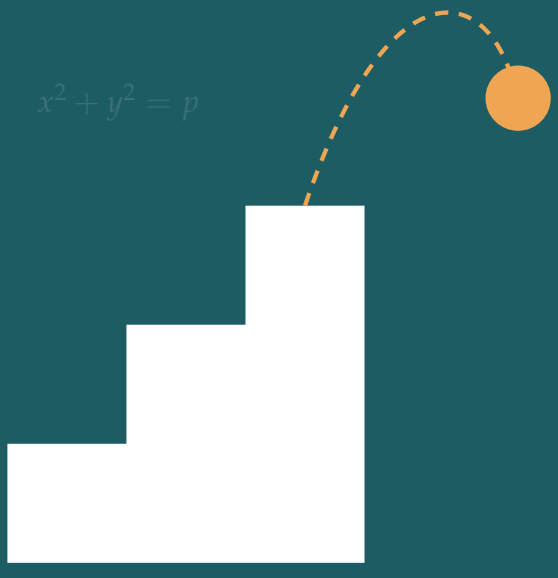
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2016 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

2016 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2016 年详细解答

2016 年 STEP 试卷（STEP I, II 及 III）全部 37 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 多项式恒等式与因式分解 纯数学 . . . 3

第 4 题 曲线的曲率 纯数学 . . . 6

第 5 题 相切圆 纯数学 . . . 9

第 6 题 向量几何 纯数学 . . . 12

第 7 题 模 4 的因式分解 纯数学 . . . 16

第 8 题 生成函数 纯数学 . . . 20

第 9 题 极限摩擦力平衡 力学 . . . 24

第 10 题 碰撞与恢复系数 力学 . . . 28

第 11 题 抛体的最大射程 力学 . . . 32

第 12 题 项分布的对称性 概率与统计 . . . 36

第 13 题 整数分布的顺序统计量 概率与统计 . . . 39

STEP II

第 1 题 参数曲线与切线 纯数学 . . . 42

第 2 题 因式定理 纯数学 . . . 46

第 3 题 多项式的根 纯数学 . . . 50

第 4 题 有理函数的值域 纯数学 . . . 53

第 5 题 二项式系数恒等式 纯数学 . . . 57

第 6 题 切比雪夫多项式方程 纯数学 . . . 61

第 7 题 利用对称性求定积分 纯数学 . . . 65

第 8 题 以积分近似求和 纯数学 . . . 69

第 9 题 动量与能量损失 力学 . . . 73

第 10 题 壁擦力与平衡 力学 . . . 76

第 11 题 抛体的碰撞 力学 . . . 79

第 12 题 容斥原理与错排 概率与统计 83
第 13 题 泊松近似 概率与统计 87

STEP III

第 1 题 归约公式 纯数学 91
第 2 题 抛物线的法线 纯数学 95
第 3 题 积分与多项式 纯数学 99
第 4 题 裂项相消级数 纯数学 102
第 5 题 质数的乘积 纯数学 105
第 6 题 双曲函数 纯数学 108
第 7 题 单位根 纯数学 112
第 8 题 函数方程 纯数学 116
第 9 题 弹簧与简谐运动 力学 120
第 10 题 斜面上的圆周运动 力学 124
第 11 题 功率与阻力 力学 128
第 12 题 切比雪夫不等式 概率与统计 132
第 13 题 和的峰度 概率与统计 135

详细解答

STEP I 2016，第 1 题

纯数学 · 多项式恒等式与因式分解

> **STEP I 2016，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/16-s1-q1

关键思路

商（quotient）$q_n(x)$ 其实是一个多项式（polynomial）：由于 $2n+1$ 是奇数，$x+1$ 整除 $x^{2n+1}+1$，而商是一个正负交替的等比数列之和。因此第 (i) 部分只需展开 p_n 并比较系数（coefficients）。第 (ii) 部分的关键在于 p_n 的形式：当 $2n+1$ 乘以 x 是完全平方数（perfect square）时，p_n 是两个平方数之差（difference of two squares），因此可以因式分解（factorise）。

解答

第 (i) 部分

当 $x \neq -1$ 时，把首项为 1、公比为 $-x$ 的等比级数（geometric series）求和，得

$$1 - x + x^2 - \cdots + x^{2n} = \sum_{k=0}^{2n} (-x)^k = \frac{1-(-x)^{2n+1}}{1+x} = \frac{1+x^{2n+1}}{1+x},$$

这里用到了 $(-x)^{2n+1} = -x^{2n+1}$，因为 $2n+1$ 是奇数。所以

$$q_n(x) = x^{2n} - x^{2n-1} + x^{2n-2} - \cdots - x + 1,$$

它的 $2n+1$ 个系数全部是 ± 1，而且正负交替。接下来我们逐一展开 p_n。

$n = 1$.

$$p_1(x) = (x+1)^2 - 3x = x^2 + 2x + 1 - 3x = x^2 - x + 1 = q_1(x).$$

$n = 2$.

$$\begin{aligned} p_2(x) &= (x+1)^4 - 5x(x^2+x+1) \\ &= (x^4+4x^3+6x^2+4x+1) - (5x^3+5x^2+5x) \\ &= x^4 - x^3 + x^2 - x + 1 = q_2(x). \end{aligned}$$

$n = 3$. 首先 $(x^2+x+1)^2 = x^4+2x^3+3x^2+2x+1$，所以

$$\begin{aligned} p_3(x) &= (x+1)^6 - 7x(x^2+x+1)^2 \\ &= (x^6+6x^5+15x^4+20x^3+15x^2+6x+1) - (7x^5+14x^4+21x^3+14x^2+7x) \\ &= x^6 - x^5 + x^4 - x^3 + x^2 - x + 1 = q_3(x). \end{aligned}$$

因此当 $n = 1, 2, 3$ 时，对 $x \neq -1$ 有 $p_n(x) \equiv q_n(x)$，正如所求。

$n = 4$. 只需一个 x 值便足以推翻一条恒等式（identity）。当 $x = 1$ 时，

$$p_4(1) = 2^8 - 9 \cdot 1 \cdot 3^3 = 256 - 243 = 13, \qquad q_4(1) = \frac{1+1}{1+1} = 1,$$

所以 $p_4 \neq q_4$。（完全展开可得 $p_4(x) = x^8 - x^7 + x^6 + 2x^5 + 7x^4 + 2x^3 + x^2 - x + 1$，从 x^5 项起便出错。）

第 (ii) 部分

> **想一想** 应该用哪个 n 和哪个 x？p_n 又怎样才可能因式分解？

把 $\dfrac{x^{2n+1}+1}{x+1}$ 与题目给出的数对应起来，便确定了 x 和 n。这时 p_n 的形式就派上用场：$(x+1)^{2n} = ((x+1)^n)^2$ 必定是平方数，而当 n 是奇数时，$(x^2+x+1)^{n-1}$ 也是平方数。因此，若 $(2n+1)x$ 同样是完全平方数，$p_n(x)$ 便是两个平方数之差。

(a) 取 $n = 1$ 及 $x = 300$，于是 $2n+1 = 3$，且 $3x = 900 = 30^2$。由第 (i) 部分，

$$\begin{aligned} \frac{300^3+1}{301} = q_1(300) = p_1(300) &= 301^2 - 3 \cdot 300 \\ &= 301^2 - 30^2 \\ &= (301-30)(301+30) = 271 \times 331. \end{aligned}$$

两个因数都不是 1。（直接计算，$300^2 - 300 + 1 = 89\,701 = 271 \times 331$。）

(b) 由于 $7^{49} = (7^7)^7$，取 $n = 3$（于是 $2n+1 = 7$）及 $x = 7^7$。那么 $(2n+1)x = 7 \cdot 7^7 =$

$7^8 = (7^4)^2$ 是完全平方数，而且 $n-1=2$ 是偶数。由第 (i) 部分，

$$\frac{7^{49}+1}{7^7+1} = q_3(7^7) = p_3(7^7) = (7^7+1)^6 - 7^8\,(7^{14}+7^7+1)^2$$
$$= [(7^7+1)^3]^2 - [7^4(7^{14}+7^7+1)]^2,$$

所以

$$\frac{7^{49}+1}{7^7+1} = \Big[(7^7+1)^3 - 7^4(7^{14}+7^7+1)\Big]\Big[(7^7+1)^3 + 7^4(7^{14}+7^7+1)\Big].$$

第二个因数显然大于 1。至于第一个因数，$(7^7+1)^3 > 7^{21}$，而

$$7^4(7^{14}+7^7+1) = 7^{18}+7^{11}+7^4 < 7^{19},$$

所以第一个因数大于 $7^{21}-7^{19} = 48\cdot 7^{19} > 1$。两个因数都不是 1，于是得到所求的因式分解。（把括号乘开，两个因数亦可写成 $7^{21}+3\cdot 7^{14}+3\cdot 7^7+1 \mp (7^{18}+7^{11}+7^4)$。）

验算

- 在第 (i) 部分代入 $x=1$: $p_1(1) = 4-3 = 1, p_2(1) = 16-15 = 1, p_3(1) = 64-63 = 1$，全部等于 $q_n(1) = 1$，而 $p_4(1) = 13$。
- 在 (a) 中，$271\times 331 = 89\,701 = 300^2 - 300 + 1$。
- 在 (b) 中，用电脑检验准确的整数值，证实两个因数之积等于 $(7^{49}+1)/(7^7+1)$；两个因数分别约为 5.57×10^{17} 和 5.60×10^{17}。

常见错误

- 忘记说明为何两个因数都不是 1；在 (b) 中这需要简短地估计大小。
- 在 (b) 中用 $n=24$ 或 $x=7$: 指数 49 必须写成 7×7，从而得到 $x=7^7$ 及 $2n+1=7$。

STEP 之外

这里的技巧是奥里费伊因式分解（Aurifeuillean factorisation）的一个特例：例如 $x^4+4y^4 = (x^2+2y^2-2xy)(x^2+2y^2+2xy)$ 这类恒等式，能把形如 $a^k\pm 1$ 的数分解得比明显的代数因式更细。它们源自分圆多项式（cyclotomic polynomials），这是大学一年级代数和数论（number theory）课程会出现的内容，而且曾被用于分解形如 $2^k\pm 1$ 的大数。Davenport 的 *The Higher Arithmetic*（《高等算术》）是进一步了解因式分解和质数（primes）的好读物。

详细解答

STEP I 2016，第 4 题

纯数学 · 曲线的曲率

> **STEP I 2016，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/16-s1-q4

关键思路

κ 的表达式看似一条艰深的二阶微分方程（second-order differential equation），但第 (i) 部分正是要说明它其实是某个函数的导数，可直接积分（exact derivative）。设 $p = \mathrm{f}'(x)$，曲率（curvature）就是 $p/\sqrt{1+p^2}$ 对 x 的导数，因此常数曲率一眼便可积分一次。第二次积分则属例行工夫，而答案原来是一个熟悉的形状。

解答

第 (i) 部分

对 $z(1+z^2)^{-1/2}$ 使用乘积法则（product rule），

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{z}{(1+z^2)^{1/2}}\right) &= (1+z^2)^{-1/2} - z\cdot z(1+z^2)^{-3/2} \\ &= \frac{(1+z^2)-z^2}{(1+z^2)^{3/2}} = \frac{1}{(1+z^2)^{3/2}}.\end{aligned}$$

第 (ii) 部分

设 κ 为非零常数，并记 $p = \mathrm{f}'(x)$，于是 $\mathrm{f}''(x) = \dfrac{\mathrm{d}p}{\mathrm{d}x}$。由第 (i) 部分及链式法则（chain rule），

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{p}{\sqrt{1+p^2}}\right) = \frac{1}{(1+p^2)^{3/2}}\frac{\mathrm{d}p}{\mathrm{d}x} = \frac{\mathrm{f}''(x)}{\left(1+(\mathrm{f}'(x))^2\right)^{3/2}} = \kappa.$$

对 x 积分，存在一个常数，我们把它写成 $-\kappa a$（因 $\kappa \neq 0$，这样做是可行的），使得

$$\frac{p}{\sqrt{1+p^2}} = \kappa(x-a). \qquad (*)$$

想一想 如何从 $(*)$ 解出 p，而又不会弄错正负号?

只要记住平方会丢失甚么信息，平方便是安全的。由于 $\sqrt{1+p^2} > 0$，$(*)$ 表示 p 与 $\kappa(x-a)$ 同号；又因 $|p| < \sqrt{1+p^2}$，它亦表示 $|\kappa(x-a)| < 1$。两边平方，得 $p^2 = \kappa^2(x-a)^2(1+p^2)$，所以

$$p^2 = \frac{\kappa^2(x-a)^2}{1-\kappa^2(x-a)^2}, \qquad \text{并按上述选取正负号,} \qquad p = \frac{\kappa(x-a)}{\sqrt{1-\kappa^2(x-a)^2}}.$$

反过来，这个 p 确实满足 $(*)$，因为 $1+p^2 = 1/\left(1-\kappa^2(x-a)^2\right)$。

现在再积分一次。由于 $\dfrac{\mathrm{d}}{\mathrm{d}x}\sqrt{1-\kappa^2(x-a)^2} = \dfrac{-\kappa^2(x-a)}{\sqrt{1-\kappa^2(x-a)^2}}$，对某常数 b，我们得到

$$y = b - \frac{1}{\kappa}\sqrt{1-\kappa^2(x-a)^2}, \qquad |x-a| < \frac{1}{|\kappa|}.$$

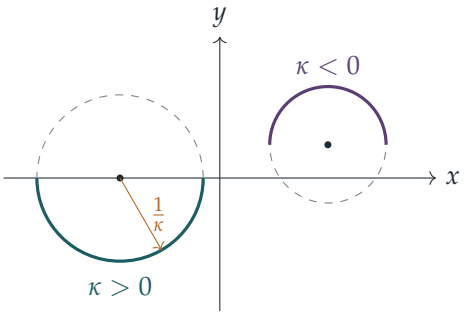
整理后得 $\kappa(y-b) = -\sqrt{1-\kappa^2(x-a)^2}$，再平方便得

$$(x-a)^2 + (y-b)^2 = \frac{1}{\kappa^2}.$$

因此，每条具有非零常数有号曲率（signed curvature）κ 的曲线，都是以 (a,b) 为圆心、半径为 $1/|\kappa|$ 的圆的一部分；而且每段这样的圆弧都会出现：把步骤倒过来，上述每个函数的有号曲率都恰好是 κ。更准确地说，$\kappa(y-b) \leq 0$，所以：

- 若 $\kappa > 0$, 曲线是下半圆 $y = b - \sqrt{\kappa^{-2} - (x-a)^2}$, 位于圆心下方, 向上弯曲 ($\mathrm{f}'' > 0$);
- 若 $\kappa < 0$, 曲线是上半圆 $y = b + \sqrt{\kappa^{-2} - (x-a)^2}$, 位于圆心上方, 向下弯曲 ($\mathrm{f}'' < 0$)。

端点 $x = a \pm 1/|\kappa|$ 须排除，因为该处斜率为无限大；曲线 $y = \mathrm{f}(x)$ 只能是一个函数的图像，所以它最多只能是一个开半圆（或其中一段）。



几何意义

对这些曲线而言，$|\kappa|$ 是圆半径的倒数，所以 $|\kappa|$ 越大，圆就越小、弯得越紧。κ 的正负号记录曲线向哪一方弯曲：当圆心位于曲线上方（曲线上凹（concave up））时 $\kappa > 0$；当圆心位于曲线下方时 $\kappa < 0$。

验算

- 取 $\kappa = \frac{1}{2}$，$a = b = 0$：曲线是 $y = -\sqrt{4 - x^2}$，即半径为 2 的圆的下半部分。在 $x = 1$ 处，$y = -\sqrt{3}$，$y' = x/\sqrt{4 - x^2} = 1/\sqrt{3}$，而 $y'' = 4/(4 - x^2)^{3/2} = 4/3^{3/2}$，所以 $1 + y'^2 = \frac{4}{3}$，且 $\kappa = \dfrac{4/3^{3/2}}{(4/3)^{3/2}} = \frac{1}{2}$，与预期相符。
- 直线有 $\mathrm{f}'' = 0$，所以 $\kappa = 0$：这是半径 $1/|\kappa| \to \infty$ 时圆的极限，亦是题目排除 $\kappa = 0$ 的原因。
- 用电脑从随机初始值逐步数值求解 $\mathrm{f}'' = \kappa(1 + \mathrm{f}'^2)^{3/2}$，所得各点准确至小数点后六位都留在预测的圆上。

常见错误

- 第一次积分时忘记积分常数（constant of integration），因而失去把圆心左右移动的自由度。
- 为 p 选错平方根，结果得到位于圆心错误一侧的半圆，以及错误的 κ 正负号。
- 只说“κ 是半径”：其实它是半径的倒数，并带有表示弯曲方向的正负号。

STEP 之外

曲率是曲线微分几何（differential geometry）的起点。对以弧长（arc length）s 参数化的曲线，$\kappa = \mathrm{d}\psi/\mathrm{d}s$，其中 ψ 是切线（tangent）与 x 轴所成的角；这里 $p/\sqrt{1 + p^2} = \sin\psi$，所以 $(*)$ 实际上是说 ψ 沿曲线以恒定速率变化。在三维空间中，曲率之外还有挠率（torsion），而 Frenet–Serret 公式（Frenet–Serret formulae）显示，这两个函数在相差一个刚体运动（rigid motion）的意义下决定了一条曲线。Pressley 的 *Elementary Differential Geometry*（《初等微分几何》）是一本大学一年级程度、平易近人的入门书。

详细解答

STEP I 2016，第 5 题

纯数学 · 相切圆

STEP I 2016，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/16-s1-q5

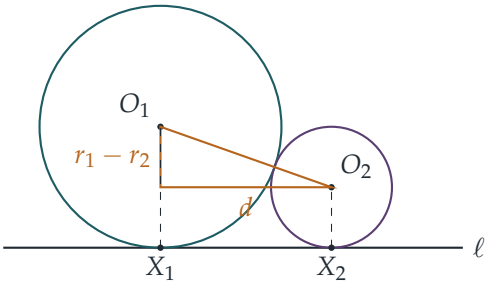
关键思路

一切都建基于一个事实：若两个圆互相相切（touch），而且都与同一条直线相切，那么它们与该直线的两个切点之间的距离是 $2\sqrt{r_1 r_2}$。这只需要一个直角三角形（right-angled triangle）便可得出。之后，代换（substitution）$x = 1/\sqrt{a}$、$y = 1/\sqrt{b}$、$z = 1/\sqrt{c}$ 把 $(**)$ 化成一个对称（symmetric）的四次式（quartic），而它可以因式分解（factorise）成四个一次因式（linear factor），其中一个因式恰好就是 $(*)$。

解答

第 (i) 部分

在直线上互相相切的两个圆。取半径分别为 r_1 和 r_2 的两个圆，它们互相外切（touch externally），都位于直线 ℓ 的同一侧，并分别在点 X_1 和 X_2 与 ℓ 相切。连到 X_1 和 X_2 的半径都垂直于 ℓ，所以圆心 O_1、O_2 位于 ℓ 上方高度 r_1 和 r_2 之处。由于两圆外切，$O_1O_2 = r_1 + r_2$。



从 O_2 向半径 O_1X_1 作垂线（图中 $r_1 \geq r_2$）。这样便得到一个直角三角形，其斜边（hy-

potenuse）为 r_1+r_2，铅垂边为 $|r_1-r_2|$，水平边为 $d=X_1X_2$。由毕氏定理（Pythagoras' theorem），

$$d^2=(r_1+r_2)^2-(r_1-r_2)^2=4r_1r_2,\qquad \text{所以}\qquad X_1X_2=2\sqrt{r_1r_2}.$$

应用三次。在图中，每两个圆都互相相切，而三个圆都与直线相切，切点为 P、Q、R，其中 Q 位于 P 与 R 之间。所以

$$PQ=2\sqrt{ab},\qquad QR=2\sqrt{bc},\qquad PR=2\sqrt{ac},$$

而由 $PQ+QR=PR$ 得 $\sqrt{ab}+\sqrt{bc}=\sqrt{ac}$。两边除以 $\sqrt{abc}>0$，

$$\frac{1}{\sqrt{c}}+\frac{1}{\sqrt{a}}=\frac{1}{\sqrt{b}},$$

这正是 $(*)$。

推导 $(**)$。设 $x=\dfrac{1}{\sqrt{a}}$、$y=\dfrac{1}{\sqrt{b}}$、$z=\dfrac{1}{\sqrt{c}}$，于是 $\dfrac{1}{a}=x^2$，如此类推，而 $(*)$ 即是 $y=x+z$。那么 $(**)$ 变成 $2(x^4+y^4+z^4)=(x^2+y^2+z^2)^2$；展开右方并移项，它等价于

$$F:=x^4+y^4+z^4-2x^2y^2-2y^2z^2-2z^2x^2=0.$$

想一想 这个四次式可以怎样因式分解?

把它组合成平方差（difference of two squares）。由于 $(x^2+y^2-z^2)^2=x^4+y^4+z^4+2x^2y^2-2y^2z^2-2z^2x^2$，

$$\begin{aligned}F&=(x^2+y^2-z^2)^2-(2xy)^2\\&=\left(x^2-2xy+y^2-z^2\right)\left(x^2+2xy+y^2-z^2\right)\\&=\left((x-y)^2-z^2\right)\left((x+y)^2-z^2\right)\\&=(x-y-z)(x-y+z)(x+y-z)(x+y+z).\end{aligned}$$

由 $(*)$，$x-y+z=0$，所以 $F=0$，即 $(**)$ 成立。

第 (ii) 部分

现在 a、b、c 是任意满足 $(**)$ 且 $b<c<a$ 的正数。如前定义 x、y、z；它们都是正数，而由于 $t\mapsto 1/\sqrt{t}$ 在 $t>0$ 上是递减（decreasing）的，条件 $b<c<a$ 给出

$$y>z>x>0.$$

如第 (i) 部分所示，$(**)$ 等价于

$$(x-y-z)(x-y+z)(x+y-z)(x+y+z)=0,$$

所以四个因式之中有一个为零。其中三个不可能为零：

- $x+y+z>0$，因为三者都是正数；
- $x+y-z>0$，因为 $y>z$ 且 $x>0$；
- $x-y-z<0$，因为 $x<y$ 且 $z>0$。

因此 $x-y+z=0$，即 $y=x+z$，亦即

$$\frac{1}{\sqrt{b}}=\frac{1}{\sqrt{a}}+\frac{1}{\sqrt{c}},$$

这正是 $(*)$。（其实只需要用到 b 是三者中最小这一点。）

验算

- 取 $a=4$、$c=1$。那么由 $(*)$ 得 $1/\sqrt{b}=\frac{1}{2}+1=\frac{3}{2}$，所以 $b=\frac{4}{9}$。此时 $2\left(\frac{1}{16}+\frac{81}{16}+1\right)=\frac{49}{4}$，而 $\left(\frac{1}{4}+\frac{9}{4}+1\right)^2=\left(\frac{7}{2}\right)^2=\frac{49}{4}$，正如 $(**)$ 所要求。
- 为甚么 (ii) 需要大小次序的条件？取 $a=4$、$c=1$，关系式 $(**)$ 是 $1/b$ 的二次方程，有两个根：$b=\frac{4}{9}$ 和 $b=4$（由 $1/\sqrt{b}=|x-z|=\frac{1}{2}$ 得出）。第二个根使 b 成为最大而非最小，它满足 $(**)$ 却不满足 $(*)$：此时以 $1/\sqrt{c}=1/\sqrt{a}+1/\sqrt{b}$ 联系起来的是 a 和 b。对调字母的角色，对应于对调哪一个圆位于中间。
- 以电脑对数以千计随机生成、在一条直线上互相相切的三圆构形作检验，$(*)$ 和 $(**)$ 均得到证实。

常见错误

- 把圆心之间的距离 r_1+r_2 当作切点之间的距离：两个圆心的高度并不相同。
- 在第 (ii) 部分，未排除其余三个因式便从因式分解得出 $(*)$；这正是需要用到 $b<c<a$ 的地方。

STEP 之外

关系式 $(**)$ 是笛卡儿圆定理（Descartes' circle theorem）的退化情况（degenerate case）：对于四个两两相切、曲率（curvature）为 $k_i=1/r_i$ 的圆，有 $2\sum k_i^2=\left(\sum k_i\right)^2$，而一条直线可视为曲率为 0 的圆。不断在空隙中填入新的圆，便得到阿波罗尼斯垫片（Apollonian gasket），其中整数曲率引出令人惊讶的数论（number theory）结果。F 的因式分解也正是希罗公式（Heron's formula）背后的那一个：$16\Delta^2=-F$，其中 x,y,z 为三角形的三边。若想进一步了解圆的几何以及能令这些结果一目了然的反演（inversion），可参阅 Needham 的 *Visual Complex Analysis*（《直观复分析》）。

详细解答

STEP I 2016，第 6 题

纯数学 · 向量几何

STEP I 2016，第 6 题

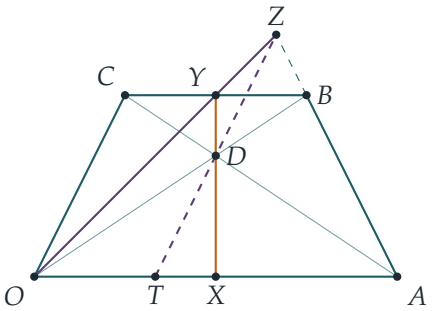
继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/16-s1-q6

关键思路

图中每一点都是两条直线的交点，因此我们把每条直线写成向量形式（vector form），然后令两者相等。由于 $\mathbf{a}$ 与 $\mathbf{b}$ 不平行（parallel），任何向量都只能以唯一一种方式写成 $\alpha\mathbf{a}+\beta\mathbf{b}$，所以把同一点的两个表达式相等，就可以分别比较 $\mathbf{a}$ 与 $\mathbf{b}$ 的系数（coefficients）。全程以 $\mathbf{a}$、$\mathbf{b}$ 为基（basis），每一步都只需解两条简短的线性方程（linear equations）。

解答



c 与 x 的形式

由于 CB 与 OA 平行，向量 $\overrightarrow{BC}=\mathbf{c}-\mathbf{b}$ 是 $\mathbf{a}$ 的纯量倍数（scalar multiple），设 $\mathbf{c}-\mathbf{b}=k\mathbf{a}$，于是 $\mathbf{c}=k\mathbf{a}+\mathbf{b}$。沿四边形（quadrilateral）$O\to A\to B\to C$ 走一圈，边 BC 的方向必须与 OA 相反（若方向相同，边 AB 与 CO 便会相交，$OABC$ 就不是真正的四边形），而且 $C\neq B$。所以

$$k<0.$$

任何负值都有可能；$k=-1$ 时就是平行四边形（parallelogram）。

由于 X 在 OA 上，$\mathbf{x}$ 是 $\mathbf{a}$ 的倍数，设 $\mathbf{x}=m\mathbf{a}$；而 X 严格位于 O 与 A 之间，即

$$0<m<1.$$

此外，由于 O、A、B 不共线（collinear），$\mathbf{a}$ 与 $\mathbf{b}$ 不平行。

点 Z

求 D。OB 上的点为 $\lambda\mathbf{b}$；AC 上的点为

$$\mathbf{a}+\mu(\mathbf{c}-\mathbf{a})=(1-\mu+\mu k)\mathbf{a}+\mu\mathbf{b}.$$

比较系数，得 $\lambda=\mu$ 及 $1-\mu(1-k)=0$。因为 $k<0$，所以 $1-k>0$，故 $\mu=\frac{1}{1-k}$，且

$$\mathbf{d}=\frac{\mathbf{b}}{1-k}.$$

求 Y。BC 上的点为 $\mathbf{b}+t\mathbf{a}$。XD 上的点为

$$m\mathbf{a}+s(\mathbf{d}-m\mathbf{a})=m(1-s)\mathbf{a}+\frac{s}{1-k}\,\mathbf{b}.$$

由 $\mathbf{b}$ 的系数得 $s=1-k$，于是 $\mathbf{a}$ 的系数为 $m(1-(1-k))=mk$。所以

$$\mathbf{y}=\mathbf{b}+mk\,\mathbf{a}.$$

求 Z。OY 上的点为 $\rho(\mathbf{b}+mk\mathbf{a})$；$AB$ 上的点为 $\mathbf{a}+\sigma(\mathbf{b}-\mathbf{a})=(1-\sigma)\mathbf{a}+\sigma\mathbf{b}$。比较系数，得 $\rho=\sigma$ 及 $\rho mk=1-\sigma$，所以 $\rho(mk+1)=1$。只要 $mk+1\neq 0$，便有

$$\mathbf{z}=\frac{\mathbf{b}+mk\,\mathbf{a}}{mk+1},$$

此即所求。（若 $mk=-1$——这在 $k<-1$ 时可能发生——则 $\mathbf{y}=\mathbf{b}-\mathbf{a}$，所以 OY 与 AB 平行，Z 不存在。）

> ? **想一想** Z 是否在边 AB 本身之上?

不是：因为 $-1<mk<0$，所以 $\sigma=\frac{1}{mk+1}>1$，即 Z 位于 AB 越过 B 的延长线上，正如图中所示。这并不要紧，因为题目讨论的是直线。

点 T 与两条关系式

DZ 上的点为 $\mathbf{d}+\tau(\mathbf{z}-\mathbf{d})$。其 $\mathbf{b}$ 的系数为

$$\frac{1}{1-k}+\tau\left(\frac{1}{mk+1}-\frac{1}{1-k}\right)=\frac{1}{1-k}-\tau\,\frac{k(1+m)}{(mk+1)(1-k)},$$

这里用了 $(1-k)-(mk+1) = -k(1+m)$。要 T 在 OA 上，此系数必须为零，所以 $\tau = \dfrac{mk+1}{k(1+m)}$（留意 $k \neq 0$）。于是 $\mathbf{a}$ 的系数为

$$\tau \cdot \frac{mk}{mk+1} = \frac{mk+1}{k(1+m)} \cdot \frac{mk}{mk+1} = \frac{m}{1+m}.$$

因此

$$\mathbf{t} = \frac{m}{1+m}\,\mathbf{a}.$$

值得注意的是，k 消失了：T 只取决于 X 的位置。

现设 $a = |\mathbf{a}| = OA$。由于 $0 < \frac{m}{1+m} < 1$，T 位于 O 与 A 之间，所以

$$OT = \frac{ma}{1+m}, \qquad TA = a - \frac{ma}{1+m} = \frac{a}{1+m}, \qquad OX = ma.$$

因此

$$OT \times OA = \frac{ma^2}{1+m} = ma \times \frac{a}{1+m} = OX \times TA,$$

以及

$$\frac{1}{OX} + \frac{1}{OA} = \frac{1}{ma} + \frac{1}{a} = \frac{1+m}{ma} = \frac{1}{OT},$$

这两条正是所要证明的结果。

验算

试用图中的梯形（trapezium）：$\mathbf{a} = (4,0)$，$\mathbf{b} = (3,2)$，$k = -\frac{1}{2}$，$m = \frac{1}{2}$。于是 $D = (2, \frac{4}{3})$，$Y = (2,2)$，$Z = (\frac{8}{3}, \frac{8}{3})$，$T = (\frac{4}{3}, 0)$；你可以验证 T、D、Z 共线，而且 $\frac{1}{OT} = \frac{3}{4} = \frac{1}{2} + \frac{1}{4} = \frac{1}{OX} + \frac{1}{OA}$。以电脑对 20,000 个随机梯形作检验，结果与上述每条公式的误差都在 10^{-10} 以内。极限情况 $m \to 1$（$X \to A$）给出 $OT = \frac{1}{2}OA$，与 $\frac{1}{OT} = \frac{2}{OA}$ 一致。

常见错误

- 取 $k > 0$：若 $\mathbf{c} = \mathbf{b} + k\mathbf{a}$ 且 $k > 0$，这个“四边形”会自我相交。
- 比较系数时没有说明为何可以这样做：这要依赖 $\mathbf{a}$ 与 $\mathbf{b}$ 不平行。
- 对两条不同的直线使用同一个参数（parameter）字母，这会强加一个虚假的重合。

STEP 之外

T 不取决于 k(亦不取决于 $\mathbf{a}$ 与 $\mathbf{b}$ 之间的夹角),这说明它其实是仿射几何(affine geometry),甚至是射影几何（projective geometry）中的定理：要紧的只是点线的关联关系，以及同

一直线上的比。关系式 $\frac{1}{OT} = \frac{1}{OX} + \frac{1}{OA}$ 正是交比（cross-ratio）与调和点列（harmonic range）能优雅处理的那类命题；而重复同一作图，便可依次得到位于 $\frac{m}{1+2m}\mathbf{a}$、$\frac{m}{1+3m}\mathbf{a}$ 等处的点。如欲深入了解，可参阅考克斯特（Coxeter）与 Greitzer 合著的 *Geometry Revisited*（《几何学的新探索》），或考克斯特的 *Projective Geometry*（《射影几何》）。

详细解答

STEP I 2016，第 7 题

纯数学 • 模 4 的因式分解

> **STEP I 2016，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/16-s1-q7

本题在 *Beyond STEP* 第 1.7 节因数分解失效之时中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch01/factorisation-fails.html

关键思路

一切都源于模（modulo）4 运算。奇数必定 $\equiv 1$ 或 $\equiv 3 \equiv -1 \pmod 4$，因此一组奇数的乘积属于 S 还是 T，取决于其中来自 T 的因数个数是偶数还是奇数。单凭这一条奇偶性（parity）规则，便可解决第 (ii)、(iii) 及 (iv)(a) 部分。第 (iv)(b) 部分则要求你看出 S-质数不一定是质数，而这就令唯一分解（unique factorisation）失效。

解答

第 (i) 部分

每个正奇数除以 4 的余数都是 1 或 3，而偶数则不会有这样的余数。所以 $S \cup T$ 是所有正奇数组成的集合。任何整数都不会有两个不同的余数，所以 $S \cap T$ 是空集（empty set）。

第 (ii) 部分

取 S 中的 $4a+1$ 和 $4b+1$，其中 $a, b \geq 0$ 为整数。那么

$$(4a+1)(4b+1) = 16ab + 4a + 4b + 1 = 4(4ab + a + b) + 1,$$

这是一个余数为 1 的正整数，所以它属于 S。

至于 T，答案是否。例如 $3 \in T$，但 $3 \times 3 = 9 = 4 \times 2 + 1 \in S$，而 $S \cap T$ 是空集。事实

上这种情况必定发生：

$$(4a+3)(4b+3) = 16ab+12a+12b+9 = 4(4ab+3a+3b+2)+1 \in S.$$

我们亦需要混合的乘积：

$$(4a+1)(4b+3) = 16ab+12a+4b+3 = 4(4ab+3a+b)+3 \in T.$$

简而言之，把 T 中的数写成 $\equiv -1 \pmod 4$：一组奇数的乘积 $\equiv (-1)^k \pmod 4$，其中 k 是来自 T 的因数个数。所以若 k 为偶数，乘积属于 S；若 k 为奇数，乘积属于 T。

第 (iii) 部分

设 $n \in T$ 为合数（composite），并把 $n = p_1 p_2 \cdots p_r$ 写成质数的乘积。由于 n 是奇数，每个 p_i 都是奇数，所以由第 (i) 部分，每个 p_i 都属于 S 或 T。

用反证法（proof by contradiction），假设每个 p_i 都属于 S。反复应用第 (ii) 部分，得 $p_1 p_2 \in S$，然后 $(p_1 p_2)p_3 \in S$，如此类推，所以 $n \in S$。但 $n \in T$ 而 $S \cap T = \varnothing$，这是矛盾。因此 n 至少有一个质因数属于 T。（用奇偶性的说法：要令 $n \equiv -1$，需要奇数个、因而至少一个 $\equiv -1$ 的因数。）

第 (iv)(a) 部分

首先留意 $1 \notin T$，所以 T 中每个元素都至少是 3。

*每个 $n \in T$ 要么是 T-质数，要么是若干个 T-质数的乘积。*我们对 n 使用强归纳法（strong induction）。若 n 是 T-质数，则已完成。否则 $n = t_1 t_2 \cdots t_k$，其中 $k \geq 2$ 且每个 $t_i \in T$。每个 $t_i \geq 3$，所以每个 $t_i \leq n/3 < n$，由归纳假设，每个 t_i 要么是 T-质数，要么是若干个 T-质数的乘积。把这些代入 $n = t_1 \cdots t_k$，便把 n 写成 T-质数的乘积。

> ? **想一想** 我们得到了一个 T-质数的乘积，但为甚么它们的个数必定是奇数？

每个 T-质数都属于 T，所以它 $\equiv -1 \pmod 4$。根据第 (ii) 部分末尾的规则，m 个这样的数的乘积因此 $\equiv (-1)^m \pmod 4$。若 m 为偶数，乘积便 $\equiv 1$，因而属于 S 而非 T。由于乘积是 $n \in T$，所以 m 是奇数。因此每个 $n \in T$ 要么是 T-质数，要么是奇数个 T-质数的乘积，得证。

留意 T-质数不一定是质数：$15 = 3 \times 5$ 是 T-质数，因为 $5 \notin T$，而分解成大于 1 的数的唯一方法是 3×5。

第 (iv)(b) 部分

与上面关于 15 的观察相同的想法，在 S 中同样适用：两个来自 T 的质数的乘积属于 S，但在 S 之内无法分解。取

$$9 = 3 \times 3, \qquad 21 = 3 \times 7, \qquad 49 = 7 \times 7.$$

每个都属于 S。每个都是 S-质数：除了 1 和它本身以外，它的因数只有 3 或 7，而这些属于 T 而非 S。（对于 21，因数是 $1, 3, 7, 21$。）现在

$$441 = 9 \times 49 = 21 \times 21,$$

而 $441 = 4 \times 110 + 1 \in S$。这是 441 分解成 S-质数的两种真正不同的方法：第一种用了 9 和 49，第二种用了两次 21。所以 441 是一个例子。

? **想一想** 你会怎样找到这个例子?

