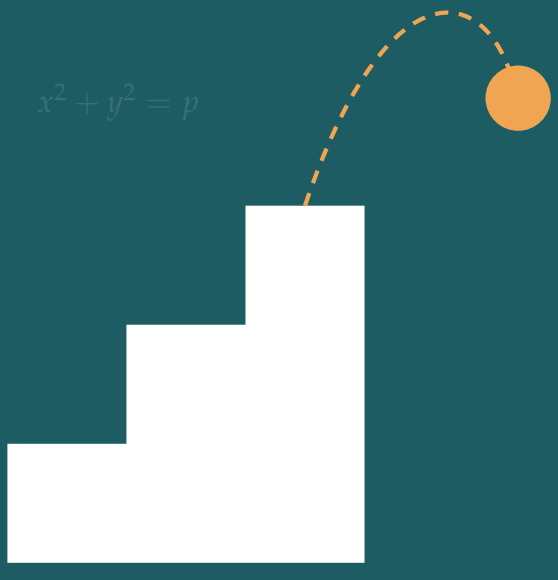
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2003 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

2003 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2003 年详细解答

2003 年 STEP 试卷（STEP I, II 及 III）全部 42 题的详细解答。自样本试卷以来每一道 STEP 题目的详细解答、*Beyond STEP* 全书以及各章的单行本，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

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

由机器从英文翻译而成；尚未经审阅。

2026 年 11 月首次于网上发表。

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

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

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

目录

STEP I

第 1 题 幂和 纯数学 . . . 3

第 2 题 代数恒等式 纯数学 . . . 7

第 3 题 半角三角方程 纯数学 . . . 10

第 4 题 三角不等式 纯数学 . . . 14

第 5 题 二项式级数 纯数学 . . . 17

第 6 题 以部分分式积分 纯数学 . . . 20

第 7 题 数字与整除性 纯数学 . . . 23

第 8 题 逻辑斯谛微分方程 纯数学 . . . 26

第 9 题 通过定点的抛体 力学 . . . 29

第 10 题 薄片的质心 力学 . . . 33

第 11 题 斜面上的碰撞 力学 . . . 37

第 12 题 乘积的期望值 概率与统计 . . . 41

第 13 题 几何分布的期望值 概率与统计 . . . 44

第 14 题 泊松分布 概率与统计 . . . 47

STEP II

第 1 题 联立线性方程 纯数学 . . . 50

第 2 题 反三角函数 纯数学 . . . 54

第 3 题 无理数与数学归纳法 纯数学 . . . 57

第 4 题 弓形的面积 纯数学 . . . 60

第 5 题 直线的向量几何 纯数学 . . . 64

第 6 题 迭代函数与积分 纯数学 . . . 67

第 7 题 广义积分与级数 纯数学 . . . 70

第 8 题 可分离变量微分方程 纯数学 . . . 73

第 9 题 极限摩擦力平衡 力学 77
第 10 题 涉及摩擦力的约束运动 力学 81
第 11 题 抛体运动 力学 85
第 12 题 分支过程 概率与统计 89
第 13 题 截尾泊松分布 概率与统计 93
第 14 题 假设检验的错误 概率与统计 96

STEP III

第 1 题 反双曲函数积分法 纯数学 100
第 2 题 二项式级数 纯数学 104
第 3 题 曲线描绘 纯数学 107
第 4 题 参数曲线的切线 纯数学 111
第 5 题 三次方程的根 纯数学 115
第 6 题 三角函数裂项相消求和 纯数学 120
第 7 题 西姆松线 纯数学 124
第 8 题 一阶微分方程 纯数学 128
第 9 题 平衡的稳定性 力学 132
第 10 题 变加速度 力学 136
第 11 题 截运动学 力学 140
第 12 题 指数分布的等候时间 概率与统计 144
第 13 题 吸收马可夫链 概率与统计 148
第 14 题 概率生成函数 概率与统计 152

详细解答

STEP I 2003，第 1 题

纯数学 · 幂和

> **STEP I 2003，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/03-s1-q1

关键思路

题目已经告诉我们答案的形式：一个以 n 为变量、系数未知的多项式（polynomial），所以余下的工作只是把这些系数确定下来。一个对所有 n 都成立的多项式恒等式（identity），对几个方便计算的 n 值当然也成立，而每个值都给出一条线性方程（linear equation）。让求和由负数下标开始是刻意安排的技巧：这样我们便可以使用 $n=-1$ 和 $n=-2$ 等较小的值，令方程组呈现悦目的对称性。

解答

题目中 r 既用作求和变量，又用作系数。为免混淆，我们把求和变量写成 k，而 p,q,r,s 和 a,b,c,d,e 则保留作系数。

平方和

设 $S(n)=\sum_{k=-1}^{n} k^2$，于是

$$S(n)=pn^3+qn^2+rn+s.$$

由 $k=-1$ 开始的和对所有 $n\geq -1$ 都有意义，而且

$$S(-1)=1,\qquad S(0)=1+0=1,\qquad S(1)=1+0+1=2,\qquad S(2)=1+0+1+4=6.$$

把 $n=-1,0,1,2$ 代入题目给出的形式，得到四条方程

$$-p+q-r+s=1, \tag{1}$$
$$s=1, \tag{2}$$
$$p+q+r+s=2, \tag{3}$$
$$8p+4q+2r+s=6. \tag{4}$$

由 (2) 得 $s=1$，所以 (1) 和 (3) 变成 $-p+q-r=0$ 和 $p+q+r=1$。两式相加得 $2q=1$，所以 $q=\frac{1}{2}$，从而 $p+r=\frac{1}{2}$。方程 (4) 变成

$$8p+2+2r+1=6 \quad\Longrightarrow\quad 8p+2r=3 \quad\Longrightarrow\quad 6p+2(p+r)=3,$$

所以 $6p=2$，得 $p=\frac{1}{3}$ 和 $r=\frac{1}{2}-\frac{1}{3}=\frac{1}{6}$。因此

$$\sum_{k=-1}^{n} k^2=\tfrac{1}{3}n^3+\tfrac{1}{2}n^2+\tfrac{1}{6}n+1.$$

想一想 题目要求的是由 $k=0$ 开始的和，但我们求得的是由 $k=-1$ 开始的和。两者有何分别?

两者只相差 $k=-1$ 那一项，即 $(-1)^2=1$。所以

$$\sum_{k=0}^{n} k^2=\tfrac{1}{3}n^3+\tfrac{1}{2}n^2+\tfrac{1}{6}n=\tfrac{1}{6}n\left(2n^2+3n+1\right)=\tfrac{1}{6}n(n+1)(2n+1),$$

这正是所要证明的结果。

立方和

现在设 $T(n)=\sum_{k=-2}^{n}k^3=an^4+bn^3+cn^2+dn+e$。这里有五个未知数，所以我们取 $n=-2,-1,0,1,2$ 这五个值:

$$T(-2)=-8,\quad T(-1)=-9,\quad T(0)=-9,\quad T(1)=-8,\quad T(2)=0.$$

由此得到

$$16a-8b+4c-2d+e=-8, \tag{5}$$
$$a-b+c-d+e=-9, \tag{6}$$
$$e=-9, \tag{7}$$
$$a+b+c+d+e=-8, \tag{8}$$
$$16a+8b+4c+2d+e=0. \tag{9}$$

由 (7) 得 $e=-9$。这些方程按 $\pm n$ 成对出现，所以把每一对相加和相减最为划算：偶次幂和奇次幂便会分开。

利用 $e=-9$，(6) 和 (8) 变成 $a-b+c-d=0$ 和 $a+b+c+d=1$。相加和相减，得

$$a+c=\tfrac{1}{2}, \qquad b+d=\tfrac{1}{2}.$$

同样地，(5) 和 (9) 变成 $16a-8b+4c-2d=1$ 和 $16a+8b+4c+2d=9$。相加和相减，得

$$32a+8c=10, \qquad 16b+4d=8,$$

即 $16a+4c=5$ 和 $4b+d=2$。代入 $c=\frac{1}{2}-a$ 得 $12a+2=5$，所以 $a=\frac{1}{4}$ 和 $c=\frac{1}{4}$。代入 $d=\frac{1}{2}-b$ 得 $3b+\frac{1}{2}=2$，所以 $b=\frac{1}{2}$ 和 $d=0$。因此

$$\sum_{k=-2}^{n} k^3=\tfrac{1}{4}n^4+\tfrac{1}{2}n^3+\tfrac{1}{4}n^2-9.$$

与由 $k=0$ 开始的和相比，由 $k=-2$ 开始的和多了 $(-2)^3+(-1)^3=-9$ 这些项，所以我们加上 9：

$$\sum_{k=0}^{n} k^3=\tfrac{1}{4}n^4+\tfrac{1}{2}n^3+\tfrac{1}{4}n^2=\tfrac{1}{4}n^2\left(n^2+2n+1\right)=\tfrac{1}{4}n^2(n+1)^2,$$

这正是所要证明的结果。

验算

当 $n=10$ 时，两条公式给出 $\frac{1}{6}\cdot 10\cdot 11\cdot 21=385$ 和 $\frac{1}{4}\cdot 100\cdot 121=3025$，而直接把 $0^2+1^2+\cdots+10^2$ 和 $0^3+1^3+\cdots+10^3$ 加起来，也得到相同的数字。还可留意 $3025=55^2=(0+1+\cdots+10)^2$：首 n 个立方数之和，等于首 n 个整数之和的平方。我们亦用电脑核对过，两条公式对由 0 至 59 的每个 n 都准确成立。

常见错误

- 由我们算出的和转到题目要求的和时，忘记了 $k=-1$（或 $k=-2$ 和 $k=-1$）的项。注意 $(-2)^3+(-1)^3=-9$，而不是 $+9$。
- 处理立方和时只打算用四个 n 值：五个未知数需要五条方程。
- 把系数 r 与题目中的求和变量 r 混淆；把其中一个改名便可避免出错。

STEP 之外

为甚么我们可以假设答案是一个次数高一次的多项式？因为一个 $m+1$ 次多项式的差分 $P(n)-P(n-1)$ 是 m 次的，所以总可以解出一个差分为 n^m 的多项式；这就是有限差分演算（calculus of finite differences）的起点，而一般情况下出现的系数由伯努利数（Bernoulli numbers）决定，即福尔哈伯公式（Faulhaber's formula）。这里所用的方法，即凭多项式在若干点上的值把它确定下来，称为多项式插值（polynomial interpolation），它在数值分析（numerical analysis）中会以拉格朗日插值（Lagrange interpolation）的形式再度出现。想做更多同类风格的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

STEP I 2003，第 2 题

纯数学 · 代数恒等式

STEP I 2003，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/03-s1-q2

关键思路

考生的答案正确，当且仅当它满足方程，所以把它代入（substitute）方程，再消去分母。这样每一部分都会化为一个关于已知数的多项式（polynomial）条件。在第一部分，这个条件是一个二次型（quadratic form），对非零实数而言不可能等于零。在第二部分，它是一个三次式，可以漂亮地因式分解（factorise），而所得的因式正正就是题目问及的那三条等式。

解答

第一问

在整个解答中，a 和 b 都是非零实数。考生的答案正确，恰好就是 $x = a + b$ 满足方程的时候。要令这句话有意义，首先需要 $a + b \neq 0$，然后需要

$$\frac{1}{a+b} = \frac{1}{a} + \frac{1}{b} = \frac{a+b}{ab}.$$

两边同乘以非零的 $ab(a+b)$，这等价于

$$ab = (a+b)^2, \qquad \text{即} \qquad a^2 + ab + b^2 = 0.$$

配方（completing the square）得

$$a^2 + ab + b^2 = \left(a + \tfrac{1}{2}b\right)^2 + \tfrac{3}{4}b^2.$$

两项都是实数的平方，所以都 ≥ 0，而且因为 $b \neq 0$，第二项严格大于零。因此对任意实数 a 和任意非零实数 b，都有 $a^2 + ab + b^2 > 0$，所以这个条件永远不可能成立。

（若 $a+b=0$，考生的“答案”$x=0$ 代入方程后根本没有意义，所以这个情况也好不到哪里去。）因此不存在任何 a 和 b 的值使 $x=a+b$ 正确，这正是所要证明的。

第二问

现在 a、b、c 都非零。答案 $x=a+b+c$ 正确，当且仅当（if and only if）$a+b+c\neq 0$ 并且

$$\frac{1}{a+b+c}=\frac{1}{a}+\frac{1}{b}+\frac{1}{c}=\frac{bc+ca+ab}{abc}.$$

两边同乘以 $abc(a+b+c)\neq 0$，这等价于

$$(a+b+c)(ab+bc+ca)-abc=0. \qquad (*)$$

想一想 怎样才能看出 $(*)$ 的左边可以因式分解?

要证明的结果提示我们, 因式应该是 $a+b$、$b+c$ 和 $c+a$。先试其中一个: 设 $b=-a$。这时 $a+b+c=c$,而 $ab+bc+ca=-a^2-ac+ca=-a^2$,所以左边变成 $c(-a^2)-(-a^2c)=0$。因此 $a+b$ 是一个因式，而根据对称性（symmetry），$b+c$ 和 $c+a$ 也是因式。左边的次数是 3，所以它是 $(a+b)(b+c)(c+a)$ 的常数倍；比较 a^2b 的系数（两边都是 1），可知这个常数是 1。另一个做法是把两边都展开:

$$\begin{aligned}(a+b+c)(ab+bc+ca)-abc&=a^2b+a^2c+b^2a+b^2c+c^2a+c^2b+3abc-abc\\&=a^2b+a^2c+b^2a+b^2c+c^2a+c^2b+2abc\\&=(a+b)(b+c)(c+a).\end{aligned}$$

所以 $(*)$ 等同于

$$(a+b)(b+c)(c+a)=0.$$

现在我们证明两个方向。

若答案正确。那么 $a+b+c\neq 0$ 而且 $(*)$ 成立，所以 $(a+b)(b+c)(c+a)=0$。若干个数的乘积为零，其中必有一个因数为零，所以 $a+b=0$、$b+c=0$、$c+a=0$ 之中至少有一条成立。

反过来，假设这三条等式中至少有一条成立；根据对称性，不妨设 $a+b=0$。那么 $a+b+c=c\neq 0$，所以考生的答案是一个合理的非零数，而且 $(*)$ 的因式分解形式成立。由于从方程推到 $(*)$ 的每一步都是可逆的（我们只是乘以非零的数 $abc(a+b+c)$），所以 $x=a+b+c$ 满足方程。这一点也可以直接看出：$\frac{1}{a}+\frac{1}{b}=\frac{1}{a}-\frac{1}{a}=0$，所以方程变成 $\frac{1}{x}=\frac{1}{c}$，它的解是 $x=c=a+b+c$。

因此，考生的答案正确，当且仅当 $a+b=0$、$b+c=0$、$c+a=0$ 之中至少有一条成立。

■

留意这个论证从来没有用到 a、b、c 是实数；这个因式分解是一个代数恒等式（identity），所以结果对非零复数（complex numbers）同样成立。相比之下，第一部分确实需要 a 和 b 是实数：在复数范围内，$a^2+ab+b^2=0$ 是有解的，例如 $b=a\,\mathrm{e}^{2\pi\mathrm{i}/3}$。

验算

- 取 $a=2$、$b=-2$、$c=5$：这时 $\frac{1}{2}-\frac{1}{2}+\frac{1}{5}=\frac{1}{5}$，而 $a+b+c=5$，所以考生是对的。
- 取 $a=b=c=1$：正确答案是 $x=\frac{1}{3}$，但考生的答案是 3；而事实上 $(a+b)(b+c)(c+a)=8\neq 0$。
- 在 $a=1$、$b=2$、$c=3$ 时展开 $(a+b)(b+c)(c+a)$：得 $3\cdot 5\cdot 4=60$，而 $(a+b+c)(ab+bc+ca)-abc=6\cdot 11-6=60$。
- 我们用精确的分数运算，对全部 17 576 个由非零有理数 n/d（其中 $|n|\leq 6$ 且 $d\in\{1,2,3\}$）组成的三元组，穷举检验了两个方向，没有发现任何例外。

常见错误

- 写下 $a^2+ab+b^2=(a+b)^2-ab$ 便停笔：这并不能说明它是正数。配方才可以。
- 在第二部分只证明一个方向。“当且仅当”需要证明两个方向，而且逆命题（converse）必须检查 $a+b+c\neq 0$。
- 忘记在两边同乘之前，必须先确定 $abc(a+b+c)$ 非零；正是这一点令各步可逆。

STEP 之外

表达式 $(a+b)(b+c)(c+a)$ 是一个对称多项式（symmetric polynomial），而恒等式 $(a+b)(b+c)(c+a)=e_1e_2-e_3$（其中 $e_1=a+b+c$、$e_2=ab+bc+ca$、$e_3=abc$）正是对称多项式基本定理（fundamental theorem of symmetric polynomials）的一个例子：每个对称多项式都可以写成这些基本对称多项式的多项式。除了正负号之外，这些 e_1、e_2、e_3 也就是以 a、b、c 为根的三次多项式的系数，这正是它们在方程论（theory of equations）中随处可见的原因。Allenby 的 *Rings, Fields and Groups* 本着这种精神讨论多项式环（polynomial rings），而 Siklos 的 *Advanced Problems in Mathematics* 则有更多辨认因式分解的 STEP 练习。

详细解答

STEP I 2003，第 3 题

纯数学 · 半角三角方程

STEP I 2003，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/03-s1-q3

关键思路

每一部分都是把 $\frac{1}{2}\theta$ 的某个函数与 θ 的同一个函数作比较，所以可利用倍角公式（double-angle formula），把所有式子都以半角（half-angle）表示。这样，每条方程都会变成只含单一个量（$\sin\frac{1}{2}\theta$, $\tan\frac{1}{2}\theta$ 或 $\cos\frac{1}{2}\theta$）的多项式方程（polynomial equation），可以用因式分解（factorise）或当作二次方程（quadratic）来解。唯一要小心的，是留意半角函数实际上可以取哪些值，以及各函数在哪里有定义。

解答

第 (i) 部分

利用 $\sin\theta = 2\sin\frac{1}{2}\theta\cos\frac{1}{2}\theta$，方程 $2\sin\frac{1}{2}\theta = \sin\theta$ 等价（equivalent）于

$$2\sin\tfrac{1}{2}\theta - 2\sin\tfrac{1}{2}\theta\cos\tfrac{1}{2}\theta = 0, \qquad \text{即} \qquad 2\sin\tfrac{1}{2}\theta\left(1-\cos\tfrac{1}{2}\theta\right) = 0.$$

所以方程成立，当且仅当（if and only if）$\sin\frac{1}{2}\theta = 0$ 或 $\cos\frac{1}{2}\theta = 1$。

但若 $\cos\frac{1}{2}\theta = 1$，则 $\sin^2\frac{1}{2}\theta = 1-\cos^2\frac{1}{2}\theta = 0$，所以无论如何都有 $\sin\frac{1}{2}\theta = 0$。因此第二种可能已包含在第一种之内，于是

$$2\sin\tfrac{1}{2}\theta = \sin\theta \iff \sin\tfrac{1}{2}\theta = 0,$$

正如所求。（换句话说，方程的解恰好是 $\theta = 2n\pi$，其中 n 为整数。）

第 (ii) 部分

设 $t = \tan\frac{1}{2}\theta$。要令方程有意义，$\tan\frac{1}{2}\theta$ 和 $\tan\theta$ 都必须有定义，所以 $\cos\frac{1}{2}\theta \neq 0$ 且 $\cos\theta \neq 0$。由于 $\cos\theta = \dfrac{1-t^2}{1+t^2}$，第二个条件即 $t^2 \neq 1$。在这些限制之下，$\tan\theta = \dfrac{2t}{1-t^2}$，方程变成

$$2t = \frac{2t}{1-t^2}.$$

两边乘以不等于零的 $1-t^2$，得 $2t - 2t^3 = 2t$，所以 $t^3 = 0$，即 $t = 0$。

> ? **想一想** 两边乘以 $1-t^2$ 是否稳妥？我们有没有因此失去了解，或多出了解？

乘以一个非零的量是可逆的步骤，所以在两个正切都有定义的集合上，这条关于 t 的方程与 $t = 0$ 完全等价。反过来，$t = 0$ 给出 $\tan\theta = 0$ 和 $\tan\frac{1}{2}\theta = 0$，两边都等于 0。

而 $\tan\frac{1}{2}\theta = 0$ 恰好在 $\frac{1}{2}\theta = n\pi$ 时成立。所以方程的解是

$$\theta = 2n\pi, \qquad n \in \mathbb{Z},$$

与第 (i) 部分的解集相同。

第 (iii) 部分

设 $c = \cos\frac{1}{2}\theta$。利用 $\cos\theta = 2\cos^2\frac{1}{2}\theta - 1 = 2c^2 - 1$，方程 $2\cos\frac{1}{2}\theta = \cos\theta$ 等价于

$$2c = 2c^2 - 1 \iff 2c^2 - 2c - 1 = 0 \iff c = \frac{2 \pm \sqrt{4+8}}{4} = \frac{1 \pm \sqrt{3}}{2}.$$

由于 c 是一个余弦值，所以 $-1 \leq c \leq 1$。根 $\frac{1}{2}(1+\sqrt{3}) \approx 1.37$ 太大，而 $\frac{1}{2}(1-\sqrt{3}) \approx -0.37$ 则在容许范围内。所以方程成立，当且仅当

$$\cos\tfrac{1}{2}\theta = \tfrac{1}{2}(1-\sqrt{3}) = -\cos\phi,$$

其中 ϕ 的定义与题目所给的相同。这个 ϕ 存在而且唯一，因为 $0 < \frac{1}{2}(\sqrt{3}-1) < 1$，而 cos 在 $[0, \frac{1}{2}\pi]$ 上由 1 递减至 0。

> ? **想一想** 怎样把 $\cos\frac{1}{2}\theta = -\cos\phi$ 化成题目所给的形式？

写成 $-\cos\phi = \cos(\pi - \phi)$。方程 $\cos x = \cos\alpha$ 的通解（general solution）是 $x = 2m\pi \pm \alpha$，所以

$$\tfrac{1}{2}\theta = 2m\pi \pm (\pi - \phi) \qquad \text{即} \qquad \theta = 4m\pi \pm (2\pi - 2\phi), \quad m \in \mathbb{Z}.$$

把两个正负号分开处理。

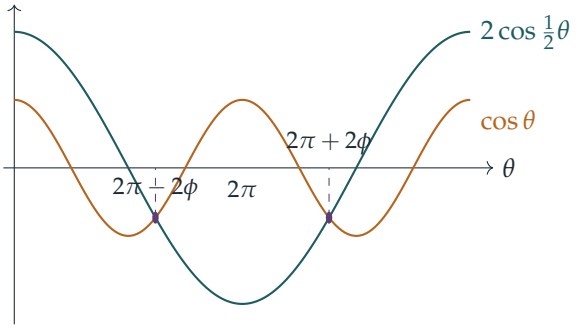
▶ 取 + 号：$\theta = 4m\pi + 2\pi - 2\phi = (4n+2)\pi - 2\phi$，其中 $n = m$。

- 取 $-$ 号：$\theta = 4m\pi - 2\pi + 2\phi = \big(4(m-1)+2\big)\pi + 2\phi = (4n+2)\pi + 2\phi$，其中 $n = m - 1$。

当 m 取遍所有整数时，两种情况下的 n 也取遍所有整数。因此

$$2\cos\tfrac{1}{2}\theta = \cos\theta \iff \theta = (4n+2)\pi \pm 2\phi，\text{对某个整数}n.$$

