$，所以 $a^x = e^{x \ln a}$ 对所有实数 $x$ 都有定义，而且是正数。若 $x \le 0$，则 $a^x > 0 \ge x$，所以没有满足 $x \le 0$ 的根。当 $x > 0$ 时，两边都是正数，而 $\ln$ 严格递增，故

$$x = a^x \iff \ln x = x \ln a \iff f(x) = 0, \qquad \text{其中} f(x) = \ln x - x \ln a.$$

所以 $x = a^x$ 的实根正好就是 $f$ 在 $(0, \infty)$ 上的零点。

### 当 $a > e^{1/e}$ 时没有根

这时 $a > e^{1/e} > 1$，所以 $\ln a > 0$。求导数（derivative），得

$$f'(x) = \frac{1}{x} - \ln a,$$

它在 $0 < x < 1/\ln a$ 时为正，在 $x > 1/\ln a$ 时为负。所以 $f$ 先递增后递减，它在 $(0, \infty)$ 上的极大值是

$$f\left(\frac{1}{\ln a}\right) = \ln\left(\frac{1}{\ln a}\right) - 1 = -\ln(\ln a) - 1.$$

现在由 $a > e^{1/e}$ 得 $\ln a > 1/e$，再取一次对数（两边都是正数），得 $\ln(\ln a) > \ln(1/e) = -1$。因此

$$f(x) \leq -\ln(\ln a) - 1 < 0 \quad \text{对所有} x > 0.$$

所以 $f$ 永不为零，方程 $x = a^x$ 没有实根，这正是所要证明的。

**想一想** 当 $a = e^{1/e}$ 时，以及当 $1 < a < e^{1/e}$ 时，情况究竟如何？

当 $a = e^{1/e}$ 时，极大值刚好是 0，在 $x = 1/\ln a = e$ 处取得，所以有一个（重）根 $x = e$：直线 $y = x$ 与 $y = a^x$ 在 $(e, e)$ 相切（tangent）。当 $1 < a < e^{1/e}$ 时，极大值为正，而在两端都有 $f \to -\infty$，所以有两个根。（当 $a = 1$ 时，唯一的根是 $x = 1$。）题目并没有要求讨论这些情况，但它们说明 $e^{1/e}$ 正好就是分界点。

### 当 $0 < a < 1$ 时恰有一个根

现在 $\ln a < 0$，所以

$$f'(x) = \frac{1}{x} - \ln a = \frac{1}{x} + |\ln a| > 0 \quad \text{对所有} x > 0,$$

故 $f$ 在 $(0, \infty)$ 上严格递增。因此 $f$ 最多只有一个零点。

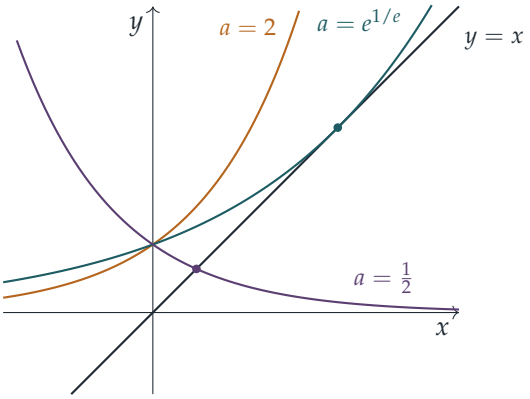
要说明它确实有一个零点，注意 $f$ 在 $(0, \infty)$ 上连续（continuous），而且

$$f(1) = -\ln a > 0, \qquad f(a) = \ln a - a \ln a = (1 - a) \ln a < 0,$$

两式都用到 $0 < a < 1$。根据介值定理（intermediate value theorem），$f$ 在 $a$ 与 1 之间有一个零点。

所以当 $0 < a < 1$ 时，方程 $x = a^x$ **恰有一个**实根，而且它位于区间 $(a, 1)$ 之内。

从图像看便一目了然：$y = a^x$ 递减而且恒为正，而 $y = x$ 递增，所以两个图像恰好相交一次。（等价地说，$g(x) = a^x - x$ 严格递减，且 $g(0) = 1 > 0$，$g(1) = a - 1 < 0$。）



## 验算

- 在分界处，$(e^{1/e})^e = e$，所以当 $a = e^{1/e} \approx 1.4447$ 时，$x = e$ 的确是一个根，与相切的情况吻合。
- 当 $a = \frac{1}{2}$ 时，用二分法（bisection）求得唯一的根 $x \approx 0.6412$，一如所料位于 $(\frac{1}{2}, 1)$ 之内。
- 数值上，随机选取 200 个 $a > e^{1/e}$，函数 $a^x - x$ 在 $[-20, 20]$ 的细密格点上始终为正；$\ln x - x \ln a$ 的极大值与 $-1 - \ln \ln a$ 吻合；当 $a$ 略小于 $e^{1/e}$ 时有两个根；而随机选取 500 个 $a \in (0, 1)$，每次都恰好只有一次变号。

## 常见错误

- 没有先指出任何根都满足 $x > 0$ 就取对数：否则 $\ln x$ 没有定义，所以 $x \leq 0$ 的情况需要另作（一行的）论证。
- 当 $0 < a < 1$ 时只证明 $f$ 递增。这只证明了最多有一个根；你还需要找到一次变号，才能证明确实有一个根。
- 对 $\ln a > 1/e$ 取 ln 时把不等号的方向弄反；这一步之所以稳妥，是因为 ln 递增，而且两边都是正数。

## STEP 之外

当 $0 < a < 1$ 时，那个根是映射 $x \mapsto a^x$ 的不动点（fixed point），而反复迭代 $x_{n+1} = a^{x_n}$ 就引出“无穷幂塔”（infinite power tower）$a^{a^{a^{\cdots}}}$。欧拉（Euler）证明了这个幂塔恰好在 $e^{-e} \leq a \leq e^{1/e}$ 时收敛（converge），所以同一个分界点 $e^{1/e}$ 再次出现。研究迭代何时收敛，会引向大学一年级分析学中的压缩映射定理（contraction mapping theorem）；Abbott 的 *Understanding Analysis*（分析学入门教材）涵盖了介值定理及所需的背景知识。

详细解答

# STEP I 1997，第 9 题

力学 · 相对速度

> **STEP I 1997，第 9 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s1-q9

## 关键思路

站在路边看，一切都在移动；但从车上看，一切都是静止的：汽车变成一排固定的方块，方块之间有长度为 $c$ 的空隙，而小鸡则以牠的相对速度（relative velocity）沿直线走过去。小鸡要顺利通过，唯一的条件是：当牠身处阔度为 $a$ 的行车线内时，牠（相对于汽车）的横向漂移不超过空隙长度 $c$。单凭这一个条件就能得出所求的不等式，而最小速率则来自求 $c\sin\theta + a\cos\theta$ 的最大值，这是标准的 $R\cos(\theta - \alpha)$ 做法。

## 解答

我们把小鸡看成质点（particle），把汽车看成以速率 $V$ 移动的长方形，每辆车都占据横跨道路、阔度为 $a$ 的同一条带状区域。取 $x$ 轴沿道路指向车流方向，$y$ 轴横跨道路，于是小鸡的速度是 $(u\cos\theta,\, u\sin\theta)$，其中 $0 < \theta < \pi$。只有汽车所在的那条带状区域才需要考虑：在道路的其他地方，小鸡毫无危险。

### 不等式

相对于汽车，小鸡的速度是

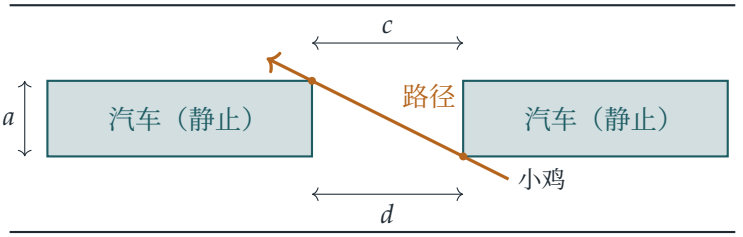
$$\mathbf{w} = (u\cos\theta - V,\ u\sin\theta).$$

在汽车的参考系（frame of reference）中，汽车静止不动，小鸡则以恒定速度 $\mathbf{w}$ 沿直线运动。牠横过阔度为 $a$ 的带状区域所需的时间是

$$\tau = \frac{a}{u\sin\theta}$$

而在这段时间内，牠（相对于汽车，逆着车流方向）向后滑移了

$$d = (V - u\cos\theta)\,\tau = \frac{a(V - u\cos\theta)}{u\sin\theta}.$$




图中所示的是极端情况 $d = c$，此时路径刚好沿着空隙的对角线。

**想一想** 为甚么漂移量 $d$ 最多只能是 $c$?

在汽车的参考系中，小鸡穿过带状区域的路径沿道路方向覆盖了长度为 $|d|$ 的一段，而这一段上的每一点，小鸡都会在行车线内的某个高度经过。静止的汽车之间只留下长度为 $c$ 的空隙，所以整段必须容纳在同一个空隙之内：$|d| \le c$。（反过来说，若 $|d| \le c$，小鸡可以选择适当的出发时刻，使这一段落在某个空隙之内，从而安全横过。）

特别地，$d \le c$，即

$$a(V - u\cos\theta) \le c\,u\sin\theta, \qquad \text{所以} \qquad Va \le u\,(c\sin\theta + a\cos\theta).$$

左边是正数，所以凡是安全的横过方式都有 $c\sin\theta + a\cos\theta > 0$，我们可以把它除过去：

$$u \ge \frac{Va}{c\sin\theta + a\cos\theta},$$

这正是所要证明的。（条件的另一半 $d \ge -c$，是防止跑得快的小鸡越过空隙而撞上前面的汽车；它在下面求最小值的过程中不起作用。）

### 最小速率与横过时间

写成 $c\sin\theta + a\cos\theta = \sqrt{a^2 + c^2}\,\cos(\theta - \alpha)$，其中 $\alpha$ 是满足以下条件的锐角：

$$\cos\alpha = \frac{a}{\sqrt{a^2 + c^2}}, \qquad \sin\alpha = \frac{c}{\sqrt{a^2 + c^2}}, \qquad \tan\alpha = \frac{c}{a}.$$

所以 $c\sin\theta + a\cos\theta \le \sqrt{a^2 + c^2}$，等号仅在 $\theta = \alpha$ 时成立。因此对每一个方向都有

$$u \ge \frac{Va}{c\sin\theta + a\cos\theta} \ge \frac{Va}{\sqrt{a^2 + c^2}},$$

而当 $\theta = \alpha$ 且 $u = Va/\sqrt{a^2 + c^2}$ 时，两个不等式都取等号。这个选择确实是安全的：它恰好给出 $d = c$（即图中由一角到对角的路径），而且 $-c \le d \le c$ 成立。所以最小可能的匀速率是 $Va/\sqrt{a^2 + c^2}$，方向角为 $\alpha = \arctan(c/a)$。

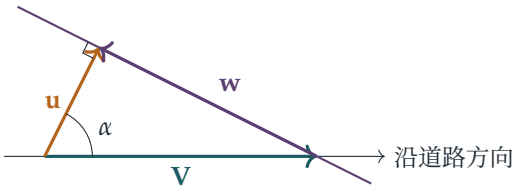
小鸡以横跨道路方向的速度分量 $u\sin\alpha$ 走过整个路阔 $b$，所以所需时间是

$$\frac{b}{u\sin\alpha} = b\cdot\frac{\sqrt{a^2+c^2}}{Va}\cdot\frac{\sqrt{a^2+c^2}}{c} = \frac{b\,(a^2+c^2)}{Vac} = \frac{b}{V}\left(\frac{c}{a}+\frac{a}{c}\right),$$

这就是所求的结果。

**想一想** 为甚么最小速率出现在 $\tan\theta = c/a$ 的时候?

看看各个速度。小鸡的速度是 $\mathbf{u} = \mathbf{V} + \mathbf{w}$，而在最理想的情况下，相对速度 $\mathbf{w}$ 沿着空隙的对角线，方向为 $(-c, a)$。所以 $\mathbf{u}$ 的末端落在经过 $\mathbf{V}$ 的末端、沿该方向的直线上，而最短的 $\mathbf{u}$ 就是由原点到这条直线的垂线。它的方向 $(a, c)$ 与道路成 $\arctan(c/a)$ 的角，而它的长度是 $V\sin(\text{对角线的倾角}) = Va/\sqrt{a^2+c^2}$。




## 验算

- 单位：$Va/\sqrt{a^2+c^2}$ 是速率，而 $\frac{b}{V}\left(\frac{c}{a}+\frac{a}{c}\right)$ 是长度除以速率，即时间。
- 极限情况：若 $\theta = \frac{\pi}{2}$（垂直横过），所得的界是 $u \geq Va/c$：小鸡必须在一个空隙经过所需的时间 $c/V$ 内走过阔度 $a$。这显然是对的。
- 数值：取 $a = 1$，$c = 2$，$V = 2$，$b = 5$，最小速率是 $2/\sqrt{5} \approx 0.894$，此时 $\tan\theta = 2$，而横过时间是 $\frac{5}{2}\left(2+\frac{1}{2}\right) = 6.25$。在道路参考系中直接模拟汽车和小鸡的运动，结果与此吻合：以这个速率的 1.01 倍能找到安全的出发时刻，以 0.99 倍则找不到。

## 常见错误

- 在安全条件中使用横过整条道路的时间 $b/(u\sin\theta)$。其实只有身处阔度为 $a$ 的行车线内的那段时间才有关系。
- 弄错相对速度的正负号：相对于汽车，小鸡沿道路方向的速度是 $u\cos\theta - V$，而不是 $u\cos\theta + V$。

## STEP 之外

转换到障碍物静止的参考系，是一个伽利略变换（Galilean transformation）；选取令问题变成静态的参考系，是力学中最有用的习惯之一。这里求最小值的过程也是一个小型的最优化（optimisation）问题：由一点到一条直线的最短向量就是垂线，你日后会在线性代数（linear algebra）中以正交投影（orthogonal projection）的形式再次遇上它。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学导论）中有很多这类关于相对运动的题目。

详细解答

# STEP I 1997，第 10 题

力学 · 圆周运动

> **STEP I 1997，第 10 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s1-q10

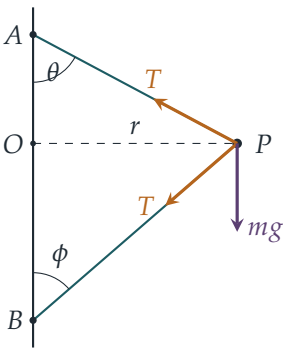
## 关键思路

由于小环是光滑 (smooth) 的, 绳子可以自由地穿过它滑动, 所以 $P$ 两边的张力 (tension) 是相同的。沿垂直方向分解 (没有垂直加速度), 再沿水平方向分解 (加速度为 $r\omega^2$, 指向转轴), 然后把一条方程除以另一条, 令 $T$ 和 $m$ 消去。接着利用和差化积公式 (sum-to-product formulae)，这个比值便化为单一个半角的正切。

## 解答

### 模型与受力

我们把 $P$ 视为一个质点（particle），系于一条轻而不可伸长（light inextensible）的绳子上；小环是光滑的，所以 $PA$ 和 $PB$ 中的张力同为 $T$。设 $O$ 为圆心，它在直线 $AB$ 上，且 $OP = r$。作用于 $P$ 的力有：它的重量 $mg$、沿 $PA$ 的张力 $T$ 和沿 $PB$ 的张力 $T$。




绳段 $PA$ 与垂直线成角 $\theta$（即在 $A$ 点 $AB$ 与 $AP$ 之间的夹角），所以它对 $P$ 的拉力有向上

的分量 $T\cos\theta$ 和指向转轴的分量 $T\sin\theta$。同样，$PB$ 在 $B$ 点与垂直线成角 $\phi$，它对 $P$ 的拉力有向下的分量 $T\cos\phi$ 和指向转轴的分量 $T\sin\phi$。

### 运动方程

$P$ 以恒定速率沿水平圆周运动，所以它没有垂直加速度，而有一个指向 $O$、大小为 $r\omega^2$ 的加速度。分解各力，得

$$\begin{aligned} &\text{垂直方向：} && T\cos\theta - T\cos\phi = mg, \\ &\text{水平方向：} && T\sin\theta + T\sin\phi = mr\omega^2. \end{aligned}$$

无论三角形的形状如何，这两条方程都成立（例如，假如 $P$ 在 $A$ 的上方，$\cos\theta$ 只不过变成负数而已）。由于 $mg > 0$，第一条方程给出 $\cos\theta > \cos\phi$，所以 $\phi > \theta$：下方的绳段与转轴的倾斜角较大，正如图中所示。

### 证明结果

把垂直方向的方程除以水平方向的方程，

$$\frac{\cos\theta - \cos\phi}{\sin\theta + \sin\phi} = \frac{g}{r\omega^2}.$$

由和差化积公式，

$$\begin{aligned} \cos\theta - \cos\phi &= 2\sin\tfrac{1}{2}(\theta+\phi)\,\sin\tfrac{1}{2}(\phi-\theta), \\ \sin\theta + \sin\phi &= 2\sin\tfrac{1}{2}(\theta+\phi)\,\cos\tfrac{1}{2}(\phi-\theta). \end{aligned}$$

由于 $0 < \theta < \pi$ 及 $0 < \phi < \pi$（它们是一个真正的三角形 $APB$ 的内角），我们有 $0 < \frac{1}{2}(\theta+\phi) < \pi$，所以 $\sin\frac{1}{2}(\theta+\phi) \neq 0$，可以约去：

$$\tan\frac{\phi-\theta}{2} = \frac{g}{r\omega^2},$$

这正是所要证明的结果。

### 张力

> **想一想** 答案可以含有 $\sin\frac{1}{2}(\theta+\phi)$，但不可含有 $\phi-\theta$。我们怎样才能消去 $\cos\frac{1}{2}(\phi-\theta)$?

由水平方向的方程和第二条和差化积公式，

$$2T\sin\tfrac{1}{2}(\theta+\phi)\cos\tfrac{1}{2}(\phi-\theta) = mr\omega^2.$$

我们已知 $\alpha = \frac{1}{2}(\phi - \theta)$ 的正切，所以可以借助一个对边为 $g$、邻边为 $r\omega^2$ 的直角三角形求出它的余弦。我们有 $0 < \phi - \theta < \pi$，所以 $0 < \alpha < \frac{\pi}{2}$，从而 $\cos\alpha > 0$：

$$\cos\tfrac{1}{2}(\phi - \theta) = \frac{r\omega^2}{\sqrt{r^2\omega^4 + g^2}}.$$

代入并约去 $r\omega^2$，

$$T = \frac{m\sqrt{r^2\omega^4 + g^2}}{2\sin\frac{1}{2}(\theta + \phi)}.$$

这个结果有一个漂亮的解释：$\sqrt{(mg)^2 + (mr\omega^2)^2}$ 是两个张力的合力（resultant）的大小，而两个大小同为 $T$、彼此夹角为 $\pi - (\theta + \phi)$（即角 $APB$）的力，其合力的大小为 $2T\sin\frac{1}{2}(\theta + \phi)$。

## 验算

- 当 $\omega \to \infty$ 时，右边趋向 0，所以 $\phi \to \theta$。小环被甩向外，直至与 $AB$ 的中点处于同一水平，这正是重力可以忽略时我们预期的情况。
- 当 $\omega \to 0$ 时，$\tan\frac{1}{2}(\phi - \theta) \to \infty$，所以 $\phi - \theta \to \pi$：绳子笔直下垂，这是没有旋转时唯一可能的情况。
- 数值方面，我们随机选取 $A$、$B$ 和 $P$ 的位置，直接由向量形式的运动方程求出 $T$ 和 $\omega^2$，再由坐标量度 $\theta$ 和 $\phi$，然后与两个结果比较：在约 8000 个个案中，最大的相对误差为 $3 \times 10^{-12}$。

## 常见错误

- 使用两个不同的张力 $T_1$、$T_2$。小环光滑，就只有一条绳子和一个张力；若用两个不同的张力，问题便无法确定。
- 把 $PB$ 中张力的垂直分量的正负号弄错：这段绳子是把 $P$ 向下拉向 $B$ 的。
- 为 $\cos\frac{1}{2}(\phi - \theta)$ 取负平方根，这样会得出负的张力。

## STEP 之外

对于给定的绳长，小环位于以 $A$ 和 $B$ 为焦点（foci）的椭圆（ellipse）上，而绳子与 $P$ 点的切线成相等的角：这就是椭圆的反射性质（reflection property）。因此两个张力的合力沿椭圆的法线（normal）方向，就像一条弯成这个形状的光滑铁线所施的反作用力一样。把约束（constraints）看成垂直于曲面的力，正是拉格朗日力学（Lagrangian mechanics）

的起点。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）收录了许多这类旋转系统的题目。

详细解答

# STEP I 1997，第 11 题

力学 · 有阻力的运动

**STEP I 1997，第 11 题**

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

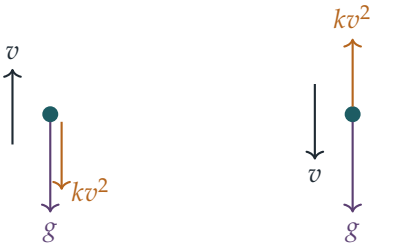
step.maths.org/questions/97-s1-q11

## 关键思路

阻力取决于速度，但我们想要的是速度作为距离的函数，而不是时间的函数。因此我们把加速度写成 $v\frac{\mathrm{d}v}{\mathrm{d}s} = \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{1}{2}v^2\right)$。以 $w = v^2$ 为未知量，运动方程（equation of motion）便成为一条线性一阶方程，用积分因子（integrating factor）即可一举解出。最后一部分的关键在于一个物理事实：阻力永远与运动方向相反，所以质点掉头时，阻力的方向也随之改变。

## 解答

我们把质点（particle）模拟为一个单位质量的点，沿铅垂线运动，受其重量 $g$（向下）以及大小为 $kv^2$、方向与速度相反的阻力（resistance）作用。由抛射点起向上量度 $s$。




上升时　　下降时

### 上升运动

质点上升时,其速度 $v$ 向上,所以阻力 $kv^2$ 与重量一同向下作用。由牛顿第二定律(Newton's second law)（质量为 1）得

$$v\frac{\mathrm{d}v}{\mathrm{d}s} = -g - kv^2,$$

这里用到 $\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t} = v\frac{\mathrm{d}v}{\mathrm{d}s}$。设 $w = v^2$，则 $\frac{\mathrm{d}w}{\mathrm{d}s} = 2v\frac{\mathrm{d}v}{\mathrm{d}s}$:

$$\frac{\mathrm{d}w}{\mathrm{d}s} + 2kw = -2g.$$

这条方程对 $w$ 而言是线性的。乘以积分因子 $\mathrm{e}^{2ks}$,

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(w\,\mathrm{e}^{2ks}\right) = -2g\,\mathrm{e}^{2ks} \quad\Longrightarrow\quad w\,\mathrm{e}^{2ks} = -\frac{g}{k}\,\mathrm{e}^{2ks} + C.$$

当 $s = 0$ 时，$w = u^2$，所以 $C = u^2 + \frac{g}{k}$。因此

$$v^2 = \left(u^2 + \frac{g}{k}\right)\mathrm{e}^{-2ks} - \frac{g}{k} = u^2\mathrm{e}^{-2ks} + \frac{g}{k}\left(\mathrm{e}^{-2ks} - 1\right),$$

这正是所求的 $(*)$。（在 $v\,\mathrm{d}v/(g + kv^2) = -\mathrm{d}s$ 中分离变量（separating variables），得出一个对数，同样可行。）

### 最大高度

当 $v = 0$ 时质点停止上升；由于 $(*)$ 的右方随 $s$ 严格递减，这种情况恰好出现一次，就在最大高度 $H$ 处。在 $(*)$ 中令 $v = 0$:

$$\left(u^2 + \frac{g}{k}\right)\mathrm{e}^{-2kH} = \frac{g}{k} \quad\Longrightarrow\quad \mathrm{e}^{2kH} = 1 + \frac{ku^2}{g},$$

所以

$$H = \frac{1}{2k}\ln\left(1 + \frac{ku^2}{g}\right).$$

### 下降路径

> ? **想一想** 推导 $(*)$ 的过程从没有提及运动方向。那么它为甚么不会继续成立?

其实推导确实用到了方向，就在阻力的正负号上。下降时速度向下，所以阻力 $kv^2$ 此时向上作用，与重量相反。仍然以向上量度 $s$，并以 $v$ 表示向上的速度，方程变成

$$v\frac{\mathrm{d}v}{\mathrm{d}s} = -g + kv^2, \qquad \text{所以} \qquad \frac{\mathrm{d}w}{\mathrm{d}s} - 2kw = -2g,$$

这是另一条微分方程（differential equation）。所以 $(*)$ 在下降路径上并不成立。

为了具体看清这一点，我们以 $s = H$ 时 $w = 0$ 为条件解这条新方程。由积分因子 $\mathrm{e}^{-2ks}$ 得 $w = \frac{g}{k} + D\mathrm{e}^{2ks}$，而 $w(H) = 0$ 则迫使 $D = -\frac{g}{k}\mathrm{e}^{-2kH}$：

$$v^2 = \frac{g}{k}\left(1 - \mathrm{e}^{-2k(H-s)}\right) \qquad (0 \le s \le H).$$

利用 $u^2 + \frac{g}{k} = \frac{g}{k}\mathrm{e}^{2kH}$，方程 $(*)$ 可以写成 $v^2 = \frac{g}{k}\left(\mathrm{e}^{2k(H-s)} - 1\right)$。设 $x = 2k(H-s) \ge 0$，这两个式子分别是 $\frac{g}{k}(1 - \mathrm{e}^{-x})$ 和 $\frac{g}{k}(\mathrm{e}^{x} - 1)$，而

$$(\mathrm{e}^{x} - 1) - (1 - \mathrm{e}^{-x}) = \mathrm{e}^{x} + \mathrm{e}^{-x} - 2 = \left(\mathrm{e}^{x/2} - \mathrm{e}^{-x/2}\right)^2 > 0 \quad \text{当} x > 0.$$

所以在最高点以下的每一点，$(*)$ 给出的速率都是错的（两者只在 $s = H$ 处相符，该处两者皆为零）；它高估了下降时的速率。

还有一个简捷的物理论证。方程 $(*)$ 把 $v^2$ 表示为只与高度有关的函数，所以假如它在下降时也成立，质点经过每个高度时的速率便会与上升时相同，并以速率 $u$ 返回起点。但阻力在整个运动过程中都作负功，所以质点损失了能量，返回时必定较慢。事实上，在 $s = 0$ 处，正确的公式给出

$$v^2 = \frac{g}{k}\left(1 - \frac{1}{1 + ku^2/g}\right) = \frac{u^2}{1 + ku^2/g} < u^2.$$

## 验算

- 当 $k \to 0$ 时，$\ln(1 + ku^2/g) \approx ku^2/g$，所以 $H \to u^2/(2g)$，即大家熟悉的无阻力高度；而 $(*)$ 则趋向 $v^2 = u^2 - 2gs$（展开 $\mathrm{e}^{-2ks} \approx 1 - 2ks$）。
- 数值上，取 $u = 10$、$k = 0.05$、$g = 9.8$，对运动方程逐步计算得 $H \approx 4.1224$，与 $\frac{1}{0.1}\ln(1 + \frac{5}{9.8}) = 4.12245$ 吻合；质点返回时 $v^2 \approx 66.22$，与 $u^2/(1 + ku^2/g)$ 吻合，而不是 $u^2 = 100$。下降至一半时，$v^2 \approx 36.51$，而 $(*)$ 则会给出 44.87。

## 常见错误

- 用单一方程描述整个运动，并把阻力写成 $-kv^2$：由于 $v^2 \ge 0$，这一项永远指向同一方向，这正是一条公式无法同时涵盖上升和下降两段的原因。（若要用单一方程，便需要 $-kv|v|$。）
- 使用 $\frac{\mathrm{d}v}{\mathrm{d}t}$ 并解出以 $t$ 表示的 $v$，然后苦于消去 $t$；用 $v\frac{\mathrm{d}v}{\mathrm{d}s}$ 则可直达答案。

## STEP 之外

对于在空气中高速运动的球，二次阻力（quadratic drag）是一个很好的模型，而下降方程显示出*终端速度*（terminal velocity）：随着下落距离增加，$v^2 \to g/k$。把阻力写成

$-kv|v|$, 便得到一条描述整个运动的非线性方程; 大学一年级力学有很大部分就是研究这类方程, 其中力取决于速度, 而能量并不守恒。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）对取决于速度的力和终端速度有细致的论述。

详细解答

# STEP I 1997，第 12 题

概率与统计 · 中心极限定理

**STEP I 1997，第 12 题**

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

step.maths.org/questions/97-s1-q12

## 关键思路

直接解出这个二次方程（quadratic equation），令 $X$ 成为 $T$ 的一个简单函数。这样，每个关于 $X$ 的问题都变成关于 $T$ 的问题，而后者的分布（distribution）我们完全清楚。一个巧妙的简化是：$1-T$ 同样在 $[0,1]$ 上服从均匀分布（uniform distribution）。最后一部分涉及大量独立（independent）副本之和，这正是提示我们使用中心极限定理（central limit theorem）。

## 解答

### $X$ 的分布

由二次公式（quadratic formula），$x^2+2x+T=0$ 的根为 $x=-1\pm\sqrt{1-T}$。由于 $0\le T\le 1$，这两个根都是实数，而较大的根取正号：

$$X=-1+\sqrt{1-T}.$$

当 $T$ 取遍 $[0,1]$ 时，$X$ 取遍 $[-1,0]$。

设 $U=1-T$。由于 $T$ 在 $[0,1]$ 上服从均匀分布，$U$ 亦然：把区间绕其中点反射，并不会改变均匀的密度（density）。所以 $X=\sqrt{U}-1$，其中 $U\sim \mathrm{U}[0,1]$。

### $X>-\frac{1}{3}$ 的概率

由于 $X+1=\sqrt{U}$，而且两边皆为非负，

$$X>-\tfrac{1}{3}\iff\sqrt{U}>\tfrac{2}{3}\iff U>\tfrac{4}{9}.$$

因为两边皆为非负，平方后不等式的方向保持不变。因此

$$\mathrm{P}\left(X > -\tfrac{1}{3}\right) = \mathrm{P}\left(U > \tfrac{4}{9}\right) = 1 - \tfrac{4}{9} = \frac{5}{9}.$$

（等价地，$T < \frac{5}{9}$。）

### 平均值与方差

对于 $U \sim \mathrm{U}[0,1]$，其密度在 $[0,1]$ 上为 1，所以

$$\mathrm{E}(\sqrt{U}) = \int_0^1 u^{1/2}\,\mathrm{d}u = \frac{2}{3}, \qquad \mathrm{E}\big((\sqrt{U})^2\big) = \mathrm{E}(U) = \frac{1}{2}.$$

因此

$$\mathrm{E}(X) = \mathrm{E}(\sqrt{U}) - 1 = \frac{2}{3} - 1 = -\frac{1}{3}.$$

减去一个常数并不会改变方差（variance），所以

$$\mathrm{Var}(X) = \mathrm{Var}(\sqrt{U}) = \mathrm{E}(U) - \big(\mathrm{E}\sqrt{U}\big)^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18},$$

正如所求。

> **想一想** 要求出这些矩（moments），我们需要 $X$ 的概率密度函数（probability density function）吗？

不需要。对任意函数 $g$，$\mathrm{E}\big(g(U)\big) = \int_0^1 g(u)\,\mathrm{d}u$，所以我们可以直接用 $U$ 来计算。如果你仍想知道密度：当 $-1 \le x \le 0$ 时，$\mathrm{P}(X \le x) = \mathrm{P}\big(U \le (x+1)^2\big) = (x+1)^2$，所以 $X$ 在 $[-1,0]$ 上的密度为 $2(x+1)$。把 $x \cdot 2(x+1)$ 在该区间上积分，同样得到 $-\frac{1}{3}$。

### 800 个根之和

由于各次实验互相独立，各个 $X_i$ 亦互相独立，而且每个的平均值（mean）为 $-\frac{1}{3}$，方差为 $\frac{1}{18}$。所以

$$\mathrm{E}(Y) = 800 \cdot \left(-\tfrac{1}{3}\right) = -\frac{800}{3}, \qquad \mathrm{Var}(Y) = 800 \cdot \frac{1}{18} = \frac{400}{9},$$