在普通整数中，$441 = 3^2 \times 7^2$，而质数 $3, 7$ 属于 T。在 S 之内，我们不能单独分出一个 3 或 7，所以 T 中的质数必须两两配对。有两个 3 和两个 7 时，配对方法有两种：$(3 \cdot 3)(7 \cdot 7)$ 或 $(3 \cdot 7)(3 \cdot 7)$。任何 p^2q^2（其中 $p, q \in T$ 为不同质数）都同样可行，例如 $9 \times 121 = 33 \times 33$。

验算

- 试试小的例子：$5 \times 9 = 45 \in S$，$3 \times 7 = 21 \in S$，$5 \times 7 = 35 \in T$。
- 对于 (iii)：$15 = 3 \times 5$ 和 $35 = 5 \times 7$ 都属于 T，而且各自都有一个属于 T 的质因数（3、7）。
- 用电脑验证：对每个小于 3000 的 $n \in T$，每种分解成 T-质数的方法都有奇数个因数。T-质数的开头是 $3, 7, 11, 15, 19, 23, 31, 35, \ldots$。留意 $27 = 3 \times 3 \times 3$ 不是 T-质数：它有三个因数，是奇数。

常见错误

- 没有测试任何例子便声称 T 中两个元素的乘积属于 T：$3 \times 3 = 9$ 立即否定了这说法。
- 忘记 X-质数不一定是质数。这正是 (iv)(b) 的重点：9 和 21 都是 S-质数。
- 在 (iv)(a) 中只证明了因数个数是奇数，却忘记证明分解成 T-质数的方法确实存在。

STEP 之外

集合 S 对乘法封闭，而它的“质数”并不能唯一分解。这是一个经典的玩具模型（有时称为希尔伯特数（Hilbert numbers），以希尔伯特（Hilbert）命名），说明了在 $\mathbb{Z}[\sqrt{-5}]$ 这类环（ring）中出现的问题：在那里 $6 = 2 \times 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$。为了在这类环中修补唯一分解，数学家发展出理想（ideals）和代数数论（algebraic number theory）。Davenport 的 *The Higher Arithmetic*（《高等算术》）解释了为甚么 $\mathbb{Z}$ 中的唯一分解是一条需要证明的定理，而 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）则讨论了唯一分解失效的环。

详细解答

STEP I 2016，第 8 题

纯数学 · 生成函数

STEP I 2016，第 8 题

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

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

本题在 *Beyond STEP* 第 9.4 节把数列看成函数中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/generating-functions.html

关键思路

生成函数（generating function）把整个数列包装成单一个函数，而对数列的运算就变成对这个函数的代数运算。在第 (i) 部分，$1-x$ 的负数次幂的二项级数（binomial series）给出我们想要的系数，然后我们用已知的数列组合出 n^2。在第 (ii) 部分，我们把递推关系（recurrence）乘以 x^n 再求和，从而化成关于 f 的方程：把数列的下标平移，等同于把它的生成函数乘以 x。

解答

第 (i) 部分

由指数为 -2 的二项级数，

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

因为 $\binom{-2}{n}(-1)^n = \frac{2\cdot 3\cdots(n+1)}{n!} = n+1$。乘以 x，

$$x(1-x)^{-2} = \sum_{n\geq 0} (n+1)x^{n+1} = \sum_{n\geq 0} nx^n,$$

这里重新编排求和下标后，可以无碍地把 $n=0$ 也包括在内（该项为 0）。所以 $u_n = n$ 的生成函数是 $x(1-x)^{-2}$，如题所求。

同理，取指数 -3，

$$\binom{-3}{n}(-1)^n = \frac{3\cdot 4\cdots(n+2)}{n!} = \frac{(n+1)(n+2)}{2},$$

所以

$$(1-x)^{-3} = \sum_{n\geq 0}\frac{(n+1)(n+2)}{2}x^n.$$

乘以 x 并如前重新编排下标，

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

所以该数列是 $u_n = \frac{1}{2}n(n+1)$，即 $0,1,3,6,10,\ldots$，也就是三角数（triangular numbers）。

想一想 怎样从我们现在已知的两个数列得出 n^2?

生成函数是线性（linear）的：若 u_n 和 v_n 的生成函数分别是 f 和 g，则 $\alpha u_n + \beta v_n$ 的生成函数是 $\alpha \mathrm{f} + \beta \mathrm{g}$。由于

$$n^2 = 2\cdot\tfrac{1}{2}n(n+1) - n,$$

$u_n = n^2$ 的生成函数是

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

第 (ii)(a) 部分

对每个 $n \geq 1$，有 $u_n x^n = k u_{n-1} x^n$。对所有 $n \geq 1$ 求和：

$$\sum_{n\geq 1} u_n x^n = kx \sum_{n\geq 1} u_{n-1}x^{n-1}$$

$$\mathrm{f}(x) - u_0 = kx\,\mathrm{f}(x),$$

因为左边的和是 $\mathrm{f}(x)$ 去掉常数项，而右边的和在令 $m = n-1$ 后是 $\sum_{m\geq 0} u_m x^m = \mathrm{f}(x)$。由于 $u_0 = a$，这就是 $\mathrm{f}(x) = a + kx\,\mathrm{f}(x)$，如题所求。因此

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

（作为验算，展开得 $a + akx + ak^2x^2 + \cdots$，而确实 $u_n = ak^n$。）

第 (ii)(b) 部分

把 $u_n = u_{n-1} + u_{n-2}$ 乘以 x^n，并对 $n \geq 2$ 求和（递推关系只由 $n = 2$ 起成立，所以必须由这里开始）：

$$\sum_{n\geq 2} u_n x^n = x\sum_{n\geq 2} u_{n-1}x^{n-1} + x^2\sum_{n\geq 2} u_{n-2}x^{n-2}.$$

想一想 这三个和各自缺少了 f 的哪些项?

左边缺少 $u_0 + u_1x$ 这些项；右边第一个和是 $u_1x + u_2x^2 + \cdots$，所以只缺少 u_0；第二个和就是整个 $\mathrm{f}(x)$。因此

$$\mathrm{f}(x) - u_0 - u_1x = x\big(\mathrm{f}(x) - u_0\big) + x^2\,\mathrm{f}(x).$$

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

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

这是斐波那契数（Fibonacci numbers）$0, 1, 1, 2, 3, 5, 8, \ldots$ 的生成函数。

验算

- 展开 $x(1+x)(1-x)^{-3} = (x+x^2)(1+3x+6x^2+10x^3+\cdots)$ 得 $x + 4x^2 + 9x^3 + 16x^4 + \cdots$，正是平方数，符合预期。
- 以长除法（long division）把 x 除以 $1 - x - x^2$，得 $x + x^2 + 2x^3 + 3x^4 + 5x^5 + \cdots$。
- 在 $x = 0.1$ 时和 $\sum n^2(0.1)^n$ 与 $0.11/0.9^3 = 0.150891\ldots$ 相符，而斐波那契数的 $\sum u_n(0.1)^n$ 与 $0.1/0.89 = 10/89 = 0.112359\ldots$ 吻合。这些结果，连同题目中每个生成函数的首 30 个系数，均已用电脑核实。

常见错误

- 弄错负指数的二项式系数（binomial coefficients）：来自 $\binom{-r}{n}$ 和来自 $(-x)^n$ 的正负号互相抵消，所以每个系数都是正数。
- 在 (ii)(b) 中由 $n = 0$ 或 $n = 1$ 开始求和，而递推关系在那里并不成立；或者忘记减去首项 u_0 和 u_1x。
- 把三角数写成 $\frac{1}{2}n(n-1)$：检查一下 $u_1 = 1$。

STEP 之外

把 $1 - x - x^2 = (1 - \phi x)(1 - \psi x)$ 因式分解，其中 $\phi, \psi = \frac{1}{2}(1 \pm \sqrt{5})$，再运用部分分式（partial fractions）配合第 (ii)(a) 部分，便得到比奈公式（Binet's formula）$u_n = (\phi^n - \psi^n)/\sqrt{5}$。同一方法可以解任何常系数线性递推关系，而生成函数是组合数学（combinatorics）和概率中的核心工具（在概率中它们成为概率生成函数（probability generating functions））。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）阐述概率方面的内容；

Wilf 的 *generatingfunctionology*（《生成函数学》）则是一本生动有趣、可免费取得的书，讲述组合方面的内容。

详细解答

STEP I 2016，第 9 题

力学 · 极限摩擦力平衡

> **STEP I 2016，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/16-s1-q9

关键思路

对 A 取力矩（moments）：墙上的两个力因此消去，栏杆处的法向反作用力（normal reaction）可直接以 θ 表示，完全不涉及摩擦力（friction）。只要知道每个摩擦力的方向，再沿水平和铅垂方向分解（resolving）便可完成。这正是本题的真正重点：摩擦力与杆即将发生的运动方向相反，所以我们必须判断端点 A 是即将沿墙向下滑，还是向上滑。

解答

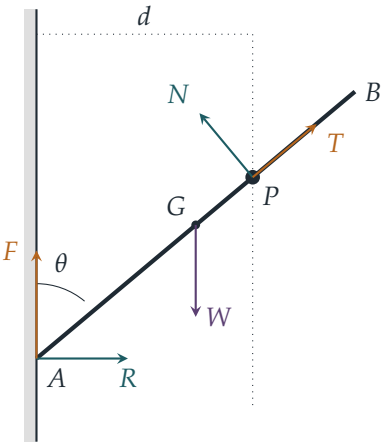
把杆模型化为重 W 的均匀刚性杆（uniform rigid rod），其中心 G 与 A 的距离为 a；把栏杆模型化为一条固定直线，与杆只相交于一点 P。由于 A 到栏杆的水平距离为 d，而杆与铅垂线成角 θ，

$$AP = \frac{d}{\sin\theta}.$$

条件 $a\sin\theta < d$ 表示 G 位于 A 与 P 之间。

各力如下：

- 在 A：法向反作用力 R，方向水平并离开墙壁；以及沿墙的摩擦力 F；
- 在 P：垂直于杆的法向反作用力 N；以及沿杆的摩擦力 T；
- 作用于 G 的重量 W。



图中 F 和 T 的箭头方向对应下文的第一种情况。

对 A 取力矩

由于 N 垂直于杆，N 对 A 的力矩为 $N \cdot AP$；T 沿杆作用，对 A 没有力矩，R 和 F 亦然。重量的力臂（lever arm）为 $a\sin\theta$。因此

$$N\frac{d}{\sin\theta} = Wa\sin\theta \quad \Longrightarrow \quad N = \frac{Wa\sin^2\theta}{d}. \tag{1}$$

无论杆即将向哪个方向滑动，此式均成立。

杆向哪个方向滑动?

? **想一想** A 是即将沿墙向下滑，还是向上滑?

先假设完全没有摩擦力。沿铅垂方向分解，需要 $N\sin\theta = W$，结合 (1) 即 $a\sin^3\theta = d$。但此处 $a\sin^3\theta < a\sin\theta < d$，故由 (1) 得 $N\sin\theta < W$：单靠法向反作用力无法撑住杆，其余部分必须由摩擦力提供。杆即将向下滑动：A 沿墙下滑，θ 减小，$AP = d/\sin\theta$ 增大，于是杆穿过栏杆向墙壁方向滑下。摩擦力阻止这个运动，所以 F 沿墙向上，T 沿杆向上（指向 B），如图所示。在两个接触点均处于极限平衡（limiting equilibrium）时，

$$F = \mu R, \qquad T = \lambda N.$$

分解各力

沿杆向上的单位向量（unit vector）为 $(\sin\theta, \cos\theta)$，而 N 的方向为 $(-\cos\theta, \sin\theta)$（指向墙壁并向上）。沿水平方向分解：

$$R - N\cos\theta + \lambda N\sin\theta = 0 \quad \Longrightarrow \quad R = N(\cos\theta - \lambda\sin\theta).$$

（墙壁只能推，故需 $R \geq 0$，所以此情况要求 $\tan\theta \leq 1/\lambda$。）沿铅垂方向分解：

$$\begin{aligned} W &= \mu R + N\sin\theta + \lambda N\cos\theta \\ &= N\big(\mu\cos\theta - \lambda\mu\sin\theta + \sin\theta + \lambda\cos\theta\big) \\ &= N\big((\lambda+\mu)\cos\theta + (1-\lambda\mu)\sin\theta\big). \end{aligned}$$

代入 (1) 中的 N，再除以 $W\sin^2\theta$，得

$$d\operatorname{cosec}^2\theta = a\big((\lambda+\mu)\cos\theta + (1-\lambda\mu)\sin\theta\big),$$

这就是所求的结果。

$a\sin\theta > d$ 的情况

此时 G 位于栏杆之外。方程 (1) 和几何关系都不变，所以唯一的问题仍然是滑动的方向。

> **想一想** G 移到栏杆之外后，有甚么改变？

上面的论证显示，当 $a\sin^3\theta < d$ 时杆向下滑，当 $a\sin^3\theta > d$ 时杆向上滑。在 $a\sin\theta < d$ 的条件下只可能出现前者。在 $a\sin\theta > d$ 的条件下，后者变得可能：伸出栏杆之外部分的重量使 $N\sin\theta > W$，于是杆倾向穿过栏杆向远离墙壁的方向滑下，而 A 沿墙向上移动。这是新的情况，此时两个摩擦力的方向都相反：$F = \mu R$ 沿墙向下，$T = \lambda N$ 沿杆向下。在上面的推导中把 λ 换成 $-\lambda$、μ 换成 $-\mu$，得

$$R = N(\cos\theta + \lambda\sin\theta), \qquad W = N\big((1-\lambda\mu)\sin\theta - (\lambda+\mu)\cos\theta\big),$$

再结合 (1)，

$$d\operatorname{cosec}^2\theta = a\big((1-\lambda\mu)\sin\theta - (\lambda+\mu)\cos\theta\big).$$

此时 $R > 0$ 自动成立。（若 $a\sin^3\theta < d < a\sin\theta$，杆仍然即将向下滑，第一个结果照样适用，因为其推导从未用到 G 的位置。）

验算

- 没有摩擦力时（$\lambda = \mu = 0$），两个结果都化为 $d = a\sin^3\theta$，即上文所得的无摩擦平衡。
- 设 $\lambda = \tan\alpha$ 及 $\mu = \tan\beta$（即摩擦角（angle of friction）），右方分别变为 $a\sin(\theta+\alpha+\beta)/(\cos\alpha\cos\beta)$ 和 $a\sin(\theta-\alpha-\beta)/(\cos\alpha\cos\beta)$：两种情况只在摩擦角的正负号上有分别，理应如此。
- 取 $\lambda = \mu = \frac{1}{2}$ 及 $\theta = \frac{1}{3}\pi$，两种情况分别给出 $d/a = \frac{3}{8} + \frac{9\sqrt{3}}{32} \approx 0.862$ 和 $d/a = \frac{9\sqrt{3}}{32} - \frac{3}{8} \approx 0.112$。按上述摩擦力方向，把每个值代回力和力矩的方程，均得合力为零及力矩为零。

常见错误

- 以为栏杆的法向反作用力是铅垂的。事实上它垂直于杆，因为杆与栏杆只在一点接触。
- 在力矩方程中，以长度 d 而非 $d/\sin\theta$ 作为 A 到栏杆的距离。
- 未先想清楚杆实际上会向哪个方向移动，便定下摩擦力的方向。

STEP 之外

在向下和向上两个极限情况之间，杆对一整个区间的角度 θ 都能保持平衡：摩擦力带来的是一系列平衡状态，而不是单一个。描述这类集合（每个接触点的摩擦锥（friction cone），以及各接触点能够维持的状态集合）是接触力学（contact mechanics）以及机械人抓取分析的起点。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）以大学一年级程度，仔细讲解摩擦力与刚体静力学（rigid-body statics）。

详细解答

STEP I 2016，第 10 题

力学 · 碰撞与恢复系数

STEP I 2016，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/16-s1-q10

关键思路

题中共有三次碰撞（collision），每一次都用同样的两条方程处理：动量守恒（conservation of momentum）和牛顿恢复定律（Newton's law of restitution）。头两次碰撞同时发生，但彼此独立，所以可以分开计算，求出 B 和 C 相遇前一刻的速度。再由 B 最终静止这个条件，得到一条联系 λ 与 e 的方程。

解答

我们把四个物体视为光滑桌面上的质点（particle），因此除碰撞外没有水平力作用，而且每次碰撞都是正碰（direct collision，即沿同一直线发生）。取由 A 指向 D 的方向为正方向，于是 A 的初速度为 u，D 的初速度为 $-u$。



第 (i) 部分

A 撞 B。设碰撞后 A 和 B 的速度分别为 v_A 和 v_B。则

$$\begin{aligned} \text{动量：} &\quad \lambda m u = \lambda m v_A + m v_B, \\ \text{恢复：} &\quad v_B - v_A = eu. \end{aligned}$$

把 $v_B = v_A + eu$ 代入第一条方程，得 $(\lambda+1)v_A = (\lambda-e)u$，所以

$$v_A = \frac{(\lambda-e)u}{\lambda+1}, \qquad v_B = v_A + eu = \frac{\lambda(1+e)u}{\lambda+1}.$$

D **撞** C。设碰撞后 C 和 D 的速度分别为 v_C 和 v_D。则

$$\begin{aligned} &\text{动量：} && -mu = mv_C + mv_D, \\ &\text{恢复：} && v_D - v_C = e\big(0-(-u)\big) = eu. \end{aligned}$$

（碰撞前 C 与 D 以速率 u 互相接近；碰撞后 D 在 C 的右方，所以两者以 $v_D - v_C$ 的速率分离。）解这两条方程：

$$v_C = -\tfrac{1}{2}(1+e)u, \qquad v_D = -\tfrac{1}{2}(1-e)u.$$

因此 C 以速率 $\frac{1}{2}(1+e)u$ 向 B 移动，而 D 则以较小的速率 $\frac{1}{2}(1-e)u$ 跟在后面向左移动。

B **撞** C。此时 B 以 v_B 向右移动，C 向左移动，所以两者的接近速率为 $v_B - v_C$。碰撞后 B 静止；设 C 的速度为 w。则

$$\begin{aligned} &\text{动量：} && mv_B + mv_C = 0 + mw, \\ &\text{恢复：} && w - 0 = e(v_B - v_C). \end{aligned}$$

消去 w，得 $v_B + v_C = e(v_B - v_C)$，即

$$(1-e)v_B = -(1+e)v_C.$$

代入上面求得的速度：

$$(1-e)\cdot\frac{\lambda(1+e)u}{\lambda+1} = (1+e)\cdot\tfrac{1}{2}(1+e)u.$$

由于 $u \neq 0$ 且 $1+e>0$，可以两边除以 $(1+e)u$：

$$\begin{aligned} 2\lambda(1-e) &= (1+e)(\lambda+1) \\ 2\lambda - 2\lambda e &= \lambda + 1 + e\lambda + e \\ \lambda - 1 &= e(3\lambda+1), \end{aligned}$$

又因 $3\lambda+1>0$，

$$e = \frac{\lambda-1}{3\lambda+1},$$

即为所求。

至于推论部分，

$$\frac{1}{3} - e = \frac{(3\lambda+1)-3(\lambda-1)}{3(3\lambda+1)} = \frac{4}{3(3\lambda+1)} > 0,$$

因为 $\lambda>1$。故 $e<\frac{1}{3}$。（另外，由于 $\lambda>1$，有 $e>0$，所以这些碰撞确实不是完全非弹性碰撞（perfectly inelastic collision）。）

第 (ii) 部分

第三次碰撞后，C 的速度为

$$w = v_B + v_C = (1+e)u\left(\frac{\lambda}{\lambda+1} - \frac{1}{2}\right) = \frac{(1+e)(\lambda-1)u}{2(\lambda+1)},$$

由于 $\lambda > 1$，此值为正：C 现在向右移动，折返向 D，而 D 仍以速率 $\frac{1}{2}(1-e)u$ 向左移动。

> **想一想** “速率相同”给出哪一条方程？又怎样消去 e？

两者速率相等时，

$$\frac{(1+e)(\lambda-1)}{2(\lambda+1)} = \frac{1-e}{2}.$$

由第 (i) 部分，$1+e = \dfrac{4\lambda}{3\lambda+1}$，$1-e = \dfrac{2(\lambda+1)}{3\lambda+1}$。代入并约去公因子 $\dfrac{1}{3\lambda+1}$：

$$\begin{aligned} \frac{4\lambda(\lambda-1)}{\lambda+1} &= 2(\lambda+1) \\ 2\lambda^2 - 2\lambda &= \lambda^2 + 2\lambda + 1 \\ \lambda^2 - 4\lambda - 1 &= 0, \end{aligned}$$

所以 $\lambda = 2 \pm \sqrt{5}$。只有 $\lambda = 2+\sqrt{5}$ 满足 $\lambda > 1$（另一个根是负数）。于是

$$e = \frac{1+\sqrt{5}}{7+3\sqrt{5}} = \frac{(1+\sqrt{5})(7-3\sqrt{5})}{49-45} = \frac{4\sqrt{5}-8}{4} = \sqrt{5}-2.$$

因此

$$\lambda = 2+\sqrt{5}, \qquad e = \sqrt{5}-2.$$

验算

- $e = \sqrt{5}-2 \approx 0.236$，确实介乎 0 与 $\frac{1}{3}$ 之间，符合第 (i) 部分的要求。
- 取 $\lambda = 2+\sqrt{5} \approx 4.236$、$u = 1$、$e \approx 0.236$，以数值逐一计算三次碰撞，得第三次碰撞前 $v_B = 1$（恰巧是精确值）及 $v_C \approx -0.618$；碰撞后 B 静止，C 以约 0.382 的速率向右移动，而 D 以约 0.382 的速率向左移动。共同速率为 $\frac{1}{2}(1-e)u = \frac{1}{2}(3-\sqrt{5})u$。
- 极限情况：当 $\lambda \to 1$ 时，$e \to 0$。质量相等且 $e = 0$ 时，B 与 C 以大小相等、方向相反的速度相遇，因此两者黏合后静止，结果合理。

常见错误

- 在 C 与 D 碰撞的恢复方程中弄错正负号：所有速度都应以同一个正方向表示，并小心运用“分离速率 $= e\times$ 接近速率”。

- 忘记 C 到达第三次碰撞时是朝向 B 移动的，所以接近速率是 $v_B - v_C = v_B + \frac{1}{2}(1+e)u$，而不是 $v_B - \frac{1}{2}(1+e)u$。
- 保留了负数根 $\lambda = 2 - \sqrt{5}$。

STEP 之外

带恢复系数的碰撞是能量损失最简单的模型：正碰中损失的动能（kinetic energy）与 $1 - e^2$ 成正比，而 $e = 1$ 就是能量守恒的弹性碰撞（elastic collision）。在大学一年级的课程中，这些概念会引出质心参考系（centre-of-mass frame）：在这个参考系中，任何二体碰撞看起来都是对称的，上面的代数运算几乎变得微不足道。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）就是以这种方式讲解碰撞的。

详细解答

STEP I 2016，第 11 题

力学 • 抛体的最大射程

STEP I 2016，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：

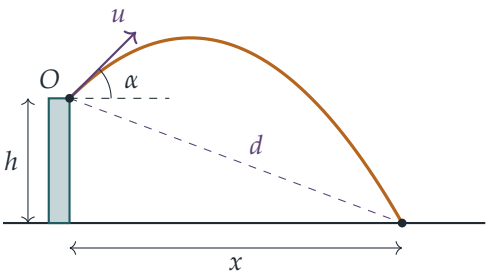
step.maths.org/questions/16-s1-q11

关键思路

写出水平和铅直方向的运动，消去时间，再把 $\sin\alpha$ 和 $\cos\alpha$ 的平方化成倍角（double angle）。至于最大射程，把落地方程视为以隐函数（implicit function）方式把 x 定义为 α 的函数，然后对它求导：在最大值处 $\mathrm{d}x/\mathrm{d}\alpha = 0$，余下的就只是关于单一未知量 $\cos 2\alpha$ 的三角代数运算。

解答

我们把 P 模拟为只受重力作用的质点（particle）（不计空气阻力（air resistance）），并假设平地是水平的。取坐标轴，原点在塔底，x 轴沿发射方向水平，y 轴铅直向上，因此 O 为点 $(0, h)$。



落地方程

发射后时间 t 时，

$$X = ut\cos\alpha, \qquad Y = h + ut\sin\alpha - \tfrac{1}{2}gt^2.$$

塔是直立的，而 P 落在某个距离 $x>0$ 处，所以 $\cos\alpha\neq 0$，且在落地时（即 $Y=0$ 时）$t=\dfrac{x}{u\cos\alpha}$：

$$0=h+x\tan\alpha-\frac{gx^2}{2u^2\cos^2\alpha}.$$

两边乘以 $2\cos^2\alpha$，并利用 $2\cos^2\alpha=1+\cos 2\alpha$ 及 $2\sin\alpha\cos\alpha=\sin 2\alpha$：

$$\frac{gx^2}{u^2}=2h\cos^2\alpha+2x\sin\alpha\cos\alpha=h(1+\cos 2\alpha)+x\sin 2\alpha,$$

这正是所求的方程。

最大水平距离

角 α 的取值范围为 $-\frac{\pi}{2}<\alpha<\frac{\pi}{2}$（容许向下发射）。对每个这样的 α，落地距离 x 是 α 的正值光滑函数，而且当 $\alpha\to\pm\frac{\pi}{2}$ 时 $x\to 0$，因为水平速率 $u\cos\alpha$ 趋于 0，而飞行时间保持有界。因此 x 在某个内部的 α 处取得最大值，而在该处 $\mathrm{d}x/\mathrm{d}\alpha=0$。

把落地方程对 α 求导，记 $x'=\mathrm{d}x/\mathrm{d}\alpha$：

$$\frac{2gx}{u^2}x'=-2h\sin 2\alpha+x'\sin 2\alpha+2x\cos 2\alpha.$$

令 $x'=0$，得

$$x\cos 2\alpha=h\sin 2\alpha.$$

这里 $\cos 2\alpha\neq 0$：否则由 $h\sin 2\alpha=0$ 及 $h>0$，便会迫使 $\sin 2\alpha=0$ 同时成立，这是不可能的。所以在最大值处

$$x=h\tan 2\alpha,$$

即为所求。

? **想一想** 现在我们有两条关于 x 和 α 的方程。怎样才能干净俐落地消去 x?

把 $x=h\tan 2\alpha$ 代入落地方程，并记 $c=\cos 2\alpha$、$s=\sin 2\alpha$。右方变为

$$h(1+c)+h\frac{s^2}{c}=\frac{h\,(c+c^2+s^2)}{c}=\frac{h(1+c)}{c},$$

而左方为

$$\frac{gh^2}{u^2}\cdot\frac{s^2}{c^2}=\frac{gh^2(1-c)(1+c)}{u^2c^2}.$$

令两者相等，再除以 $h(1+c)/c$——它不为零，因为 $c\neq -1$（我们已证明 $c\neq 0$，而 $c=-1$ 会给出 $\alpha=\pm\frac{\pi}{2}$）：

$$\frac{gh(1-c)}{u^2c}=1\quad\Longrightarrow\quad gh-ghc=u^2c\quad\Longrightarrow\quad\cos 2\alpha=\frac{gh}{u^2+gh}.$$

由于这是驻点（stationary point）处唯一可能的值，而 $\tan 2\alpha = x/h > 0$ 确定了 $0 < 2\alpha < \frac{\pi}{2}$，所以恰好只有一个驻点，而它就是最大值。（顺带一提，$\tan 2\alpha = \dfrac{u\sqrt{u^2+2gh}}{gh}$，所以最大距离为 $x = \dfrac{u\sqrt{u^2+2gh}}{g}$。）

与 O 的最大距离

落地点在 $(x, 0)$，所以它与 $O = (0, h)$ 的距离为

$$d = \sqrt{x^2 + h^2}.$$

在 h 固定的情况下，d 随 x 递增，所以 d 恰好在 x 最大时取得最大值。在该处 $x = h\tan 2\alpha$，所以

$$d = \sqrt{h^2\tan^2 2\alpha + h^2} = h\sec 2\alpha = h \cdot \frac{u^2+gh}{gh} = \frac{u^2}{g} + h,$$

这里用到 $\sec 2\alpha > 0$。这就是所求的最大距离。

验算

- 当 $h = 0$ 时，得 $\cos 2\alpha = 0$，所以 $\alpha = 45°$，最大射程为 u^2/g：正是在平地上我们熟悉的结果。
- 取 $u = 10$、$g = 10$、$h = 5$（国际单位制）。那么 $\cos 2\alpha = \frac{50}{150} = \frac{1}{3}$，最大射程为 $\sqrt{100\cdot 200}/10 = 10\sqrt{2} \approx 14.14$，而 $d = \sqrt{200+25} = 15 = \frac{100}{10} + 5$，与预测相符。用电脑对 α 进行搜索，亦证实 14.14 是最大的落地距离。
- 当 h 增大时，$\cos 2\alpha \to 1$，所以最佳角度趋于 0：从极高的塔上发射，应该差不多水平地射出。

常见错误

- 写出 $x = h\tan 2\alpha$ 后，忘记它只在最大值处成立，而非对所有 α 都成立。
- 除以 $1 + \cos 2\alpha$ 或 $\cos 2\alpha$ 时，没有说明它们为何不为零。
- 另行对 d 求最大值，而没有留意到 $d^2 = x^2 + h^2$ 与 x 在同一角度取得最大值。

STEP 之外

答案 $d = u^2/g + h$ 有一个优美的几何解释。所有从 O 以速率 u 射出的轨迹，都与同一条包络线（envelope）相切，这就是“安全抛物线”（parabola of safety），它的焦点（focus）

在 O，准线（directrix）位于 O 上方高度 u^2/g 处。最远的落地点位于这条包络线上，所以根据焦点—准线性质，它与 O 的距离等于它在准线下方的距离，即 $u^2/g + h$。曲线族的包络线是大学一年级微积分和力学的标准课题。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）是下一步的好选择，书中以同样严谨的方式处理抛射体问题。

详细解答

STEP I 2016，第 12 题

概率与统计 · 二项分布的对称性

STEP I 2016，第 12 题

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

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

本题在 *Beyond STEP* 第 12.6 节多一枚硬币中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/one-more-coin.html

关键思路

第 (i) 部分和第 (ii) 部分是直接运用二项概率（binomial probabilities）的计算，而两者都得出同一个简单得惊人的答案。第 (iii) 部分则解释个中原因。把 Bob 的掷币分成前 n 次和额外的一次：在前 n 次掷币中，Bob 和 Alice 处于完全对等的地位，所以对称性（symmetry）便能解决问题；而额外的一次掷币则公平地打破平局。

解答

全文中，设 A 为 Alice 掷得正面的次数，B 为 Bob 掷得正面的次数。把一枚公正硬币掷 m 次，正面次数 X 满足

$$\mathrm{P}(X = k) = \binom{m}{k}\frac{1}{2^m}, \qquad k = 0, 1, \ldots, m,$$

而 Alice 和 Bob 的掷币互相独立（independent），所以可以把 A 和 B 的概率相乘。

第 (i) 部分

Alice 掷两次，Bob 掷三次，所以

k	0	1	2	3
$\mathrm{P}(A=k)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	
$\mathrm{P}(B=k)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

以 Alice 的得分为条件（condition on）。给定 $A=a$，我们需要 $B>a$，因此需要尾概率（tail probabilities）$\mathrm{P}(B>0)=\frac{7}{8}$、$\mathrm{P}(B>1)=\frac{1}{2}$、$\mathrm{P}(B>2)=\frac{1}{8}$。故此

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

第 (ii) 部分

现在对 $a=0,1,2,3$，有 $\mathrm{P}(A=a)=\frac{1}{8},\frac{3}{8},\frac{3}{8},\frac{1}{8}$；而对 $k=0,\ldots,4$，有 $\mathrm{P}(B=k)=\frac{1}{16},\frac{4}{16},\frac{6}{16},\frac{4}{16},\frac{1}{16}$。尾概率为

$$\mathrm{P}(B>0)=\tfrac{15}{16},\quad \mathrm{P}(B>1)=\tfrac{11}{16},\quad \mathrm{P}(B>2)=\tfrac{5}{16},\quad \mathrm{P}(B>3)=\tfrac{1}{16}.$$

所以

$$\begin{aligned}\mathrm{P}(B>A) &= \tfrac{1}{8}\cdot\tfrac{15}{16}+\tfrac{3}{8}\cdot\tfrac{11}{16}+\tfrac{3}{8}\cdot\tfrac{5}{16}+\tfrac{1}{8}\cdot\tfrac{1}{16}\\ &= \frac{15+33+15+1}{128}=\frac{64}{128}=\frac{1}{2}.\end{aligned}$$

答案再一次恰好是 $\frac{1}{2}$。

第 (iii) 部分

首先，看看掷币次数相同的游戏的一个事实。当 Alice 和 Bob 各掷 n 次时，“Bob 较多”、“相同”、“Alice 较多”三者中恰有一个发生。两名玩家的处境是对称的（他们进行完全相同的试验），所以“Alice 较多”的概率也是 p_2。因此

$$2p_2+p_1=1. \qquad (*)$$

现在让 Alice 掷 n 次，Bob 掷 $n+1$ 次。设 B' 为 Bob 前 n 次掷币中正面的次数，并设 H 为他最后一次掷得正面的事件。那么若 H 发生，则 $B=B'+1$；否则 $B=B'$。

> ? **想一想** 比较 B 与 A 的结果，怎样取决于比较 B' 与 A 的结果?