每一步都是等价关系，所以两个方向都已证明。我们也可以直接验证逆命题：若 $\theta = (4n+2)\pi \pm 2\phi$，则 $\cos\tfrac{1}{2}\theta = \cos\big((2n+1)\pi \pm \phi\big) = -\cos\phi = c$，而它满足 $2c = 2c^2 - 1$。



在 $2\cos\tfrac{1}{2}\theta$ 的一个周期（period）$0 \le \theta \le 4\pi$ 内，两条曲线相交两次，交点关于 $\theta = 2\pi$ 对称，而 $2\cos\tfrac{1}{2}\theta$ 正是在该处取得最小值 -2。

验算

- 以数值计算，$\cos\phi \approx 0.3660$ 给出 $\phi \approx 1.1961$，所以 (iii) 的第一个正数解是 $2\pi - 2\phi \approx 3.8911$。在该处 $2\cos\tfrac{1}{2}\theta \approx -0.7321$ 而 $\cos\theta \approx -0.7321$，符合要求。在 $[-6\pi, 6\pi]$ 上对 $2\cos\tfrac{1}{2}\theta - \cos\theta$ 进行求根搜索，找到的恰好是该范围内形如 $(4n+2)\pi \pm 2\phi$ 的六个值。
- 对 (ii) 作类似的搜索（避开极点（poles）），只在 2π 的倍数处找到根。
- 对 (i) 和 (ii) 的合理性检查：$\theta = \pi$ 不满足 (i)，因为 $2\sin\tfrac{1}{2}\pi = 2 \neq 0 = \sin\pi$；它亦不在 (ii) 的定义域（domain）之内，因为 $\tan\tfrac{1}{2}\pi$ 没有定义。

常见错误

- 在 (i) 中，约去 $\sin\tfrac{1}{2}\theta$ 而没有说明它可能等于零；或把 $\cos\tfrac{1}{2}\theta = 1$ 当作另一族独立的解，而没有留意到它已包含在前者之内。
- 在 (ii) 中，去分母之前忘记了 θ 必须避开 $\tan\theta$ 或 $\tan\tfrac{1}{2}\theta$ 没有定义的点。

- 在 (iii) 中,保留了大于 1 的根 $\frac{1}{2}(1+\sqrt{3})$;或在把 $4m\pi \pm (2\pi - 2\phi)$ 化成 $(4n+2)\pi \pm 2\phi$ 的形式时，错误处理 $\pm$ 号。

STEP 之外

每一部分都是以半角的单一个函数作代换（substitution）来解方程，从而把三角方程化为代数方程。(ii) 中的代换 $t = \tan\frac{1}{2}\theta$ 尤其强大：它把 $\sin\theta$, $\cos\theta$ 和 $\tan\theta$ 都表示成 t 的有理函数（rational function），这正是大学一年级微积分用它来求 $\sin\theta$ 和 $\cos\theta$ 的任何有理函数的积分的原因。从几何上看，它以有理点（rational points）把单位圆参数化（parametrise）。Siklos 的 *Advanced Problems in Mathematics* 一书中，有更多运用这些想法的 STEP 式练习。

详细解答

STEP I 2003，第 4 题

纯数学 · 三角不等式

STEP I 2003，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/03-s1-q4

关键思路

最容易想到的做法是把两边同乘 $\cos\theta$，但 $\cos\theta$ 在 $[0,2\pi)$ 上会变号，而乘以负数会令不等式（inequality）的方向逆转。因此我们分成 $\cos\theta>0$ 和 $\cos\theta<0$ 两种情况讨论。其中一种情况其实十分简单，因为 $1+\sin\theta$ 永不为负。另一种情况，只要把 $\sin\theta-\cos\theta$ 写成 $R\sin(\theta-\alpha)$ 的形式，便可化为单一个正弦不等式。

解答

设 $f(\theta)=\dfrac{1+\sin\theta}{\cos\theta}$。以下一律假设 $0\le\theta<2\pi$ 且 $\theta\neq\frac{\pi}{2},\frac{3\pi}{2}$。

情况 1：$\cos\theta<0$，即 $\frac{\pi}{2}<\theta<\frac{3\pi}{2}$

由于 $\sin\theta\ge-1$，分子 $1+\sin\theta$ 最少为 0，而分母为负数。所以 $f(\theta)\le0<1$。

$\left(\frac{\pi}{2},\frac{3\pi}{2}\right)$ 内的每一个 θ 都满足该不等式。

情况 2：$\cos\theta>0$，即 $0\le\theta<\frac{\pi}{2}$ 或 $\frac{3\pi}{2}<\theta<2\pi$

乘以正数 $\cos\theta$ 不会改变不等式的方向，所以

$$f(\theta)\le1\iff1+\sin\theta\le\cos\theta\iff\sin\theta-\cos\theta\le-1.$$

现在 $\sin\theta-\cos\theta=\sqrt{2}\left(\sin\theta\cos\frac{\pi}{4}-\cos\theta\sin\frac{\pi}{4}\right)=\sqrt{2}\sin\left(\theta-\frac{\pi}{4}\right)$，于是条件变成

$$\sin\left(\theta-\tfrac{\pi}{4}\right)\le-\tfrac{1}{\sqrt{2}}.$$

想一想 我们需要考虑 $\phi = \theta - \frac{\pi}{4}$ 的哪些值?

我们只需考虑这个情况所容许的两个范围:

- 若 $0 \le \theta < \frac{\pi}{2}$，则 $-\frac{\pi}{4} \le \phi < \frac{\pi}{4}$。在这个区间（interval）上，$\sin\phi$ 是递增的，且 $\sin\phi \ge \sin(-\frac{\pi}{4}) = -\frac{1}{\sqrt{2}}$，等号只在 $\phi = -\frac{\pi}{4}$ 时成立。所以这里唯一的解是 $\theta = 0$。（验算：$f(0) = 1/1 = 1$，等号成立。）
- 若 $\frac{3\pi}{2} < \theta < 2\pi$，则 $\frac{5\pi}{4} < \phi < \frac{7\pi}{4}$。在这个开区间（open interval）上 $\sin\phi < -\frac{1}{\sqrt{2}}$，因为 $\sin\phi$ 在两个端点都等于 $-\frac{1}{\sqrt{2}}$，而在两者之间的 $\phi = \frac{3\pi}{2}$ 处降至 -1。所以 $\left(\frac{3\pi}{2}, 2\pi\right)$ 内的每一个 θ 都是解。

综合结果

$\theta = \frac{3\pi}{2}$ 这个值要剔除，因为该处 $\cos\theta = 0$。综合各个情况，不等式成立，当且仅当

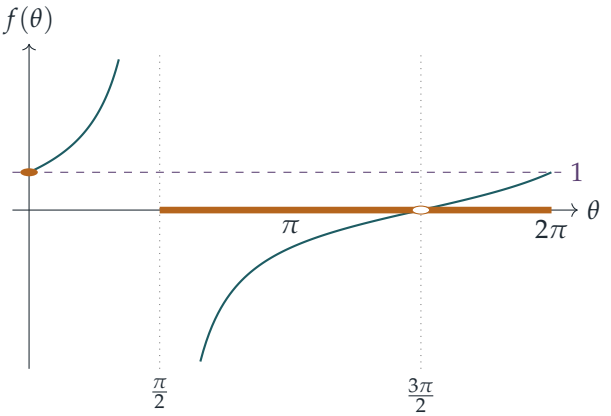
$$\theta = 0 \quad \text{或} \quad \frac{\pi}{2} < \theta < \frac{3\pi}{2} \quad \text{或} \quad \frac{3\pi}{2} < \theta < 2\pi.$$

更巧妙的看法

利用 $\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$、$\cos\theta = \cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2}$ 和 $1 = \cos^2\frac{\theta}{2} + \sin^2\frac{\theta}{2}$，

$$f(\theta) = \frac{\left(\cos\frac{\theta}{2} + \sin\frac{\theta}{2}\right)^2}{\left(\cos\frac{\theta}{2} + \sin\frac{\theta}{2}\right)\left(\cos\frac{\theta}{2} - \sin\frac{\theta}{2}\right)} = \frac{1 + \tan\frac{\theta}{2}}{1 - \tan\frac{\theta}{2}} = \tan\left(\frac{\theta}{2} + \frac{\pi}{4}\right),$$

这里约去了公因式，而由于 $\cos\theta \ne 0$，该公因式不等于零。（当 $\theta = \pi$ 时，中间的式子没有定义，但首尾两式在该处仍然相等。）当 θ 取遍 $[0, 2\pi)$ 时，$\psi = \frac{\theta}{2} + \frac{\pi}{4}$ 取遍 $[\frac{\pi}{4}, \frac{5\pi}{4})$。从 $\tan\psi$ 的图像一看便知，$\tan\psi \le 1$ 只在 $\psi = \frac{\pi}{4}$ 以及 $\frac{\pi}{2} < \psi < \frac{5\pi}{4}$ 时成立。只要剔除 $\psi = \pi$（即 $\theta = \frac{3\pi}{2}$），这就是同一个答案。



解集（琥珀色）就是曲线位于虚线 $f = 1$ 上或其下方的部分。留意 $\theta = \frac{3\pi}{2}$ 处的空心点：在该处 $f(\theta) \to 0$，但 f 没有定义。

验算

- $\theta = 0$：$f(0) = 1$，刚好在边界上。在略大于 0 之处，分子大于 1 而分母小于 1，所以 $f > 1$。这确认了 $\theta = 0$ 是一个孤立的解。
- $\theta = \frac{7\pi}{4}$：$f = \left(1 - \frac{1}{\sqrt{2}}\right) / \frac{1}{\sqrt{2}} = \sqrt{2} - 1 \approx 0.414 \leq 1$。正确。
- $\theta = \pi$：$f = 1/(-1) = -1 \leq 1$。正确。
- 我们在 $[0, 2\pi)$ 内随机抽取了 2×10^5 个 θ 作测试，结果与上述解集完全吻合。在 $\theta = \frac{3\pi}{2}$ 附近，我们得到 $f \approx 5 \times 10^{-5}$，与 $f \to 0$ 一致。

常见错误

- 没有分情况便把整条式乘以 $\cos\theta$。这样会在所有地方都得出 $1 + \sin\theta \leq \cos\theta$，错误地舍弃了整个区间 $\left(\frac{\pi}{2}, \frac{3\pi}{2}\right)$。
- 遗漏孤立的解 $\theta = 0$。题目容许 $\theta = 0$，而等号成立也算是解。
- 把 $\theta = \frac{3\pi}{2}$ 也包括在内。该处的式子是 $\frac{0}{0}$，所以没有定义，尽管它的极限（limit）是 0。

STEP 之外

这里的教训是：只有*递增*的运算才会保持不等式的方向，例如乘以正数、代入递增函数（increasing function）等等。时刻留意这一点，正是实分析（real analysis）的日常功夫。$\theta = \frac{3\pi}{2}$ 处的可去奇点（removable singularity），即 f 有极限但没有函数值的地方，是实分析的另一个主题：连续性（continuity）所讲的，就是极限与函数值是否一致。Abbott 的 *Understanding Analysis* 把这两个概念讲解得既严谨又易读。

详细解答

STEP I 2003，第 5 题

纯数学 · 二项式级数

STEP I 2003，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/03-s1-q5

关键思路

在任何二项展开式（binomial expansion）中，先写出通项（general term），再追踪它所带的 x 的幂次；常数项（constant term）就是该幂次为零的那一项。第 (ii) 部分的展开式并非有限项，所以要先抽出主导因子：当 $|x|$ 很大时，主导因子是 x^6，剩下一个以 $3/x$ 为变量的级数（series）；当 $|x|$ 很小时，主导因子是 $3x^5$，剩下一个以 $x/3$ 为变量的级数。之后只要看清楚 x 的哪些幂次可能出现，问题便迎刃而解。

解答

第 (i) 部分

$(2x + x^{-2})^6$ 的通项是

$$\binom{6}{r}(2x)^{6-r}\left(x^{-2}\right)^r = \binom{6}{r}2^{6-r}x^{6-3r}, \qquad r = 0, 1, \ldots, 6.$$

这一项与 x 无关，当且仅当 $6 - 3r = 0$，即 $r = 2$，此时该项为

$$\binom{6}{2}2^4 = 15 \times 16 = 240,$$

正如题目所求。

同样地，$(ax^3 + bx^{-2})^{5n}$ 的通项是

$$\binom{5n}{r}\left(ax^3\right)^{5n-r}\left(bx^{-2}\right)^r = \binom{5n}{r}a^{5n-r}b^r\,x^{15n-5r}.$$

当 $15n = 5r$ 时，x 的幂次为零，所以 $r = 3n$（它符合 $0 \le r \le 5n$）。因此与 x 无关的项是

$$\binom{5n}{3n}a^{2n}b^{3n} = \frac{(5n)!}{(2n)!\,(3n)!}\,a^{2n}b^{3n}.$$

第 (ii) 部分

$|x|$ **很大时**。抽出因子 x^6，得

$$f(x) = \left(x^6(1+3/x)\right)^{1/2} = x^3\left(1+\frac{3}{x}\right)^{1/2},$$

这里按题目的原意取 $(x^6)^{1/2} = x^3$（严格来说应是 $|x|^3$；见下文的说明）。当 $|x| > 3$ 时，$|3/x| < 1$，所以取 $u = 3/x$ 时，$(1+u)^{1/2}$ 的二项级数 (binomial series) 收敛 (converge)：

$$(1+u)^{1/2} = 1 + \tfrac{1}{2}u - \tfrac{1}{8}u^2 + \tfrac{1}{16}u^3 - \tfrac{5}{128}u^4 + \cdots,$$

其中 u^3 的系数是

$$\binom{1/2}{3} = \frac{\frac{1}{2}\cdot(-\frac{1}{2})\cdot(-\frac{3}{2})}{3!} = \frac{3/8}{6} = \frac{1}{16}.$$

把级数乘以 x^3，项 $\binom{1/2}{k}3^k x^{-k}$ 便变成 $\binom{1/2}{k}3^k x^{3-k}$，所以只有 $k = 3$ 会给出常数：

$$f(x) = x^3 + \tfrac{3}{2}x^2 - \tfrac{9}{8}x + \frac{1}{16}\cdot 27 - \frac{405}{128x} + \cdots,$$

而与 x 无关的项是 $\frac{27}{16}$，正如题目所求。

? **想一想** 当 $x < -3$ 时会怎样?

平方根指的是非负的那个根，而 $(x^6)^{1/2} = |x|^3$，当 $x < 0$ 时它等于 $-x^3$。所以当 $x < -3$ 时，展开式的每一项都会变号，常数项是 $-\frac{27}{16}$。题目所述的结果 $\frac{27}{16}$ 在 $x > 3$ 时完全正确；在 $x < -3$ 时，若把 $(x^6)^{1/2}$ 理解为 x^3，这个结果才成立，而这显然就是题目的原意。

$|x|$ **很小时**。这时括号内的主导项是 $3x^5$，所以我们改为抽出它：

$$x^6 + 3x^5 = 3x^5\left(1+\frac{x}{3}\right).$$

要 $f(x)$ 是实数，必须有 $x^5(x+3) \geq 0$；由于 $|x| < 3$，因子 $x+3$ 是正数，所以必须有 $x \geq 0$。（当 $-3 < x < 0$ 时，f 根本不能定义为实函数。）当 $0 < x < 3$ 时，

$$f(x) = \sqrt{3}\,x^{5/2}\left(1+\frac{x}{3}\right)^{1/2} = \sqrt{3}\,x^{5/2}\sum_{k\geq 0}\binom{1/2}{k}\frac{x^k}{3^k},$$

而由于 $|x/3| < 1$，这个级数收敛。每一项都是

$$x^{5/2+k}, \qquad k = 0, 1, 2, \ldots,$$

的常数倍，而 $\frac{5}{2}+k$ 永远不是整数，当然也就永远不是 0。因此，当 $|x| < 3$ 时，$f(x)$ 按 x 的幂展开的式子中没有与 x 无关的项。（在 $x = 0$ 这一点，$f(0) = 0$。）

验算

- 当 x 是很大的正数时，$f(x)-\left(x^3+\frac{3}{2}x^2-\frac{9}{8}x\right)$ 应趋近 $\frac{27}{16}=1.6875$。取 $x=1000$，计算机给出 1.68434，而级数的下一项 $-\frac{405}{128x}\approx -0.00316$ 正好解释了这个差距。
- 在 0 附近，$f(x)\approx\sqrt{3}\,x^{5/2}$：取 $x=0.01$，$f(0.01)=\sqrt{10^{-12}+3\times10^{-10}}\approx 1.734\times10^{-5}$，与 $\sqrt{3}\times10^{-5}$ 相当接近，而且肯定不是趋向一个非零常数。
- 对于一般公式，取 $n=1$、$a=b=1$，得 $\binom{5}{3}=10$，即 $(x^3+x^{-2})^5$ 的常数项；直接展开便可确认。

常见错误

- 写通项时把幂次的次序调转了，例如写成 $(2x)^r(x^{-2})^{6-r}$，然后又把 r 与二项式系数（binomial coefficient）配错。两种标记方式都可以，只要前后一致便行。
- 弄错 $\binom{1/2}{3}$ 的正负号：两个负因子 $(-\frac{1}{2})(-\frac{3}{2})$ 相乘，结果是正数。
- 在 x 很小时，以抽出 x^6 的方法展开 $(x^6+3x^5)^{1/2}$：这样会得到一个以 $3/x$ 为变量的级数，而它在 $|x|<3$ 时发散（diverge）。

STEP 之外

f 的两个展开式在不同的区域内成立，而抽出哪一个因子，就决定了你得到的是哪一个区域。这正是把函数围绕不同的点展开的想法，它通往复分析（complex analysis）中的洛朗级数（Laurent series）和皮瑟级数（Puiseux series，即按分数幂展开的级数，就像这里的 $x^{5/2+k}$）。在 $x<-3$ 时引起正负号问题的平方根分支（branch），也是复分析的一个核心课题。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）以图像方式解释了这两个概念。

详细解答

STEP I 2003，第 6 题

纯数学 · 以部分分式积分

> **STEP I 2003，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/03-s1-q6

关键思路

两个二次式都可以因式分解，所以最自然的工具就是部分分式（partial fractions）。所谓“不同情况”，出现在两个一次因式重合的那一刻：这时部分分式失效，我们改为对一个完全平方（perfect square）积分。动手做第 (ii) 部分之前，先仔细看清楚那两个二次式：只要把变量简单平移一下，其中一个就变成另一个。

解答

第 (i) 部分

由于 $2 \times a = 2a$ 及 $2 + a = a + 2$，这个二次式可以因式分解：

$$x^2 + (a+2)x + 2a = (x+2)(x+a).$$

因为 $a > 0$，当 $0 \le x \le 1$ 时两个因式都是正数，所以被积函数（integrand）在整个区间上连续（continuous），而对每一个 $a > 0$，积分都存在。两个因式相等，当且仅当 $a = 2$，这就是分界线。

情况 $a \neq 2$。两个因式相异，且

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

这可以通分来验证：分子变成 $(x+a) - (x+2) = a - 2$。由于 $x+2$ 和 $x+a$ 在 $[0,1]$ 上

都是正数，对数（logarithm）里不需要加绝对值符号：

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

想一想 当 $0<a<2$ 时，因子 $\frac{1}{a-2}$ 是负数。被积函数是正的，答案必须是正数；它仍然是正数吗？

是的。我们有 $\dfrac{3a}{2(a+1)}<1 \iff 3a<2a+2 \iff a<2$，所以当 $0<a<2$ 时，对数也是负数，两者相除便得正数。同理，当 $a>2$ 时两者皆为正数。如果你希望每一部分都是正数，可以把 $0<a<2$ 时的答案写成 $\dfrac{1}{2-a}\ln\left(\dfrac{2(a+1)}{3a}\right)$；这是同一个数。

情况 $a=2$。这时被积函数是 $(x+2)^{-2}$，而

$$\int_0^1 \frac{\mathrm{d}x}{(x+2)^2}=\left[-\frac{1}{x+2}\right]_0^1=-\frac{1}{3}+\frac{1}{2}=\frac{1}{6}.$$

第 (ii) 部分

这里的二次式是 $u^2+au+a-1=(u+1)(u+a-1)$，但在重复上面的工夫之前，先作代换（substitution）$u=x+1$：

$$(x+1)^2+a(x+1)+a-1=x^2+(a+2)x+2a.$$

这正正就是第 (i) 部分的二次式。此外 $\mathrm{d}u=\mathrm{d}x$，而 $u=1,2$ 对应 $x=0,1$。所以

$$\int_1^2 \frac{\mathrm{d}u}{u^2+au+a-1}=\int_0^1 \frac{\mathrm{d}x}{x^2+(a+2)x+2a},$$

答案亦与第 (i) 部分相同：

$$\int_1^2 \frac{\mathrm{d}u}{u^2+au+a-1}=\begin{cases}\dfrac{1}{a-2}\ln\left(\dfrac{3a}{2(a+1)}\right), & a\neq 2,\\[2ex] \dfrac{1}{6}, & a=2.\end{cases}$$

直接计算：在 $[1,2]$ 上有 $u+1>0$ 及 $u+a-1\geq a>0$，所以被积函数同样连续，而两个因式重合，当且仅当 $a-1=1$，即 $a=2$。部分分式给出 $\frac{1}{a-2}\big[\ln\frac{u+1}{u+a-1}\big]_1^2=\frac{1}{a-2}\big(\ln\frac{3}{a+1}-\ln\frac{2}{a}\big)$，这是同一个表达式。

验算

- **在 $a = 2$ 处的连续性**。当 $a \to 2$ 时，对数趋向 $\ln 1 = 0$，所以公式呈 $0/0$ 的形式。它的极限（limit）就是 $\ln 3a - \ln(2a + 2)$ 在 $a = 2$ 处的导数（derivative），即 $\frac{1}{2} - \frac{1}{3} = \frac{1}{6}$，与特殊情况的结果吻合。两种情况顺畅地衔接起来，本应如此。
- **一个特殊值**。取 $a = 1$，第 (i) 部分的被积函数是 $\frac{1}{(x+1)(x+2)}$，而公式给出 $-\ln\frac{3}{4} = \ln\frac{4}{3} \approx 0.2877$。
- **数值验证**。对两个积分使用辛普森法则（Simpson's rule），取 $a = 1$、$a = 2$、a 接近 0 和 2 的值，以及 200 个不超过 20 的随机 a 值，结果与公式相符，误差小于 10^{-10}。

常见错误

- 忘记了 $a = 2$ 的情况，这时部分分式的系数 $\frac{1}{a-2}$ 没有定义。
- 把部分分式的正负号弄反了。通分检查一下。
- 把 $0 < a < 2$ 和 $a > 2$ 当成真正不同的答案。只要留意正负号，一条公式已涵盖两者。

STEP 之外

在 $a = 2$ 处的表现，是可去奇点（removable singularity）的一个小例子：公式在某一点看似没有定义，但它在该点的极限恰好就是积分的值。在分析学（analysis）中，你会学到一个依赖参数的积分在甚么条件下对该参数连续（或可微），这就解释了为何两种情况必须吻合。Bartle 与 Sherbert 的 *Introduction to Real Analysis*（实分析入门）对极限、连续性和黎曼积分（Riemann integral）有严谨的论述。

详细解答

STEP I 2003，第 7 题

纯数学 · 数字与整除性

STEP I 2003，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/03-s1-q7

关键思路

设 $N = 100a + 10b + c$，并令 $S = N$，便得到一条关于三个数字的方程。把它整理成

$$99a + (10b - b^2) = c^3 - c,$$

题目第一段的两个事实就可以直接派上用场：含 b 的一项被困在 0 与 25 之间，而 $99a$ 和 $c^3 - c$ 都是 3 的倍数。第一个事实表示 $c^3 - c$ 必定刚好略高于 99 的某个倍数；第二个事实则减少了 b 的可能取值。之后只剩下寥寥几种情况，用手逐一检查即可。

解答

两个预备事实

配方（completing the square）得

$$10k - k^2 = 25 - (5 - k)^2.$$

由于 $(5-k)^2 \geq 0$，可得 $10k - k^2 \leq 25$。由于 $0 \leq k \leq 9$，我们有 $-4 \leq 5 - k \leq 5$，所以 $(5-k)^2 \leq 25$，因此 $10k - k^2 \geq 0$。（当 $k = 0$ 时取等号得 0，当 $k = 5$ 时取等号得 25。）

其次，$k-1$、k、$k+1$ 是三个连续整数（consecutive integers），所以当中恰有一个是 3 的倍数。因此它们的积 $k(k-1)(k+1)$ 可被 3 整除。（这对所有整数 k 都成立，不限于 $0 \leq k \leq 9$。）

找出这些数

设 $N = 100a + 10b + c$，其中各数字满足 $1 \le a \le 9$（因为 $N \ge 100$）及 $0 \le b, c \le 9$。则 $S = a + b^2 + c^3$，而

$$\begin{aligned} S = N &\iff a + b^2 + c^3 = 100a + 10b + c \\ &\iff 99a + (10b - b^2) = c^3 - c = (c-1)c(c+1). \qquad (*) \end{aligned}$$

想一想 为甚么要把 $10b - b^2$ 与 $99a$ 放在同一边?