其中把方差相加时用到了独立性。$Y$ 的标准差（standard deviation）为 $\frac{20}{3}$。

由于 800 很大，中心极限定理指出 $Y$ 近似服从 $\mathrm{N}\left(-\frac{800}{3}, \frac{400}{9}\right)$。我们要求

$$\mathrm{P}(Y \le K) \approx \Phi\left(\frac{K + \frac{800}{3}}{20/3}\right) = 0.08.$$

查表得 $\Phi(1.405) \approx 0.92$，所以由对称性 $\Phi(-1.405) \approx 0.08$。因此

$$\frac{K + \frac{800}{3}}{20/3} \approx -1.405,$$

$$K \approx -\frac{800}{3} - 1.405 \times \frac{20}{3} \approx -266.67 - 9.37 \approx -276.$$

**想一想** 这里不是应该作连续性校正（continuity correction）吗?

不需要。每个 $X_i$ 都是连续随机变量（continuous random variable），所以 $Y$ 也是，因而无须作连续性校正。这种校正只用于以正态分布（normal distribution）近似离散分布（discrete distribution），例如二项分布（binomial distribution）。

## 验算

- 根是正确的：当 $T = 0$ 时，方程为 $x(x+2) = 0$，其较大的根为 $0 = -1 + \sqrt{1}$。
- $\mathrm{E}(X) = -\frac{1}{3}$ 位于 $[-1, 0]$ 之内，而密度 $2(x+1)$ 偏向 0 的一方，所以平均值高于中点 $-\frac{1}{2}$ 是合理的。
- 以 400 000 个 $T$ 值进行模拟(simulation),得到 $\mathrm{P}(X > -\frac{1}{3}) \approx 0.555$,平均值 $\approx -0.3337$,方差 $\approx 0.0556$。直接模拟 $Y$ 共 4000 次，得到 $\mathrm{P}(Y \leq -276.03) \approx 0.083$，与 0.08 吻合。

## 常见错误

- 在二次公式中取了负号。这样得到的是较小的根。
- 写成 $\mathrm{Var}(Y) = 800^2\,\mathrm{Var}(X)$。这是 $800X_1$ 的方差，而不是 800 个独立副本之和的方差。
- 用了 $+1.405$ 而不是 $-1.405$。概率 0.08 位于下尾，所以 $K$ 必定低于平均值。

## STEP 之外

把 $X = g(U)$ 写成均匀变量 $U$ 的函数，这个技巧也可以反过来用。如果 $F$ 是任意一个连续的分布函数（distribution function），那么 $F^{-1}(U)$ 的分布函数就是 $F$。这个*逆变换法*（*inverse transform method*）正是电脑利用均匀随机数模拟随机变量的方法。中心极限定理在这里被当作黑箱使用，而在大学的第一个概率课程中，会用矩母函数（moment generating function）或特征函数（characteristic function）来证明它。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率入门教材）涵盖了这两个概念。

详细解答

# STEP I 1997，第 13 题

概率与统计 · 条件概率

**STEP I 1997，第 13 题**

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

step.maths.org/questions/97-s1-q13

## 关键思路

Blond 先生应否开灯，只取决于等着他的是哪一种陷阱，所以整道题归结为一个条件概率（conditional probability）：在没有金丝雀的情况下，陷阱是森蚺而不是炸弹的机会。刺客是谁之所以重要，只因为这会影响所用的武器。因此，我们用全概率公式（law of total probability）把两层（先是刺客，然后是武器）结合起来，以“没有金丝雀”为条件，再比较两种做法的死亡风险。

## 解答

### 设置

设 $M_i$ 为 $i$ 先生布下陷阱的事件，其中 $i = 1, 2, 3, 4$，并设 $A$、$B$ 和 $C$ 分别为陷阱是森蚺、炸弹或金丝雀的事件。已知

$$P(M_i) = \frac{i}{10}, \qquad P(A \mid M_i) = \frac{i+1}{10}, \quad P(B \mid M_i) = \frac{i}{10}, \quad P(C \mid M_i) = \frac{9-2i}{10}.$$

这些数值互相吻合：$\frac{1+2+3+4}{10} = 1$，而且对每个 $i$，三个武器概率均为非负，其和为 $\frac{(i+1)+i+(9-2i)}{10} = 1$。牵涉其中的刺客恰好一名，武器亦恰好一种，所以事件 $M_i \cap A$、$M_i \cap B$ 和 $M_i \cap C$ 构成所有可能情况的一个划分（partition）。

### 寂静告诉他甚么

根据全概率公式，

$$P(A)=\sum_{i=1}^{4}P(M_i)P(A\mid M_i)=\frac{1\cdot 2+2\cdot 3+3\cdot 4+4\cdot 5}{100}=\frac{40}{100},$$
$$P(B)=\sum_{i=1}^{4}P(M_i)P(B\mid M_i)=\frac{1+4+9+16}{100}=\frac{30}{100}.$$

所以 $P(C')=P(A)+P(B)=\frac{70}{100}$。（验算一下：$P(C)=\frac{7+10+9+4}{100}=\frac{30}{100}$。）

**想一想** 我们需要找出现在哪一名刺客的可能性最大吗?

不需要。听不到歌声确实会改变每名刺客的可能性，但死亡风险只取决于武器。由于 $A\subseteq C'$ 及 $B\subseteq C'$，以 $C'$ 为条件只是把这两个概率按比例重新缩放：

$$P(A\mid C')=\frac{P(A)}{P(C')}=\frac{40}{70}=\frac{4}{7},\qquad P(B\mid C')=\frac{P(B)}{P(C')}=\frac{30}{70}=\frac{3}{7}.$$

有关刺客的信息已经包含在这两个数之中。

### 比较两个选择

设 $K$ 为 Blond 先生被杀的事件。一旦以 $C'$ 为条件，陷阱不是森蚺就是炸弹，而他的命运只取决于陷阱和他的选择。

*开灯*。炸弹有 $\frac{1}{2}$ 的概率杀死他，森蚺则有 $1-\frac{2}{3}=\frac{1}{3}$ 的概率杀死他。所以

$$P(K\mid C',\text{开灯})=\frac{3}{7}\cdot\frac{1}{2}+\frac{4}{7}\cdot\frac{1}{3}=\frac{9}{42}+\frac{8}{42}=\frac{17}{42}.$$

*不开灯*。炸弹必定不会杀死他，森蚺则有 $\frac{1}{2}$ 的概率杀死他。所以

$$P(K\mid C',\text{不开灯})=\frac{3}{7}\cdot 0+\frac{4}{7}\cdot\frac{1}{2}=\frac{2}{7}=\frac{12}{42}.$$

由于 $\frac{12}{42}<\frac{17}{42}$，**Blond 先生应该不开灯便进入寓所**。这样他的生还机会是 $\frac{5}{7}$（约 71%），而开灯的话则是 $\frac{25}{42}$（约 60%）。

原因在于：森蚺虽然是较有可能出现的陷阱，但开灯真正令情况恶化的却是炸弹。开灯把森蚺的风险由 $\frac{1}{2}$ 降至 $\frac{1}{3}$，却把炸弹的风险由 0 升至 $\frac{1}{2}$。陷阱是森蚺的概率为 $\frac{4}{7}$，开灯可减少 $\frac{4}{7}\cdot\frac{1}{6}=\frac{4}{42}$ 的风险；陷阱是炸弹的概率为 $\frac{3}{7}$，开灯却要付出 $\frac{3}{7}\cdot\frac{1}{2}=\frac{9}{42}$ 的代价。代价较大。

## 验算

- 在他细听之前，开灯的风险是 $\frac{30}{100}\cdot\frac{1}{2}+\frac{40}{100}\cdot\frac{1}{3}$，不开灯的风险是 $\frac{40}{100}\cdot\frac{1}{2}$。这正是上面的答案乘以 $\frac{7}{10}$，理应如此，因为金丝雀对两边的比较都毫无影响。
- 只要 $p=P(A\mid C')$ 满足 $\frac{1}{2}p<\frac{1}{2}(1-p)+\frac{1}{3}p$，即只要 $p<\frac{3}{4}$，不开灯的策略便较佳。这里 $p=\frac{4}{7}$，远低于这个数值。
- 模拟 $10^6$ 个晚上，得出开灯时的死亡率约为 0.404，不开灯时约为 0.286，与 $\frac{17}{42}\approx 0.405$ 及 $\frac{2}{7}\approx 0.286$ 吻合。

## 常见错误

- 忘记以“没有金丝雀”为条件，直接把 $P(A)=0.4$ 和 $P(B)=0.3$ 当作全部事实来计算。得出的建议相同，但死亡概率却是错的。
- 混淆生还与死亡：开灯时森蚺的数字是 $\frac{2}{3}$ 的*生还*机会，所以死亡机会是 $\frac{1}{3}$。
- 把各名刺客当作可能性均等，或者忽略刺客这一层，直接把武器概率取平均。

## STEP 之外

这是*决策论*(decision theory)中的一道小题：我们选择在后验分布(posterior distribution)下令期望损失（expected loss）最小的行动，而这里的“损失”就是死亡的概率。只有 $P(A\mid C')$ 才起作用这一观察，以及最佳行动随之改变的临界值 $p<\frac{3}{4}$，都是分析这类决策的典型手法：数据改变后验分布，后验分布决定行动。想进一步了解条件概率和贝叶斯更新（Bayesian updating），可参阅 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门教材）。

详细解答

# STEP I 1997，第 14 题

概率与统计 · 期望成本的最小化

**STEP I 1997，第 14 题**

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

step.maths.org/questions/97-s1-q14

## 关键思路

成本 $T$ 是随机变量（random variable）$X$ 的函数，所以它的期望值（expectation）就是该函数乘以概率密度函数（density）后的积分。在 $x = y$ 处把积分拆开，便可将 $\mathrm{E}(T)$ 写成 $y$ 的普通函数；之后求其最小值只是大学一年级程度的微积分练习：求导、令导数为零，再检查驻点（stationary point）确实是极小值。

## 解答

我们取 $y \geq 0$，因为准备工作的份量为负数并无意义。记 $T = g(X)$，其中

$$g(x) = \begin{cases} y & x \leq y, \\ y + r + s(x - y) & x > y. \end{cases}$$

于是

$$\begin{aligned} \mathrm{E}(T) &= \int_0^\infty g(x)\,\mathrm{e}^{-x}\,\mathrm{d}x \\ &= \int_0^\infty y\,\mathrm{e}^{-x}\,\mathrm{d}x + \int_y^\infty \big(r + s(x - y)\big)\mathrm{e}^{-x}\,\mathrm{d}x, \end{aligned}$$

这里我们把成本 $y$ 抽了出来，因为无论发生甚么事都要支付这笔费用。由于密度函数的积分为 1，第一个积分等于 $y$。

至于第二个积分，作代换 $u = x - y$，则 $\mathrm{e}^{-x} = \mathrm{e}^{-y}\mathrm{e}^{-u}$，而 $u$ 由 0 变到 $\infty$：

$$\begin{aligned}\int_y^\infty (r + s(x-y))\mathrm{e}^{-x}\,\mathrm{d}x &= \mathrm{e}^{-y}\int_0^\infty (r+su)\,\mathrm{e}^{-u}\,\mathrm{d}u \\ &= \mathrm{e}^{-y}\Big(r\int_0^\infty \mathrm{e}^{-u}\,\mathrm{d}u + s\int_0^\infty u\,\mathrm{e}^{-u}\,\mathrm{d}u\Big) \\ &= \mathrm{e}^{-y}(r+s),\end{aligned}$$

其中用到 $\int_0^\infty u\,\mathrm{e}^{-u}\,\mathrm{d}u = \big[-u\mathrm{e}^{-u}\big]_0^\infty + \int_0^\infty \mathrm{e}^{-u}\,\mathrm{d}u = 1$（分部积分法，integration by parts）。因此

$$C(y) = \mathrm{E}(T) = y + (r+s)\,\mathrm{e}^{-y}.$$

**想一想** 有没有方法不用积分就能看出 $\mathrm{e}^{-y}(r+s)$ 这个因子？

有。水灾造成损失的概率是 $\mathrm{P}(X > y) = \mathrm{e}^{-y}$。在已知损失发生的条件下，超出的部分 $X - y$ 仍然服从平均值为 1 的指数分布（exponential distribution），因为指数分布具有无记忆性（memoryless），所以损失的期望值是 $r + s\cdot 1$。两者相乘便得 $(r+s)\mathrm{e}^{-y}$。

现在求最小值。求导得

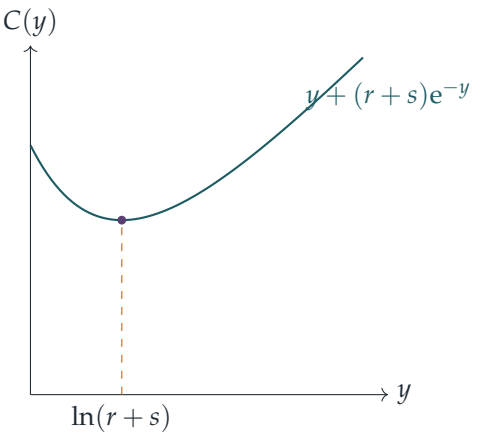
$$C'(y) = 1 - (r+s)\mathrm{e}^{-y}, \qquad C''(y) = (r+s)\mathrm{e}^{-y} > 0.$$

所以 $C'(y) = 0$ 当且仅当 $\mathrm{e}^y = r+s$，即 $y = \ln(r+s)$。由于对所有 $y$ 都有 $C'' > 0$，$C$ 是凸函数（convex），$C'$ 递增，因此 $C$ 在 $y < \ln(r+s)$ 时递减，在 $y > \ln(r+s)$ 时递增。故此这个驻点就是全域最小值（global minimum），而要令总成本的期望值最小，应取

$$y = \ln(r+s),$$

这正是所求。成本期望值的最小值为 $C\big(\ln(r+s)\big) = \ln(r+s) + 1$。

以上需要 $\ln(r+s) \geq 0$，即 $r+s \geq 1$，题目隐含了这个假设。若 $r+s < 1$，则对每个 $y \geq 0$ 都有 $C'(y) > 0$，所以最佳选择是 $y = 0$：损失实在太便宜，根本不值得作任何准备。



（草图以 $r+s=3$ 绘画。）

## 验算

- 当 $y=0$ 时，我们必定要支付损失，而 $\mathrm{E}(T)=r+s\,\mathrm{E}(X)=r+s$，与 $C(0)$ 吻合。
- 取 $r=2$、$s=3$，公式给出 $y=\ln 5\approx 1.609$，最小成本为 $1+\ln 5\approx 2.609$；对 $C(y)$ 作精细的网格搜索，找到的最小值点相同；另外在 $y=1$ 下模拟 $4\times 10^5$ 年，得出平均成本为 2.844，而公式给出的值是 $1+5\mathrm{e}^{-1}\approx 2.839$。
- 当 $r+s$ 增大时，$y$ 亦随之增加：水灾的代价愈高，就愈值得多作准备，这与常理相符。

## 常见错误

- 忘记准备成本 $y$ 是每一年都要支付的，而不是只在 $X\le y$ 时才支付。
- 对 $sx\,\mathrm{e}^{-x}$ 积分时漏掉了 $-sy\,\mathrm{e}^{-x}$ 一项，结果得出 $\mathrm{e}^{-y}(r+s+sy)$ 而不是 $\mathrm{e}^{-y}(r+s)$。
- 找到驻点后，没有证明它是极小值。

## STEP 之外

这是不确定情况下作决策（decision-making under uncertainty）的一个小例子：选择一个行动，使损失的期望值最小。同样的结构，即在一笔确定的成本与一项预期的惩罚之间取得平衡，正是运筹学（operations research）中“报童问题”（newsvendor problem）以及统计学中贝叶斯决策（Bayes decisions）这个概念的基础。上文用到的无记忆性，在连续分布之中刻划了指数分布的特征；Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教科书）涵盖了这个性质，以及随机变量函数的期望值。

详细解答

# STEP II 1997，第 1 题

纯数学 · 数字和与计数

> **STEP II 1997，第 1 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s2-q1

## 关键思路

不要把这些数逐一列出：先数清楚有多少个，再把每个位值（place value）的贡献分开相加。在末三个数位上，每个可用的数字出现的次数都相同，所以这些和几乎不用怎样计算就出来了。更妙的是，这里有一种对称性（symmetry）：把每个数字 $d$ 换成 $7-d$，就会把 $0 \leftrightarrow 7$ 和 $2 \leftrightarrow 5$ 互换，于是这些数可以两两配对，每对的和都相同，因数分解（factorisation）也随即得出。

## 解答

### 哪些数符合条件？

介乎 1000 与 6000 之间而符合条件的数有四个数位，而它的千位数字必定是 $0, 2, 5, 7$ 其中之一。千位不可以是 0（否则不足四位），也不可以是 7（否则这个数会超过 7000）。所以千位数字是 2 或 5，而由 2000 至 5777 之间按此方式组成的每一个数都在范围之内。两个端点 1000 和 6000 本身含有数字 1 和 6，所以“之间”是否包括端点并无影响。

其余三个数位各自可以独立地取四个可用数字中的任何一个。所以共有

$$2 \times 4 \times 4 \times 4 = 128$$

个数。

### 按位值相加

千位。这 128 个数之中有一半，即 64 个，以 2 开首，另外 64 个以 5 开首。千位数字的贡献是

$$64 \times (2+5) \times 1000 = 448\,000.$$

百位、十位和个位。固定其中一个数位。四个数字各自恰好在 $128/4 = 32$ 个数的这个数位上出现，因为不论这个数位上放的是哪个数字，其余数位的选法数目都相同。因此这个数位上的数字总和是 $32 \times (0+2+5+7) = 32 \times 14 = 448$。这三个数位的权重分别是 100、10 和 1，所以它们合共的贡献是

$$448 \times (100+10+1) = 448 \times 111 = 49\,728.$$

相加之后，所求的和是

$$S = 448 \times 1000 + 448 \times 111 = 448 \times 1111 = 497\,728.$$

### 写成质数的乘积

我们有 $448 = 64 \times 7 = 2^6 \times 7$，以及 $1111 = 11 \times 101$，其中 101 是质数（prime）（它不能被 2、3、5 或 7 整除，而且 $11^2 > 101$）。因此

$$S = 2^6 \times 7 \times 11 \times 101.$$

**想一想** 有没有方法可以不必逐个位值记帐，就看出答案?

有。如果 $n$ 符合条件，那么 $7777 - n$ 也符合条件：从 7777 逐位相减时不需要借位，而且每个数字 $d$ 都会变成 $7-d$，这个变换把 $\{0,2,5,7\}$ 映射到它自身，并把首位数字 2 和 5 互换。所以这 128 个数可分成 64 对，每对的和都是 7777（没有任何数会与自己配对，因为 7777 是奇数）。因此

$$S = 64 \times 7777 = 2^6 \times 7 \times 1111 = 2^6 \times 7 \times 11 \times 101,$$

与上面的计算结果一致。

## 验算

用电脑在 1000 至 6000 之间作简短的搜索，可找到恰好 128 个符合条件的数，其和为 497 728，而 $2^6 \times 7 \times 11 \times 101 = 64 \times 7777 = 497\,728$。再从大小上检查是否合理：符合条件的数的平均值是 $7777/2 = 3888.5$，正好稳稳落在 2000 至 5777 这个范围的中间。

## 常见错误

- 容许 7 作为首位数字（得出超过 6000 的数），或者忘记 0 不可以放在首位。
- 以为每个数字在千位上也出现 128/4 次：千位上只会出现 2 和 5，各 64 次。
- 停在 497 728 或 $448 \times 1111$，没有把 1111 分解成 $11 \times 101$。

## STEP 之外

数算每个数字在每个数位上出现多少次，是算两次（double counting）论证最简单的例子：把同一个量按两种次序求和，这里就是按数和按数位。$7777 - n$ 的配对是一种对称性论证，高斯（Gauss）的公式 $1 + 2 + \cdots + n = \frac{1}{2}n(n+1)$ 也可以用同类的论证得出。数字取自某个固定集合的数，以及它们的和，是初等数论（elementary number theory）和组合数学（combinatorics）中的常见题材；Davenport 的 *The Higher Arithmetic*（高等算术）是很值得一读的下一步。

详细解答

# STEP II 1997，第 2 题

纯数学 · 递推关系

**STEP II 1997，第 2 题**

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

step.maths.org/questions/97-s2-q2

## 关键思路

这一串等式其实是乔装了的递推关系（recurrence）：每一环都在说 $x_n + 2/x_{n+1} = 3$，所以每一项都决定了下一项。先把头几项以分数形式算出来，再仔细观察分子和分母。2 的幂很快就会出现；一旦有了猜想，递推关系正好提供归纳法（induction）所需的那一步。

## 解答

### 找出规律

把这串等式逐个读出来，得

$$\frac{2}{x_1} = 3 \qquad \text{及} \qquad x_n + \frac{2}{x_{n+1}} = 3 \quad (n \geq 1),$$

因此

$$x_1 = \frac{2}{3}, \qquad x_{n+1} = \frac{2}{3 - x_n}.$$

（这个递推关系只在 $x_n \neq 3$ 时才有意义；我们稍后会看到，每个 $x_n$ 都介乎 0 与 1 之间。）算出头几项：

$$x_1 = \frac{2}{3}, \qquad x_2 = \frac{2}{3 - \frac{2}{3}} = \frac{6}{7}, \qquad x_3 = \frac{2}{3 - \frac{6}{7}} = \frac{14}{15}, \qquad x_4 = \frac{2}{3 - \frac{14}{15}} = \frac{30}{31}.$$

分母 $3, 7, 15, 31$ 各自都比 2 的某个幂少一，即 $2^{n+1} - 1$，而每个分子又比它的分母少一。于是我们猜想

$$x_n = \frac{2^{n+1} - 2}{2^{n+1} - 1} = 1 - \frac{1}{2^{n+1} - 1}.$$

### 以归纳法证明

设 $P(n)$ 为命题 $x_n = \dfrac{2^{n+1}-2}{2^{n+1}-1}$。

**基础情况（base case）**。当 $n=1$ 时，公式给出 $\frac{4-2}{4-1} = \frac{2}{3} = x_1$，所以 $P(1)$ 成立。

**归纳步骤（inductive step）**。假设对某个 $k \geq 1$，$P(k)$ 成立。为了令代数运算整洁一些，记 $N = 2^{k+1}$，于是 $x_k = \dfrac{N-2}{N-1}$。那么

$$3 - x_k = \frac{3(N-1)-(N-2)}{N-1} = \frac{2N-1}{N-1},$$

它不等于零（事实上它大于 2，因为 $x_k < 1$），所以我们可以除以它：

$$x_{k+1} = \frac{2}{3-x_k} = \frac{2(N-1)}{2N-1} = \frac{2N-2}{2N-1}.$$

> ? **想一想** 为甚么这就是 $n = k+1$ 时的公式?

因为 $2N = 2^{k+2} = 2^{(k+1)+1}$。所以

$$x_{k+1} = \frac{2^{(k+1)+1}-2}{2^{(k+1)+1}-1},$$

这正是 $P(k+1)$。

由于 $P(1)$ 成立，而且对每个 $k \geq 1$ 都有 $P(k) \Rightarrow P(k+1)$，根据归纳法，$P(n)$ 对所有 $n \geq 1$ 都成立。这同时确认了对每个 $n$ 都有 $0 < x_n < 1$，所以分母 $3 - x_n$ 永远不会是零。

### “其他方法”：线性化

这个猜想其实也可以推导出来。对于形如 $x_{n+1} = 2/(3-x_n)$ 的递推关系，只要引入一个各项均非零的新数列 $(a_n)$，并写成 $x_n = 2a_n/a_{n+1}$，它就会变成线性（linear）的。代入后得

$$\frac{2a_{n+1}}{a_{n+2}} = \frac{2}{3-\frac{2a_n}{a_{n+1}}} = \frac{2a_{n+1}}{3a_{n+1}-2a_n}, \qquad \text{所以} \qquad a_{n+2} = 3a_{n+1} - 2a_n.$$

辅助方程（auxiliary equation）$t^2 - 3t + 2 = 0$ 的根是 1 和 2，所以 $a_n = A + B\cdot 2^n$。取 $a_1 = 1$（我们可以随意按比例缩放 $a$）及 $a_2 = 3$（使得 $x_1 = 2a_1/a_2 = \frac{2}{3}$），便得 $a_n = 2^n - 1$，从而 $x_n = \dfrac{2(2^n-1)}{2^{n+1}-1}$，与之前的公式相同。

## 验算

- 在公式中取 $n = 0$，得 $x_0 = 0$；而第一条等式 $3 = 2/x_1$ 确实就是一般的那一环 $x_0 + 2/x_1 = 3$ 在 $x_0 = 0$ 时的情况。这是一个漂亮的一致性检查。

- 当 $n \to \infty$ 时，$x_n \to 1$。递推关系的极限（limit）$L$ 必须满足 $L = 2/(3-L)$，即 $L^2 - 3L + 2 = 0$，所以 $L = 1$ 或 $L = 2$，而我们的数列从下方趋近的正是 1。
- 由 $x_1 = \frac{2}{3}$ 出发，以精确分数反复计算 $x_{n+1} = 2/(3 - x_n)$，结果对每个不超过 40 的 $n$ 都与公式吻合；例如 $x_5 = \frac{62}{63}$。

## 常见错误

- 读错那串等式：方程是 $x_n + 2/x_{n+1} = 3$，而不是 $x_{n+1} + 2/x_n = 3$。弄错了就会得到另一个数列。
- 写到“规律如此类推”就停笔。题目要求的是证明，而几项数值只算是佐证。
- 在归纳步骤中，忘记把 $2 \cdot 2^{k+1}$ 改写成 $2^{(k+1)+1}$，以致结论始终没有明确显示为命题 $P(k+1)$。

## STEP 之外

形如 $x_{n+1} = (ax_n + b)/(cx_n + d)$ 的递推关系，就是对一个莫比乌斯变换（Möbius transformation）作迭代（iteration），而且总可以像“其他方法”那样线性化。等价地，我们可以追踪 $2 \times 2$ 矩阵 $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 的幂，它的特征值（eigenvalues，这里是 1 和 2）支配着长远的变化。不动点（fixed points）1 和 2 分别是吸引的（attracting）和排斥的（repelling），这就是数列最终稳定于 1 的原因。关于矩阵的观点，可参阅 Strang 的 *Introduction to Linear Algebra*（线性代数入门）；至于数列收敛的一般理论，可参阅 Abbott 的 *Understanding Analysis*（数学分析入门）。

详细解答

# STEP II 1997，第 3 题

纯数学 · 部分分式积分法

**STEP II 1997，第 3 题**

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

step.maths.org/questions/97-s2-q3

## 关键思路

分母中的两个二次式相乘得 $x^4+4$。这是经典的“乔装了的平方差”（difference of two squares）：$x^4+4=(x^2+2)^2-(2x)^2$。一旦求得部分分式（partial fractions），每一项都是标准积分。我们把每个分子拆成两部分：一部分是其分母的导数（derivative）的倍数，积分后得对数（logarithm）；另一部分是常数，配方（completing the square）后得反正切（arctangent）。

## 解答

### 部分分式

首先留意

$$(x^2+2x+2)(x^2-2x+2)=(x^2+2)^2-(2x)^2=x^4+4x^2+4-4x^2=x^4+4.$$

因此，把所求的恒等式两边同乘 $x^4+4$，我们需要

$$(ax+b)(x^2-2x+2)+(cx+d)(x^2+2x+2)=1 \quad \text{对所有} x.$$

展开左方得

$$(a+c)x^3+(b-2a+d+2c)x^2+(2a-2b+2c+2d)x+(2b+2d).$$

与常数多项式 1 比较系数（comparing coefficients）：

$$\begin{aligned} x^3: &\quad a+c=0,\\ x^2: &\quad b+d-2a+2c=0,\\ x^1: &\quad 2(a+c)+2(d-b)=0,\\ x^0: &\quad 2(b+d)=1. \end{aligned}$$

由第一及第三条方程得 $c=-a$ 及 $d=b$。再由第四条方程得 $b=d=\frac{1}{4}$。最后，第二条方程变成 $\frac{1}{2}-4a=0$，所以 $a=\frac{1}{8}$。因此

$$a=\tfrac{1}{8}, \qquad b=\tfrac{1}{4}, \qquad c=-\tfrac{1}{8}, \qquad d=\tfrac{1}{4},$$

即

$$\frac{1}{x^4+4}=\frac{1}{8}\left(\frac{x+2}{x^2+2x+2}+\frac{2-x}{x^2-2x+2}\right).$$

（其实我们毋须任何代数运算，也可预料 $c=-a$ 及 $d=b$。以 $-x$ 代替 $x$，$\frac{1}{x^4+4}$ 保持不变，而两个分母则互相对调。）

## 积分

**想一想** 我们怎样求 $\dfrac{x+2}{x^2+2x+2}$ 的积分?

分母的导数是 $2x+2$。于是我们把分子写成 $2x+2$ 的倍数，再加上余下的部分：

$$x+2=\tfrac{1}{2}(2x+2)+1, \qquad 2-x=-\tfrac{1}{2}(2x-2)+1.$$

配方得 $x^2\pm 2x+2=(x\pm 1)^2+1$。因此

$$\frac{1}{x^4+4}=\frac{1}{16}\cdot\frac{2x+2}{x^2+2x+2}-\frac{1}{16}\cdot\frac{2x-2}{x^2-2x+2}+\frac{1}{8}\cdot\frac{1}{(x+1)^2+1}+\frac{1}{8}\cdot\frac{1}{(x-1)^2+1}.$$

两个二次式对所有实数 $x$ 均为正数，所以每一项在 $[0,1]$ 上都是连续的。首两项的形式为 $f'/f$，其中 $f>0$，积分后得 $\ln f$。最后两项积分后得反正切。所以

$$\begin{aligned} \int_0^1 \frac{\mathrm{d}x}{x^4+4} &= \frac{1}{16}\Big[\ln(x^2+2x+2)-\ln(x^2-2x+2)\Big]_0^1+\frac{1}{8}\Big[\tan^{-1}(x+1)+\tan^{-1}(x-1)\Big]_0^1\\ &= \frac{1}{16}\big[(\ln 5-\ln 1)-(\ln 2-\ln 2)\big]\\ &\quad +\frac{1}{8}\big[(\tan^{-1}2+\tan^{-1}0)-(\tan^{-1}1+\tan^{-1}(-1))\big]\\ &= \frac{1}{16}\ln 5+\frac{1}{8}\big[\tan^{-1}2-\big(\tfrac{\pi}{4}-\tfrac{\pi}{4}\big)\big]\\ &= \frac{1}{16}\ln 5+\frac{1}{8}\tan^{-1}2, \end{aligned}$$

正如所求。各个 $\ln 2$ 项和各个 $\frac{\pi}{4}$ 项都成对抵消，这正是上文留意到的对称性所带来的美妙结果。

## 验算

- 当 $x=0$ 时，部分分式给出 $\frac{1}{8}\left(\frac{2}{2}+\frac{2}{2}\right)=\frac{1}{4}=\frac{1}{0+4}$。当 $x=1$ 时，则给出 $\frac{1}{8}\left(\frac{3}{5}+\frac{1}{1}\right)=\frac{1}{5}$，同样正确。
- 被积函数（integrand）在 $[0,1]$ 上由 $\frac{1}{4}$ 递减至 $\frac{1}{5}$，所以积分值必定介乎 0.2 与 0.25 之间。以数值计算，$\frac{1}{16}\ln 5+\frac{1}{8}\tan^{-1}2\approx 0.100589+0.138394=0.238983$。对该积分使用辛普森法则（Simpson's rule）得 0.2389834593，两者吻合至约 16 位有效数字。

## 常见错误

- 试图寻找实根，借此在实数范围内把 $x^4+4$ 因式分解（factorise）。它并没有实根，所以因式必定是二次式。窍门在于加上再减去 $4x^2$ 来配方。
- 忘记反正切项的下限。$\tan^{-1}(0+1)=\frac{\pi}{4}$ 与 $\tan^{-1}(0-1)=-\frac{\pi}{4}$ 必须两者都计算在内，才会互相抵消。

## STEP 之外

因式分解 $x^4+4=(x^2+2x+2)(x^2-2x+2)$ 是热尔曼（Sophie Germain）恒等式 $a^4+4b^4=(a^2+2ab+2b^2)(a^2-2ab+2b^2)$ 的特例。举例说，它说明了 $n^4+4$ 只有在 $n=1$ 时才是质数。更一般而言，每个实系数多项式都可分解成实系数的一次因式和二次因式，因为复数根总是以共轭对（conjugate pairs）的形式出现。所以，只要懂得部分分式，从而懂得对数和反正切，便足以求出任何有理函数（rational function）的积分。利用围道积分（contour integration），你几乎毋须计算便可求得 $\int_{-\infty}^{\infty}\frac{\mathrm{d}x}{x^4+4}=\frac{\pi}{4}$；Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）解释了个中原因。