由于 B' 和 A 都是整数，而 B 不是 B' 就是 $B'+1$：

- 若 $B'>A$，则无论最后一次结果如何，必然有 $B>A$；
- 若 $B'=A$，则 $B>A$ 当且仅当最后一次掷得正面；

- 若 $B' < A$，则 $B' \le A - 1$，所以 $B \le A$，Bob 不会胜出。

Bob 的前 n 次掷币对 Alice 的 n 次掷币，正正就是掷币次数相同的游戏，所以 $\mathrm{P}(B' > A) = p_2$ 及 $\mathrm{P}(B' = A) = p_1$。最后一次掷币与其他一切独立，且 $\mathrm{P}(H) = \frac{1}{2}$。故此

$$\mathrm{P}(\text{Bob 掷得的正面比 Alice 多}) = p_2 + \tfrac{1}{2}p_1.$$

由 $(*)$，$p_2 + \frac{1}{2}p_1 = \frac{1}{2}(2p_2 + p_1) = \frac{1}{2}$。

这就是所求的推广（generalisation）：*若 Alice 把一枚公正硬币掷 n 次，而 Bob 把一枚公正硬币掷 $n+1$ 次，则对每个 $n \ge 0$，Bob 掷得的正面比 Alice 多的概率都是 $\frac{1}{2}$*。第 (i) 部分和第 (ii) 部分分别是 $n = 2$ 和 $n = 3$ 的情况。

验算

- $n = 0$：Alice 不掷，Bob 掷一次；Bob 胜出当且仅当他掷得正面，概率为 $\frac{1}{2}$。
- $n = 2$：用手计算，$p_1 = \frac{1}{16} + \frac{4}{16} + \frac{1}{16} = \frac{3}{8}$，再由 $(*)$ 得 $p_2 = \frac{5}{16}$。于是 $p_2 + \frac{1}{2}p_1 = \frac{5}{16} + \frac{3}{16} = \frac{1}{2}$，与第 (i) 部分吻合。
- 用电脑精确穷举（exact enumeration）证实，对每个不超过 29 的 n，都有 $\mathrm{P}(B > A) = \frac{1}{2}$；而对第 (ii) 部分进行 200 000 次试验的模拟（simulation），得出 0.4997。

常见错误

- 在掷币次数相同的游戏中假设 $p_2 = \frac{1}{2}$：平局的概率是正数，所以 $p_2 = \frac{1}{2}(1 - p_1) < \frac{1}{2}$。
- 在第 (iii) 部分漏掉 $B' = A$ 的情况，或把它整个计入，而没有乘上额外一次掷币的因子 $\frac{1}{2}$。
- 用了尾概率 $\mathrm{P}(B \ge a)$ 而不是 $\mathrm{P}(B > a)$。

STEP 之外

这个简洁的答案来自对称性论证，再加上以额外一步为条件——这种模式贯穿整个概率论，从随机游走（random walks）到选票问题（ballot problems）都可见到。还有一个更巧妙的版本：由于 Bob 多掷一次，他未能掷得比 Alice 多的正面，当且仅当他掷得比她多的反面。把正面和反面互换，可见“正面较多”与“反面较多”的概率相等，而两者互补（complementary），所以各自的概率都是 $\frac{1}{2}$。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）一书，对这些条件化和对称性技巧有深入的发展。

详细解答

STEP I 2016，第 13 题

概率与统计 · 指数分布的顺序统计量

> **STEP I 2016，第 13 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/16-s1-q13

关键思路

对于第一封和第二封邮件的到达，要问的是它们何时尚未发生。第一封电子邮件在时间 t 之后才到达，当且仅当全部 n 封都在 t 之后到达；第二封在 t 之后才到达，当且仅当到那时至多只有一封已到达。利用独立性（independence），这两个事件都很容易计算，再把生存概率（survival probability）求导，便得到概率密度函数（density）。

解答

设 $X_1, \ldots, X_n$ 为各封邮件的到达时间。它们互相独立，而且每一个都满足

$$\mathrm{P}(X_i > t) = \int_t^\infty \lambda \mathrm{e}^{-\lambda s}\,\mathrm{d}s = \mathrm{e}^{-\lambda t}, \qquad \mathrm{P}(X_i \le t) = 1 - \mathrm{e}^{-\lambda t}$$

其中 $t \ge 0$。我们会用到标准积分 $\int_0^\infty t k \mathrm{e}^{-kt}\,\mathrm{d}t = \frac{1}{k}$（$k > 0$），这正是参数为 k 的指数分布（exponential distribution）的平均值。（用分部积分法（integration by parts）：$\left[-t\mathrm{e}^{-kt}\right]_0^\infty + \int_0^\infty \mathrm{e}^{-kt}\,\mathrm{d}t = 0 + \frac{1}{k}$。）

第 (i) 部分

第一封邮件在 t 之后才到达，当且仅当每一封邮件都在 t 之后才到达。由独立性，

$$\mathrm{P}(T > t) = \prod_{i=1}^{n} \mathrm{P}(X_i > t) = \left(\mathrm{e}^{-\lambda t}\right)^n = \mathrm{e}^{-n\lambda t}.$$

所以累积分布函数（cumulative distribution function）为 $F(t) = 1 - \mathrm{e}^{-n\lambda t}$，其中 $t \ge 0$。求导便得到密度函数

$$f(t) = F'(t) = n\lambda\,\mathrm{e}^{-n\lambda t} \qquad (t \ge 0),$$

正如所求。这是参数为 $n\lambda$ 的指数密度函数，所以取 $k = n\lambda$ 代入标准积分，得

$$\mathrm{E}(T) = \frac{1}{n\lambda}.$$

第 (ii) 部分

设 S 为第二封邮件的到达时间（因此需要 $n \geq 2$）。第二封邮件在 t 之后才到达，当且仅当到时间 t 为止至多一封邮件已到达：要么一封也没有，要么恰好一封。

? **想一想** 到时间 t 为止恰好有一封邮件已到达，有多少种可能情况?

n 封邮件中任何一封都可以是那封提早到达的。那一封已到达（概率为 $1 - \mathrm{e}^{-\lambda t}$），而其余 $n-1$ 封尚未到达（每封概率为 $\mathrm{e}^{-\lambda t}$）。由独立性，并且由于这 n 种情况互斥（mutually exclusive），

$$\mathrm{P}(S > t) = \mathrm{e}^{-n\lambda t} + n\left(1 - \mathrm{e}^{-\lambda t}\right)\mathrm{e}^{-(n-1)\lambda t}.$$

展开第二项得 $n\mathrm{e}^{-(n-1)\lambda t} - n\mathrm{e}^{-n\lambda t}$，所以

$$\mathrm{P}(S > t) = n\,\mathrm{e}^{-(n-1)\lambda t} - (n-1)\,\mathrm{e}^{-n\lambda t}.$$

简单验证一下：当 $t = 0$ 时它等于 1，而当 $t \to \infty$ 时它趋于 0。

S 的密度函数 g 是 $1 - \mathrm{P}(S > t)$ 的导数：

$$\begin{aligned} g(t) &= -\frac{\mathrm{d}}{\mathrm{d}t}\left(n\,\mathrm{e}^{-(n-1)\lambda t} - (n-1)\,\mathrm{e}^{-n\lambda t}\right) \\ &= n(n-1)\lambda\,\mathrm{e}^{-(n-1)\lambda t} - n(n-1)\lambda\,\mathrm{e}^{-n\lambda t} \\ &= n(n-1)\lambda\left(\mathrm{e}^{-(n-1)\lambda t} - \mathrm{e}^{-n\lambda t}\right) \qquad (t \geq 0). \end{aligned}$$

为求期望值（expectation），把 g 写成指数密度函数的组合：

$$g(t) = n \cdot (n-1)\lambda\,\mathrm{e}^{-(n-1)\lambda t} \;-\; (n-1) \cdot n\lambda\,\mathrm{e}^{-n\lambda t}.$$

分别取 $k = (n-1)\lambda$ 和 $k = n\lambda$ 代入标准积分（这是成立的，因为 $n \geq 2$，所以两种情况下都有 $k > 0$），

$$\begin{aligned} \mathrm{E}(S) = \int_0^\infty t\,g(t)\,\mathrm{d}t &= \frac{n}{(n-1)\lambda} - \frac{n-1}{n\lambda} \\ &= \frac{1}{\lambda} \cdot \frac{n^2 - (n-1)^2}{n(n-1)} = \frac{1}{\lambda} \cdot \frac{2n-1}{n(n-1)}. \end{aligned}$$

由于 $\dfrac{1}{n-1} + \dfrac{1}{n} = \dfrac{n + (n-1)}{n(n-1)} = \dfrac{2n-1}{n(n-1)}$，所以

$$\mathrm{E}(S) = \frac{1}{\lambda}\left(\frac{1}{n-1} + \frac{1}{n}\right),$$

这正是所求的结果。

想一想 为甚么答案是两个简单项之和?

指数分布具有无记忆性（memoryless）。第一封邮件在一段速率为 $n\lambda$ 的指数时间后到达，平均值为 $\frac{1}{n\lambda}$。之后，其余 $n-1$ 封邮件重新开始，就好像时钟刚被重设一样，所以额外的等待时间是速率为 $(n-1)\lambda$ 的指数分布，平均值为 $\frac{1}{(n-1)\lambda}$。把两段等待时间相加，立即得到答案，这也是记住它的好方法。

验算

- 当 $n=2$ 且 $\lambda=1$ 时，第二封邮件的到达时间是两个独立指数时间中较晚的一个。它的平均值应为 $1+\frac{1}{2}=1.5$，而以电脑模拟 10^5 次试验得到 1.501。
- 密度函数 g 的积分为 1：$\int_0^\infty g = n-(n-1)=1$。
- 当 $n=5$ 且 $\lambda=1$ 时，公式给出 $\mathrm{E}(S)=\frac{1}{4}+\frac{1}{5}=0.45$，与 $\int_0^\infty t\,g(t)\,\mathrm{d}t$ 的数值计算结果吻合。

常见错误

- 在恰好一封邮件已到达的情况中漏掉因子 n。n 封中任何一封都可以是提早到达的那封。
- 把 $\mathrm{P}(S>t)$ 写成恰好一封已到达的概率，这样便漏掉了一封也未到达的情况。

STEP 之外

T 和 S 是一个指数样本的首两个顺序统计量（order statistics）。上面的无记忆性论证可以推广，证明第 k 封邮件的到达时间平均值为 $\frac{1}{\lambda}\left(\frac{1}{n}+\frac{1}{n-1}+\cdots+\frac{1}{n-k+1}\right)$，而这正是泊松过程（Poisson process）和排队论（queueing theory）的基础。Blitzstein 和 Hwang 的 *Introduction to Probability*（《概率导论》）对顺序统计量和无记忆性都有清晰的讲解。

详细解答

STEP II 2016，第 1 题

纯数学 · 参数曲线与切线

> **STEP II 2016，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/16-s2-q1

关键思路

用参数（parameter）而不是坐标来处理。O 处的直角会变成两点的参数 p 和 q 之间一条非常简单的关系，而一般点上的切线（tangent）用 t 表示时有一条简洁的方程。求出两条切线的交点后，交点的坐标会是 p 和 q 的对称函数（symmetric function），而 p 与 q 之间的关系正好足以把它们消去（eliminate）。

解答

一般点上的切线

在参数为 t 的点上，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y/\mathrm{d}t}{\mathrm{d}x/\mathrm{d}t} = \frac{3t^2}{2t} = \frac{3t}{2} \qquad (t \neq 0).$$

所以 (t^2, t^3) 处的切线是 $y - t^3 = \frac{3t}{2}(x - t^2)$，即

$$2y = 3tx - t^3. \qquad (*)$$

（当 $t = 0$ 时，这给出 $y = 0$，即尖点（cusp）处的切线，但我们不需要用到该点。）

直角

设 $P = (p^2, p^3)$ 及 $Q = (q^2, q^3)$。要令 $\angle POQ$ 有意义，两点都不能是 O，所以 $p \neq 0$ 且 $q \neq 0$。该角为直角当且仅当 $\overrightarrow{OP} \cdot \overrightarrow{OQ} = 0$：

$$p^2q^2 + p^3q^3 = p^2q^2(1 + pq) = 0.$$

由于 $p^2q^2 \neq 0$，这迫使

$$pq = -1.$$

特别地，$p \neq q$（它们的积是负数），所以 P 和 Q 处的切线斜率（gradient）不同，$\frac{3p}{2} \neq \frac{3q}{2}$，确实相交于唯一一点。

切线的交点

把 $(*)$ 分别取 $t = p$ 和 $t = q$ 的两个版本相减：

$$3(p-q)x = p^3 - q^3 = (p-q)(p^2+pq+q^2).$$

除以 $p - q \neq 0$，并利用 $pq = -1$，

$$3x = p^2 + pq + q^2 = p^2 + q^2 - 1.$$

代回 $t = p$ 时的 $(*)$：

$$\begin{aligned} 2y &= p(p^2+pq+q^2) - p^3 \\ &= p^2q + pq^2 = pq(p+q) = -(p+q). \end{aligned}$$

想一想 我们已用两个参数表示 x 和 y。怎样才能得到一条只含 x 和 y 的方程?

把 y 平方：出现的交叉项 $2pq$ 是已知常数，剩下的是 $p^2 + q^2$，而这正是 $3x$ 所含的部分。事实上，

$$4y^2 = (p+q)^2 = p^2 + q^2 + 2pq = p^2 + q^2 - 2 = (3x+1) - 2 = 3x - 1.$$

所以 P 和 Q 处的切线相交于 C_2: $4y^2 = 3x - 1$ 上，得证。

（事实上 C_2 上每一点都会出现：当 p 取遍所有正实数时，$2y = -(p+q) = \frac{1}{p} - p$ 取遍所有实数值，而 x 则由 C_2 的方程确定。）

C_1 与 C_2 是否相交?

把 $x = t^2$、$y = t^3$ 代入 C_2 的方程：

$$4t^6 = 3t^2 - 1 \iff 4u^3 - 3u + 1 = 0, \quad \text{其中} u = t^2 \geq 0.$$

显然 $u = -1$ 是一个根，除去因式 $u + 1$ 得

$$4u^3 - 3u + 1 = (u+1)(4u^2 - 4u + 1) = (u+1)(2u-1)^2.$$

由于 $u \geq 0$，唯一可能是 $u = \frac{1}{2}$，所以 $t = \pm\frac{1}{\sqrt{2}}$。因此两曲线确实相交，而且恰好交于两点

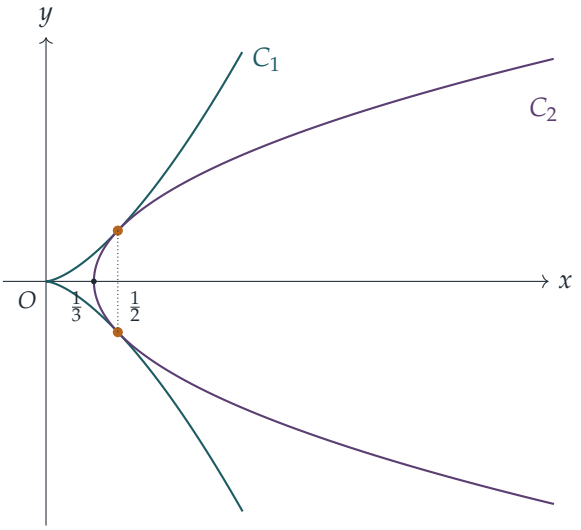
$$\left(\tfrac{1}{2}, \tfrac{1}{2\sqrt{2}}\right) \quad \text{及} \quad \left(\tfrac{1}{2}, -\tfrac{1}{2\sqrt{2}}\right).$$

重根（repeated root）告诉我们更多：两曲线在该处相切，而不是相交穿过。我们可以直接验证。在 $t=\frac{1}{\sqrt{2}}$ 处，C_1 的斜率是 $\frac{3t}{2}=\frac{3}{2\sqrt{2}}$。对 $4y^2=3x-1$ 进行隐函数求导（implicit differentiation）得 $8y\frac{\mathrm{d}y}{\mathrm{d}x}=3$，所以在 $y=\frac{1}{2\sqrt{2}}$ 处，C_2 的斜率是 $\frac{3}{8y}=\frac{3}{2\sqrt{2}}$，两者相同。由于图形关于 x 轴对称，另一点亦然。

由于对 $u\geq 0$ 有 $4u^3-3u+1=(u+1)(2u-1)^2\geq 0$，曲线 C_1 永远不会走进抛物线（parabola）内部：除了两个切点外，C_2 都位于 C_1 的两条分支之间。

草图

C_1 满足 $y^2=x^3$：它位于 $x\geq 0$ 的区域，关于 x 轴对称，并在 O 处有一个尖点（两条分支的斜率 $\frac{3t}{2}\to 0$，所以它们沿正 x 轴离开 O）。C_2 是抛物线 $x=\frac{1}{3}(4y^2+1)$，顶点（vertex）为 $(\frac{1}{3},0)$，开口向右，同样关于 x 轴对称。抛物线位于尖点之内、C_1 的两条分支之间，并在 $x=\frac{1}{2}$ 处与每条分支各相切一次。



验算

- 取 $p=1$、$q=-1$：$P=(1,1)$，$Q=(1,-1)$，且 $\overrightarrow{OP}\cdot\overrightarrow{OQ}=1-1=0$。切线为 $2y=3x-1$ 及 $2y=-3x+1$，交于 $(\frac{1}{3},0)$，即 C_2 的顶点。
- 取 $p=2$、$q=-\frac{1}{2}$：切线 $2y=6x-8$ 与 $2y=-\frac{3}{2}x+\frac{1}{8}$ 交于 $x=\frac{13}{12}$、$y=-\frac{3}{4}$，而 $4y^2=\frac{9}{4}=3\cdot\frac{13}{12}-1$。
- 切点 $(\frac{1}{2},\frac{1}{2\sqrt{2}})\approx(0.5,\,0.35355)$ 满足 $4y^2=\frac{1}{2}=3x-1$，而该处两条曲线的斜率都是 ≈ 1.06066。用电脑检验了 10^5 对满足 $pq=-1$ 的随机参数，所有交点在舍入误差范围内都落在 C_2 上。

常见错误

- 忘记排除 $p = 0$ 或 $q = 0$：在除去因式以得到 $pq = -1$ 之前，必须确定 p^2q^2 不为零。
- 除以 $p - q$ 时没有说明它为何不为零。
- 停在“$4u^3 - 3u + 1 = 0$ 有根”，而没有检查它是否给出 $u = t^2 \geq 0$；根 $u = -1$ 并不对应 C_1 上的任何点。

STEP 之外

曲线 $y^2 = x^3$ 是经典的*尖点三次曲线*（cuspidal cubic），而把一条曲线与一条参数化曲线联立、再从重根看出相切的技巧，正是代数几何（algebraic geometry）的起点：相交重数（intersection multiplicity）计算曲线相交了多少次，而贝祖（Bézout）定理指出，一条三次曲线与一条二次曲线（conic）在正确计算重数下相交于六点。这里的六点是二重根 $t = \pm\frac{1}{\sqrt{2}}$（共四个）和一对复数根 $t = \pm i$。一本浅白的入门书是 Kirwan 的 *Complex Algebraic Curves*（《复代数曲线》）。

详细解答

STEP II 2016，第 2 题

纯数学 · 因式定理

> **STEP II 2016，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/16-s2-q2

关键思路

把 $(*)$ 看成以 c 为变量的多项式（polynomial），其系数（coefficients）恰好涉及 a 和 b。因式定理（factor theorem）指出：$c-(a+b)$ 是因式（factor），当且仅当代入 $c=a+b$ 后结果为零。一旦找到一个因式，$(*)$ 对 a、b、c 的对称性（symmetry）便免费送上另外两个，再数一数次数（degree），因式分解（factorisation）就完成了。之后的两条方程其实是经过伪装的特殊情况：要做的就是选取 a、b、c（以及后来的 d、e）来配合它们。

解答

因式 $a+b-c$

记 $(*)$ 为 $F(a,b,c)$。视之为 c 的多项式，F 有因式 $c-(a+b)$（因而亦有因式 $a+b-c$），当且仅当 $F(a,b,a+b)=0$。设 $p=a+b$ 及 $c=p$。那么 $a+b+c=2p$，并利用 $a^2+b^2=p^2-2ab$ 和 $a^3+b^3=p^3-3abp$，得

$$a^2+b^2+c^2=2p^2-2ab, \qquad a^3+b^3+c^3=2p^3-3abp.$$

所以

$$\begin{aligned} F(a,b,a+b) &= 8p^3-6(2p)(2p^2-2ab)+8(2p^3-3abp) \\ &= 8p^3-24p^3+24abp+16p^3-24abp=0, \end{aligned}$$

因此 $a+b-c$ 是 $(*)$ 的因式，正如所求。

将 $(*)$ 完全因式分解

无论怎样重新排列 a、b、c，F 都不变。交换 b 与 c，因式 $a+b-c$ 变成 $a+c-b$；交换 a 与 c，它变成 $b+c-a$。所以以下三者

$$a+b-c, \qquad b+c-a, \qquad c+a-b$$

都是因式。它们是互不相同的线性（linear）多项式，没有一个是另一个的常数倍，所以它们的乘积整除 F。F 的每一项次数都是 3，所以余下的因式次数为 0：它是一个常数 k，而

$$F(a,b,c) = k(a+b-c)(b+c-a)(c+a-b).$$

为求 k，设 $a=b=c=1$：左方为 $27-6\cdot 3\cdot 3+8\cdot 3=-3$，而乘积为 1。因此 $k=-3$，且

$$\begin{aligned}(a+b+c)^3-6(a+b+c)(a^2+b^2+c^2)+8(a^3+b^3+c^3)\\ = -3(a+b-c)(b+c-a)(c+a-b).\end{aligned}$$

第 (i) 部分

把方程的中间项和最后一项除开，化成 $(*)$ 的形式：

$$(x+1)^3-6(x+1)\left(x^2+\tfrac{5}{2}\right)+8\left(x^3+\tfrac{13}{4}\right)=0.$$

想一想 a、b、c 取甚么值，才能令这式子恰好就是 $(*)$?

试取 $a=x$，而 b、c 为常数。我们需要

$$b+c=1, \qquad b^2+c^2=\tfrac{5}{2}, \qquad b^3+c^3=\tfrac{13}{4}.$$

由前两式得 $2bc=(b+c)^2-(b^2+c^2)=-\frac{3}{2}$，所以 $bc=-\frac{3}{4}$。于是

$$b^3+c^3=(b+c)^3-3bc(b+c)=1+\tfrac{9}{4}=\tfrac{13}{4},$$

所以第三个条件自动成立。现在 b、c 是 $t^2-t-\frac{3}{4}=0$ 的根，即 $(2t-3)(2t+1)=0$，所以 $b=\frac{3}{2}$，$c=-\frac{1}{2}$。

取 $a=x$、$b=\frac{3}{2}$、$c=-\frac{1}{2}$，因式分解给出

$$a+b-c=x+2, \qquad b+c-a=1-x, \qquad c+a-b=x-2,$$

所以方程为 $-3(x+2)(1-x)(x-2)=0$，而

$$x=-2, \quad x=1, \quad x=2.$$

第 (ii) 部分

记该四变量表达式为 $G(a,b,d,e)$。(设 $c=d+e$，则 $a+b-d-e$ 看起来就像 $a+b-c$，由此可猜到这个因式；不过直接验证也一样快。) 根据因式定理，把 G 视为 e 的多项式，只需证明当 $a+b=d+e$ 时 $G=0$。设 $p=a+b=d+e$。与之前完全一样，$a^2+b^2=p^2-2ab$、$d^2+e^2=p^2-2de$、$a^3+b^3=p^3-3abp$ 及 $d^3+e^3=p^3-3dep$，所以设 $q=ab+de$，得

$$\begin{aligned} G &= (2p)^3-6(2p)(2p^2-2q)+8(2p^3-3pq)\\ &= 8p^3-24p^3+24pq+16p^3-24pq=0. \end{aligned}$$

所以 $a+b-d-e$ 是 G 的因式。

无论怎样重新排列 a、b、d、e，G 都不变。因式 $a+b-d-e$ 来自把四个字母分成 $\{a,b\}$ 和 $\{d,e\}$ 两对；另外两种配对方法给出因式

$$a+d-b-e \quad (\text{交换}b,d), \qquad a+e-b-d \quad (\text{交换}b,e).$$

这三个线性因式互不相同，没有一个是另一个的倍数，而 G 是 3 次齐次 (homogeneous) 多项式，所以对某常数 k，有 $G=k(a+b-d-e)(a+d-b-e)(a+e-b-d)$。设 $a=1$、$b=d=e=0$，左方得 $1-6+8=3$，右方得 k，所以

$$G=3(a+b-d-e)(a+d-b-e)(a+e-b-d).$$

至于方程，取 $a=x$，并取常数 b、d、e 使得

$$b+d+e=6, \qquad b^2+d^2+e^2=14, \qquad b^3+d^3+e^3=36.$$

数字 1、2、3 符合要求：$1+2+3=6$，$1+4+9=14$，$1+8+27=36$。所以取 $a=x$、$b=1$、$d=2$、$e=3$，方程恰好就是 $G=0$，其因式分解为

$$3(x+1-2-3)(x+2-1-3)(x+3-1-2)=3(x-4)(x-2)x=0.$$

因此

$$x=0, \quad x=2, \quad x=4.$$

验算

- 把 $a=b=1$、$c=2$ 代入 $(*)$：$64-6\cdot4\cdot6+8\cdot10=0$，正如因式 $a+b-c$ 所预示。
- 在 (i) 中，$x=1$ 给出 $8-3\cdot2\cdot7+2\cdot17=8-42+34=0$。
- 在 (ii) 中，$x=0$ 给出 $216-6\cdot6\cdot14+8\cdot36=216-504+288=0$，而 $x=4$ 给出 $1000-6\cdot10\cdot30+8\cdot100=0$。
- 这些因式分解亦已用电脑在变量取几千个随机有理数值时验证过。

常见错误

- 忘记常数：$(*)$ 是各因式乘积的 -3 倍，而不是乘积本身。务必用方便的代入来确定 k。
- 在 (ii) 中列出多于三个因式。例如交换 a 与 d，得到 $d+b-a-e$，这不过是 $-(a+e-b-d)$；三个真正不同的因式来自把四个字母两两配对的三种方法。
- 用蛮力把所有东西展开。这样做行得通，但又长又容易出错；把方程与恒等式配对才是这道题的重点。

STEP 之外

因式定理在任何域（field）上都成立；这里我们把除一个以外的所有变量视为常数，从而把它用于多变量多项式。对称性加上多项式的唯一因式分解（unique factorisation）是威力强大的组合：对称多项式（symmetric polynomials）的理论指出，每个对称多项式都可以用基本对称多项式（elementary symmetric polynomials）$a+b+c$、$ab+bc+ca$、abc 表示，这正是 (i) 中关于 b、c 的条件背后隐藏的结构。顺带一提，乘积 $(a+b-c)(b+c-a)(c+a-b)$ 正是三角形面积的海伦公式（Heron's formula）中出现的式子。想进一步了解多项式环（polynomial rings）与因式分解，可参阅 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）。

详细解答

STEP II 2016，第 3 题

纯数学 · 多项式的根

> **STEP II 2016，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/16-s2-q3

关键思路

整道题建基于一条恒等式（identity）：$\mathrm{f}_n(x) = \mathrm{f}_{n-1}(x) + \dfrac{x^n}{n!}$。配合第 (i) 部分，它告诉我们：在 f_n 的根（root）a 处，斜率（gradient）为 $\mathrm{f}_n'(a) = -a^n/n!$。因此，即使不知道根是甚么，我们也知道每个根处的斜率。由于所有根的正负号相同，这些斜率的正负号亦全部相同；而一条曲线不可能连续两次向上穿过坐标轴，中间却没有一次向下穿过。

解答

全题中记 $\mathrm{f}_n(x) = \displaystyle\sum_{k=0}^{n} \frac{x^k}{k!}$。

第 (i) 部分

逐项微分（differentiating term by term），常数项 1 消失，而对 $1 \le k \le n$，

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x^k}{k!}\right) = \frac{kx^{k-1}}{k!} = \frac{x^{k-1}}{(k-1)!}.$$

所以

$$\mathrm{f}_n'(x) = \sum_{k=1}^{n} \frac{x^{k-1}}{(k-1)!} = \sum_{j=0}^{n-1} \frac{x^j}{j!} = \mathrm{f}_{n-1}(x),$$

得证。

第 (ii) 部分

若 $x \geq 0$，则 $\mathrm{f}_n(x)$ 的每一项都非负，而首项是 1，所以 $\mathrm{f}_n(x) \geq 1 > 0$。因此没有根能满足 $a \geq 0$，任何实根都有 $a < 0$。（当 $n = 0$ 时，$\mathrm{f}_0 = 1$ 根本没有根，与此一致。）

第 (iii) 部分

由定义，$\mathrm{f}_n(x) = \mathrm{f}_{n-1}(x) + \dfrac{x^n}{n!}$。若 $\mathrm{f}_n(a) = 0$，则由第 (i) 部分，

$$\mathrm{f}'_n(a) = \mathrm{f}_{n-1}(a) = \mathrm{f}_n(a) - \frac{a^n}{n!} = -\frac{a^n}{n!},$$

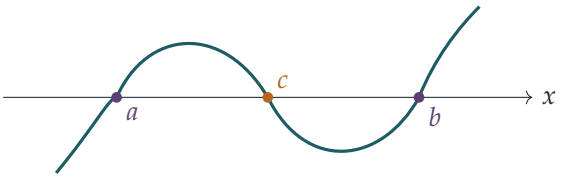
同理 $\mathrm{f}'_n(b) = -b^n/n!$。因此

$$\mathrm{f}'_n(a)\,\mathrm{f}'_n(b) = \frac{(ab)^n}{(n!)^2} > 0,$$

这是因为由第 (ii) 部分得 $a < 0$ 及 $b < 0$，所以 $ab > 0$。这正是所求的不等式。特别地，两个斜率都不为零，而且正负号相同。

> ? **想一想** 为甚么两个斜率正负号相同，就迫使两者之间必有另一个根?