因为这正是我们刚刚学会控制的组合。设 $t = 10b - b^2$，第一个事实指出 $0 \le t \le 25$，而 $99a$ 是 99 的倍数，远大于 25。所以 $(*)$ 表示 $c^3 - c$ 等于 99 的某个倍数加上一个细小的“余数” t，这样很快就能把情况锁定。

限制 b。由第二个事实，3 整除 $c^3 - c = (c-1)c(c+1)$，而 3 亦整除 $99a$。所以由 $(*)$，3 整除 $t = b(10 - b)$。由于 $10 - b \equiv 1 - b \pmod 3$，这需要 $b \equiv 0$ 或 $b \equiv 1 \pmod 3$。可能的情况如下：

b	0	1	3	4	6	7	9
$t = 10b - b^2$	0	9	21	24	24	21	9

所以 $t \in \{0, 9, 21, 24\}$。（$b = 2, 5, 8$ 时 $t = 16, 25, 16$，全都不能被 3 整除。）

限制 c。由于 $a \ge 1$ 及 $t \ge 0$，$(*)$ 给出 $c^3 - c \ge 99$。$c^3 - c$ 的值如下：

c	0	1	2	3	4	5	6	7	8	9
$c^3 - c$	0	0	6	24	60	120	210	336	504	720

所以 $c \ge 5$。

各种情况。对每个 $c \ge 5$，我们需要一个倍数 $99a$（其中 $1 \le a \le 9$）落在范围 $[c^3 - c - 25,\ c^3 - c]$ 之内，然后 $t = c^3 - c - 99a$ 必须是 $0, 9, 21, 24$ 其中之一。

- $c = 5$：$c^3 - c = 120$。在 $[95, 120]$ 内 99 的倍数只有 99，得 $a = 1$ 及 $t = 21$，所以 $b = 3$ 或 $b = 7$。由此得 $N = 135$ 和 $N = 175$。
- $c = 6$：$c^3 - c = 210$。在 $[185, 210]$ 内 99 的倍数只有 198，得 $a = 2$ 及 $t = 12$。但 12 并非 $0, 9, 21, 24$ 其中之一（事实上 $10b - b^2 = 12$ 没有整数根）。无解。
- $c = 7$：$c^3 - c = 336$。附近的倍数是 297 和 396，两者都不在 $[311, 336]$ 内。无解。
- $c = 8$：$c^3 - c = 504$。在 $[479, 504]$ 内 99 的倍数只有 495，得 $a = 5$ 及 $t = 9$，所以 $b = 1$ 或 $b = 9$。由此得 $N = 518$ 和 $N = 598$。
- $c = 9$：$c^3 - c = 720$。附近的倍数是 693 和 792，两者都不在 $[695, 720]$ 内。无解。

以上每一步都可以逆推，所以这四个数确实满足 $S = N$；直接验证：

$$135 : 1 + 3^2 + 5^3 = 1 + 9 + 125 = 135, \quad 175 : 1 + 7^2 + 5^3 = 1 + 49 + 125 = 175,$$
$$518 : 5 + 1^2 + 8^3 = 5 + 1 + 512 = 518, \quad 598 : 5 + 9^2 + 8^3 = 5 + 81 + 512 = 598.$$

所以满足 $S = N$ 的数恰好是

$$N = 135, \quad 175, \quad 518, \quad 598.$$

验算

对全部 900 个三位数作穷举搜索（brute-force search），逐一计算 $a + b^2 + c^3$，找到的恰好是 135、175、518 和 598，与分情况讨论的结果一致。还有一个巧妙的结构性检验：每一对解之中，b 的两个值相加都等于 10（$3 + 7$ 和 $1 + 9$）。这是必然的，因为把 b 换成 $10 - b$ 时，$t = 10b - b^2$ 保持不变。

常见错误

- 忘记 $a \geq 1$，因而接受了 000 或 001 这类并非三位数的“解”。
- 对给定的 a 和 c 找到一个 b 之后便停下来：由于 $t(b) = t(10 - b)$，解是成对出现的。
- 只用被 3 整除这一性质，而没有用上界限 $0 \leq t \leq 25$：正是这个界限令每个 c 最多只有一个可能的 a。

STEP 之外

等于自身各数字的幂之和的数，是趣味数学中的经典题材（其中最著名的一类是阿姆斯特朗数（Armstrong numbers），又称“自恋数”（narcissistic numbers），例如 $153 = 1^3 + 5^3 + 3^3$）。背后真正重要的想法，正是这里所用的方法：先估计两边的大小，使搜索范围变成有限，再利用同余（congruence）来删减。这种把大小估计与模算术（modular arithmetic）结合起来的手法，是解决许多丢番图方程（Diophantine equations）的关键。Davenport 的 *The Higher Arithmetic*（一本数论入门著作）对同余及其应用有极佳的介绍。

详细解答

STEP I 2003，第 8 题

纯数学 · 逻辑斯谛微分方程

STEP I 2003，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/03-s1-q8

关键思路

把“A 转化为 B 的速率”翻译成导数（derivative）：它就是 B 的体积增长的速率。由于 $x = 1 - y$，这样便得到一条只含 y 的可分离变量（separable variables）微分方程（differential equation），再用部分分式（partial fractions）完成积分。最后一部分问的是 $x = 1 - y$ 的公式会否到达零，而答案取决于一个事实：指数函数（exponential）的值永远不会是无限大。

解答

建立方程

B 的体积是 yV，而 V 是固定的。A 以速率 $kVxy$ 转化为 B，所以 B 的体积以该速率增加：

$$\frac{\mathrm{d}}{\mathrm{d}t}(yV) = kVxy \quad \Longrightarrow \quad \frac{\mathrm{d}y}{\mathrm{d}t} = kxy = ky(1-y),$$

这里除以 V 并用了 $x = 1 - y$。（等价地，A 的体积以这个速率减少：$V\,\mathrm{d}x/\mathrm{d}t = -kVxy$，由于 $\mathrm{d}x/\mathrm{d}t = -\mathrm{d}y/\mathrm{d}t$，这说的是同一件事。）

解方程

开始时 $x > 0$ 且 $y > 0$，所以 $0 < y < 1$。只要 $0 < y < 1$，我们便可以除以 $y(1-y)$，把变量分离：

$$\int \frac{\mathrm{d}y}{y(1-y)} = \int k\,\mathrm{d}t.$$

用部分分式，$\dfrac{1}{y(1-y)} = \dfrac{1}{y} + \dfrac{1}{1-y}$，所以

$$\ln y - \ln(1-y) = kt + c,$$
$$\ln \frac{y}{1-y} = kt + c,$$

这里用到 y 和 $1-y$ 都是正数，所以不需要加绝对值符号。两边取指数，并记 $D = \mathrm{e}^c$，

$$\frac{y}{1-y} = D\mathrm{e}^{kt}.$$

于是 $y = D\mathrm{e}^{kt} - yD\mathrm{e}^{kt}$，所以 $y\left(1 + D\mathrm{e}^{kt}\right) = D\mathrm{e}^{kt}$，而

$$y = \frac{D\mathrm{e}^{kt}}{1 + D\mathrm{e}^{kt}},$$

正是所求。代入 $t = 0$ 可得 $D = \dfrac{y_0}{1-y_0} = \dfrac{y_0}{x_0}$，其中 x_0 和 y_0 是初始值；特别地，$D > 0$。

想一想 我们在积分时假设了 $0 < y < 1$。怎样知道这个条件在任何时刻都一直成立?

假设不然，并设 T 为 y 首次到达 0 或 1 的时刻。在 $0 \le t < T$ 上，以上的推导是成立的，所以 y 在该处由公式给出。由于 y 是连续（continuous）的，$y(T)$ 就是公式在 $t \to T$ 时的极限（limit），即 $D\mathrm{e}^{kT}/(1 + D\mathrm{e}^{kT})$。因为 $D > 0$，这个值严格介乎 0 与 1 之间，于是出现矛盾。所以对所有 t 都有 $0 < y < 1$，而公式对所有 t 都成立。

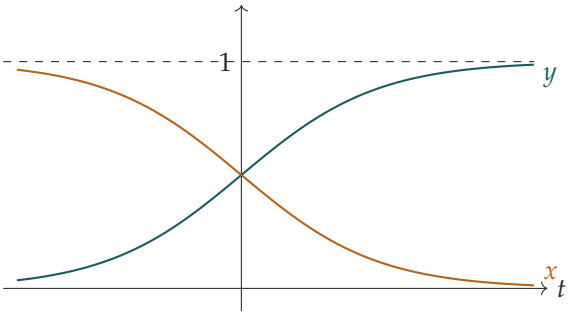
A 会否完全转化?

不会。由公式，

$$x = 1 - y = \frac{1}{1 + D\mathrm{e}^{kt}},$$

而对每一个 t，$D\mathrm{e}^{kt}$ 都是有限的，所以在任何时刻都有 $x > 0$：A 总会剩下一些。不过，当 $t \to \infty$ 时，$D\mathrm{e}^{kt} \to \infty$（因为 $k > 0$ 且 $D > 0$），所以 $x \to 0$ 而 $y \to 1$。转化过程愈来愈接近完成，但永远不会在有限时间内结束。

这在物理上也说得通：反应速率 $kVxy$ 与剩余的 A 的份量成正比，所以 A 快用完时，反应便会慢下来。



草图（取 $D = k = 1$）显示 y 的 S 形逻辑斯谛曲线（logistic curve）趋近 1，以及 x 的曲线趋近 0，两者都永远不会到达各自的极限。

验算

- 对答案求导：设 $u = De^{kt}$，则 $y = u/(1+u)$，而 $\mathrm{d}y/\mathrm{d}t = ku/(1+u)^2 = ky(1-y)$，与预期相符。
- 取 $D = k = 1$（即开始时两者体积相等），公式给出在 $t = \ln 3$ 时 $y = \frac{3}{4}$。我们亦随机选取了 200 组 k 和初始值，用数值方法把方程逐步向前推算，结果与公式相差不超过 10^{-12}，而 x 自始至终都保持正值。

常见错误

- 把“在极限中 $y = 1$”当成“在某个时刻 $y = 1$”。$t \to \infty$ 时的极限，在任何有限的时刻都不会达到。
- 忘记 $\ln(y/(1-y))$ 这一步需要 $0 < y < 1$，或者没有说明这个条件为何一直成立。

STEP 之外

方程 $\dot{y} = ky(1-y)$ 就是逻辑斯谛方程（logistic equation），是描述受资源限制的增长的标准模型。这里 $y = 0$ 和 $y = 1$ 是平衡解（equilibrium solutions），而微分方程的唯一性定理（uniqueness theorem）指出，其他解永远不能穿越或触及它们，这是解决最后一部分的一个更简洁的方法。Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科书）详细讨论了逻辑斯谛方程和这类稳定性分析（stability analysis）。

详细解答

STEP I 2003，第 9 题

力学 • 通过定点的抛体

STEP I 2003，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：

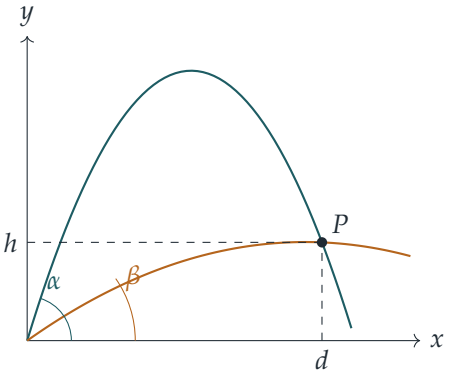
step.maths.org/questions/03-s1-q9

关键思路

写出水平位移和垂直位移，消去时间，再把 $\sec^2\theta$ 换成 $1+\tan^2\theta$：轨迹方程（trajectory equation）就变成 $T = \tan\theta$ 的二次方程。此后一切都围绕这个二次方程。有两个角度，即有两个相异实根，所以要看判别式（discriminant）。而 $\alpha+\beta$ 则通过 $\tan(\alpha+\beta)$ 的公式，由两根之和与两根之积决定。

解答

我们把质点（particle）视为只受重力作用而自由运动（不计空气阻力），以抛射点为原点，x 轴水平指向 P，y 轴铅垂向上。于是 P 是点 (d,h)，其中 $d>0$。我们取 $h>0$，即 P 在抛射点之上；$h\le 0$ 的情况留待最后再谈。



T 的二次方程

水平速度恒为 $V\cos\theta$，而垂直方向的运动有恒定加速度 $-g$，所以在时间 t

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

质点在时间 $t = d/(V\cos\theta)$ 到达 $x = d$（这里 $\cos\theta > 0$，因为质点朝 P 运动）。把它代入 y，并令 $y = h$，

$$h = d\tan\theta - \frac{gd^2}{2V^2\cos^2\theta} = dT - \frac{gd^2}{2V^2}\left(1 + T^2\right),$$

其中用了 $\sec^2\theta = 1 + \tan^2\theta$。现在全式乘以 $\dfrac{2V^2}{gd^2} = \dfrac{2k}{d}$：

$$\frac{2kh}{d} = 2kT - \left(1 + T^2\right), \qquad \text{即} \qquad T^2 - 2kT + \frac{2kh}{d} + 1 = 0,$$

这正是所求。

两个相异的角度

对于给定的 V（因而给定的 k），可行的抛射角（angle of projection）就是那些 $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$，其正切是这个二次方程的根。由于 tan 在该区间上严格递增（strictly increasing），相异的实根 T 对应相异的角度。所以我们需要这个二次方程有两个相异实根，而这恰好在判别式为正时成立：

$$4k^2 - 4\left(\frac{2kh}{d} + 1\right) > 0 \iff k^2d^2 - 2h(kd) - d^2 > 0 \iff (kd - h)^2 > h^2 + d^2,$$

这里先乘以 $d^2 > 0$，再对 kd 配方（completing the square）。

> **想一想** 把不等式两边平方有时会出错，那么为甚么由题目给出的条件可以推出 $(kd - h)^2 > h^2 + d^2$？

若 $kd > h + \sqrt{h^2 + d^2}$，则 $kd - h > \sqrt{h^2 + d^2} > 0$。两边皆为正数，所以平方后不等式仍然成立：$(kd - h)^2 > h^2 + d^2$。因此判别式为正，$T$ 有两个相异的根，从而有两个相异的抛射角。

（逆命题（converse）同样成立。要令 $(kd - h)^2 > h^2 + d^2$，另一个可能是 $kd < h - \sqrt{h^2 + d^2}$，但由于 $\sqrt{h^2 + d^2} > |h|$，右边是负数，而 $kd = V^2/g > 0$。所以当且仅当 $kd > h + \sqrt{h^2 + d^2}$ 时，才可以有两个角度。）

两个角度都确实在水平线之上。两根之和为 $2k > 0$，两根之积为 $1 + 2kh/d > 0$（因为 $h > 0$），所以两根皆为正数。

两角之和

设 $\tan\alpha$ 和 $\tan\beta$ 为两根。由二次方程可得

$$\tan\alpha+\tan\beta=2k,\qquad \tan\alpha\tan\beta=1+\frac{2kh}{d}.$$

两根之积不等于 1（因为 $kh/d\neq 0$），所以 $\alpha+\beta\neq\frac{\pi}{2}$，复角公式（compound-angle formula）适用：

$$\tan(\alpha+\beta)=\frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta}=\frac{2k}{-2kh/d}=-\frac{d}{h}.$$

留意 k 已经约去：两角之和与速率无关。

> ? **想一想** $\tan(\alpha+\beta)=-d/h$ 有无限多个解。哪一个才是 $\alpha+\beta$?

我们已证明 $\alpha,\beta\in(0,\frac{\pi}{2})$，所以 $0<\alpha+\beta<\pi$。此外 $\tan(\alpha+\beta)<0$，所以 $\alpha+\beta$ 位于 $(\frac{\pi}{2},\pi)$ 之内。在 $(\frac{\pi}{2},\pi)$ 内正切为 $-d/h$ 的角只有一个，就是 $\pi-\arctan(d/h)$，因为 $\arctan(d/h)\in(0,\frac{\pi}{2})$，而且 $\tan(\pi-u)=-\tan u$。因此

$$\alpha+\beta=\pi-\arctan\left(\frac{d}{h}\right),$$

这正是所求。

这个结果有一个漂亮的看法。$\arctan(d/h)$ 是 OP 与向上铅垂线之间的夹角，所以 $\pi-\arctan(d/h)=\frac{\pi}{2}+\gamma$，其中 γ 是 P 的仰角（angle of elevation）。于是 $\frac{1}{2}(\alpha+\beta)=\frac{\pi}{4}+\frac{\gamma}{2}$：两个抛射方向关于 OP 与铅垂线夹角的角平分线对称。

验算

- 若 $h=0$（目标在同一水平），公式给出 $\alpha+\beta=\pi-\frac{\pi}{2}=\frac{\pi}{2}$：这就是大家熟悉的、射程（range）相同的一对余角。
- 取 $d=3$、$h=4$、$k=4$。则 $kd=12>4+5$，两根为 $T=4\pm\sqrt{13/3}$，得出 $\alpha\approx 1.4078$ 和 $\beta\approx 1.0903$。所以 $\alpha+\beta\approx 2.4981$，等于 $\pi-\arctan(0.75)\approx 2.4981$。
- 对数千组随机的 d、h 和 k，分别以两个角度对飞行过程作数值模拟，轨迹都通过 (d,h)，并印证了该条件和两角之和。

常见错误

- 忘记 $\sec^2\theta=1+\tan^2\theta$，以致始终得不到 T 的二次方程。
- 写成 $\alpha+\beta=\arctan(-d/h)=-\arctan(d/h)$，这是负数。反正切（inverse tangent）只会给出 $(-\frac{\pi}{2},\frac{\pi}{2})$ 内的角，所以必须利用 $\alpha+\beta$ 的范围来选取正确的分支（branch）。

- 不加小心地取 $h < 0$（目标在抛射点之下）。这时其中一根可能是负数，而正确的一般结论是 $\alpha + \beta = \frac{\pi}{2} + \arctan(h/d)$。只有当 $h > 0$ 时，它才与题目中的形式一致。

STEP 之外

在临界情况 $kd = h + \sqrt{h^2 + d^2}$，两个角度合而为一，这个情况描绘出以速率 V 抛射的所有轨迹的包络线（envelope）：“安全抛物线”（parabola of safety）$y = \frac{V^2}{2g} - \frac{gx^2}{2V^2}$。在它之内的点有两种方法可以击中，在它之上的点只有一种方法，在它之外的点则完全无法击中。曲线族的包络线，可以把一条方程与它对参数的导数联立求解而得出，在微分方程（differential equations）和光学中还会再次出现。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）正是以这种严谨的态度讲解抛体运动和约束运动。

详细解答

STEP I 2003，第 10 题

力学 • 薄片的质心

> **STEP I 2003，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/03-s1-q10

关键思路

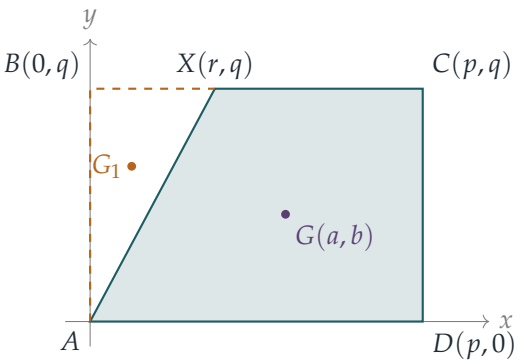
处理切去了一块的图形，最好的方法是相减：长方形对每条轴的矩（moment）等于梯形（trapezium）的矩加上三角形的矩，而各部分的质量与面积成正比。至于悬挂的部分，一块薄片（lamina）从某一点自由悬挂，静止时其重心（centre of gravity）必在该点的正下方，所以直线 AG 是铅垂的。这样，45° 这个条件就化为 a 与 b 之间一个非常简单的关系。

解答

我们把薄片当作均匀而且很薄，因此任何一部分的质量都与其面积成正比，而三角形的重心位于其形心（centroid），即三个顶点的平均位置。

重心

按题目所给的坐标轴，$A=(0,0)$，$D=(p,0)$，$C=(p,q)$，因此 $B=(0,q)$。由于 X 在 BC 上，且 $BX=r$，所以 $X=(r,q)$，其中 $0\le r\le p$。



三个部分如下。

- 长方形 $ABCD$：面积为 pq，重心为 $(\frac{p}{2}, \frac{q}{2})$。
- 三角形 ABX（被切去的部分）：面积为 $\frac{1}{2}qr$，重心 G_1 是 $(0,0)$、$(0,q)$ 和 (r,q) 的平均位置，即 $(\frac{r}{3}, \frac{2q}{3})$。
- 余下的梯形 $AXCD$：面积为 $pq - \frac{1}{2}qr = \frac{1}{2}q(2p-r)$，重心为 $G = (a,b)$。

长方形对每条轴的矩，等于它两个部分的矩之和。对 y-轴取矩（即比较 x-坐标），

$$pq \cdot \frac{p}{2} = \frac{qr}{2} \cdot \frac{r}{3} + \frac{q(2p-r)}{2}\, a.$$

两边除以 $\frac{q}{2}$，得 $p^2 = \frac{r^2}{3} + (2p-r)a$；由于 $2p - r \geq p > 0$，我们可以除以它：

$$a = \frac{3p^2 - r^2}{3(2p-r)}.$$

同样地，对 x-轴取矩，

$$pq \cdot \frac{q}{2} = \frac{qr}{2} \cdot \frac{2q}{3} + \frac{q(2p-r)}{2}\, b,$$

所以 $pq = \frac{2qr}{3} + (2p-r)b$，而

$$b = \frac{q(3p-2r)}{3(2p-r)}.$$

快速验算一下：$r = 0$（即没有切去任何部分）时，$a = \frac{p}{2}$，$b = \frac{q}{2}$，正是长方形的中心。

从 A 悬挂

当薄片从 A 自由悬挂并处于平衡（equilibrium）时，只有两个力作用在它上面：通过 G 的重量，以及在 A 的反作用力。两个力若互相平衡，必定沿同一直线作用，所以 G 在 A 的正下方。

想一想 “AD 在水平线以下 45°”怎样变成一条关于 a 和 b 的方程?