详细解答

# STEP II 1997，第 4 题

纯数学 · 余式定理

> **STEP II 1997，第 4 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s2-q4

## 关键思路

以一个次数（degree）为 $d$ 的多项式作除式，所得余式（remainder）的次数必小于 $d$，而且恒等式（identity）被除式 = 除式 × 商式 + 余式 对每一个 $x$ 都成立。因此，我们代入令除式变为零的 $x$ 值：商式（quotient）那一项随即消失，只剩下余式。整道题目用的就是这一招，前后共用了三次。

## 解答

### 余式定理（remainder theorem）

把 $\mathrm{p}(x)$ 除以一次多项式 $x-a$，得到商式 $\mathrm{s}(x)$ 和余式 $R$，而余式的次数小于 1，所以 $R$ 是常数：

$$\mathrm{p}(x) = (x-a)\,\mathrm{s}(x) + R \qquad \text{对所有} x.$$

由于这是关于 $x$ 的恒等式，我们可以代入 $x=a$。右边第一项变成 $0\cdot \mathrm{s}(a)=0$，于是剩下 $\mathrm{p}(a)=R$。所以余式就是 $\mathrm{p}(a)$，正如所求。

### 第 (i) 部分

根据余式定理，题目给出的余式表示

$$\mathrm{p}(1)=3, \qquad \mathrm{p}(2)=1, \qquad \mathrm{p}(3)=5.$$

把 $x=1,2,3$ 代入 $\mathrm{p}(x)=(x-1)(x-2)(x-3)\mathrm{q}(x)+\mathrm{r}(x)$，每次第一项都会变成零，所以

$$\mathrm{r}(1)=3, \qquad \mathrm{r}(2)=1, \qquad \mathrm{r}(3)=5.$$

**想一想** 我们对 $\mathrm{q}(x)$ 一无所知。这会构成问题吗?

不会：当 $x = 1, 2, 3$ 时，因式 $(x-1)(x-2)(x-3)$ 等于零，所以无论 q 是甚么，它都没有任何贡献。三个函数值足以确定 r，因为它最多只有三个系数。

次数不超过二的多项式，可以方便地写成

$$\mathrm{r}(x) = A(x-1)(x-2) + B(x-1) + C,$$

因为这样每次代入都会逐一消去某些项：

$$\begin{aligned} x = 1: &\quad C = 3, \\ x = 2: &\quad B + C = 1 \;\Rightarrow\; B = -2, \\ x = 3: &\quad 2A + 2B + C = 5 \;\Rightarrow\; 2A - 4 + 3 = 5 \;\Rightarrow\; A = 3. \end{aligned}$$

因此

$$\mathrm{r}(x) = 3(x-1)(x-2) - 2(x-1) + 3 = 3x^2 - 11x + 11.$$

验算：$\mathrm{r}(1) = 3 - 11 + 11 = 3$，$\mathrm{r}(2) = 12 - 22 + 11 = 1$，$\mathrm{r}(3) = 27 - 33 + 11 = 5$。

**第 (ii) 部分**

（题目有一处写作 $\mathrm{P}_n(x)$；它指的是同一个多项式 $\mathrm{P}(x)$。）根据余式定理，我们需要

$$\mathrm{P}(a) = a \qquad \text{其中} a = 0, 1, \ldots, n.$$

多项式 $x$ 本身已经符合这个要求，但它的次数不对。所以我们改为考虑 $\mathrm{D}(x) = \mathrm{P}(x) - x$：我们需要 $\mathrm{D}(a) = 0$ 对 $a = 0, 1, \ldots, n$ 成立，于是根据因式定理（factor theorem，即余式为 0 时的余式定理），$x, x-1, \ldots, x-n$ 各自都整除 $\mathrm{D}(x)$。这 $n+1$ 个一次因式互不相同，所以它们的乘积整除 $\mathrm{D}(x)$，而这个乘积的次数是 $n+1$。

**想一想** 怎样才能令 $\mathrm{P}(x)$ 的次数刚好是 $n+1$?

取 $\mathrm{D}(x)$ 为该乘积本身（喜欢的话也可以乘上一个非零常数）。加上 $x$ 不可能抵销它的首项（leading term）$x^{n+1}$，因为 $n+1 \geq 2$。所以

$$\mathrm{P}(x) = x + x(x-1)(x-2)\cdots(x-n)$$

的次数是 $n+1$，并且 $\mathrm{P}(a) = a + 0 = a$ 对每个满足 $0 \leq a \leq n$ 的整数 $a$ 都成立。这就是一个符合要求的多项式。

事实上，这个论证找出了所有这样的多项式：如果 P 的次数是 $n+1$，并且有题目所述的余式，那么 $\mathrm{D}(x) = \mathrm{P}(x) - x$ 的次数最多是 $n+1$，又能被那个次数为 $n+1$ 的乘积整除，

所以 $\mathrm{D}(x) = k\,x(x-1)\cdots(x-n)$，其中 $k$ 是常数，而且必须有 $k \neq 0$，次数才会是 $n+1$。一般的答案是

$$\mathrm{P}(x) = x + k\,x(x-1)(x-2)\cdots(x-n), \qquad k \neq 0.$$

## 验算

- 当 $n = 1$ 时，$\mathrm{P}(x) = x + x(x-1) = x^2$：除以 $x$ 余 0，除以 $x-1$ 余 1，符合要求。
- 在 (i) 中，任取一个 q，例如 $\mathrm{q}(x) = x$，构作出 p，再用综合除法（synthetic division）分别除以 $x-1$、$x-2$、$x-3$：所得余式正是 $3, 1, 5$。
- 我们用电脑验证了以下各项：以 1000 个随机多项式检验余式定理；第 (i) 部分取 100 个随机选择的 q；第 (ii) 部分取 $n = 1, \ldots, 11$ 及 $k = 1, -3, 7$。每一个余式都完全吻合。

## 常见错误

- 在第 (ii) 部分，忘记了多项式 $\mathrm{P}(x) = x$ 本身虽然满足各项条件，但次数是 1 而不是 $n+1$；正是那个乘积项把次数修正过来。
- 就 $ax^2 + bx + c$ 的系数解三条联立方程并无不妥，只是较慢；采用嵌套形式 $A(x-1)(x-2) + B(x-1) + C$ 则可免去消元的步骤。

## STEP 之外

第 (i) 部分是多项式插值（polynomial interpolation）的一个例子：通过任意 $m$ 个 $x$ 坐标互不相同的点，恰好有一个次数小于 $m$ 的多项式，而我们所用的嵌套形式就是它的牛顿（Newton）形式。同样的推理，用环（ring）的语言来说，就是多项式的中国余数定理（Chinese remainder theorem）：知道一个多项式模 $x-1$、$x-2$ 和 $x-3$ 的结果，等同于知道它模这三者乘积的结果。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）详细讲述了多项式环和这个定理。

详细解答

# STEP II 1997，第 5 题

纯数学 · 复指数映射

> **STEP II 1997，第 5 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s2-q5

本题在 *Beyond STEP* 第 5.7 节作为映射的复指数函数中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch05/complex-exponential.html

## 关键思路

这个关系式其实就是 $z = \mathrm{e}^{u}\mathrm{e}^{\mathrm{i}v} = \mathrm{e}^{w}$，只不过写成了模—辐角形式（modulus–argument form）：$u$ 决定 $z$ 的模（modulus），$v$ 决定它的辐角（argument）。一旦看出这一点，每一部分其实都是极坐标（polar coordinates）的问题。唯一要留神的是，辐角只能确定到相差 $2\pi$ 的整数倍，所以这个映射是多对一（many-to-one）的；要得到一一对应（one-to-one correspondence），就必须把 $u$–$v$ 平面裁成一条高度恰当的水平带状区域，并且小心决定哪些边界要包括在内。

## 解答

**当 $z = \mathrm{i}\,\mathrm{e}$ 时求 $w$**

由于 $\mathrm{e}^{u} > 0$ 而 $|\cos v + \mathrm{i}\sin v| = 1$，这个关系式表示

$$|z| = \mathrm{e}^{u}, \qquad \arg z = v \pmod{2\pi}.$$

当 $z = \mathrm{i}\,\mathrm{e}$ 时，$|z| = \mathrm{e}$ 而 $\arg z = \frac{\pi}{2}$。所以 $\mathrm{e}^{u} = \mathrm{e}$，得 $u = 1$（实指数函数是一对一的）；又 $\cos v + \mathrm{i}\sin v = \mathrm{i}$，所以 $\cos v = 0$ 且 $\sin v = 1$，得 $v = \frac{\pi}{2} + 2k\pi$。因此

$$w = 1 + \mathrm{i}\left(\tfrac{\pi}{2} + 2k\pi\right), \qquad k \in \mathbb{Z}.$$

这样的 $w$ 有无穷多个，它们在垂直线 $u = 1$ 上每隔 $2\pi$ 出现一个。

### 直线 $u =$ 常数 的像

在直线 $u = c$ 上，点 $z$ 的模是 $\mathrm{e}^c$，辐角是 $v$。当 $v$ 取遍所有实数时，$\cos v + \mathrm{i}\sin v$ 会绕单位圆走（无穷多圈），所以 $z$ 描出的轨迹是

$$\text{圆} x^2 + y^2 = \mathrm{e}^{2c}, \quad \text{圆心} O, \text{ 半径} \mathrm{e}^c.$$

$c$ 愈大，圆就愈大；$c = 0$ 给出单位圆；当 $c \to -\infty$ 时，这些圆向 $O$ 收缩，但永远不会到达该点，因为 $\mathrm{e}^c > 0$。

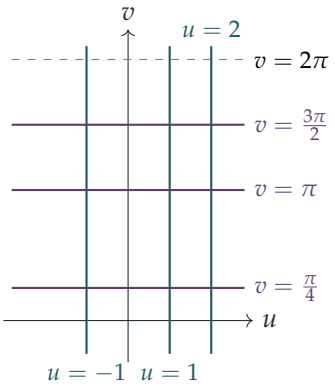
### 直线 $v =$ 常数 的像

在直线 $v = c$ 上，我们有 $z = \mathrm{e}^u(\cos c + \mathrm{i}\sin c)$，即固定的单位向量 $(\cos c, \sin c)$ 的正数倍。当 $u$ 取遍 $\mathbb{R}$ 时，倍数 $\mathrm{e}^u$ 取 $(0, \infty)$ 中的每一个值恰好一次。所以它的像是

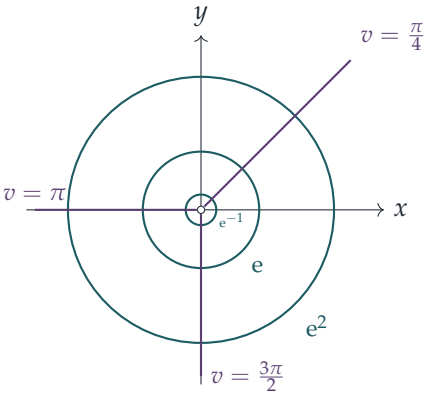
$$\text{由} O \text{ 出发、以角度} c \text{ 与正} x \text{ 轴相夹的半直线（half-line），但不包括} O \text{ 本身。}$$

直线 $v = c$ 与 $v = c + 2k\pi$ 给出同一条半直线。举例说，用直角坐标表示时，$v = \frac{\pi}{4}$ 给出 $y = x$ 中 $x > 0$ 的部分，而 $v = \pi$ 给出负 $x$ 轴。

下图显示 $u$–$v$ 平面上由直线组成的网格及其像：垂直线变成圆，水平线变成射线。




$u$–$v$ 平面




$x$–$y$ 平面

（图中的圆并非按比例绘画：半径 $\mathrm{e}^{-1}$、e、$\mathrm{e}^2$ 增长得非常快。空心圆圈标示 $O$，这一点永远不会到达。）

### 与第一象限一一对应的子集

点 $z$ 位于开的第一象限（first quadrant），当且仅当 $z \neq 0$ 而且它有一个严格介乎 0 与 $\frac{\pi}{2}$ 之间的辐角。取

$$W_1 = \left\{(u, v) : u \in \mathbb{R},\ 0 < v < \tfrac{\pi}{2}\right\}, \qquad W_2 = \left\{(u, v) : u \in \mathbb{R},\ 2\pi < v < \tfrac{5\pi}{2}\right\}.$$

$W_1$ 的每一点都映射到第一象限内，因为在那里 $e^u\cos v > 0$ 且 $e^u \sin v > 0$。反过来，对第一象限的一点 $z$，等式 $z = e^u(\cos v + i\sin v)$ 成立当且仅当 $u = \ln|z|$ 且 $v = \theta + 2k\pi$，其中 $\theta \in (0, \frac{\pi}{2})$ 是它的主辐角（principal argument）。这些值之中恰好只有一个，即 $k = 0$ 的那个，落在 $(0, \frac{\pi}{2})$ 内，所以 $z$ 在 $W_1$ 中恰有一个原像（preimage）。因此 $W_1$ 与该象限一一对应。取 $k = 1$ 作同样的论证，便适用于 $W_2$，它就是 $W_1$ 向上平移 $2\pi$ 而成。

（任何带状区域 $\{2k\pi < v < 2k\pi + \frac{\pi}{2}\}$ 都同样可行。）

### 与 $0 < |z| < 1$ 一一对应的子集

现在我们需要 $|z| = e^u < 1$，即 $u < 0$，而且每一个辐角都必须恰好取到一次，所以 $v$ 必须取遍一个长度恰好是 $2\pi$ 的区间。

> **想一想** 这条带状区域应该包括上边界、下边界、两者都包括，还是两者都不包括？

恰好包括其中一条。假如两条都包括，例如 $0 \le v \le 2\pi$，那么 $(u, 0)$ 和 $(u, 2\pi)$ 都会映射到正实轴上的同一点 $e^u$，这个对应就不是一对一的了。假如两条都不包括，正实轴就完全取不到。只包括一条边界，才能令每个辐角恰好出现一次。所以取

$$W_3 = \{(u,v) : u < 0,\ 0 \le v < 2\pi\}, \qquad W_4 = \{(u,v) : u < 0,\ -\pi < v \le \pi\}.$$

对于 $W_3$：若 $0 < |z| < 1$，则必然有 $u = \ln|z| < 0$，而在 $z$ 的各个辐角 $\theta + 2k\pi$ 之中，恰好有一个落在 $[0, 2\pi)$ 内，因为这个区间的长度是 $2\pi$，而且只包含它的一个端点。所以每一个这样的 $z$ 在 $W_3$ 中恰有一个原像，而 $W_3$ 的每一点都映射到一个满足 $0 < e^u < 1$ 的 $z$。对于 $W_4$，把区间换成 $(-\pi, \pi]$，同样的论证依然成立，这个区间对应的正是主辐角。两者都是把半平面 $u < 0$ 裁成高度为 $2\pi$ 的半开带状区域。

## 验算

- 代入 $u = 1$、$v = \frac{\pi}{2} + 2\pi$：$z = e(\cos\frac{5\pi}{2} + i\sin\frac{5\pi}{2}) = ie$，符合要求。
- 数值上，我们对数以千计的随机 $(u, v)$ 作检验，发现 $|z| = e^u$ 及 $\arg z \equiv v$ 的误差都在 $10^{-9}$ 之内。对每个目标集合中的随机点，我们发现它在 $W_1, \dots, W_4$ 中相应的集合内都恰有一个原像。此外，点 $(-1, 0)$ 和 $(-1, 2\pi)$ 都给出 $z = e^{-1} \approx 0.3679$，这证实了闭的带状区域会造成重复计算。

## 常见错误

- 只写出 $w = 1 + \frac{\pi}{2}i$，忘记了 $+2k\pi$。

- 把原计数进像之中：$e^u$ 永远不等于 0，所以那些圆永远不会缩成一点，而那些射线的起点是不包括在内的。
- 为 $W_3$ 选用闭的带状区域 $0 \leq v \leq 2\pi$（或开的带状区域 $0 < v < 2\pi$）：前者把正实轴重复计算，后者则漏掉了它。

## STEP 之外

这里的映射就是复指数函数（complex exponential）$w \mapsto e^w$。它把直角坐标的网格线变成极坐标的网格线，而且是周期函数，周期为 $2\pi i$。选择一条令它成为一对一的带状区域，正正就是为它的反函数选择一个复对数的分支（*branch of the complex logarithm*）。你舍去的那条边界就是分支切割（*branch cut*），这解释了为甚么 $\log z$ 不能在整个 $\mathbb{C} \setminus \{0\}$ 上连续地定义。Tristan Needham 的 *Visual Complex Analysis*（以图像阐释复分析的著作）对这个映射以及一般的保角映射（conformal maps）都有十分精美的图示。

详细解答

# STEP II 1997，第 6 题

纯数学 · 三角方程

> **STEP II 1997，第 6 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s2-q6

## 关键思路

把所有式子都用 $t = \tan\theta$ 表示。三倍角公式（triple-angle formula）把方程化为 $t$ 的三次方程（cubic），而一旦看出一个容易的根，余下的二次方程正好就是第一部分的那一个，因此它的根可以用一个二倍角（double-angle）条件来处理，而不必计算 $\tan^{-1}(1 \pm \sqrt{2})$。余切（cotangent）方程其实是同一个方程的变相：代换（substitution）$\theta = \frac{\pi}{2} - \psi$ 把 cot 换成 tan。

## 解答

### 开首的结果

设 $t = \tan\phi$，其中 $t^2 = 2t + 1$。那么 $1 - t^2 = -2t$。此外 $t \neq 0$，因为 $t = 0$ 不满足 $t^2 = 2t + 1$；而且 $t^2 \neq 1$（否则 $2t = 0$）。所以 $\tan 2\phi$ 有定义，并且

$$\tan 2\phi = \frac{2t}{1 - t^2} = \frac{2t}{-2t} = -1,$$

这正是所要证明的。

逆命题（converse）同样成立，而我们稍后会用到它。若 $\tan 2\phi = -1$，则 $2t = -(1 - t^2)$，即 $t^2 = 2t + 1$。所以

$$\tan^2\phi = 2\tan\phi + 1 \iff \tan 2\phi = -1.$$

**方程** $\tan\theta = 2 + \tan 3\theta$

要成为解，$\tan\theta$ 和 $\tan 3\theta$ 都必须有定义。设 $t = \tan\theta$。那么

$$\tan 3\theta = \frac{3t - t^3}{1 - 3t^2},$$

而且因为 $\tan 3\theta$ 有定义，所以 $1 - 3t^2 \neq 0$。因此

$$\tan\theta - \tan 3\theta = \frac{t(1 - 3t^2) - (3t - t^3)}{1 - 3t^2} = \frac{-2t - 2t^3}{1 - 3t^2}.$$

令上式等于 2，并把两边乘以 $1 - 3t^2$（它不等于零），得

$$-t - t^3 = 1 - 3t^2 \quad \Longleftrightarrow \quad t^3 - 3t^2 + t + 1 = 0.$$

各项系数之和为零，所以 $t = 1$ 是一个根，并且

$$t^3 - 3t^2 + t + 1 = (t - 1)(t^2 - 2t - 1).$$

$t = 1$ 和 $t = 1 \pm \sqrt{2}$ 都不满足 $3t^2 = 1$，所以每一个根都给出真正的解。

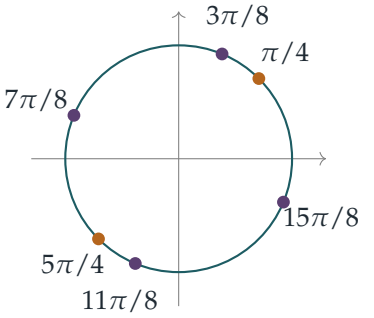
**想一想** 怎样把 $\tan\theta = 1 \pm \sqrt{2}$ 化为 $\pi$ 的有理数倍?

我们利用第一部分。$t^2 - 2t - 1 = 0$ 的根正好就是满足 $t^2 = 2t + 1$ 的 $\tan\theta$ 的值，而根据上面的等价关系（equivalence），这些正好就是满足 $\tan 2\theta = -1$ 的 $\theta$。

- $\tan\theta = 1$：$\theta = \frac{\pi}{4} + n\pi$。在 $(0, 2\pi)$ 内，这给出 $\theta = \frac{\pi}{4}, \frac{5\pi}{4}$。
- $\tan 2\theta = -1$：$2\theta = \frac{3\pi}{4} + n\pi$，所以 $\theta = \frac{3\pi}{8} + \frac{n\pi}{2}$。这里 $2\theta$ 的取值范围是 $(0, 4\pi)$，所以 $n = 0, 1, 2, 3$，而 $\theta = \frac{3\pi}{8}, \frac{7\pi}{8}, \frac{11\pi}{8}, \frac{15\pi}{8}$。

（例如 $\tan\frac{3\pi}{8} = 1 + \sqrt{2}$，而 $\tan\frac{7\pi}{8} = -\tan\frac{\pi}{8} = 1 - \sqrt{2}$。）所以在 $0 < \theta < 2\pi$ 内的解是

$$\theta = \frac{\pi}{4}, \frac{3\pi}{8}, \frac{7\pi}{8}, \frac{5\pi}{4}, \frac{11\pi}{8}, \frac{15\pi}{8}.$$




琥珀色的点来自 $\tan\theta = 1$；梅紫色的点彼此相隔四分之一圈，来自 $\tan 2\theta = -1$。

**方程** $\cot\theta = 2 + \cot 3\theta$

> ? **想一想** 我们是否要用 cot 把整个计算重做一遍?

不必。设 $\theta = \frac{\pi}{2} - \psi$。那么 $\cot\theta = \cot\left(\frac{\pi}{2} - \psi\right) = \tan\psi$。此外

$$\cot 3\theta = \cot\left(\frac{3\pi}{2} - 3\psi\right) = \cot\left(\frac{\pi}{2} - 3\psi\right) = \tan 3\psi,$$

这里用到 cot 的周期（period）为 $\pi$ 这个事实。每一边有定义，当且仅当对应的 tan 有定义。所以方程变成

$$\tan\psi = 2 + \tan 3\psi,$$

这就是前一个方程。范围 $-\frac{3\pi}{2} < \theta < \frac{\pi}{2}$ 变成 $0 < \psi < 2\pi$，正好就是前面的范围。所以 $\psi$ 取上面求得的六个值，而由 $\theta = \frac{\pi}{2} - \psi$ 得

$$\theta = \frac{\pi}{4}, \frac{\pi}{8}, -\frac{3\pi}{8}, -\frac{3\pi}{4}, -\frac{7\pi}{8}, -\frac{11\pi}{8}.$$

## 验算

- 当 $\theta = \frac{\pi}{4}$：$\tan 3\theta = \tan\frac{3\pi}{4} = -1$，而 $2 + (-1) = 1 = \tan\frac{\pi}{4}$。当 $\theta = \frac{3\pi}{8}$：$\tan\frac{9\pi}{8} = \tan\frac{\pi}{8} = \sqrt{2} - 1$，而 $2 + \sqrt{2} - 1 = 1 + \sqrt{2} = \tan\frac{3\pi}{8}$。
- $t$ 的三次方程最多有三个根，而 tan 的周期为 $\pi$，所以在长度为 $2\pi$ 的区间内，我们预期最多有 $3 \times 2 = 6$ 个解。我们找到了六个。
- 以数值方法，在 $(0, 2\pi)$ 上用细密的格点扫描 $\tan x - \tan 3x - 2$，发现（在远离极点（pole）之处）只在 $x = \frac{2\pi}{8}, \frac{3\pi}{8}, \frac{7\pi}{8}, \frac{10\pi}{8}, \frac{11\pi}{8}, \frac{15\pi}{8}$ 出现变号。对余切方程作同样的扫描，找到的正好就是上面列出的六个值。

## 常见错误

- 在 $\tan 2\theta = -1$ 这一步只取 $n = 0, 1$ 就停下来。由于 $\theta$ 的取值范围是 $(0, 2\pi)$，$2\theta$ 的取值范围就是 $(0, 4\pi)$，这样会得到四个解，而不是两个。
- 未经检查就除以某个式子，因而漏掉根 $t = 1$；或者忘记了 $\frac{5\pi}{4}$。
- 在最后一部分从头解起,并把范围弄错。这个代换把题目所给的区间恰好映射到 $(0, 2\pi)$ 上，而这绝非巧合。

## STEP 之外

恒等式（identity）$\tan 3\theta = (3t - t^3)/(1 - 3t^2)$ 其实就是 $(1 + \mathrm{i}t)^3$ 的实部与虚部；一般而言，$\tan n\theta$ 是 $\tan\theta$ 的两个多项式之比，可以从 $(1 + \mathrm{i}t)^n$ 直接读出。这些多项式的根是 $\pi$ 的有理数倍的正切值，这就是为甚么它们的因式分解（factorisation），例如这里的 $t^2 - 2t - 1$，会如此整齐。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）展示了倍角公式（multiple-angle formulae）如何从复数乘法的几何中自然得出。

详细解答

# STEP II 1997，第 7 题

纯数学 · 曲线描绘

> **STEP II 1997，第 7 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s2-q7

## 关键思路

右边只通过 $x^2$ 依赖于 $x$，因此可把它看成 $u = x^2$ 的二次式，然后配方（complete the square）：这样一行就得出 $y^2$ 的最大可能值，完全不需要微积分。至于作图，把 $x \to -x$ 或 $y \to -y$ 时方程不变，所以只需弄清楚第一象限的情况，再作反射即可。值得确定的地方有：曲线与两轴相交之处、在原点的斜率，以及曲线在原点以外与 $x$ 轴相交之处的斜率。

## 解答

### 最大值与最小值

若 $a = 0$，方程变成 $y^2 = -x^4$，这迫使 $x = y = 0$：这条“曲线”只是单独一点 $(0,0)$，而 $y$ 唯一的值是 0。以下设 $a \neq 0$。方程中只出现 $a^2$，所以不妨设 $a > 0$，到最后才写成 $|a|$。

由于 $y^2 \geq 0$，我们需要 $x^2(a^2 - x^2) \geq 0$，即 $x = 0$ 或 $x^2 \leq a^2$。所以曲线位于带状区域 $-|a| \leq x \leq |a|$ 之内。对 $x^2$ 配方，

$$y^2 = a^2x^2 - x^4 = \frac{a^4}{4} - \left(x^2 - \frac{a^2}{2}\right)^2 \leq \frac{a^4}{4},$$

等号成立当且仅当 $x^2 = \frac{1}{2}a^2$。因此在曲线上处处有 $|y| \leq \frac{1}{2}a^2$，而 $y = \frac{1}{2}a^2$ 和 $y = -\frac{1}{2}a^2$ 两个值都能取得，取于 $x = \pm|a|/\sqrt{2}$。所以

$$\boxed{y_{\max} = \frac{a^2}{2}, \qquad y_{\min} = -\frac{a^2}{2}, \qquad \text{两者皆取于} x = \pm\frac{|a|}{\sqrt{2}}.}$$

**想一想** 这与微积分的结果是否吻合？$\mathrm{d}y/\mathrm{d}x = 0$ 在哪里成立？

作隐函数求导（implicit differentiation），$2y\,\dfrac{\mathrm{d}y}{\mathrm{d}x} = 2a^2x - 4x^3 = 2x(a^2 - 2x^2)$。在 $y \neq 0$ 之处，斜率为零当且仅当 $x^2 = \frac{1}{2}a^2$（根 $x = 0$ 对应 $y = 0$），这正是上面找到的四个点。$y = 0$ 的点需要另行小心处理，因为在那里不能除以 $y$；我们接着就处理它们。两种方法结果一致，但配方能即时证明这些驻值（stationary values）就是整体的最大值和最小值。

### 曲线的形状

取 $a > 0$。解出 $y$ 得到两个分支，

$$y = x\sqrt{a^2 - x^2} \qquad \text{及} \qquad y = -x\sqrt{a^2 - x^2}, \qquad -a \leq x \leq a,$$

两者互为对 $x$ 轴的反射。在第一个分支上，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \sqrt{a^2 - x^2} - \frac{x^2}{\sqrt{a^2 - x^2}} = \frac{a^2 - 2x^2}{\sqrt{a^2 - x^2}}.$$

- **在原点**，斜率是 $a$，而在另一个分支上则是 $-a$。所以曲线两次经过原点，并在那里与自身相交，切线为 $y = \pm ax$。
- **在** $x = \pm a$，我们有 $y = 0$，而当 $x \to a^-$ 时斜率趋向 $-\infty$。所以曲线在 $(\pm a, 0)$ 与 $x$ 轴相交，切线是铅垂的；曲线在那里平滑而圆浑，并非尖角。
- **在两者之间**，第一个分支由 $(0,0)$ 上升至 $x = a/\sqrt{2}$ 处的最大值 $\frac{1}{2}a^2$，再下降回到 $(a, 0)$。其余象限的情况可由对两轴的对称性得出。

结果是一个横卧的 8 字形，在 $x \geq 0$ 和 $x \leq 0$ 各有一个环，阔度为 $2|a|$，总高度为 $a^2$。

要一次过看清整条曲线，有一个简洁的方法：设 $x = a\sin t$。那么 $a^2 - x^2 = a^2\cos^2 t$，所以 $y^2 = a^4 \sin^2 t \cos^2 t$，而曲线可由

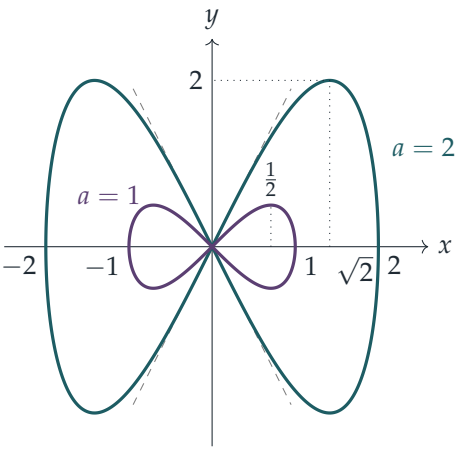
$$x = a\sin t, \qquad y = \tfrac{1}{2}a^2 \sin 2t, \qquad 0 \leq t < 2\pi,$$

描出，这再次显示 $|y| \leq \frac{1}{2}a^2$。

### $a = 1$ 与 $a = 2$ 的情况

| | 与 $x$ 轴相交于 | 在 $O$ 的切线 | 极值点 |
|---|---|---|---|
| $a = 1$ | $x = 0,\ \pm 1$ | $y = \pm x$ | $\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{2}\right)$ |
| $a = 2$ | $x = 0,\ \pm 2$ | $y = \pm 2x$ | $\left(\pm\sqrt{2}, \pm 2\right)$ |

要确定两条曲线的相对位置：对于 $0 < |x| \leq 1$，我们有 $x^2(1 - x^2) < x^2(4 - x^2)$，所以 $a = 1$ 的曲线严格位于 $a = 2$ 的曲线之内，唯一的公共点是原点。例如，在 $x = 1$ 处，$a = 1$ 的曲线位于 $y = 0$，而 $a = 2$ 的曲线则位于 $y = \pm\sqrt{3}$。




虚线是 $a = 2$ 时的切线 $y = \pm 2x$；当 $a = 1$ 时，在原点的切线是 $y = \pm x$，倾角为 45°。

## 验算

- 对 50 个随机的 $a$ 值，在 $|x| \le |a|$ 上作网格搜索，每次找到的最大 $|y|$ 都是 $\frac{1}{2}a^2$，位于 $|x| = |a|/\sqrt{2} \approx 0.7071|a|$。当 $a = 2$ 时，结果是 $y_{\max} = 2.000000$，位于 $|x| \approx 1.41421$。
- 最大值按 $a^2$ 的比例变化，这是必然的：把 $(x, y)$ 换成 $(ax, a^2y)$，就把 $a = 1$ 的曲线变成一般的曲线。所以 $a = 2$ 的曲线就是把 $a = 1$ 的曲线沿水平方向拉伸 2 倍、沿铅垂方向拉伸 4 倍而成，这就是它高出那么多的原因。

## 常见错误

- 只画出 $y = x\sqrt{a^2 - x^2}$。方程定义的是 $y^2$，所以 $y$ 的正负两个值都属于图像，得出的是两个环，而不是 S 形。
- 在 $(\pm a, 0)$ 画出尖角。那里的斜率是无限大，所以曲线是铅垂而平滑地穿过 $x$ 轴的。
- 忘记曲线在原点与自身相交，斜率为 $\pm a$。正是这些斜率让我们在 $O$ 附近分辨出两条曲线。