设 $a < b$，并假设两个斜率皆为正（负的情况与之对称）。由于 $\mathrm{f}'_n(a) > 0$，曲线在 a 处向上穿过坐标轴，所以当 x 略大于 a 时 $\mathrm{f}_n(x) > 0$。由于 $\mathrm{f}'_n(b) > 0$，曲线在 b 处亦向上穿过坐标轴，所以当 x 略小于 b 时 $\mathrm{f}_n(x) < 0$。一个连续函数若在某点为正、在较后的一点为负，则必在两者之间取值零（介值定理（intermediate value theorem)），所以存在某个 c 满足 $a < c < b$ 且 $\mathrm{f}_n(c) = 0$。



曲线上升穿过 a，又上升穿过 b，所以在两者之间某处必定向下穿过坐标轴。

*最多只有一个实根。*假设 $(*)$ 有两个或以上相异的实根。它是 n 次多项式 (polynomial)，所以只有有限个根，我们可以从中选出两个*相邻* (consecutive) 的根 $a < b$，即两者之间严格地没有其他根。但上面的论证在 a 与 b 之间严格地找到一个根 c，矛盾。所以 $(*)$ 最多只有一个实根。（论证按题目所设使用了 $n \geq 2$；当 $n = 1$ 时唯一的根是 -1，当 $n = 0$ 时没有根，所以结论对每个 n 都成立。）

*n 为奇数。*此时 f_n 的次数为奇数，首项系数 (leading coefficient) $1/n!$ 为正，所以当 $x \to -\infty$ 时 $\mathrm{f}_n(x) \to -\infty$，当 $x \to +\infty$ 时 $\mathrm{f}_n(x) \to +\infty$。由介值定理，它最少有一个实根，因此**恰好有一个**。

n 为偶数。当 $n = 0$ 时没有根。对偶数 $n \geq 2$，当 $x \to \pm\infty$ 时 $\mathrm{f}_n(x) \to +\infty$，所以 f_n 在某点 x_0 取得全域最小值（global minimum），而该处 $\mathrm{f}_n'(x_0) = \mathrm{f}_{n-1}(x_0) = 0$。由于 $\mathrm{f}_{n-1}(0) = 1 \neq 0$，故 $x_0 \neq 0$，于是

$$\mathrm{f}_n(x_0) = \mathrm{f}_{n-1}(x_0) + \frac{x_0^n}{n!} = \frac{x_0^n}{n!} > 0,$$

因为 n 是偶数。最小值为正，所以对所有 x 都有 $\mathrm{f}_n(x) > 0$，$(*)$ **没有实根**。

（另一方法：由于 $\mathrm{f}_n'(a) \neq 0$，任何根 a 都会是单根（simple root），所以 f_n 在该处会变号；但一个首项系数为正的偶次多项式若恰好只变号一次，则它对所有很大的 x 或对所有很负的 x 必为负，这是不可能的。）

验算

- $n = 1$：$1 + x = 0$ 只有一个根 -1。$n = 2$：$1 + x + \frac{1}{2}x^2 = \frac{1}{2}\big((x+1)^2 + 1\big) > 0$，没有根。两者皆吻合。
- $n = 3$：根约为 -1.596，而该处 $\mathrm{f}_3'(a) = \mathrm{f}_2(a) \approx 0.678 = -a^3/3!$，与预测相符。
- 用电脑检查 f_n 对每个 $n \leq 25$ 的变号情况，发现 n 为奇数时恰有一个根，n 为偶数时没有根。

常见错误

- 证明斜率之积为正时，没有说明为甚么 $ab > 0$：这需要用到第 (ii) 部分。
- 由“任意两根之间必有一根”推出“最多只有一个根”，却没有选取两个相邻的根；正是这个选取才导致矛盾。
- n 为偶数时，忘记检查 $x_0 \neq 0$，而这是 $x_0^n/n! > 0$ 严格成立所必需的。

STEP 之外

多项式 f_n 是 e^x 的泰勒多项式（Taylor polynomials），而第 (i) 部分正是 $(\mathrm{e}^x)' = \mathrm{e}^x$ 在有限情况下的影子。由于 e^x 永不为零，f_n 只有很少实根也就顺理成章；它们的复数根（complex roots）按 $1/n$ 缩放后，当 $n \to \infty$ 时会聚集在一条称为塞格曲线（Szegő curve）的曲线上。这里用到的工具——介值定理，以及连续函数必有最小值——会在第一门分析（analysis）课程中严谨地证明；可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

STEP II 2016，第 4 题

纯数学 · 有理函数的值域

> **STEP II 2016，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/16-s2-q4

关键思路

去分母后，y 的定义便化为一条关于 x 的二次方程（quadratic equation），而“x 是实数”就变成“判别式（discriminant）非负”。要消去 θ，我们需要一个对所有 θ 都成立的 $y\cos\theta - \sin\theta$ 的上界（upper bound）；把它写成 $(y,-1)$ 与单位向量（unit vector）$(\cos\theta, \sin\theta)$ 的点积（dot product），便恰好得到 $\sqrt{y^2+1}$。在第 (ii) 部分，不等式链中的每一个不等式都必须取等号，而这正好把 x 和 θ 都确定下来。

解答

第 (i) 部分

首先留意分母永不为零：把 $x^2 + x\cos\theta + 1$ 看作 x 的二次式，其判别式为 $\cos^2\theta - 4 < 0$，所以它恒为正。因此可以放心地两边相乘：

$$y\left(x^2 + x\cos\theta + 1\right) = x^2 + x\sin\theta + 1 \iff (y-1)x^2 + (y\cos\theta - \sin\theta)\,x + (y-1) = 0.$$

若 $y \neq 1$，这是一条真正关于 x 的二次方程，而它有实根 x 当且仅当其判别式非负：

$$(y\cos\theta - \sin\theta)^2 - 4(y-1)^2 \geq 0.$$

若 $y = 1$，所求的不等式变为 $(\cos\theta - \sin\theta)^2 \geq 0$，这本来就成立。所以对实数 x，

$$(y\cos\theta - \sin\theta)^2 \geq 4(y-1)^2,$$

即为所求。

想一想 怎样才能消去 θ？我们需要一个比 $(y\cos\theta-\sin\theta)^2$ 大、而又不依赖 θ 的式子。

式子 $y\cos\theta-\sin\theta$ 是 $(y,-1)$ 与单位向量 $(\cos\theta,\sin\theta)$ 的纯量积（scalar product），所以其大小至多为 $|(y,-1)|=\sqrt{y^2+1}$。等价地，由柯西—施瓦茨不等式（Cauchy–Schwarz inequality），或直接由

$$(y^2+1)(\cos^2\theta+\sin^2\theta)-(y\cos\theta-\sin\theta)^2=(y\sin\theta+\cos\theta)^2\geq 0,$$

可得 $(y\cos\theta-\sin\theta)^2\leq y^2+1$。把这与第一个结果结合，

$$y^2+1\geq(y\cos\theta-\sin\theta)^2\geq 4(y-1)^2,$$

这就是第二个不等式。

展开得 $y^2+1\geq 4y^2-8y+4$，即

$$3y^2-8y+3\leq 0.$$

$3y^2-8y+3=0$ 的根为 $y=\dfrac{8\pm\sqrt{64-36}}{6}=\dfrac{4\pm\sqrt{7}}{3}$，而由于这个二次式的首项系数（leading coefficient）为正，它恰好在两根之间取非正值。因此

$$\frac{4-\sqrt{7}}{3}\leq y\leq\frac{4+\sqrt{7}}{3}.$$

第 (ii) 部分

设 $y=\frac{1}{3}(4+\sqrt{7})$，则 $y-1=\frac{1}{3}(1+\sqrt{7})>0$。于是

$$y^2+1=\frac{16+8\sqrt{7}+7}{9}+1=\frac{32+8\sqrt{7}}{9},$$
$$4(y-1)^2=\frac{4(1+2\sqrt{7}+7)}{9}=\frac{32+8\sqrt{7}}{9}.$$

两者相等，而且 $2(y-1)>0$，所以取正平方根便得 $\sqrt{y^2+1}=2(y-1)$，即为所求。（这是理所当然的：y 是 $3y^2-8y+3=0$ 的根，而这条方程正是 $y^2+1=4(y-1)^2$。）

想一想 这个等式对第 (i) 部分的不等式链说明了甚么？

不等式链 $y^2+1\geq(y\cos\theta-\sin\theta)^2\geq 4(y-1)^2$ 的两端现在相等，所以两个不等式都取等号。

判别式为零，所以关于 x 的二次方程有重根（repeated root）

$$x=-\frac{y\cos\theta-\sin\theta}{2(y-1)}.$$

关于 θ 的上界被达到，所以 $y\sin\theta+\cos\theta=0$，从而

$$\tan\theta=-\frac{1}{y}=-\frac{3}{4+\sqrt{7}}=-\frac{3(4-\sqrt{7})}{16-7}=\frac{\sqrt{7}-4}{3}.$$

（此处 $\cos\theta\neq 0$，否则 $\sin\theta=0$ 亦成立，这是不可能的。）此外 $(y\cos\theta-\sin\theta)^2=y^2+1=4(y-1)^2$，所以 $y\cos\theta-\sin\theta=\pm 2(y-1)$，因此

$$x=\mp 1.$$

具体来说，$(\cos\theta,\sin\theta)$ 与 $(y,-1)$ 平行。若两者方向相同，即 $\cos\theta>0$ 且 $\sin\theta<0$，则 $y\cos\theta-\sin\theta=+2(y-1)$，$x=-1$。若两者方向相反，即 $\cos\theta<0$ 且 $\sin\theta>0$，则 $x=1$。两种情况下都有

$$x=\pm 1,\qquad \tan\theta=\frac{\sqrt{7}-4}{3}.$$

这两种情况其实是同一回事：把 x 换成 $-x$、θ 换成 $\theta+\pi$，y 的值不变。

验算

- 当 $x=1$ 时，公式为 $y=\dfrac{2+\sin\theta}{2+\cos\theta}$。取第二象限（quadrant）中满足 $\tan\theta=\frac{1}{3}(\sqrt{7}-4)\approx -0.4514$ 的 θ，得 $\theta\approx 2.7176$，于是 $y\approx 2.21525$，与 $\frac{1}{3}(4+\sqrt{7})\approx 2.21525$ 吻合。
- 用电脑随机抽取大量实数 x 和 θ，所得 y 值都落在 0.45142 与 2.21524 之间，刚好在界限 $\frac{1}{3}(4\mp\sqrt{7})\approx 0.451416$ 和 2.215250 之内。
- 一个值得留意的巧合：$\tan\theta=-\frac{1}{3}(4-\sqrt{7})$ 正是下界的相反数，因为两个界互为倒数（reciprocal）（它们的积是 $\frac{1}{9}(16-7)=1$）。

常见错误

- 忘记 $y=1$ 的情况：此时关于 x 的方程不是二次方程，判别式的论证不能直接套用。
- 取 $\sqrt{y^2+1}=-2(y-1)$，或没有检查正负号：这里取正根之所以正确，只是因为 $y>1$。
- 只给出一个 x 值；其实 $x=1$ 和 $x=-1$ 都会出现，对应的 θ 位于不同象限。

STEP 之外

步骤 $(y\cos\theta-\sin\theta)^2\leq(y^2+1)(\cos^2\theta+\sin^2\theta)$ 是柯西－施瓦茨不等式 $|\mathbf{u}\cdot\mathbf{v}|\leq|\mathbf{u}|\,|\mathbf{v}|$ 的二维情况，等号恰好在两向量平行时成立。这个不等式在所有内积空间（inner-product

space）中都成立，是大学一年级分析（analysis）与线性代数（linear algebra）中最常用的工具之一。Strang 的 *Introduction to Linear Algebra*（线性代数导论）从几何角度发展这个不等式；Abbott 的 *Understanding Analysis*（理解分析）则展示它在分析中的应用。

详细解答

STEP II 2016，第 5 题

纯数学 · 二项式系数恒等式

STEP II 2016，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/16-s2-q5

关键思路

三条恒等式（identity）都源自同一个技巧：把一个函数用两种方式写成乘积，然后比较 x 某一个幂的系数（coefficient）。若一个和式由二项式系数（binomial coefficient）的乘积组成，而各下标之和固定，那正正就是把两个幂级数（power series）相乘后读出某一项系数所得的结果。题目约定在 $p \geq q \geq 0$ 以外 $\binom{p}{q} = 0$，正因如此，和式可以在一个整齐的 j 范围内求和，而不必费神理会哪些项存在。

解答

第 (i) 部分

当 $|x| < 1$ 时，由二项式级数（binomial series）可得 $(1-x)^{-N}$ 中 x^n 的系数为

$$\frac{(-N)(-N-1)\cdots(-N-n+1)}{n!}(-1)^n = \frac{N(N+1)\cdots(N+n-1)}{n!} = \frac{(N+n-1)!}{n!\,(N-1)!} = \binom{N+n-1}{n}.$$

各因子带来的 n 个负号，与 $(-x)^n$ 带来的 n 个负号互相抵消。因此

$$(1-x)^{-N} = \sum_{n=0}^{\infty} \binom{N+n-1}{n} x^n \qquad (|x| < 1).$$

（当 $n = 0$ 时，系数为 $\binom{N-1}{0} = 1$。由于 $N \geq 1$，这没有问题。）

现在 $(1-x)^{-1}(1-x)^{-N} = (1-x)^{-(N+1)}$。把上述展开式中的 N 换成 $N+1$，可知右边 x^n 的系数为 $\binom{N+n}{n}$。

在左边，$(1-x)^{-1} = \sum_{i\geq 0} x^i$ 的每一项系数都等于 1。$(1-x)^{-N}$ 中的 x^j 项必须与 $(1-x)^{-1}$ 中的 x^{n-j} 配对，而 j 由 0 取到 n。所以乘积中 x^n 的系数为

$$\sum_{j=0}^{n} 1 \cdot \binom{N+j-1}{j}.$$

两个幂级数在 $|x| < 1$ 时都收敛（converge），并在该处表示同一个函数，因此它们的系数相同。于是得到

$$\sum_{j=0}^{n} \binom{N+j-1}{j} = \binom{N+n}{n},$$

即为所求。

第 (ii) 部分

利用 $(1+x)^{m+n} = (1+x)^m(1+x)^n$，三者皆为多项式（polynomial）。按题目的约定，可以写成

$$(1+x)^m = \sum_{j\geq 0} \binom{m}{j} x^j, \qquad (1+x)^n = \sum_{k\geq 0} \binom{n}{k} x^k,$$

因为 $j > m$ 或 $k > n$ 的系数均为零。乘积中 x^r 的系数来自满足 $j+k=r$ 的配对，即 $k = r-j$ 且 $0 \leq j \leq r$：

$$\sum_{j=0}^{r} \binom{m}{j}\binom{n}{r-j}.$$

? **想一想** 那么 $j > m$ 或 $r-j > n$ 的项又如何？这些幂并不出现在多项式之中。

这些项含有因子 $\binom{m}{j}$（其中 $j > m$）或 $\binom{n}{r-j}$（其中 $r-j > n$），按约定它们都是零。它们没有任何贡献，同时也不对应任何实际的项的乘积，所以在 $0 \leq j \leq r$ 上求和恰好正确。$(1+x)^{m+n}$ 中 x^r 的系数为 $\binom{m+n}{r}$。相等的多项式有相同的系数，因此

$$\binom{m+n}{r} = \sum_{j=0}^{r} \binom{m}{j}\binom{n}{r-j}.$$

第 (iii) 部分

因子 $(-1)^j\binom{m+j-1}{j}$ 提示我们用第 (i) 部分，并以 $-x$ 代替 x。在第 (i) 部分的展开式中以 $-x$ 代替 x，并以 m 代替 N，可得

$$(1+x)^{-m} = \sum_{j=0}^{\infty} (-1)^j \binom{m+j-1}{j} x^j \qquad (|x| < 1).$$

? **想一想** 应该用哪一个乘积？

另一个因子 $\binom{N+m}{n-j}$ 是 $(1+x)^{N+m}$ 的系数，而答案 $\binom{N}{n}$ 是 $(1+x)^N$ 的系数。由于 m 是正整数，当 $|x|<1$ 时有

$$(1+x)^{N+m}(1+x)^{-m}=(1+x)^N.$$

在左边，$(1+x)^{N+m}=\sum_{k\geq 0}\binom{N+m}{k}x^k$（对 $k>N+m$ 使用上述约定）。x^n 的系数来自 $(1+x)^{-m}$ 中的 x^j 乘以 $(1+x)^{N+m}$ 中的 x^{n-j}，其中 $0\leq j\leq n$：

$$\sum_{j=0}^{n}\binom{N+m}{n-j}(-1)^j\binom{m+j-1}{j}.$$

在右边，$(1+x)^N$ 中 x^n 的系数为 $\binom{N}{n}$。当 $n>N$ 时，按约定它等于 0，这也是正确的，因为 $(1+x)^N$ 的次数是 N。两边在 $|x|<1$ 上相等，所以它们的幂级数系数相同：

$$\sum_{j=0}^{n}(-1)^j\binom{N+m}{n-j}\binom{m+j-1}{j}=\binom{N}{n}.$$

验算

- 第 (i) 部分，取 $N=2$、$n=3$：$1+2+3+4=10=\binom{5}{3}$。
- 第 (ii) 部分，取 $m=n=2$、$r=2$：$1\cdot 1+2\cdot 2+1\cdot 1=6=\binom{4}{2}$。
- 第 (iii) 部分，取 $m=2$、$N=3$、$n=2$：$\binom{5}{2}\binom{1}{0}-\binom{5}{1}\binom{2}{1}+\binom{5}{0}\binom{3}{2}=10-10+3=3=\binom{3}{2}$。当 $n>N$ 时，这个正负交替的和应该等于 0。例如取 $m=1$、$N=1$、$n=2$，得 $\binom{2}{2}-\binom{2}{1}+\binom{2}{0}=1-2+1=0$。

三条恒等式亦已用电脑对所有 $m,N\leq 9$ 及 $n\leq 24$ 验证过。

常见错误

- 在 $(1-x)^{-N}$ 的系数中弄乱了正负号。那里有两组 $(-1)^n$，两者互相抵消。
- 在第 (iii) 部分，忘记 $(1+x)^{-m}$ 的级数是无穷级数，只在 $|x|<1$ 时成立。比较系数之所以成立，是因为在某区间上相等的两个幂级数，其系数相同。

STEP 之外

这就是*生成函数*（generating functions）的方法：把一个数列编码为某幂级数的系数，于是级数相乘便化为这里所见的卷积（convolution）和。第 (ii) 部分是范德蒙恒等式（Vandermonde's identity），而第 (i) 部分常被称为曲棍球棒恒等式（hockey-stick identity）。

两者都有漂亮的组合计数证明（counting proof），值得自己找找看。Wilf 的 *generating-functionology*（可在网上免费阅读），以及 Graham、高德纳（Knuth）与 Patashnik 合著的 *Concrete Mathematics*（《具体数学》），都把这套工具发展得更深入。

详细解答

STEP II 2016，第 6 题

纯数学 · 切比雪夫多项式方程

> **STEP II 2016，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/16-s2-q6

关键思路

每一部分都只需把一个候选函数代入微分方程（differential equation），并检查它在 $x = 1$ 处的条件；其余的就交给唯一性（uniqueness）假设。第 (iii) 及 (iv) 部分的窍门是：永远不要展开 y_n（我们并不知道它的明确表达式），而是直接利用方程 $(1-x^2)y_n'^2 = n^2(1-y_n^2)$ 本身来消去导数（derivative）。

解答

在整个解答中，y_k 满足

$$(1-x^2)\big(y_k'(x)\big)^2 = k^2\big(1-y_k(x)^2\big), \qquad y_k(1) = 1,$$

这不过是 $(*)$ 移项后的形式。要证明某函数等于 y_k，（根据所假设的唯一性）只需证明它在该 k 值下满足 $(*)$，并且在 $x = 1$ 处取值 1。

第 (i) 部分

当 $k = 1$ 时，方程为

$$(1-x^2)\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 + y^2 = 1.$$

对于 $y = x$，我们有 $y' = 1$，而左方为 $(1-x^2) + x^2 = 1$，符合要求。此外 $y(1) = 1$，所以 $y_1(x) = x$。

第 (ii) 部分

当 $k = 2$ 时，方程为

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

对于 $y = 2x^2 - 1$，我们有 $y' = 4x$，所以

$$\begin{aligned}(1 - x^2)(4x)^2 + 4(2x^2 - 1)^2 &= 16x^2 - 16x^4 + 16x^4 - 16x^2 + 4\\ &= 4.\end{aligned}$$

由于 $y(1) = 2 - 1 = 1$，我们得到 $y_2(x) = 2x^2 - 1$。

第 (iii) 部分

为简便起见，记 $y = y_n(x)$。以链式法则（chain rule）对 $z = 2y^2 - 1$ 求导，得 $z' = 4yy'$，所以

$$(1 - x^2)z'^2 = 16y^2\,(1 - x^2)y'^2 = 16y^2 \cdot n^2(1 - y^2),$$

这里用到了 y_n 所满足的方程。此外

$$4n^2z^2 = 4n^2(2y^2 - 1)^2 = 4n^2(4y^4 - 4y^2 + 1).$$

两式相加，

$$(1 - x^2)z'^2 + 4n^2z^2 = 16n^2y^2 - 16n^2y^4 + 16n^2y^4 - 16n^2y^2 + 4n^2 = 4n^2,$$

这正是所求的方程。

? **想一想** 这条方程与 y_{2n} 有甚么关系?

由于 $(2n)^2 = 4n^2$，它恰好就是 $k = 2n$ 时的 $(*)$。而且 $z(1) = 2y_n(1)^2 - 1 = 2 - 1 = 1$。所以 z 是 $k = 2n$ 时 $(*)$ 的一个解，并且在 $x = 1$ 处取值 1；根据唯一性，

$$y_{2n}(x) = 2\big(y_n(x)\big)^2 - 1.$$

作为验证，$n = 1$ 给出 $y_2 = 2x^2 - 1$，与 (ii) 一致；而 $n = 2$ 给出 $y_4 = 2(2x^2 - 1)^2 - 1 = 8x^4 - 8x^2 + 1$。

第 (iv) 部分

记 $u = y_m(x)$，于是 $v = y_n(u)$。根据链式法则，

$$v'(x) = y_n'(u)\,y_m'(x).$$

因此，先用 y_m 的方程（在点 x），再用 y_n 的方程（在点 u），得

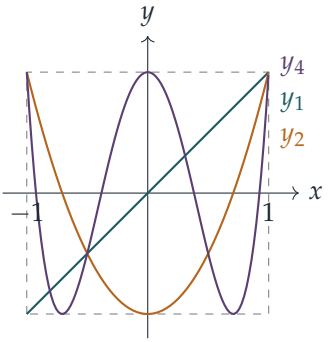
$$\begin{aligned}(1-x^2)v'^2 &= y_n'(u)^2\,(1-x^2)y_m'(x)^2\\ &= y_n'(u)^2\cdot m^2(1-u^2)\\ &= m^2\,(1-u^2)y_n'(u)^2\\ &= m^2\cdot n^2\big(1-y_n(u)^2\big) = m^2n^2(1-v^2).\end{aligned}$$

所以 $(1-x^2)v'^2+(mn)^2v^2=(mn)^2$，即 $k=mn$ 时的 $(*)$。最后

$$v(1)=y_n\big(y_m(1)\big)=y_n(1)=1,$$

所以根据唯一性，$v(x)=y_{mn}(x)$，得证。

y_n 的方程是在点 u 而非点 x 应用的：该方程对其变量的每一个值都成立，所以我们可以用 u 代替 x。正是这一步令 (iv) 得以成立。



三条曲线都通过 $(1,1)$，并且在 $[-1,1]$ 上都介乎 -1 与 1 之间，这是方程所迫使的：在该区间上 $(1-x^2)y'^2\geq 0$，所以 $k^2y^2\leq k^2$。

验算

- 第 (iii) 部分取 $n=2$，预测 $y_4(x)=8x^4-8x^2+1$。直接代入，$y_4'=32x^3-16x$，而 $(1-x^2)(32x^3-16x)^2+16(8x^4-8x^2+1)^2$ 化简后等于 16。在 $x=0.3$ 处，$y_4=0.3448$。
- 第 (iv) 部分取 $m=n=2$：$y_2(y_2(x))=2(2x^2-1)^2-1$，与 (iii) 得出的 y_4 相同，理应如此。
- 以电脑用精确算术验证过：以这种方法构造的多项式对每个不超过 24 的 k 都满足 $(*)$，而且对 $m,n\leq 4$ 有 $y_n(y_m(x))=y_{mn}(x)$。

常见错误

- 忘记 $x = 1$ 处的条件。单单满足 $(*)$ 并不足够：$-y_k$ 同样满足它。
- 在 (iii) 或 (iv) 中试图明确展开 y_n。我们对 y_n 并没有任何明确的认识；应利用它的微分方程来取代 $(1 - x^2)y_n'^2$。
- 在 (iv) 中，把 y_n 的方程用在 x 上，而不是 $u = y_m(x)$ 上，结果令因子 $(1 - x^2)$ 出现在错误的位置。

STEP 之外

函数 y_k 就是切比雪夫多项式（Chebyshev polynomials）T_k：设 $x = \cos\theta$，$(*)$ 便化为 $(\mathrm{d}y/\mathrm{d}\theta)^2 = k^2(1 - y^2)$，其解为 $y = \cos k\theta$，因此 $T_k(\cos\theta) = \cos k\theta$。于是第 (iii) 部分就是二倍角公式，而第 (iv) 部分说的是 $\cos\big(n(m\theta)\big) = \cos(mn\theta)$。不过要提醒一句：题目容许你假设的唯一性其实并不完全成立，因为常数函数 $y = 1$ 同样满足 $(*)$ 及 $y(1) = 1$。像这样导数只能确定至正负号的方程，正是使用唯一性定理时须格外小心的地方；Boyce 与 DiPrima 合著的 *Elementary Differential Equations*（《初等微分方程》）讨论了解的存在性与唯一性。

详细解答

STEP II 2016，第 7 题

纯数学 · 利用对称性求定积分

> **STEP II 2016，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/16-s2-q7

关键思路

把区间 $[0, a]$ 关于其中点作反射（reflection），即 $x \mapsto a - x$，积分的值不会改变，但被积函数（integrand）却可能变成一个能与原式漂亮地结合的形式。对于 $[0, \frac{1}{2}\pi]$ 上的三角被积函数，这个反射会把 sin 与 cos 互换；在 $[0, \frac{1}{4}\pi]$ 上，则靠 $\tan(\frac{1}{4}\pi - x)$ 的复角公式（compound-angle formula）来完成工作。最后一部分看似全新，但一次分部积分（integration by parts）就能把它化为第 (iii) 部分。

解答

结果 $(*)$

在 $\int_0^a f(a-x)\,\mathrm{d}x$ 中作代换（substitution）$u = a - x$，于是 $\mathrm{d}u = -\mathrm{d}x$；当 $x = 0$ 时 $u = a$，当 $x = a$ 时 $u = 0$。因此

$$\int_0^a f(a-x)\,\mathrm{d}x = \int_a^0 f(u)\,(-\mathrm{d}u) = \int_0^a f(u)\,\mathrm{d}u = \int_0^a f(x)\,\mathrm{d}x,$$

因为积分变量用甚么名称并不重要。这就是 $(*)$。

第 (i) 部分

设该积分为 I。取 $a = \frac{1}{2}\pi$，有 $\sin(\frac{1}{2}\pi - x) = \cos x$ 及 $\cos(\frac{1}{2}\pi - x) = \sin x$，所以由 $(*)$ 得

$$I = \int_0^{\frac{1}{2}\pi} \frac{\cos x}{\sin x + \cos x}\,\mathrm{d}x.$$

把它与 I 的原来形式相加，

$$2I = \int_0^{\frac{1}{2}\pi} \frac{\sin x + \cos x}{\cos x + \sin x}\,\mathrm{d}x = \int_0^{\frac{1}{2}\pi} 1\,\mathrm{d}x = \frac{\pi}{2},$$

所以

$$I = \frac{\pi}{4}.$$

（分母 $\cos x + \sin x$ 在 $[0, \frac{1}{2}\pi]$ 上为正，因此不会出问题。）

第 (ii) 部分

现在 $a = \frac{1}{4}\pi$，第 (i) 部分的技巧不再能互换 sin 与 cos。改为把分子分母同除以 $\cos x$（在此区间上为正），被积函数变成

$$\frac{\sin x}{\cos x + \sin x} = \frac{\tan x}{1 + \tan x}.$$

设 $t = \tan x$。由于 $\tan(\frac{1}{4}\pi - x) = \dfrac{1-t}{1+t}$，

$$\frac{\tan(\frac{1}{4}\pi - x)}{1 + \tan(\frac{1}{4}\pi - x)} = \frac{(1-t)/(1+t)}{2/(1+t)} = \frac{1-t}{2} = \frac{1}{2} - \frac{1}{2}\tan x.$$

于是设该积分为 J，由 $(*)$ 得

$$\begin{aligned} J &= \int_0^{\frac{1}{4}\pi} \Big(\frac{1}{2} - \frac{1}{2}\tan x\Big)\mathrm{d}x = \frac{\pi}{8} - \frac{1}{2}\Big[-\ln\cos x\Big]_0^{\frac{1}{4}\pi} \\ &= \frac{\pi}{8} + \frac{1}{2}\ln\frac{1}{\sqrt{2}} = \frac{\pi}{8} - \frac{1}{4}\ln 2. \end{aligned}$$

想一想 有没有方法完全不用 $(*)$ 就能做第 (ii) 部分?

有。设 $K = \int_0^{\frac{1}{4}\pi} \frac{\cos x}{\cos x + \sin x}\,\mathrm{d}x$。那么 $J + K = \frac{1}{4}\pi$；又因为分子 $\cos x - \sin x$ 正是分母的导数，

$$K - J = \Big[\ln(\cos x + \sin x)\Big]_0^{\frac{1}{4}\pi} = \ln\sqrt{2} = \tfrac{1}{2}\ln 2.$$

两式相减，得 $2J = \frac{1}{4}\pi - \frac{1}{2}\ln 2$，与上面的答案一致。

第 (iii) 部分

设该积分为 L。同一条复角公式给出

$$1 + \tan(\tfrac{1}{4}\pi - x) = 1 + \frac{1 - \tan x}{1 + \tan x} = \frac{2}{1 + \tan x}.$$

所以由 $(*)$，

$$L=\int_0^{\frac{1}{4}\pi}\ln\frac{2}{1+\tan x}\,\mathrm{d}x=\int_0^{\frac{1}{4}\pi}\big(\ln 2-\ln(1+\tan x)\big)\,\mathrm{d}x=\frac{\pi}{4}\ln 2-L.$$

因此 $2L=\frac{1}{4}\pi\ln 2$，而

$$L=\frac{\pi}{8}\ln 2.$$

第 (iv) 部分

把分子分母同除以 $\cos^2 x$：

$$\frac{x}{\cos x\,(\cos x+\sin x)}=\frac{x\sec^2 x}{1+\tan x}.$$

? **想一想** $\dfrac{\sec^2 x}{1+\tan x}$ 跟第 (iii) 部分有甚么关系?

它恰好是 $\ln(1+\tan x)$ 的导数，因为 $\frac{\mathrm{d}}{\mathrm{d}x}(1+\tan x)=\sec^2 x$；而且在 $[0,\frac{1}{4}\pi]$ 上 $1+\tan x>0$。于是取 $u=x$ 及 $\mathrm{d}v=\dfrac{\sec^2 x}{1+\tan x}\,\mathrm{d}x$、$v=\ln(1+\tan x)$ 作分部积分：

$$\begin{aligned}\int_0^{\frac{1}{4}\pi}\frac{x\sec^2 x}{1+\tan x}\,\mathrm{d}x&=\Big[x\ln(1+\tan x)\Big]_0^{\frac{1}{4}\pi}-\int_0^{\frac{1}{4}\pi}\ln(1+\tan x)\,\mathrm{d}x\\&=\frac{\pi}{4}\ln 2-\frac{\pi}{8}\ln 2\qquad\text{（利用第 (iii) 部分）}\\&=\frac{\pi}{8}\ln 2.\end{aligned}$$

所以第 (iv) 部分的积分与第 (iii) 部分的积分数值相同。

验算

- 第 (i) 部分：被积函数由 0 升至 1，并关于点 $(\frac{1}{4}\pi,\frac{1}{2})$ 对称，所以其平均值为 $\frac{1}{2}$，积分等于区间长度的一半，即 $\frac{1}{4}\pi$。
- 第 (ii) 部分应小于第 (i) 部分的一半，因为被积函数在 $[0,\frac{1}{4}\pi]$ 上低于 $\frac{1}{2}$：的确 $\frac{1}{8}\pi-\frac{1}{4}\ln 2\approx 0.2194<0.3927$。
- 数值积分（numerical quadrature）（辛普森法则（Simpson's rule））给出四个部分的值分别为 0.785398、0.219412、0.272198 及 0.272198，与 $\frac{1}{4}\pi$、$\frac{1}{8}\pi-\frac{1}{4}\ln 2$ 及 $\frac{1}{8}\pi\ln 2$（两次）吻合。

常见错误

- 忘记代换 $u=a-x$ 会把积分上下限对调；而 $\mathrm{d}u=-\mathrm{d}x$ 中的负号会把它们调回来。

- 在第 (ii) 部分仿照第 (i) 部分试图互换 sin 与 cos：在 $[0, \frac{1}{4}\pi]$ 上，反射是 $x \mapsto \frac{1}{4}\pi - x$，而不是 $\frac{1}{2}\pi - x$。
- 写成 $\int \tan x \, \mathrm{d}x = \ln \cos x$；正确应为 $-\ln \cos x$（或 $\ln \sec x$）。

STEP 之外

第 (iii) 部分是经典题：同一个反射技巧可以求出像 $\int_0^{\frac{1}{2}\pi} \ln \sin x \, \mathrm{d}x = -\frac{1}{2}\pi \ln 2$ 这类积分，而这些积分无法通过初等函数（elementary functions）的原函数（antiderivative）求得。到了大学，这类对称性论证与围道积分（contour integration）及积分号下求导（differentiation under the integral sign）并列，都是处理没有初等原函数的定积分（definite integral）的工具。想看更多富 STEP 风格的例子，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《数学进阶难题》）；至于为代换法与分部积分提供严格依据的分析（analysis），可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

STEP II 2016，第 8 题

纯数学 • 以积分近似求和

STEP II 2016，第 8 题

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

step.maths.org/questions/16-s2-q8

关键思路

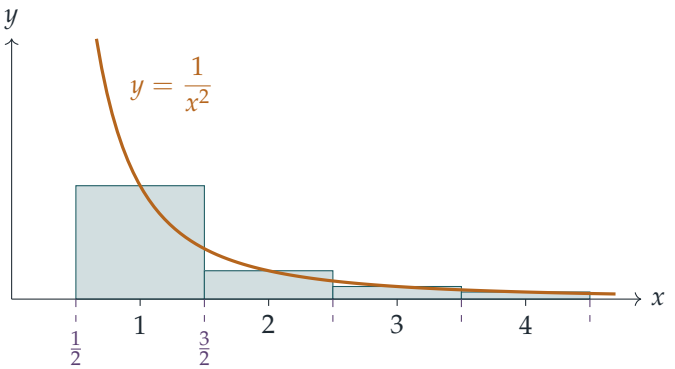
项 $\frac{1}{r^2}$ 是一个宽 1、高 $\frac{1}{r^2}$ 的长方形的面积。若这个长方形以 $x = r$ 为中心，它的面积就非常接近 $y = \frac{1}{x^2}$ 在 $r - \frac{1}{2}$ 与 $r + \frac{1}{2}$ 之间曲线下的面积，这就是中点法则（midpoint rule）。把这些窄条拼接起来，一个和式就变成一个积分。接着可以准确算出每一窄条的误差（error），而把这些误差加起来，几乎不费吹灰之力就得到第二个级数（series）$\sum \frac{1}{r^4}$。

解答

积分与草图

当 $m > \frac{1}{2}$ 时，积分下限为正，所以被积函数（integrand）在积分范围内连续，并且

$$\int_{m-\frac{1}{2}}^{\infty} \frac{1}{x^2}\,\mathrm{d}x = \lim_{X\to\infty}\left[-\frac{1}{x}\right]_{m-\frac{1}{2}}^{X} = \frac{1}{m-\frac{1}{2}} = \frac{2}{2m-1}.$$



以 $x = r$ 为中心的长方形宽 1、高 $\frac{1}{r^2}$，所以面积是 $\frac{1}{r^2}$。在窄条 $r - \frac{1}{2} \le x \le r + \frac{1}{2}$ 上，曲线在左半边位于长方形顶边之上，在右半边则位于其下，而这两小块几乎互相抵消。因此

$$\frac{1}{r^2} \approx \int_{r-\frac{1}{2}}^{r+\frac{1}{2}} \frac{1}{x^2}\,\mathrm{d}x.$$

对 $r = m, m+1, \ldots, n$ 把这些式子相加，各窄条首尾相接，刚好覆盖 $[m - \frac{1}{2}, n + \frac{1}{2}]$，于是得到 $(*)$。

第 (i) 部分

在 $(*)$ 中令 $n \to \infty$。右边趋向我们刚才算出的积分，所以

$$\sum_{r=m}^{\infty} \frac{1}{r^2} \approx \frac{2}{2m-1}.$$

我们可以把这个近似用于整个级数，也可以先准确地加上头几项，只对尾部（tail）作近似：

$$\begin{aligned}
m = 1: &\quad E \approx \frac{2}{1} = 2,\\
m = 2: &\quad E = 1 + \sum_{r=2}^{\infty} \frac{1}{r^2} \approx 1 + \frac{2}{3} = \frac{5}{3},\\
m = 3: &\quad E = 1 + \frac{1}{4} + \sum_{r=3}^{\infty} \frac{1}{r^2} \approx \frac{5}{4} + \frac{2}{5} = \frac{33}{20}.
\end{aligned}$$

这就是题目要求的三个近似值。随着 m 增大，近似应该愈来愈好，因为曲线在愈远处愈平坦，中点近似在那里也愈准确。

第 (ii) 部分

在一个窄条上的积分是

$$\int_{r-\frac{1}{2}}^{r+\frac{1}{2}} \frac{1}{x^2}\,\mathrm{d}x = \frac{1}{r-\frac{1}{2}} - \frac{1}{r+\frac{1}{2}} = \frac{1}{r^2 - \frac{1}{4}} = \frac{4}{4r^2-1}.$$

所以用这个积分近似 $\frac{1}{r^2}$ 的误差是

$$\frac{4}{4r^2-1} - \frac{1}{r^2} = \frac{4r^2 - (4r^2-1)}{r^2(4r^2-1)} = \frac{1}{r^2(4r^2-1)} = \frac{1}{4r^4} \cdot \frac{1}{1 - \frac{1}{4r^2}}.$$

当 r 很大时，$\frac{1}{4r^2}$ 很小，所以最后一个因子接近 1，误差约为 $\frac{1}{4r^4}$，正如题目所要求。（积分值总是较大，这与图像吻合：$\frac{1}{x^2}$ 是凸函数（convex），所以每个窄条左边多出的面积胜过右边不足的面积。）

至于最后的结果，把由 $r=m$ 起的误差全部加起来。各窄条的积分加起来正是我们最先算出的积分，所以

$$\frac{2}{2m-1}-\sum_{r=m}^{\infty}\frac{1}{r^2}=\sum_{r=m}^{\infty}\frac{1}{r^2(4r^2-1)}\approx\frac{1}{4}\sum_{r=m}^{\infty}\frac{1}{r^4}.$$

想一想 为甚么不取看来最简单的 $m=1$?