由于 AG 是铅垂的，而 AD 与水平线成 45°，所以 AD 与 AG 亦成 45°；即是说，角 GAD 是 45°。在我们的坐标轴中，AD 是正 x-轴，而 $G=(a,b)$，其中 $a,b>0$（它位于梯形之内），所以 $\tan\angle GAD = b/a$。因此 $b/a=1$，即

$$a=b.$$

代入上面的表达式，并约去共同的正分母 $3(2p-r)$，

$$\begin{aligned} & 3p^2-r^2 = q(3p-2r) \\ \iff\quad & r^2-2qr+3pq-3p^2=0. \end{aligned}$$

由二次公式（quadratic formula），

$$r = q\pm\sqrt{q^2-(3pq-3p^2)} = q\pm\sqrt{q^2-3pq+3p^2}.$$

根号内的量是 $(q-\frac{3}{2}p)^2+\frac{3}{4}p^2>0$，所以两个根都是实数，而且互不相同。记 $s=\sqrt{q^2-3pq+3p^2}>0$。

选择正负号。由于 X 在边 BC 上，而这条边的长度是 p，我们需要 $0\le r\le p$。现在证明根 $q+s$ 总是大于 p。

- 若 $q\ge p$，则 $q+s>q\ge p$，因为 $s>0$。
- 若 $q<p$，则 $p-q>0$，而且

$$s^2-(p-q)^2 = (q^2-3pq+3p^2)-(p^2-2pq+q^2) = 2p^2-pq = p(2p-q)>0,$$

因为 $q<p<2p$。所以 $s>p-q$，从而 $q+s>p$。

无论哪种情况，取正号都会令 X 越过 C，落在薄片之外，所以不可能。因此

$$r = q-\sqrt{q^2-3pq+3p^2},$$

这正是所要证明的结果。

为完整起见补充一点：这个根确实位于 $[0,p]$ 之内，当且仅当 $p\le q\le 2p$。事实上，$q-s\ge 0 \iff q^2\ge s^2 \iff 3p(q-p)\ge 0 \iff q\ge p$，而且（当 $q>p$ 时）$q-s\le p \iff (q-p)^2\le s^2 \iff p(2p-q)\ge 0 \iff q\le 2p$。所以题目描述的情况只会在这类形状中出现，这与直觉吻合：未经切割的正方形（$q=p$，$r=0$）从一角悬挂时，对角线是铅垂的；而在另一个极端（$q=2p$，$r=p$），图形就是直角三角形 ACD，其形心 $(\frac{2p}{3},\frac{q}{3})$ 同样满足 $a=b$。

验算

- 取 $r = 0$ 便得回长方形的中心 $(\frac{p}{2}, \frac{q}{2})$，而 $r = p$ 则给出 $(\frac{2p}{3}, \frac{q}{3})$，即三角形 ACD 的形心。
- 取 $p = 1$，$q = 1.5$，公式给出 $r = 1.5 - \sqrt{0.75} \approx 0.634$，于是 $a = b \approx 0.634$，所以 AG 确实与 AD 成 45°。另一个根 ≈ 2.37，超过了 $p = 1$。
- 我们亦以一般的多边形形心公式，用数千组随机的 p, q, r 核对过重心公式。

常见错误

- 把三角形的形心写成 $(\frac{r}{3}, \frac{q}{3})$：它有两个顶点位于高度 q，所以形心的高度是 $\frac{2q}{3}$。
- 把三角形的矩加上去而不是减去：梯形是长方形*减去*三角形。
- 以一句含糊的“r 一定很小”来舍去正号。真正的理由是 $r \le p$，而这需要一段简短的论证，按 $q \ge p$ 与否分开两种情况处理。

STEP 之外

用相减法求质心（centre of mass），其实是计算 $\iint x \, \mathrm{d}A \big/ \iint \mathrm{d}A$ 的离散版本；到了大学，你会在二重积分（double integral）中再遇上同一个想法，用来处理密度不均匀的薄片以及立体。悬挂条件则是一个普遍原理的最简单情况：一个以某一点为支点的刚体（rigid body），当其质心位于支点正下方时处于稳定平衡，此时它的势能（potential energy）最小。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教科书）对质心和刚体平衡有仔细的论述。

详细解答

STEP I 2003，第 11 题

力学 · 斜面上的碰撞

STEP I 2003，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/03-s1-q11

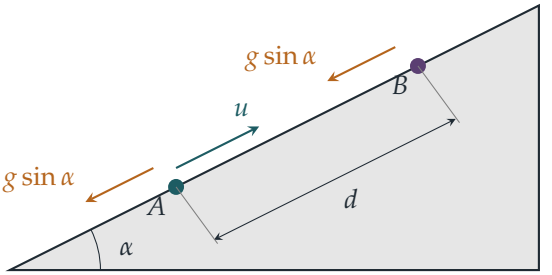
关键思路

在光滑（smooth）斜面上，无论速度如何，每个质点的加速度都相同，都是沿最大斜率线（line of greatest slope）向下的 $g\sin\alpha$。因此重力无法改变两个质点之间的相对运动：从其中一个质点看，另一个质点以恒定速度运动。这样一来，碰撞时间便很容易求出，而且不必经过繁琐的运动学计算，就能知道碰撞一刻的相对速度（relative velocity）。碰撞中的能量损失只取决于这个相对速度，这正是 α、g 和 d 全都从答案中消失的原因。

解答

模型

我们把两个质点都视为点质量，沿着通过 A 和 B 的最大斜率线运动。斜面是光滑的，所以每个质点所受的力只有重量和法向反作用力（normal reaction）。沿垂直于斜面的方向分解，反作用力与 $mg\cos\alpha$ 平衡。沿斜面方向分解，每个质点的加速度都是沿斜面向下的 $g\sin\alpha$。取 x 为由 A 起沿斜面向上的距离。



碰撞时间

设 $a = g\sin\alpha$。以沿斜面向上为正方向，从 A 出发的质点（称为 P）和从 B 出发的质点（称为 Q）的位置分别为

$$x_P = ut - \tfrac{1}{2}at^2,$$
$$x_Q = d - \tfrac{1}{2}at^2.$$

两者之间的距离为

$$x_Q - x_P = d - ut,$$

因为 $\frac{1}{2}at^2$ 项互相抵消。这段距离由 d 开始均匀地缩短，并且恰好在

$$t = \frac{d}{u},$$

时变为零，这正是我们要证明的结果。这次抵消正是关键思路的体现：两个质点以同样的方式沿斜面下落，所以只有初始相对速率 u 才有影响。

动能损失

在 $t = d/u$ 时，两个质点沿斜面向上的速度为

$$v_P = u - \frac{ad}{u}, \qquad v_Q = -\frac{ad}{u}.$$

为简便起见，记 $w = ad/u$。于是 $v_P = u - w$ 及 $v_Q = -w$，而 P 以相对速率 $v_P - v_Q = u$ 接近 Q。

> **想一想** 重力正在作用，为甚么我们仍可使用动量守恒（conservation of momentum）？

碰撞是瞬时发生的。在极短的碰撞时间内，重力和法向反作用力这些有限的力所产生的冲量（impulse），与质点之间的巨大冲量相比可以忽略不计。因此在碰撞中，沿斜面方向的动量守恒。

设碰撞后沿斜面向上的速度为 v'_P 和 v'_Q。由于两个质点的质量相等，动量守恒给出

$$v'_P + v'_Q = v_P + v_Q = u - 2w. \tag{1}$$

牛顿实验定律（Newton's experimental law）指出，分离速率等于接近速率的 e 倍：

$$v'_Q - v'_P = e(v_P - v_Q) = eu. \tag{2}$$

我们不分别解出 v'_P 和 v'_Q，而是利用恒等式

$$p^2 + q^2 = \tfrac{1}{2}\left[(p+q)^2 + (p-q)^2\right].$$

把它分别应用于碰撞前和碰撞后，总动能（kinetic energy，简称 KE）为

$$\text{碰撞前动能} = \tfrac{1}{2}m\left(v_P^2 + v_Q^2\right) = \tfrac{1}{4}m\left[(u-2w)^2 + u^2\right],$$
$$\text{碰撞后动能} = \tfrac{1}{2}m\left(v_P'^2 + v_Q'^2\right) = \tfrac{1}{4}m\left[(u-2w)^2 + e^2u^2\right],$$

其中碰撞后的和与差分别用了 (1) 和 (2)。两式中“和”的一项相同，所以

$$\text{动能损失} = \tfrac{1}{4}mu^2 - \tfrac{1}{4}me^2u^2 = \tfrac{1}{4}mu^2\left(1-e^2\right),$$

正如所求。

想一想 $(u-2w)^2$ 这一项在物理上代表甚么？为甚么它不会改变？

$v_P + v_Q$ 的一半就是质心（centre of mass）的速度。$\frac{1}{4}m(u-2w)^2$ 这一项是总质量 $2m$ 以该速度运动时的动能，而动量守恒令它保持不变。只有相对于质心的运动的动能才可能损失，而这部分动能只取决于相对速率：碰撞前是 u，碰撞后是 eu。

验算

- 若 $e=1$（完全弹性，perfectly elastic），损失为零，理应如此。若 $e=0$，两个质点在碰撞后一起运动，损失为 $\frac{1}{4}mu^2$，即相对运动的全部动能。
- 若 $\alpha=0$，斜面变成水平面，Q 保持静止，而 P 经过时间 d/u 后以速率 u 到达它那里，与结果吻合。
- 数值上，取 $\alpha=30^\circ$、$g=9.8$、$d=2$、$u=4$、$e=\frac{1}{2}$ 及 $m=1$：在 $t=0.5$ 时，两个速度为 1.55 和 -2.45。碰撞后它们变为 -1.45 和 0.55，动能由 4.2025 降至 1.2025，损失为 3，等于 $\frac{1}{4}\cdot 16\cdot\frac{3}{4}$。另外以数值方法逐步模拟运动，随机测试了 1000 个情况，结果同样吻合。

常见错误

- 把 Q 在碰撞一刻的速度当作零。它已经沿斜面下滑了 d/u 这么长的时间，正如 P 在这段时间内一直在减速。
- 在牛顿定律中弄错正负号。应把两个速度都按同一个正方向写出，然后用分离速率 $= e\times$ 接近速率。
- 完整解出 v_P' 和 v_Q'，然后展开四个平方。这样做是可行的，但利用和与差的恒等式可以省去几乎所有代数运算。

STEP 之外

两个部分都是转换到质心参考系（centre-of-mass frame）的例子。均匀重力场令每个质点都有相同的加速度，所以它在相对运动中完全消失。此外，任何系统的动能都可以分拆为质心的动能，加上相对于质心的动能（柯尼希定理，König's theorem）。在二体问题中，第二部分是 $\frac{1}{2}\mu v_{\text{rel}}^2$，其中折合质量（reduced mass）为 $\mu = m_1 m_2/(m_1 + m_2)$，在这里是 $\frac{1}{2}m$。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教科书）对这两个想法都有仔细的阐述。

详细解答

STEP I 2003，第 12 题

概率与统计 • 乘积的期望值

STEP I 2003，第 12 题

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

step.maths.org/questions/03-s1-q12

关键思路

期望值（expected value）就是对所有等可能的结果，把（数值 × 概率）加起来，所以两部分都归结为列出所有有序数对（ordered pairs），再把它们的乘积相加。二重和 $\sum_r \sum_s rs$ 可以分解为 $\left(\sum r\right)^2$，第 (i) 部分要做的就只有这一步。第 (ii) 部分唯一的分别，是 $r = s$ 的数对不再可能出现，所以我们取同一个二重和，再除去它的对角线项。

解答

设 X 为第一个球上的数字，Y 为第二个球上的数字。

第 (i) 部分

由于第一个球会放回袋中，第二次抽取时袋中仍是全部的球，结果与第一次无关。所以对任意 $r, s \in \{1, \ldots, n\}$，

$$\mathrm{P}(X = r,\ Y = s) = \mathrm{P}(X = r)\,\mathrm{P}(Y = s) = \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2}.$$

这样的有序数对 (r, s) 共有 n^2 个，每个给出乘积 rs，所以按期望值的定义，

$$\mathrm{E}(XY) = \sum_{r=1}^{n} \sum_{s=1}^{n} rs\,\mathrm{P}(X = r,\ Y = s) = \frac{1}{n^2} \sum_{r=1}^{n} \sum_{s=1}^{n} rs,$$

这正是题目要求我们解释的式子。

要化简这个式子，留意 r 与 s 无关，所以可以把它提到内层求和之外：

$$\sum_{r=1}^{n} \sum_{s=1}^{n} rs = \sum_{r=1}^{n} r\left(\sum_{s=1}^{n} s\right) = \left(\sum_{r=1}^{n} r\right)^2 = \tfrac{1}{4} n^2 (n+1)^2.$$

因此

$$\mathrm{E}(XY) = \frac{(n+1)^2}{4}.$$

（等价地说，X 与 Y 独立（independent），所以 $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y) = \left(\frac{n+1}{2}\right)^2$。）

第 (ii) 部分

要抽出两个球，须有 $n \geq 2$。现在第一个球不放回，所以可能的结果是满足 $r \neq s$ 的有序数对 (r,s)。对这样的一个数对，

$$\mathrm{P}(X = r,\ Y = s) = \mathrm{P}(X = r)\,\mathrm{P}(Y = s \mid X = r) = \frac{1}{n} \cdot \frac{1}{n-1},$$

因为取走 r 号球之后，第二个球是余下 $n-1$ 个球中任何一个的可能性均等。这样的数对共有 $n(n-1)$ 个，全部等可能，而 $r = s$ 的数对概率为 0。因此

$$\mathrm{E}(XY) = \frac{1}{n(n-1)} \sum_{r \neq s} rs.$$

? **想一想** 怎样只对 $r \neq s$ 的数对把 rs 加起来?

先取对所有数对的和（已在第 (i) 部分求出），再减去 $r = s$ 的对角线项，这些项各自贡献 $r \cdot r = r^2$：

$$\sum_{r \neq s} rs = \sum_{r=1}^{n} \sum_{s=1}^{n} rs - \sum_{r=1}^{n} r^2 = \tfrac{1}{4}n^2(n+1)^2 - \tfrac{1}{6}n(n+1)(2n+1).$$

提出公因式 $\frac{1}{12}n(n+1)$，

$$\begin{aligned}\sum_{r \neq s} rs &= \tfrac{1}{12}n(n+1)\big[3n(n+1) - 2(2n+1)\big] \\ &= \tfrac{1}{12}n(n+1)\big(3n^2 - n - 2\big) \\ &= \tfrac{1}{12}n(n+1)(n-1)(3n+2).\end{aligned}$$

除以 $n(n-1)$（因为 $n \geq 2$，它不等于零），

$$\mathrm{E}(XY) = \frac{(n+1)(3n+2)}{12},$$

这就是所求的结果。

验算

取 $n = 3$。有放回时，九个乘积的平均值是 $\frac{1}{9}(1+2+3)^2 = 4$，而 $\frac{(3+1)^2}{4} = 4$。无放回时，无序数对 $\{1,2\},\{1,3\},\{2,3\}$ 给出乘积 $2,3,6$，平均值为 $\frac{11}{3}$，而 $\frac{4 \cdot 11}{12} = \frac{11}{3}$。当 $n = 2$ 时，

无放回的乘积只有 2 一个，而公式给出 $\frac{3\cdot 8}{12} = 2$。我们亦以逐一列举的方式，对 12 或以下的每个 n 准确验证了两条公式。

比较两个答案也是一个很好的合理性检查：

$$\frac{(n+1)(3n+2)}{12} - \frac{(n+1)^2}{4} = -\frac{n+1}{12} < 0.$$

无放回时答案较小，这是合理的：我们失去了 $n \cdot n$ 这类数值大的对角线乘积。

常见错误

- 在第 (ii) 部分除以 n^2。可能的有序数对只有 $n(n-1)$ 个，每个的概率是 $\frac{1}{n(n-1)}$。
- 求和时用无序数对，计数时却用有序数对（或者反过来）。对无序数对求和，会令总和与个数同时减半，所以比值不变；但把两者混用，结果就会相差一个因子 2。
- 在第 (ii) 部分写 $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$。这需要独立性，而球不放回时 X 与 Y 并不独立。

STEP 之外

我们求得的差正好就是协方差 (covariance)：$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) = -\frac{n+1}{12}$，它是负数，因为第一个数字大，第二个数字也大的可能性就会稍低。在无放回抽样（sampling without replacement）中求和的方差（variance）时，需要的正是这类协方差，由此可以得出超几何分布（hypergeometric distribution）的方差，以及抽样调查中使用的“有限总体校正”（finite population correction）。Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门教材）正是以这种方式发展协方差与指示变量（indicator）方法。

详细解答

STEP I 2003，第 13 题

概率与统计 · 几何分布的期望值

STEP I 2003，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/03-s1-q13

关键思路

第一部分的做法，是数每队射失的次数，而不是入球数：两队各射五球之后，比数相同的充要条件是两队射失的次数相同。至于期望值（expectation），留意突然死亡（sudden death）阶段是以每回合两球的方式进行的，而每一回合都是一次独立试验，以相同的概率结束这场互射十二码。因此回合数服从几何分布（geometric distribution），而射球次数的期望值可以拆成“必定要射的十球”加上“额外的射球，并按需要加射的机会加权”。

解答

我们按照模型的原意，假设每一次射球都与其他各次射球独立（independent），并且 A 队每次射入的概率恒为 p_A，B 队则恒为 p_B。记 $q_A = 1 - p_A$ 及 $q_B = 1 - p_B$。

射完十球后打和的概率 α

设 X 为 A 队在首五球中射失的球数，Y 为 B 队射失的球数。每队的五次射球都是射失概率相同的独立试验，所以

$$X \sim \mathrm{B}(5, q_A), \qquad Y \sim \mathrm{B}(5, q_B),$$

而且由于两队的射球互相独立，X 与 Y 亦互相独立。

A 队射入 $5 - X$ 球,B 队射入 $5 - Y$ 球,所以比数相同的充要条件是 $X = Y$。对 $i = 0, 1, \ldots, 5$,

事件 $\{X = Y = i\}$ 互不相交（disjoint），所以

$$\begin{aligned}\alpha = \mathrm{P}(X = Y) &= \sum_{i=0}^{5} \mathrm{P}(X = i)\,\mathrm{P}(Y = i) \\ &= \sum_{i=0}^{5} \binom{5}{i} q_A^i p_A^{5-i} \binom{5}{i} q_B^i p_B^{5-i} \\ &= \sum_{i=0}^{5} \binom{5}{i}^2 (1-p_A)^i (1-p_B)^i\, p_A^{5-i} p_B^{5-i},\end{aligned}$$

这就是所求的表达式。（若改为数入球数，即两队各入 $5-i$ 球，会得到相同的和，因为 $\binom{5}{i} = \binom{5}{5-i}$。）

射球次数的期望值

如果射完十球后比数仍然相同，突然死亡阶段便以回合进行：每一回合中 A 队射一球，B 队射一球。若两球之中恰好只有一球射入，该回合便分出胜负，而这个情况出现的概率是

$$p_A q_B + q_A p_B = p_A(1-p_B) + p_B(1-p_A) = p_A + p_B - 2p_A p_B = \beta.$$

否则（两队都射入或都射失）便要再进行一个回合。我们需要 $\beta > 0$，也就是说，我们排除 $p_A = p_B = 0$ 和 $p_A = p_B = 1$ 这两个平凡情况，因为在这两个情况下，任何回合都永远无法分出胜负。

> **想一想** 为甚么互射十二码不可以在一个回合的中途结束?

规则订明，当一队射入而另一队未能射入时，两队的射球次数必须相同。即使 A 队射失，B 队仍然要射自己的一球（B 队也可能射失）。所以突然死亡阶段的射球总是成对出现，射球总数是 $10 + 2R$，其中 R 是所进行的回合数（若比赛在首十球内已分出胜负，则 $R = 0$）。

在已知进入突然死亡阶段的条件下，各回合是独立试验，每个回合分出胜负的概率都是 β，所以 R 服从几何分布：

$$\mathrm{P}(R = k \mid \text{十球后打和}) = (1-\beta)^{k-1}\beta, \qquad k = 1, 2, 3, \ldots$$

它的平均值是

$$\mathrm{E}(R \mid \text{十球后打和}) = \sum_{k=1}^{\infty} k\beta(1-\beta)^{k-1} = \beta \cdot \frac{1}{\big(1-(1-\beta)\big)^2} = \frac{1}{\beta},$$

这里用到当 $|x| < 1$ 时 $\sum_{k\geq 1} kx^{k-1} = (1-x)^{-2}$，而由于 $0 < \beta \leq 1$，这条公式适用。

如果比赛在首十球内已分出胜负，则 $R = 0$。按射完十球后比数是否相同来取条件，

$$\mathrm{E}(R) = \alpha \cdot \frac{1}{\beta} + (1-\alpha) \cdot 0 = \frac{\alpha}{\beta}.$$

因此射球次数的期望值是

$$\mathrm{E}(10+2R) = 10 + \frac{2\alpha}{\beta},$$

正如所求。

验算

- 两队实力相同，$p_A = p_B = \frac{1}{2}$：由范德蒙恒等式（Vandermonde's identity）$\sum_i \binom{5}{i}^2 = \binom{10}{5} = 252$，所以 $\alpha = 252/2^{10} = \frac{63}{256}$，而 $\beta = \frac{1}{2}$。射球次数的期望值是 $10 + \frac{63}{64} = \frac{703}{64} \approx 10.98$。
- 如果一队百发百中而另一队一球不入（$p_A = 1$，$p_B = 0$），则 $\alpha = 0$（只有 $i = 0$ 的一项有可能不为零，但它含有 $p_B^5 = 0$），所以恰好射 10 球，与常理相符。
- 我们把 α 的公式与直接列举首十球全部 2^{10} 个结果所得的答案作了比对，又把 $10 + 2\alpha/\beta$ 与模拟 2×10^5 场互射十二码的结果作了比对；当 $p_A = \frac{3}{4}$、$p_B = \frac{2}{3}$ 时，公式给出 11.264，模拟则给出 11.25。

常见错误

- 把 β 当成 A 队赢得一个回合的概率 $p_A(1 - p_B)$：其实任何一队赢得一个回合，互射十二码都会结束。
- 把答案写成 $10 + 2/\beta$，忘记了突然死亡阶段只以概率 α 出现。
- 把突然死亡阶段的每一步算作一球，而不是两球。

STEP 之外

$\mathrm{E}(R) = \alpha\,\mathrm{E}(R \mid \text{打和}) + (1-\alpha)\,\mathrm{E}(R \mid \text{未打和})$ 这一步就是全期望定律（law of total expectation），它是大学一年级概率论中最有用的工具之一。突然死亡阶段也是一条带有吸收态（absorbing state）的简单马可夫链（Markov chain），而这里“到达吸收的期望时间”的计算，可以推广到随机漫步（random walk）和赌徒破产（gambler's ruin）问题。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）对条件期望（conditional expectation）和这类马可夫链都有论述。

详细解答

STEP I 2003，第 14 题

概率与统计 · 泊松分布

STEP I 2003，第 14 题

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

step.maths.org/questions/03-s1-q14

关键思路

Jane 见到的朋友数目并不是来访的人数，而是这个人数以 2 为上限后的值。一个设有上限的变量只需考虑三种情况：无人来访、恰好一人来访、两人或以上来访。因此我们根本不必把整个泊松（Poisson）级数加起来，只需求出 $P(0)$ 和 $P(1)$，而“两人或以上”的概率就是余下的部分。至于新朋友，我们按其是否来访作条件分析（conditioning）。

解答

原有的朋友

设 N 为某一天来访的朋友数目，则 $N \sim \mathrm{Po}(2)$，且

$$P(N=k)=\frac{\mathrm{e}^{-2}2^k}{k!}, \qquad k=0,1,2,\ldots$$

Jane 见到 $S=\min(N,2)$ 位朋友。所以当 $N=0$ 时 $S=0$，当 $N=1$ 时 $S=1$，当 $N\geq 2$ 时 $S=2$。现在

$$P(N=0)=\mathrm{e}^{-2}, \qquad P(N=1)=2\mathrm{e}^{-2}, \qquad P(N\geq 2)=1-3\mathrm{e}^{-2},$$

因此

$$\mathrm{E}(S)=0\cdot\mathrm{e}^{-2}+1\cdot 2\mathrm{e}^{-2}+2\left(1-3\mathrm{e}^{-2}\right)=2-4\mathrm{e}^{-2},$$

这正是所要证明的结果。

同一个计算还有一个更简洁的写法，下文会再次用到。对任何取值于 $\{0,1,2,\ldots\}$ 的随机变量（random variable）X，

$$\mathrm{E}\big(\min(X,2)\big)=2-2P(X=0)-P(X=1),$$

因为 $\min(X,2)$ 与 2 的差距，在 $X=0$ 时是 2，在 $X=1$ 时是 1，其余情况则为零。取 $X=N$，便再次得到 $2-2\mathrm{e}^{-2}-2\mathrm{e}^{-2}$。

加入新朋友之后

若新朋友来访，设 $B=1$，否则设 $B=0$，于是 $P(B=1)=p$。题目指出 B 与 N 独立（independent）。现在来访的总人数是 $M=N+B$，而 Jane 见到 $\min(M,2)$ 位朋友。

> **想一想** 我们需要 M 的完整分布吗?