## STEP 之外

这里的原点是一个结点（node），即代数曲线（algebraic curve）与自身相交的点：在 $O$ 附近，方程看起来就像 $y^2 \approx a^2x^2$，即一对直线 $y = \pm ax$。像这样按最低次项把奇点（singular point）分类，是研究代数曲线的第一步。$x = a\sin t$ 这个技巧显示这条曲线可以用三角函数参数化（parametrise）；这类曲线称为利萨茹图形（Lissajous figures）（这

里的频率比为 1 : 2）。若想多练习描绘这类曲线，以及更广泛地练习 STEP 的思考方式，可阅读 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

# STEP II 1997，第 8 题

纯数学 · 以积分证明不等式

> **STEP II 1997，第 8 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s2-q8

## 关键思路

每个形如 $(x^p-1)/p$ 的式子，其实都是乔装了的积分（integral）：它就是 $\int_1^x t^{p-1}\,\mathrm{d}t$。因此，比较两个这样的式子，归根究底就是比较被积函数（integrand）$t^{p-1}$ 与 $t^{q-1}$；只要留意到两者的大小关系在 $t=1$ 处对调，这个比较便很容易。至于最后的不等式（inequality），只需再积分一次，这次是对刚证明的不等式积分。

## 解答

### 对不等式积分

若在 $[a,b]$ 上有 $\mathrm{f}(t)\geq \mathrm{g}(t)$，则在该区间上 $\mathrm{f}-\mathrm{g}\geq 0$，所以 $\mathrm{f}-\mathrm{g}$ 的图像与 $t$-轴之间的区域位于轴上或轴的上方，其有向面积（signed area）为非负：

$$\int_a^b \mathrm{f}(t)\,\mathrm{d}t-\int_a^b \mathrm{g}(t)\,\mathrm{d}t=\int_a^b \big(\mathrm{f}(t)-\mathrm{g}(t)\big)\,\mathrm{d}t\geq 0.$$

注意这里需要 $a\leq b$：若上下限的次序颠倒，两个积分之间的不等号便会反转。

### $x\geq 1$ 的情况

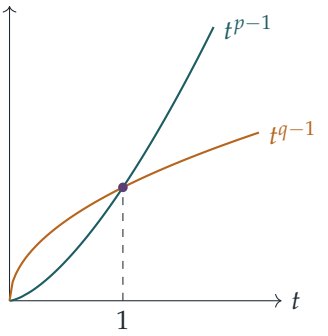
当 $t\geq 1$ 且 $p>q$ 时，有 $t^{p-1}=t^{q-1}\,t^{p-q}\geq t^{q-1}$，因为当 $t\geq 1$ 且 $p-q>0$ 时 $t^{p-q}\geq 1$。由于 $x\geq 1$，区间 $[1,x]$ 的上下限次序正确，所以由第一部分可得

$$\int_1^x t^{p-1}\,\mathrm{d}t\;\geq\;\int_1^x t^{q-1}\,\mathrm{d}t.$$

因为 $p,q>0$，两个积分都可以直接计算：

$$\int_1^x t^{p-1}\,\mathrm{d}t=\Big[\frac{t^p}{p}\Big]_1^x=\frac{x^p-1}{p},\qquad \int_1^x t^{q-1}\,\mathrm{d}t=\frac{x^q-1}{q}.$$

因此对 $x \geq 1$ 有 $\dfrac{x^p-1}{p} \geq \dfrac{x^q-1}{q}$，正如所求。




这幅草图（以 $p = 2.5$、$q = 1.5$ 绘画）说明了要点：当 $t > 1$ 时，次数较高的幂较大；当 $0 < t < 1$ 时则较小。

### $0 \leq x \leq 1$ 的情况

**想一想** 当 $0 < t < 1$ 时，$t^{p-1}$ 与 $t^{q-1}$ 之间的不等号方向相反。为甚么结论没有跟着反转?

两次反转互相抵消。当 $0 < t \leq 1$ 时，有 $t^{p-q} \leq 1$，所以 $t^{p-1} \leq t^{q-1}$。但现在 $x \leq 1$，所以次序正确的区间是 $[x, 1]$，而不是 $[1, x]$。

先考虑 $0 < x \leq 1$。在 $[x, 1]$ 上对 $t^{q-1} \geq t^{p-1}$ 积分（由于 $x > 0$，两者在该区间上都连续 (continuous)），得

$$\int_x^1 t^{q-1}\,\mathrm{d}t \;\geq\; \int_x^1 t^{p-1}\,\mathrm{d}t, \qquad \text{即} \qquad \frac{1-x^q}{q} \;\geq\; \frac{1-x^p}{p}.$$

两边乘以 $-1$，不等号反转，得 $\dfrac{x^p-1}{p} \geq \dfrac{x^q-1}{q}$。

当 $x = 0$ 时，我们直接验证：不等式变成 $-\frac{1}{p} \geq -\frac{1}{q}$，这是成立的，因为由 $p > q > 0$ 可得 $\frac{1}{p} < \frac{1}{q}$。(我们把 $x = 0$ 分开处理，是因为若 $q < 1$，则 $t^{q-1}$ 在 0 附近无界 (unbounded)。) 所以不等式对所有 $0 \leq x \leq 1$ 成立；连同上一个情况，对每个 $x \geq 0$ 都成立。

### 最后的不等式

记 $\phi_p(t) = \dfrac{t^p-1}{p}$。我们已经证明：只要 $p > q > 0$，对每个 $t \geq 0$ 都有 $\phi_p(t) \geq \phi_q(t)$。两个函数在 $[0, \infty)$ 上都连续，所以当 $x > 0$ 时，我们可以在 $[0, x]$ 上对这个不等式积分：

$$\int_0^x \frac{t^p-1}{p}\,\mathrm{d}t \;\geq\; \int_0^x \frac{t^q-1}{q}\,\mathrm{d}t.$$

而

$$\int_0^x \frac{t^p-1}{p}\,\mathrm{d}t = \frac{1}{p}\Big(\frac{x^{p+1}}{p+1}-x\Big) = \frac{x}{p}\Big(\frac{x^p}{p+1}-1\Big),$$

对 $q$ 亦同样。所以

$$\frac{x}{p}\Big(\frac{x^p}{p+1}-1\Big) \;\geq\; \frac{x}{q}\Big(\frac{x^q}{q+1}-1\Big),$$

两边除以 $x>0$（不等号方向不变），得

$$\frac{1}{p}\Big(\frac{x^p}{p+1}-1\Big) \;\geq\; \frac{1}{q}\Big(\frac{x^q}{q+1}-1\Big).$$

在 $x=0$ 处，这个不等式变成 $-\frac{1}{p}\geq -\frac{1}{q}$，我们已经知道它成立。至此，对所有 $x\geq 0$ 的证明完成。

## 验算

- 在 $x=1$ 处，中间的不等式是 $0\geq 0$：两条曲线在该处相交，与等号成立吻合。
- 在最后的不等式中取 $p=2$、$q=1$、$x=3$：左边是 $\frac{1}{2}(3-1)=1$，右边是 $\frac{3}{2}-1=\frac{1}{2}$，所以 $1\geq\frac{1}{2}$。
- 随机测试了 200,000 组满足 $0<q<p$ 及 $x\geq 0$ 的三元组 $(p,q,x)$，当中包括 $x=0$、$x=1$ 及极小的 $x$，两个不等式都找不到反例（counterexample）。

## 常见错误

- 当 $x<1$ 时在 $[1,x]$ 上积分，却保持不等号方向不变：对不等式积分之前，上下限必须按递增次序排列。
- 忘记端点 $x=0$，在该处 $t^{q-1}$ 甚至可能没有定义。
- 试图用微积分从头证明最后一部分，而不是对刚建立的不等式积分。

## STEP 之外

当 $q\to 0$ 时，$(x^q-1)/q\to\ln x$，所以这个结果表明：当 $p$ 递减至 0 时，$(x^p-1)/p$ 递减至 $\ln x$。这是对数的一族精巧的界，也是统计学中 Box–Cox 变换（Box–Cox transformation）的基础。积分保持不等式这个原理（积分的单调性，monotonicity of the integral），是分析（analysis）课程中关于黎曼积分（Riemann integral）最先证明的定理之一。参见 Abbott, *Understanding Analysis*。

详细解答

# STEP II 1997，第 9 题

力学 · 摩擦力与刚体平衡

**STEP II 1997，第 9 题**

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

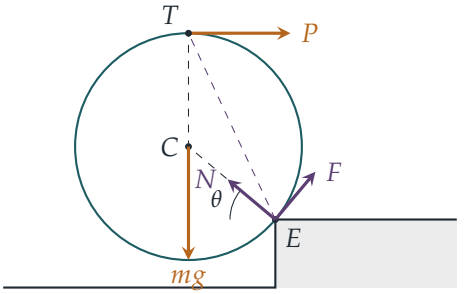
step.maths.org/questions/97-s2-q9

## 关键思路

由于球是被缓慢地拉上去的，它在每一瞬间都处于平衡（equilibrium）状态，所以在每个位置我们面对的都是一道静力学（statics）问题。球一旦离开地面，就只有三个力作用在它身上：重量、顶点处的拉力，以及台阶边缘的反作用力（reaction）。对边缘取力矩（moments），便可消去未知的反作用力，从而求出拉力。然后，三力定则（three-force rule，即三条作用线（lines of action）必须交于一点）确切地告诉我们反作用力指向哪个方向，而摩擦力（friction）的条件也就几乎唾手可得。

## 解答

设半径为 $r = \frac{1}{2}d$，则台阶的高度为 $\frac{1}{4}d = \frac{1}{2}r$。我们把球看成均匀的刚体（rigid body），它只在边缘 $E$ 处与台阶接触（边缘是一条水平线，在通过球心的铅垂截面中是一个点）。设 $C$ 为球心，$T$ 为最高点，水平力 $P$ 作用于该点，方向指向台阶。设 $\theta$ 为 $EC$ 与水平线所成的角。



### 几何关系

以 $E$ 为原点，$C$ 比 $E$ 高出 $r\sin\theta$，与它的水平距离为 $r\cos\theta$，而 $T$ 则在 $C$ 之上再高出 $r$。开始时，球静止在地面上，$C$ 离地面的高度为 $r$，即比 $E$ 高出 $\frac{1}{2}r$；因此 $\sin\theta = \frac{1}{2}$，$\theta = \frac{\pi}{6}$。当 $C$ 位于 $E$ 的正上方，即 $\theta = \frac{\pi}{2}$ 时，球便完全登上台阶。所以整个运动过程对应于 $\theta$ 由 $\frac{\pi}{6}$ 增加至 $\frac{\pi}{2}$。

### 最大的力

当球仍然接触地面时，地面承担了部分重量，而地面只能向上推。当 $P$ 刚好大得足以令地面的反作用力降至零时，球便开始上升。从这时起，作用于球的力是：在 $T$ 处的 $P$、在 $C$ 处的重量 $mg$，以及边缘在 $E$ 处的反作用力，后者有沿 $EC$ 方向的法向分量（normal component）$N$ 和垂直于 $EC$ 的摩擦力 $F$。

对 $E$ 取力矩（反作用力对该点没有力矩），$P$ 的力臂是 $T$ 高出 $E$ 的高度，即 $r(1+\sin\theta)$，而重量的力臂是 $r\cos\theta$。由平衡条件得

$$P\,r(1+\sin\theta) = mg\,r\cos\theta \quad\Longrightarrow\quad P = \frac{mg\cos\theta}{1+\sin\theta}.$$

在 $\frac{\pi}{6} \le \theta \le \frac{\pi}{2}$ 上，分子递减而分母递增，所以 $P$ 随 $\theta$ 递减。（在离地之前，$P$ 只是逐渐增大至这个值。）因此最大值出现在 $\theta = \frac{\pi}{6}$，即球离开地面的一刻：

$$P_{\max} = \frac{mg\cdot\frac{\sqrt{3}}{2}}{1+\frac{1}{2}} = \frac{mg}{\sqrt{3}}.$$

### 摩擦力的条件

> **想一想** 边缘处的反作用力指向哪个方向?

三个互不平行的力若处于平衡，它们的作用线必定共点（concurrent）。$P$ 的作用线是通过 $T$ 的水平线，而重量的作用线是通过 $C$ 的铅垂线，这条线同样通过 $T$。所以两者交于 $T$，而 $E$ 处的总反作用力必定沿 $ET$ 作用。

现在 $CE = CT = r$，所以三角形 $CET$ 是等腰三角形。在 $C$ 处，$CT$（铅垂向上）与 $CE$（向下，与水平线成 $\theta$ 角）之间的夹角是 $\frac{\pi}{2}+\theta$，所以每个底角是

$$\angle CET = \tfrac{1}{2}\left(\pi - \tfrac{\pi}{2} - \theta\right) = \tfrac{\pi}{4} - \tfrac{\theta}{2}.$$

这就是总反作用力与法线 $EC$ 之间的夹角，所以

$$\frac{F}{N} = \tan\left(\frac{\pi}{4} - \frac{\theta}{2}\right).$$

我们可以用分解（resolving）力的方法来验证这个结果。反作用力是 $(-P,\ mg)$，因为它要抵消 $P$ 和重量。它沿单位法向量 $(-\cos\theta, \sin\theta)$ 和沿单位切向量 $(\sin\theta, \cos\theta)$ 的分量

分别是

$$N = P\cos\theta + mg\sin\theta = mg\,\frac{\cos^2\theta + \sin\theta + \sin^2\theta}{1+\sin\theta} = mg,$$
$$F = mg\cos\theta - P\sin\theta = mg\cos\theta\left(1 - \frac{\sin\theta}{1+\sin\theta}\right) = \frac{mg\cos\theta}{1+\sin\theta},$$

所以 $F/N = \cos\theta/(1+\sin\theta)$,而根据半角公式(half-angle formulae),这等于 $\tan(\frac{\pi}{4} - \frac{\theta}{2})$。在整个过程中 $F$ 和 $N$ 都是正数，所以边缘确实是在推着球，而摩擦力沿切线向上作用，正如图中所画。

只有当在每个位置都有 $F \le \mu N$ 时，球才不会在 $E$ 处滑动，即

$$\mu \ge \tan\left(\frac{\pi}{4} - \frac{\theta}{2}\right) \quad \text{对所有} \frac{\pi}{6} \le \theta \le \frac{\pi}{2}.$$

右边随 $\theta$ 增加而递减，所以要求最严格的位置同样是 $\theta = \frac{\pi}{6}$，这时它等于 $\tan\frac{\pi}{6} = \frac{1}{\sqrt{3}}$。在那一瞬间，摩擦力必须至少是法向反作用力的 $1/\sqrt{3}$ 倍；若 $\mu = 1/\sqrt{3}$ 恰好成立，摩擦力便达到极限（limiting），球正处于即将滑动的状态，所以要安全地把球拉上去而不滑动，我们需要 $\mu > 1/\sqrt{3}$，这正是所要求的结果。

## 验算

- 当 $\theta = \frac{\pi}{2}$ 时，球心在边缘的正上方：$P = 0$ 而且 $F = 0$，理应如此。
- 整个过程中 $N = mg$，这是合理的：反作用力是 $(-P, mg)$，而它沿法线方向的分量算出来恰好是 $mg$。
- 以数值方法验算：让 $\theta$ 由 $\frac{\pi}{6}$ 逐步增加至 $\frac{\pi}{2}$，分别用取力矩和分解力的方法计算，$P/mg$ 和 $F/N$ 的最大值都是 $0.57735\ldots = 1/\sqrt{3}$，两者都出现在起始位置。

## 常见错误

- 把 $P$ 对 $E$ 的力臂当作 $r\sin\theta$（$C$ 的高度），而不是 $r(1+\sin\theta)$（$T$ 的高度）。
- 假设边缘处的反作用力纯粹沿法线方向。这是不可能的：法线通过 $C$，这样三个力便不能交于一点。

## STEP 之外

三力的论证是一个普遍事实的特例：刚体处于平衡，当且仅当合力以及对任意一点的合力矩都等于零，而对于三个力的情况，这就迫使它们共点。反作用力与法线之间的夹角

（这里是 $\frac{\pi}{4} - \frac{\theta}{2}$）就是“摩擦角”（angle of friction）的概念，它把摩擦力问题转化为几何问题。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教材）从这些原理出发，建立刚体的静力学和动力学。

详细解答

# STEP II 1997，第 10 题

力学 · 爆炸与弹性碰撞

> **STEP II 1997，第 10 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s2-q10

## 关键思路

把每个速度分解成沿轴的分量和垂直于轴的分量。圆柱面是光滑的，所以它只能沿径向施力：它永远不会改变轴向速度，而弹性反弹只是把横向速度反转。爆炸时的动量守恒（conservation of momentum）迫使两块碎片沿同一条直径飞开，于是整个问题就变成两面墙之间的一维运动，轴向运动只是顺带跟着走。

## 解答

### 建立模型

把碎片当作质点（particle）（它们很小），圆柱视为光滑而固定，并按题目指示忽略重力。取 $z$ 轴沿圆柱的轴，并设爆炸在时间 $t = 0$ 发生于原点。

紧接爆炸之后，把质量为 $qM$ 的碎片的速度写成轴向分量 $w_1$ 加上横向分量 $\mathbf{u}_1$（垂直于轴），质量为 $(1-q)M$ 的碎片则同样写成 $w_2$、$\mathbf{u}_2$。爆炸期间没有外力作用，所以动量守恒：

$$qMw_1 + (1-q)Mw_2 = Mv, \qquad qM\mathbf{u}_1 + (1-q)M\mathbf{u}_2 = \mathbf{0}.$$

第二条方程说明两个横向速度方向相反，所以两块碎片沿同一条直径、以相反方向离开轴。把这条直径称为 $x$ 方向，碎片 1（质量 $qM$）以速率 $u_1$ 向 $x = +L$ 运动，碎片 2 以速率 $u_2$ 向 $x = -L$ 运动，其中

$$qu_1 = (1-q)u_2. \tag{1}$$

由于 $q \le \frac{1}{2}$，我们有 $u_1 \ge u_2$：较轻的碎片较快。

每块碎片都是正面撞向墙壁（它沿半径运动），所以光滑墙壁的冲量（impulse）沿该半径方向。这会把横向速度反转（反弹是完全弹性的），而轴向速度则不受影响。因此每块

碎片一直沿同一条直径运动，而它的轴向速度自始至终保持不变。

**证明** $q = \frac{3}{8}$

**轴向运动**。轴向位置为 $z_1 = w_1 t$ 和 $z_2 = w_2 t$。两块碎片在某个时间 $t > 0$ 相遇，所以 $w_1 = w_2$，再由动量方程得 $w_1 = w_2 = v$。两块碎片都与原来的物体本应到达的位置并肩前进。

**横向运动**。碎片 1 在 $t = L/u_1$ 到达 $x = L$，然后折返，所以反弹之后

$$x_1 = L - u_1\left(t - \frac{L}{u_1}\right) = 2L - u_1 t.$$

同理，碎片 2 反弹之后位于 $x_2 = -2L + u_2 t$。它们在两者都反弹之后才碰撞，所以由 $x_1 = x_2$ 得

$$t = \frac{4L}{u_1 + u_2}, \qquad x = 2L - \frac{4Lu_1}{u_1 + u_2} = 2L\,\frac{u_2 - u_1}{u_1 + u_2}.$$

由 (1)，$u_2 = \frac{q}{1-q}u_1$，所以 $u_1 + u_2 = \frac{u_1}{1-q}$ 及 $u_2 - u_1 = \frac{(2q-1)u_1}{1-q}$。因此碰撞点为

$$x = 2L(2q - 1).$$

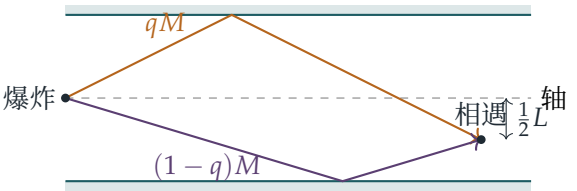
我们需要 $|x| = \frac{1}{2}L$，所以 $2q - 1 = \pm\frac{1}{4}$，得 $q = \frac{5}{8}$ 或 $q = \frac{3}{8}$。由于 $q \le \frac{1}{2}$，

$$q = \tfrac{3}{8},$$

正如所求。碰撞发生在 $x = -\frac{1}{2}L$，位于轴的较重碎片那一侧。

**想一想** 这真的是第一次相遇，而且两者都刚好各反弹了一次吗?

当 $q = \frac{3}{8}$ 时，由 (1) 得 $u_2 = \frac{3}{5}u_1$。碎片 1 在 $t = L/u_1$ 反弹，碎片 2 在 $t = \frac{5}{3}L/u_1$ 反弹，而碰撞时间为 $4L/(u_1 + u_2) = \frac{5}{2}L/u_1$，比两者都迟。它们会否更早相遇？在碎片 1 反弹之前，两者正在互相远离；而碎片 1 在回程中要到 $t = 2L/(u_1 - u_2) = 5L/u_1$ 才会追上碎片 2，那时碎片 2 早已掉头。所以我们找到的相遇就是第一次相遇，而每块碎片都刚好反弹了一次。




图中所示是包含轴和该公共直径的平面（轴向为横向）；此图按 $q = \frac{3}{8}$ 及下文求得的速率依比例绘画。

### 爆炸所给予的能量

碰撞时间为

$$t = \frac{4L}{u_1 + u_2} = \frac{4L(1-q)}{u_1} = \frac{5L}{2u_1}.$$

令它等于 $5L/v$，得 $u_1 = \frac{1}{2}v$，于是 $u_2 = \frac{3}{5}u_1 = \frac{3}{10}v$。

轴向速度维持在 $v$ 不变，所以爆炸中所增加的动能（kinetic energy）纯粹是横向动能：

$$\begin{aligned}\Delta E &= \tfrac{1}{2}\cdot\tfrac{3}{8}M\left(\tfrac{1}{2}v\right)^2 + \tfrac{1}{2}\cdot\tfrac{5}{8}M\left(\tfrac{3}{10}v\right)^2 \\ &= \tfrac{1}{2}Mv^2\left(\tfrac{3}{32} + \tfrac{9}{160}\right) = \tfrac{1}{2}Mv^2\cdot\tfrac{3}{20} = \frac{3Mv^2}{40}.\end{aligned}$$

### 合并后的速度

紧接碰撞之前，每块碎片的横向速度都已因反弹而反转：碎片 1 的横向动量为 $-qMu_1$，碎片 2 的为 $+(1-q)Mu_2$。由 (1)，两者之和为零。轴向动量为 $qMv + (1-q)Mv = Mv$。合并（coalescence）过程中动量守恒，所以质量为 $M$ 的合并体的速度为

$$v，\text{沿轴的方向},$$

与原来的物体相同，但它沿着一条与轴平行、距离轴 $\frac{1}{2}L$ 的直线运动。来自爆炸的 $\frac{3}{40}Mv^2$ 能量在合并中全部损失。

## 验算

- $q = \frac{1}{2}$ 给出 $x = 0$：两块相等的碎片互为镜像，必定在轴上相遇，合乎预期。
- 取 $L = v = 1$，$u_1 = 0.5$ 及 $u_2 = 0.3$：当 $t = 5$ 时，碎片 1 位于 $2 - 0.5\times 5 = -0.5$，碎片 2 位于 $-2 + 0.3\times 5 = -0.5$。一个在 $x = \pm 1$ 处设有反射的逐步时间模拟，重现了两者各反弹一次后在 $t = 5.000$、$x = -0.500$ 相遇的结果。
- 最终的横向动量为 $-\frac{3}{8}\cdot\frac{1}{2} + \frac{5}{8}\cdot\frac{3}{10} = 0$，所以最终速度纯粹是轴向的。

## 常见错误

- 让墙壁改变了轴向速度。光滑的墙壁只会施加法向（径向）力。
- 忘记反弹会把横向动量反转。总横向动量要待两块碎片都反弹之后才回复为零。
- 保留 $q = \frac{5}{8}$：那是把两块碎片的标号对调后的同一个情况，已被 $q \le \frac{1}{2}$ 排除。

## STEP 之外

把运动分解成沿对称方向和垂直于对称方向的分量，是一个重要思想的开端：系统的一种对称性（symmetry）会带来一个守恒量（conserved quantity）。这里圆柱沿其长度处处看来都一样，所以即使有墙壁存在，轴向动量仍然守恒（诺特定理（Noether's theorem）把这一点说得精确）。本题亦显示，爆炸之后接着一次完全非弹性碰撞（perfectly inelastic collision），可以把能量以热的形式交还，而轴向运动却丝毫不受影响，照样继续。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）正是以这种方式阐述动量和质心（centre of mass）运动。

详细解答

# STEP II 1997，第 11 题

力学 • 抛体运动

**STEP II 1997，第 11 题**

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

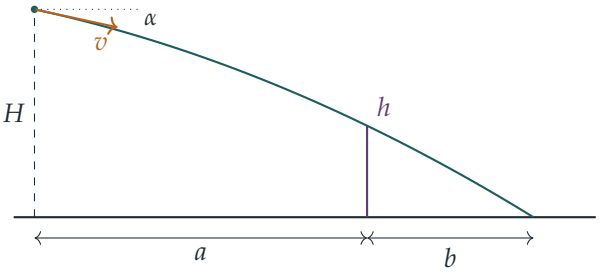
step.maths.org/questions/97-s2-q11

## 关键思路

先把轨迹（trajectory）写成一条方程，以水平移动距离表示球比发球者击球高度下降了多少。这样，题目给出的两项事实（球水平移动了 $a$ 时高度为 $h$，水平移动了 $a+b$ 时高度为 0）就是关于两个未知数 $v$ 和 $\alpha$ 的两条方程。至于上下界，则来自这个问题：这两条方程的解在甚么时候才真正具有物理意义？

## 解答

我们把球看成只受重力作用的质点（particle），即不考虑空气阻力（air resistance）和旋转。取击球点正下方的地面为原点，$x$ 轴沿水平方向指向球网，$y$ 轴铅垂向上。




### $v^2$ 的公式

初速度为 $(v\cos\alpha, -v\sin\alpha)$。经过时间 $t$ 后，

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

以 $t = x/(v\cos\alpha)$ 消去时间，并利用 $\sec^2\alpha = 1 + \tan^2\alpha$，得

$$y = H - x\tan\alpha - \frac{gx^2(1+\tan^2\alpha)}{2v^2}. \qquad (*)$$

为方便起见，记 $T = \tan\alpha$ 及 $k = \dfrac{g(1+T^2)}{2v^2}$，于是 $y = H - xT - kx^2$。

球在 $x = a + b$ 处着地，所以在该处代入 $y = 0$：

$$0 = H - (a+b)T - k(a+b)^2 \quad\Longrightarrow\quad k = \frac{H-(a+b)T}{(a+b)^2}. \qquad (1)$$

把 $k$ 换回它的定义，再整理成 $v^2$ 的表达式，得

$$v^2 = \frac{g(a+b)^2(1+\tan^2\alpha)}{2[H-(a+b)\tan\alpha]},$$

这就是第一个结果。

### $\tan\alpha$ 的公式

球刚好越过球网，所以在 $x = a$ 处 $y = h$。由 $(*)$ 得

$$H - h = aT + ka^2.$$

代入 $(1)$ 中的 $k$，并将全式乘以 $(a+b)^2$：

$$\begin{aligned}(H-h)(a+b)^2 &= aT(a+b)^2 + a^2\big[H-(a+b)T\big]\\ &= aT(a+b)\big[(a+b)-a\big] + a^2H\\ &= abT(a+b) + a^2H.\end{aligned}$$

把含 $H$ 的项合并，并利用 $(a+b)^2 - a^2 = b(2a+b)$，得

$$ab(a+b)T = b(2a+b)H - (a+b)^2h.$$

两边除以 $ab(a+b)$（此数为正），得

$$\tan\alpha = \frac{2a+b}{a(a+b)}H - \frac{a+b}{ab}h,$$

正是所求。

### $H$ 的范围

给定 $a$、$b$、$h$ 和 $H$ 后，第二条公式决定了 $\tan\alpha$，然后第一条公式决定了 $v^2$。当且仅当这些值都合理时，这样的发球才存在：即 $\tan\alpha > 0$（球是向下击出的，所以 $0 < \alpha < \frac{\pi}{2}$）以及 $v^2 > 0$。

**想一想** $v^2$ 的分子显然是正数。那么 $v^2 > 0$ 实际上要求甚么?

它要求分母为正：$H - (a+b)\tan\alpha > 0$。代入 $\tan\alpha$ 的公式，

$$\begin{aligned} H - (a+b)\tan\alpha &= H - \frac{2a+b}{a}H + \frac{(a+b)^2}{ab}h \\ &= -\frac{a+b}{a}H + \frac{(a+b)^2}{ab}h = \frac{a+b}{a}\left[\frac{(a+b)h}{b} - H\right]. \end{aligned}$$

由于 $(a+b)/a > 0$，上式为正当且仅当

$$H < \frac{(a+b)h}{b}.$$

要使 $\tan\alpha > 0$，我们需要

$$\frac{2a+b}{a(a+b)}H > \frac{a+b}{ab}h \quad\Longleftrightarrow\quad H > \frac{(a+b)^2 h}{b(2a+b)},$$

这里两边乘以了正数 $a(a+b)/(2a+b)$。

两个条件合起来就是

$$\boxed{\frac{(a+b)^2}{b(2a+b)}h \;<\; H \;<\; \frac{a+b}{b}h}$$

反过来，在这个范围内的任何 $H$ 都给出合理的 $\alpha$ 和 $v$。这个区间非空，因为 $\dfrac{a+b}{2a+b} < 1$。

两个界限都有清晰的意义。上界是这样的一个高度：由着地点经过网顶的直线，与发球者所在的铅垂线相交之处。重力使路径向下弯曲，所以球必须从这条直线下方出发（在 $H = (a+b)h/b$ 时，我们需要无限大的速率）。下界则对应 $\alpha = 0$ 的情况，即水平发球；若从比这更低的位置发球，球就必须向上击出，才能越过球网而仍然落在指定的着地点。

## 验算

- 取 $a = 12\,\text{m}$、$b = 6\,\text{m}$ 及 $h = 1\,\text{m}$，范围为 $1.8\,\text{m} < H < 3\,\text{m}$。当 $H = 1.8$ 时，公式给出 $\tan\alpha = \frac{30}{216}(1.8) - \frac{18}{72} = 0$，与下界应有的结果一致。
- 在范围内随机选取 $a$、$b$、$h$ 和 $H$，用两条公式算出 $\alpha$ 和 $v$，再以数值方法逐步模拟运动，结果球每次都经过网顶，并落在 $x = a + b$ 处（就精确路径而言，相对误差低于 $10^{-14}$）。
- 量纲（dimensions）：$\tan\alpha$ 的每一项都是长度除以长度，而 $v^2$ 的量纲是 $g\times$ 长度，正如所料。

## 常见错误

- 因角度向下而弄错正负号：初始铅垂速度是 $-v\sin\alpha$，而 $(*)$ 中的 $x\tan\alpha$ 一项是减去的。
- 写到“$v^2 > 0$”就停下来，没有留意到这其实是对分母的条件；要代入 $\tan\alpha$ 之后，它才变成对 $H$ 的条件。

## STEP 之外

在这题中，是两个边界条件（球网和着地点）决定了发射条件，而不是反过来：这是一个简单的*边值问题*（boundary-value problem），也是以数值方法求解微分方程（differential equations）的“打靶法”（shooting methods）背后的想法。一旦考虑空气阻力和旋转，方程就不再有闭合形式（closed-form）的解，那时正正要用这种方法来处理。Kleppner and Kolenkow, *An Introduction to Mechanics*（力学入门教科书）讨论了带阻力的抛体运动，以及更多其他课题。

详细解答

# STEP II 1997，第 12 题

概率与统计 · 二项分布的奇偶概率

**STEP II 1997，第 12 题**

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

step.maths.org/questions/97-s2-q12

## 关键思路

只需追踪信息此刻是对还是错。每听错一次，信息便对错互换，所以传回 Albert 的信息正确，当且仅当耳语中出错的次数是偶数。要抽出二项展开式（binomial expansion）中的偶数项，正好可以用 $(a+b)^n+(a-b)^n$，因为奇数项会互相抵消。至于修改后的游戏，要留意一旦失败一次，Albert 传出去的便是错的信息，所以从那时起，每一轮都需要有奇数次听错。

## 解答

### 三名玩家

若有 $n$ 名玩家，便有 $n$ 次耳语：Albert 传给 Bertha，Bertha 传给下一位，如此类推绕一圈传到 Zebedee，然后 Zebedee 传回 Albert。每次耳语被正确听到的概率（probability）是 $p$，而且各次互相独立（independent）。Albert 听到的是否真相，只取决于有多少次耳语被听错，与硬币掷出的结果无关。

三名玩家时有三次耳语，每次正确的概率是 $\frac{2}{3}$。传到 Albert 的信息正确，条件是没有耳语出错，或者恰好有两次出错（第二次错误把第一次抵消了）。所以

$$\Pr(\text{获胜})=\left(\tfrac{2}{3}\right)^3+3\left(\tfrac{1}{3}\right)^2\left(\tfrac{2}{3}\right)=\frac{8}{27}+\frac{6}{27}=\frac{14}{27}.$$

### $n$ 名玩家

设 $q = 1 - p$。听错的耳语次数服从 $\mathrm{B}(n, q)$ 分布（distribution），而当这个次数是偶数时游戏便获胜：

$$P = \sum_{k\text{ 为偶数}} \binom{n}{k} q^k p^{n-k}.$$

现在

$$(p+q)^n + (p-q)^n = \sum_{k=0}^{n} \binom{n}{k} p^{n-k} q^k \left(1 + (-1)^k\right) = 2 \sum_{k\text{ 为偶数}} \binom{n}{k} p^{n-k} q^k,$$

因为 $k$ 为奇数的项互相抵消，而 $k$ 为偶数的项则出现两次。由于 $p + q = 1$ 及 $p - q = 2p - 1$，

$$\boxed{P = \tfrac{1}{2}\left(1 + (2p-1)^n\right).}$$

当 $n = 3$、$p = \frac{2}{3}$ 时，上式给出 $\frac{1}{2}\left(1 + \frac{1}{27}\right) = \frac{14}{27}$，与第一部分的结果一致。

### 修改后的游戏

设 $N$ 为 Albert 向 Bertha 耳语的次数。把信息绕圈传足一周称为一轮，所以每一轮都由 Albert 耳语开始。

> **想一想** 一轮失败之后，下一轮是否只是把第一轮原封不动地重来一次?