取 $m=1$ 时，左边是 $2-1.645=0.355$，得出 $\sum\frac{1}{r^4}\approx 1.42$，这大得太多了。问题在于“$\approx\frac{1}{4r^4}$”只在 r 很大时才准确：在 $r=1$ 时，真正的误差是 $\frac{1}{3}$ 而非 $\frac{1}{4}$，而这第一项主导了整个和。所以我们应该把头几项准确保留，只对尾部使用近似。取 $m=3$（即得出 $\frac{33}{20}$ 的那个选择），最差的因子 $\frac{1}{1-1/(4r^2)}$ 是 $\frac{36}{35}$，与 1 相差不足 3%：

$$\frac{1}{4}\sum_{r=3}^{\infty}\frac{1}{r^4}\approx\frac{2}{5}-\left(E-1-\frac{1}{4}\right)\approx 0.4-0.395=0.005,$$

所以 $\sum_{r=3}^{\infty}\frac{1}{r^4}\approx 0.02$，而

$$\sum_{r=1}^{\infty}\frac{1}{r^4}\approx 1+\frac{1}{16}+0.02=1.0825\approx 1.08.$$

这就是所要求的结果。由于 E 只给到三位小数，E 中最多 0.0005 的误差在这里会变成最多 0.002，结果仍然是 1.08。

验算

- 近似值 2、$\frac{5}{3}\approx 1.667$、$\frac{33}{20}=1.65$ 逐步递减，趋向 $E=\frac{\pi^2}{6}\approx 1.6449$，而且每一个都偏大，正如凸性论证所预测。
- 在 $r=10$ 时，准确误差 $\frac{1}{100\times 399}\approx 2.506\times 10^{-5}$ 非常接近 $\frac{1}{4\times 10^4}=2.5\times 10^{-5}$。
- 准确值是 $\sum\frac{1}{r^4}=\frac{\pi^4}{90}\approx 1.0823$，与我们的 1.0825 吻合。若改用 $m=2$，得 $1+4\left(\frac{2}{3}-0.645\right)\approx 1.087$，明显较差；$m=1$ 则得 1.42。

常见错误

- 在最后一部分使用 $m=1$，而大 r 近似 $\frac{1}{4r^4}$ 在那里并不准确，结果得出 1.42。
- 把误差的正负号弄反：积分大于 $\frac{1}{r^2}$，所以 $\frac{2}{2m-1}$ 是尾部的高估值。
- 忘记把准确计算的头几项（$1+\frac{1}{4}$ 及 $1+\frac{1}{16}$）加回去。

STEP 之外

这就是中点法则，而误差项 $\frac{1}{4r^4}$ 让我们窥见欧拉–麦克劳林公式（Euler–Maclaurin formula）：它以一个积分加上一连串导数项来修正一个和式，而像 $\sum \frac{1}{r^2}$ 这样的和，正是这样计算到许多位小数的。准确值 $\frac{\pi^2}{6}$ 和 $\frac{\pi^4}{90}$ 来自傅里叶级数（Fourier series），或来自欧拉（Euler）对 $\sin x$ 的乘积展开。Abbott 的 *Understanding Analysis*（《理解分析》）仔细地处理级数与积分，而 Siklos 的 *Advanced Problems in Mathematics*（《高等数学问题》）则有更多同类精神的题目。

详细解答

STEP II 2016，第 9 题

力学 · 动量与能量损失

STEP II 2016，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：

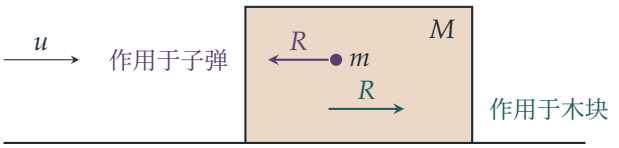
step.maths.org/questions/16-s2-q9

关键思路

阻力 R 是子弹与木块之间的内力（internal force）：它令子弹减速，而根据牛顿第三定律（Newton's third law），它同时以相同大小把木块向前推。因此两者组成的系统动量守恒（conservation of momentum），但动能（kinetic energy）并不守恒。损失的能量恰好等于 R 沿子弹在木块内移动的相对距离所做的功。对木块单独应用功能原理（work–energy principle），便可求出木块移动的距离。

解答

我们把子弹和木块视为沿同一条水平直线运动的质点，并假设子弹相对木块运动期间阻力 R 保持不变。取子弹的初始运动方向为正方向。



铅垂方向的力（重力及桌面的法向反作用力）互相平衡，不起任何作用。在水平方向上，子弹受向后的力 R，木块则受向前的力 R。

第 (i) 部分

木块固定时，子弹的减速度恒为 R/m。利用 $v^2 = u^2 + 2\alpha s$，代入末速 0 及 $\alpha = -R/m$：

$$0 = u^2 - \frac{2R}{m}a \quad \Longrightarrow \quad \boxed{a = \frac{mu^2}{2R}.}$$

（等价地说，克服阻力所做的功 Ra 等于子弹的初始动能 $\frac{1}{2}mu^2$。）

第 (ii) 部分

共同末速度。由于桌面光滑，子弹和木块组成的系统不受任何水平外力。当子弹相对木块静止时，两者以相同速度 V 运动，由动量守恒得

$$mu = (M+m)V \quad \Longrightarrow \quad V = \frac{mu}{M+m}.$$

> **想一想** 阻力在子弹于木块内的路径上作用了距离 b，但子弹本身相对桌面移动了 $b+c$。能量方程中应该用哪一个距离？

求 b。相对桌面而言，作用于子弹的力 R 所做的功为 $-R(b+c)$，因为子弹移动了 $b+c$；作用于木块的力 R 所做的功为 $+Rc$，因为木块移动了 c。因此对整个系统所做的净功为 $-Rb$：只有相对位移才算数。由功能原理，

$$\begin{aligned} Rb &= \tfrac{1}{2}mu^2 - \tfrac{1}{2}(M+m)V^2 \\ &= \tfrac{1}{2}mu^2 - \frac{m^2u^2}{2(M+m)} \\ &= \tfrac{1}{2}mu^2\left(1 - \frac{m}{M+m}\right) = \tfrac{1}{2}mu^2 \cdot \frac{M}{M+m}. \end{aligned}$$

由于第 (i) 部分得 $\frac{1}{2}mu^2 = Ra$，两边除以 R 得

$$\boxed{b = \frac{M}{M+m}\,a.}$$

求 c。现在对木块单独应用功能原理。木块由静止开始，最终速率为 V，而它所受的唯一水平力是向前的 R，作用距离为 c：

$$Rc = \tfrac{1}{2}MV^2 = \frac{Mm^2u^2}{2(M+m)^2} = \tfrac{1}{2}mu^2 \cdot \frac{mM}{(M+m)^2}.$$

再次利用 $\frac{1}{2}mu^2 = Ra$，

$$\boxed{c = \frac{mM}{(M+m)^2}\,a.}$$

另一方法：运动学。子弹以 R/m 减速，木块以 R/M 加速。当 $u - Rt/m = Rt/M$ 时两者速度相等，所以

$$t = \frac{mMu}{R(M+m)}, \qquad c = \frac{1}{2}\cdot\frac{R}{M}\,t^2 = \frac{m^2Mu^2}{2R(M+m)^2},$$

与上面的结果一致；而 b 则等于子弹的位移减去 c。

验算

- 当 $M \to \infty$ 时，木块的表现犹如固定不动：$b \to a$ 且 $c \to 0$，正合所料。
- 由于 $\frac{M}{M+m} < 1$，故 $b < a$：可移动的木块把较少的子弹能量转化为热，所以子弹穿入得没有那么深。
- 取 $M = 4m$，得 $b = \frac{4}{5}a$ 及 $c = \frac{4}{25}a$。以细小的时间步长逐步数值模拟运动，并随机选取多组 m、M、u 和 R，所得的 a、b 和 c 均准确至四位有效数字。

常见错误

- 在能量损失方程中用了子弹相对桌面的距离 $b + c$，而非相对距离 b。
- 假设碰撞中动能守恒；事实并非如此，损失的 Rb 就是所产生的热。
- 忘记木块同样受到力 R 的作用（牛顿第三定律），以致无法解释木块的运动。

STEP 之外

内力只有在没有相对运动时才做零净功，这个事实正是严谨处理质点系能量的起点。在质心参考系（centre-of-mass frame）中处理，问题就变得简洁：可供耗散的动能为 $\frac{1}{2}\mu u^2$，其中约化质量（reduced mass）$\mu = mM/(M+m)$，而这恰好就是本题中的 Rb。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对这些概念有详尽的阐述。

详细解答

STEP II 2016，第 10 题

力学 · 摩擦力与平衡

STEP II 2016，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/16-s2-q10

关键思路

作用在三角形上的只有两个力：作用于质心（centre of mass）G 的重量，以及钉在接触点 P 给予的单一接触力（contact force）。两个力处于平衡（equilibrium）时，必须大小相等、方向相反，并且作用在同一直线上，所以三角形会摆动，直至 G 位于 P 的正下方。于是问题变成：钉的反作用力要偏离 BC 的法线多少，而摩擦力（friction）是否足以承受？因此第一步是找出 G 的位置。

解答

模型与质心

把金属线视为三根单位长度质量相同的均匀杆（uniform rod），并把钉视为单一接触点，提供一个法向反作用力（normal reaction）R（垂直于 BC）和一个摩擦力 F（沿 BC），且 $F \le \mu R$。

设 $AB = AC = b$，并设 M 为 BC 的中点。由于 A 处的角是 2θ，直线 AM 平分此角并垂直于 BC，所以

$$BM = MC = b\sin\theta, \qquad AM = b\cos\theta, \qquad BC = 2b\sin\theta.$$

由对称性，G 位于 AM 上。从 BC 向 A 的方向量度距离。杆 AB 和 AC 的质量各与 b 成正比，其中心位于各自的中点，与 BC 的距离为 $\frac{1}{2}b\cos\theta$；杆 BC 的质量与 $2b\sin\theta$ 成正比，其中心位于 BC 上。所以距离 $d = GM$ 为

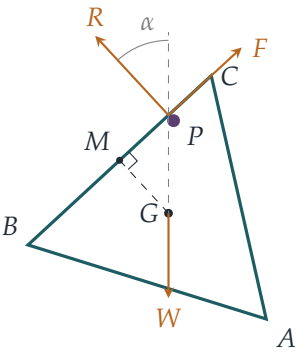
$$d = \frac{b\cdot\frac{1}{2}b\cos\theta + b\cdot\frac{1}{2}b\cos\theta + 2b\sin\theta\cdot 0}{b + b + 2b\sin\theta} = \frac{b\cos\theta}{2(1+\sin\theta)}.$$

钉位于一般位置时的平衡

假设钉在 P 点接触 BC，其中 $MP = x$，且 $0 \le x \le b\sin\theta$（由对称性，可取 P 位于 M 与 C 之间）。三角形悬挂在钉的下方，所以钉位于三角形内部，而 BC 搁在钉上。

> ? **想一想** 三角形可以以任何角度悬挂。我们怎样知道它会取哪一个角度?

作用在三角形上的力是作用于 G 的重量 W，以及钉在 P 的总接触力。只有两个力时，平衡要求它们大小相等、方向相反（所以接触力是铅直向上的 W），并且作用在同一直线上（否则它们会构成一个力偶（couple））。所以 PG 必须是铅直的：三角形会稳定在 G 位于 P 正下方的位置。



设 α 为 PG（铅直线）与 BC 的法线 GM 之间的夹角。在直角三角形 PMG 中，

$$\tan\alpha = \frac{MP}{GM} = \frac{x}{d}.$$

BC 的法线与铅直线成 α 角，而 BC 与水平线成 α 角。沿垂直于 BC 及沿 BC 的方向分解（resolve）：

$$R = W\cos\alpha, \qquad F = W\sin\alpha.$$

由于 $0 \le \alpha < \frac{1}{2}\pi$，两者皆非负，且 $R > 0$。只要 $F \le \mu R$，三角形便能保持不动而不滑动，即

$$\mu \ge \tan\alpha = \frac{x}{d}.$$

BC 上的每一点

右方的 x/d 随 x 递增，所以当 P 位于端点，即 $x = b\sin\theta$ 时，条件最难满足。当且仅当在该处可以平衡时，在 BC 上每一点都可以平衡：

$$\mu \ge \frac{b\sin\theta}{d} = b\sin\theta \cdot \frac{2(1+\sin\theta)}{b\cos\theta} = 2\tan\theta\,(1+\sin\theta),$$

这就是所求的条件。若此条件成立，则对 $[0, b\sin\theta]$ 中每一个 x 都有 $\mu \ge x/d$，所以无论钉位于 BC 上哪一点，三角形都可以静止。

验算

- 当 $x = 0$（钉位于 M）时，得 $\alpha = 0$：三角形悬挂时 BC 保持水平，完全不需要摩擦力，正如对称性所预期。
- 当 $\theta = 30^\circ$ 时，三角形是等边三角形，$d = \frac{1}{3} b\cos 30^\circ$，而 G 应为形心（centroid），即从 BC 沿高 $b\cos 30^\circ$ 向上三分之一处。结果确实如此。条件变成 $\mu \geq 2 \cdot \frac{1}{\sqrt{3}} \cdot \frac{3}{2} = \sqrt{3} \approx 1.732$：在角位时，$BC$ 须与水平线成 60°。
- 当 $\theta \to 0$ 时，三角形变成一根细长的尖刺，BC 缩小至零，界限趋于 0；当 $\theta \to \frac{1}{2}\pi$ 时，三角形变成一条又长又扁的窄条，界限趋于无穷大。两者都符合直觉。
- 以电脑验算：把三根杆分割成许多小段来求 G，并对随机的 θ 和 b 在 BC 上 201 个点分解力，结果与 $2\tan\theta(1+\sin\theta)$ 的差在 10^{-11} 以内。

常见错误

- 用了三角形薄板（lamina）的形心（AM 向上三分之一处），而不是金属线的质心。金属线框的质量分布在三边上，所以 BC 的权重是 $2b\sin\theta$。
- 假设 BC 保持水平，然后尝试取力矩。三角形会倾斜，直至 G 位于钉的下方；摩擦力要应付的正是这个倾斜角。
- 只检查 BC 上的一点。最坏的情况在 BC 的端点，而正是这一点决定了界限。

STEP 之外

写出 $\tan\alpha \leq \mu$，即表示总接触力必须位于摩擦锥（cone of friction）之内，这是一个以法线为轴、半顶角为 $\arctan\mu$ 的圆锥。这个图像令许多摩擦问题（梯子、楔子、粗糙的钉）几乎变成几何问题，并且可以推广至有多个接触点的刚体（rigid body）：问题变成能否在每个锥内选取某些力，使之与负载平衡。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）便以这种严谨的态度处理摩擦与刚体平衡。

详细解答

STEP II 2016，第 11 题

力学 • 抛体的碰撞

STEP II 2016，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/16-s2-q11

关键思路

两个质点碰撞（collide），即是在同一时刻两个坐标都相等，因此我们得到两条共用同一个未知数 t 的方程；把其中一条除以另一条，便可消去 t。第 (ii) 部分的窍门在于留意到：在每一瞬间，重力把子弹和靶子往下拉的距离完全相同，都是 $\frac{1}{2}gt^2$。所以两者能否相遇，取决于第 (i) 部分那种没有重力的运动；重力只在问“相遇时靶子在哪里”时才有关系。

解答

第 (i) 部分

把第一个质点记作 P（从 $(a,0)$ 出发），第二个记作 Q（从 $(0,b)$ 出发）。设两者在时刻 t 碰撞。令 x 坐标和 y 坐标分别相等，

$$\begin{aligned} a + ut\cos\alpha &= vt\cos\beta, & \text{所以} \quad t(v\cos\beta - u\cos\alpha) &= a, \\ ut\sin\alpha &= b + vt\sin\beta, & \text{所以} \quad t(u\sin\alpha - v\sin\beta) &= b. \end{aligned}$$

由于 $a > 0$，第一条方程显示 $t \neq 0$（两个质点从不同的点出发），因此 $t > 0$，我们可以把第二条方程除以第一条：

$$\frac{u\sin\alpha - v\sin\beta}{v\cos\beta - u\cos\alpha} = \frac{b}{a} = \tan\theta = \frac{\sin\theta}{\cos\theta}.$$

（分母等于 $a/t \neq 0$。）交叉相乘，

$$\cos\theta\,(u\sin\alpha - v\sin\beta) = \sin\theta\,(v\cos\beta - u\cos\alpha),$$

再把含 u 的项移到左边、含 v 的项移到右边，

$$u(\sin\alpha\cos\theta + \cos\alpha\sin\theta) = v(\sin\beta\cos\theta + \cos\beta\sin\theta).$$

由 $\sin(A+B)$ 的复角公式（compound-angle formula），上式即是

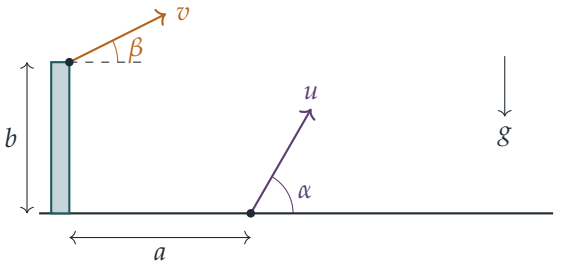
$$u\sin(\theta+\alpha) = v\sin(\theta+\beta),$$

此即所求。我们另外记下两个稍后会用到的事实：由于 $a, b, t > 0$，

$$v\cos\beta > u\cos\alpha \qquad 及 \qquad u\sin\alpha > v\sin\beta. \tag{$*$}$$

第 (ii) 部分

把子弹和靶子视为只受重力作用、自由运动的质点，不计空气阻力（air resistance）。取塔底为原点，x 轴为通过靶子发射点的水平线（方向为远离塔的一方），y 轴铅直向上。子弹必须朝靶子的方向发射才有可能命中，所以两者都在这个铅直平面内运动。



在时刻 t（两者落地之前），靶子和子弹的位置分别是

$$T = \left(a + ut\cos\alpha,\ ut\sin\alpha - \tfrac{1}{2}gt^2\right), \qquad B = \left(vt\cos\beta,\ b + vt\sin\beta - \tfrac{1}{2}gt^2\right).$$

这恰好是第 (i) 部分中的位置，各自向下平移了相同的距离 $\frac{1}{2}gt^2$。

想一想 第 (i) 部分没有重力，为甚么我们仍可以用第 (i) 部分的结果?

命中发生在 $T = B$ 的时候。在 y 方程中，$-\frac{1}{2}gt^2$ 这一项同时出现在两边，可以互相抵消，所以碰撞条件与第 (i) 部分一字不差：

$$t\left(v\cos\beta - u\cos\alpha\right) = a, \qquad t\left(u\sin\alpha - v\sin\beta\right) = b.$$

特别地，第 (i) 部分及 $(*)$ 都适用，而命中的时刻是

$$t = \frac{b}{u\sin\alpha - v\sin\beta},$$

其中由 $(*)$ 知分母为正。

此刻靶子仍在空中，当且仅当它的高度为正：

$$ut\sin\alpha - \tfrac{1}{2}gt^2 > 0 \iff t < \frac{2u\sin\alpha}{g},$$

这里除以了 $\frac{1}{2}t > 0$。代入命中的时刻，

$$\frac{b}{u\sin\alpha - v\sin\beta} < \frac{2u\sin\alpha}{g}.$$

两个分母都是正数，所以两边同乘 $g(u\sin\alpha - v\sin\beta)$ 不会改变不等式（inequality）的方向，于是得到

$$2u\sin\alpha\,(u\sin\alpha - v\sin\beta) > bg,$$

此即所求。（命中时子弹与靶子处于同一高度，所以子弹也在地面之上。）

为甚么需要 $\alpha > \beta$

我们已经看到，要命中，两个质点必须按第 (i) 部分的意义碰撞，所以

$$u\sin(\theta+\alpha) = v\sin(\theta+\beta),$$

其中 $\tan\theta = b/a$；而且由 $(*)$，还有 $u\sin\alpha > v\sin\beta$。这里四个正弦值都是正数，因为 θ、α、β 都是锐角，所以 $\theta+\alpha$ 和 $\theta+\beta$ 严格介乎 0 与 π 之间。把等式除以这个不等式，不等号方向会倒转：

$$\frac{\sin(\theta+\alpha)}{\sin\alpha} < \frac{\sin(\theta+\beta)}{\sin\beta}.$$

展开得 $\dfrac{\sin(\theta+\alpha)}{\sin\alpha} = \sin\theta\cot\alpha + \cos\theta$，$\beta$ 的情况同理，所以

$$\sin\theta\cot\alpha < \sin\theta\cot\beta \iff \cot\alpha < \cot\beta,$$

因为 $\sin\theta > 0$。在 $(0, \frac{\pi}{2})$ 上余切函数（cotangent）严格递减（strictly decreasing），所以 $\alpha > \beta$。

从物理角度看：相对于子弹，靶子以恒定速度沿直线运动，因为重力已经抵消了。这个相对速度（relative velocity）必须由靶子的起点 $(a,0)$ 指向位于 $(0,b)$ 的枪，所以它必须指向后方及上方；由 $(*)$，靶子在水平方向必须比子弹慢，在铅直方向却比子弹快，这迫使它的路径较陡，即 $\tan\alpha > \tan\beta$。（把 $(*)$ 的两部分相除，也可以直接得到同一结论。）

验算

试取 $u = 20$、$\alpha = 60°$、$v = 15$、$\beta = 30°$，并设在 $t = 1$ 时命中，$g = 9.8$。碰撞方程给出 $a \approx 2.990$ 及 $b \approx 9.821$，所以 $\theta \approx 73.06°$。于是 $u\sin(\theta+\alpha) \approx 14.612$，而 $v\sin(\theta+\beta) \approx 14.612$，两者吻合。命中时靶子的高度为 $17.32 - 4.9 \approx 12.42 > 0$，而确实 $2u\sin\alpha(u\sin\alpha - v\sin\beta) \approx 340.2 > bg \approx 96.2$。我们用电脑检验了数以千计的随机发射情况，证实该不等式成立当且仅当命中时靶子仍在空中，而且每一次命中都有 $\alpha > \beta$。

常见错误

- 忘记两个质点必须在同一时刻处于同一位置，结果解出两个不同的时刻。
- 未先确定 $u\sin\alpha - v\sin\beta$ 为正，便把两边同乘以它；它确实为正，因为 $t > 0$ 且 $b > 0$。
- 单凭 $u\sin\alpha > v\sin\beta$ 推出 $\alpha > \beta$；除非同时比较速率，否则这条不等式对角度毫无启示。

STEP 之外

重力互相抵消，是伽利略相对性（Galilean relativity）的一个例子：在随子弹一同自由下落的参考系中，靶子以恒定速度沿直线运动，所以“直接瞄准它”就行。这正是经典的“猴子与猎人”（monkey and hunter）示范实验；而惯性参考系（inertial frame）（或自由下落参考系）的概念，是大学一年级力学的起点，日后亦是广义相对论（general relativity）中等效原理（equivalence principle）的起点。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对相对运动和参考系有细致的阐述。

详细解答

STEP II 2016，第 12 题

概率与统计 · 容斥原理与错排

STEP II 2016，第 12 题

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

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

本题在 *Beyond STEP* 第 9.8 节没有人拿到自己的恤衫中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/derangements.html

关键思路

对于三个事件（event），把 $A \cup B$ 看作单一事件，然后两次运用两个事件的公式。同样的正负交替规律（加上单个事件、减去两两之交、加上三个之交，...）对任意数目的事件都成立。把这规律应用于“第 i 张卡在第 i 个位置”这些事件时，每个由 r 个事件组成的项都有相同的概率，所以和式的每一层都会化简成一个分数。至于“恰好一个”，先固定哪一张卡是配对的那张，再要求其余的卡完全没有配对。

解答

三个事件

写成 $A \cup B \cup C = (A \cup B) \cup C$，然后以 $A \cup B$ 代替 A、以 C 代替 B，运用两个事件的公式：

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

第一项是 $\mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$。至于最后一项，由分配律（distributive law）得 $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$，再运用一次两个事件的公式，

$$\begin{aligned}\mathrm{P}\big((A \cap C) \cup (B \cap C)\big) &= \mathrm{P}(A \cap C) + \mathrm{P}(B \cap C) - \mathrm{P}\big((A \cap C) \cap (B \cap C)\big) \\ &= \mathrm{P}(A \cap C) + \mathrm{P}(B \cap C) - \mathrm{P}(A \cap B \cap C).\end{aligned}$$

把两者代入，

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

得证。

四个事件

加上四个单一事件的概率，减去六个两两交集（intersection）的概率，加上四个三重交集的概率，再减去四重交集的概率：

$$\begin{aligned}\mathrm{P}(A\cup B\cup C\cup D)=\;&\mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)+\mathrm{P}(D)\\&-\mathrm{P}(A\cap B)-\mathrm{P}(A\cap C)-\mathrm{P}(A\cap D)\\&-\mathrm{P}(B\cap C)-\mathrm{P}(B\cap D)-\mathrm{P}(C\cap D)\\&+\mathrm{P}(A\cap B\cap C)+\mathrm{P}(A\cap B\cap D)+\mathrm{P}(A\cap C\cap D)+\mathrm{P}(B\cap C\cap D)\\&-\mathrm{P}(A\cap B\cap C\cap D).\end{aligned}$$

第 (i)–(iii) 部分

全部 $n!$ 种排列的可能性相等。若要求某 r 张指定的卡各自在自己的位置上，余下的 $n-r$ 张卡可以有 $(n-r)!$ 种排法，所以概率是 $(n-r)!/n!$。因此

$$\text{(i) }\mathrm{P}(E_i)=\frac{1}{n},\qquad \text{(ii) }\mathrm{P}(E_i\cap E_j)=\frac{1}{n(n-1)},\qquad \text{(iii) }\mathrm{P}(E_i\cap E_j\cap E_k)=\frac{1}{n(n-1)(n-2)}.$$

（第 (ii) 部分亦可这样看：第 i 张卡落在第 i 个位置的概率是 $\frac{1}{n}$，然后第 j 张卡在余下的 $n-1$ 个位置中落在第 j 个位置的概率是 $\frac{1}{n-1}$。）

至少一个配对

至少有一张卡配对，正正就是 $E_1\cup E_2\cup\cdots\cup E_n$ 发生。上述规律可推广至 n 个事件：第 r 层是 $(-1)^{r+1}$ 乘以所有由 r 个不同事件组成的交集的概率之和。

> **想一想** 在不知道是哪些事件的情况下，怎样计算一个遍及所有 r 个事件选法的和？

由对称性（symmetry），每个由 r 个不同事件组成的交集都有相同的概率 $(n-r)!/n!$，而选出 r 个事件的方法有 $\binom{n}{r}$ 种。所以第 r 层贡献

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

对 $r=1,2,\ldots,n$ 求和，

$$\mathrm{P}(\text{至少一个配对})=1-\frac{1}{2!}+\frac{1}{3!}-\cdots+(-1)^{n+1}\frac{1}{n!},$$

得证。（当 $r = 1, 2, 3$ 时，这不过是第 (i)–(iii) 部分得出的 $n \cdot \frac{1}{n}$、$\binom{n}{2} \cdot \frac{1}{n(n-1)}$ 和 $\binom{n}{3} \cdot \frac{1}{n(n-1)(n-2)}$。）

恰好一个配对

以 p_m 表示 m 张洗乱了的卡中没有任何一张在自己位置上的概率。由上一个结果，

$$p_m = 1 - \left(1 - \frac{1}{2!} + \cdots + (-1)^{m+1}\frac{1}{m!}\right) = \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^m \frac{1}{m!}.$$

“恰好一个配对”这事件是各事件“第 i 张卡配对而其他卡都不配对”对 i 的不相交并集（disjoint union）。在 E_i 发生的条件下，其余 $n-1$ 张卡在其余 $n-1$ 个位置上的任何排列都有相等的可能性，所以这等同于重新洗一副 $n-1$ 张的卡，而其中没有任何一张配对的概率是 p_{n-1}。因此

$$\mathrm{P}(\text{恰好一个配对}) = \sum_{i=1}^{n} \mathrm{P}(E_i)\, p_{n-1} = n \cdot \frac{1}{n} \cdot p_{n-1} = p_{n-1},$$

即是

$$\mathrm{P}(\text{恰好一个配对}) = \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-1}\frac{1}{(n-1)!} \qquad (n \geq 2),$$

而当 $n = 1$ 时概率为 1。等价地，它等于 $\sum_{r=0}^{n-1}(-1)^r/r!$，这个写法连 $n = 1$ 的情况也包括在内。

验算

当 $n = 3$ 时，排列有 123、132、213、231、312、321。其中四个有配对，得 $\frac{4}{6} = \frac{2}{3} = 1 - \frac{1}{2} + \frac{1}{6}$；而三个（即对换 132、213、321）恰好有一个配对，得 $\frac{1}{2}$，与公式相符。当 $n = 2$ 时，恰好一个配对是不可能的（若一张卡在其位置上，另一张也必然如此），而公式给出 0。以电脑列举 $n \leq 7$ 的所有排列，结果与两条公式都相符；例如当 $n = 7$ 时，恰好一个配对的概率是 $\frac{53}{144} \approx 0.368$。当 $n \to \infty$ 时，两个答案都趋向熟悉的极限：至少一个配对趋向 $1 - \mathrm{e}^{-1} \approx 0.632$，而恰好一个配对趋向 e^{-1}。

常见错误

- 写成 $\mathrm{P}(E_i \cap E_j) = \frac{1}{n^2}$：这些事件并不独立（independent），因为一旦第 i 张卡在第 i 个位置，第 j 张卡便只剩下 $n-1$ 个位置。
- 把“恰好一个”答成 $n \cdot \frac{1}{n} = 1$ 或 $\mathrm{P}(E_1)$：我们需要第 i 张卡在其位置上，*而且*没有其他卡在其位置上。
- 忘记“至少一个”的和式中 $r = 1$ 的项是 $n \cdot \frac{1}{n} = 1$，这正是级数以 1 开始的原因。

STEP 之外

这是经典的错排（derangement）问题（即 *problème des rencontres*, “配对问题”），而一般的容斥（inclusion–exclusion）公式可以用归纳法（induction）或指示随机变量（indicator random variables）简洁地证明。原来当 n 很大时，配对的数目近似服从平均值为 1 的泊松分布（Poisson distribution）：恰好有 k 个配对的概率趋向 $\mathrm{e}^{-1}/k!$。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》），以及 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》），都有讨论容斥原理和配对问题。

详细解答

STEP II 2016，第 13 题

概率与统计 • 正态近似

STEP II 2016，第 13 题

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

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

本题在 *Beyond STEP* 第 12.23 节由概率得出斯特灵公式中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/stirling-probability.html

关键思路

离散分布（discrete distribution）的单一概率 $\mathrm{P}(X = k)$，是一条以 k 为中心、宽度为 1 的长条的面积。若该分布接近正态分布（normal distribution），这个面积大约等于 k 处的正态概率密度（density）乘以宽度 1。在平均值（mean）处应用这一点（标准正态分布在该处的密度为 $\frac{1}{\sqrt{2\pi}}$），便能把一个已知的精确概率化为一个近似式。只要巧妙地选取分布，就能令精确概率包含我们想要的阶乘（factorial）。

解答

全题通用的近似

设 W 取整数值，平均值为 μ，方差（variance）为 σ^2，并且可用 $V \sim \mathrm{N}(\mu, \sigma^2)$ 很好地近似。V 的概率密度为

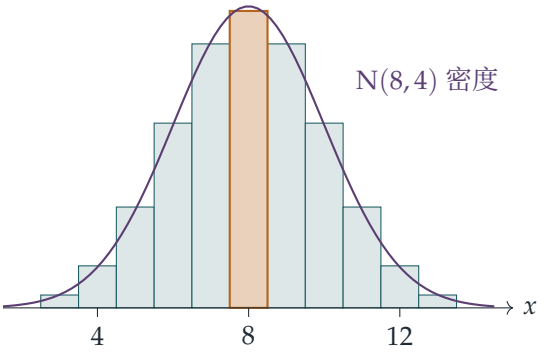
$$f(v) = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2}.$$

加上连续性校正（continuity correction），

$$\mathrm{P}(W = k) \approx \mathrm{P}\left(k - \tfrac{1}{2} < V < k + \tfrac{1}{2}\right) = \int_{k-1/2}^{k+1/2} f(v)\,\mathrm{d}v \approx 1 \times f(k),$$

最后一步成立，是因为区间宽度为 1，而当 σ 不太小时，f 在区间上几乎不变。在 $k=\mu$ 处，指数部分为 $\mathrm{e}^0=1$，所以

$$\mathrm{P}(W=\mu)\approx\frac{1}{\sigma\sqrt{2\pi}}. \qquad (*)$$



第 (i) 部分

对于 $X\sim\mathrm{B}(16,\frac{1}{2})$，

$$\mathrm{E}(X)=np=8,\qquad \mathrm{Var}(X)=np(1-p)=16\cdot\tfrac{1}{2}\cdot\tfrac{1}{2}=4,$$

所以 $\sigma=2$。由于 $p=\frac{1}{2}$，分布关于 8 对称，而且 n 相当大，因此 $\mathrm{N}(8,4)$ 是很好的近似（图中把各长条与正态曲线作了比较）。由 $(*)$，阴影长条的面积为