不需要。根据上面的恒等式，我们只需要 $P(M=0)$ 和 $P(M=1)$。利用独立性把概率相乘：

$$\begin{aligned}P(M=0)&=P(N=0)\,P(B=0)=(1-p)\mathrm{e}^{-2},\\P(M=1)&=P(N=1)\,P(B=0)+P(N=0)\,P(B=1)=2(1-p)\mathrm{e}^{-2}+p\,\mathrm{e}^{-2}.\end{aligned}$$

因此

$$\begin{aligned}\mathrm{E}\big(\min(M,2)\big)&=2-2(1-p)\mathrm{e}^{-2}-2(1-p)\mathrm{e}^{-2}-p\,\mathrm{e}^{-2}\\&=2-(4-3p)\mathrm{e}^{-2}.\end{aligned}$$

另一个做法是以 B 为条件。若新朋友没有来访（概率为 $1-p$），平均值一如之前，是 $2-4\mathrm{e}^{-2}$。若新朋友来访（概率为 p），Jane 见到 $\min(N+1,2)$ 位朋友，这个数在 $N=0$ 时是 1，其余情况是 2，其平均值为 $\mathrm{e}^{-2}+2(1-\mathrm{e}^{-2})=2-\mathrm{e}^{-2}$。所以平均值是

$$(1-p)(2-4\mathrm{e}^{-2})+p(2-\mathrm{e}^{-2})=2-(4-3p)\mathrm{e}^{-2}.$$

所以 Jane 现在平均每天见到 $2-(4-3p)\mathrm{e}^{-2}$ 位朋友，比之前多了 $3p\,\mathrm{e}^{-2}$。

验算

- 当 $p=0$ 时，新朋友永不来访，我们重新得到 $2-4\mathrm{e}^{-2}\approx 1.4587$。
- 当 $p=1$ 时，我们得到 $2-\mathrm{e}^{-2}\approx 1.8647$，这正是 $\min(N+1,2)$ 的平均值，与预期相符。
- 答案随 p 递增，而且恒小于 2，这是合理的。
- 按精确分布对 $\min(k+b,2)$ 求和，以及取 $p=0.3$ 模拟 4×10^5 天，两者都得出约 1.580，与 $2-3.1\mathrm{e}^{-2}$ 吻合。

常见错误

- 忘记了最多只见两位朋友的上限，答成 $\mathrm{E}(N) = 2$ 或 $2 + p$。
- 把 $N + B$ 当作泊松变量。伯努利（Bernoulli）变量并非泊松变量，所以两者之和也不是泊松变量。
- 恰好一人来访有两种情况（一位旧朋友来访而新朋友没来，或没有旧朋友来访而新朋友来了），却漏掉其中一种。

STEP 之外

恒等式 $\mathrm{E}(\min(X,2)) = P(X \geq 1) + P(X \geq 2)$ 是尾和公式（tail-sum formula）$\mathrm{E}(X) = \sum_{k\geq 1} P(X \geq k)$ 的特殊情况，该公式适用于取非负整数值的随机变量。它是对生存函数（survival function）积分的离散版本。这类设有上限的平均值常见于排队模型和存货模型，在容量有限的情况下，它们给出获得服务的顾客数目的期望值（expected value）。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）涵盖尾和公式和条件期望（conditional expectation），正是本题所用的两件工具。

详细解答

STEP II 2003，第 1 题

纯数学 · 联立线性方程

> **STEP II 2003，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/03-s2-q1

关键思路

未知数 x 只以 ax 的形式出现，而 y 在每条方程中的系数都是 -1。因此，把相邻的方程相减，就可以完全消去 y，而且所得的两个差式左边相同。这样一来，方程组只在 b 取某一个值时才相容（consistent）；一旦相容，就有一个自由参数（free parameter）。题目余下的部分，就是看怎样好好选取这个参数。

解答

把三条方程编号为 (1)、(2)、(3)。从 (2) 减去 (1)，再从 (3) 减去 (2)，得

$$ax - 2z = 4, \tag{A}$$

$$ax - 2z = b - 7. \tag{B}$$

方程组的任何解都同时满足 (A) 和 (B)，所以只有当 $4 = b - 7$，即 $b = 11$ 时，才可能有解。这一点对 a 的每一个值都成立。

第 (i) 部分

当 $a = 0$ 时，方程变成 $-y - z = 3$、$-y - 3z = 7$、$-y - 5z = b$，x 完全没有出现。

若方程组有解，则由上面的说明可知 $b = 11$。

反过来，假设 $b = 11$。此时方程 (A) 变成 $-2z = 4$，所以 $z = -2$，再由 (1) 得 $y = -3 - z = -1$。以 $y = -1$、$z = -2$ 验算全部三条方程：

$$1 + 2 = 3, \qquad 1 + 6 = 7, \qquad 1 + 10 = 11 = b,$$

所以（不论 x 取甚么值）我们都得到一个解。因此，方程组有解若且唯若（if and only if）$b = 11$。

第 (ii) 部分

现在 $a \neq 0$ 而 $b = 11$。设 $z = \lambda$。由方程 (A) 得 $ax = 4 + 2\lambda$，而由于 $a \neq 0$，我们可以两边相除：

$$x = \frac{4 + 2\lambda}{a}.$$

然后由 (1) 得 $y = ax - z - 3 = 4 + 2\lambda - \lambda - 3 = 1 + \lambda$。

我们只用了 (1) 和 (2) 来求出这些值，所以还必须验证它们也满足 (3)：

$$3ax - y - 5z = 3(4 + 2\lambda) - (1 + \lambda) - 5\lambda = 12 + 6\lambda - 1 - 6\lambda = 11 = b.$$

所以对每一个数 λ，

$$x = \frac{4 + 2\lambda}{a}, \qquad y = 1 + \lambda, \qquad z = \lambda$$

都是一个满足 $z = \lambda$ 的解，正如题目所求。（事实上，这些就是全部的解：任何解的 z 都有某个值 λ，而上面的推导随即确定了 x 和 y。）

想一想 为甚么我们要另外验证方程 (3)?

我们用了 (1) 和 (2)（以 (1) 和 (A) 的形式）来求 x 和 y。三条含三个未知数的方程不一定会自动互相吻合，而 (3) 正是 b 出场的地方。$b = 11$ 就是在这一步验证中用上的。

第 (iii) 部分

当 $a = 2$ 时，由第 (ii) 部分可知所有解都是 $(x, y, z) = (2 + \lambda, 1 + \lambda, \lambda)$。于是

$$\begin{aligned} x^2 + y^2 + z^2 &= (2 + \lambda)^2 + (1 + \lambda)^2 + \lambda^2 \\ &= 3\lambda^2 + 6\lambda + 5 \\ &= 3(\lambda + 1)^2 + 2, \end{aligned}$$

它恰好在 $\lambda = -1$ 时取得最小值 2。所以所求的解是

$$x = 1, \qquad y = 0, \qquad z = -1.$$

从几何角度看，所有解构成一条方向为 $(1, 1, 1)$ 的直线，而我们要找的是直线上最接近原点的一点。事实上 $(1, 0, -1) \cdot (1, 1, 1) = 0$，所以位置向量（position vector）与直线垂直（perpendicular），正如所料。

第 (iv) 部分

根据开首的说明，b 必须是 11。若 $a = 0$，由第 (i) 部分必有 $y = -1$、$z = -2$，于是 $y^2 + z^2 = 5$，太大了。所以 $a \neq 0$，而由第 (ii) 部分，解为 $x = (4 + 2\lambda)/a$、$y = 1 + \lambda$、$z = \lambda$。

> **想一想** x 怎样才可以非常大，而 y 和 z 却保持很小?

a 唯一出现的地方是 $x = (4 + 2\lambda)/a$。选定 λ 就确定了 y 和 z，然后只要 a 足够小，x 就可以要多大有多大。

我们需要 $y^2 + z^2 = (1 + \lambda)^2 + \lambda^2 = 2\lambda^2 + 2\lambda + 1 < 1$, 即 $2\lambda(\lambda + 1) < 0$, 即 $-1 < \lambda < 0$。取 $\lambda = -\frac{1}{2}$：此时 $y = \frac{1}{2}$、$z = -\frac{1}{2}$ 而 $y^2 + z^2 = \frac{1}{2} < 1$，同时 $x = 3/a$。选取

$$a = 10^{-6}$$

便得 $x = 3 \times 10^6 > 10^6$。所以 $a = 10^{-6}$（连同 $b = 11$）符合要求。

更一般地说，当 $-1 < \lambda < 0$ 时，我们有 $2 < 4 + 2\lambda < 4$，所以恰好在 $0 < a < 4 \times 10^{-6}$ 时才存在合适的解。负的 a 不可行，因为那时 $x < 0$。

验算

- 把 $a = 2$、$(x, y, z) = (1, 0, -1)$ 代入原来的方程：$2 - 0 + 1 = 3$、$4 - 0 + 3 = 7$、$6 - 0 + 5 = 11$。
- 对于第 (iv) 部分，取 $a = 10^{-6}$ 和 $(x, y, z) = (3 \times 10^6, \frac{1}{2}, -\frac{1}{2})$：$3 - \frac{1}{2} + \frac{1}{2} = 3$、$6 - \frac{1}{2} + \frac{3}{2} = 7$、$9 - \frac{1}{2} + \frac{5}{2} = 11$。
- 我们用精确分数验证过，对数百组随机选取的 a 和 λ，第 (ii) 部分的公式都满足全部三条方程；另外，对 λ 作网格搜索，也确认了第 (iii) 部分的最小值 2 在 $\lambda = -1$ 时取得。

常见错误

- 由其中两条方程求出 x 和 y 后，忘记验证第三条。
- 在第 (iv) 部分，以为 y 和 z “不可能同时很小”，就断定题目所求不可能做到。其实是可以的：$y - z = 1$ 只能推出 $y^2 + z^2 \geq \frac{1}{2}$。

STEP 之外

这是对线性代数（linear algebra）中秩（rank）与相容性的初步认识。这里的系数矩阵（coefficient matrix）是奇异（singular）的（它的各行满足 $r_1 - 2r_2 + r_3 = 0$），所以方程组要么无解，要么有一整条直线的解，视乎右边的常数项是否满足同一个关系。第 (iv) 部分也说明了为甚么接近奇异的方程组在数值计算中很危险：一个极小的系数 a 会产生一个极大的解。Strang 的 *Introduction to Linear Algebra*（线性代数入门教材）涵盖了以上所有内容，包括像第 (iii) 部分那样的最小范数解（least-norm solution）。

详细解答

STEP II 2003，第 2 题

纯数学 · 反三角函数

STEP II 2003，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/03-s2-q2

关键思路

$a\cos\theta + b\sin\theta$ 这类式子，总可以写成单一个余弦 $R\cos(\theta - \alpha)$。这样一来，方程就变成一个关于 arccos 的命题，而 α 本身则是一个 arctan。因此，只要找到一个漂亮的解 θ，关于 π 的恒等式（identity）不过是把“$\theta = (\theta - \alpha) + \alpha$”改写一下而已。唯一真正要小心的地方，是验证每个角都落在所用反函数（inverse function）的值域（range）之内。

解答

第一个恒等式

先试 $(\frac{1}{4}\pi, \frac{1}{2}\pi)$ 内最容易想到的角。取 $\theta = \frac{1}{3}\pi$，

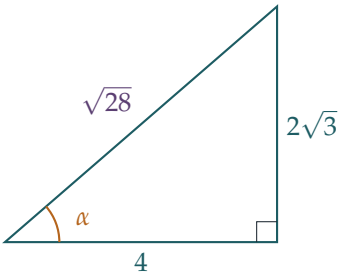
$$4\cos\tfrac{\pi}{3} + 2\sqrt{3}\,\sin\tfrac{\pi}{3} = 4\cdot\tfrac{1}{2} + 2\sqrt{3}\cdot\tfrac{\sqrt{3}}{2} = 2 + 3 = 5,$$

所以 $\theta = \frac{1}{3}\pi$ 满足方程，而且 $\frac{1}{4}\pi < \frac{1}{3}\pi < \frac{1}{2}\pi$。

现在把左方写成 $R\cos(\theta - \alpha) = R\cos\theta\cos\alpha + R\sin\theta\sin\alpha$ 的形式。比较系数，

$$R\cos\alpha = 4, \qquad R\sin\alpha = 2\sqrt{3},$$

所以 $R^2 = 16 + 12 = 28$，得 $R = \sqrt{28}$，而 $\tan\alpha = \frac{2\sqrt{3}}{4} = \frac{\sqrt{3}}{2}$。由于 $\cos\alpha$ 和 $\sin\alpha$ 皆为正数，我们可以取 α 为锐角，所以 $\alpha = \arctan(\sqrt{3}/2)$。



于是方程变成

$$\sqrt{28}\cos(\theta-\alpha)=5,\qquad \text{即}\qquad \cos(\theta-\alpha)=\frac{5}{\sqrt{28}},$$

而 $\theta=\frac{1}{3}\pi$ 满足此方程。

> **想一想** 我们可以立即断定 $\frac{1}{3}\pi-\alpha=\arccos(5/\sqrt{28})$ 吗?

还不可以：$\cos x=c$ 有很多个解，而 $\arccos c$ 只是其中位于 $[0,\pi]$ 的那一个。所以我们必须验证 $\frac{1}{3}\pi-\alpha$ 位于 $[0,\pi]$ 之内。由于 tan 在 $(0,\frac{1}{2}\pi)$ 上是递增的，而 $\tan\alpha=\frac{\sqrt{3}}{2}<\sqrt{3}=\tan\frac{\pi}{3}$，故有 $0<\alpha<\frac{1}{3}\pi$。因此

$$0<\tfrac{1}{3}\pi-\alpha<\tfrac{1}{3}\pi\le\pi,$$

所以 $\frac{1}{3}\pi-\alpha=\arccos(5/\sqrt{28})$。移项后乘以 3,

$$\pi=3\arccos(5/\sqrt{28})+3\alpha=3\arccos(5/\sqrt{28})+3\arctan(\sqrt{3}/2),$$

这正是所要证明的结果。

第二个恒等式

沿用同一个想法：我们想得到 $\frac{1}{4}\pi=\arcsin(7\sqrt{2}/10)-\arctan(3/4)$，所以设 $\beta=\arctan(3/4)$，然后考虑 $\sin(\frac{1}{4}\pi+\beta)$。

由于 $0<\frac{3}{4}$，β 是锐角；再由 3–4–5 直角三角形，可得 $\sin\beta=\frac{3}{5}$ 及 $\cos\beta=\frac{4}{5}$。于是，根据复角公式（addition formula），

$$\begin{aligned}\sin\left(\tfrac{1}{4}\pi+\beta\right)&=\sin\tfrac{\pi}{4}\cos\beta+\cos\tfrac{\pi}{4}\sin\beta\\&=\frac{1}{\sqrt{2}}\left(\frac{4}{5}+\frac{3}{5}\right)=\frac{7}{5\sqrt{2}}=\frac{7\sqrt{2}}{10}.\end{aligned}$$

> **想一想** $\frac{1}{4}\pi+\beta$ 真的就是 arcsin 所给出的那个角吗?

arcsin 的值域是 $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$，所以我们需要 $\frac{1}{4}\pi + \beta$ 落在这个区间内。因为 $\tan\beta = \frac{3}{4} < 1 = \tan\frac{\pi}{4}$，而 tan 在 $(0, \frac{1}{2}\pi)$ 上是递增的，故有 $0 < \beta < \frac{1}{4}\pi$，所以

$$\tfrac{1}{4}\pi < \tfrac{1}{4}\pi + \beta < \tfrac{1}{2}\pi.$$

因此 $\arcsin(7\sqrt{2}/10) = \frac{1}{4}\pi + \arctan(3/4)$，乘以 4 再移项，

$$\pi = 4\arcsin(7\sqrt{2}/10) - 4\arctan(3/4),$$

这就是所求的结果。

验算

以数值计算，$\arctan(\sqrt{3}/2) \approx 0.713724$ 及 $\arccos(5/\sqrt{28}) \approx 0.333473$；两者之和为 $1.047198 \approx \pi/3$，其三倍为 3.141593。至于第二个恒等式，$\arcsin(7\sqrt{2}/10) \approx 1.428899$ 及 $\arctan(3/4) \approx 0.643501$；两者之差为 $0.785398 \approx \pi/4$。快速检查一下亦可见，$\frac{1}{3}\pi$ 是方程在 $(\frac{1}{4}\pi, \frac{1}{2}\pi)$ 内的唯一解：附近的另一个解 $\alpha - \arccos(5/\sqrt{28}) \approx 0.380$ 小于 $\frac{1}{4}\pi$。

常见错误

- 没有验证 $x \in [0, \pi]$ 就写下 $\arccos(\cos x) = x$（或没有验证 $x \in [-\frac{1}{2}\pi, \frac{1}{2}\pi]$ 就写下 $\arcsin(\sin x) = x$）。严谨性的分数正正就在这里。
- 把 $\tan\alpha$ 上下颠倒：由 $R\cos\alpha = 4$、$R\sin\alpha = 2\sqrt{3}$ 得到的是 $\tan\alpha = 2\sqrt{3}/4$，而不是 $4/(2\sqrt{3})$。

STEP 之外

把 π 表示成反正切（inverse tangent）组合的恒等式，是梅钦类公式（Machin-type formulae）的基础，例如 $\frac{1}{4}\pi = 4\arctan\frac{1}{5} - \arctan\frac{1}{239}$。由于 $\arctan x$ 的级数（series）在 x 很小时收敛（converge）得很快，这类公式曾被用来计算 π 至许多个位数。要看清这些恒等式，最简洁的方法是通过复数（complex numbers）：$\arctan(b/a)$ 是 $a + b\mathrm{i}$ 的辐角（argument），所以反正切相加就对应于复数相乘。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）把这个几何观点阐述得十分精彩。

详细解答

STEP II 2003，第 3 题

纯数学 · 无理数与数学归纳法

STEP II 2003，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/03-s2-q3

关键思路

无理性（irrationality）是一种“不是”型的性质，所以最自然的工具是逆否命题（contrapositive）：与其证明无理数（irrational number）的立方根（cube root）是无理数，不如证明立方根若是有理数（rational number），原数本身就必定是有理数。之后的归纳法（induction）全靠这一个事实推动，因为每个 u_{n+1} 都是 u_n 的立方根。至于最后一部分，我们留意到 $u_n \to 1$，于是把整个数列（sequence）平移一个整数，这样每一项仍然是无理数。

解答

无理数的立方根

设 x 为实数，并设 $y = \sqrt[3]{x}$ 为它的实立方根，即 $y^3 = x$。我们证明逆否命题：若 y 是有理数，则 x 是有理数。

假设 $y = \dfrac{p}{q}$，其中 p, q 为整数且 $q \neq 0$。那么

$$x = y^3 = \frac{p^3}{q^3},$$

而 p^3 和 q^3 都是整数，且 $q^3 \neq 0$。所以 x 是两个整数之比，也就是有理数。

因此，有理的立方根只能来自有理数。换言之，若 x 是无理数，则 $\sqrt[3]{x}$ 不可能是有理数，所以它是无理数。这正是我们要证明的结论。

? **想一想** 为甚么不直接处理原命题?

直接证明要从“x 不是整数之比”出发，但这句话没有提供任何具体的东西可供计算。逆否命题则从“$y = p/q$”出发，马上给了我们可以运算的整数，于是一行就证完了。每当假设说某个数*不是*有理数时，都值得试试逆否命题或反证法（proof by contradiction）。

归纳法

由于 $3^{n+1} = 3 \cdot 3^n$，根据指数定律（laws of indices），

$$u_{n+1} = 5^{1/3^{n+1}} = \left(5^{1/3^n}\right)^{1/3} = \sqrt[3]{u_n}.$$

所以每一项都是前一项的立方根。

基础情况。$u_1 = 5^{1/3} = \sqrt[3]{5}$，题目已告诉我们它是无理数。

归纳步骤。假设对某个正整数 k，u_k 是无理数。那么 $u_{k+1} = \sqrt[3]{u_k}$ 是一个无理数的立方根，所以根据第一部分，它是无理数。

因此，根据归纳法，对所有正整数 n，u_n 都是无理数。

收敛至 m 的数列

首先求 u_n 的极限（limit）。取对数（logarithm），

$$\ln u_n = \frac{\ln 5}{3^n} \to 0 \quad \text{当} n \to \infty,$$

而由于指数函数是连续（continuous）的，$u_n = \mathrm{e}^{\ln u_n} \to \mathrm{e}^0 = 1$。此外 $\ln u_n > 0$，所以对所有 n 都有 $u_n > 1$；各项递减并趋向 1：

$$u_1 \approx 1.710, \quad u_2 \approx 1.196, \quad u_3 \approx 1.061, \quad u_4 \approx 1.020, \ldots$$

> ? **想一想** 怎样把极限由 1 移到 m，而又不破坏无理性?