不是。在第一轮，Albert 传出的是正确的结果。如果他从 Zebedee 那里听错了，他传下去的是他听到的内容，而那是错的。由错的信息出发，这一轮结束时 Albert 听到真相，当且仅当有奇数次耳语出错，其概率为 $1 - P$。如果这一轮又失败，Albert 便再一次听到（并传出）错的信息，所以之后的每一轮都与这一轮相同。由于各次聆听互相独立，各轮也互相独立。

因此第一轮成功的概率是 $P$，而之后每一轮成功的概率是 $1 - P$，失败的概率是 $P$。故此 $\Pr(N = 1) = P$，而对于 $k \geq 2$，

$$\Pr(N = k) = \underbrace{(1-P)}_{\text{第 1 轮失败}} \underbrace{P^{k-2}}_{\text{第 2,…,}k-1\text{ 轮失败}} \underbrace{(1-P)}_{\text{第 }k\text{ 轮成功}} = (1-P)^2 P^{k-2}.$$

假设 $0 < p < 1$，于是 $0 < P < 1$。那么，在题目给出的和式中取 $x = P$，可得

$$\sum_{k=2}^{\infty} kP^{k-2} = \frac{1}{P}\left(\sum_{k=1}^{\infty} kP^{k-1} - 1\right) = \frac{1}{P}\left(\frac{1}{(1-P)^2} - 1\right),$$

因此

$$\begin{aligned}\mathrm{E}(N) &= P + (1-P)^2 \cdot \frac{1}{P}\left(\frac{1}{(1-P)^2} - 1\right)\\ &= P + \frac{1-(1-P)^2}{P}\\ &= P + (2-P) = 2.\end{aligned}$$

Albert 向 Bertha 耳语次数的期望值（expectation）是 2，不论玩家有多少人，也不论 $p$ 是多少。

还有一个更快捷的方法看出这个结果。已知第一轮失败，其后所需的轮数服从几何分布（geometric distribution），成功概率为 $1-P$，所以其平均值是 $1/(1-P)$。故此 $\mathrm{E}(N) = 1 + (1-P)\cdot\frac{1}{1-P} = 2$。

（边界情况：若 $p=1$，或 $p=0$ 而 $n$ 为偶数，则 $P=1$，Albert 只耳语一次。若 $p=0$ 而 $n$ 为奇数，则 $P=0$，第二轮必定成功，所以 $N=2$。）

## 验算

- 以准确分数列出听对与听错的全部 $2^n$ 种组合，对 $n \le 8$ 及若干个 $p$ 值都得出 $\frac{1}{2}(1+(2p-1)^n)$，当中包括三名玩家时的 $\Pr(\text{获胜}) = \frac{14}{27}$。
- $p = \frac{1}{2}$ 给出 $P = \frac{1}{2}$：信息纯粹是杂讯，理应如此。
- 以 $(n,p) = (3,\frac{2}{3})$、$(5,0.9)$ 及 $(4,0.3)$ 模拟 $2\times10^5$ 局修改后的游戏，所得的平均耳语次数分别为 2.001、1.996 及 1.994。

## 常见错误

- 把耳语次数数成 $n-1$ 而不是 $n$：最后一次耳语，即由 Zebedee 传回 Albert 的那一次，也要计算在内。
- 把修改后的游戏中每一轮的获胜概率都当作 $P$。这样会得出 $1/P$，是错的，因为失败之后 Albert 传出的是错的信息。

## STEP 之外

信息的对／错状态是一条两状态的马可夫链（Markov chain），而 $2p-1$ 是其转移矩阵（transition matrix）的第二个特征值（eigenvalue）。这就是为甚么 $(2p-1)^n$ 决定了信息“忘记”硬币结果的速度。同一个模型称为二元对称通道（binary symmetric channel），是信息论（information theory）和纠错码（error-correcting codes）的起点。Grimmett 与

Welsh 的 *Probability: An Introduction*（概率论入门）以这个程度讲解生成函数（generating functions）和马可夫链。

详细解答

# STEP II 1997，第 13 题

概率与统计 · 布丰投针问题

**STEP II 1997，第 13 题**

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

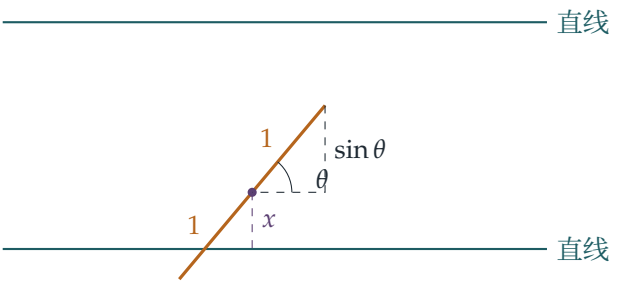
step.maths.org/questions/97-s2-q13

## 关键思路

这是布丰投针问题（Buffon's needle）。用两个独立（independent）的随机量来描述针的位置：针的中心与最近一条直线的距离，以及针与直线所成的锐角。中心一旦固定，一切就取决于针在垂直于直线的方向上伸展多远，也就是针的半长乘以该角的正弦。于是每一部分都化为关于一个均匀分布（uniformly distributed）的角的问题。

## 解答

我们按惯常的方式为“随机投下”建立模型：中心与最近一条直线的距离 $X$ 在 $[0,1]$ 上均匀分布（直线相距 2 cm，所以最近的一条至多相距 1 cm），而针与直线所成的锐角 $\theta$ 在 $[0,\frac{\pi}{2}]$ 上均匀分布，且与 $X$ 独立。（若取角在 $[0,\pi)$ 上均匀分布，答案也一样，因为两种取法下 $\sin\theta$ 的分布相同。）




针的每一半长度为 1，所以两端分别在中心的上方和下方，垂直距离为 $\sin\theta$。

**第 (i) 部分**

设中心与最近一条直线相距 $x$，则针与该直线相交，当且仅当其中一端越过该直线，即

$$\sin\theta > x \iff \theta > \arcsin x,$$

因为 sin 在 $[0, \frac{\pi}{2}]$ 上递增。由于 $\theta$ 在 $[0, \frac{\pi}{2}]$ 上均匀分布，

$$\mathrm{P}(\text{与最近的直线相交} \mid X = x) = \frac{\frac{\pi}{2} - \arcsin x}{\frac{\pi}{2}} = 1 - \frac{2}{\pi}\arcsin x = \frac{2}{\pi}\arccos x.$$

**第 (ii) 部分**

假设针与最近的直线相交，即 $\theta > \arcsin x$。沿着针量度，交点与中心相距 $x/\sin\theta$（这是一个直角三角形的斜边，其中 $\theta$ 的对边为 $x$）。所以两段的长度分别为 $1 + x/\sin\theta$ 和 $1 - x/\sin\theta$，较短的一段是

$$S = 1 - \frac{x}{\sin\theta}.$$

当 $\theta$ 由 $\arcsin x$ 变到 $\frac{\pi}{2}$ 时，$S$ 由 0 递增至 $1 - x$，所以 $S$ 在 $[0, 1 - x]$ 内取值。

在相交的条件下，$\theta$ 在 $(\arcsin x, \frac{\pi}{2}]$ 上均匀分布，这个区间的长度为 $\frac{\pi}{2} - \arcsin x = \arccos x$。对于 $0 \le s \le 1 - x$，

$$S \le s \iff \frac{x}{\sin\theta} \ge 1 - s \iff \sin\theta \le \frac{x}{1-s} \iff \theta \le \arcsin\frac{x}{1-s},$$

这里用到了 $1 - s > 0$ 和 $x/(1-s) \le 1$（因为 $s \le 1 - x$），所以反正弦有意义。因此较短一段的累积分布函数（cumulative distribution function）为

$$F(s) = \mathrm{P}(S \le s) = \begin{cases} 0, & s < 0, \\ \dfrac{\arcsin\left(\frac{x}{1-s}\right) - \arcsin x}{\arccos x}, & 0 \le s \le 1 - x, \\ 1, & s > 1 - x. \end{cases}$$

作为合理性检查，中间的公式在 $s = 0$ 时给出 0，在 $s = 1 - x$ 时给出 $(\frac{\pi}{2} - \arcsin x)/\arccos x = 1$，所以 $F$ 是连续的。

**想一想** 为甚么得知针与直线相交之后，$\theta$ 仍然是均匀分布的?

把一个均匀分布的变量限制在某个子区间内（即以它落在该子区间为条件），它在该子区间上仍然是均匀分布的：对于 $\arcsin x \le a \le \frac{\pi}{2}$，$\mathrm{P}(\theta \le a \mid \theta > \arcsin x) = (a - \arcsin x)/(\frac{\pi}{2} - \arcsin x)$，这是 $a$ 的线性函数。这正是我们所用到的。

### 第 (iii) 部分

最近的直线相距 $X \le 1$，而其他每一条直线都相距至少 $2 - X \ge 1$。针从中心最多伸展 $\sin\theta \le 1$，所以它要与最近那条以外的直线相交，只可能在 $X = 1$ 且 $\theta = \frac{\pi}{2}$ 恰好成立的情况下发生，而这个情况的概率为零。所以针不碰到任何直线，当且仅当它不碰到最近的直线，于是由第 (i) 部分得

$$\mathrm{P}(\text{不相交} \mid X = x) = \frac{2}{\pi}\arcsin x.$$

由于 $X$ 在 $[0,1]$ 上均匀分布（密度为 1），我们对 $x$ 取平均：

$$\mathrm{P}(\text{不相交}) = \int_0^1 \frac{2}{\pi}\arcsin x \,\mathrm{d}x.$$

用分部积分法（integration by parts），取 $u = \arcsin x$ 及 $\mathrm{d}v = \mathrm{d}x$，

$$\begin{aligned}\int_0^1 \arcsin x \,\mathrm{d}x &= \Big[x\arcsin x\Big]_0^1 - \int_0^1 \frac{x}{\sqrt{1-x^2}}\,\mathrm{d}x \\ &= \frac{\pi}{2} - \Big[-\sqrt{1-x^2}\Big]_0^1 = \frac{\pi}{2} - 1.\end{aligned}$$

因此

$$\mathrm{P}(\text{针不碰到任何直线}) = \frac{2}{\pi}\Big(\frac{\pi}{2} - 1\Big) = 1 - \frac{2}{\pi} \approx 0.363.$$

## 验算

- 第 (i) 部分在两端的表现合理：当 $x = 0$ 时，中心在直线上，概率为 1；当 $x \to 1$ 时，只有垂直于直线的针才碰得到，概率趋向 0。
- 相交的概率为 $2/\pi \approx 0.637$，即布丰的经典答案 $2\ell/(\pi d)$，其中 $\ell = d = 2$。
- 模拟 $10^6$ 次随机投针，得出的不相交比率为 0.3626，而 $1 - 2/\pi = 0.3634$；对 $x = 0.2, 0.5, 0.7$ 模拟所得的 $F(s)$ 值与公式相符，误差约为 0.002。

## 常见错误

- 在第 (ii) 部分忘记取条件：除以 $\frac{\pi}{2}$ 而不是 $\arccos x$，会得出一个永远达不到 1 的“分布函数”。
- 把较短的一段取为 $1 - x\sin\theta$ 或 $1 - x/\cos\theta$：要查清楚三角形的哪一条边是 $x$，哪一个角是 $\theta$。
- 在第 (iii) 部分乘以 $\frac{1}{2}$，或取 $X$ 在 $[0,2]$ 上均匀分布而没有相应调整相交的条件；与最近直线的距离是在 $[0,1]$ 上均匀分布的。

## STEP 之外

布丰投针问题（1777 年）是几何概率（geometric probability）中最早的问题之一，而 $\pi$ 的出现意味着投针可以给出 $\pi$ 的一个（缓慢的）蒙地卡罗（Monte Carlo）估计。最简洁的现代处理方法利用期望值的线性性质（linearity of expectation）：相交次数的期望值与针的长度成正比，即使针是弯曲的也一样（“布丰面条”，Buffon's noodle），由此可通往积分几何（integral geometry）和克罗夫顿公式（Crofton formula）。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率入门教材）讨论了联合分布（joint distribution）和这一类几何概率问题。

详细解答

# STEP II 1997，第 14 题

概率与统计 · 泊松过程

**STEP II 1997，第 14 题**

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

step.maths.org/questions/97-s2-q14

## 关键思路

虽然题目以力学为背景，这其实是一道泊松过程（Poisson process）的题目。运动学（kinematics）只是告诉我们哪些车辆要计算在内：一辆较快的车要在隧道内追上我，它进入隧道的时间必须在我之后某段时间之内。一旦知道这个时限，余下的一切都变成关于泊松计数的问题。在后半部分，我们会用到以下事实：若把泊松车流中的每辆车随机标记为快车或慢车，车流便会分拆成两个独立（independent）的泊松车流。我们还需要留意，一辆慢车会替我挡住它后面的所有快车。

## 解答

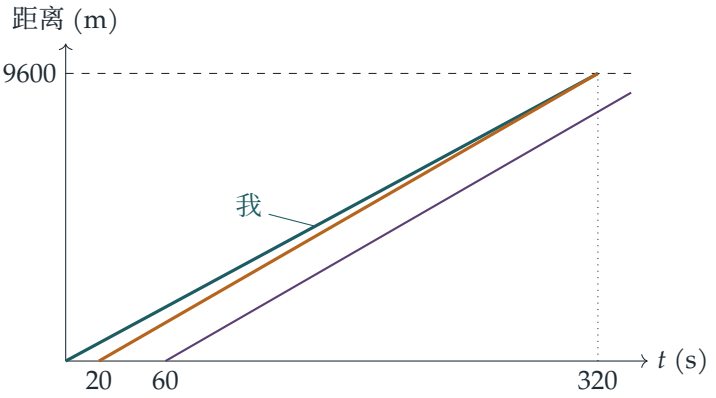
全题把车辆视为质点（particle）（不计车身长度），所以一辆车追上前车之后，便会一直以前车的速率行驶。以秒为单位量度时间 $t$，由我进入隧道的一刻起计。车辆以每分钟 6 辆的率进入，即每秒 0.1 辆，所以在任何长度为 $\tau$ 秒的时段内进入的车辆数目服从 $\mathrm{Po}(0.1\tau)$，而且互不相交的时段内的数目互相独立。

### 第一部分：谁能追上我？

我通过隧道需时 $9600/30 = 320\,\mathrm{s}$。考虑一辆以 $32\,\mathrm{m\,s^{-1}}$ 行驶、在时间 $t > 0$ 进入的车辆。在它追上我之前，于时间 $s$ 我已驶入隧道 $30s$，而它则驶入了 $32(s-t)$。两者在 $2s = 32t$ 时相等，即 $s = 16t$，此时距离隧道入口 $480t$ 米。所以它在我离开之前追上我，当且仅当

$$480t \le 9600 \iff t \le 20.$$

（等价地说：它通过隧道需时 $9600/32 = 300\,\mathrm{s}$，所以它不迟于我到达出口，当且仅当 $t + 300 \le 320$。）




这幅距离－时间图显示了我（蓝绿色）、一辆在 $t = 20$ 进入的快车（琥珀色）和一辆在 $t = 60$ 进入的快车（紫红色）。琥珀色的车刚好在我们同时到达出口的一刻追上我；任何更迟进入的快车，例如紫红色那一辆，在我离开时仍然在我后面。

其他车辆全部以 $32\,\mathrm{m\,s^{-1}}$ 行驶，所以在 $(0, 20]$ 内进入的车辆在到达我身后之前，不会被任何东西阻慢：除我以外，在它前面的任何车辆同样以 $32\,\mathrm{m\,s^{-1}}$ 行驶，直至那辆车也追上我为止。比我先进入的车辆在我前面，与问题无关。所以我离开时身后的车龙，正好由在 $(0, 20]$ 内进入的车辆组成，其数目服从 $\mathrm{Po}(0.1 \times 20) = \mathrm{Po}(2)$。我身后车龙中车辆数目的期望值（expected value）是

$$\boxed{2}.$$

**第二部分：恰好有两辆车进入并追上我**

现在其他每辆车各有 $\frac{1}{2}$ 的概率是快车（$32\,\mathrm{m\,s^{-1}}$）或慢车（$30\,\mathrm{m\,s^{-1}}$），并且与其他一切独立。

> ? **想一想** 在 20 秒内进入的每一辆快车，是否仍然都能追上我?

不是。一辆在时间 $t' > 0$ 进入的慢车，会一直与我保持 $30t'$ 米的固定距离，所以它永远追不上我。在它之后进入的快车会追上这辆慢车（或被慢车后面的其他车辆挡住），此后只能以 $30\,\mathrm{m\,s^{-1}}$ 行驶，所以它同样永远到不了我身后。因此，一辆在 $(0, 20]$ 内进入的车辆能追上我，当且仅当它是快车，*而且*在我与它之间进入的每一辆车都是快车。在 $t = 20$ 之后进入的快车，即使前面毫无阻挡，也追不上我。

设 $K \sim \mathrm{Po}(2)$ 为在 $(0, 20]$ 内进入的车辆数目。这一部分所求的事件，是恰好有两辆车在该时段内进入，而且两辆都追上我。这需要 $K = 2$ 且两辆车都是快车：

$$\mathrm{P} = \mathrm{P}(K = 2) \times \left(\tfrac{1}{2}\right)^2 = \frac{2^2 \mathrm{e}^{-2}}{2!} \cdot \frac{1}{4} = \boxed{\tfrac{1}{2}\mathrm{e}^{-2}} \approx 0.0677.$$

### 最后部分：恰好有两辆车在我身后排队

现在，头 20 秒内进入的车辆数目可以是任意的。把在 $(0,20]$ 内进入的 $K$ 辆车按进入的先后次序排列。根据上面的论证，我在出口时身后的车龙，就是这个列表开首那一串连续不断的快车。（在 $t=20$ 之后进入的车辆不可能加入车龙，因为它永远追不上我。）所以车龙中恰好有两辆车，当且仅当

- $K=2$ 且两辆都是快车：已知 $K=2$ 时，概率为 $\frac{1}{4}$；或
- $K\geq 3$，且头两辆是快车、第三辆是慢车：已知 $K=k\geq 3$ 时，概率为 $\frac{1}{8}$。

因此

$$\begin{aligned}\mathrm{P}(\text{车龙有两辆车}) &= \tfrac{1}{4}\,\mathrm{P}(K=2)+\tfrac{1}{8}\,\mathrm{P}(K\geq 3)\\ &= \tfrac{1}{4}\cdot 2\mathrm{e}^{-2}+\tfrac{1}{8}\left(1-\mathrm{e}^{-2}-2\mathrm{e}^{-2}-2\mathrm{e}^{-2}\right)\\ &= \tfrac{1}{2}\mathrm{e}^{-2}+\tfrac{1}{8}-\tfrac{5}{8}\mathrm{e}^{-2}\\ &= \boxed{\tfrac{1}{8}\left(1-\mathrm{e}^{-2}\right)}\approx 0.108.\end{aligned}$$

这个答案比上一个大，而这是必然的：上一个事件只是出现两辆车车龙的其中一种情况。

> 一个巧妙的另解：把车流分拆成快车和慢车，得到两个独立的泊松过程，各自的率为 $0.05\,\mathrm{s}^{-1}$。车龙恰好有两辆车，意思是在头 20 秒内，最先到达的两辆是快车，而下一辆到达的车（如果有的话）是慢车。遇到事件带有标记的泊松题目时，从这两个车流的角度去想，往往是最快捷的途径。

## 验算

- 极限情况：假如每辆车都是慢车，便没有车能追上我；果然，当快车的概率为 0 时，车龙的公式给出 0。
- 若快车的概率是 $p$ 而不是 $\frac{1}{2}$，同样的论证给出 $p^2\mathrm{P}(K=2)+p^2(1-p)\mathrm{P}(K\geq 3)$。当 $p=1$ 时，这等于 $\mathrm{P}(K=2)=2\mathrm{e}^{-2}$，与第一部分吻合，因为那里的车龙长度正好服从 $\mathrm{Po}(2)$。
- 我们对不准超车的交通情况模拟了 $4\times 10^5$ 次。结果是第一部分的平均车龙长度为 2.003，第二部分的两个概率为 0.0674 和 0.1076，而准确值则是 0.0677 和 0.1081。

## 常见错误

- 忘记了一辆慢车会替我挡住它后面的所有快车,于是在最后部分答 $\mathrm{P}(20\text{ s 内恰好有两辆快车}) = \frac{1}{2}\mathrm{e}^{-1}$。
- 把每分钟 6 辆的率与以秒计的时间一并使用：20 秒内的平均数是 2，不是 120。
- 以为任何时间进入的车辆都会令车龙增长：只有在头 20 秒内进入的车辆才能到达我身后。

## STEP 之外

第二部分用到*稀疏化*（thinning）：若率为 $\lambda$ 的泊松过程中，每个事件各自独立地以概率 $p$ 被保留，则被保留和被舍弃的事件分别构成率为 $p\lambda$ 和 $(1-p)\lambda$ 的两个独立泊松过程。不能超车的车龙是交通流理论（traffic flow theory）中的经典模型，其中慢车会造成我们在这里见到的那种“车队”（platoon）。关于泊松过程和稀疏化，可参阅 Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）。

详细解答

# STEP III 1997，第 1 题

纯数学 · 指数幂级数

**STEP III 1997，第 1 题**

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

step.maths.org/questions/97-s3-q1

## 关键思路

把 $\mathrm{e}^x = \sum x^n/n!$ 乘以一个多项式（polynomial），相当于把级数（series）逐项向后平移。于是 $x^n$ 的系数（coefficient）就是 $1/n!$、$1/(n-1)!$、$1/(n-2)!$,… 的组合；以 $n!$ 作公分母通分后，分子是一个关于 $n$ 的多项式。代入 $x=1$，幂级数（power series）的恒等式就变成一个数值级数的恒等式。因此整道题的关键，在于选出能产生所需分子的多项式。下降阶乘（falling factorials）$n$、$n(n-1)$、$n(n-1)(n-2)$ 是最自然的基本组件，因为 $n(n-1)\cdots(n-k+1)/n! = 1/(n-k)!$。

## 解答

在整个解答中，$\mathrm{e}^x = \sum_{n\geq 0} x^n/n!$ 对所有实数 $x$ 都绝对收敛（converges absolutely）。把它乘以一个多项式，得到的是有限个这类级数之和，所以我们可以把这些级数逐项相加，然后代入 $x=1$。我们约定：当 $m$ 为负整数时，$1/m! = 0$。这样，下面每一个和式都可以由 $n=0$ 开始。

### 第 (i) 部分

我们有

$$(x^2+5x+4)\,\mathrm{e}^x = \sum_{n\geq 0}\frac{x^{n+2}}{n!} + 5\sum_{n\geq 0}\frac{x^{n+1}}{n!} + 4\sum_{n\geq 0}\frac{x^n}{n!}.$$

要合并 $x^n$ 的系数，先把头两个和式重新编排下标（reindex）。该系数为

$$\begin{aligned}\frac{1}{(n-2)!}+\frac{5}{(n-1)!}+\frac{4}{n!} &= \frac{n(n-1)+5n+4}{n!}\\ &= \frac{n^2+4n+4}{n!}=\frac{(n+2)^2}{n!}.\end{aligned}$$

这对所有 $n \geq 0$ 都成立，包括 $n=0$ 和 $n=1$：在这两个情况下，约定 $1/m!=0$ 会把不存在的项消去。（当 $n=0$ 时，系数为 $4=2^2/0!$；当 $n=1$ 时，系数为 $5+4=9=3^2/1!$。）因此

$$(x^2+5x+4)\,\mathrm{e}^x=\sum_{n\geq 0}\frac{(n+2)^2}{n!}\,x^n,$$

代入 $x=1$ 得

$$10\,\mathrm{e}=\frac{2^2}{0!}+\frac{3^2}{1!}+\frac{4^2}{2!}+\frac{5^2}{3!}+\cdots=4+\frac{3^2}{1!}+\frac{4^2}{2!}+\cdots,$$

这正是所要证明的结果。

### 第 (ii) 部分

现在通项（general term）是 $(n+1)^2/n!$，所以我们要找一个二次式 $x^2+bx+c$，使其级数的系数为

$$\frac{n(n-1)+bn+c}{n!}=\frac{(n+1)^2}{n!}.$$

**想一想** 怎样才能不靠猜测而求出 $b$ 和 $c$?

把目标分子写成下降阶乘的形式：$n^2+2n+1=n(n-1)+3n+1$。比较系数得 $b=3$ 和 $c=1$。于是，按照与第 (i) 部分完全相同的论证，

$$(x^2+3x+1)\,\mathrm{e}^x=\sum_{n\geq 0}\frac{(n+1)^2}{n!}\,x^n,$$

代入 $x=1$ 得

$$(1+3+1)\,\mathrm{e}=5\,\mathrm{e}=1+\frac{2^2}{1!}+\frac{3^2}{2!}+\frac{4^2}{3!}+\cdots.$$

### 第 (iii) 部分

通项是 $(n+1)^3/n!$，所以这次需要一个三次式。$(ax^3+bx^2+cx+d)\,\mathrm{e}^x$ 中 $x^n$ 的系数为

$$\frac{a}{(n-3)!}+\frac{b}{(n-2)!}+\frac{c}{(n-1)!}+\frac{d}{n!}=\frac{a\,n(n-1)(n-2)+b\,n(n-1)+c\,n+d}{n!}.$$

我们把目标分子展开成下降阶乘。由于 $n^3 = n(n-1)(n-2) + 3n(n-1) + n$ 及 $n^2 = n(n-1) + n$，

$$\begin{aligned}(n+1)^3 &= n^3 + 3n^2 + 3n + 1 \\ &= \big[n(n-1)(n-2) + 3n(n-1) + n\big] + 3\big[n(n-1) + n\big] + 3n + 1 \\ &= n(n-1)(n-2) + 6n(n-1) + 7n + 1.\end{aligned}$$

（以 $n = 1$ 快速验算：两边都等于 8。以 $n = 2$ 验算：$27 = 0 + 12 + 14 + 1$。）所以 $a = 1$、$b = 6$、$c = 7$、$d = 1$，而

$$(x^3 + 6x^2 + 7x + 1)\,\mathrm{e}^x = \sum_{n \geq 0} \frac{(n+1)^3}{n!}\, x^n.$$

代入 $x = 1$，

$$1 + \frac{2^3}{1!} + \frac{3^3}{2!} + \frac{4^3}{3!} + \cdots = (1 + 6 + 7 + 1)\,\mathrm{e} = 15\,\mathrm{e}.$$

**想一想** 为甚么不能直接把级数逐项平方或立方？

因为 $\sum (n+1)^3/n!$ 并不能以任何简单的方式由 $\sum (n+1)/n!$ 构造出来。和式对分子而言是线性的，所以我们能做的，是把分子拆成若干个 $n(n-1)\cdots(n-k+1)$ 形式的部分；每一部分都与 $n!$ 约分，得出 e 的级数平移后的版本。每一个这样的部分在 $x = 1$ 时都恰好贡献 e。因此答案就是 e 乘以各下降阶乘系数之和，即 $1 + 6 + 7 + 1 = 15$。

## 验算

- 计算至 $n = 60$ 的部分和（partial sums），得 27.1828…、13.5914… 和 40.7742…，在机器精度内与 10e、5e 和 15e 吻合。
- 三个分子恒等式中的系数，已就 $n = 0, \ldots, 39$ 逐一精确验证。
- 一致性：第 (ii) 部分的级数是 $\sum (n+1)^2/n!$，第 (i) 部分的是 $\sum (n+2)^2/n!$。两者之差 $\sum (2n+3)/n! = 2\sum n/n! + 3\sum 1/n! = 2\mathrm{e} + 3\mathrm{e} = 5\mathrm{e}$，确实等于 $10\mathrm{e} - 5\mathrm{e}$。

## 常见错误

- 忘记在重新编排和式的下标时，头两项（$n = 0, 1$）需要另行处理。约定 $m < 0$ 时 $1/m! = 0$ 可以处理这两项，但你仍应亲自检查。
- 把 $(n+1)^3$ 与幂 $n^3, n^2, n$ 比较系数，而不是与下降阶乘比较，结果选错了三次式。

## STEP 之外

和式 $\sum_{n\geq 0} n^k/n!$ 等于 $B_k\,\mathrm{e}$，其中 $B_k$ 是第 $k$ 个贝尔数（Bell number），即把一个含 $k$ 个元素的集合划分（partition）的方法数目：$B_0,\ldots,B_4 = 1,1,2,5,15$。这就是 Dobiński 公式（Dobiński's formula）。事实上，设 $m = n+1$，第 (iii) 部分就是 $\sum_{m\geq 1} m^4/m! = B_4\,\mathrm{e} = 15\,\mathrm{e}$，而系数 $1,6,7,1$ 正是第二类斯特灵数（Stirling numbers of the second kind）$1,7,6,1$ 倒过来的次序。幂级数之所以可以逐项重排和合并，依据是绝对收敛，这是大学一年级分析学（analysis）的一个主题。理论方面可参阅 Abbott 的 *Understanding Analysis*；想做更多同类题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

# STEP III 1997，第 2 题

纯数学 · 曲线描绘

> **STEP III 1997，第 2 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s3-q2

## 关键思路

取对数（logarithm）后，$x^y = y^x$ 变成 $y\ln x = x\ln y$，再除以 $xy$ 便得 $\mathrm{f}(x) = \mathrm{f}(y)$。所以题目的后半其实是在问前半的事：只要知道 f 的每个输出值对应多少个输入值，就知道每个 $x$ 有多少个“拍档” $y$。这条方程总有一个明显的解 $y = x$；有趣的是另外那一个。

## 解答

### f 的图像

用商法则（quotient rule）求导，

$$\mathrm{f}'(t) = \frac{t\cdot\frac{1}{t} - \ln t}{t^2} = \frac{1-\ln t}{t^2}, \qquad \mathrm{f}''(t) = \frac{2\ln t - 3}{t^3}.$$

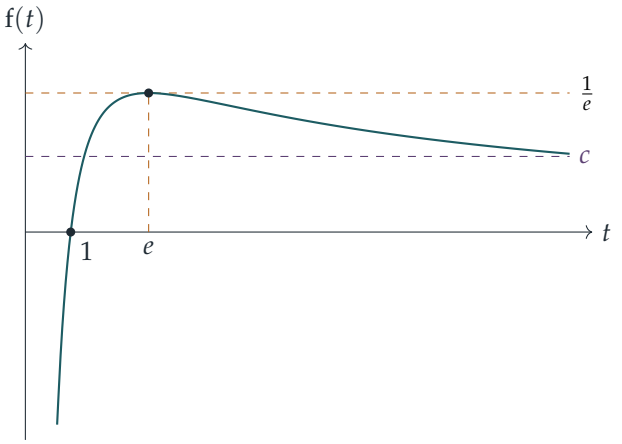
所以当 $0 < t < e$ 时 $\mathrm{f}'(t) > 0$，当 $t > e$ 时 $\mathrm{f}'(t) < 0$：f 在 $(0, e]$ 上严格递增（strictly increasing），在 $[e, \infty)$ 上严格递减（strictly decreasing）。极大值为

$$\mathrm{f}(e) = \frac{1}{e}.$$

草图还需要的其他特征：

- 当 $0 < t < 1$ 时 $\mathrm{f}(t) < 0$，$\mathrm{f}(1) = 0$，而当 $t > 1$ 时 $\mathrm{f}(t) > 0$。
- 当 $t \to 0^+$ 时，$\ln t \to -\infty$ 而 $1/t \to +\infty$，所以 $\mathrm{f}(t) \to -\infty$：$y$-轴是一条渐近线（asymptote）。
- 当 $t \to \infty$ 时，$\mathrm{f}(t) \to 0^+$，因为 $\ln t$ 的增长比 $t$ 慢。所以 $t$-轴是一条渐近线。

- 在 $t = e^{3/2}$ 处有一个拐点（point of inflection），该处 $\mathrm{f} = \frac{3}{2}e^{-3/2} \approx 0.33$。计数解的数目时用不着它，但有了它，草图才算如实。




**有多少个 $t$ 给出数值 $c$?**

由于 f 在 $(0, e]$ 和 $[e, \infty)$ 上各自连续（continuous）且严格单调（strictly monotonic），每一段对它取得到的任何值都恰好取一次。在 $(0, e]$ 上，f 由 $-\infty$ 递增至 $\frac{1}{e}$；在 $[e, \infty)$ 上，它由 $\frac{1}{e}$ 向 0 递减，但永远不会达到这个值。因此，对给定的数值 $c$：

- $c > \frac{1}{e}$：**没有** $t$ 值；
- $c = \frac{1}{e}$：**一个**值，$t = e$；
- $0 < c < \frac{1}{e}$：**两个**值，一个在 $(1, e)$ 内，一个在 $(e, \infty)$ 内（即草图中的水平线 $c$）；
- $c \leq 0$：**一个**值，在 $(0, 1]$ 内，因为递减的那一支始终严格为正。

**计数 $x^y = y^x$ 的解**

对 $x, y > 0$，两边皆为正数，所以可以取对数；ln 是一对一（one-to-one）的，所以既不会多出解，也不会遗漏解：

$$x^y = y^x \iff y \ln x = x \ln y \iff \frac{\ln x}{x} = \frac{\ln y}{y} \iff \mathrm{f}(y) = \mathrm{f}(x),$$

其中我们除以了 $xy > 0$。所以，对固定的 $x$，正数 $y$ 的数目就是满足 $\mathrm{f}(t) = c$ 的 $t$ 的数目，其中 $c = \mathrm{f}(x)$。其中一个必定是 $y = x$。

? **想一想** $c = \mathrm{f}(x)$ 实际上可以落入上述四种情况中的哪几种?