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

即为所求。

第 (ii) 部分

设 $Y\sim\mathrm{B}(2n,\frac{1}{2})$。则 $\mathrm{E}(Y)=n$，$\mathrm{Var}(Y)=2n\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{n}{2}$，所以 $\sigma=\sqrt{n/2}$。精确地，

$$\mathrm{P}(Y=n)=\binom{2n}{n}\left(\tfrac{1}{2}\right)^n\left(\tfrac{1}{2}\right)^n=\frac{(2n)!}{(n!)^2\,2^{2n}}.$$

当 n 很大时，正态近似适用，由 $(*)$ 得

$$\mathrm{P}(Y=n)\approx\frac{1}{\sqrt{n/2}\,\sqrt{2\pi}}=\frac{1}{\sqrt{n\pi}}.$$

令两个表达式相等，再乘以 $2^{2n}(n!)^2$：

$$(2n)!\approx\frac{2^{2n}(n!)^2}{\sqrt{n\pi}},$$

这正是所求的结果。（第 (i) 部分就是 $n=8$ 的情况：$\sqrt{8\pi}=2\sqrt{2\pi}$。）

第 (iii) 部分

设 $Z \sim \mathrm{Po}(n)$，于是 $\mathrm{E}(Z) = \mathrm{Var}(Z) = n$，$\sigma = \sqrt{n}$。精确地，

$$\mathrm{P}(Z = n) = \frac{\mathrm{e}^{-n} n^n}{n!}.$$

想一想 为甚么这里可以用正态分布来近似泊松分布（Poisson distribution）？

一个 $\mathrm{Po}(n)$ 随机变量与 n 个独立的 $\mathrm{Po}(1)$ 随机变量之和同分布，所以由中心极限定理（central limit theorem），当 n 很大时，它近似服从 $\mathrm{N}(n, n)$。这就是“当 λ 很大时 $\mathrm{Po}(\lambda) \approx \mathrm{N}(\lambda, \lambda)$”这条标准法则。

因此由 $(*)$，

$$\frac{\mathrm{e}^{-n} n^n}{n!} \approx \frac{1}{\sqrt{n}\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi n}},$$

整理后得

$$n! \approx \sqrt{2\pi n}\, \mathrm{e}^{-n}\, n^n,$$

即为所求。

验算

- 第 (i) 部分：精确值为 $\mathrm{P}(X = 8) = \binom{16}{8}/2^{16} = 12870/65536 \approx 0.1964$，而 $\frac{1}{2\sqrt{2\pi}} \approx 0.1995$，相差约 1.6% 以内。
- 第 (ii) 部分取 $n = 10$：$\binom{20}{10}/2^{20} \approx 0.1762$，对比 $1/\sqrt{10\pi} \approx 0.1784$。随着 n 增大，阶乘近似式两边之比趋向 1（在 $n = 100$ 时约为 0.9988）。
- 第 (iii) 部分取 $n = 10$：$10! = 3\,628\,800$，而 $\sqrt{20\pi}\,\mathrm{e}^{-10}\,10^{10} \approx 3\,598\,696$，偏小不足 1%。
- 一致性：把 (iii) 的结果同时代入 (ii) 中的 $(2n)!$ 和 $n!$，两边都得到 $\sqrt{4\pi n}\,\mathrm{e}^{-2n}(2n)^{2n}$，所以两个近似式互相吻合。

常见错误

- 直接使用标准正态密度 $\frac{1}{\sqrt{2\pi}}$ 而没有除以 σ。$\mathrm{N}(\mu, \sigma^2)$ 在平均值处的密度是 $\frac{1}{\sigma\sqrt{2\pi}}$。
- 在第 (ii) 部分中把标准差（standard deviation）误用为 $\frac{n}{2}$ 而不是 $\sqrt{n/2}$，或把 $\mathrm{B}(2n, \frac{1}{2})$ 的方差算错。
- 在第 (iii) 部分中忘记说明为甚么泊松分布可以用正态分布近似。

STEP 之外

第 (iii) 部分的结果就是斯特林公式（Stirling's formula），而上述论证是中心极限定理在单一点上应用的一种启发式（heuristic）形式，即所谓“局部极限定理”（local limit theorem）。要把它严格化需要小心：知道一个分布函数接近正态分布函数，本身并不能推出个别概率接近密度。斯特林公式也可以直接证明，例如把 $\ln n! = \sum \ln k$ 与 $\int \ln x \, \mathrm{d}x$ 作比较，或从伽玛函数（gamma function）出发，运用拉普拉斯（Laplace）方法。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）讲解了中心极限定理及泊松分布与正态分布之间的联系；Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率入门》）是很好的辅助读物。

详细解答

STEP III 2016，第 1 题

纯数学 · 归约公式

> **STEP III 2016，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/16-s3-q1

关键思路

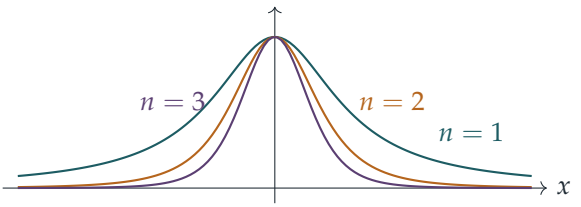
配方（completing the square）把二次式化成 $(x+a)^2+c^2$，其中 $c^2=b-a^2>0$，之后用正切代换（tangent substitution）便可立即求出 I_1。至于递推公式（reduction formula），窍门在于对 1 进行分部积分（integration by parts）：对 $(x^2+2ax+b)^{-n}$ 求导会把幂次提升到 $n+1$，这正是我们把 I_n 与 I_{n+1} 联系起来所需要的。其后的归纳法（induction）只不过是整理二项式系数（binomial coefficient）而已。

解答

全文记

$$Q(x)=x^2+2ax+b=(x+a)^2+c^2, \qquad c=\sqrt{b-a^2}>0.$$

由于对所有 x 都有 $Q(x)\geq c^2>0$，被积函数（integrand）是连续的；又由于 $Q(x)^n$ 的增长速度与 x^{2n} 相若，且 $2n\geq 2$，所以每个 I_n 都收敛（converge）。



当 $a=0$、$b=1$ 时的被积函数 $1/Q^n$：随着 n 增大，曲线变得愈来愈窄，因此我们预期当 $c=1$ 时 I_n 会逐渐变小。

第 (i) 部分

设 $x+a=c\tan u$，则 $\mathrm{d}x=c\sec^2 u\,\mathrm{d}u$。当 x 由 $-\infty$ 变到 ∞ 时，u 由 $-\frac{\pi}{2}$ 变到 $\frac{\pi}{2}$，而且

$$Q(x)=c^2\tan^2 u+c^2=c^2\sec^2 u.$$

因此

$$I_1=\int_{-\pi/2}^{\pi/2}\frac{c\sec^2 u}{c^2\sec^2 u}\,\mathrm{d}u=\frac{1}{c}\int_{-\pi/2}^{\pi/2}\mathrm{d}u=\frac{\pi}{c}=\frac{\pi}{\sqrt{b-a^2}},$$

即为所求。

第 (ii) 部分

对 I_n 进行分部积分，取 $u=Q^{-n}$ 及 $\mathrm{d}v=\mathrm{d}x$。取 $v=x+a$ 比取 $v=x$ 更巧妙（任何原函数（antiderivative）皆可），因为 $Q'(x)=2(x+a)$。于是

$$\frac{\mathrm{d}}{\mathrm{d}x}Q^{-n}=-2n(x+a)\,Q^{-n-1},$$

所以

$$I_n=\Big[\frac{x+a}{Q^n}\Big]_{-\infty}^{\infty}+2n\int_{-\infty}^{\infty}\frac{(x+a)^2}{Q^{n+1}}\,\mathrm{d}x.$$

? **想一想** 为甚么边界项（boundary term）会等于零?

分子的增长速度与 $|x|$ 相若，而分母的增长速度与 x^{2n} 相若，又因为 $n\geq 1$，所以 $2n\geq 2$。因此当 $x\to\pm\infty$ 时 $(x+a)/Q^n\to 0$，方括号内的项为零。

接着利用 $(x+a)^2=Q-c^2$ 把余下的积分拆开：

$$\begin{aligned}I_n&=2n\int_{-\infty}^{\infty}\frac{Q-c^2}{Q^{n+1}}\,\mathrm{d}x\\&=2n\,I_n-2nc^2\,I_{n+1}.\end{aligned}$$

移项得 $2nc^2I_{n+1}=(2n-1)I_n$，即

$$2n(b-a^2)\,I_{n+1}=(2n-1)\,I_n,$$

即为所求。（由于 I_n 与 I_{n+1} 均收敛，这样拆开积分是合理的。）

第 (iii) 部分

设 $P(n)$ 为命题

$$I_n=\frac{\pi}{2^{2n-2}\,c^{2n-1}}\binom{2n-2}{n-1},$$

其中 $c^{2n-1}=(b-a^2)^{n-\frac{1}{2}}$。

基本情况。当 $n=1$ 时，右方为 $\dfrac{\pi}{2^0 c^1}\binom{0}{0}=\dfrac{\pi}{c}$，由第 (i) 部分可知它等于 I_1。所以 $P(1)$ 成立。

归纳步骤。假设对某个 $n\geq 1$，$P(n)$ 成立。由第 (ii) 部分，

$$I_{n+1}=\frac{2n-1}{2nc^2}\,I_n=\frac{2n-1}{2n}\cdot\frac{\pi}{2^{2n-2}\,c^{2n+1}}\binom{2n-2}{n-1}.$$

我们要证明它等于 $\dfrac{\pi}{2^{2n}c^{2n+1}}\binom{2n}{n}$，这就是 $P(n+1)$。两个二项式系数之间的关系是

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

> **想一想** 怎样才能令 2 的幂次吻合?

写成 $\dfrac{2n-1}{2n}=\dfrac{4}{2^2}\cdot\dfrac{2n-1}{2n}=\dfrac{1}{4}\cdot\dfrac{2(2n-1)}{n}$。于是

$$\begin{aligned}I_{n+1}&=\frac{1}{4}\cdot\frac{2(2n-1)}{n}\cdot\frac{\pi}{2^{2n-2}\,c^{2n+1}}\binom{2n-2}{n-1}\\&=\frac{\pi}{2^{2n}\,c^{2n+1}}\cdot\frac{2(2n-1)}{n}\binom{2n-2}{n-1}=\frac{\pi}{2^{2n}\,c^{2n+1}}\binom{2n}{n}.\end{aligned}$$

由于 $2n=2(n+1)-2$、$n=(n+1)-1$ 及 $c^{2n+1}=(b-a^2)^{(n+1)-\frac{1}{2}}$，这正是 $P(n+1)$。由归纳法，$P(n)$ 对所有正整数 n 成立：

$$I_n=\frac{\pi}{2^{2n-2}(b-a^2)^{n-\frac{1}{2}}}\binom{2n-2}{n-1}.$$

验算

- 取 $a=0$、$b=1$ 及 $n=2$，公式给出 $I_2=\frac{\pi}{4}\binom{2}{1}=\frac{\pi}{2}\approx 1.5708$。以数值方法计算 $\int_{-\infty}^{\infty}(1+x^2)^{-2}\,\mathrm{d}x$，结果准确至小数点后十位都吻合。
- 正切代换亦给出 $I_n=c^{1-2n}\int_{-\pi/2}^{\pi/2}\cos^{2n-2}u\,\mathrm{d}u$，这是标准的沃利斯积分（Wallis integral），其已知值与公式相符。
- 对多组满足 $b>a^2$ 的随机 a 和 b，数值积分证实公式在 $n=1,\ldots,5$ 时成立，因此第 (ii) 部分的递推关系（recurrence）亦获证实。

常见错误

- 忘记说明第 (ii) 部分的边界项为何等于零；这需要 $n\geq 1$。

- 在分部积分中取 $v = x$ 而不取 $v = x + a$。这样做仍然可行，但会多出一项 $a\int (x + a)Q^{-n-1}\,\mathrm{d}x$ 需要处理。
- 在归纳步骤中弄错 $b - a^2$ 的幂次：每一步都乘以 $1/c^2$，所以 c 的指数刚好增加 2，由 $2n - 1$ 变为 $2n + 1$。

STEP 之外

有理函数在整条实数线上的积分，是展示复分析（complex analysis）威力的绝佳例子：留数定理（residue theorem）只需考虑 $1/Q^n$ 在上半平面（upper half-plane）内位于 $z = -a + ic$ 的唯一极点（pole），寥寥数行便能求出 I_n。这里出现的中心二项式系数（central binomial coefficient）亦见于 π 的沃利斯乘积（Wallis product）及随机漫步（random walk）之中。Needham 的 *Visual Complex Analysis*（《直观复分析》）以优美的几何方式阐述留数。

详细解答

STEP III 2016，第 2 题

纯数学 · 抛物线的法线

STEP III 2016，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/16-s3-q2

关键思路

先写出一般点 $(at^2, 2at)$ 处的法线（normal），再问哪些 t 值使法线经过 P。这个条件是一个关于 t 的三次方程（cubic），而且其中一个根 $t = p$ 必定存在（P 处的法线当然经过 P）。把这个根除去，剩下一个二次方程（quadratic），它的两个根正好是 q 和 r。此后的一切都由两根之和与两根之积，即 $q + r$ 和 qr，来推动。

解答

第 (i) 部分

在 $y^2 = 4ax$ 上取 $x = at^2$、$y = 2at$，得 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{2a}{2at} = \dfrac{1}{t}$，所以参数为 t 的点处的法线斜率（gradient）为 $-t$：

$$y - 2at = -t\,(x - at^2), \qquad \text{即} \qquad y + tx = 2at + at^3.$$

（这个形式同样适用于顶点（vertex）$t = 0$，该处的法线就是 x 轴。）

t 处的法线经过 $P(ap^2, 2ap)$ 当且仅当

$$2ap + tap^2 = 2at + at^3.$$

除以 $a > 0$ 并整理，得

$$\begin{aligned} 0 &= 2(p - t) + t(p^2 - t^2) \\ &= (p - t)\big(2 + t(p + t)\big) \\ &= (p - t)\,(t^2 + pt + 2). \end{aligned}$$

想一想 为甚么可以舍去因子 $p-t$?

根 $t=p$ 是平凡的那一个：P 本身的法线经过 P。题目说明 $Q\neq P$，所以 $q\neq p$，因子 $p-q$ 不为零。因此

$$q^2+qp+2=0,$$

正是所求。同理 $r^2+rp+2=0$。

第 (ii) 部分

于是 q 和 r 都是 $t^2+pt+2=0$ 的根。由于 $Q\neq R$，有 $q\neq r$，所以它们就是这个方程的两个根，从而

$$q+r=-p,\qquad qr=2.$$

（特别地，判别式（discriminant）为正：$p^2>8$。另外留意，$q=p$ 需要 $2p^2+2=0$，这是不可能的，所以 Q 和 R 自动与 P 不同。）

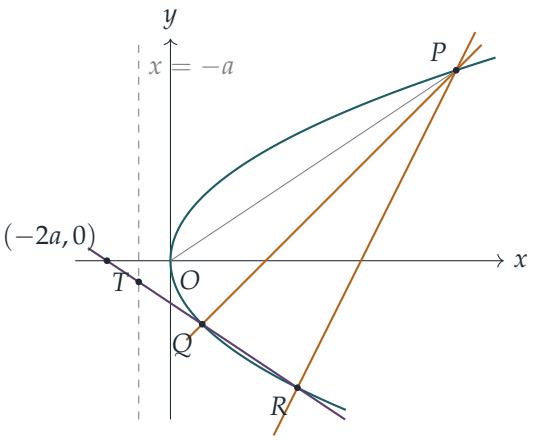
经过 $(aq^2,2aq)$ 和 $(ar^2,2ar)$ 的弦（chord）的斜率为 $\dfrac{2a(q-r)}{a(q^2-r^2)}=\dfrac{2}{q+r}$，其方程为

$$(q+r)\,y=2x+2aqr.$$

（验证：在 Q 处左边是 $2aq^2+2aqr$，与右边相等；在 R 处同理。）代入 $q+r=-p$ 和 $qr=2$：

$$-p\,y=2x+4a.$$

无论 p 取何值，点 $(-2a,0)$ 都满足这个方程。所以 QR 必定经过 $(-2a,0)$，而这一点与 P 的选取无关。



上图显示 $a=1$、$p=3$ 的情况，此时 $q=-1$、$r=-2$：两条法线（琥珀色）相交于 P，弦 QR（紫红色）经过 $(-2,0)$，而 T 位于虚线 $x=-1$ 上。

第 (iii) 部分

由于 $p^2 > 8$，$p \neq 0$，所以 $P \neq O$，直线 OP 为 $y = \dfrac{2ap}{ap^2}\,x = \dfrac{2}{p}\,x$，即 $x = \frac{1}{2}py$。（它与 QR 不平行，因为后者的斜率是 $-2/p$。）

把 $x = \frac{1}{2}py$ 代入 QR 的方程：

$$-py = py + 4a \quad \Longrightarrow \quad y = -\frac{2a}{p}, \qquad x = \tfrac{1}{2}p\Big(-\frac{2a}{p}\Big) = -a.$$

所以 $T = \big(-a, -\frac{2a}{p}\big)$，它必定位于直线 $x = -a$（抛物线的准线（directrix））上，与 P 的选取无关。

T 与 x 轴的距离是 $\dfrac{2a}{|p|}$。由第 (ii) 部分可知，经过 P 存在两条不同的法线，这迫使 $p^2 > 8$，即 $|p| > 2\sqrt{2}$。由于 $a > 0$，

$$\frac{2a}{|p|} < \frac{2a}{2\sqrt{2}} = \frac{a}{\sqrt{2}},$$

证明完毕。不等式是严格的，因为 $p^2 = 8$ 会使 $q = r$。

验算

- 在图中 $a = 1$、$p = 3$ 的情况下：$t^2 + 3t + 2$ 的根是 -1 和 -2；法线 $y = x - 3$ 和 $y = 2x - 12$ 都经过 $P(9, 6)$；经过 $(1, -2)$ 和 $(4, -4)$ 的弦是 $-3y = 2x + 4$，它经过 $(-2, 0)$；而 $T = (-1, -\frac{2}{3})$，且 $\frac{2}{3} < \frac{1}{\sqrt{2}} \approx 0.707$。
- 令 $|p| \to 2\sqrt{2}$ 从上方趋近，Q 和 R 会合而为一，T 的高度趋于 $a/\sqrt{2}$，所以这个界是最佳的。以电脑对 20,000 组随机选取的 a 和 p 进行检验，所有结果均得到确认，T 的高度最高达到 $a/\sqrt{2}$ 的 0.9996 倍，但从未超过。

常见错误

- 忘记说明为何可以除以 $p - q$：正是条件 $Q \neq P$ 容许这样做。
- 没有利用 $q \neq r$ 来说明 q 和 r 都是同一个二次方程的根，而这正是得出 $q + r = -p$ 和 $qr = 2$ 的依据。
- 证明最后的界时没有交代 $p^2 > 8$ 从何而来：它来自二次方程必须有两个相异实根这一要求。

STEP 之外

第 (i) 部分背后关于 t 的三次方程，是一个经典事实的特例：对一般点 (h,k)，条件是 $at^3+(2a-h)\,t-k=0$，所以从平面上一点最多可以作三条抛物线的法线，而“一条法线”与“三条法线”两个区域之间的分界是一条称为渐屈线（evolute）的曲线（这里是一条半立方抛物线（semicubical parabola））。渐屈线、包络线（envelope）和曲率（curvature）都是曲线微分几何（differential geometry）入门课程的标准课题。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶题》）收录了更多这种坐标几何风格的 STEP 题目。

详细解答

STEP III 2016，第 3 题

纯数学 · 积分与多项式

STEP III 2016，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/16-s3-q3

本题在 *Beyond STEP* 第 4.6 节没有公式的积分中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch04/no-formula.html

关键思路

形如 $\int g(x)\,\mathrm{e}^x\,\mathrm{d}x = \frac{P}{Q}\,\mathrm{e}^x + \text{常数}$ 的等式，其实只是一个关于导数（derivative）的陈述，所以把右边求导，再消去分母即可。这样问题便化为一个多项式恒等式（polynomial identity），而多项式恒等式有两种处理方法：代入一个精心挑选的值（这里是 $x=-1$），以及比较次数（degree）。

解答

以下把 $P(x)$、$Q(x)$ 简写为 P、Q。由商法则（quotient rule）和积法则（product rule），

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{P}{Q}\,\mathrm{e}^x\right) = \frac{PQ + P'Q - PQ'}{Q^2}\,\mathrm{e}^x.$$

因此，积分等式成立当且仅当此式等于被积函数（integrand）（其中 Q 不是零多项式）。

第 (i) 部分

令导数等于 $\frac{x^3-2}{(x+1)^2}\,\mathrm{e}^x$，约去 $\mathrm{e}^x \neq 0$ 并去分母，得多项式恒等式

$$(x+1)^2\left(PQ + P'Q - PQ'\right) = \left(x^3-2\right)Q^2. \qquad (*)$$

Q 有因式 $x+1$。在 $(*)$ 中代入 $x=-1$。左边为 0，右边为 $-3\,Q(-1)^2$。因此 $Q(-1)=0$，由因式定理（factor theorem）知 $x+1$ 整除 Q。

次数。P 不是零多项式，否则导数会是 0，而不是被积函数。设 $\deg P = p$、$\deg Q = q$，首项系数（leading coefficient）分别为 $a \neq 0$ 和 $b \neq 0$。在 $PQ + P'Q - PQ'$ 中，PQ 一项的次数为 $p+q$，首项系数为 ab；而 $P'Q$ 和 PQ' 的次数各自至多为 $p+q-1$。所以 $PQ + P'Q - PQ'$ 的次数恰好是 $p+q$，而 $(*)$ 左边的次数是 $p+q+2$。右边的次数是 $3+2q$。因此

$$p+q+2 = 2q+3 \quad \Longrightarrow \quad p = q+1,$$

即 P 的次数恰好比 Q 的次数多一。

$Q = x+1$ 的情况。此时 P 的次数为 2，设 $P = ax^2 + bx + c$。由于 $Q = x+1$、$Q' = 1$，$(*)$ 变为

$$(x+1)^2\big((x+1)P + (x+1)P' - P\big) = (x^3-2)(x+1)^2,$$

约去 $(x+1)^2$ 得

$$xP + (x+1)P' = x^3 - 2.$$

由于 $xP = ax^3 + bx^2 + cx$，而 $(x+1)P' = (x+1)(2ax+b) = 2ax^2 + (2a+b)x + b$，

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

比较系数：$a = 1$；由常数项得 $b = -2$；于是 x^2 项的条件 $2a + b = 0$ 亦相符；而由 x 项得 $c = -(2a+b) = 0$。因此

$$P(x) = x^2 - 2x, \qquad \int \frac{x^3-2}{(x+1)^2}\,\mathrm{e}^x\,\mathrm{d}x = \frac{x^2-2x}{x+1}\,\mathrm{e}^x + \text{常数}.$$

第 (ii) 部分

用反证法（proof by contradiction），假设存在这样的 P 和 Q，且两者没有公因式。和之前一样，求导并去分母，得

$$(x+1)\big(PQ + P'Q - PQ'\big) = Q^2. \qquad (**)$$

代入 $x = -1$ 得 $Q(-1)^2 = 0$，所以 $x+1$ 整除 Q。记

$$Q = (x+1)^n R, \qquad n \geq 1, \quad R(-1) \neq 0,$$

使 n 恰好是 Q 中 $x+1$ 的幂次。于是 $Q' = n(x+1)^{n-1}R + (x+1)^n R'$，而

$$\begin{aligned} PQ + P'Q - PQ' &= (x+1)^n\big(PR + P'R - PR'\big) - n(x+1)^{n-1}PR \\ &= (x+1)^{n-1}\Big[(x+1)\big(PR + P'R - PR'\big) - nPR\Big]. \end{aligned}$$

代入 $(**)$，并从两边约去公因式 $(x+1)^n$（右边是 $(x+1)^{2n}R^2$）：

$$(x+1)\big(PR + P'R - PR'\big) - nPR = (x+1)^n R^2.$$

想一想 为甚么要费功夫把 $x+1$ 的确切幂次抽出来，而不是直接再把 $x=-1$ 代入 $(**)$?

直接把 $x=-1$ 代入 $(**)$，只能告诉我们 $Q(-1)=0$；之后每一项都等于零，我们对 P 一无所知。把 Q 中所有 $x+1$ 的因式移除后，便剩下一项 $-nPR$，它在 $x=-1$ 时不会自动等于零。

现在代入 $x=-1$。由于 $n\geq 1$，右边为 0，左边第一项亦为 0。因此

$$-n\,P(-1)\,R(-1)=0.$$

但 $n\neq 0$ 且 $R(-1)\neq 0$，所以 $P(-1)=0$。于是 $x+1$ 既是 Q 的因式，也是 P 的因式，与两者没有公因式的假设矛盾。所以这样的多项式不存在。 ■

（限定 P、Q 互质（coprime）并无损一般性：任何公因式都可以从 P/Q 中约去，而不改变该函数。）

验算

- 直接对 (i) 的答案求导：$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x^2-2x}{x+1}\right)=\frac{x^2+2x-2}{(x+1)^2}$，再加上 $\frac{x^2-2x}{x+1}=\frac{x^3-x^2-2x}{(x+1)^2}$，得 $\frac{x^3-2}{(x+1)^2}$，正如所需。
- 数值上，$\int_0^1 \frac{x^3-2}{(x+1)^2}\,\mathrm{e}^x\,\mathrm{d}x\approx -1.359141$，与 $\left[\frac{x^2-2x}{x+1}\mathrm{e}^x\right]_0^1=-\frac{\mathrm{e}}{2}$ 相符。
- 在 (i) 中次数吻合：P 的次数为 2，比 $Q=x+1$ 多一。

常见错误

- 忘记 $P'Q-PQ'$ 的次数小于 $p+q$，因为首项互相抵消；决定次数的只有 PQ。
- 在 (ii) 中，只把 $x=-1$ 代入 $(**)$，求得 $Q(-1)=0$ 便停下来。要得出矛盾，必须用到 Q 中 $x+1$ 的确切幂次。

STEP 之外

第 (ii) 部分说明，$\int \frac{\mathrm{e}^x}{x+1}\,\mathrm{d}x$（本质上就是指数积分（exponential integral）Ei）在这个受限的意义下并不是“初等”（elementary）的。判定一个积分究竟能否用初等函数表示的完整理论，是刘维尔（Liouville）定理以及微分代数（differential algebra）中的 Risch 算法；这里的论证只是其中一个小例子。若想仔细了解上面用到的多项式代数（因式定理、整除性、互质），可参阅 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）。

详细解答

STEP III 2016，第 4 题

纯数学 · 裂项相消级数

STEP III 2016，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/16-s3-q4

关键思路

题目建议的两个分数之差，原来是通项（general term）的常数倍，因此这个和会伸缩相消（telescoping）：几乎所有项都互相抵消，只剩下首尾两部分。第 (ii) 部分中，把 sech 用指数函数表示并提出一个因子后，每一项看起来就与第 (i) 部分取 $x = \mathrm{e}^{-2y}$ 时的项完全一样。由于 sech 是偶函数（even function），双向无限和便可以折回单向的和。

解答

第 (i) 部分

把建议的差化为公分母（common denominator），

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

由于 $|x| \neq 1$，故 $x \neq 1$，所以可以除以 $x-1$。（这个条件亦保证 $x \neq -1$，因此没有分母 $1+x^r$ 会等于零：若 $x^r = -1$，则必有 $|x| = 1$。）因此

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

由 $r = 1$ 加到 N，每个括号中的第二个分数都与下一个括号的第一个分数抵消，剩下

$$\begin{aligned}\sum_{r=1}^{N} \frac{x^r}{(1+x^r)(1+x^{r+1})} &= \frac{1}{x-1}\left(\frac{1}{1+x} - \frac{1}{1+x^{N+1}}\right)\\ &= \frac{x\,(x^N-1)}{(x-1)(1+x)(1+x^{N+1})},\end{aligned}$$

其中第二个形式是把两个分数合并而得：分子为 $x^{N+1}-x=x(x^N-1)$。

现设 $|x|<1$。当 $N\to\infty$ 时 $x^{N+1}\to 0$，因此部分和（partial sums）趋向

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

这正是所求的无穷级数之和。

第 (ii) 部分

对任何实数 t，

$$\operatorname{sech}t=\frac{2}{\mathrm{e}^{t}+\mathrm{e}^{-t}}=\frac{2\mathrm{e}^{-t}}{1+\mathrm{e}^{-2t}}.$$

设 $x=\mathrm{e}^{-2y}$；由于 $y>0$，故 $0<x<1$。于是 $\operatorname{sech}(ry)=2\mathrm{e}^{-ry}/(1+x^r)$，而

$$\operatorname{sech}(ry)\operatorname{sech}\big((r+1)y\big)=\frac{4\,\mathrm{e}^{-(2r+1)y}}{(1+x^r)(1+x^{r+1})}=4\mathrm{e}^{-y}\,\frac{x^r}{(1+x^r)(1+x^{r+1})},$$

因为 $\mathrm{e}^{-2ry}=x^r$。

> **想一想** 为甚么要把 $\mathrm{e}^{-(2r+1)y}$ 拆成 $\mathrm{e}^{-y}\cdot\mathrm{e}^{-2ry}$?

我们希望分子是 x^r，即分母中出现的同一个 x 的幂，这样第 (i) 部分便可以逐项套用；剩下的因子 e^{-y} 与 r 无关，可以提到和号之外。

由于 $|x|<1$，由第 (i) 部分得

$$\sum_{r=1}^{\infty}\operatorname{sech}(ry)\operatorname{sech}\big((r+1)y\big)=4\mathrm{e}^{-y}\,\frac{x}{1-x^2}=\frac{4\mathrm{e}^{-y}}{x^{-1}-x}=\frac{4\mathrm{e}^{-y}}{\mathrm{e}^{2y}-\mathrm{e}^{-2y}}=\frac{4\mathrm{e}^{-y}}{2\sinh 2y},$$

即 $2\mathrm{e}^{-y}\operatorname{cosech}(2y)$，与题目所求相同。把这个和记作 S。

双向无限和

把对所有整数 r 求的和分成三部分：$r\ge 1$、$r\in\{-1,0\}$ 及 $r\le -2$。

> **想一想** r 为负数的项是怎样的?

对 $r\le -2$，设 $s=-r-1$，则 $s\ge 1$，且 $r+1=-s$、$r=-(s+1)$。由于 sech 是偶函数，

$$\operatorname{sech}(ry)\operatorname{sech}\big((r+1)y\big)=\operatorname{sech}\big((s+1)y\big)\operatorname{sech}(sy),$$

而当 r 取遍 $-2,-3,\ldots$ 时，s 取遍 $1,2,\ldots$。所以 $r\le -2$ 的各项之和亦为 S。（两个单向的和都收敛（converge），因此这样拆分双向无限和是合理的。）

中间的两项为

$$r=-1:\ \operatorname{sech}(-y)\operatorname{sech}0=\operatorname{sech}y,\qquad r=0:\ \operatorname{sech}0\operatorname{sech}y=\operatorname{sech}y.$$

因此，利用 $\sinh 2y = 2\sinh y\cosh y$，

$$\sum_{r=-\infty}^{\infty} \operatorname{sech}(ry)\operatorname{sech}\big((r+1)y\big) = 2S + 2\operatorname{sech} y = \frac{2\mathrm{e}^{-y}}{\sinh y\cosh y} + \frac{2\sinh y}{\sinh y\cosh y}$$
$$= \frac{2(\mathrm{e}^{-y}+\sinh y)}{\sinh y\cosh y}.$$

但 $\mathrm{e}^{-y} + \sinh y = \mathrm{e}^{-y} + \frac{1}{2}(\mathrm{e}^{y} - \mathrm{e}^{-y}) = \frac{1}{2}(\mathrm{e}^{y} + \mathrm{e}^{-y}) = \cosh y$，所以 $\cosh y$ 互相抵消，得

$$\sum_{r=-\infty}^{\infty} \operatorname{sech}(ry)\operatorname{sech}\big((r+1)y\big) = \frac{2}{\sinh y} = 2\operatorname{cosech} y.$$

验算

- 取 $N = 1$，有限和公式给出 $\dfrac{x(x-1)}{(x-1)(1+x)(1+x^2)} = \dfrac{x}{(1+x)(1+x^2)}$，正是 $r = 1$ 的单一项。
- 取 $x = \frac{1}{2}$：无穷级数之和应为 $\frac{1/2}{3/4} = \frac{2}{3}$，而部分和确实趋近 0.6667。
- 取 $y = 1$：$2\mathrm{e}^{-1}\operatorname{cosech} 2 \approx 0.202864$，$2\operatorname{cosech} 1 \approx 1.701836$，两者都与直接对级数求和的结果在多个小数位上吻合（已用电脑对多个 y 值验证）。
- 当 $y \to \infty$ 时，只有 $r = -1, 0$ 两项不会消失，每项约为 $2\mathrm{e}^{-y}$，合共约 $4\mathrm{e}^{-y}$；而 $2\operatorname{cosech} y \approx 4\mathrm{e}^{-y}$，亦相符。

常见错误

- 忘记因子 $\frac{1}{x-1}$，或弄错其正负号：该差等于项的 $(x-1)$ 倍，而不是 $(1-x)$ 倍。
- 在双向无限和中，只把 S 乘以二，却忘记 $r = -1$ 及 $r = 0$ 两项；这两项涉及 $\operatorname{sech} 0 = 1$，并不属于任何一个单向的和。

STEP 之外

在这里重排及拆分无穷级数是安全的，因为每一项都是正数（级数绝对收敛（converges absolutely））；对于条件收敛（conditionally convergent）的级数，黎曼（Riemann）重排定理（rearrangement theorem）指出，重新排列各项可以令和变成任何你想要的数值——这是大学第一个分析（analysis）课程中会遇到的惊人结果。像 $\operatorname{sech}(ry)\operatorname{sech}\big((r+1)y\big)$ 这类乘积对所有整数求和，亦会在椭圆函数（elliptic functions）理论中出现，而这类双边级数（bilateral series）往往有优美的闭合形式。关于级数的分析，可参阅 Abbott, *Understanding Analysis*（《理解分析》）。

详细解答

STEP III 2016，第 5 题

纯数学 · 质数的乘积

STEP III 2016，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/16-s3-q5

本题在 *Beyond STEP* 第 1.2 节质数分布得有多稀疏？中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch01/prime-distribution.html

关键思路

中间的二项式系数（binomial coefficient）$\binom{2m+1}{m}$ 并不大（它和它的“孪生”项 $\binom{2m+1}{m+1}$ 与其他各项共同分享总和 2^{2m+1}），但它可被 $m+1$ 与 $2m+1$ 之间的每一个质数（prime）整除。因此这些质数的乘积也不大。把不超过 $2m+1$ 的质数分成“不超过 $m+1$”和“大于 $m+1$”两组，便可以建立一个强归纳法（strong induction）；而偶数完全不会带来额外麻烦，因为它们（除了 2 以外）都不是质数。

解答

第 (i) 部分

在 $(1+x)^{2m+1}$ 的二项展开式（binomial expansion）中代入 $x=1$：

$$\sum_{k=0}^{2m+1}\binom{2m+1}{k}=2^{2m+1}.$$

左边每一项都是正数。由对称性，$\binom{2m+1}{m}=\binom{2m+1}{m+1}$，而这是和式中两个不同的项。由于 $m\geq 1$，项 $\binom{2m+1}{0}=1$ 是第三个不同的项。因此

$$2\binom{2m+1}{m}=\binom{2m+1}{m}+\binom{2m+1}{m+1}<2^{2m+1},$$

所以 $\binom{2m+1}{m}<2^{2m}$，得证。

第 (ii) 部分

写成

$$\binom{2m+1}{m} = \frac{(2m+1)!}{m!\,(m+1)!}, \qquad \text{所以} \qquad m!\,(m+1)!\binom{2m+1}{m} = (2m+1)!.$$

设 p 为满足 $m+1 < p \le 2m+1$ 的质数。那么 p 是 $(2m+1)!$ 的其中一个因数，所以 p 整除右边。但 $m!\,(m+1)!$ 的每个因数都不超过 $m+1 < p$，所以质数 p 不整除其中任何一个，因此 $p \nmid m!\,(m+1)!$。若一个质数整除乘积 ab，它必定整除 a 或 b，所以 $p \mid \binom{2m+1}{m}$。

? **想一想** 每个这样的质数都整除该系数。为甚么它们的乘积也整除它?