无理数加上一个整数，结果仍是无理数。事实上，若 a 是整数，而 $u + a = r$ 是有理数，那么 $u = r - a$ 也会是有理数，因为两个有理数之差是有理数。

所以，对给定的整数 m，定义

$$a_n = u_n + m - 1 = 5^{1/3^n} + m - 1.$$

由于 $m - 1$ 是整数而 u_n 是无理数，每个 a_n 都是无理数，而且

$$a_n \to 1 + m - 1 = m \quad \text{当} n \to \infty.$$

所以 (a_n) 是一个由无理数组成、收敛（converge）至 m 的无穷数列。这个构造对任何整数 m 都适用，包括 $m = 0$。

（另一个自然的选择是 $m\,u_n$，它同样是无理数并趋向 m，但只限于 $m \neq 0$ 的情况：当 $m = 0$ 时，每一项都是 0，是有理数。所以平移 $m - 1$ 才是较稳妥的答案。）

验算

- 数值上，$u_2^3 \approx 1.19581^3 \approx 1.70998 \approx u_1$，印证了 $u_{n+1} = \sqrt[3]{u_n}$。
- 取 $m = 4$：$a_1 \approx 4.710$，$a_2 \approx 4.196$，$a_3 \approx 4.061$，而 a_{20} 与 4 相符至小数点后八位以上。
- 差 $a_n - m = u_n - 1$ 是正数，大小约为 $\ln 5/3^n$，所以每一步大约缩小至原来的 3 分之一。

常见错误

- 证明了逆命题（converse）（“无理数的立方是无理数”），而这个命题是错的：$\left(\sqrt[3]{2}\right)^3 = 2$。
- 使用 $m\,u_n$，却没有留意到它在 $m = 0$ 时失效。
- 忘记说明为何 $u_n \to 1$；“指数趋向零”这句话还需要交代一下连续性，或者用上面的对数论证。

STEP 之外

这道题显示无理数在实数线上是稠密（dense）的：每一个实数，而不仅是每一个整数，都是无理数的极限（试试 $q_n + \sqrt{2}/n$，其中有理数 $q_n \to x$）。有理数同样稠密，但在某个精确的意义下，无理数远比有理数多：有理数是可数（countable）的，而实数则不是。Abbott 的 *Understanding Analysis* 在开首几章对稠密性、可数性和数列的极限有十分精彩的讲解。

详细解答

STEP II 2003，第 4 题

纯数学 · 弓形的面积

STEP II 2003，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/03-s2-q4

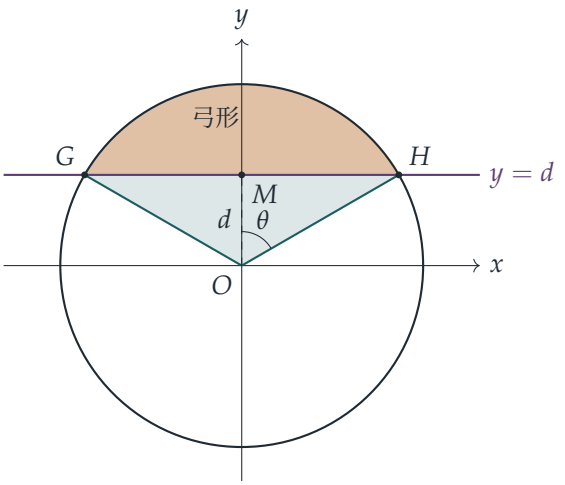
关键思路

劣弓形（minor segment）就是扇形（sector）减去一个三角形，所以第一个结果不过是“扇形减三角形”。至于其余部分，留意在 $(*)$ 中，d 这个数其实是圆心到直线的垂直距离（*perpendicular distance*）。把图形平移或旋转都不会改变任何面积，所以在每一种情况下，我们都保留 $(*)$，只把 d 换成该距离。全部工夫都在于计算这个距离。

解答

基本结果

由于直线与圆相交于两点，故 $0 < d < R$。把 $y = d$ 代入 $x^2 + y^2 = R^2$ 得 $x = \pm\sqrt{R^2 - d^2}$，所以 $G = \left(-\sqrt{R^2 - d^2},\, d\right)$ 及 $H = \left(\sqrt{R^2 - d^2},\, d\right)$。设 $M = (0, d)$ 为 GH 的中点，并设 $\theta = \angle MOH$。



在直角三角形 OMH 中，斜边（hypotenuse）为 $OH = R$，邻边为 $OM = d$，所以 $\cos\theta = d/R$。由于 $0 < d/R < 1$，我们有 $\theta = \arccos(d/R)$，其中 $0 < \theta < \frac{\pi}{2}$，而根据对称性，$\angle GOH = 2\theta < \pi$。

圆心角为 2θ 的扇形 OGH，面积为

$$\tfrac{1}{2}R^2(2\theta) = R^2\arccos\left(\frac{d}{R}\right).$$

三角形 OGH 的底为 $GH = 2\sqrt{R^2 - d^2}$，高为 $OM = d$，所以其面积为 $d\sqrt{R^2 - d^2}$。因为 $2\theta < \pi$，圆心 O 与劣弧（minor arc）分别位于 GH 的两侧，所以扇形恰好就是三角形加上劣弓形。因此

$$\text{劣弓形的面积} = R^2\arccos\left(\frac{d}{R}\right) - d\sqrt{R^2 - d^2},$$

这就是 $(*)$。

想一想 从几何的角度看，$(*)$ 中的 d 究竟代表甚么？

上述论证完全没有用到坐标轴：它只用到半径 R，以及圆心到弦（chord）的距离 $OM = d$。把整个图形平移（translation）或旋转（rotation），这两个量都保持不变，面积亦保持不变。所以对于任何半径为 R 的圆，以及任何与圆心的垂直距离为 p 的直线，其中 $0 \le p < R$，劣弓形的面积为

$$R^2\arccos\left(\frac{p}{R}\right) - p\sqrt{R^2 - p^2}. \qquad (\dagger)$$

（若 $p = 0$，直线就是直径，两部分是相等的两半，而 $(\dagger)$ 正确地给出 $\frac{1}{2}\pi R^2$。）在每一部分中，我们求出 p，再代入 $(\dagger)$。

第 (i) 部分

圆心为 (a,b)，而直线 $y=c$ 是水平线，所以圆心到直线的距离为 $p=|c-b|$。把 d 换成 $|c-b|$：

$$R^2\arccos\left(\frac{|c-b|}{R}\right)-|c-b|\sqrt{R^2-(c-b)^2}.$$

这里的绝对值（modulus）不可省略：c 可能小于 b。

第 (ii) 部分

这里圆心为 O。我们需要求 O 到 $y=mx+c$ 的距离。设直线与 x 轴的夹角为 α，故 $\tan\alpha=m$，并设直线与 y 轴相交于 $C=(0,c)$。由 O 向直线作垂线，垂足为 F。OC（铅垂）与 OF（垂直于直线）之间的夹角，等于水平线与该直线之间的夹角，即 α。所以在直角三角形 OFC 中，

$$p=OF=|c|\,|\cos\alpha|=\frac{|c|}{\sqrt{1+\tan^2\alpha}}=\frac{|c|}{\sqrt{1+m^2}}.$$

（这就是常用公式 $|Ax_0+By_0+C|/\sqrt{A^2+B^2}$ 应用于直线 $mx-y+c=0$ 和点 $(0,0)$ 的结果。）于是

$$\sqrt{R^2-p^2}=\sqrt{R^2-\frac{c^2}{1+m^2}}=\frac{\sqrt{R^2(1+m^2)-c^2}}{\sqrt{1+m^2}},$$

所以劣弓形的面积为

$$R^2\arccos\left(\frac{|c|}{R\sqrt{1+m^2}}\right)-\frac{|c|\sqrt{R^2(1+m^2)-c^2}}{1+m^2}.$$

直线与圆相交，当且仅当 $c^2<R^2(1+m^2)$，所以平方根是实数。

第 (iii) 部分

设 $X=x-a$、$Y=y-b$，把原点移到圆心。圆变成 $X^2+Y^2=R^2$，而直线变成

$$Y+b=m(X+a)+c \quad\Longleftrightarrow\quad Y=mX+k, \qquad k=c+ma-b.$$

这就是把 c 换成 k 的第 (ii) 部分，而平移不会改变面积。所以 $p=|c+ma-b|/\sqrt{1+m^2}$，劣弓形的面积为

$$R^2\arccos\left(\frac{|c+ma-b|}{R\sqrt{1+m^2}}\right)-\frac{|c+ma-b|\sqrt{R^2(1+m^2)-(c+ma-b)^2}}{1+m^2}.$$

代入 $m=0$ 便得回第 (i) 部分，代入 $a=b=0$ 便得回第 (ii) 部分。

想一想 为甚么在第 (i) 部分不能直接写 $c-b$ 而不加绝对值?

若 $c < b$，则 $(c-b)/R$ 为负数，arccos 给出的角是大于 $\frac{\pi}{2}$ 的 $\pi - \theta$，而第二项亦会变号。该式变成 $R^2(\pi-\theta) + p\sqrt{R^2-p^2}$，这是优弓形（major segment）的面积，而不是劣弓形的面积。

验算

- 取 $R = 2$、$d = 1$，得 $\theta = \frac{\pi}{3}$，而 $(*)$ 给出 $\frac{4\pi}{3} - \sqrt{3} \approx 2.4567$。这小于圆盘面积的一半（$2\pi \approx 6.28$），与预期相符。
- 当 $d \to R$ 时，面积趋向 0；而当 $d = 0$ 时，面积等于 $\frac{1}{2}\pi R^2$。
- 我们随机选取了数百组圆和直线，把全部四条公式与直接对弦长作数值积分（numerical integration）的结果比较，两者相差不超过 2×10^{-6}。

常见错误

- 漏掉绝对值。每当圆心位于直线“错”的一侧，带正负号的距离便会给出优弓形的面积。
- 在第 (ii) 和第 (iii) 部分，用了铅垂距离 $|c|$，而不是垂直距离 $|c|/\sqrt{1+m^2}$。
- 用 $\frac{1}{2}R^2\theta$ 计算扇形面积，忘记了圆心角是 2θ 而不是 θ。

STEP 之外

这道题建基于一个事实：面积在等距变换（isometry）之下保持不变，等距变换即平面上的平移、旋转和反射（reflection）。在大学一年级的课程中，这会变成以下的说法：等距变换的雅可比行列式（Jacobian determinant）的绝对值为 1，所以二重积分（double integral）的变量代换公式不会改变面积。把等距变换分类，并找出它们所保持的量，是进入几何学和群论（group theory）的一条好途径。Strang 的 *Introduction to Linear Algebra*（线性代数入门）讲述了旋转矩阵，以及行列式作为面积比例因子的角色。

详细解答

STEP II 2003，第 5 题

纯数学 · 直线的向量几何

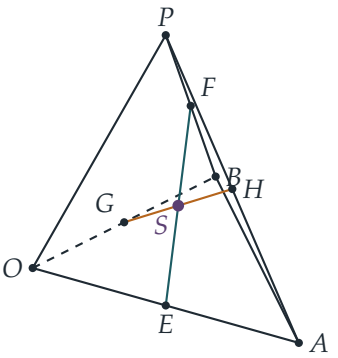
STEP II 2003，第 5 题

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/03-s2-q5

关键思路

先写出四个中点，把每条直线写成“点 + 参数 × 方向”的形式，然后令两者相等。三维空间中的两条直线通常不会相交，因此真正要做的，是证明这三条含两个未知参数的分量方程是相容（consistent）的。**k** 分量很快就能解决问题。从几何上看，O、A、B、P 构成一个四面体（tetrahedron），而 EF 和 GH 各自连接一对对棱的中点。由此可以猜想，它们相交于四个顶点的“平均”位置。



解答

直线 EF 与 GH 相交

设 $\mathbf{p} = l\mathbf{i} + m\mathbf{j} + n\mathbf{k}$。各中点的位置向量（position vector）为

$$\mathbf{e} = \tfrac{1}{2}a\mathbf{i}, \qquad \mathbf{f} = \tfrac{1}{2}\big(l\mathbf{i} + (b+m)\mathbf{j} + n\mathbf{k}\big),$$
$$\mathbf{g} = \tfrac{1}{2}b\mathbf{j}, \qquad \mathbf{h} = \tfrac{1}{2}\big((a+l)\mathbf{i} + m\mathbf{j} + n\mathbf{k}\big).$$

EF 上的一般点是 $\mathbf{e} + s(\mathbf{f} - \mathbf{e})$，而 GH 上的一般点是 $\mathbf{g} + t(\mathbf{h} - \mathbf{g})$。把两者都乘以二以消去二分之一，得

$$2\big(\mathbf{e} + s(\mathbf{f} - \mathbf{e})\big) = \big(a + s(l-a)\big)\mathbf{i} + s(b+m)\mathbf{j} + sn\,\mathbf{k},$$
$$2\big(\mathbf{g} + t(\mathbf{h} - \mathbf{g})\big) = t(a+l)\mathbf{i} + \big(b + t(m-b)\big)\mathbf{j} + tn\,\mathbf{k}.$$

两条直线相交，当且仅当存在 s 和 t 的值使以上两式相等，因此以下三式必须全部成立：

$$\mathbf{i}: \quad a + s(l-a) = t(a+l),$$
$$\mathbf{j}: \quad s(b+m) = b + t(m-b),$$
$$\mathbf{k}: \quad sn = tn.$$

想一想 三条方程、两个未知数：为甚么它们竟然会有解?

一般而言是没有解的，这正是空间中两条直线通常不会相交的原因。这里我们先解其中两条方程，再验证第三条也成立。由于 $n \neq 0$，$\mathbf{k}$ 方程给出 $s = t$。于是 $\mathbf{i}$ 方程变成 $a + sl - sa = sa + sl$，即 $a = 2sa$，而由于 $a \neq 0$，我们得到 $s = \frac{1}{2}$。最后，当 $s = t = \frac{1}{2}$ 时，$\mathbf{j}$ 方程成为

$$\tfrac{1}{2}(b+m) = b + \tfrac{1}{2}(m-b) = \tfrac{1}{2}(b+m),$$

这是成立的。所以当 $s = t = \frac{1}{2}$ 时三条方程全部成立，直线 EF 与 GH 相交，正如所求。

交点是 EF 的中点（也是 GH 的中点）：

$$\mathbf{s} = \tfrac{1}{4}\big((a+l)\mathbf{i} + (b+m)\mathbf{j} + n\mathbf{k}\big) = \tfrac{1}{4}\big(\mathbf{0} + a\mathbf{i} + b\mathbf{j} + \mathbf{p}\big),$$

即 O、A、B 和 P 的位置向量的平均。

点 T

由于 S 是 DT 的中点，我们有 $\mathbf{s} = \frac{1}{2}(\mathbf{d} + \mathbf{t})$，所以 $\mathbf{t} = 2\mathbf{s} - \mathbf{d}$：

$$\mathbf{t} = \tfrac{1}{2}(a+l)\,\mathbf{i} + \tfrac{1}{2}(b+m)\,\mathbf{j} + \big(\tfrac{1}{2}n - d\big)\mathbf{k}.$$

平面 OAB 包含 O 以及向量 $a\mathbf{i}$ 和 $b\mathbf{j}$。由于 a 和 b 均不为零，这两个向量分别沿 x 轴和 y 轴的方向，所以平面 OAB 就是平面 $z = 0$。因此 T 位于该平面上，当且仅当它的 $\mathbf{k}$ 分量

为零，即 $\frac{1}{2}n - d = 0$。故此

$$d = \tfrac{1}{2}n.$$

由于 $n \neq 0$，这个值不为零，所以与条件 $d \neq 0$ 并无矛盾。

验算

- 取 $a = 2, b = 4, \mathbf{p} = (2,0,4)$。则 $E = (1,0,0), F = (1,2,2), G = (0,2,0), H = (2,0,2)$。$EF$ 和 GH 的中点都是 $(1,1,1)$，亦即 $\frac{1}{4}\big((0,0,0) + (2,0,0) + (0,4,0) + (2,0,4)\big)$。
- 取 $d = 2 = \frac{1}{2}n$，则 $T = 2(1,1,1) - (0,0,2) = (2,2,0)$，它位于 $z = 0$ 上。
- 我们以电脑用精确的有理数运算，对 20 000 组随机选取的 a, b, l, m, n, d 作了检验，确认两条直线在 $s = t = \frac{1}{2}$ 时相交于上述的点，并且 T 位于平面 OAB 上（纯量三重积（scalar triple product）为零）当且仅当 $d = \frac{1}{2}n$。

常见错误

- 一开始就用同一个参数来令两条直线相等。这样做等于预先假设了答案 $s = t$，而在这里它是由 **k** 方程推出来的。
- 解了其中两条分量方程就停下来。要证明空间中两条直线相交，还必须验证第三条方程。
- 忘记说明在哪里用到 $a \neq 0$ 和 $n \neq 0$：正是有了这两个条件，我们才可以作除法求出 s 和 t。

STEP 之外

任何四面体中，连接对棱中点的三条直线（称为双中线（bimedians））全都通过它的形心（centroid），即四个顶点的平均，并且各自在该点被平分。这是“三角形的中线（median）交于形心”在三维空间中的近亲，也让我们初次体会到把点看成加权平均的想法，而这正是重心坐标（barycentric coordinates）和仿射坐标（affine coordinates）背后的概念。Strang 的 *Introduction to Linear Algebra*（线性代数入门）详细阐述了这里所用的向量与线性方程观点，包括方程数目多于未知数的方程组在甚么情况下相容。

详细解答

STEP II 2003，第 6 题

纯数学 · 迭代函数与积分

STEP II 2003，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/03-s2-q6

关键思路

再施行一次 f，意思就是“先减去 1，再把负的部分反射回横轴上方”。若某一段图像的值介乎 0 与 1 之间，这样做只会把该段在带状区域 $0 \le y \le 1$ 内上下倒转，而两条长长的直线臂则各多出一个锯齿。于是图像变成锯齿形（三角波，triangle wave）；看出这一点之后，最后一部分的两个积分都只是若干个相同周期上的积分之和。

解答

略图

整个解答只靠以下两项事实。

递推关系（*recurrence*）。由于 $g_{n+1} = f \circ g_n$，我们有 $g_{n+1}(x) = |g_n(x) - 1|$：把 g_n 的图像向下平移 1，再把位于 x 轴下方的部分反射上去。

对称性（*symmetry*）。由于 $f(2-x) = |1-x| = f(x)$，当 $n \ge 2$ 时可得 $g_n(2-x) = g_{n-1}\big(f(2-x)\big) = g_{n-1}\big(f(x)\big) = g_n(x)$，而 $n = 1$ 时则可直接验证。所以每个 g_n 都关于直线 $x = 1$ 对称，只需弄清楚 $x \ge 1$ 的情况便可。

先从 $g_1(x) = |x - 1|$ 开始，这是一个顶点在 $(1, 0)$ 的 V 形。

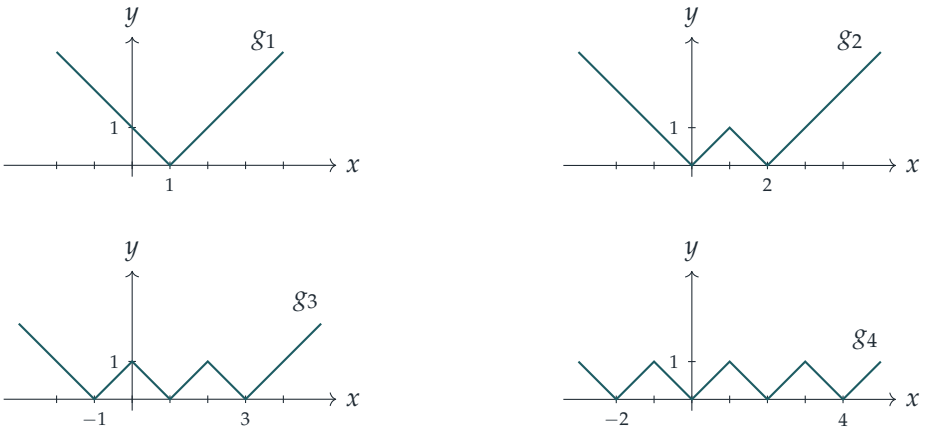
至于 g_2：在 $[0, 2]$ 上有 $0 \le g_1 \le 1$，所以 $g_2 = 1 - |x - 1|$，是一个帐篷形，在 $x = 1$ 处升至 1，在 $x = 0$ 和 $x = 2$ 处降至 0。当 $x \ge 2$ 时，$g_2 = x - 2$；由对称性，当 $x \le 0$ 时 $g_2 = -x$。所以 g_2 呈 W 形。

之后每一步都是做同样的事：锯齿部分（取值在 $[0, 1]$ 内）上下倒转，而直线臂 $x - n$ 变成 $|x - n - 1|$，在 $[n, n+1]$ 上多添一个锯齿（由对称性，在 $[1-n, 2-n]$ 上亦然）。以

$d_{\text{even}}(x)$ 和 $d_{\text{odd}}(x)$ 分别表示 x 与最接近的偶数及最接近的奇数之间的距离，便得

$$g_n(x) = \begin{cases} 2-n-x, & x \le 2-n, \\ d_{\text{even}}(x)\text{(若 } n \text{ 为偶数)}，\ d_{\text{odd}}(x)\text{(若 } n \text{ 为奇数)}， & 2-n \le x \le n, \\ x-n, & x \ge n. \end{cases}$$

这个结果可由递推关系用归纳法（induction）证明：对任意实数 x 都有 $d_{\text{even}} + d_{\text{odd}} = 1$（在两个相邻整数之间，这两个距离分别是 t 和 $1-t$），所以“1 减去锯齿”会把奇偶性对调；而在 $[n, n+1]$ 上，新增的一段 $n+1-x$ 正好是到 $n+1$ 的距离，而 $n+1$ 就是具有新奇偶性的最接近整数。左边的部分则由关于 $x=1$ 的对称性得出。



用文字描述：g_3 的零点在 $-1, 1, 3$，在 0 和 2 处有高度为 1 的尖峰，两臂为 $y = -1 - x$ 和 $y = x - 3$；g_4 的零点在 $-2, 0, 2, 4$，在 $-1, 1, 3$ 处有高度为 1 的尖峰，两臂为 $y = -2 - x$ 和 $y = x - 4$。四个图像都关于 $x = 1$ 对称。

积分

设 n 为偶数。由于 $n \ge 2$，区间 $[0, n]$ 包含在 $[2-n, n]$ 之内，所以在这个区间上 $g_n(x) = d_{\text{even}}(x)$。

> **想一想** 为甚么只需在 $[0, 2]$ 上积分便足够?