由于 $c$ 是 f 的一个值，$c > \frac{1}{e}$ 这种情况不可能出现。把其余情况转换回关于 $x$ 的条件：

- $0 < x \le 1$：这时 $c = \mathrm{f}(x) \le 0$，所以恰好有一**个** $y$，就是 $y = x$。
- $x = e$：这时 $c = \frac{1}{e}$，所以恰好有一**个** $y$，就是 $y = e$。
- $1 < x < e$ 或 $x > e$：这时 $0 < c < \frac{1}{e}$，所以恰好有**两个** $y$ 值：$y = x$ 和另外一个。若 $1 < x < e$，另外那个位于 $(e, \infty)$ 内；若 $x > e$，它位于 $(1, e)$ 内。

例如，$x = 2$ 的拍档是 $y = 4$，因为 $2^4 = 4^2 = 16$。

### 描绘解集

解集由直线 $y = x$（对所有 $x > 0$）和第二条曲线组成；第二条曲线由那些“另外的”解构成，完全位于 $x > 1, y > 1$ 的区域内。它的形状如下：

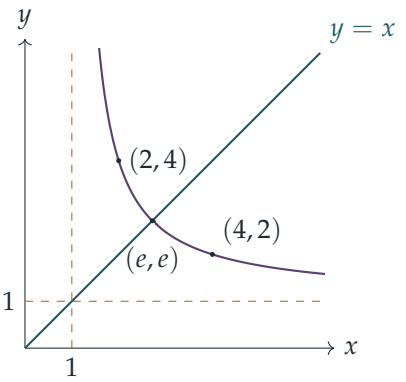
- *对称性*。把 $x$ 和 $y$ 对调，方程不变，所以曲线关于直线 $y = x$ 对称。
- *递减*。当 $x$ 在 $(1, e)$ 内增大时，$\mathrm{f}(x)$ 增大，所以拍档 $y > e$ 必须移向 f 较大之处，即 $y$ 减小。$x > e$ 时同理。所以在这一支上，$y$ 是 $x$ 的递减函数。
- *渐近线*。当 $x \to 1^+$ 时，$\mathrm{f}(x) \to 0^+$，所以拍档 $y \to \infty$：直线 $x = 1$ 是一条渐近线。由对称性，当 $x \to \infty$ 时，$y \to 1^+$，故 $y = 1$ 也是一条渐近线。
- *与直线 $y = x$ 相交*。当 $x \to e$ 时，拍档也趋向 $e$，所以曲线经过 $(e, e)$，并在该处穿过直线。在 $t = e$ 附近，利用 $\mathrm{f}'(e) = 0$ 和 $\mathrm{f}''(e) = -e^{-3}$，由泰勒定理（Taylor's theorem）得 $\mathrm{f}(t) \approx \frac{1}{e} - \frac{(t-e)^2}{2e^3}$。所以 $\mathrm{f}(x) = \mathrm{f}(y)$ 迫使 $(y-e)^2 \approx (x-e)^2$，即 $y - e \approx \pm(x-e)$。取正号就是直线 $y = x$。取负号就是我们的曲线，因此它与直线成直角相交，斜率（gradient）为 $-1$。

> ? **想一想** 我们能否把这条曲线明确地写出来?

可以，只要用两者之比作参数化（parametrise）。设 $y = tx$，其中 $t > 0, t \ne 1$。那么 $x^{tx} = (tx)^x$；两边开 $x$ 次方得 $x^t = tx$，所以 $x^{t-1} = t$，于是

$$x = t^{1/(t-1)}, \qquad y = t^{t/(t-1)}.$$

在 $t = 2$ 时得到 $(2, 4)$，在 $t = \frac{1}{2}$ 时得到 $(4, 2)$。当 $t \to 1$ 时，两个坐标都趋向 $e$，因为 $t^{1/(t-1)} \to e$。下图的曲线就是这样画出来的。




## 验算

- $2^4 = 4^2$ 是著名的整数例子，它确实落在曲线上，并如所料分别位于 $1 < x < e$ 和 $x > e$ 之内。
- 数值上，用细密的网格搜索，得出 f 的极大值位于 $t \approx 2.718$，其值为 $0.367879 \approx 1/e$。用二分法（bisection）求得 $x = 1.01$ 的拍档为 $y \approx 659$，而 $x = 100$ 的拍档为 $y \approx 1.050$。这些结果与渐近线 $x = 1$ 和 $y = 1$ 吻合。
- 在 $x = e$ 处用中心差分（central difference），得出第二支的斜率为 $-1.0000$。

## 常见错误

- 说每个 $x \neq e$ 都有两个拍档。对 $0 < x \leq 1$，$f(x)$ 的值并非正数，只有 $y = x$ 成立。
- 在最后的草图中漏画直线 $y = x$，或者把另一支画成在 $(e, e)$ 处与直线相切，而不是穿过直线。
- 除以 $\ln x$ 或 $\ln y$（它们可能为零），而不是除以 $xy$。

## STEP 之外

第二条曲线是 $f(x) - f(y)$ 的一个水平集（level set），而 $(e, e)$ 正是该函数的梯度为零的一点。这就是为何在该处有两支曲线相交，而不是只有一条光滑曲线经过。大学一年级分析课程中会遇到的隐函数定理（implicit function theorem），解释了水平集何时在局部上是一个图像，以及不是的时候会发生甚么事。Abbott 的 *Understanding Analysis* 是打好基础的入门好书。方程 $x^y = y^x$ 本身历史悠久；Siklos 的 *Advanced Problems in Mathematics* 中还有更多同类风格的题目。

详细解答

# STEP III 1997，第 3 题

纯数学 · 单位根

**STEP III 1997，第 3 题**

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

step.maths.org/questions/97-s3-q3

## 关键思路

$n$ 次单位根（roots of unity）$1, \omega, \omega^2, \ldots, \omega^{n-1}$ 正好是 $z^n - 1$ 的全部根，所以这个多项式（polynomial）可以完全因式分解（factorise）；除去因式 $z - 1$ 便得到第一个恒等式（identity）。把多边形放在复平面（complex plane）上，以 $O$ 为原点，各顶点就是 $re^{i\theta}$ 乘以各个单位根。这样，向量和就是一个等比级数（geometric series），而弦长的平方则可用点积（dot product）展开 $|\mathbf{a}_k - \mathbf{a}_1|^2$ 而得。

## 解答

### 乘积恒等式

对 $k = 0, 1, \ldots, n-1$，我们有 $(\omega^k)^n = e^{2\pi i k} = 1$，所以每个 $\omega^k$ 都是 $z^n - 1 = 0$ 的根。这 $n$ 个数互不相同，因为它们的辐角（argument）$2\pi k/n$ 互不相同，而且全部位于 $[0, 2\pi)$ 之内。一个 $n$ 次多项式，若首项系数（leading coefficient）为 1 并有 $n$ 个相异的根，就等于相应的一次因式之积，所以

$$z^n - 1 = (z-1)(z-\omega)(z-\omega^2)\cdots(z-\omega^{n-1}).$$

另一方面，由等比级数求和（或者直接乘开），

$$z^n - 1 = (z-1)(1 + z + z^2 + \cdots + z^{n-1}).$$

对每个 $z \neq 1$，我们可以把两个式子同除以 $z - 1$，得出

$$(z-\omega)(z-\omega^2)\cdots(z-\omega^{n-1}) = 1 + z + z^2 + \cdots + z^{n-1} \qquad (z \neq 1).$$

**想一想** 题目说的是“任意复数 $z$”。那么 $z=1$ 又如何？这时我们不能做除法。

两边都是 $n-1$ 次多项式，而它们在无穷多个点 $z\neq 1$ 上相等。于是两者之差是一个有无穷多个根的多项式，所以它是零多项式（zero polynomial）。（另一个做法：两边都是 $z$ 的连续（continuous）函数，所以令 $z\to 1$ 即可。）因此，恒等式对每个复数 $z$ 都成立，包括 $z=1$，这正是所要证明的。

### 向量和

把圆所在的平面视为复平面，以 $O$ 为原点，并以 $a_k$ 表示代表 $A_k$ 的复数。若 $a_1=r\mathrm{e}^{\mathrm{i}\theta}$，则把各顶点按逆时针方向依次标记（相邻两个顶点在 $O$ 所对的角为 $2\pi/n$），便有

$$a_k=r\mathrm{e}^{\mathrm{i}\theta}\,\omega^{k-1},\qquad k=1,2,\ldots,n.$$

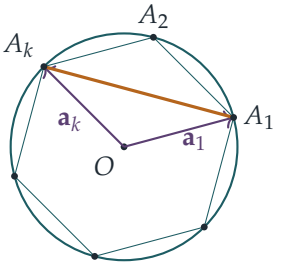
（若改为按顺时针方向标记，只不过是把 $\omega$ 换成 $\bar{\omega}=\omega^{n-1}$，而它所遍历的仍是同一组单位根；况且这个和本来就与顶点的次序无关。）

向量 $\overrightarrow{OA_k}$ 对应复数 $a_k$，而向量相加对应复数相加。所以我们要求的是

$$\sum_{k=1}^{n}a_k=r\mathrm{e}^{\mathrm{i}\theta}\left(1+\omega+\omega^2+\cdots+\omega^{n-1}\right).$$

括号内的式子就是恒等式右边在 $z=\omega$ 时的值。由于 $\omega$ 出现在左边的因式之中（当 $n\geq 2$ 时），左边含有因式 $\omega-\omega=0$，所以括号内的值为 $0$。（同样地，它是一个公比（common ratio）为 $\omega\neq 1$ 的等比级数，所以等于 $(\omega^n-1)/(\omega-1)=0$。）因此

$$\overrightarrow{OA_1}+\overrightarrow{OA_2}+\cdots+\overrightarrow{OA_n}=\mathbf{0}.$$



### 距离平方之和

现在把 $\mathbf{a}_k = \overrightarrow{OA_k}$ 写成向量，于是对每个 $k$ 都有 $|\mathbf{a}_k| = r$，并且 $\overrightarrow{A_1A_k} = \mathbf{a}_k - \mathbf{a}_1$。用点积展开，

$$\begin{aligned} |A_1A_k|^2 &= (\mathbf{a}_k - \mathbf{a}_1)\cdot(\mathbf{a}_k - \mathbf{a}_1) \\ &= |\mathbf{a}_k|^2 + |\mathbf{a}_1|^2 - 2\,\mathbf{a}_1\cdot\mathbf{a}_k \\ &= 2r^2 - 2\,\mathbf{a}_1\cdot\mathbf{a}_k. \end{aligned}$$

这对 $k = 1$ 也成立，此时两边都是 0。对 $k = 1, \ldots, n$ 求和，并利用点积对加法的分配律，

$$\begin{aligned} \sum_{k=1}^{n} |A_1A_k|^2 &= 2r^2n - 2\,\mathbf{a}_1\cdot\left(\sum_{k=1}^{n}\mathbf{a}_k\right) \\ &= 2r^2n - 2\,\mathbf{a}_1\cdot\mathbf{0} = 2r^2n, \end{aligned}$$

这里用了上一部分的结果。这就是所要证明的结果。

## 验算

- $n = 2$：恒等式变成 $z - (-1) = 1 + z$。两点是一条直径的两端，所以和为 $0 + (2r)^2 = 4r^2 = 2r^2\cdot 2$。
- $n = 4$，$r = 1$：由正方形的一个角出发，各距离的平方为 $0, 2, 4, 2$，总和为 $8 = 2\cdot 1^2\cdot 4$。
- 数值上，随机选取 2000 组 $n \le 20$、半径和旋转角度，向量和均为零，弦长平方之和与 $2r^2n$ 的误差在 $10^{-14}$ 以内；乘积恒等式也在随机的 $z$（以及 $z = 1$）上验证过，准确度相同。

## 常见错误

- 除以 $z - 1$ 之后，忘记了这样只证明了恒等式在 $z \neq 1$ 时成立。
- 写成 $|\mathbf{a}_k - \mathbf{a}_1|^2 = |\mathbf{a}_k|^2 - |\mathbf{a}_1|^2$：交叉项 $-2\,\mathbf{a}_1\cdot\mathbf{a}_k$ 正是整个论证的关键。

## STEP 之外

在乘积恒等式中代入 $z = 1$ 并取模（modulus），可得 $|1-\omega|\,|1-\omega^2|\cdots|1-\omega^{n-1}| = n$：单位圆的内接正 $n$ 边形中，由一个顶点出发的各条弦长之积恰好是 $n$，与本题的平方和相映成趣。把 $z^n - 1$ 按单位根因式分解，是分圆多项式（cyclotomic polynomials）和离散傅里叶变换（discrete Fourier transform）的起点，两者都会在大学课程的早期遇到。

Needham 的 *Visual Complex Analysis*（以图像方式讲解复分析）是一本优美的书，很适合用来见识这类以复数处理的几何。

详细解答

# STEP III 1997，第 4 题

纯数学 · 算术几何平均不等式

**STEP III 1997，第 4 题**

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

step.maths.org/questions/97-s3-q4

## 关键思路

题目原文有一处印刷错误：有四个根的多项式（polynomial）必定是四次多项式，而系数的规律 $1,4,6,4,1$ 告诉我们，题目原意的多项式是

$$\mathrm{f}(x) = x^4 - 4ax^3 + 6b^2x^2 - 4c^3x + d^4.$$

这样选取系数，是为了令 $a, b^2, c^3, d^4$ 各自都是一个*平均值*：分别是各根的平均值、两两乘积的平均值、三三乘积的平均值，以及四个根全部相乘之积的平均值。之后每一个不等式，都是把算术平均—几何平均不等式（AM–GM inequality）应用于 f、f′ 或 f″ 的根而得出，而罗尔定理（Rolle's theorem）则确保这些根仍然是相异的正数。

## 解答

设 $\mathrm{f}(x) = (x-p)(x-q)(x-r)(x-s)$。比较系数，得

$$p+q+r+s = 4a, \qquad pq+pr+ps+qr+qs+rs = 6b^2,$$
$$pqr+qrs+rsp+spq = 4c^3, \qquad pqrs = d^4.$$

### 第 (i) 部分

设 $P = pqrs = d^4$。由于没有一个根是零，

$$pqr = \frac{P}{s}, \quad qrs = \frac{P}{p}, \quad rsp = \frac{P}{q}, \quad spq = \frac{P}{r}.$$

映射 $t \mapsto P/t$ 在正实数上是一对一（one-to-one）的，而 $p, q, r, s$ 互不相同，所以这四个乘积亦互不相同。

### 第 (ii) 部分

四个三三乘积是互不相同的正数，所以根据 AM–GM 不等式，

$$c^3 = \frac{pqr + qrs + rsp + spq}{4} > \big(pqr \cdot qrs \cdot rsp \cdot spq\big)^{1/4} = \big((pqrs)^3\big)^{1/4} = (d^4)^{3/4} = d^3.$$

每个根都恰好出现在四个乘积的其中三个之中，因此它们的积是 $(pqrs)^3$。由于 $c, d > 0$，而 $t \mapsto t^3$ 是递增的，故由 $c^3 > d^3$ 可得 $c > d$。

### 第 (iii) 部分

把各根排序，使得 $0 < p < q < r < s$。多项式 f 可微（differentiable），并且在 $p$ 和 $q$ 处取值为零，所以根据罗尔定理，f′ 有一个根 $\alpha$ 满足 $p < \alpha < q$。同理，还有根 $\beta \in (q, r)$ 和 $\gamma \in (r, s)$。这些开区间（open interval）互不重叠，所以 $\alpha < \beta < \gamma$ 互不相同。由于 f′ 是三次多项式，它再没有其他根。还要留意，这三个根全是正数，因为 $\alpha > p > 0$。

### 第 (iv) 部分

求导，得

$$\mathrm{f}'(x) = 4x^3 - 12ax^2 + 12b^2x - 4c^3 = 4\big(x^3 - 3ax^2 + 3b^2x - c^3\big),$$

所以 $\alpha, \beta, \gamma$ 是 $x^3 - 3ax^2 + 3b^2x - c^3$ 的根，并且

$$\alpha + \beta + \gamma = 3a, \qquad \alpha\beta + \beta\gamma + \gamma\alpha = 3b^2, \qquad \alpha\beta\gamma = c^3.$$

这是次数降低了一次的同一种规律，所以我们照搬第 (i) 和第 (ii) 部分的做法。两两乘积是 $\alpha\beta = c^3/\gamma$、$\beta\gamma = c^3/\alpha$、$\gamma\alpha = c^3/\beta$，由于 $\alpha, \beta, \gamma$ 是互不相同的正数，这些乘积亦互不相同。根据 AM–GM 不等式，

$$b^2 = \frac{\alpha\beta + \beta\gamma + \gamma\alpha}{3} > \big((\alpha\beta\gamma)^2\big)^{1/3} = c^2,$$

又因为 $b, c > 0$，所以 $b > c$。

**想一想** 为甚么要先完成第 (iii) 部分，才可以对 f′ 使用 AM–GM?

AM–GM 不等式只适用于正实数，而且我们需要的是严格不等的版本。一个三次多项式根本不一定有三个实根。正是罗尔定理保证了 $\alpha, \beta, \gamma$ 是互不相同的正实数，因而两两乘积也是如此。

### 第 (v) 部分

把同样的论证再用一次。由于 f′ 有相异的根 $\alpha < \beta < \gamma$，根据罗尔定理，

$$\mathrm{f}''(x) = 12x^2 - 24ax + 12b^2 = 12\big(x^2 - 2ax + b^2\big),$$

有根 $\mu \in (\alpha, \beta)$ 和 $\nu \in (\beta, \gamma)$，而且这两个根是相异的正数。于是 $\mu + \nu = 2a$ 及 $\mu\nu = b^2$，再根据两个数的 AM–GM 不等式，

$$a = \frac{\mu + \nu}{2} > \sqrt{\mu\nu} = b.$$

（换一个说法：$x^2 - 2ax + b^2$ 有两个相异实根，所以它的判别式（discriminant）$4a^2 - 4b^2$ 是正数，而两者皆为正数，故 $a > b$。）综合以上各部分，我们证明了

$$a > b > c > d.$$

## 验算

- 取根为 $1, 2, 3, 4$：这时 $4a = 10$，$6b^2 = 35$，$4c^3 = 50$，$d^4 = 24$，所以 $a = 2.5$，$b \approx 2.415$，$c \approx 2.321$，$d \approx 2.213$，次序正确。
- 假如四个根全都等于 $k$，则 $\mathrm{f}(x) = (x - k)^4$ 而 $a = b = c = d = k$。这就是为甚么要有“根互不相同”这个条件，才能得到严格不等式。
- 我们随机测试了 20 000 组由四个相异正根组成的数据，当中包括几乎重合的根，每一次 $a > b > c > d$ 都成立。

## 常见错误

- 直接对 f 的根使用 AM–GM。这样可得 $a > d$，却无法分开中间的两项。
- 忘记检查代入 AM–GM 的数是否互不相同的正数，而这正是第 (i) 和第 (iii) 部分的用意所在。
- 由 $c^3 > d^3$ 或 $b^2 > c^2$ 推出 $c > d$ 或 $b > c$ 时，没有指出所有字母都代表正数。

## STEP 之外

$a, b, c, d$ 这些量是各根的*对称平均*（symmetric means），而不等式链 $a > b > c > d$ 是*麦克劳林不等式*（Maclaurin's inequality）的一个特殊情况：对于正实数，$k$ 重乘积的平均值的 $k$ 次方根，随 $k$ 增加而递减。藉求导来降低次数、同时保持各平均值不变，这个技巧也是通往牛顿（Newton）不等式的标准途径。想了解更多，可参阅 Steele 的 *The Cauchy–Schwarz Master Class*（一本以柯西－施瓦茨不等式为主线的不等式专著），或哈代（Hardy）、李特尔伍德（Littlewood）与波利亚（Pólya）合著的经典 *Inequalities*（《不等式》）。

详细解答

# STEP III 1997，第 5 题

纯数学 · 摆线的弧长

> **STEP III 1997，第 5 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s3-q5

## 关键思路

滚动车轮边缘上的一点，所描出的轨迹是一条摆线（cycloid）。只要把该点的位置用车轮转过的角度表示出来，这道题就只是一个弧长（arc length）积分。滚动条件（不打滑）把轮心的运动与转角连系起来，而半角恒等式（half-angle identity）则令弧长公式中的平方根一下子化简。

## 解答

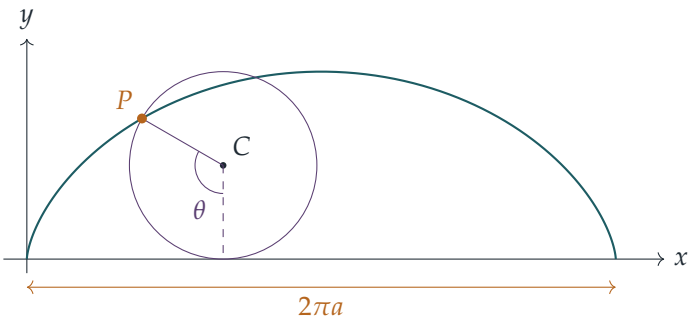
### 建立路径

设车轮半径为 $a$，沿 $x$ 轴作无滑动的滚动。开始时，标记的点 $P$ 在原点与地面接触，并设 $\theta$ 为车轮转过的角度（以弧度计）。

由于没有打滑，车轮走过的地面长度等于与地面接触过的轮缘长度，所以轮心 $C$ 移动了距离 $a\theta$，位于 $(a\theta,\, a)$。相对于 $C$ 而言，点 $P$ 起初在它的正下方，其后转过了角度 $\theta$（车轮向右滚动，故为顺时针方向），所以 $P$ 位于 $C + (-a\sin\theta,\, -a\cos\theta)$。因此，车轮转动一周时，

$$x = a(\theta - \sin\theta), \qquad y = a(1 - \cos\theta), \qquad 0 \le \theta \le 2\pi,$$

在这段时间内，车轮向前移动了距离 $2\pi a$。




### 弧长

求导数，得

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = a(1-\cos\theta), \qquad \frac{\mathrm{d}y}{\mathrm{d}\theta} = a\sin\theta,$$

所以

$$\begin{aligned}\left(\frac{\mathrm{d}x}{\mathrm{d}\theta}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}\theta}\right)^2 &= a^2\left(1 - 2\cos\theta + \cos^2\theta + \sin^2\theta\right)\\ &= 2a^2(1-\cos\theta)\\ &= 4a^2\sin^2\tfrac{\theta}{2},\end{aligned}$$

其中用到 $1-\cos\theta = 2\sin^2\frac{\theta}{2}$。

? **想一想** 取平方根时，结果是 $2a\sin\frac{\theta}{2}$ 还是 $2a\left|\sin\frac{\theta}{2}\right|$？在这里有没有分别？

平方根永远是 $2a\left|\sin\frac{\theta}{2}\right|$。当 $0 \le \theta \le 2\pi$ 时，我们有 $0 \le \frac{\theta}{2} \le \pi$，在这范围内 $\sin\frac{\theta}{2} \ge 0$，所以恰好转动一周时，绝对值（modulus）符号可以省去。（若范围更长，就必须把积分分段计算。）

因此，$P$ 所走的距离为

$$s = \int_0^{2\pi} 2a\sin\tfrac{\theta}{2}\,\mathrm{d}\theta = \Big[-4a\cos\tfrac{\theta}{2}\Big]_0^{2\pi} = -4a(-1) + 4a = 8a.$$

### 比值

车轮移动了 $2\pi a$，而轮缘上的点则走了 $8a$，所以所求的比值为

$$\frac{2\pi a}{8a} = \frac{\pi}{4} \approx 0.785.$$

这个比值与车轮的大小无关，而这亦是理所当然的：把车轮放大或缩小，两段距离都会按相同的倍数改变。

## 验算

- 该点所走的距离比车轮的长（$8a > 2\pi a \approx 6.28a$），这是合理的：它既向前移动，同时又上升和下降各 $2a$。事实上，$8a$ 恰好是四条直径的长度。
- 一个粗略的估计：路径由 $(0,0)$ 上升至 $(\pi a, 2a)$，再下降至 $(2\pi a, 0)$，所以它的长度至少是 $2\sqrt{\pi^2 a^2 + 4a^2} \approx 7.45a$，而至多是 $2\pi a + 4a \approx 10.28a$。数值 $8a$ 正好落在两者之间。
- 以数值方法验证：在车轮滚动的过程中，追踪轮缘上的点的 200 000 个位置，再把各段直线距离相加，当 $a = 1$ 时得出 7.99999999992，所以比值是 $0.78539816\ldots = \pi/4$。

## 常见错误

- 把参数式（parametrisation）中的正负号弄错（例如 $x = a(\theta + \sin\theta)$），这样得出的路径相对于车轮是向后移动的。可在 $\theta = \pi$ 处检查：$P$ 应位于最高点 $(\pi a, 2a)$。
- 把被积函数（integrand）保留为 $a\sqrt{2 - 2\cos\theta}$，并试图直接积分；半角恒等式正是令积分变得容易的一步。
- 把比值上下颠倒：题目要求的是车轮所走的距离除以该点所走的距离，即 $\pi/4$，而不是 $4/\pi$。

## STEP 之外

摆线是数学中最著名的曲线之一。它是最速降线（brachistochrone），即在重力作用下，两点之间下降得最快的曲线；它也是等时降线（tautochrone），即珠子无论从曲线上哪一点出发，都会在相同时间内到达最低点。这两项事实都要用变分法（calculus of variations）来证明，你会在大学一年级或二年级的力学或数学方法课程中学到；而本题的弧长计算，正是两者的第一步。如想尝试更多同类的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

# STEP III 1997，第 6 题

纯数学 · 切比雪夫微分方程

> **STEP III 1997，第 6 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s3-q6

## 关键思路

因子 $1-x^2$ 与 $x=\cos\theta$ 并列，只暗示一件事：$1-\cos^2\theta=\sin^2\theta$。只要用链式法则（chain rule）小心地换元，这条关于 $x$ 的棘手方程，就会变成关于 $\theta$ 的简谐运动（simple harmonic motion）方程。我们解出该方程，再用两个附带条件定出常数。至于递推关系（recurrence），其实是乔装了的复角恒等式（compound-angle identity）。全题中 $n$ 为非负整数。条件 $y_n(x)=(-1)^n y_n(-x)$ 只有在这个情况下才有意义。

## 解答

### 变量代换

设 $x=\cos\theta$，其中 $0<\theta<\pi$，故 $\sin\theta>0$，且 $\mathrm{d}x/\mathrm{d}\theta=-\sin\theta\neq 0$。为简化记号，我们把 $y_n$ 写成 $y$。由链式法则，

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\mathrm{d}y/\mathrm{d}\theta}{\mathrm{d}x/\mathrm{d}\theta}=-\frac{1}{\sin\theta}\frac{\mathrm{d}y}{\mathrm{d}\theta}.$$

要再对 $x$ 求导一次，我们先对 $\theta$ 求导，再乘以 $\mathrm{d}\theta/\mathrm{d}x=-1/\sin\theta$：

$$\begin{aligned}\frac{\mathrm{d}^2y}{\mathrm{d}x^2}&=-\frac{1}{\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(-\frac{1}{\sin\theta}\frac{\mathrm{d}y}{\mathrm{d}\theta}\right)=\frac{1}{\sin\theta}\left(\frac{1}{\sin\theta}\frac{\mathrm{d}^2y}{\mathrm{d}\theta^2}-\frac{\cos\theta}{\sin^2\theta}\frac{\mathrm{d}y}{\mathrm{d}\theta}\right)\\&=\frac{1}{\sin^2\theta}\frac{\mathrm{d}^2y}{\mathrm{d}\theta^2}-\frac{\cos\theta}{\sin^3\theta}\frac{\mathrm{d}y}{\mathrm{d}\theta}.\end{aligned}$$

现在利用 $1-x^2=\sin^2\theta$ 和 $x=\cos\theta$ 代入：

$$\begin{aligned}(1-x^2)\frac{\mathrm{d}^2y}{\mathrm{d}x^2}-x\frac{\mathrm{d}y}{\mathrm{d}x}+n^2y&=\frac{\mathrm{d}^2y}{\mathrm{d}\theta^2}-\frac{\cos\theta}{\sin\theta}\frac{\mathrm{d}y}{\mathrm{d}\theta}+\frac{\cos\theta}{\sin\theta}\frac{\mathrm{d}y}{\mathrm{d}\theta}+n^2y\\&=\frac{\mathrm{d}^2y}{\mathrm{d}\theta^2}+n^2y.\end{aligned}$$

两个一阶导数项刚好互相抵消。因此题目所给的方程等价于 $\dfrac{\mathrm{d}^2y_n}{\mathrm{d}\theta^2}+n^2y_n=0$，正是所求。

> ? **想一想** 为甚么 $\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2}$ 不就是 $\dfrac{1}{\sin^2\theta}\dfrac{\mathrm{d}^2y}{\mathrm{d}\theta^2}$？

因为 $\mathrm{d}y/\mathrm{d}x$ 是一个乘积：$-1/\sin\theta$ 乘以 $\mathrm{d}y/\mathrm{d}\theta$。两个因子都取决于 $\theta$，所以用积法则（product rule）会得出两项。多出来的那一项，恰好抵消了 $-x\,\mathrm{d}y/\mathrm{d}x$ 一项，这正是题目中有这一项的原因。