这些质数互不相同，所以它们两两互质 (pairwise coprime)，而一个能被若干个两两互质的数整除的数，必能被它们的乘积整除。(等价地，由唯一分解定理 (unique factorisation)，这些质数每一个都出现在 $\binom{2m+1}{m}$ 的质因数分解中。) 所以 $P_{m+1,2m+1}$ 整除 $\binom{2m+1}{m}$。若不存在这样的质数，则 $P_{m+1,2m+1} = 1$，它整除一切整数。

正整数的正因数不会大于该数本身，所以由第 (i) 部分，

$$P_{m+1,2m+1} \le \binom{2m+1}{m} < 2^{2m}.$$

第 (iii) 部分

由于 $m \ge 1$，我们有 $1 < m+1 < 2m+1$。每个质数 $p \le 2m+1$ 要么满足 $p \le m+1$，要么满足 $m+1 < p \le 2m+1$，但不会两者同时成立，所以

$$P_{1,2m+1} = P_{1,m+1}\,P_{m+1,2m+1}.$$

此外 $2 \le m+1 \le 2m$ (因为 $m \ge 1$)，所以 $k = m+1$ 是假设所涵盖的值之一，从而得到 $P_{1,m+1} < 4^{m+1}$。结合第 (ii) 部分，

$$P_{1,2m+1} < 4^{m+1} \cdot 2^{2m} = 4^{m+1} \cdot 4^m = 4^{2m+1}.$$

第 (iv) 部分

我们对 n 使用强归纳法。设 $S(n)$ 为命题 $P_{1,n} < 4^n$。

基本情况。 $P_{1,2} = 2 < 16 = 4^2$。

归纳步骤。设 $n \ge 3$，并假设 $S(k)$ 对 $k = 2, 3, \ldots, n-1$ 均成立。

- 若 n 是奇数，写成 $n = 2m+1$，其中 $m \ge 1$。那么 $n-1 = 2m$，所以 $S(k)$ 对 $k = 2, \ldots, 2m$ 成立。由第 (iii) 部分得 $P_{1,2m+1} < 4^{2m+1}$，这正是 $S(n)$。

- 若 n 是偶数，则 $n \geq 4$ 是大于 2 的偶数，所以它不是质数。因此不超过 n 的质数恰好就是不超过 $n-1$ 的质数，于是

$$P_{1,n} = P_{1,n-1} < 4^{n-1} < 4^n.$$

想一想 为甚么我们不能对每个 n 都直接使用第 (iii) 部分?

第 (iii) 部分只处理奇数 $2m+1$。对于偶数 n，我们需要另一个论证；而上面的论证很简单，因为大于 2 的偶数不会为乘积添加新的质数。

所以 $S(n)$ 在两种情况下都成立，由强归纳法，$P_{1,n} < 4^n$ 对所有 $n \geq 2$ 成立。

验算

- $m = 3$：$\binom{7}{3} = 35 < 64 = 2^6$。在 $(4,7]$ 中的质数是 5 和 7，而 $P_{4,7} = 35$ 恰好整除 35。
- $m = 5$：$\binom{11}{5} = 462 = 2 \cdot 3 \cdot 7 \cdot 11$，它可被 $P_{6,11} = 77$ 整除。此外 $462 < 1024$。
- $P_{1,10} = 2 \cdot 3 \cdot 5 \cdot 7 = 210$，远小于 $4^{10} = 1\,048\,576$。电脑验证确认 $P_{1,n} < 4^n$ 对 2 至 1500 的每个 n 都成立。

常见错误

- 在第 (i) 部分把 $<$ 写成 $\leq$：要得到严格不等式，除了两个中间项之外，还需要第三个正项，例如 $\binom{2m+1}{0}$。
- 在第 (iii) 部分，忘记检查 $m+1$ 是否位于假设适用的范围 $2, \ldots, 2m$ 之内。
- 在第 (iv) 部分，只处理奇数 n，或者忘记基本情况 $n = 2$。

STEP 之外

这是艾狄胥（Erdős）对上界 $\prod_{p \leq n} p < 4^n$ 的初等证明。取对数后，它说明切比雪夫（Chebyshev）函数 $\theta(n) = \sum_{p \leq n} \ln p$ 小于 $n \ln 4$。这是切比雪夫定理的其中一半；该定理指出不超过 n 的质数个数的数量级为 $n / \ln n$。艾狄胥用同类的二项式系数论证证明了伯特兰公设（Bertrand's postulate）：在 n 与 $2n$ 之间必定存在一个质数。Davenport 的 *The Higher Arithmetic*（《高等算术》）是进一步学习初等数论的好选择。

详细解答

STEP III 2016，第 6 题

纯数学 · 双曲函数

STEP III 2016，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/16-s3-q6

关键思路

把 $\operatorname{sech} x = a\tanh x + b$ 两边乘以 $\cosh x$，得 $a\sinh x + b\cosh x = 1$，这正是开首部分所处理的组合。因此第一步就是把大家熟悉的“$R\cos(x+\alpha)$”技巧搬到双曲函数（hyperbolic functions）上。与三角函数的情况不同，所得的形式取决于 B 与 $\pm A$ 的大小关系，而每种形式都能让你立即看出解的个数。

解答

$A\sinh x + B\cosh x$ 的各种形式

展开，

$$R\cosh(x+\gamma) = R\sinh\gamma\,\sinh x + R\cosh\gamma\,\cosh x,$$

所以我们需要 $R\sinh\gamma = A$ 及 $R\cosh\gamma = B$。由于 $\cosh^2\gamma - \sinh^2\gamma = 1$，这就迫使 $R^2 = B^2 - A^2$ 及 $\tanh\gamma = A/B$。当 $B > A > 0$ 时，两者都可以满足：取

$$R = \sqrt{B^2 - A^2} > 0, \qquad \gamma = \operatorname{artanh}\frac{A}{B},$$

由于 $0 < A/B < 1$，γ 是存在的。于是 $\cosh\gamma = 1/\sqrt{1 - A^2/B^2} = B/R$，且 $\sinh\gamma = \tanh\gamma\cosh\gamma = A/R$，所以 $R\cosh\gamma = B$ 及 $R\sinh\gamma = A$，符合要求。因此 $A\sinh x + B\cosh x = R\cosh(x+\gamma)$。

其他情况（设 $A > 0$）：

- $-A < B < A$：改为比较 $R\sinh(x+\gamma) = R\cosh\gamma\sinh x + R\sinh\gamma\cosh x$ 的系数，得

$$A\sinh x + B\cosh x = \sqrt{A^2 - B^2}\,\sinh(x+\gamma), \qquad \gamma = \operatorname{artanh}\frac{B}{A}.$$

- $B = A$：$A(\sinh x + \cosh x) = A\mathrm{e}^{x}$。
- $B = -A$：$A(\sinh x - \cosh x) = -A\mathrm{e}^{-x}$。
- $B < -A$：把第一个结果应用于 $-A\sinh x - B\cosh x$（此时 $-B > A$ 而 $-A < 0$；上面的推导其实只需要 $|A| < B$），得

$$A\sinh x + B\cosh x = -\sqrt{B^2 - A^2}\,\cosh(x + \gamma), \qquad \gamma = \operatorname{artanh}\frac{A}{B}.$$

交点方程

由于 $\cosh x > 0$，两曲线相交于

$$\operatorname{sech} x = a\tanh x + b \iff 1 = a\sinh x + b\cosh x. \tag{$*$}$$

第 (i) 部分

当 $b > a > 0$ 时，记 $R = \sqrt{b^2 - a^2}$；于是 $(*)$ 变成

$$\cosh\left(x + \operatorname{artanh}\tfrac{a}{b}\right) = \frac{1}{\sqrt{b^2 - a^2}}.$$

若两曲线相交，则此方程有解，所以右方至少为 1。当 $c \geq 1$ 时，$\cosh u = c$ 的解为 $u = \pm\operatorname{arcosh} c$，所以

$$x = \pm\operatorname{arcosh}\left(\frac{1}{\sqrt{b^2 - a^2}}\right) - \operatorname{artanh}\frac{a}{b},$$

这正是所要求的形式。

第 (ii) 部分

当 $a > b > 0$ 时，属于 sinh 的情况：$(*)$ 变成 $\sqrt{a^2 - b^2}\,\sinh\left(x + \operatorname{artanh}\frac{b}{a}\right) = 1$。由于 sinh 是从 $\mathbb{R}$ 到 $\mathbb{R}$ 的双射（bijection），两曲线必定恰有一个交点，位于

$$x = \operatorname{arsinh}\left(\frac{1}{\sqrt{a^2 - b^2}}\right) - \operatorname{artanh}\frac{b}{a}.$$

第 (iii) 部分

按开首部分的每种情况（取 $A = a > 0$，$B = b$）数算 $(*)$ 的解：

- $b < -a$：左方 $-\sqrt{b^2 - a^2}\cosh(x + \gamma)$ 是负数，所以无解。
- $b = -a$：$-a\mathrm{e}^{-x} = 1$ 无解。
- $-a < b < a$：恰有一个解，与 (ii) 相同；那里的论证对此范围内任何 b 都成立。

- $b = a$：$ae^x = 1$ 只有一个解 $x = -\ln a$。
- $b > a$：若 $1/\sqrt{b^2 - a^2} > 1$，则 $\cosh(x + \gamma) = 1/\sqrt{b^2 - a^2}$ 有两个解；若等于 1，则有一个解；若小于 1，则无解。

> **想一想** 为甚么 $b \le a$ 的情况永远不可能给出两个交点?

在这些情况下，$(*)$ 的左方要么是 x 的严格单调（strictly monotonic）函数（$\sinh(x + \gamma)$ 或 $e^{\pm x}$ 的倍数），要么恒为负数，所以它最多只取值 1 一次。只有先下降再上升的 cosh 形式才能两次达到 1。

所以两曲线相交于两个不同点，当且仅当 $b > a$ 且 $b^2 - a^2 < 1$，即

$$a < b < \sqrt{a^2 + 1}.$$

第 (iv) 部分

由于 $\operatorname{sech} x - (a \tanh x + b) = \operatorname{sech} x \left(1 - a \sinh x - b \cosh x\right)$ 且 $\operatorname{sech} x \neq 0$，两曲线相切（touch）恰好发生在 $g(x) = a \sinh x + b \cosh x - 1$ 有重根（repeated zero）之处，即 $g(x) = g'(x) = 0$。在 sinh 及指数的情况下，$g'(x)$ 是 $\cosh(x + \gamma)$ 或 e^x 的倍数，永不为零，所以没有重根。在 $b > a$ 的情况下，$g(x) = R \cosh(x + \gamma) - 1$ 而 $g'(x) = R \sinh(x + \gamma)$，两者只在 $x + \gamma = 0$ 且 $R = 1$ 时同时为零。因此两曲线相切，当且仅当

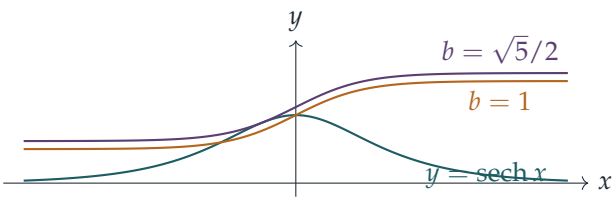
$$b > a \quad \text{且} \quad b^2 - a^2 = 1, \qquad \text{即} \quad b = \sqrt{1 + a^2}.$$

切点的 $x = -\gamma = -\operatorname{artanh}(a/b)$，所以

$$y = \operatorname{sech}\left(\operatorname{artanh} \tfrac{a}{b}\right) = \sqrt{1 - \tfrac{a^2}{b^2}} = \frac{\sqrt{b^2 - a^2}}{b} = \frac{1}{b} = \frac{1}{\sqrt{1 + a^2}},$$

这里用了 $\operatorname{sech}^2 u = 1 - \tanh^2 u$ 及 $\operatorname{sech} u > 0$。作为验证，另一条曲线同样给出 $y = -a \cdot \frac{a}{b} + b = \frac{b^2 - a^2}{b} = \frac{1}{b}$。

下图显示 $a = \frac{1}{2}$ 的情况：当 $b = 1$（琥珀色）时有两个交点；当 $b = \sqrt{5}/2$（梅红色）时，两曲线在高度 $2/\sqrt{5}$ 处相切。



验算

取 $a=\frac{3}{4}$，$b=1$，则 $b^2-a^2=\frac{7}{16}<1$，按 (iii) 应有两个交点。此时 $\operatorname{arcosh}(4/\sqrt{7})=\ln\big((4+3)/\sqrt{7}\big)=\ln\sqrt{7}$，而 $\operatorname{artanh}\frac{3}{4}=\frac{1}{2}\ln 7=\ln\sqrt{7}$，所以 (i) 的公式给出 $x=0$ 及 $x=-\ln 7\approx -1.946$。在 $x=0$ 处，两曲线都经过 $(0,1)$，与预期相符。另外，以电脑对大量随机的 a 和 b 检验 $(*)$，亦确认了每种情况下根的个数，以及相切条件确实给出重根。

常见错误

- 忘记边界情况 $b=\pm a$：此时 cosh 和 sinh 形式都不存在，但该式化为单一个指数函数。
- 在 (iii) 中只写出 $b^2-a^2<1$ 而漏掉 $b>a$：当 $b<a$ 时两曲线仍然相交，但只有一个交点。
- 弄错 γ 的正负号：解是 $\pm\operatorname{arcosh}(\cdot)$ 减去 $\operatorname{artanh}(a/b)$。

STEP 之外

情况的划分由 B^2-A^2 的正负决定：点 (A,B) 位于由直线 $B=\pm A$ 所分隔的某个区域之内，或位于这些直线上。这正是狭义相对论（special relativity）的几何：劳仑兹推动（Lorentz boost）是保持 t^2-x^2 不变的双曲“旋转”，正如普通旋转保持 x^2+y^2 不变；而 $B=\pm A$ 担当光锥（light cone）的角色——那是推动永远无法令你越过的唯一边界。想看更多双曲函数及这类题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

STEP III 2016，第 7 题

纯数学 · 单位根

> **STEP III 2016，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/16-s3-q7

本题在 *Beyond STEP* 第 5.4 节正多边形中的距离中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch05/polygon-distances.html

关键思路

距离就是模（modulus），而模的乘积等于乘积的模。若把正多边形放在阿冈图（Argand diagram）上，使其顶点为经旋转的 n 次单位根（roots of unity），那么根据开首的恒等式，距离的乘积就变成某个适当的 w 所对应的 $|w^n-1|$。唯一真正要花心思的，是选好坐标：把 O 放在原点，并令 X_0X_1 的垂直平分线（perpendicular bisector）为实轴，这样 P 就是一个实数。

解答

因式分解

对 $k=0,1,\ldots,n-1$，我们有 $(\omega^k)^n=\mathrm{e}^{2\pi ik}=1$，因此每个 ω^k 都是 z^n-1 的根。这 n 个数互不相同，因为它们的辐角（argument）$2\pi k/n$ 是 $[0,2\pi)$ 中互不相同的数。恒等式两边都是 n 次的首一多项式（monic polynomial），所以它们的差

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

的次数（degree）至多为 $n-1$。但 D 在 n 个不同的点 $1,\omega,\ldots,\omega^{n-1}$ 处取值为零，而次数至多为 $n-1$ 的非零多项式最多只有 $n-1$ 个根。因此 $D\equiv 0$，即

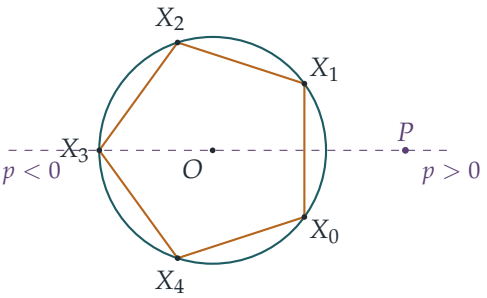
$$\prod_{k=0}^{n-1}(z-\omega^k)=z^n-1\quad\text{对所有复数}z.$$

第 (i) 部分

我们假设顶点沿圆周依次标记，因此 X_0 与 X_1 相邻，且 $\angle X_0OX_1 = 2\pi/n$。旋转或反射图形都不会改变任何距离，所以可以把 O 放在原点，并令

$$X_k = \mathrm{e}^{\mathrm{i}(2k-1)\pi/n} = \mathrm{e}^{-\mathrm{i}\pi/n}\,\omega^k, \qquad k = 0, 1, \ldots, n-1.$$

于是 $X_0 = \mathrm{e}^{-\mathrm{i}\pi/n}$ 与 $X_1 = \mathrm{e}^{\mathrm{i}\pi/n}$ 互为关于实轴的反射，所以 X_0X_1 的垂直平分线就是实轴。与 X_0 和 X_1 等距的点恰好就是这条直线上的点，所以 P 是一个实数 p，且 $|OP| = |p|$。X_0X_1 的中点是 $\cos(\pi/n) > 0$，因此 $p > 0$ 表示 P 与边 X_0X_1 位于 O 的同一侧。



（图中所示为 $n = 5$；当 n 为奇数时，通过 O 的这条直线亦会经过对面的顶点。）

现在，利用 $|ab| = |a||b|$ 及 $|\mathrm{e}^{-\mathrm{i}\pi/n}| = 1$，

$$\begin{aligned}|PX_0|\,|PX_1|\cdots|PX_{n-1}| &= \prod_{k=0}^{n-1}|p - \mathrm{e}^{-\mathrm{i}\pi/n}\omega^k| = \prod_{k=0}^{n-1}|\mathrm{e}^{-\mathrm{i}\pi/n}|\,|p\,\mathrm{e}^{\mathrm{i}\pi/n} - \omega^k| \\ &= \left|\prod_{k=0}^{n-1}\left(p\,\mathrm{e}^{\mathrm{i}\pi/n} - \omega^k\right)\right| = \left|\left(p\,\mathrm{e}^{\mathrm{i}\pi/n}\right)^n - 1\right|,\end{aligned}$$

这里用了因式分解（factorisation），并取 $z = p\,\mathrm{e}^{\mathrm{i}\pi/n}$。由于 $\mathrm{e}^{\mathrm{i}\pi} = -1$，

$$|PX_0|\,|PX_1|\cdots|PX_{n-1}| = |-p^n - 1| = |p^n + 1|.$$

> ? **想一想** 为甚么 p 的正负号对奇数 n 有影响，对偶数 n 却没有影响?

若 n 为偶数，则无论 p 的正负号如何，都有 $p^n = |p|^n \geq 0$，所以 $|p^n + 1| = |p|^n + 1$，从而

$$|PX_0| \times |PX_1| \times \cdots \times |PX_{n-1}| = |OP|^n + 1,$$

这正是所求的结果。若 n 为奇数，则 p^n 与 p 同号，需分两种情况。

- 若 P 位于由 O 出发、经过 X_0X_1 中点的射线（ray）上（$p \geq 0$），则 $p^n = |OP|^n$，乘积为 $|OP|^n + 1$。

- 若 P 位于相反方向的射线上（$p < 0$），对奇数 n 而言，这是由 O 出发、经过顶点 $X_{(n+1)/2}$（位于 $\mathrm{e}^{\mathrm{i}\pi} = -1$）的射线，则 $p^n = -|OP|^n$，乘积为

$$\big|\,|OP|^n - 1\big|.$$

在第二种情况中，模的符号是不可或缺的：当 P 在圆内时，其值为 $1 - |OP|^n$；当 $P = X_{(n+1)/2}$ 时，其值为 0，这是理所当然的，因为此时其中一段距离为零。

第 (ii) 部分

这次旋转图形，使 $X_k = \omega^k$，因此 $X_0 = 1$。对 $z \neq 1$，因式分解给出

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

两端都是多项式，而且在无限多个点（每个 $z \neq 1$）上取值相同，所以它们对所有 z 都相等，包括 $z = 1$。代入 $z = 1$：

$$(1-\omega)(1-\omega^2)\cdots(1-\omega^{n-1}) = \underbrace{1 + 1 + \cdots + 1}_{n} = n.$$

取模，并利用 $|1 - \omega^k| = |X_0 X_k|$，

$$|X_0X_1| \times |X_0X_2| \times \cdots \times |X_0X_{n-1}| = n,$$

这正是所求的结果。

验算

- 正方形（$n = 4$），取 $P = O$：每段距离都是 1，所以乘积为 $1 = 0^4 + 1$。当 P 很远时，乘积的表现与 $|OP|^n$ 相若，与公式相符。
- 等边三角形（$n = 3$），P 在远侧且与 O 相距 2：各距离为 1（到位于 -1 的对面顶点）及两次 $\sqrt{(-2 - \frac{1}{2})^2 + \frac{3}{4}} = \sqrt{7}$，得出 $7 = 2^3 - 1$。
- 正方形的第 (ii) 部分：$|X_0X_1|\,|X_0X_2|\,|X_0X_3| = \sqrt{2} \times 2 \times \sqrt{2} = 4$。三个结果亦已用电脑就随机的 n、旋转及 P 的位置加以验证。

常见错误

- 在第 (i) 部分把 X_0 放在 1：这样 X_0X_1 的垂直平分线便是倾角为 π/n 的直线，而不是实轴，代数运算会更繁复。应按 P 的位置来选取坐标。
- 在奇数情况且 P 在远侧时忘记取模：当 P 在圆内时，$|OP|^n - 1$ 是负数。
- 直接把 $z = 1$ 代入 $\dfrac{z^n - 1}{z - 1}$；能够证明其值为 n 的，是那条多项式恒等式。

STEP 之外

第 (ii) 部分其实是说 $z^n - 1$ 在 $z = 1$ 处的导数（derivative）为 n：一般而言，若 $f(z) = \prod(z - a_k)$，则 $f'(a_0) = \prod_{k \neq 0}(a_0 - a_k)$。这类对单位根所取的乘积，出现于分圆多项式（cyclotomic polynomials）的理论及判别式（discriminant）的计算之中；而第 (i) 部分本质上是罗杰·科茨（Roger Cotes）的一个经典结果，涉及正多边形对称轴上一点到各顶点的距离。Needham 的 *Visual Complex Analysis*（《直观复分析》）是欣赏复数乘法这种几何意义如何发展的绝佳读物。

详细解答

STEP III 2016，第 8 题

纯数学 · 函数方程

STEP III 2016，第 8 题

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

step.maths.org/questions/16-s3-q8

关键思路

每条方程都把未知函数在 x 的值，与它在某个经变换后的点的值联系起来。如果再施行一次（或第三次）该变换就会回到 x，那么把变换后的点代入方程，便会得到同一组未知量的第二条方程。这样你便有了联立线性方程（simultaneous linear equations），可以像处理普通未知数一样，消去不需要的值。

解答

第 (i) 部分

方程对所有 x 都成立，因此把 x 换成 $-x$ 后仍然成立：

$$\mathrm{f}(-x) + (1+x)\,\mathrm{f}(x) = (-x)^2 = x^2,$$

这正是所求的结果。

现在我们有两条方程，含两个未知量 $\mathrm{f}(x)$ 和 $\mathrm{f}(-x)$。由第二条得 $\mathrm{f}(-x) = x^2 - (1+x)\mathrm{f}(x)$，代入原方程，得

$$\begin{aligned} \mathrm{f}(x) + (1-x)\big(x^2 - (1+x)\mathrm{f}(x)\big) &= x^2 \\ \mathrm{f}(x)\big(1 - (1-x^2)\big) &= x^2 - (1-x)x^2 \\ x^2\,\mathrm{f}(x) &= x^3. \end{aligned}$$

当 $x \neq 0$ 时，可以除以 x^2，得 $\mathrm{f}(x) = x$。当 $x = 0$ 时，原方程变为 $\mathrm{f}(0) + \mathrm{f}(0) = 0$，所以同样有 $\mathrm{f}(0) = 0$。因此

$$\mathrm{f}(x) = x \quad \text{对所有} x.$$

验证：$x+(1-x)(-x)=x-x+x^2=x^2$，符合要求。

想一想 为甚么题目坚持要我们验证答案?

我们的论证只说明了：如果有函数满足该方程，它必定是 $\mathrm{f}(x)=x$。它并没有说明这样的函数存在：运算过程中可能丢失了信息，那样的话便没有任何函数可行。代回原方程便填补了这个缺口。

第 (ii) 部分

首先注意，对每个 $x\neq 1$ 都有 $\mathrm{K}(x)\neq 1$，因为 $\frac{x+1}{x-1}=1$ 需要 $x+1=x-1$。所以 $\mathrm{K}(\mathrm{K}(x))$ 有定义；把分子分母同乘以 $x-1$，

$$\mathrm{K}(\mathrm{K}(x))=\frac{\dfrac{x+1}{x-1}+1}{\dfrac{x+1}{x-1}-1}=\frac{(x+1)+(x-1)}{(x+1)-(x-1)}=\frac{2x}{2}=x.$$

g 的方程是 $\mathrm{g}(x)+x\,\mathrm{g}(\mathrm{K}(x))=x$。由于 $\mathrm{K}(x)\neq 1$，我们可以把 x 换成 $\mathrm{K}(x)$；又因 $\mathrm{K}(\mathrm{K}(x))=x$，得

$$\mathrm{g}(\mathrm{K}(x))+\mathrm{K}(x)\,\mathrm{g}(x)=\mathrm{K}(x).$$

以 K 表示 $\mathrm{K}(x)$。于是 $\mathrm{g}(K)=K-K\,\mathrm{g}(x)$，代入原方程，

$$\mathrm{g}(x)+x\big(K-K\,\mathrm{g}(x)\big)=x \quad\Longrightarrow\quad (1-xK)\,\mathrm{g}(x)=x-xK.$$

而

$$1-xK=\frac{(x-1)-x(x+1)}{x-1}=-\frac{x^2+1}{x-1},$$
$$x-xK=x\cdot\frac{(x-1)-(x+1)}{x-1}=-\frac{2x}{x-1}.$$

由于 $x^2+1>0$，因子 $1-xK$ 永不为零，所以可以相除：

$$\mathrm{g}(x)=\frac{-2x/(x-1)}{-(x^2+1)/(x-1)}=\frac{2x}{x^2+1}\qquad (x\neq 1),$$

符合所求。(可以验证这个函数确实满足方程;例如在 $x=2$ 时,$\mathrm{K}(2)=3$,而 $\frac{4}{5}+2\cdot\frac{6}{10}=2$。)

第 (iii) 部分

设 $\mathrm{M}(x)=\dfrac{1}{1-x}$。方程为

$$\mathrm{h}(x)+\mathrm{h}(\mathrm{M}(x))=1-x-\mathrm{M}(x),$$

而右方只用 x 和 $\mathrm{M}(x)$ 表示，这是个好兆头。

这次 M 不是自身的逆（inverse），所以要看重复施行的结果。对 $x \neq 0, 1$：

$$\mathrm{M}(\mathrm{M}(x)) = \frac{1}{1-\frac{1}{1-x}} = \frac{1-x}{-x} = \frac{x-1}{x}, \qquad \mathrm{M}(\mathrm{M}(\mathrm{M}(x))) = \frac{1}{1-\frac{x-1}{x}} = \frac{x}{1} = x.$$

所以 M 的周期（period）为三。我们还需要 M 让我们留在定义域（domain）之内：$\mathrm{M}(x)$ 永不为 0，而 $\mathrm{M}(x) = 1$ 只在 $x = 0$ 时发生，因此若 $x \neq 0, 1$，则 $\mathrm{M}(x) \neq 0, 1$ 亦成立。

想一想 施行两次 M 不会回到 x。我们怎样才能得到足够的方程？

利用循环中的全部三个点。记 $a = x$，$b = \mathrm{M}(x)$，$c = \mathrm{M}(b) = \frac{x-1}{x}$，于是 $\mathrm{M}(c) = a$。依次在 a、b、c 应用方程：

$$\begin{aligned} \mathrm{h}(a) + \mathrm{h}(b) &= 1 - a - b, \\ \mathrm{h}(b) + \mathrm{h}(c) &= 1 - b - c, \\ \mathrm{h}(c) + \mathrm{h}(a) &= 1 - c - a. \end{aligned}$$

三式相加再除以二，

$$\mathrm{h}(a) + \mathrm{h}(b) + \mathrm{h}(c) = \tfrac{3}{2} - (a + b + c).$$

减去第二条方程，得

$$\mathrm{h}(a) = \tfrac{3}{2} - a - b - c - (1 - b - c) = \tfrac{1}{2} - a.$$

因此

$$\mathrm{h}(x) = \tfrac{1}{2} - x \qquad (x \neq 0,\ x \neq 1).$$

验证：$\left(\frac{1}{2} - x\right) + \left(\frac{1}{2} - \frac{1}{1-x}\right) = 1 - x - \frac{1}{1-x}$，符合要求。

验算

- 在 (i) 中，试 $x = 3$：$3 + (-2)(-3) = 9 = 3^2$。
- 在 (ii) 中，试 $x = -1$：$\mathrm{K}(-1) = 0$，而 $\mathrm{g}(-1) + (-1)\,\mathrm{g}(0) = -1 - 0 = -1$。此外 $\mathrm{K}(0) = -1$，证实了周期为二。
- 在 (iii) 中，试 $x = 2$：循环为 $2 \to -1 \to \frac{1}{2} \to 2$，而 $\mathrm{h}(2) + \mathrm{h}(-1) = -\frac{3}{2} + \frac{3}{2} = 0$，同时 $1 - 2 - (-1) = 0$。

三个函数亦已经由电脑以精确算术，在两千个随机有理数值上逐一检验每条方程，全部确认无误。

常见错误

- 在 (i) 中除以 x^2 时忘记了 $x = 0$，这情况需要另作（只需一行的）论证。
- 把 $\mathrm{K}(x)$ 或 $\mathrm{M}(x)$ 代入 x 时，没有检查它是否仍在方程成立的定义域之内。
- 在 (iii) 中只代入一次，结果卡在三个未知量上：你需要整个循环。

STEP 之外

映射 $x \mapsto -x$、K 和 M 都是莫比乌斯变换（Möbius transformations）$x \mapsto \frac{ax+b}{cx+d}$，它们在复合（composition）运算下构成一个群（group）。有限阶（finite order）的变换生成一个有限循环群（cyclic group），而这里的方法就是对该群求和（或消元），这个想法在群表示论（group representation theory）中会再次出现。莫比乌斯映射与 2×2 矩阵相对应，你可以通过把一个矩阵自乘三次，得出 $\mathrm{M}^3 = $ 恒等变换。Needham 的 *Visual Complex Analysis*（《直观复分析》）对莫比乌斯变换有精彩的论述。

详细解答

STEP III 2016，第 9 题

力学 · 弹簧与简谐运动

STEP III 2016，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/16-s3-q9

关键思路

把原点放在三角形的中心，并利用对称性（symmetry）。平衡位置（equilibrium position）就是中心，在那里三条相等的张力（tension）互相抵消。当质点沿着直线向 P 移动时，弹簧 QX 和 RX 始终互为镜像，因此它们的侧向拉力互相抵消，运动保持在该直线上。求 RX 需要用到余弦定律（cosine rule）。至于周期，我们先写出准确的恢复力（restoring force），然后只保留对小位移为线性的项。

解答

我们把 X 视为光滑桌面上的质点，把弹簧视为遵守虎克定律（Hooke's law）的轻弹簧，因此长度为 r 的弹簧的张力为 $\frac{\lambda}{l}(r-l)$。

平衡

设 O 为三角形的中心。每个顶点与 O 的距离为

$$c = \frac{2}{3} \times (\text{高}) = \frac{2}{3} \cdot \sqrt{3}\, a = \frac{2a}{\sqrt{3}}$$

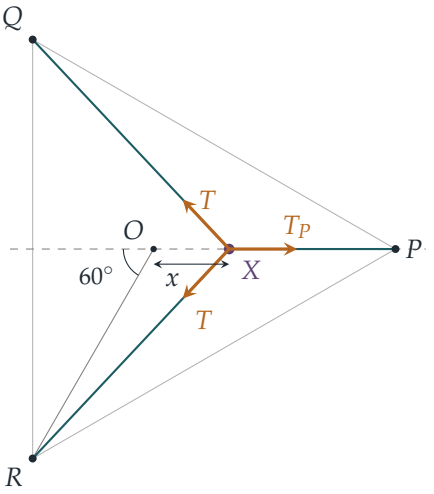
而直线 OP、OQ、OR 两两之间的夹角为 120°。当 X 在 O 时，每条弹簧的长度都是 c，所以三条张力相等。三个大小相等、两两夹角为 120° 的力的合力为零，所以 O 是一个平衡位置。我们使用的正是这个平衡位置。由于 $l < \frac{2}{\sqrt{3}}a$，每条弹簧都被拉伸，伸长量（extension）为

$$\frac{2}{\sqrt{3}}\, a - l,$$

得证。

RX 中的张力

取 x 轴沿 OP，以 P 的方向为正方向。设 X 在由 O 向 P 方向位移 x 之处。那么 $OX = x$、$OR = c$，而 $\angle XOR = 180° - 60° = 120°$。



在三角形 XOR 中应用余弦定律，

$$RX^2 = c^2 + x^2 - 2cx\cos 120° = c^2 + cx + x^2 = \frac{4a^2}{3} + \frac{2ax}{\sqrt{3}} + x^2.$$

因此由虎克定律，RX 中的张力为

$$T = \frac{\lambda}{l}\left(\sqrt{\frac{4a^2}{3} + \frac{2ax}{\sqrt{3}} + x^2} - l\right),$$

得证。由关于轴 OP 的对称性，$QX = RX$，所以 QX 的张力同样是 T。

运动方程

记 $r = RX$。弹簧 PX 的长度为 $c - x$，所以它的张力为 $T_P = \frac{\lambda}{l}(c - x - l)$，方向指向 P（即 $+x$ 方向）。

点 R 的 x 坐标为 $-c\cos 60° = -\frac{c}{2}$，所以由 X 到 R 的位移的 x 分量为 $-\left(x + \frac{c}{2}\right)$。因此沿 XR 的力 T 的 x 分量为 $-T\,\frac{x + c/2}{r}$。来自 QX 的力有相同的 x 分量，而两个垂直分量由对称性互相抵消。所以没有横越直线 OP 的力，X 沿该直线运动。

? **想一想** 怎样从这个平方根得出一条线性方程?