两个函数的周期（period）都是 2：$d_{\text{even}}(x+2) = d_{\text{even}}(x)$，因为平移 2 会把偶数映到偶数；而 $\left|\sin \frac{\pi(x+2)}{2}\right| = \left|-\sin \frac{\pi x}{2}\right| = h(x)$。由于 n 是偶数，$[0, n]$ 正好由 $\frac{n}{2}$ 个完整周期 $[0,2], [2,4], \ldots, [n-2, n]$ 组成，所以在 $[0, n]$ 上的每个积分都是在 $[0, 2]$ 上的积分的 $\frac{n}{2}$ 倍。

在 $[0, 2]$ 上，g_n 就是帐篷形 $1 - |x-1|$：一个底为 2、高为 1 的三角形，所以

$$\int_0^2 g_n(x)\,\mathrm{d}x = \tfrac{1}{2} \cdot 2 \cdot 1 = 1.$$

在 $[0,2]$ 上，$\frac{\pi x}{2} \in [0,\pi]$，所以 $\sin\frac{\pi x}{2} \geq 0$，绝对值（modulus）符号可以略去：

$$\int_0^2 h(x)\,\mathrm{d}x = \int_0^2 \sin\frac{\pi x}{2}\,\mathrm{d}x = \Big[-\frac{2}{\pi}\cos\frac{\pi x}{2}\Big]_0^2 = \frac{2}{\pi}+\frac{2}{\pi} = \frac{4}{\pi}.$$

因此

$$\int_0^n \big(h(x)-g_n(x)\big)\,\mathrm{d}x = \frac{n}{2}\Big(\frac{4}{\pi}-1\Big) = \frac{2n}{\pi}-\frac{n}{2},$$

这正是所求。

验算

- 取一个数值，亲手迭代一次：$g_4(-1)$：$-1 \to 2 \to 1 \to 0 \to 1$，所以 $g_4(-1)=1$，是一个尖峰，与略图相符。
- 答案是正数，因为 $4/\pi \approx 1.27 > 1$：正弦拱形比同底同高的直边帐篷形向外鼓出，这与两个图像的形状吻合。
- 数值方面，取 $n=2$ 时用辛普森法则（Simpson's rule）计算得 0.273240，与 $\frac{4}{\pi}-1 \approx 0.273240$ 一致；当 $n=4,6,8,10$ 时同样吻合。

常见错误

- 忘记左边的部分：当 $n \geq 3$ 时，在 $x<0$ 的范围内图像并不是直线。利用对称性 $g_n(2-x)=g_n(x)$ 是把这部分画对的最快方法。
- 在 $[0,n]$ 上积分 $\sin\frac{\pi x}{2}$ 时漏掉绝对值，这样对偶数 n 会得出 0，而不是 $\frac{2n}{\pi}$。

STEP 之外

把一个简单的映射反复迭代，观察图像如何变化，正是离散动力系统（discrete dynamical systems）的起点。这里的锯齿最终稳定成一个三角波，而与之密切相关的“帐篷映射”（tent map）则是混沌（chaos）的经典例子。把 h 与三角波作比较，也引向傅里叶级数（Fourier series）：在那里，三角波会写成正弦与余弦之和，而 $|\sin|$ 是最早遇到的例子之一。Boyce 与 DiPrima 合著的 *Elementary Differential Equations*（初等微分方程）对傅里叶级数有一段易读的入门介绍。

详细解答

STEP II 2003，第 7 题

纯数学 · 广义积分与级数

STEP II 2003，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/03-s2-q7

关键思路

形如 $\ln x \times$ (幂，底数为x) 的被积函数（integrand），最适合用分部积分法（integration by parts）处理：对对数求导，它便会消失。第一个积分就这样解决了。题目余下的部分把精确值 $2/(n^2\mathrm{e})$ 反过来用：它让我们可以把 $1/n^2$ 写成一个积分。然后我们把所有积分下限移到同一位置，这样只会令每个积分变大，再在积分号内把一个等比级数（geometric series）加起来。

解答

积分

对 $a > 1$ 及 $n > 0$，设

$$I_n(a) = \int_a^\infty \frac{\ln x}{x^{n+1}}\,\mathrm{d}x.$$

取 $u = \ln x$ 及 $\mathrm{d}v = x^{-n-1}\,\mathrm{d}x$ 作分部积分，于是 $v = -\frac{1}{n}x^{-n}$（这里用到 $n \neq 0$）。在 $[a, X]$ 上，

$$\begin{aligned}\int_a^X \frac{\ln x}{x^{n+1}}\,\mathrm{d}x &= \left[-\frac{\ln x}{nx^n}\right]_a^X + \frac{1}{n}\int_a^X x^{-n-1}\,\mathrm{d}x \\ &= \frac{\ln a}{na^n} - \frac{\ln X}{nX^n} + \frac{1}{n^2}\left(\frac{1}{a^n} - \frac{1}{X^n}\right).\end{aligned}$$

现在令 $X \to \infty$。由于 $n > 0$，$X^{-n} \to 0$。至于另一项，设 $Y = X^n$，则 $\ln X = \frac{1}{n}\ln Y$，并且

$$\frac{\ln X}{X^n} = \frac{1}{n}\cdot\frac{\ln Y}{Y} \to 0,$$

因为当 $X \to \infty$ 时 $Y \to \infty$，而题目容许我们假设 $(\ln Y)/Y \to 0$。（这里不能直接引用 $(\ln X)/X \to 0$：当 $0 < n < 1$ 时，X^n 增长得比 X 慢。）所以这个积分收敛（converge），并且

$$I_n(a) = \frac{\ln a}{na^n} + \frac{1}{n^2 a^n}.$$

当 $a = \mathrm{e}^{1/n}$ 时，$a^n = \mathrm{e}$ 且 $\ln a = 1/n$，所以

$$I_n\left(\mathrm{e}^{1/n}\right) = \frac{1}{n^2\mathrm{e}} + \frac{1}{n^2\mathrm{e}} = \frac{2}{n^2\mathrm{e}},$$

这正是所求的结果。

比较两个积分

若 $1 < a < b$，则

$$I_n(a) - I_n(b) = \int_a^b \frac{\ln x}{x^{n+1}}\,\mathrm{d}x.$$

对 $x > 1$，有 $\ln x > 0$ 及 $x^{n+1} > 0$，所以被积函数在 $[a, b]$ 上严格为正。它同时是连续的，而且 $b > a$，所以这个积分严格为正。因此 $I_n(b) < I_n(a)$：把积分下限向右移，等于舍弃了一块正的面积。

不等式

由第一部分，对每个正整数 n，

$$\frac{1}{n^2} = \frac{\mathrm{e}}{2} I_n\left(\mathrm{e}^{1/n}\right).$$

想一想 和式中每一项的积分下限 $\mathrm{e}^{1/n}$ 各不相同，但右边只有单一个下限 $\mathrm{e}^{1/N}$。我们怎样由前者过渡到后者？

取 $1 \le n \le N$。那么 $1/N \le 1/n$，所以 $1 < \mathrm{e}^{1/N} \le \mathrm{e}^{1/n}$。当 $n < N$ 时不等式是严格的，而第二部分（取 $a = \mathrm{e}^{1/N}$，$b = \mathrm{e}^{1/n}$）给出

$$I_n\left(\mathrm{e}^{1/n}\right) < I_n\left(\mathrm{e}^{1/N}\right);$$

当 $n = N$ 时两边相等。对 $n = 1, \ldots, N$ 求和，所得的不等式是严格的，因为 $N > 1$ 意味着最少有一项（例如 $n = 1$）满足 $n < N$：

$$\sum_{n=1}^{N} \frac{1}{n^2} = \frac{\mathrm{e}}{2} \sum_{n=1}^{N} I_n\left(\mathrm{e}^{1/n}\right) < \frac{\mathrm{e}}{2} \sum_{n=1}^{N} I_n\left(\mathrm{e}^{1/N}\right).$$

这正是需要用到 $N > 1$ 的地方。若 $N = 1$，两边便会相等。

余下的工作是把这些积分加起来。有限个在同一范围上收敛的积分之和，等于各被积函数之和的积分，所以

$$\sum_{n=1}^{N} I_n\left(\mathrm{e}^{1/N}\right) = \int_{\mathrm{e}^{1/N}}^{\infty} \left(\sum_{n=1}^{N} x^{-(n+1)} \right) \ln x\,\mathrm{d}x.$$

括号内是一个等比级数，首项为 x^{-2}，公比为 $x^{-1} \neq 1$（因为 $x > 1$），共 N 项：

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

这里分子和分母同乘以 x^2。因此

$$\sum_{n=1}^{N} \frac{1}{n^2} < \frac{\mathrm{e}}{2} \int_{\mathrm{e}^{1/N}}^{\infty} \left(\frac{1 - x^{-N}}{x^2 - x}\right) \ln x \, \mathrm{d}x,$$

正如所求。

验算

- 把 $I_n(a)$ 的公式对 a 求导：你应该得到 $-\ln a / a^{n+1}$，即被积函数的相反数，正如微积分基本定理（fundamental theorem of calculus）所要求的。
- 数值上，利用代换（substitution）$x = \mathrm{e}^u$，最后一部分的积分变成 $\int_{1/N}^{\infty} u(1 - \mathrm{e}^{-Nu})/(\mathrm{e}^u - 1)\,\mathrm{d}u$。当 $N = 2$ 时，右边约为 1.4865，而左边是 $1 + \frac{1}{4} = 1.25$。当 $N = 10$ 时，右边约为 2.0579，而左边是 1.5498。
- 当 $N \to \infty$ 时，右边趋向 $\frac{\mathrm{e}}{2} \int_1^{\infty} \frac{\ln x}{x^2 - x}\,\mathrm{d}x$，而代换 $x = 1/t$ 把它化为 $\frac{\mathrm{e}}{2} \int_0^1 \frac{-\ln t}{1-t}\,\mathrm{d}t = \frac{\mathrm{e}}{2} \cdot \frac{\pi^2}{6} \approx 2.236$。这个值明显大于 $\pi^2/6 \approx 1.645$，而这是必然的。

常见错误

- 直接用 $(\ln X)/X \to 0$ 来消去 $(\ln X)/X^n$。这对 $n \geq 1$ 行得通，但题目的结果是对所有 $n > 0$ 而言的。
- 把积分下限移错方向。由于 $\mathrm{e}^{1/N}$ 是各个 $\mathrm{e}^{1/n}$ 之中最小的，每个积分都会变大。
- 声称不等式是严格的，却没有说明原因。$n = N$ 的一项给出等式，所以严格性来自 $n < N$ 的各项，而这些项之所以存在，是因为 $N > 1$。

STEP 之外

把一个数写成积分，然后在积分号内求和，是分析（analysis）中的标准手法。把这个做法推到极限，便得到 $\sum 1/n^2 = \int_0^1 \frac{-\ln t}{1-t}\,\mathrm{d}t$，这是通往 $\pi^2/6$ 的其中一条路径。令 $N \to \infty$，就是把无穷和与反常积分（improper integral）交换次序。这个交换需要论证，要用到单调收敛定理或控制收敛定理（monotone or dominated convergence theorems），你会在大学的第一门分析或测度论（measure theory）课程中遇到它们。Abbott 的 *Understanding Analysis* 写得平易近人，很适合用来了解这类交换为何可能失败，以及何时才是安全的。

详细解答

STEP II 2003，第 8 题

纯数学 · 可分离变量微分方程

STEP II 2003，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/03-s2-q8

关键思路

这是一阶可分离变量（separable）的微分方程：$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}t}$ 等于 t 的一个有理函数（rational function）。把那个二次式除以 $t+1$，这个函数便分拆成一个多项式加一个简单分式，两者都可以立即积分。至于驻点（stationary point），直接由微分方程本身便可得出，无须对解求导：$\frac{\mathrm{d}y}{\mathrm{d}t}=0$ 恰好在二次式 t^2-3t+2 等于零之处成立。

解答

求 y

由多项式除法（polynomial division）得

$$\frac{t^2-3t+2}{t+1}=t-4+\frac{6}{t+1},$$

因为 $(t+1)(t-4)+6=t^2-3t-4+6=t^2-3t+2$。

解必须由 $y=A>0$ 出发。假如 y 在某处等于零，则根据这个线性方程的解的唯一性（uniqueness），它必定是零解，这与 $y(0)=A$ 矛盾。所以自始至终 $y>0$（在 $t>-1$ 上，即系数有定义的范围内），因此可以除以 y：

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}t}=-k\left(t-4+\frac{6}{t+1}\right).$$

由 0 积分至 t，并利用 $\ln 1=0$，

$$\ln\frac{y}{A}=-k\left(\frac{t^2}{2}-4t+6\ln(t+1)\right),$$

所以

$$\boxed{y = A\,(t+1)^{-6k}\,\mathrm{e}^{-k(t^2/2-4t)} = A\,(t+1)^{-6k}\,\mathrm{e}^{k(8t-t^2)/2}}.$$

驻点值

由微分方程得

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -k\,\frac{(t-1)(t-2)}{t+1}\,y.$$

我们假设 $k \neq 0$（否则 $y \equiv A$，每一点都是驻点）。由于 $y > 0$ 及 $t+1 > 0$，导数只在 $t = 1$ 和 $t = 2$ 处等于零，而且因为两者都是单根（simple root），导数在这两处都会变号。所以恰好有两个驻点值（stationary value）：

$$y(1) = A \cdot 2^{-6k}\,\mathrm{e}^{-k(1/2-4)} = A \cdot 2^{-6k}\,\mathrm{e}^{7k/2},$$
$$y(2) = A \cdot 3^{-6k}\,\mathrm{e}^{-k(2-8)} = A \cdot 3^{-6k}\,\mathrm{e}^{6k}.$$

两者之比为

$$\frac{y(1)}{y(2)} = \frac{3^{6k}}{2^{6k}}\,\mathrm{e}^{7k/2-6k} = \left(\frac{3}{2}\right)^{6k}\mathrm{e}^{-5k/2},$$

正是所求。

> ? **想一想** 两个驻点值之中，哪一个较大?

取对数：

$$\ln\frac{y(1)}{y(2)} = k\left(6\ln\tfrac{3}{2} - \tfrac{5}{2}\right),$$

而 $6\ln 1.5 \approx 2.433 < 2.5$，所以括号内的数是负数（约为 -0.067）。因此当 $k > 0$ 时 $y(1) < y(2)$，当 $k < 0$ 时 $y(1) > y(2)$。两者相差很小，所以图像上的起伏十分平缓。同样地，

$$\ln\frac{y(1)}{A} = k\left(\tfrac{7}{2} - 6\ln 2\right) \approx -0.659k, \qquad \ln\frac{y(2)}{A} = k(6 - 6\ln 3) \approx -0.592k,$$

所以当 $k > 0$ 时两个驻点值都小于 A，当 $k < 0$ 时两个都大于 A。

当 $t \to +\infty$ 时的表现

写成

$$\ln\frac{y}{A} = -k\left(\frac{t^2}{2} - 4t + 6\ln(t+1)\right).$$

当 $t \to \infty$ 时，括号内的式子趋向 $+\infty$，并由 $\frac{1}{2}t^2$ 主导（ln 项的增长慢得多）。所以：

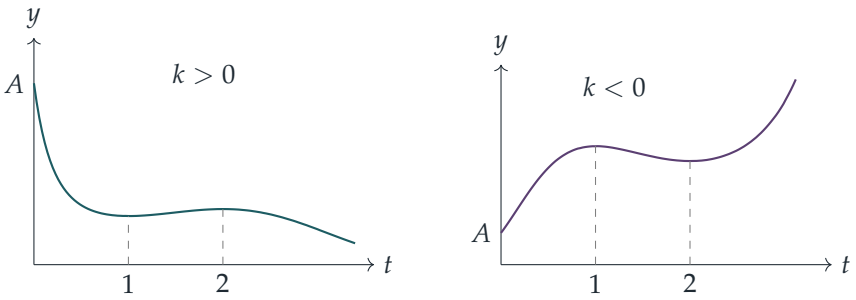
- 若 $k > 0$，则 $\ln(y/A) \to -\infty$，所以 $y \to 0$（速度极快，有如 $\mathrm{e}^{-kt^2/2}$）；
- 若 $k < 0$，则 $\ln(y/A) \to +\infty$，所以 $y \to \infty$（有如 $\mathrm{e}^{|k|t^2/2}$）。

草图

当 $t > 0$ 时，$\frac{\mathrm{d}y}{\mathrm{d}t}$ 的正负号与 $-k(t-1)(t-2)$ 的正负号相同。

- $k > 0$：y 在 $(0,1)$ 上由 A 递减，在 $t = 1$ 处达到局部极小值（local minimum），然后上升，在 $t = 2$ 处达到局部极大值（local maximum）（仍然低于 A），其后递减并趋向 0。
- $k < 0$：y 由 A 递增，在 $t = 1$ 处达到局部极大值，然后下降，在 $t = 2$ 处达到局部极小值（仍然高于 A），其后无止境地递增。

以下图像是取 $A = 1$ 及 $k = \pm 2$ 准确绘制的。



验算

- 在 $t = 0$ 处，公式给出 $A \cdot 1 \cdot \mathrm{e}^0 = A$。对 $\ln y = \ln A - 6k\ln(t+1) - k(t^2/2 - 4t)$ 求导，得 $\frac{y'}{y} = -\frac{6k}{t+1} - k(t-4)$，合并后便回到原来的方程。
- 数值方面，取 $A = 1.5$ 及 $k = \pm 2$，以龙格—库塔法（Runge–Kutta）逐步对微分方程积分，在 $t = 3$ 处所得结果与公式吻合至八位有效数字。取 $k = 2$ 时，比值 $y(1)/y(2)$ 为 0.8742，与 $(1.5)^{12}\mathrm{e}^{-5}$ 相符。

常见错误

- 除法出错：余数是 6，不是 2。可乘回 $t+1$ 来检查。
- 把比值倒转：题目要求的是 $y(1)/y(2)$，而指数中的正负号很容易令人误得 $y(2)/y(1)$。
- 草图中把两个转向点（turning point）画在 A 的错误一侧：先前的计算表明，当 $k > 0$ 时两者都低于 A，当 $k < 0$ 时两者都高于 A。

STEP 之外

这个方程是线性的，形如 $y' + p(t)y = 0$，而它的解 $A\exp\left(-\int_0^t p\right)$ 正是积分因子法（integrating-factor method）最简单的情形。它也说明了你在第一门常微分方程（ODE）课程中会遇到的一个普遍原则：很多时候无须解出方程，便可直接从方程本身看出解的定性特征（驻点、解的正负号、增长或衰减）。Boyce 与 DiPrima 合著的 *Elementary Differential Equations*（常微分方程入门教科书）对这两个想法都有详细阐述，当中包括我们用来证明 y 永不为零的唯一性定理（uniqueness theorem）。

详细解答

STEP II 2003，第 9 题

力学 · 极限摩擦力平衡

STEP II 2003，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/03-s2-q9

关键思路

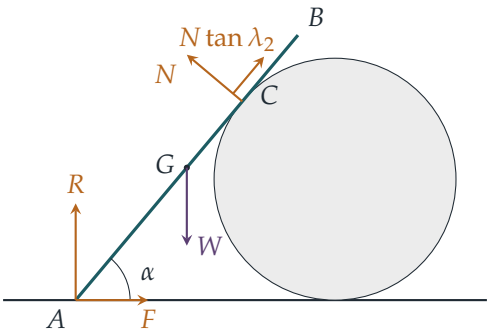
由于摩擦系数（coefficient of friction）是以 $\tan\lambda_1$ 和 $\tan\lambda_2$ 的形式给出，最自然的工具就是*总反作用力*（total reaction，即把法向反作用力（normal reaction）与摩擦力合成为一个力）。当摩擦达到极限（limiting）时，总反作用力与法线的夹角就是摩擦角（angle of friction）λ。这样一来，杆只受三个力作用，而且三个力的方向都已知；沿水平和垂直方向分解（resolve），便得到两条方程，其中只有两个未知的大小。真正要花工夫的，只是判断每个摩擦力指向哪一边。

解答

模型与摩擦力的方向

我们把杆视为均匀（uniform）的刚体（rigid body）杆，并把圆柱视为固定不动。杆沿切线（tangent）与圆柱接触，所以在 C 处的法向反作用力垂直于杆。设圆柱位于 A 的水平正方向一侧。于是作用在杆上的力有：

- 杆的重量 W，作用于中点 G，垂直向下；
- 在 A 处：垂直向上的法向反作用力 R，以及沿地面的摩擦力 F；
- 在 C 处：垂直于杆的法向反作用力 N，指向上方并向后偏往 A 的一边，另有一个沿杆方向的摩擦力。



想一想 在 A 处和在 C 处，摩擦力分别指向哪一边？杆会不会是即将向相反方向滑动？

杆要同时保持两个接触点，就只有一种运动方式：当 A 沿地面滑动时，角 α 随之改变，杆亦在圆柱上滑过。若 A 远离圆柱，杆便会下降（即 α 减小），而杆上位于 C 的那一点相对于圆柱沿杆向下滑动。若 A 移向圆柱，一切就反过来。

在第二种情况下，A 处的摩擦力会指向远离圆柱的方向，而 C 处的摩擦力会沿杆向下。这样三个水平方向的力（A 处的摩擦力、N 的水平分量，以及 C 处摩擦力的水平分量）全都指向远离圆柱的一边，水平方向便不可能平衡。所以在极限平衡（limiting equilibrium）时，杆是即将向下滑：杆在 A 处所受的摩擦力指向圆柱，而 C 处的摩擦力沿杆向上，正如图中所示。

总反作用力

设 S 为 A 处的总反作用力，T 为 C 处的总反作用力。由于两个接触点的摩擦都达到极限：

- S 与向上的垂直线成 λ_1 角，并向圆柱一方倾斜，因为 $F = R\tan\lambda_1$。
- C 处的法线与向上的垂直线成 α 角，并向后偏往 A 的一边，因为杆与水平线成 α 角。沿杆向上的摩擦力使 T 朝垂直线转回 λ_2。所以 T 与向上的垂直线成 $\alpha - \lambda_2$ 角，这个角是在远离圆柱的一侧量度的。

对杆分解各力，

$$\begin{aligned} \text{水平方向：} \quad & S\sin\lambda_1 = T\sin(\alpha - \lambda_2), \\ \text{垂直方向：} \quad & S\cos\lambda_1 + T\cos(\alpha - \lambda_2) = W. \end{aligned}$$

为了消去 T，我们把第一条方程乘以 $\cos(\alpha - \lambda_2)$，把第二条乘以 $\sin(\alpha - \lambda_2)$，再从第二条减去第一条：

$$\begin{aligned} S\left[\cos\lambda_1\sin(\alpha - \lambda_2) + \sin\lambda_1\cos(\alpha - \lambda_2)\right] &= W\sin(\alpha - \lambda_2), \\ S\sin(\alpha + \lambda_1 - \lambda_2) &= W\sin(\alpha - \lambda_2), \end{aligned}$$

这里用了 sin 的复角公式（addition formula）。题目告诉我们 $\alpha \neq \lambda_2 - \lambda_1$，所以 $\sin(\alpha + \lambda_1 - \lambda_2) \neq 0$（所有角都是锐角，所以 $\alpha + \lambda_1 - \lambda_2$ 严格介乎 $-\frac{\pi}{2}$ 与 π 之间）。因此可以相除：

$$S = \frac{W \sin(\alpha - \lambda_2)}{\sin(\alpha + \lambda_1 - \lambda_2)}.$$

A 处的摩擦力就是 S 的水平分量，即 $F = S \sin \lambda_1$，所以

$$F = \frac{W \sin \lambda_1 \, \sin(\alpha - \lambda_2)}{\sin(\alpha + \lambda_1 - \lambda_2)},$$

这正是所求。同样地，$T = W \sin \lambda_1 / \sin(\alpha + \lambda_1 - \lambda_2)$。

想一想 我们从头到尾都没有取力矩（moment）。这不是很可疑吗?