### 求解 $y_n$

当 $n\geq 1$ 时，$y''+n^2y=0$ 的通解（general solution）为

$$y_n=A\cos n\theta+B\sin n\theta.$$

当 $n=0$ 时，方程是 $y''=0$，故 $y_0=A+B\theta$。

*在 $x=1$ 的条件。* $x=1$ 对应 $\theta=0$，而由 $y_n(1)=1$ 可知两种情况下都有 $A=1$。

*对称条件。* 把 $x$ 换成 $-x$，即是把 $\theta$ 换成 $\pi-\theta$，因为 $\cos(\pi-\theta)=-\cos\theta$，而且 $\pi-\theta$ 仍然在 $[0,\pi]$ 之内。对于整数 $n\geq 1$，

$$\cos n(\pi-\theta)=(-1)^n\cos n\theta,\qquad \sin n(\pi-\theta)=-(-1)^n\sin n\theta.$$

所以 $(-1)^n y_n(-x)=(-1)^n\left[A(-1)^n\cos n\theta-B(-1)^n\sin n\theta\right]=A\cos n\theta-B\sin n\theta$。令此式对所有 $\theta$ 都等于 $y_n(x)=A\cos n\theta+B\sin n\theta$，得 $2B\sin n\theta=0$，故 $B=0$。当 $n=0$ 时，条件变成对所有 $\theta$ 都有 $1+B\theta=1+B(\pi-\theta)$，同样必须有 $B=0$。

因此，对每个非负整数 $n$，

$$\boxed{y_n=\cos n\theta},\qquad \text{即}\quad y_n(x)=\cos(n\arccos x)\quad(|x|\leq 1).$$

关系式 $\theta=\arccos x$ 把 $[-1,1]$ 映满 $[0,\pi]$，所以结果对 $|x|\leq 1$ 成立。端点 $x=\pm 1$（$\theta=0,\pi$）则由连续性（continuity）补上。

### 推论

取 $n=0$ 和 $n=1$：$y_0=\cos 0=1$，$y_1=\cos\theta=x$。

至于递推关系，把以下两条复角公式相加：

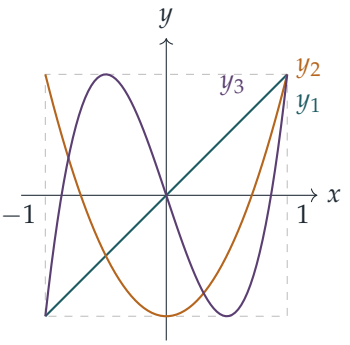
$$\cos(n+1)\theta = \cos n\theta\cos\theta - \sin n\theta\sin\theta, \qquad \cos(n-1)\theta = \cos n\theta\cos\theta + \sin n\theta\sin\theta.$$

正弦项互相抵消，余下 $\cos(n+1)\theta + \cos(n-1)\theta = 2\cos\theta\cos n\theta$。以 $x$ 表示，即

$$y_{n+1} - 2xy_n + y_{n-1} = 0 \qquad (n \geq 1,\ |x| \leq 1),$$

这就是所要求的结果。

由 $y_0 = 1$ 和 $y_1 = x$ 出发，递推关系会产生一个个多项式（polynomial）：$y_2 = 2x^2 - 1$、$y_3 = 4x^3 - 3x$、$y_4 = 8x^4 - 8x^2 + 1$，如此类推。




每个 $y_n$ 在 $[-1, 1]$ 上都介乎 $-1$ 与 $1$ 之间；它既然是余弦，这是必然的。

## 验算

- $y_2 = 2x^2 - 1$：此时 $(1-x^2)\cdot 4 - x\cdot 4x + 4(2x^2-1) = 0$。它满足 $n = 2$ 时的方程，而且 $y_2(1) = 1$，并且是偶函数（even function）。
- $y_n$ 的奇偶性（parity）与 $(-1)^n$ 吻合。例如 $y_3 = 4x^3 - 3x$ 是奇函数（odd function）。
- 数值上，取 $x = 0.3$：$\cos(2\arccos 0.3) = -0.82 = 2(0.3)^2 - 1$。再以有限差分（finite difference）检验 $y_n = \cos(n\arccos x)$（$0 \leq n \leq 8$）是否满足原方程，所得残差（residual）均小于 $10^{-5}$。

## 常见错误

- 求二阶导数时忘记用积法则，以致一阶导数项无法抵消。
- 保留了 $\sin n\theta$ 这个解。它满足微分方程（differential equation）和 $y_n(1) = 1$（因为 $\sin 0 = 0$），但不符合对称条件。题目设下这个条件，正是为了把它剔除。
- 忘记 $n = 0$ 要另外处理：这时通解是 $A + B\theta$，而不是一对余弦和正弦。

## STEP 之外

多项式 $y_n$ 就是**切比雪夫多项式**（Chebyshev polynomials）$T_n(x)$。题目所给的方程就是切比雪夫方程。我们舍弃的第二个解 $\sin(n \arccos x)$，是 $\sqrt{1-x^2}$ 乘以一个多项式。切比雪夫多项式在逼近理论（approximation theory）中占有核心地位。在所有 $n$ 次首一多项式（monic polynomial）之中，$2^{1-n}T_n$ 在 $[-1,1]$ 上的最大绝对值最小，因此对许多数值方法而言，$T_n$ 的零点是最佳的插值（interpolation）节点。至于这类方程的级数解（series solution），以及由此衍生的特殊函数（special functions），可参阅 Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科书）。

详细解答

# STEP III 1997，第 7 题

纯数学 • e 的无理性

> **STEP III 1997，第 7 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s3-q7

## 关键思路

级数（series）$a_n$ 正好就是 $n!\,\mathrm{e}$ 去掉整数部分（integer part）之后余下的东西：把级数 $\mathrm{e} = \sum 1/k!$ 乘以 $n!$，直至 $k = n$ 为止的每一项都变成整数，而尾段则变成 $a_n$。等比级数（geometric series）$b_n$ 的作用，是逐项为 $a_n$ 提供一个容易处理的上界（upper bound）。一旦知道尾段严格介乎 0 与 1 之间，假如 e 是有理数（rational number），就会有某个 $n!\,\mathrm{e}$ 既是整数，小数部分（fractional part）却又不等于零，这是荒谬的。

## 解答

### 第 (i) 部分

级数 $b_n$ 是等比级数，首项为 $\frac{1}{n+1}$，公比（common ratio）为 $r = \frac{1}{n+1}$。由于 $n \geq 1$，我们有 $0 < r \leq \frac{1}{2} < 1$，所以它收敛（converges），而且

$$b_n = \frac{\frac{1}{n+1}}{1 - \frac{1}{n+1}} = \frac{\frac{1}{n+1}}{\frac{n}{n+1}} = \frac{1}{n}.$$

### 第 (ii) 部分

把两个级数的第 $k$ 项分别写成

$$u_k = \frac{1}{(n+1)(n+2)\cdots(n+k)}, \qquad v_k = \frac{1}{(n+1)^k}.$$

$u_k$ 的分母中有 $k$ 个因子 $n+1, n+2, \ldots, n+k$，每一个都至少是 $n+1$，所以 $(n+1)(n+2)\cdots(n+k) \geq (n+1)^k$，因此

$$0 < u_k \leq v_k \quad \text{对所有} k \geq 1.$$

等号只在 $k=1$ 时成立；当 $k\geq 2$ 时，因子 $n+2$ 严格大于 $n+1$，所以 $u_k<v_k$。

所有项都是正数，并且不超过收敛级数 $b_n$ 的对应项，所以 $a_n$ 收敛（根据比较判别法（comparison test）），而且 $a_n\geq u_1>0$。至于上界，

$$b_n-a_n=\sum_{k\geq 1}(v_k-u_k)\ \geq\ v_2-u_2=\frac{1}{(n+1)^2}-\frac{1}{(n+1)(n+2)}>0,$$

因为每一项 $v_k-u_k$ 都是非负的。于是 $a_n<b_n=\frac{1}{n}$，合起来便得

$$0<a_n<\frac{1}{n},$$

这正是所要证明的。

> **想一想** 为甚么单凭“$a_n$ 的每一项都不超过 $b_n$ 的对应项”，并不足以得出严格不等式（strict inequality）？

逐项的 $\leq$ 在取极限后只能给出 $a_n\leq b_n$；把无穷多项相加时，严格不等式有可能变成等式。正因如此，我们挑出 $k=2$ 这一项，它的差距是一个确定的正数，再指出其余各项的差距全都 $\geq 0$。

### 第 (iii) 部分

利用级数 $\mathrm{e}=\sum_{k=0}^{\infty}\frac{1}{k!}$，乘以 $n!$，并在 $k=n$ 处把和式拆开：

$$n!\,\mathrm{e}=\underbrace{\sum_{k=0}^{n}\frac{n!}{k!}}_{N}+\sum_{k=n+1}^{\infty}\frac{n!}{k!}.$$

当 $k\leq n$ 时，$\frac{n!}{k!}=(k+1)(k+2)\cdots n$ 是整数，所以 $N$ 是正整数。当 $k=n+j$ 而 $j\geq 1$ 时，

$$\frac{n!}{(n+j)!}=\frac{1}{(n+1)(n+2)\cdots(n+j)}=u_j,$$

所以尾段正好就是 $a_n$。因此

$$n!\,\mathrm{e}=N+a_n,\qquad N\in\mathbb{Z},\quad 0<a_n<\tfrac{1}{n}\leq 1,$$

这里用到第 (ii) 部分及 $n\geq 1$。一个形如（整数）+（$[0,1)$ 内的某数）的数，其整数部分就是那个整数，所以 $[n!\,\mathrm{e}]=N$，因此

$$a_n=n!\,\mathrm{e}-[n!\,\mathrm{e}].$$

### 第 (iv) 部分

采用反证法（proof by contradiction），假设 $\mathrm{e}=\frac{p}{q}$，其中 $p,q$ 为正整数。取 $n=q$，则

$$n!\,\mathrm{e}=q!\cdot\frac{p}{q}=(q-1)!\,p$$

是整数，所以它等于自身的整数部分，而由第 (iii) 部分得 $a_q = 0$。但第 (ii) 部分指出 $a_q > 0$。这个矛盾说明 e 是无理数（irrational number）。

（任何 $n \geq q$ 同样可行，因为 $q$ 整除 $n!$；重点只在于必定有某个 $n!\,\mathrm{e}$ 是整数。）

## 验算

当 $n = 1$ 时：$1!\,\mathrm{e} = 2.71828\ldots$，所以小数部分是 $0.71828\ldots$，而把 $\frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \cdots$ 加起来同样得到 $0.71828\ldots$，明显小于 $b_1 = 1$。当 $n = 5$ 时，$a_5 \approx 0.19382$，仅仅略小于 $\frac{1}{5}$：随着 $n$ 增大，这个界相对而言愈来愈紧，因为 $a_n$ 与 $b_n$ 的首项相同。用电脑以精确分数验算，可确认对 1 至 29 的每一个 $n$，都有 $0 < a_n < \frac{1}{n}$ 以及 $n!\,\mathrm{e} = N + a_n$，其中 $[n!\,\mathrm{e}] = N$。

## 常见错误

- 仅仅因为对所有 $k$ 都有 $u_k \leq v_k$，就声称 $a_n < b_n$。你需要有一项是严格较小的（这里是 $k = 2$），才能保住严格不等式。
- 忘记了 $\sum_{k \leq n} n!/k!$ 为何是整数：原因是当 $k \leq n$ 时，$k!$ 整除 $n!$。
- 在第 (iv) 部分只用 $0 < a_n$，或只用 $a_n < 1$。两边都不可或缺：$a_n < 1$（来自第 (ii) 部分）令我们能够确定 $a_n$ 就是小数部分，而 $a_n > 0$ 则带来矛盾。

## STEP 之外

这是傅里叶（Fourier）对 e 是无理数的证明，而它的模式（证明一个本应是整数的量严格介乎 0 与 1 之间）是许多无理性证明的蓝本，包括 $\pi$ 的无理性，以及 $r \neq 0$ 为有理数时 $\mathrm{e}^r$ 的无理性。在大学第一门分析（analysis）课程中，你会严格地论证比较判别法和 e 的级数，并且明白这个论证之所以行得通，全靠 $\sum 1/k!$ 收敛得够快：分数逼近 e 的效果，远远好于逼近任何有理数时所能做到的。Abbott 的 *Understanding Analysis* 涵盖级数方面的内容，而 Davenport 的 *The Higher Arithmetic* 则是有理逼近（rational approximation）的一本出色入门书。

详细解答

# STEP III 1997，第 8 题

纯数学 · 二次型的对角化

**STEP III 1997，第 8 题**

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

step.maths.org/questions/97-s3-q8

## 关键思路

一个对称矩阵（symmetric matrix）描述一个二次型（quadratic form）$ax^2+2bxy+cy^2$，而共轭（conjugate）后的矩阵 $R_{-\alpha}AR_\alpha$ 则描述同一个二次型在旋转了 $\alpha$ 的坐标轴下的样子。把它化成对角矩阵（diagonal matrix），$xy$ 项便会消失，于是在新坐标轴下椭圆（ellipse）处于标准位置，其半轴（semi-axes）长度可以直接读出。由于旋转保持各点与原点的距离不变，这两条半轴的长度正是我们要找的极端距离。

## 解答

全题记

$$R_\alpha=\begin{pmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{pmatrix},\qquad R_{-\alpha}=\begin{pmatrix}\cos\alpha & \sin\alpha\\ -\sin\alpha & \cos\alpha\end{pmatrix}=R_\alpha^{\mathsf T}.$$

### 第 (i) 部分

设 $C=\cos\alpha$ 及 $S=\sin\alpha$。首先

$$AR_\alpha=\begin{pmatrix}aC+bS & -aS+bC\\ bC+cS & -bS+cC\end{pmatrix},$$

再从左边乘以 $R_{-\alpha}$，右上角的元素是

$$\begin{aligned}C(-aS+bC)+S(-bS+cC)&=b(C^2-S^2)-(a-c)SC\\&=b\cos 2\alpha-\tfrac12(a-c)\sin 2\alpha.\end{aligned}$$

因为 $A$ 是对称的，所以 $R_{-\alpha}AR_\alpha=R_\alpha^{\mathsf T}AR_\alpha$ 也是对称的（它的转置（transpose）就是它本身），故左下角的元素与右上角相同，一个条件便同时处理了两者。这个矩阵是对角矩

阵，当且仅当

$$\tan 2\alpha = \frac{2b}{a-c} \quad (a \neq c), \qquad \text{故可取} \qquad \alpha = \tfrac{1}{2}\arctan\frac{2b}{a-c}.$$

若 $a = c$，条件变成 $b\cos 2\alpha = 0$，取 $\alpha = \frac{1}{4}\pi$ 即可（若同时有 $b = 0$，则 $A$ 本身已是对角矩阵，任何 $\alpha$ 都可以）。

**第 (ii) 部分**

把椭圆的方程展开，得

$$x^2 + y^2 + 4xy\cot 2\theta + 4x^2\cot^2 2\theta = 1,$$

而 $\begin{pmatrix} x & y \end{pmatrix} A \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + 2bxy + cy^2$。比较系数，得

$$a = 1 + 4\cot^2 2\theta, \qquad b = 2\cot 2\theta, \qquad c = 1.$$

于是 $a - c = 4\cot^2 2\theta \neq 0$（因为 $0 < 2\theta < \frac{1}{2}\pi$），并且

$$\frac{2b}{a-c} = \frac{4\cot 2\theta}{4\cot^2 2\theta} = \tan 2\theta,$$

所以 $\alpha = \theta$ 满足第 (i) 部分的条件：$R_{-\theta}AR_\theta$ 是对角矩阵。

至于对角线上的元素，按第 (i) 部分同样的乘法可得

$$\text{左上角} = aC^2 + 2bSC + cS^2, \qquad \text{右下角} = aS^2 - 2bSC + cC^2,$$

这里 $C = \cos\theta$, $S = \sin\theta$。设 $k = \cot 2\theta$ 及 $\mu = \cos 2\theta$, 并利用 $2SC = \sin 2\theta$、$k\sin 2\theta = \mu$、$C^2 = \frac{1}{2}(1+\mu)$、$S^2 = \frac{1}{2}(1-\mu)$ 以及 $k^2 = \mu^2/(1-\mu^2)$。对于右下角的元素，

$$\begin{aligned}(1+4k^2)S^2 - 2k\sin 2\theta + C^2 &= 1 + 4k^2S^2 - 2\mu \\ &= 1 - 2\mu + \frac{2\mu^2(1-\mu)}{(1-\mu)(1+\mu)} = \frac{(1-2\mu)(1+\mu)+2\mu^2}{1+\mu} \\ &= \frac{1-\mu}{1+\mu} = \frac{2\sin^2\theta}{2\cos^2\theta} = \tan^2\theta.\end{aligned}$$

用完全相同的方法，左上角的元素是 $1 + 2\mu + \dfrac{2\mu^2}{1-\mu} = \dfrac{1+\mu}{1-\mu} = \cot^2\theta$。所以

$$R_{-\theta}AR_\theta = \begin{pmatrix} \cot^2\theta & 0 \\ 0 & \tan^2\theta \end{pmatrix}.$$

**想一想** 有没有快捷的方法，不必重做一次代数运算，也能确认这两个元素是对的?

有。共轭不会改变矩阵的迹（trace）和行列式（determinant）。这里 $\det A = ac - b^2 = 1 + 4k^2 - 4k^2 = 1$，而的确 $\cot^2\theta \tan^2\theta = 1$。另外 $\operatorname{tr} A = 2 + 4\cot^2 2\theta$，而由于 $\cot\theta - \tan\theta = \dfrac{\cos^2\theta - \sin^2\theta}{\sin\theta\cos\theta} = 2\cot 2\theta$，

$$\cot^2\theta + \tan^2\theta = (\cot\theta - \tan\theta)^2 + 2 = 4\cot^2 2\theta + 2.$$

**第 (iii) 部分**

以 $\begin{pmatrix} x \\ y \end{pmatrix} = R_\theta \begin{pmatrix} X \\ Y \end{pmatrix}$ 引入新坐标，使 $X$ 轴和 $Y$ 轴就是原来的坐标轴旋转 $\theta$ 所得。由于 $\begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} X & Y \end{pmatrix} R_\theta^{\mathsf{T}} = \begin{pmatrix} X & Y \end{pmatrix} R_{-\theta}$，椭圆的方程变成

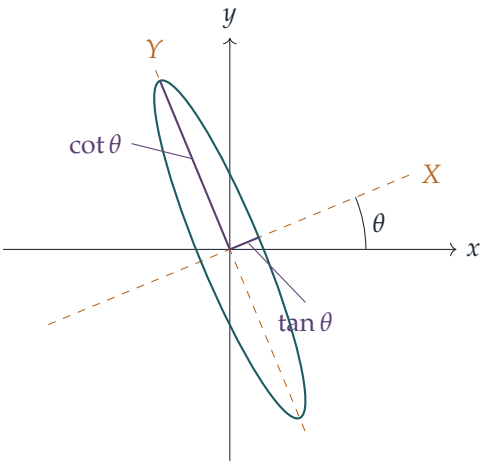
$$\begin{pmatrix} X & Y \end{pmatrix} R_{-\theta} A R_\theta \begin{pmatrix} X \\ Y \end{pmatrix} = 1, \qquad \text{即} \qquad X^2\cot^2\theta + Y^2\tan^2\theta = 1.$$

这个方程在 $(X, Y) \mapsto (-X, -Y)$ 之下不变，所以椭圆的中心是原点。旋转保持长度不变，因此中心到某一点的距离是 $r = \sqrt{x^2 + y^2} = \sqrt{X^2 + Y^2}$。

因为 $0 < \theta < \frac{1}{4}\pi$，我们有 $0 < \tan\theta < 1 < \cot\theta$，所以 $\tan^2\theta < \cot^2\theta$。对椭圆上任何一点，

$$\tan^2\theta\,(X^2 + Y^2) \;\le\; X^2\cot^2\theta + Y^2\tan^2\theta = 1 \;\le\; \cot^2\theta\,(X^2 + Y^2),$$

所以 $\tan^2\theta \le r^2 \le \cot^2\theta$，即 $\tan\theta \le r \le \cot\theta$。左边的不等式在 $X = 0$ 时取等号，即在 $(X, Y) = (0, \pm\cot\theta)$；右边的不等式在 $Y = 0$ 时取等号，即在 $(X, Y) = (\pm\tan\theta, 0)$。两个界限都能达到，所以中心到椭圆的最短和最长距离分别是 $\tan\theta$ 和 $\cot\theta$，正如题目所求。



上图所示为 $\theta = \frac{1}{8}\pi$ 的情况，虚线是旋转后的坐标轴。

## 验算

- $(0, \pm 1)$ 在椭圆上，而 $\tan\theta < 1 < \cot\theta$，与上述界限吻合。
- 当 $\theta = \frac{1}{8}\pi$ 时，长轴（major axis）的端点在原坐标下是 $\cot\theta\,(-\sin\theta, \cos\theta) = (-\cos\theta, \cos^2\theta/\sin\theta) \approx (-0.924,\, 2.230)$。这时 $\cot 2\theta = 1$，而 $0.924^2 + (2.230 - 1.848)^2 \approx 0.854 + 0.146 = 1$，正如所料。
- 数值上，对若干个 $\theta$ 值沿椭圆扫描 200 000 个方向，在 $\theta = 0.3$ 时得出最短距离 0.309336 和最长距离 3.232728，与 $\tan 0.3$ 和 $\cot 0.3$ 相符。

## 常见错误

- 把 $xy$ 的系数当作 $b$ 而不是 $2b$：二次型是 $ax^2 + 2bxy + cy^2$。
- 搞乱了哪个旋转放在哪一边：若 $\mathbf{x} = R_\theta \mathbf{X}$，行向量便是 $\mathbf{X}^\mathsf{T} R_{-\theta}$，这就是出现 $R_{-\theta} A R_\theta$ 的原因。
- 以为系数较大的那条轴较长：$\cot^2\theta X^2$ 的系数较大，所以椭圆在 $X$ 方向上反而是短的。

## STEP 之外

这是谱定理（spectral theorem）的二维情况：每个实对称矩阵都可以用旋转对角化（diagonalise），而对角线上的元素就是它的特征值（eigenvalues）。二次型的特征值可用来为圆锥曲线（conics）和二次曲面（quadric surfaces）分类；同一个想法在力学中给出惯量主轴（principal axes of inertia），在统计学中给出主成分（principal components）。Strang 的 *Introduction to Linear Algebra*（线性代数入门）对对称矩阵和二次型的讲解十分易读。

详细解答

# STEP III 1997，第 9 题

力学 • 涉及弹簧的静力学

> **STEP III 1997，第 9 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s3-q9

## 关键思路

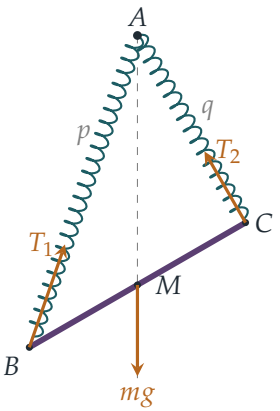
作用在杆上的力只有三个：两条弹簧的张力（tension）和杆本身的重量。当一个物体在三个互不平行的力作用下处于平衡（equilibrium）时，这三个力的作用线（line of action）必定通过同一点。两个张力的作用线都通过 $A$，所以重量的作用线也必须通过 $A$，亦即杆的中点位于 $A$ 的正下方。之后，沿水平方向分解力，几乎就完成余下的工作。

## 解答

我们把杆视为均匀（uniform）的刚体（rigid body），质量为 $m$，中点为 $M$。两条弹簧都是轻弹簧，遵从虎克定律（Hooke's law），弹簧常数同为 $\kappa$，而且都处于拉伸状态。设弹簧 $AB$ 和 $AC$ 的自然长度（natural length）分别为 $a$ 和 $b$，拉伸后的长度分别为 $p$ 和 $q$，则张力为

$$T_1 = \kappa(p - a), \qquad T_2 = \kappa(q - b),$$

由于弹簧都被拉长，两者皆为正数。




**第 1 步：$M$ 位于 $A$ 的正下方**

对杆取关于 $A$ 的力矩（moment）。张力 $T_1$ 沿 $BA$ 作用，$T_2$ 沿 $CA$ 作用，所以两条作用线都通过 $A$，两者关于 $A$ 的力矩均为零。因此，垂直通过 $M$ 的重量关于 $A$ 的力矩也必须为零，即通过 $M$ 的铅垂线通过 $A$。于是 $M$ 位于 $A$ 的正下方；设它在 $A$ 下方的深度为 $h > 0$。

**第 2 步：沿水平方向分解**

以 $A$ 为原点建立坐标轴，$x$ 轴为水平方向（在杆所在的铅垂平面内），$y$ 轴铅垂向上。则 $M = (0, -h)$，而由于 $M$ 是 $BC$ 的中点，我们可以写成

$$B = (-d,\ -h-e), \qquad C = (d,\ -h+e)$$

其中 $d$、$e$ 为某些常数。这里 $d$ 是杆在水平方向上的半阔度，而 $d \neq 0$ 正是因为杆并非铅垂。

$AB$ 中的张力作用于杆上的 $B$ 点，方向由 $B$ 指向 $A$，即沿向量（vector）$\overrightarrow{BA} = (d,\ h+e)$，其长度为 $p$。所以这个力是 $\dfrac{T_1}{p}(d,\ h+e)$。同理，在 $C$ 点的张力是 $\dfrac{T_2}{q}(-d,\ h-e)$。

沿水平方向分解，得

$$\frac{T_1}{p}\,d - \frac{T_2}{q}\,d = 0,$$

而由于 $d \neq 0$，

$$\frac{T_1}{p} = \frac{T_2}{q}.$$

? **想一想** 我们究竟在哪里用到“杆并非铅垂”这个事实?

只在约去 $d$ 的时候。假如杆铅垂悬挂，两条弹簧便会落在同一条铅垂线上，水平方向的

分解式只会变成 $0=0$，而两个张力可以成任何比例，只需满足铅垂方向的平衡。这种铅垂的平衡状态确实会出现，而在这些情况下，结论一般并不成立。

### 第 3 步：引入虎克定律

代入张力，得

$$\frac{\kappa(p-a)}{p}=\frac{\kappa(q-b)}{q}\quad\Longrightarrow\quad 1-\frac{a}{p}=1-\frac{b}{q}\quad\Longrightarrow\quad \frac{a}{p}=\frac{b}{q},$$

所以

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

拉伸后长度之比等于自然长度之比，这正是所要证明的。 ■

为完整起见，沿铅垂方向分解可得 $\dfrac{T_1}{p}(h+e)+\dfrac{T_2}{q}(h-e)=mg$，即 $2h\dfrac{T_1}{p}=mg$。我们用不着这条式，但它确定了 $T/$长度 这个共同的值。

## 验算

- 诠释。由于 $p/q=a/b$，两条弹簧都按相同的倍数伸长：对某个 $\lambda>1$，有 $p=\lambda a$ 及 $q=\lambda b$。所以这两条弹簧就像一条均匀的弹簧按比例放大一样，这是记住答案的一个好方法。
- 数值检验。我们随机选取 $m,\kappa,a,b$ 和杆的长度，以数值方法求势能（potential energy）$mgy_M+\frac{1}{2}\kappa(p-a)^2+\frac{1}{2}\kappa(q-b)^2$ 的最小值，从而找出平衡状态。例如，当 $a/b=0.709972$ 时，一个倾斜的平衡状态给出 $p/q=0.709972$，而 $M$ 位于 $A$ 的正下方，误差在 $10^{-7}$ 之内。同一次搜索亦找到一些 $p/q\neq a/b$ 的铅垂平衡状态（例如 $a/b=1.526$ 但 $p/q=1.192$），这证实了题目的假设确有必要。

## 常见错误

- 假设两个张力相等，或假设两条弹簧与铅垂线所成的角相等。两者一般都不成立；相等的是张力除以长度。
- 对 $M$ 取力矩，结果纠缠于各个角度之中。对 $A$ 取力矩简洁得多，因为两个张力的作用线都通过 $A$。
- 从没用到“杆并非铅垂”这个条件。完整的答案应指出哪里需要这个条件。

## STEP 之外

我们用到的向量 $\frac{T}{p}\overrightarrow{BA}$ 是表示弹簧力的自然写法。自然长度为零的弹簧所施的力，恰好与位移向量成正比，而这类“理想”弹簧能把许多平衡问题化为线性代数（linear algebra）。更一般地说，平衡状态是势能的驻点（stationary point），这正是我们数值检验背后的想法，也是拉格朗日力学（Lagrangian mechanics）的起点。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）对静力学（statics）和能量观点都有细致的阐述。

详细解答

# STEP III 1997，第 10 题

力学 · 有滑动的滚动

**STEP III 1997，第 10 题**

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

step.maths.org/questions/97-s3-q10

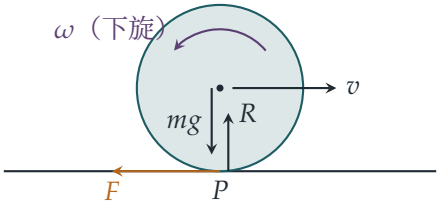
本题在 *Beyond STEP* 第 11.6 节*后旋*中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch11/backspin.html

## 关键思路

关键在于弹珠的底部是否正在桌面上滑动。只要它在滑动，摩擦力（friction）的大小和方向都固定不变，因此线运动和角运动都是匀减速的，可以准确地写出来。当接触点相对于桌面静止时，滑动便告结束，此后弹珠作滚动，完全不受摩擦力。所以整道题目问的就是哪一件事最先发生：$v = 0$、$\omega = 0$，还是滑动速度（slip velocity）降至零。

## 解答

**模型。** 弹珠是质量为 $m$、半径为 $a$ 的均匀实心球，所以它对直径的转动惯量（moment of inertia）为 $I = \frac{2}{5}ma^2$。取 $v$ 为球心的速度（以 $v_0$ 的方向为正），$\omega$ 为*向后*转动方向的角速度（angular velocity）（下旋（backspin）为正），于是 $v(0) = v_0 > 0$ 及 $\omega(0) = \omega_0 > 0$。手指只负责令弹珠开始运动；其后作用于弹珠的力只有重量 $mg$、法向反作用力（normal reaction）$R = mg$ 以及接触点 $P$ 处的摩擦力 $F$。



在下旋的情况下，弹珠底部相对于球心以速率 $a\omega$ 向前运动，所以接触点的速度为

$$s = v + a\omega.$$

起初 $s = v_0 + a\omega_0 > 0$：弹珠在桌面上向前打滑，所以摩擦力向后作用，大小恒为 $F$。记 $F = mk$。

**滑动期间的运动方程**。 沿水平方向分解，得 $m\dot{v} = -F$。对球心取力矩（moment），底部向后的摩擦力会使弹珠朝向前滚动的方向转动，因此它会减弱下旋：

$$\tfrac{2}{5}ma^2\,\dot{\omega} = -Fa \quad\Longrightarrow\quad \dot{\omega} = -\frac{5k}{2a}.$$

两者皆为常数，所以只要弹珠继续向前滑动，

$$v = v_0 - kt, \qquad a\omega = a\omega_0 - \tfrac{5}{2}kt, \qquad s = v_0 + a\omega_0 - \tfrac{7}{2}kt.$$

因此滑动速度降至零的时刻为

$$t_1 = \frac{2(v_0 + a\omega_0)}{7k},$$

而在此之前摩擦力的方向一直不变，所以这些公式在整个区间 $0 \le t \le t_1$ 上都成立，即使 $v$ 或 $\omega$ 在途中改变了符号亦然。

> ? **想一想** 在 $t = t_1$ 时会发生甚么事？为甚么弹珠不会随即向相反方向滑动？

在 $t_1$ 时接触点静止，所以 $v = -a\omega$：弹珠正在滚动。假如它继续以恒定速率滚动，便不需要任何力去推动它，所以摩擦力（一旦没有滑动，它只需要有多大便有多大）为零，而以恒定速度滚动与所有方程都相容。因此由 $t_1$ 起，弹珠作匀速滚动，速度为

$$V = v(t_1) = v_0 - \tfrac{2}{7}(v_0 + a\omega_0) = \frac{5v_0 - 2a\omega_0}{7}.$$

### 第 (i) 部分

要弹珠最终完全静止，最后的滚动速度必须为零：$V = 0$，所以 $5v_0 = 2a\omega_0$，即

$$\frac{v_0}{a\omega_0} = \frac{2}{5}.$$

检查一下这个情景是否合理：取这个比值时，$v = 0$ 发生于 $t = v_0/k$，而 $\omega = 0$ 发生于 $t = 2a\omega_0/(5k) = v_0/k$，是同一时刻，而这亦正是 $t_1$。所以线运动和角运动在弹珠仍在滑动时一同消失，此后弹珠保持静止（没有滑动，也没有摩擦力）。

### 第 (ii) 部分

现在弹珠最终必须以速率 $v_0/7$ 向后滚动，所以 $V = -v_0/7$：

$$\frac{5v_0 - 2a\omega_0}{7} = -\frac{v_0}{7} \quad\Longrightarrow\quad 6v_0 = 2a\omega_0 \quad\Longrightarrow\quad \frac{v_0}{a\omega_0} = \frac{1}{3}.$$

取这个比值时，$v$ 在 $t = v_0/k$ 时降至零，这早于 $t_1 = 2(v_0 + 3v_0)/(7k) = 8v_0/(7k)$。在那一瞬间弹珠暂时不动，但仍带有下旋 $\omega = \omega_0 - \frac{5v_0}{2a} = \frac{v_0}{2a} > 0$，所以底部仍在向前打滑，摩擦力继续把它往后拉。在 $t_1$ 时我们得到 $v = -v_0/7$ 及 $a\omega = v_0/7$，所以它以速率 $v_0/7$ 朝起点滚回去，符合要求。

> ? **想一想** 有没有更快捷的方法，毋须解出 $t_1$ 便可求得 $V$?