先写出准确的力，再对小 x 作展开。沿 OP 分解，

$$\begin{aligned} m\ddot{x} &= T_P - 2T\,\frac{x+c/2}{r} = \frac{\lambda}{l}\left[(c-x-l)-(2x+c)\left(1-\frac{l}{r}\right)\right] \\ &= \frac{\lambda}{l}\left[-3x-l+\frac{l(2x+c)}{r}\right]. \end{aligned}$$

这条方程是准确的。现在 $r^2 = c^2\left(1+\frac{x}{c}+\frac{x^2}{c^2}\right)$。利用二项式展开（binomial expansion），并舍去 x^2 及更高次的项，

$$\frac{1}{r} \approx \frac{1}{c}\left(1-\frac{x}{2c}\right), \qquad \frac{2x+c}{r} \approx \left(1+\frac{2x}{c}\right)\left(1-\frac{x}{2c}\right) \approx 1+\frac{3x}{2c}.$$

代入，

$$m\ddot{x} \approx \frac{\lambda}{l}\left[-3x-l+l+\frac{3lx}{2c}\right] = -\frac{3\lambda}{l}\left(1-\frac{l}{2c}\right)x.$$

由于 $\frac{l}{2c} = \frac{\sqrt{3}\,l}{4a}$，

$$\ddot{x} \approx -\omega^2 x, \qquad \omega^2 = \frac{3\lambda(4a-\sqrt{3}\,l)}{4mla}.$$

由于 $\sqrt{3}\,l < 2a < 4a$，这个 ω^2 是正数。所以这个力是恢复力，而对于小 x，质点绕 O 作近似的简谐运动（simple harmonic motion），周期为

$$\frac{2\pi}{\omega} = 2\pi\sqrt{\frac{4mla}{3(4a-\sqrt{3}\,l)\lambda}},$$

得证。

验算

- 当 $x = 0$ 时，准确的力为 $\frac{\lambda}{l}\left[-l+\frac{lc}{c}\right] = 0$，与最初求得的平衡位置一致。
- 单位：λ 是力，l、a 是长度，所以 $\frac{mla}{a\lambda}$ 的单位为 $\mathrm{kg\,m/N} = \mathrm{s}^2$，而周期是时间。
- 求 ω^2 的另一途径。对于小位移，一条弹簧因拉伸而贡献劲度（stiffness）$\frac{\lambda}{l}\cos^2\phi$，因转向而贡献 $\frac{T_0}{c}\sin^2\phi$，其中 ϕ 是弹簧与 OP 的夹角，$T_0 = \frac{\lambda}{l}(c-l)$ 是平衡时的张力。取 $\phi = 0°, 60°, 60°$，总劲度为 $\frac{3}{2}\cdot\frac{\lambda}{l} + \frac{3}{2}\cdot\frac{T_0}{c} = \frac{3\lambda}{l}\left(1-\frac{l}{2c}\right)$，与之前相同。
- 取 $a = \lambda = m = 1$ 及 $l = 0.8$，公式给出的周期约为 2.838。以电脑模拟完整的二维运动，从向 P 方向的小位移开始，得出的数值在四位有效数字内相同。

常见错误

- 用了 $\cos 120° = +\frac{1}{2}$，或在余弦定律中用错角。这会在平方根内得出 $-\frac{2ax}{\sqrt{3}}$。

- 把 QX 和 RX 中的张力当作常数。它们方向的改变和大小的改变都会贡献对 x 为线性的项。
- 求中心到顶点的距离时，把三角形的边长写成 a 而不是 $2a$。

STEP 之外

在任何稳定平衡位置附近，力对位移都近似为线性。这里的恢复力是弹性势能的梯度（gradient），而 $m\omega^2$ 是它的二阶导数（second derivative）。在二维情况下，这个二阶导数变成一个 2×2 矩阵（黑森矩阵（Hessian）），而它的特征向量（eigenvector）就是小振动的简正模（normal modes）。由于三重对称性，这个系统绕 O 在每个方向的振动周期都相同，而不只是沿 OP 的方向。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）讲述了小振动和耦合系统。

详细解答

STEP III 2016，第 10 题

力学 · 平面上的圆周运动

> **STEP III 2016，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/16-s3-q10

关键思路

只要绳子保持拉紧而质点留在斜面上，P 与 A 的距离恒为 a，且 P 位于斜面上，因此它在斜面内的一个圆上运动，圆心是由 A 向斜面所作垂线的垂足。在斜面之内，这不过是在减弱了的重力 $g\sin\alpha$ 下的铅垂圆周运动；在垂直于斜面的方向，绳子把 P 拉离斜面，而法向反作用力（normal reaction）必须保持为正。所以我们先求张力（tension），再求反作用力，然后问两者各自在何处最小。

解答

建立模型

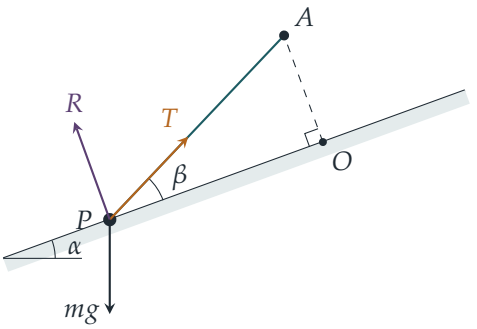
把 P 视为质点，绳子视为轻而不可伸长（inextensible），斜面视为光滑（smooth）。设 O 为由 A 向斜面所作垂线的垂足。由于绳长为 a，且与斜面成角 β，

$$AO = a\sin\beta, \qquad OP = a\cos\beta.$$

在平衡（equilibrium）状态下，除重力外，唯一有沿斜面分量的力是张力，因此绳子必须沿着一条最大斜率线（line of greatest slope），P 位于 O 沿斜面正下方。沿斜面及垂直于斜面方向分解力（张力 T_0，反作用力 R_0）：

$$T_0\cos\beta = mg\sin\alpha, \qquad R_0 + T_0\sin\beta = mg\cos\alpha,$$

所以 $R_0 = mg(\cos\alpha - \sin\alpha\tan\beta) = mg\cos(\alpha+\beta)/\cos\beta$。因此，平衡状态要存在，就必须有 $\cos(\alpha+\beta) > 0$。



上图是开始时经过一条最大斜率线的横切面。AO 垂直于斜面。

施加冲量（impulse）之后，只要绳子保持拉紧而 P 留在斜面上，就有 $AP = a$ 及 $AO = a\sin\beta$，所以由毕氏定理（Pythagoras′ theorem），$OP = a\cos\beta$ 恒成立：P 在斜面内一个以 O 为圆心、半径为

$$r = a\cos\beta$$

的圆上运动。绳子与斜面所成的角恒为 β，而绳子在斜面上的投影由 P 指向 O。冲量是水平的，且平行于斜面，因此它在这个圆的最低点与圆相切。

设 θ 为 OP 与经过 O 向下的最大斜率线所成的角，于是开始时 $\theta = 0$；又设 v 为速率。重量沿斜面向下（在斜面内）的分量为 $mg\sin\alpha$，垂直压入斜面的分量为 $mg\cos\alpha$。

垂直于斜面方向，P 没有加速度，所以

$$R + T\sin\beta = mg\cos\alpha. \tag{1}$$

在斜面内指向 O 的方向，加速度为 v^2/r。张力贡献 $T\cos\beta$，而沿斜面向下的重量分量贡献 $-mg\sin\alpha\cos\theta$：

$$T\cos\beta - mg\sin\alpha\cos\theta = \frac{mv^2}{a\cos\beta}. \tag{2}$$

能量：反作用力和张力都不做功（两者各自垂直于速度），而 P 沿斜面上升的距离为 $r(1-\cos\theta)$，即上升高度为 $r(1-\cos\theta)\sin\alpha$。因此

$$v^2 = u^2 - 2ga\cos\beta\sin\alpha\,(1-\cos\theta). \tag{3}$$

不会立即离开斜面

刚施加冲量后，$\theta = 0$ 且 $v = u$。由 (2) 得 $T\cos\beta = mg\sin\alpha + \dfrac{mu^2}{a\cos\beta}$，再由 (1) 得

$$\begin{aligned} R &= mg\cos\alpha - \tan\beta\left(mg\sin\alpha + \frac{mu^2}{a\cos\beta}\right) \\ &= \frac{m}{\cos\beta}\left(g\cos\alpha\cos\beta - g\sin\alpha\sin\beta\right) - \frac{mu^2\tan\beta}{a\cos\beta} \\ &= \frac{m}{a\cos\beta}\Big(ag\cos(\alpha+\beta) - u^2\tan\beta\Big). \end{aligned}$$

由于 $\cos\beta > 0$，若

$$ag\cos(\alpha+\beta) > u^2\tan\beta,$$

则反作用力为正，质点不会立即离开斜面，此即所求。

完整的圆周

现在把 (3) 代入 (2)，得出一般位置的张力：

$$T\cos\beta = \frac{mu^2}{a\cos\beta} - 2mg\sin\alpha + 3mg\sin\alpha\cos\theta,$$

于是由 (1)，

$$R = mg\cos\alpha - \tan\beta\left(\frac{mu^2}{a\cos\beta} - 2mg\sin\alpha + 3mg\sin\alpha\cos\theta\right).$$

想一想 质点最有可能在圆上哪处离开斜面：顶部还是底部?

从铅垂圆周运动得来的直觉会说“顶部”，但在这里，顶部是绳子的危险位置，而不是斜面的。绳子以 $T\sin\beta$ 的力把 P 拉离斜面，而张力在底部最大，因为 P 在那里速度最快，而斜面内的重力分量也把它拉离 O。事实上，R 在 $\cos\theta$ 最大时最小，即在 $\theta = 0$；而 T 在 $\cos\theta = -1$，即顶部时最小。所以要保持接触地走完整个圆周，我们需要两个条件。

- **底部保持接触**。在 $\theta = 0$ 时 $R > 0$，这正是第一部分的条件：$u^2\tan\beta < ag\cos(\alpha+\beta)$。
- **顶部绳子拉紧**。在 $\theta = \pi$ 时 $T \geq 0$：$\dfrac{u^2}{a\cos\beta} - 5g\sin\alpha \geq 0$，即 $u^2 \geq 5ag\sin\alpha\cos\beta$。（由 (3)，这同时保证顶部有 $v^2 > 0$。）

把第二个条件乘以 $\tan\beta > 0$，可见两者能同时成立，仅当

$$5ag\sin\alpha\cos\beta\tan\beta \leq u^2\tan\beta < ag\cos(\alpha+\beta),$$

即

$$5\sin\alpha\sin\beta < \cos\alpha\cos\beta - \sin\alpha\sin\beta \iff 6\sin\alpha\sin\beta < \cos\alpha\cos\beta.$$

两边除以 $\cos\alpha\cos\beta > 0$，得到必要条件

$$6\tan\alpha\tan\beta < 1.$$

（这条件也是充分的：若它成立，则任何满足 $5ag\sin\alpha\cos\beta \le u^2 < ag\cos(\alpha+\beta)/\tan\beta$ 的 u 都可行，因为此时 R 和 T 在圆上每一处都为正。）

验算

- 当 $u = 0$ 时，第一部分的反作用力变成 $mg\cos(\alpha+\beta)/\cos\beta$，与平衡值 R_0 吻合。
- 若 $\beta \to 0$（绳子几乎沿着斜面），绳子无法把质点提起，而条件 $6\tan\alpha\tan\beta < 1$ 自动成立，理应如此。
- 取 $\alpha = 30°$、$\beta = 5°$（于是 $6\tan\alpha\tan\beta \approx 0.30$）、$a = 1\,\mathrm{m}$ 及 $g = 9.8\,\mathrm{m\,s^{-2}}$，容许范围是 $24.4 \le u^2 < 91.8$（单位为 $\mathrm{m^2\,s^{-2}}$）。取 $u^2 = 50$，则最小反作用力约为 $0.37mg$，出现在底部；最小张力约为 $2.63mg$，出现在顶部。我们用电脑在三维空间中沿整个圆周求解牛顿第二定律（Newton's second law），证实 R 和 T 保持为正，数值与上述相符。

常见错误

- 把整条绳长 a 当作圆的半径：半径是 $a\cos\beta$。
- 以为质点最有可能在圆的顶部离开斜面。反作用力其实在底部最小。
- 在能量方程中用 g 而不是 $g\sin\alpha$：质点在铅垂方向只上升 $r(1-\cos\theta)\sin\alpha$。

STEP 之外

这是约束（constraints）下的运动：斜面和绳子各自提供一个反作用力，其大小恰好是维持约束所需的大小，而运动只有在这些力的正负号正确时才成立。在大学一年级的课程中，你会学到如何有系统地处理这类问题，就是利用拉格朗日力学（Lagrangian mechanics）消去约束，之后再求回反作用力。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）收录了许多正是这类“何时失去接触?”的问题。

详细解答

STEP III 2016，第 11 题

力学 · 功率与阻力

STEP III 2016，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：

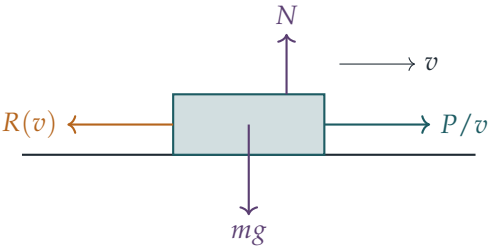
step.maths.org/questions/16-s3-q11

关键思路

以恒定功率（power）P 运作的引擎，在速率 v 时产生的牵引力为 P/v。由于我们要求的是以速率表示的距离，而非以时间表示，所以把加速度写成 $v\frac{\mathrm{d}v}{\mathrm{d}x}$。这样，牛顿第二定律（Newton's second law）就变成一条关于 x 与 v 的可分离变量微分方程（separable equation）。在 $4U$ 时加速度为零这个条件，可以把阻力常数以 P 和 U 表示出来，之后每段距离都归结为一个定积分。

解答

我们把汽车模拟为一个质量为 m 的质点，沿水平直线运动，向前的牵引力为 P/v，向后的阻力为 $R(v)$。在铅垂方向，重量与法向反作用力互相抵消，此后不再起任何作用。



利用 $a = v\frac{\mathrm{d}v}{\mathrm{d}x}$，水平方向的运动方程为

$$mv\frac{\mathrm{d}v}{\mathrm{d}x} = \frac{P}{v} - R(v), \qquad \text{所以} \qquad \frac{\mathrm{d}x}{\mathrm{d}v} = \frac{mv^2}{P - vR(v)}.$$

在两个模型中，$vR(v)$ 都随 v 递增，并在 $v = 4U$ 时等于 P。因此当 $U \le v \le 2U$ 时分母

为正，汽车在整个过程中确实一直在加速，即 x 随 v 递增，而所求距离为

$$X = \int_U^{2U} \frac{mv^2}{P - vR(v)}\,\mathrm{d}v.$$

第 (i) 部分

设 $R = kv$。在 $4U$ 时加速度为零，得 $\dfrac{P}{4U} = 4kU$，所以 $k = \dfrac{P}{16U^2}$。于是

$$P - kv^2 = k\,(16U^2 - v^2), \qquad \frac{m}{k} = \frac{16mU^2}{P} = \frac{1}{\lambda U}.$$

因此

$$\lambda X_1 = \frac{1}{U}\int_U^{2U} \frac{v^2}{16U^2 - v^2}\,\mathrm{d}v = \int_1^2 \frac{w^2}{16 - w^2}\,\mathrm{d}w,$$

其中代入了 $v = Uw$。把被积函数写成 $\dfrac{w^2}{16 - w^2} = -1 + \dfrac{16}{16 - w^2} = -1 + \dfrac{2}{4 - w} + \dfrac{2}{4 + w}$，得

$$\begin{aligned}\lambda X_1 &= \Big[-w + 2\ln\frac{4 + w}{4 - w}\Big]_1^2 \\ &= -1 + 2\ln 3 - 2\ln\tfrac{5}{3} = 2\ln\tfrac{9}{5} - 1,\end{aligned}$$

此即所求。

第 (ii) 部分

现设 $R = cv^2$。在 $4U$ 时加速度为零，得 $\dfrac{P}{4U} = 16cU^2$，所以 $c = \dfrac{P}{64U^3}$。于是

$$P - cv^3 = c\,(64U^3 - v^3), \qquad \frac{m}{c} = \frac{64mU^3}{P} = \frac{4}{\lambda}.$$

? **想一想** 这里有没有简洁的原函数？还是需要再用一次部分分式？

分子 v^2 除了相差一个常数倍外，正是分母 $64U^3 - v^3$ 的导数。所以这个积分可以直接得出对数：

$$\begin{aligned}\lambda X_2 &= 4\int_U^{2U} \frac{v^2}{64U^3 - v^3}\,\mathrm{d}v = 4\Big[-\tfrac{1}{3}\ln\big(64U^3 - v^3\big)\Big]_U^{2U} \\ &= \tfrac{4}{3}\ln\frac{63U^3}{56U^3} = \tfrac{4}{3}\ln\tfrac{9}{8},\end{aligned}$$

此即所求。

第 (iii) 部分

由于 $\lambda > 0$，$X_1 > X_2$ 当且仅当 $\lambda X_1 - \lambda X_2 > 0$。题目只给出 $\ln 24$ 和 $\ln 5$ 的界，所以我们把差改写成以这两者表示：

$$\begin{aligned}\lambda(X_1 - X_2) &= 2\ln 9 - 2\ln 5 - 1 - \tfrac{4}{3}\ln 9 + \tfrac{4}{3}\ln 8 \\ &= \tfrac{4}{3}\ln 3 + 4\ln 2 - 2\ln 5 - 1 \\ &= \tfrac{4}{3}(\ln 3 + 3\ln 2) - 2\ln 5 - 1 = \tfrac{4}{3}\ln 24 - 2\ln 5 - 1.\end{aligned}$$

想一想 每个界应该怎样用，是用上界还是下界？

要证明差为正，我们需要一个下界，所以对 $\ln 24$（带正号）用其下界，对 $\ln 5$（带负号）用其上界：

$$\lambda(X_1 - X_2) > \tfrac{4}{3}(3.17) - 2(1.61) - 1 = 4.22\dot{6} - 3.22 - 1 = 0.00\dot{6} > 0.$$

因此 $X_1 > X_2$：汽车由 U 加速至 $2U$ 期间，阻力与速率成正比时所走的距离，比阻力与速率平方成正比时更远。

这在物理上是合理的。两种阻力在 $v = 4U$ 时相等，但当 $v < 4U$ 时，有 $cv^2 = kv \cdot \frac{v}{4U} < kv$。所以在所考虑的速率范围内，平方定律的阻力较小，汽车加速得更猛，因而更快到达 $2U$。

验算

- 数值上，$\lambda X_1 = 2\ln 1.8 - 1 \approx 0.1756$，$\lambda X_2 = \frac{4}{3}\ln 1.125 \approx 0.1570$，所以差约为 0.0185，是正数，与所得结果一致。用电脑由 $v = U$ 开始按时间逐步模拟运动方程，可把两个数值重现至小数点后五位。
- 单位：$P/(mU^3)$ 的单位是 $\mathrm{W}/(\mathrm{kg\,m^3\,s^{-3}}) = \mathrm{m^{-1}}$，所以 λX 是无量纲的，理应如此。
- 第 (iii) 部分末尾的物理论证，独立地预测了同样的大小次序。

常见错误

- 使用 $a = \frac{\mathrm{d}v}{\mathrm{d}t}$，这样得到的是时间而不是距离。这里需要的是 $v\,\frac{\mathrm{d}v}{\mathrm{d}x}$。
- 在第 (iii) 部分把界用反了。用 3.18 和 1.60 只会得到一个上界，对差的正负号毫无证明作用。
- 忘记余量极小（$0.00\dot{6}$），以致把 $\frac{4}{3}(3.17)$ 四舍五入为 4.2，结果便证不出来。

STEP 之外

把加速度写成 $a = v\,\mathrm{d}v/\mathrm{d}x$，是迈向大学力学中能量方法（energy method）的第一步。把运动方程乘以 v，便得到动能的变化率，它等于输入功率减去耗散功率。像 $4U$ 这种两者平衡的速率，是一阶微分方程的平衡点（equilibria），而研究解如何趋近平衡点，正是动力系统（dynamical systems）的起点。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）便以这种风格讨论功、功率及与速度相关的力。

详细解答

STEP III 2016，第 12 题

概率与统计 · 切比雪夫不等式

STEP III 2016，第 12 题

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

step.maths.org/questions/16-s3-q12

关键思路

切比雪夫不等式（Chebyshev's inequality）给出的是远离均值（mean）的概率的上界，所以要证明某个量接近均值，就要从补事件（complement）入手。每一部分的工作都是：找出合适的随机变量（random variable），求出它的均值和标准差（standard deviation），再选取 k，使 $k\sigma$ 恰好等于我们关心的距离。在第 (ii) 部分，和式 $\sum n^r/r!$ 与 e^n 并列出现，正是泊松分布（Poisson distribution）的标志。

解答

第 (i) 部分

设 X 为 $100n$ 次抛掷中得到正面的次数。假设各次抛掷互相独立（independent），而每次得到正面的概率为 0.2，则 $X \sim \mathrm{B}(100n, 0.2)$，因此

$$\mu = 100n \times 0.2 = 20n, \qquad \sigma^2 = 100n \times 0.2 \times 0.8 = 16n, \qquad \sigma = 4\sqrt{n}.$$

事件 $16n \le X \le 24n$ 正好就是 $|X - 20n| \le 4n$，所以

$$\alpha = \mathrm{P}\big(|X-\mu| \le 4n\big) = 1 - \mathrm{P}\big(|X-\mu| > 4n\big).$$

我们希望 $k\sigma = 4n$，即 $k \cdot 4\sqrt{n} = 4n$，故 $k = \sqrt{n}$，这是正数。由切比雪夫不等式得

$$\mathrm{P}\big(|X-\mu| > 4n\big) = \mathrm{P}\big(|X-\mu| > \sqrt{n}\,\sigma\big) \le \frac{1}{(\sqrt{n})^2} = \frac{1}{n},$$

因此

$$\alpha \ge 1 - \frac{1}{n},$$

得证。（条件 $n > 1$ 只是为了让这个界有实质内容：当 $n = 1$ 时，它只说 $\alpha \ge 0$。）

第 (ii) 部分

想一想 哪一个随机变量的概率形如 $n^r/r!$?

若 $Y \sim \mathrm{Po}(n)$，则对 $r = 0, 1, 2, \ldots$ 有 $\mathrm{P}(Y = r) = \mathrm{e}^{-n} n^r / r!$，所以

$$\mathrm{e}^{-n}\left(1 + n + \frac{n^2}{2!} + \cdots + \frac{n^{2n}}{(2n)!}\right) = \mathrm{P}(Y \le 2n).$$

所求不等式两边除以 $\mathrm{e}^n > 0$ 后，便等价于

$$\mathrm{P}(Y \le 2n) \ge 1 - \frac{1}{n}.$$

泊松分布的均值和方差（variance）都是 n，所以 $\mu = n$，$\sigma = \sqrt{n}$。

想一想 切比雪夫不等式是双侧的，但 $Y \le 2n$ 是单侧事件。两者如何对应起来?

由于 $Y \ge 0$，必有 $Y - n \ge -n$，所以 $Y - n < -n$ 不可能发生。因此

$$|Y - n| > n \iff Y - n > n \iff Y > 2n.$$

取 $k = \sqrt{n}$，使 $k\sigma = n$，由切比雪夫不等式得

$$\mathrm{P}(Y > 2n) = \mathrm{P}\left(|Y - n| > \sqrt{n}\,\sigma\right) \le \frac{1}{n},$$

所以 $\mathrm{P}(Y \le 2n) = 1 - \mathrm{P}(Y > 2n) \ge 1 - \frac{1}{n}$。两边乘以 e^n，

$$1 + n + \frac{n^2}{2!} + \cdots + \frac{n^{2n}}{(2n)!} \ge \left(1 - \frac{1}{n}\right)\mathrm{e}^n,$$

这正是所求结果。

验算

这些界是正确的，但相当宽松。当 $n = 2$ 时，第 (i) 部分的准确二项分布（binomial distribution）概率为 $\alpha \approx 0.868$，而界只是 0.5；当 $n = 10$ 时，$\alpha \approx 0.9986$，而界是 0.9。在第 (ii) 部分取 $n = 2$，左边为 $1 + 2 + 2 + \frac{4}{3} + \frac{2}{3} = 7$，而 $\frac{1}{2}\mathrm{e}^2 \approx 3.69$。我们已就第 (i) 部分中 n 由 2 至 30、第 (ii) 部分中 n 由 1 至 59 的每一个值，准确验证了两个不等式。

常见错误

- 选取 k 时，误用方差 $16n$ 代替标准差 $4\sqrt{n}$。
- 忘记切比雪夫不等式给出的是落在区间以外的概率的上界，因此要用补事件才能得到 α。
- 在第 (ii) 部分没有解释为何下尾 $Y < 0$ 没有任何贡献。

STEP 之外

切比雪夫不等式是集中不等式（concentration inequality）最简单的例子；把它应用于 m 次独立试验的平均值，便得到弱大数定律（weak law of large numbers）：离散程度以 $1/\sqrt{m}$ 的速度收窄。第 (ii) 部分还告诉我们，指数级数（exponential series）在 $n^{2n}/(2n)!$ 之后的尾项之和至多为 e^n/n——这是一个用概率证明的分析学事实。更精细的工具，例如切尔诺夫界（Chernoff bounds），可以证明真正的概率与 1 的差距是指数级地小。如欲深入了解，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率论导引》）。

详细解答

STEP III 2016，第 13 题

概率与统计 · 和的峰度

STEP III 2016，第 13 题

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

step.maths.org/questions/16-s3-q13

关键思路

峰度（kurtosis）只由偏离平均值的差构成，所以任何平移变量的操作都不会改变它；这让我们可以把每个变量都中心化至 0，然后直接处理普通的矩（moments）。对于正态分布（Normal distribution），四阶矩可用分部积分法（integration by parts）求得，方法是分出一个 x 因子与指数函数配对。对于若干个互相独立（independent）且平均值为零的变量之和，展开 T^4 再取期望值，凡是含有某个变量的一次方的项都会变成零。这就是第 (ii) 部分的全部内容，而它会直接用于第 (iii) 部分。

解答

平移。 设 $W = X - a$。由期望值的线性性质（linearity of expectation），$\mathrm{E}(W) = \mu - a$，所以

$$W - \mathrm{E}(W) = (X - a) - (\mu - a) = X - \mu.$$

W 偏离其平均值的差，正好等于 X 偏离其平均值的差。因此 W 有相同的方差 $\mathrm{E}\big((X-\mu)^2\big) = \sigma^2$、相同的四阶中心矩（central moment）$\mathrm{E}\big((X-\mu)^4\big)$，从而有相同的峰度。特别地，取 $a = \mu$，则每个随机变量都与其中心化版本 $X - \mu$ 有相同的峰度，而后者的平均值为 0。

第 (i) 部分

设 $X \sim \mathrm{N}(0, \sigma^2)$，其概率密度函数（probability density function）为

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-x^2/(2\sigma^2)}, \qquad \text{因此} \qquad f'(x) = -\frac{x}{\sigma^2}\,f(x).$$

由于 $\mu = 0$，我们需要求 $\mathrm{E}(X^4) = \int_{-\infty}^{\infty} x^4 f(x)\,\mathrm{d}x$。

想一想 既然 $\mathrm{e}^{-x^2/(2\sigma^2)}$ 没有初等的原函数，我们怎样积分 $x^4\mathrm{e}^{-x^2/(2\sigma^2)}$?

函数 $xf(x)$ 确实有原函数（antiderivative），就是 $-\sigma^2 f(x)$。所以写成 $x^4 f(x) = x^3 \cdot xf(x)$，并取 $u = x^3$ 及 $\mathrm{d}v = xf(x)\,\mathrm{d}x$、$v = -\sigma^2 f(x)$ 进行分部积分：

$$\int_{-\infty}^{\infty} x^4 f(x)\,\mathrm{d}x = \Big[-\sigma^2 x^3 f(x)\Big]_{-\infty}^{\infty} + 3\sigma^2 \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = 3\sigma^2 \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x.$$

边界项为零，因为当 $x \to \pm\infty$ 时 $x^3\mathrm{e}^{-x^2/(2\sigma^2)} \to 0$（指数函数衰减的速度比任何幂的增长都快）。以 $u = x$ 重复同一步骤，得

$$\int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \Big[-\sigma^2 x f(x)\Big]_{-\infty}^{\infty} + \sigma^2 \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \sigma^2,$$

因为 f 是概率密度函数，其积分为 1。（这也证实了方差是 σ^2。）因此 $\mathrm{E}(X^4) = 3\sigma^2 \cdot \sigma^2 = 3\sigma^4$，而

$$\kappa = \frac{3\sigma^4}{\sigma^4} - 3 = 0.$$

由平移的结果，每个正态分布，无论其平均值是多少，峰度都是 0。

第 (ii) 部分

用多项式定理（multinomial theorem）展开 $T^4 = (Y_1 + \cdots + Y_n)^4$。每一项都是四个因子 Y_r 的乘积，而按照这四个因子如何分配到不同的下标，各项可分为五类：

$$Y_r^4, \qquad 4\,Y_r^3 Y_s, \qquad 6\,Y_r^2 Y_s^2, \qquad 12\,Y_r^2 Y_s Y_t, \qquad 24\,Y_r Y_s Y_t Y_u,$$

其中不同的字母代表不同的下标。系数计算的是排列方式的数目：例如，$Y_r^2 Y_s^2$ 来自从 4 个括号中选出哪 2 个提供 Y_r，共有 $\binom{4}{2} = 6$ 种方法。每个无序对 $\{r, s\}$ 在 $Y_r^2 Y_s^2$ 这一类中只出现一次。

现在取期望值。各个 Y 互相独立，所以不同 Y 的幂的乘积的期望值，等于各自期望值的乘积。在 $Y_r^3 Y_s$、$Y_r^2 Y_s Y_t$ 和 $Y_r Y_s Y_t Y_u$ 这几类中，总有某个变量只以一次方出现，而它的因子 $\mathrm{E}(Y_s) = 0$ 令整个期望值为零。例如，

$$\mathrm{E}(Y_r^3 Y_s) = \mathrm{E}(Y_r^3)\,\mathrm{E}(Y_s) = 0.$$

只有第一类和第三类保留下来，而同样由独立性，当 $r \neq s$ 时 $\mathrm{E}(Y_r^2 Y_s^2) = \mathrm{E}(Y_r^2)\,\mathrm{E}(Y_s^2)$。把各对写成 $r < s$，得

$$\mathrm{E}(T^4) = \sum_{r=1}^{n} \mathrm{E}(Y_r^4) + 6 \sum_{r=1}^{n-1} \sum_{s=r+1}^{n} \mathrm{E}(Y_r^2)\,\mathrm{E}(Y_s^2),$$

此即所求。（注意这部分用到了独立性和平均值为零，但没有用到各个 Y 同分布（identically distributed）。）

第 (iii) 部分

设每个 X_r 的平均值为 μ、标准差为 σ，并设 $S = X_1 + \cdots + X_n$。令 $Y_r = X_r - \mu$。各个 Y_r 互相独立且同分布，平均值为 0，而

$$\mathrm{E}(Y_r^2) = \sigma^2, \qquad \mathrm{E}(Y_r^4) = (\kappa + 3)\sigma^4,$$

第二式由移项 κ 的定义得出。于是 $T = \sum Y_r = S - n\mu$。由平移的结果，S 与 T 有相同的峰度，所以我们处理平均值为 0 的 T。

由独立性，$\mathrm{Var}(T) = \sum \mathrm{Var}(Y_r) = n\sigma^2$。第 (ii) 部分中的双重和共有 $\binom{n}{2} = \frac{1}{2}n(n-1)$ 项，每项都等于 σ^4，所以

$$\begin{aligned}\mathrm{E}(T^4) &= n(\kappa+3)\sigma^4 + 6\cdot\tfrac{1}{2}n(n-1)\sigma^4 \\ &= n\sigma^4\big(\kappa + 3 + 3(n-1)\big) = n\sigma^4(\kappa + 3n).\end{aligned}$$

因此 T 的峰度，亦即 S 的峰度，为

$$\frac{\mathrm{E}(T^4)}{(\mathrm{Var}\,T)^2} - 3 = \frac{n\sigma^4(\kappa+3n)}{n^2\sigma^4} - 3 = \frac{\kappa}{n} + 3 - 3 = \frac{\kappa}{n},$$

此即所求。

验算

- 取 $Y = \pm 1$，概率各为 $\frac{1}{2}$。则 $\mathrm{E}(Y^4) = \sigma^4 = 1$，所以 $\kappa = 1 - 3 = -2$。当 $n = 2$ 时，和的取值为 $-2, 0, 2$，概率分别为 $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$，所以 $\mathrm{E}(T^4) = 8$ 及 $\mathrm{Var}\,T = 2$。由此得峰度为 $8/4 - 3 = -1 = \kappa/2$，与预测相符。
- 当 $n = 1$ 时，第 (iii) 部分表示甚么都不变，理应如此。
- 由于互相独立的正态变量之和仍是正态变量，第 (iii) 部分取 $\kappa = 0$ 时必须再次得出 0，而事实确是如此。
- 我们用电脑以精确分数验证了：一个取三个值的变量，其最多六个独立副本之和的峰度恰好是 κ/n。我们亦以数值方法验证了，对多个 σ 值，$\int x^4 f(x)\,\mathrm{d}x = 3\sigma^4$。

常见错误

- 弄错 $Y_r^2 Y_s^2$ 的系数：每个无序对的系数是 $\binom{4}{2} = 6$（或每个有序对 3），不是 4 或 12。
- 在使用第 (ii) 部分之前忘记把 X_r 中心化，而第 (ii) 部分要求平均值为 0。
- 除以 σ^4 而不是 $(n\sigma^2)^2$：和的方差是 $n\sigma^2$。

STEP 之外

这里定义的峰度，是四阶累积量（cumulant）除以 σ^4。累积量是矩母函数（moment generating function）的对数展开式中的系数，而对于独立变量之和，它们直接相加。这就是为甚么和的标准化四阶累积量会像 $1/n$ 那样缩小，而三阶累积量（偏度（skewness））则像 $1/\sqrt{n}$ 那样缩小。这让我们定量地窥见中心极限定理（central limit theorem）：和看起来愈来愈像正态分布，而正态分布二阶以上的累积量全部为零。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）在这个程度上讲解生成函数和中心极限定理。

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

2016 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。