有。在桌面上沿运动路线取一个固定点 $O$，考虑对该点的角动量（angular momentum）。摩擦力沿桌面作用，所以它对 $O$ 没有力矩，而 $R$ 与 $mg$ 的力矩互相抵消。以向前滚动的方向为正，对 $O$ 的角动量等于 $mva$（球心位于高度 $a$ 处）减去自旋部分 $\frac{2}{5}ma^2\omega$，所以这个量守恒：

$$mv_0a - \tfrac{2}{5}ma^2\omega_0 = maV + \tfrac{2}{5}ma^2 \cdot \frac{V}{a} = \tfrac{7}{5}maV,$$

再次得出 $V = (5v_0 - 2a\omega_0)/7$。

## 验算

- 没有旋转（$\omega_0 = 0$）时，公式给出 $V = \frac{5}{7}v_0$，这是不带旋转、打滑射出的球的标准结果。
- 若 $v_0 = 0$ 而只有下旋，则 $V = -\frac{2}{7}a\omega_0 < 0$：弹珠向后运动，任何弹过带下旋弹珠的人都知道这一点。
- 以数值方法逐步推算这些方程（滑动速度一降至零便关掉摩擦力），当 $v_0 = 1$、$a\omega_0 = 3$ 时最终速度为 $-0.142857\ldots = -\frac{1}{7}$，而当 $v_0 = 2$、$a\omega_0 = 5$ 时则完全停下。

## 常见错误

- 在 $v = 0$ 时便停止计时。弹珠要到 $v + a\omega = 0$ 时才停止滑动；在 $v = 0$ 时它可能仍在旋转，正如第 (ii) 部分的情况。
- 弄错力矩（torque）的符号：底部向后的摩擦力会*减弱*下旋。
- 滚动开始后仍然保留摩擦力。一旦没有滑动，保持恒定速度并不需要摩擦力，所以摩擦力降为零。

## STEP 之外

这里的摩擦力采用的是库仑（Coulomb）的滑动摩擦模型：表面相对滑动时大小恒定，不滑动时则需要多少便有多少（不超过某个上限）。角动量的捷径是刚体动力学（rigid-body

dynamics）中一条普遍原则的例子：选取适当的参考点，使那个麻烦的未知力没有力矩。桌球（snooker）和英式撞球（billiards）球手打缩杆（screw-back）时，利用的正是这套物理。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）正是以这种思路处理刚体转动和带滑动的滚动。

详细解答

# STEP III 1997，第 11 题

力学 • 刚体转动

> **STEP III 1997，第 11 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s3-q11

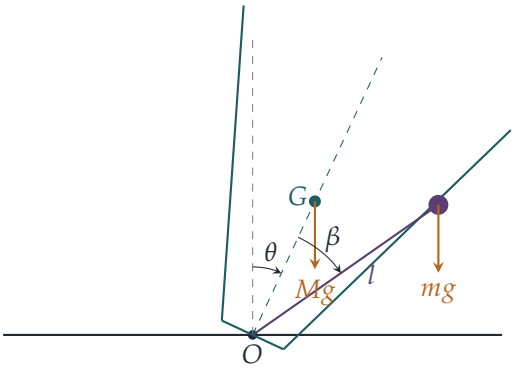
## 关键思路

只要钟与钟锤保持接触，两者便以相同的角速度（angular velocity）绕同一点 $O$ 转动，因此唯一能令它们分开的，就是角*加速度*（angular acceleration）的差异。不妨问：两者各自独立运动时，加速有多快？当自由摆动的钟的角加速度刚好等于自由钟锤的角加速度时，钟锤与钟便在那一瞬间分离。此时接触力恰好为零，而角速度从头到尾都没有出现。

## 解答

**模型**。钟是一个刚体（rigid body），可绕通过 $O$ 的水平轴自由转动，对该轴的转动惯量（moment of inertia）为 $Mk^2$。钟锤是一个质量为 $m$ 的质点（particle），装在长度为 $l$ 的轻质刚杆上，同样以 $O$ 为枢轴。枢轴光滑，且没有空气阻力。所有角度均由向上的铅垂线量起，以钟锤相对于钟轴所在的一侧为正方向。设 $\theta$ 为钟轴 $OG$ 的角度。两者接触期间，钟锤的杆与钟轴一直保持夹角 $\beta$，所以杆的角度是 $\theta + \beta$。

在钟内，钟锤的球靠在内壁上，所靠的正是它本身会倒下去的那一侧。因此内壁把球向后推（朝角度较小的方向），而根据牛顿第三定律（Newton's third law），球则把钟向前推。设 $\Gamma \geq 0$ 为这个接触力（包括任何摩擦力）对 $O$ 的力矩（moment）。它对钟锤的作用是 $-\Gamma$，对钟的作用是 $+\Gamma$。




**接触期间的运动方程**。分别对每个物体取对 $O$ 的力矩（枢轴处的反作用力和杆中的张力对 $O$ 均无力矩）：

$$\text{钟锤：}\quad ml^2\ddot{\theta} = mgl\sin(\theta+\beta) - \Gamma,$$
$$\text{钟：}\quad Mk^2\ddot{\theta} = Mgh\sin\theta + \Gamma.$$

重量 $mg$ 的作用点与 $O$ 相距 $l$,所在的直线与铅垂线成角 $\theta+\beta$,所以它的力矩是 $mgl\sin(\theta+\beta)$；作用于 $G$ 的 $Mg$ 亦同理。两个物体一起运动，所以有相同的 $\ddot{\theta}$。

把两式相加，便可消去 $\Gamma$：

$$\ddot{\theta} = \frac{g(ml\sin(\theta+\beta) + Mh\sin\theta)}{ml^2+Mk^2}.$$

代回钟锤的方程并化简，得

$$\Gamma = \frac{mMgl}{ml^2+Mk^2}\Big(k^2\sin(\theta+\beta) - hl\sin\theta\Big).$$

释放时 $\theta = 0$，而 $\Gamma = \dfrac{mMglk^2\sin\beta}{ml^2+Mk^2} > 0$，所以钟锤起初确实压在钟上，正如所料。

**想一想** 为甚么不论角速度是多少，分离都恰好在 $\Gamma$ 降至零的时候发生?

接触只能推、不能拉，所以必须有 $\Gamma \geq 0$。当 $\Gamma > 0$ 时，自由的钟锤会比钟加速得更快，所以它被压在内壁上，两者一起运动。由于两者以相同的角速度绕同一点 $O$ 转动，球与内壁之间的空隙只能因角加速度不同而出现。一旦 $\Gamma$ 要变为负值，钟本身的角加速度 $gh\sin\theta/k^2$ 便超过钟锤的 $g\sin(\theta+\beta)/l$，内壁于是摆离球。所以分离发生在第一个令 $\Gamma = 0$ 的 $\theta$。（在 $\theta = \pi$ 之前必定存在这样的 $\theta$，因为括号内的式子在 $\theta = 0$ 时是 $k^2\sin\beta > 0$，在 $\theta = \pi$ 时是 $-k^2\sin\beta < 0$。）

**分离时的角度**。设 $\alpha = \theta+\beta$，即分离时钟锤的杆与向上铅垂线所成的角。于是由 $\Gamma = 0$ 得

$$k^2\sin\alpha = hl\sin(\alpha-\beta) = hl(\sin\alpha\cos\beta - \cos\alpha\sin\beta).$$

这里 $0 < \alpha < \pi$，所以 $\sin\alpha \neq 0$，而且 $\sin\beta \neq 0$。两边除以 $hl\sin\alpha\sin\beta$：

$$\frac{k^2}{hl\sin\beta} = \cot\beta - \cot\alpha,$$

即

$$\cot\alpha = \cot\beta - \frac{k^2}{hl}\operatorname{cosec}\beta,$$

这正是所求的结果。留意两个质量都没有出现。

## 验算

- 短钟锤。当 $l \to 0$ 时，修正项趋向无穷大，所以 $\cot\alpha \to -\infty$，$\alpha \to \pi$。这很合理：短的摆有很大的自由角加速度 $g\sin\alpha/l$，所以装在极短杆上的钟锤会一直推着钟，直至 $\sin\alpha$ 接近零，即直至两者几乎倒转为止。
- 等价形式。这个条件其实就是 $\dfrac{g\sin\alpha}{l} = \dfrac{gh\sin(\alpha-\beta)}{k^2}$：自由的钟锤与自由的钟有相等的角加速度。
- 数值验证。取 $k=1$、$h=0.6$、$l=0.8$ 及 $\beta=30°$，公式给出 $\cot\alpha = \sqrt{3} - 25/6 \approx -2.43$，所以 $\alpha \approx 157.7°$。由释放一刻起对合并运动逐步作数值积分，到 $\Gamma$ 首次降至零时停止，所得的角度与此相同，准确至小数点后多个位，而且 $M/m = 5$ 与 $M/m = 500$ 的情况皆然。

## 常见错误

- 用能量守恒（conservation of energy）求出 $\dot\theta$，然后试图由圆周运动求“法向反作用力”。角速度是两个物体共有的，对它们是否分离毫无影响。
- 由钟轴而不是由向上的铅垂线量度 $\alpha$。在整段接触期间，杆与钟轴的夹角都是 $\beta$，所以 $\alpha = \theta + \beta$。
- 忘记 $Mk^2$ 是对 $O$ 而不是对 $G$ 的转动惯量。这里无须用平行轴定理（parallel-axis theorem）作修正。

## STEP 之外

要找出约束（constraint）何时失效，有一个通用的原则，而本题正是一个简洁的例子：先解出受约束的运动，再计算约束力，然后找出它在何处变号。拉格朗日力学（Lagrangian mechanics）利用拉格朗日乘数（Lagrange multipliers）把这个做法系统化，而这些乘数原来就是约束力。同样的想法可以处理质点离开球面这个经典问题。Kleppner 与 Kolenkow

合著的 *An Introduction to Mechanics*（力学入门）涵盖刚体绕固定轴的转动，是极佳的预备读物。

详细解答

# STEP III 1997，第 12 题

概率与统计 • 赠券收集问题

> **STEP III 1997，第 12 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s3-q12

## 关键思路

第 (i) 部分是几何分布（geometric distribution），而它的概率生成函数（probability generating function）只需求导，便可给出平均值和方差：$\mathrm{E}(K) = \mathrm{G}'(1)$ 及 $\mathrm{E}\big(K(K-1)\big) = \mathrm{G}''(1)$。在第 (ii) 部分，窍门是把集齐全套的漫长等待拆成 52 段短等待，每段对应一张新牌。每段短等待都是一个“掷硬币直至首次出现正面”的实验，所以第 (i) 部分已经把工夫全做好了。

## 解答

### 第 (i) 部分

首次正面出现在第 $k$ 次投掷，当且仅当头 $k-1$ 次投掷都是反面，而第 $k$ 次是正面。各次投掷互相独立（independent），所以

$$\mathrm{P}(K = k) = q^{k-1}p, \qquad k = 1, 2, 3, \ldots$$

因此

$$\mathrm{G}(s) = \sum_{k=1}^{\infty} pq^{k-1}s^k = ps\sum_{j=0}^{\infty}(qs)^j = \frac{ps}{1-qs},$$

这里用了等比级数（geometric series）求和，该级数在 $|qs| < 1$ 时收敛（特别是当 $p > 0$ 时，对接近 1 的 $s$ 收敛）。这就是所要求的结果。

求导得

$$\mathrm{G}'(s) = \frac{p(1-qs)+psq}{(1-qs)^2} = \frac{p}{(1-qs)^2},$$
$$\mathrm{G}''(s) = \frac{2pq}{(1-qs)^3}.$$

由于 $\mathrm{G}'(s) = \sum k\,\mathrm{P}(K=k)s^{k-1}$ 及 $\mathrm{G}''(s) = \sum k(k-1)\mathrm{P}(K=k)s^{k-2}$，代入 $s=1$ 并利用 $1-q=p$，可得

$$\mathrm{E}(K) = \mathrm{G}'(1) = \frac{p}{p^2} = \frac{1}{p}, \qquad \mathrm{E}\big(K(K-1)\big) = \mathrm{G}''(1) = \frac{2pq}{p^3} = \frac{2q}{p^2}.$$

所以

$$\begin{aligned}\mathrm{Var}(K) &= \mathrm{E}\big(K(K-1)\big) + \mathrm{E}(K) - \big(\mathrm{E}(K)\big)^2 \\ &= \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{q}{p^2}.\end{aligned}$$

故此 $\mathrm{E}(K) = \dfrac{1}{p}$ 及 $\mathrm{Var}(K) = \dfrac{q}{p^2}$。

**第 (ii) 部分**

设 $N_i$ 为已经见过恰好 $i-1$ 张不同的牌之后所抽的次数，计至并包括抽出第 $i$ 张不同的牌的那一次。于是 $N_1 = 1$（第一张牌必定是新的），而

$$N = N_1 + N_2 + \cdots + N_{52}.$$

? **想一想** 为甚么每个 $N_i$ 都是像 $K$ 那样的“首次正面”变量?

一旦见过 $i-1$ 张不同的牌，52 张牌之中仍有 $53-i$ 张未见过。由于抽牌后放回（with replacement），每次抽牌都与过去的结果独立，而且抽出未见过的牌的概率都相同，即

$$p_i = \frac{53-i}{52}.$$

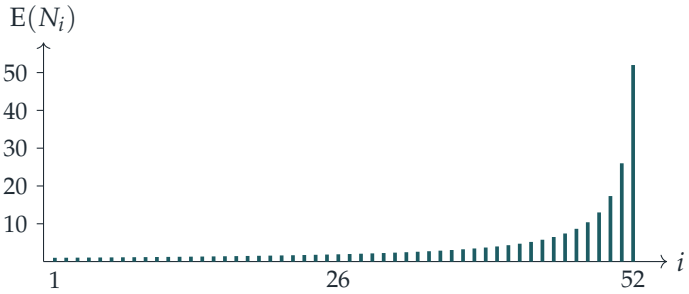
把这样的一次抽牌称为“正面”。那么 $N_i$ 就是计至并包括首次正面的抽牌次数，正正是第 (i) 部分中的变量 $K$，其中 $p = p_i$。因此

$$\mathrm{E}(N_i) = \frac{1}{p_i} = \frac{52}{53-i}.$$

期望值（expectation）是线性的（这里毋须各个 $N_i$ 互相独立），所以

$$\begin{aligned}\mathrm{E}(N) &= \sum_{i=1}^{52} \frac{52}{53-i} = 52\left(\frac{1}{52} + \frac{1}{51} + \cdots + \frac{1}{2} + \frac{1}{1}\right) \\ &= 52\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{52}\right).\end{aligned}$$

早段的等待很短，最后几段才是主导：单是最后一张牌，平均就要抽 52 次。




要作估算，可用 $1+\frac{1}{2}+\cdots+\frac{1}{n}\approx 0.5+\ln n$，取 $n=52$。至于 $\ln 52$：$2.7^2=7.29$ 而 $7.29^2\approx 53.1$，所以 $e^4\approx 53\approx 52$，得出 $\ln 52\approx 4$。因此

$$\mathrm{E}(N)\approx 52\,(0.5+4)=52\times 4.5=234.$$

所以平均而言，我们大约要抽 234 次，即约为整副牌张数的四倍半，才能见齐每一张牌。

## 验算

- 在 (i) 中，$\mathrm{G}(1)=p/(1-q)=1$，任何概率生成函数都必须如此。若硬币是公平的 $(p=\frac{1}{2})$，便得 $\mathrm{E}(K)=2$ 及 $\mathrm{Var}(K)=2$；而当 $p\to 1$ 时，方差 $q/p^2\to 0$，这正是第一次投掷几乎必然是正面时应有的结果。
- 准确值是 $52H_{52}\approx 235.98$，而模拟 $10^5$ 次完整收集所得的平均值是 235.7。我们的估计值 234 相当接近；那少许差额来自近似式 $0.5+\ln n$（较精确的是 $\ln n+0.577$），以及把 $\ln 52\approx 3.95$ 上舍入为 4。

## 常见错误

- 把 $N_i$ 的成功概率弄颠倒：抽到新牌的机会是 $(53-i)/52$，它随着收集到的牌增多而减小。
- 取 $\mathrm{Var}(K)=\mathrm{G}''(1)-\mathrm{G}'(1)^2$，忘记了 $+\mathrm{G}'(1)$ 这一项；$\mathrm{G}''(1)$ 是 $\mathrm{E}(K(K-1))$，而不是 $\mathrm{E}(K^2)$。
- 以为各个 $N_i$ 必须互相独立，才可以把期望值相加。无论独立与否，期望值的线性性质都成立（不过在这里它们其实是独立的）。

## STEP 之外

这就是*优惠券收集问题*（coupon collector's problem）：若有 $n$ 种出现机会均等的类型，集齐全部所需的期望时间是 $nH_n\approx n\ln n$。大学本科课程会更进一步，证明这个时间高

度集中在 $n \ln n$ 附近，波动的量级为 $n$；而概率生成函数亦会成为处理这类独立变量之和的标准工具。Grimmett 与 Welsh 的 *Probability: An Introduction* 仔细讲解生成函数和几何分布；Blitzstein 与 Hwang 的 *Introduction to Probability* 则讨论优惠券收集问题及其多种变化。

详细解答

# STEP III 1997，第 13 题

概率与统计 · 正态分布的矩生成函数

> **STEP III 1997，第 13 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/97-s3-q13

## 关键思路

一切都建基于一个积分：标准正态分布（standard normal distribution）的矩母函数（moment generating function，简称 mgf），只要把指数部分配方（completing the square）便可求得。之后，独立性（independence）把和的矩母函数化为乘积，而矩母函数的唯一性则让我们可以直接读出分布。最后一部分其实是同一个计算换了个面貌：$Z = \mathrm{e}^{\mu+\sigma X}$ 的幂，不过是 $X$ 的某个倍数的指数函数而已。

## 解答

### 第 (i) 部分

设 $M(\theta) = \mathrm{E}(\mathrm{e}^{\theta X})$。则

$$M(\theta) = \int_{-\infty}^{\infty} \mathrm{e}^{\theta x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(x^2-2\theta x)}\,\mathrm{d}x.$$

配方得 $x^2 - 2\theta x = (x-\theta)^2 - \theta^2$，所以

$$M(\theta) = \mathrm{e}^{\frac{1}{2}\theta^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(x-\theta)^2}\,\mathrm{d}x = \mathrm{e}^{\frac{1}{2}\theta^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}u^2}\,\mathrm{d}u,$$

其中用了代换 $u = x - \theta$。最后一个积分是标准正态概率密度函数（density）在整条实数线上的积分，其值为 1。因此，对任意实数 $\theta$，

$$\boxed{\mathrm{E}(\mathrm{e}^{\theta X}) = \mathrm{e}^{\frac{1}{2}\theta^2}.}$$

### 第 (ii) 部分

由于 $X$ 与 $Y$ 独立，$\mathrm{e}^{\theta aX}$ 与 $\mathrm{e}^{\theta bY}$ 亦独立，而独立随机变量（random variable）之积的期望值（expectation），等于各自期望值之积。把第 (i) 部分的 $\theta$ 分别换成 $a\theta$ 和 $b\theta$（并利用 $Y$ 与 $X$ 分布相同这一事实），可得

$$\mathrm{E}\big(\mathrm{e}^{\theta(aX+bY)}\big) = \mathrm{E}(\mathrm{e}^{a\theta X})\,\mathrm{E}(\mathrm{e}^{b\theta Y}) = \mathrm{e}^{\frac{1}{2}a^2\theta^2}\,\mathrm{e}^{\frac{1}{2}b^2\theta^2} = \mathrm{e}^{\frac{1}{2}(a^2+b^2)\theta^2}.$$

**?** **想一想** 为甚么矩母函数相同，就足以断定分布相同?

若矩母函数对 0 附近的所有 $\theta$ 都存在，它便唯一地决定了分布。所以，$aX+bY$ 与 $X$ 分布相同，当且仅当其矩母函数对所有 $\theta$ 都等于 $\mathrm{e}^{\frac{1}{2}\theta^2}$，即

$$\mathrm{e}^{\frac{1}{2}(a^2+b^2)\theta^2} = \mathrm{e}^{\frac{1}{2}\theta^2} \quad \text{对所有}\theta.$$

取对数，并任选一个 $\theta \neq 0$，可知上式成立当且仅当

$$\boxed{a^2+b^2=1.}$$

（“仅当”这个方向不用唯一性也显而易见：$\mathrm{Var}(aX+bY) = a^2+b^2$ 必须等于 $\mathrm{Var}(X) = 1$。）一般而言，$aX+bY$ 服从平均值（mean）为 0、方差（variance）为 $a^2+b^2$ 的正态分布；例如 $(X+Y)/\sqrt{2}$ 仍然是标准正态。

### 第 (iii) 部分

对任意实数 $\theta$，

$$Z^\theta = \big(\mathrm{e}^{\mu+\sigma X}\big)^\theta = \mathrm{e}^{\mu\theta}\,\mathrm{e}^{(\sigma\theta)X},$$

而 $\mathrm{e}^{\mu\theta}$ 是常数，所以把第 (i) 部分的 $\theta$ 换成 $\sigma\theta$，便得

$$\mathrm{E}(Z^\theta) = \mathrm{e}^{\mu\theta}\,\mathrm{E}\big(\mathrm{e}^{(\sigma\theta)X}\big) = \mathrm{e}^{\mu\theta}\,\mathrm{e}^{\frac{1}{2}\sigma^2\theta^2} = \mathrm{e}^{\mu\theta+\frac{1}{2}\sigma^2\theta^2},$$

这正是所要证明的结果。

取 $\theta = 1$ 便得平均值：

$$\boxed{\mathrm{E}(Z) = \mathrm{e}^{\mu+\frac{1}{2}\sigma^2}.}$$

取 $\theta = 2$ 得 $\mathrm{E}(Z^2) = \mathrm{e}^{2\mu+2\sigma^2}$，所以

$$\begin{aligned}\mathrm{Var}(Z) &= \mathrm{E}(Z^2) - \big(\mathrm{E}(Z)\big)^2 = \mathrm{e}^{2\mu+2\sigma^2} - \mathrm{e}^{2\mu+\sigma^2} \\ &= \mathrm{e}^{2\mu+\sigma^2}\big(\mathrm{e}^{\sigma^2}-1\big).\end{aligned}$$

## 验算

- 对 $M(\theta) = \mathrm{e}^{\frac{1}{2}\theta^2}$ 求导，得 $M'(0) = 0$ 及 $M''(0) = 1$：平均值为 0，方差为 1，与应有的结果一致。
- 若 $\sigma = 0$，则 $Z = \mathrm{e}^{\mu}$ 是常数；我们的公式给出平均值 $\mathrm{e}^{\mu}$ 和方差 0。
- 数值上，取 $\mu = 0.3$ 和 $\sigma = 0.5$，对正态密度函数积分，得 $\mathrm{E}(Z) \approx 1.5296$ 及 $\mathrm{Var}(Z) \approx 0.6645$，与 $\mathrm{e}^{0.425}$ 和 $\mathrm{e}^{0.85}(\mathrm{e}^{0.25} - 1)$ 吻合。对 $0.6X + 0.8Y$ 进行模拟，所得样本方差为 1.000。

## 常见错误

- 以“代入 $X$ 的平均值”的方式写出 $\mathrm{E}(Z) = \mathrm{e}^{\mu}$。一般而言 $\mathrm{E}(g(X)) \neq g(\mathrm{E}(X))$，而这里指数函数的凸性（convexity）令平均值增大至 $\mathrm{e}^{\frac{1}{2}\sigma^2}$ 倍。
- 把条件写成 $a + b = 1$：相加的是方差，而不是系数。

## STEP 之外

随机变量 $Z$ 服从*对数正态*（log-normal）分布；凡是由许多细小的随机因子连乘而增长的量，都以它为标准模型，例如布莱克–舒尔斯（Black–Scholes）模型中的股价。第 (ii) 部分让我们窥见一个引人注目的事实：在方差有限的分布之中，基本上只有正态分布在适当缩放后求和，仍然留在同一个分布族内，而这与中心极限定理（central limit theorem）息息相关。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）正好以这个程度讲解矩母函数、正态分布和中心极限定理。

详细解答

# STEP III 1997，第 14 题

概率与统计 · 随机变量的乘积

**STEP III 1997，第 14 题**

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

step.maths.org/questions/97-s3-q14

## 关键思路

周界是长度与阔度的线性函数，因此它的平均值和方差可直接由独立变量之和的法则得出。面积则是一个乘积，而对于独立变量，乘积的期望值等于期望值的乘积；把这一点应用于 $LB$ 和 $L^2B^2$，便得到平均值和方差。要证明两者并非独立，最干净利落的做法是找出一个非零的协方差（covariance），因为独立变量的协方差必定为零。

## 解答

设 $L$ 和 $B$ 分别为随机抽出的一块板的长度和阔度，则 $L$ 与 $B$ 独立（independent），且

$$\mathrm{E}(L)=\mu_1,\quad \mathrm{Var}(L)=\sigma_1^2,\qquad \mathrm{E}(B)=\mu_2,\quad \mathrm{Var}(B)=\sigma_2^2.$$

设 $P=2(L+B)$ 为周界，$A=LB$ 为面积。我们会多次用到 $\mathrm{E}(X^2)=\mathrm{Var}(X)+[\mathrm{E}(X)]^2$，从而有 $\mathrm{E}(L^2)=\mu_1^2+\sigma_1^2$ 及 $\mathrm{E}(B^2)=\mu_2^2+\sigma_2^2$。

### 周界

由期望值的线性性质（linearity of expectation），

$$\mathrm{E}(P)=2(\mu_1+\mu_2).$$

由于 $L$ 与 $B$ 独立，两者之和的方差（variance）等于各自方差之和，所以

$$\mathrm{Var}(P)=2^2\big(\mathrm{Var}(L)+\mathrm{Var}(B)\big)=4(\sigma_1^2+\sigma_2^2),\qquad \text{标准差}(P)=2\sqrt{\sigma_1^2+\sigma_2^2}.$$

### 面积

因为 $L$ 与 $B$ 独立，$\mathrm{E}(LB) = \mathrm{E}(L)\,\mathrm{E}(B)$，所以

$$\mathrm{E}(A) = \mu_1\mu_2.$$

要求方差，我们需要 $\mathrm{E}(A^2) = \mathrm{E}(L^2B^2)$。

> **想一想** 我们还可以把 $\mathrm{E}(L^2B^2)$ 拆成乘积吗?

可以：独立变量的函数仍然互相独立，所以 $L^2$ 与 $B^2$ 独立，而 $\mathrm{E}(L^2B^2) = \mathrm{E}(L^2)\,\mathrm{E}(B^2)$。因此

$$\begin{aligned}\mathrm{Var}(A) &= \mathrm{E}(L^2)\,\mathrm{E}(B^2) - \mu_1^2\mu_2^2\\ &= (\mu_1^2+\sigma_1^2)(\mu_2^2+\sigma_2^2) - \mu_1^2\mu_2^2\\ &= \sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2,\end{aligned}$$

以及

$$\text{标准差}(A) = \sqrt{\sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2}.$$

留意乘积的方差并非只是 $\sigma_1^2\sigma_2^2$：每一个维度上的离散程度，都会被另一个维度的典型大小放大。

### 周界与面积并非独立

假如 $P$ 与 $A$ 独立，便会有 $\mathrm{E}(PA) = \mathrm{E}(P)\,\mathrm{E}(A)$，即 $\mathrm{Cov}(P, A) = 0$。我们来计算协方差。首先，

$$\mathrm{E}(PA) = 2\,\mathrm{E}(L^2B) + 2\,\mathrm{E}(LB^2).$$

再次利用独立性（这次是 $L^2$ 与 $B$ 独立，以及 $L$ 与 $B^2$ 独立），

$$\mathrm{E}(L^2B) = (\mu_1^2+\sigma_1^2)\mu_2, \qquad \mathrm{E}(LB^2) = \mu_1(\mu_2^2+\sigma_2^2).$$

故此

$$\begin{aligned}\mathrm{Cov}(P, A) &= \mathrm{E}(PA) - \mathrm{E}(P)\,\mathrm{E}(A)\\ &= 2\big[(\mu_1^2+\sigma_1^2)\mu_2 + \mu_1(\mu_2^2+\sigma_2^2)\big] - 2(\mu_1+\mu_2)\mu_1\mu_2\\ &= 2\big(\mu_2\sigma_1^2 + \mu_1\sigma_2^2\big).\end{aligned}$$

长度和阔度都是正数，所以 $\mu_1 > 0$ 及 $\mu_2 > 0$；题目已知标准差（standard deviation）不为零，所以 $\sigma_1^2 > 0$ 及 $\sigma_2^2 > 0$。于是 $\mathrm{Cov}(P, A) > 0$。

由于独立变量的协方差为零，$P$ 与 $A$ 不可能独立。这是合理的：一块板若在任何一个方向上比平均大，它的周界和面积往往都会较大。

**想一想** 假如算出来的协方差是零，这能否说明 $P$ 与 $A$ 独立?

不能。这个推论只是单向的：独立必然导致协方差为零，但协方差为零并不保证独立。所以非零的协方差能证明两者相依，这正是我们现在的情况；若协方差为零，则甚么结论也得不到。

## 验算

- 取 $L$ 与 $B$ 独立，各自以概率 $\frac{1}{2}$ 取值 1 或 3。则 $\mu_i = 2$, $\sigma_i^2 = 1$，而公式给出 $\mathrm{Cov}(P, A) = 2(2+2) = 8$，与直接列举四块等可能的板所得的结果一致。两者相依亦可直接看出：$\mathrm{P}(P = 4) = \mathrm{P}(A = 1) = \frac{1}{4}$，但 $\mathrm{P}(P = 4 \text{ 且 } A = 1) = \frac{1}{4} \neq \frac{1}{16}$，因为两个事件都表示 $L = B = 1$。
- 以 $\mu_1 = 5$、$\mu_2 = 3$、$\sigma_1 = 0.7$、$\sigma_2 = 0.4$ 进行模拟，得到 标准差$(A) \approx 2.919$，而公式给出 2.913；又得到 $\mathrm{Cov}(P, A) \approx 4.56$，而公式给出 4.54。
- 若 $\sigma_2 \to 0$（阔度固定为 $\mu_2$），则 标准差$(A) \to \mu_2\sigma_1$，这正是 $A = \mu_2 L$ 时应有的结果。

## 常见错误

- 写成 $\mathrm{Var}(LB) = \sigma_1^2\sigma_2^2$，或写成 $\mathrm{Var}(LB) = \mathrm{Var}(L)\mathrm{Var}(B) + \ldots$ 而漏掉了某些项；务必经由 $\mathrm{E}(L^2B^2)$ 来计算。
- 计算 $\mathrm{Var}(2(L+B))$ 时忘记把因子 2 平方。
- 以“因为两者都取决于 $L$ 和 $B$”为理由，断言 $P$ 与 $A$ 相依。这只是直觉，不是证明；真正能下定论的，是协方差的计算（或一个概率不相等的具体例子）。

## STEP 之外

不相关（uncorrelated）与独立的随机变量（random variable）之间的分别，是大学一年级概率课程的核心：独立意味着联合分布（joint distribution）可以分解成乘积，这比单单一个矩（moment）的条件强得多。$\mathrm{Var}(LB)$ 的计算，也是实验科学中广泛使用的“误差传递”（propagation of errors）公式的起点。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门教科书）对独立性、协方差及随机变量的乘积有严谨细致的讨论。

# Beyond STEP

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

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

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

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

**1997 年详细解答.** 全书各章及
所有详细解答均可免费阅读、下载及分享，网址为
**beyondstep.keithmcnulty.org**

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