在这里并不可疑。沿两个方向分解，已足以求出两个未知的大小。对 A 取力矩（S 对该点没有力矩，C 处的摩擦力沿杆作用，同样没有力矩），得 $N \cdot AC = W \cdot AG \cos \alpha$，其中 $N = T \cos \lambda_2$。这只告诉我们，要达到极限平衡，C 必须位于哪里：

$$\frac{AC}{AB} = \frac{\cos \alpha \, \sin(\alpha + \lambda_1 - \lambda_2)}{2 \sin \lambda_1 \cos \lambda_2}.$$

求摩擦力时，并不需要用到题目所给的条件 $AC > CB$。从物理上说，我们还需要 $S > 0$ 和 $T > 0$，而这恰好在 $\alpha > \lambda_2$ 时成立：杆若以小于 λ_2 的倾角搁在圆柱上，就不可能处于即将沿圆柱滑下的状态。

验算

- 地面光滑，$\lambda_1 = 0$：公式给出 $F = 0$，理应如此。
- $\alpha = \lambda_2$：此时 T 是垂直的，单靠它就把杆托住，而 $F = 0$。这是杆刚好能搁在圆柱上的临界角度。
- 数值验算：取 $\lambda_1 = 0.3$、$\lambda_2 = 0.5$、$\alpha = 1.0$（弧度），直接从分量方程解出 R 和 N，得 $F \approx 0.1975W$，与公式吻合。在这个例子中，再对 A 取力矩，可知 C 大约位于杆长的 0.75 处，与 $AC > CB$ 相符。

常见错误

- 把 C 处的反作用力当成垂直的，或当成垂直于圆柱的“顶部”。反作用力应垂直于公切线，而公切线就是杆本身。
- 选定摩擦力的方向却不加检验。选错了那一对方向，水平方向便不可能平衡。
- 把 C 处的总反作用力向错误的一边倾斜，结果得出 $\alpha + \lambda_2$ 而不是 $\alpha - \lambda_2$。

STEP 之外

取总反作用力与法线成 λ 角，正是摩擦锥（cone of friction）背后的想法：只要总反作用力留在以法线为轴、半顶角为 λ 的圆锥之内，接触点就不会滑动。若只有三个力，它们的作用线（line of action）还必须交于一点，这就把许多静力学（statics）问题变成几何问题。有多个粗糙接触点、而且各接触点滑动的先后次序有影响的问题，是刚体静力学的上佳入门。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对摩擦和刚体平衡有仔细的论述。

详细解答

STEP II 2003，第 10 题

力学 · 涉及摩擦力的约束运动

STEP II 2003，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/03-s2-q10

关键思路

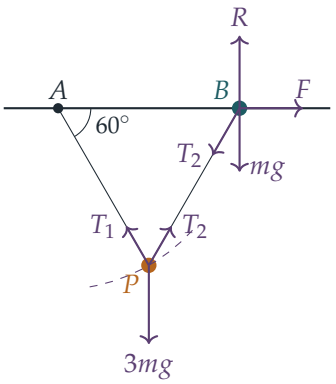
开始时，$AP = PB = AB = \ell$，所以 APB 是等边三角形（equilateral triangle）。由于 AP 一直拉紧，P 沿着以 A 为圆心的圆运动。整个系统由静止开始，所以没有向心（centripetal）项，P 的初始加速度（acceleration）必定沿着该圆的切线（tangent）方向，这就确定了它的方向。其次，P 始终位于以 AB 为底的等腰三角形（isosceles triangle）的顶点，所以 P 的水平位置永远是 B 的一半。至于最后一部分，分别对 B 和 P 写出牛顿第二定律（Newton's second law）。对 P 沿其加速度方向（垂直于 AP）分解，AP 中的张力（tension）便完全不会出现。

解答

我们把 B 和 P 视为质点（particle），并把绳视为轻而不可伸长（light and inextensible）。绳的两半都保持拉紧。取 A 为原点，金属线沿 x 轴，B 最初位于 $(\ell, 0)$。设 θ 为角 PAB。由于 $AP = PB = \ell$，

$$B = (2\ell\cos\theta,\, 0), \qquad P = (\ell\cos\theta,\, -\ell\sin\theta),$$

而开始时 $\theta = \frac{\pi}{3}$。



P 的加速度方向

由于 $AP = \ell$ 固定不变，P 沿着以 A 为圆心、半径为 ℓ 的圆运动（图中以虚线表示）。圆上一点的加速度有一个指向圆心的分量 $\ell\dot{\theta}^2$，以及一个沿切线的分量 $\ell\ddot{\theta}$。开始时 $\dot{\theta} = 0$，因为系统是由静止释放的。所以 P 的初始加速度沿着切线方向，即垂直于 AP。

AP 与水平线成 $60°$ 角，所以 AP 的垂直线与水平线成 $30°$ 角。若加速度的大小为 a，则

$$a_1 = a\cos 30° = \tfrac{\sqrt{3}}{2}a, \qquad a_2 = a\sin 30° = \tfrac{1}{2}a,$$

因此 $a_1 = \sqrt{3}\,a_2$，正如所求。

以 θ 表示：对 P 的位置求导，并令 $\dot{\theta} = 0$，可得初始加速度为 $\ell\ddot{\theta}(-\sin\theta, -\cos\theta)$。当 $\ddot{\theta} > 0$ 时，P 向下并朝 A 的方向运动。其大小为 $a = \ell\ddot{\theta}$。

B 的加速度

B 的 x 坐标永远恰好是 P 的两倍，因为 P 位于 AB 的垂直平分线（perpendicular bisector）上。求导两次可知，B 的加速度永远是 P 的水平加速度的两倍。所以 B 的初始加速度大小为 $2a_1$，方向指向 A，正如所求。

想一想 为甚么可以说 B 朝 A 移动？摩擦力又朝哪个方向作用？

P 下落时，θ 增大，所以 $x_B = 2\ell\cos\theta$ 减小。因此 B 朝 A 滑动，而作用于 B 的摩擦力（friction）则背离 A（沿正 x 方向）。这里我们假设 B 确实会移动。题目给出摩擦力为 $\frac{\sqrt{3}}{6}R$，这正是滑动时的数值，从而确认了这一点。

初始加速度的大小

设 T_1 和 T_2 分别为 AP 和 PB 中的张力。设 R 为作用于 B 的法向反作用力（normal reaction），$F = \frac{\sqrt{3}}{6}R$ 为摩擦力。以 a 表示 P 的初始加速度的大小，故 $a_1 = \frac{\sqrt{3}}{2}a$，而 B 的

加速度为 $2a_1 = \sqrt{3}\,a$，方向指向 A。

珠子 B。绳 BP 以水平线以下 60° 的方向把 B 拉向 A 的一方。B 没有垂直方向的加速度，所以沿垂直方向分解可得

$$R = mg + \tfrac{\sqrt{3}}{2}T_2.$$

沿水平方向分解，取指向 A 的方向为正：

$$m(\sqrt{3}\,a) = \tfrac{1}{2}T_2 - \tfrac{\sqrt{3}}{6}\left(mg + \tfrac{\sqrt{3}}{2}T_2\right) = \tfrac{1}{4}T_2 - \tfrac{\sqrt{3}}{6}mg,$$

所以

$$T_2 = 4\sqrt{3}\,ma + \tfrac{2\sqrt{3}}{3}mg. \tag{1}$$

质点 P。沿 P 的加速度方向分解，即向下并朝向 A、在水平线以下 30° 的方向。这个方向垂直于 AP，所以 T_1 没有任何贡献。重量 $3mg$ 与这个方向成 60° 角，所以它的贡献是 $3mg\cos 60^\circ = \tfrac{3}{2}mg$。张力 T_2 沿 PB 作用，即向上并背离 A、在水平线以上 60° 的方向。它与加速度方向成 150° 角，所以它的贡献是 $T_2\cos 150^\circ = -\tfrac{\sqrt{3}}{2}T_2$。因此

$$3ma = \tfrac{3}{2}mg - \tfrac{\sqrt{3}}{2}T_2. \tag{2}$$

把 (1) 代入 (2)：

$$3ma = \tfrac{3}{2}mg - \tfrac{\sqrt{3}}{2}\left(4\sqrt{3}\,ma + \tfrac{2\sqrt{3}}{3}mg\right) = \tfrac{3}{2}mg - 6ma - mg,$$

$$9ma = \tfrac{1}{2}mg,$$

所以 P 的初始加速度的大小为

$$a = \frac{g}{18},$$

正如所求。

这个答案前后一致：由 (1) 得 $T_2 = \tfrac{8\sqrt{3}}{9}mg > 0$，所以绳 PB 是拉紧的。于是 $R = \tfrac{7}{3}mg > 0$。对 P 沿水平方向分解，得 $T_1 = T_2 + 6ma_1 = \tfrac{19\sqrt{3}}{18}mg > 0$，所以 AP 也是拉紧的。

验算

- P 的垂直方向方程，我们之前没有用到：$\tfrac{\sqrt{3}}{2}(T_1 + T_2) - 3mg = \tfrac{35}{12}mg - 3mg = -\tfrac{1}{12}mg$。这等于 $-3ma_2 = -3m\cdot\tfrac{g}{36}$，与预期相符。
- 数值上，我们对 B 和 P 的位置求导，得到 $a_1/a_2 \approx 1.7320508$ 及 $a_B/a_1 = 2$。然后取 $g = 9.81$，从三条力的方程解出 T_1、T_2 和 $\ddot{\theta}$。结果是 $|a| = 0.545 = g/18$，并有 $T_2/mg \approx 1.5396$ 及 $R/mg \approx 2.3333$。
- 若没有摩擦力，(1) 会变成 $T_2 = 2\sqrt{3}\,ma$，而 (2) 会给出 $6ma = \tfrac{3}{2}mg$，所以 $a = g/4$。摩擦力理应令运动减慢，而的确 $g/18 < g/4$。

常见错误

- 加入了向心项。在释放的一刻 $\dot{\theta} = 0$，所以加速度纯粹沿切线方向。
- 忘记了 T_2 会把 B 向下拉，令 R 增大，摩擦力亦随之增大。取 $R = mg$ 会得出错误答案。
- 对 P 沿水平和垂直方向分解。这样做是可行的，但会引入 T_1，并需要多一条方程。沿垂直于 AP 的方向分解便可把它消去。

STEP 之外

这是一个只有一个自由度（degree of freedom）的约束系统（constrained system）：一切都由单一个角 θ 决定。把位置以 θ 表示然后求导，正是拉格朗日力学（Lagrangian mechanics）的第一步。在拉格朗日力学中，像 T_1 这类约束力（constraint force）是以有系统的方法消去的，而不是靠巧妙地选择方向。摩擦力并不能妥贴地纳入这个框架，这也是这类题目适合用来练习牛顿力学方法的原因之一。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门）对约束运动和非匀速圆周运动（non-uniform circular motion）有细致的讲解。

详细解答

STEP II 2003，第 11 题

力学 · 抛体运动

STEP II 2003，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：

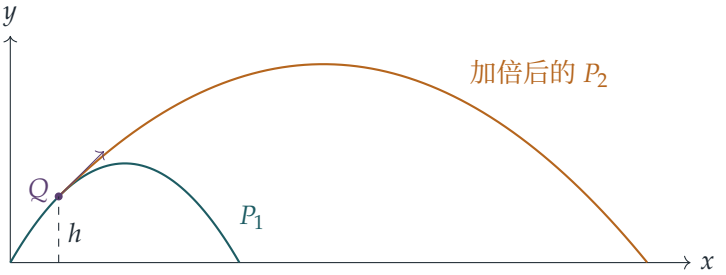
step.maths.org/questions/03-s2-q11

关键思路

把速度的水平分量和垂直分量分开处理。运动方向与水平线成 45°，正正就是两个分量相等的时候，由此可直接求出第一个时刻。速率加倍时方向不变，所以两个分量都加倍，而余下的飞行过程就是一个从已知高度出发的新抛体问题。

解答

我们把每个质点（particle）都当作只受重力作用的点状抛体（projectile），不计空气阻力。记 $s = \sin\alpha$ 及 $c = \cos\alpha$。由于 $\alpha > 45°$，我们有 $s > c > 0$。



此草图取 $\alpha = 60°$。两个质点都沿青绿色的路径运动至 Q，运动方向在该处首次与水平线成 45°。其后 P_2 沿琥珀色的路径运动。

P_1 的飞行时间

在时间 t 的垂直位移（displacement）是 $y = Vs\,t - \frac{1}{2}gt^2 = t\left(Vs - \frac{1}{2}gt\right)$。它再次为零的时刻是

$$T_1 = \frac{2V\sin\alpha}{g}.$$

首个 45° 时刻

在时间 t，速度的分量（component）是

$$\dot{x} = Vc, \qquad \dot{y} = Vs - gt.$$

水平分量始终不变。发射时运动方向比 45° 更陡，因为 $\dot{y} = Vs > Vc = \dot{x}$。随着 $\dot{y}$ 减小，当 $\dot{y} = \dot{x}$ 时，角度首次降至 45°，此时质点仍在上升：

$$Vs - gt_1 = Vc \quad\Longrightarrow\quad t_1 = \frac{V(s-c)}{g}.$$

（下降途中当 $\dot{y} = -Vc$ 时，运动方向会再次成 45°，但那是较后的事。）由于 $\frac{1}{2}T_1 = Vs/g$，

$$t_1 = \frac{Vs}{g}\left(1 - \frac{c}{s}\right) = \tfrac{1}{2}(1 - \cot\alpha)\,T_1,$$

正如所求。

P_2 的飞行

在时间 t_1 之前，P_2 的运动与 P_1 完全相同。在 t_1 时，速度的两个分量都等于 Vc，而高度是

$$\begin{aligned} h = Vs\,t_1 - \tfrac{1}{2}gt_1^2 &= \frac{V^2}{g}\left[s(s-c) - \tfrac{1}{2}(s-c)^2\right] \\ &= \frac{V^2}{2g}(s-c)\left[2s - (s-c)\right] = \frac{V^2}{2g}\,(s^2 - c^2). \end{aligned}$$

由于 $s > c$，这是正数。

> **想一想** “速率加倍”对速度的每个分量有甚么影响?

速率改变了，但方向不变，所以整个速度向量（velocity vector）乘以 2。两个分量都变成 $2Vc$：质点从高度 h 出发，方向在水平线之上 45°，速率为 $2\sqrt{2}\,Vc$。

设 τ 为由速率加倍至 P_2 着地所经过的时间。以 Q 为起点、向上为正量度垂直位移，着地即位移为 $-h$：

$$2Vc\,\tau - \tfrac{1}{2}g\tau^2 = -\frac{V^2}{2g}(s^2 - c^2).$$

两边乘以 $-2/g$，并记 $\tau = (V/g)u$，其中 u 是无量纲（dimensionless）的量：

$$u^2 - 4cu - (s^2 - c^2) = 0 \quad \Longrightarrow \quad u = 2c \pm \sqrt{4c^2 + s^2 - c^2} = 2c \pm \sqrt{s^2 + 3c^2}.$$

两根之积是 $-(s^2 - c^2) < 0$，所以其中一个根是负数。我们需要 $\tau > 0$，故取正号：

$$\tau = \frac{V}{g}\left(2c + \sqrt{s^2 + 3c^2}\right).$$

因此

$$T_2 = t_1 + \tau = \frac{V}{g}\left(s - c + 2c + \sqrt{s^2 + 3c^2}\right) = \frac{V}{g}\left(s + c + \sqrt{s^2 + 3c^2}\right).$$

除以 $\frac{1}{2}T_1 = Vs/g$，并利用 $\sqrt{s^2 + 3c^2}/s = \sqrt{1 + 3c^2/s^2}$（因 $s > 0$ 而成立）：

$$\frac{2T_2}{T_1} = \frac{s + c + \sqrt{s^2 + 3c^2}}{s} = 1 + \cot\alpha + \sqrt{1 + 3\cot^2\alpha},$$

这就是所要求的结果。

验算

- 当 $\alpha \to 90^\circ$ 时，$\cot\alpha \to 0$，比值趋向 2。这合乎常理：45° 时刻此时就是飞行的最高点，该处速率为零，所以加倍毫无作用，$T_2 = T_1$。
- 当 $\alpha = 45^\circ$（边界情况）时，加倍在发射一刻发生，所以 P_2 其实就是以 $2V$ 抛出，而 $T_2 = 2T_1$。公式给出 $1 + 1 + 2 = 4$，两者吻合。
- 当 $\alpha = 60^\circ$ 时，$t_1/T_1 = \frac{1}{2}(1 - 1/\sqrt{3}) \approx 0.2113$，而 $2T_2/T_1 = 1 + 1/\sqrt{3} + \sqrt{2} \approx 2.9916$。对 P_2 的飞行以数值方法逐步模拟，随机选取 2000 组 α 和 V，结果均与 T_2 的公式相符。

常见错误

- 只把垂直分量加倍，或只把水平分量加倍。方向不变，所以两个分量都加倍。
- 让 P_2 从地面重新开始运动，而不是从高度 h 开始。
- 没有检查正负号就取了 τ 的负根。

STEP 之外

速度的瞬时改变是一个*冲量*（impulse）：一个极大的力作用了极短的时间。大学一年级的力学正是用冲量来处理碰撞、火箭点火和踢球等情况。把一条轨迹分成若干段，各段

之间以冲量相接，正是规划轨道机动（orbital manoeuvre）的方法：引擎短暂点火，就能把太空船从一条圆锥曲线（conic）轨道转移到另一条。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）对冲量和动量（momentum）有严谨的阐述，并收录了许多这类题目。

详细解答

STEP II 2003，第 12 题

概率与统计 · 分支过程

STEP II 2003，第 12 题

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

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

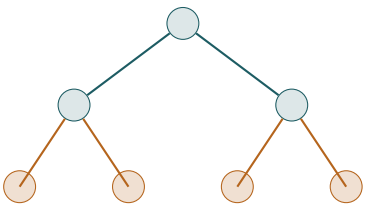
本题在 *Beyond STEP* 第 12.15 节灭绝还是爆发中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/branching.html

关键思路

这是一个分支过程（branching process）：每一代都由上一代产生，而每颗粒子各自独立地分裂。整道题只靠一个事实：若 m 颗粒子各自独立地产生后代，而每颗的后代数目服从 $\mathrm{B}(2, p)$，则后代总数服从 $\mathrm{B}(2m, p)$。因此，无论是求概率还是求期望值（expectation），我们都以上一代的大小作为条件。

解答

我们按照模型的原意，假设不同粒子的衰变彼此独立，亦与过去的情况独立。记 $X_0 = 1$，代表在 $0 < t < T$ 期间存活的那一颗粒子。在 $t = nT$ 时，存活的 X_{n-1} 颗粒子各自被自己的后代取代，所以 X_n 就是第 $n-1$ 代的后代总数。



$X_0 = 1$

$X_1 = 2$

$X_2 \leq 4$

图中所示为 $X_1 = 2$ 的情况：这两颗粒子合共最多有四个后代，而四个可能的后代各自以

概率 p 独立地诞生。

联合概率

由于 $X_1 \sim \mathrm{B}(2,p)$，我们有 $\mathrm{P}(X_1=2)=p^2$。已知 $X_1=2$ 时，两颗粒子各自独立地产生后代，每颗的后代数目服从 $\mathrm{B}(2,p)$，所以 X_2 就是 4 次独立试验中成功的次数，而每次成功的概率为 p；即 $X_2 \mid X_1=2 \sim \mathrm{B}(4,p)$。因此

$$\mathrm{P}(X_1=2 \text{及} X_2=2)=\mathrm{P}(X_1=2)\,\mathrm{P}(X_2=2 \mid X_1=2)=p^2\cdot\binom{4}{2}p^2q^2=6p^4q^2,$$

正如题目所求。

条件概率

要运用条件概率（conditional probability）的定义，我们需要 $\mathrm{P}(X_2=2)$，这可以通过以 X_1 为条件求得：

- 若 $X_1=0$，整个群体已经灭绝，所以 $X_2=0$；
- 若 $X_1=1$（概率为 $2pq$），则 $X_2\sim\mathrm{B}(2,p)$，而 $\mathrm{P}(X_2=2 \mid X_1=1)=p^2$；
- 若 $X_1=2$，其贡献就是刚才求得的 $6p^4q^2$。

所以

$$\mathrm{P}(X_2=2)=2pq\cdot p^2+6p^4q^2=2p^3q\,(1+3pq),$$

因此，当 $0<p<1$ 时，

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

令上式等于 $\frac{9}{25}$，得 $75pq=9+27pq$，所以 $48pq=9$，即 $pq=\frac{3}{16}$。代入 $q=1-p$，这就是

$$p^2-p+\tfrac{3}{16}=0 \quad\Longleftrightarrow\quad (p-\tfrac14)(p-\tfrac34)=0,$$

所以 $p=\frac14$ 或 $p=\frac34$。（$p=0$ 和 $p=1$ 这两种情况须排除：当 $p=0$ 时，事件 $X_2=2$ 不可能发生；当 $p=1$ 时，条件概率为 0。）

想一想 为甚么 p 的两个值会给出相同的答案?

条件概率只通过 $pq=p(1-p)$ 依赖于 p，而把 p 换成 $1-p$ 时，这个量保持不变。所以方程的根必定成对出现，即 p 与 $1-p$，在这里就是 $\frac14$ 和 $\frac34$。

期望值

设 $X_{n-1}=m$。这 m 颗粒子每颗平均产生 $\mathrm{E}(X)=2p$ 个后代，所以下一代大小的期望值为 $2pm$：

$$\mathrm{E}(X_n \mid X_{n-1}=m)=2pm.$$

对 m 的所有可能取值求平均（全期望值定律，law of total expectation），

$$\mathrm{E}(X_n)=\sum_m \mathrm{P}(X_{n-1}=m)\,2pm=2p\,\mathrm{E}(X_{n-1}).$$

用文字来说：在某一段时间内存活的每颗粒子，平均为下一段时间贡献 $2p$ 颗粒子。由于 $\mathrm{E}(X_1)=2p$（或 $\mathrm{E}(X_0)=1$），这个递推关系（recurrence）给出一个等比数列（geometric sequence）：

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

极限

当 $p=\frac{1}{4}$ 时，我们得到 $\mathrm{E}(X_n)=\left(\frac{1}{2}\right)^n\to 0$（当 $n\to\infty$），因为 $0<\frac{1}{2}<1$。

当 $p=\frac{3}{4}$ 时，我们得到 $\mathrm{E}(X_n)=\left(\frac{3}{2}\right)^n\to+\infty$，因为 $\frac{3}{2}>1$。

由此可见，p 的这两个值在条件概率中看似毫无分别，长远表现却截然不同：平均而言，当 $p=\frac{1}{4}$ 时群体逐渐消亡，而当 $p=\frac{3}{4}$ 时群体则急剧膨胀。

验算

- 取 $p=\frac{3}{4}$：$6p^4q^2=\frac{243}{2048}\approx 0.1187$，而模拟 10^6 个两代群体所得的结果是 0.1186。
- 当 $p=\frac{1}{4}$ 时，$pq=\frac{3}{16}$，所以 $3pq=\frac{9}{16}$，而 $\dfrac{9/16}{25/16}=\frac{9}{25}$，与预期相符。
- 对若干个 p 值计算 $X_1,\ldots,X_5$ 的精确分布（已知上一代时，每一代服从 $\mathrm{B}(2m,p)$），所得的平均值恰好是 $(2p)^n$。

常见错误

- 在求 $\mathrm{P}(X_2=2)$ 时，遗漏了 $X_1=1$、$X_2=2$ 这条路径；至于经过 $X_1=0$ 的路径，其贡献确实为零。
- 把 $X_2 \mid X_1=2$ 当作 $2\times\mathrm{B}(2,p)$（即一颗粒子后代数目的两倍），而不是两个独立副本之和，后者才是 $\mathrm{B}(4,p)$。

STEP 之外

这是一个高尔顿–沃森（Galton–Watson）分支过程，最初是为了解释姓氏为何会消失而提出的。后代数目的平均值 $\mu = 2p$ 决定了一切：若 $\mu \leq 1$，群体以概率 1 灭绝；若 $\mu > 1$，群体则有正概率永远存活下去。即使 $p = \frac{3}{4}$，群体仍然有 $\frac{1}{9}$ 的概率灭绝，这个数值是 $s = (q + ps)^2$ 的最小根，而这条方程来自概率母函数（probability generating function）。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率论入门教材）以母函数处理分支过程，其讲法正好可以直接承接本题。

详细解答

STEP II 2003，第 13 题

概率与统计 · 截尾泊松分布

STEP II 2003，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/03-s2-q13

关键思路

这是一个剔除了 0 这个值的泊松分布（Poisson distribution），余下的概率按比例放大，使它们的总和仍然是 1。因此，我们需要的每一个和，都是缺了 $k = 0$ 一项的泊松和；而在计算平均值（mean）和方差（variance）时，$k = 0$ 一项本来就没有任何贡献。所以我们可以借用熟悉的泊松分布结果 $\sum k\,p_k = \lambda$ 和 $\sum k^2 p_k = \lambda + \lambda^2$，然后乘以 A 就可以了。

解答

常数 A

记 $p_k = \dfrac{\lambda^k \mathrm{e}^{-\lambda}}{k!}$ 为普通 $\mathrm{Po}(\lambda)$ 的概率。由于 $\sum_{k\ge 0} \lambda^k/k! = \mathrm{e}^{\lambda}$，

$$\sum_{k=1}^{\infty} p_k = \mathrm{e}^{-\lambda}\left(\mathrm{e}^{\lambda} - 1\right) = 1 - \mathrm{e}^{-\lambda}.$$

概率 $\mathrm{P}(X = k) = A p_k$ 的总和必须是 1，所以 $A(1 - \mathrm{e}^{-\lambda}) = 1$，即

$$A = (1 - \mathrm{e}^{-\lambda})^{-1},$$

这正是题目要求的结果。留意当 $\lambda > 0$ 时 $0 < 1 - \mathrm{e}^{-\lambda} < 1$，所以 $A > 1$：剩下来的概率 p_0 已经分摊给 $k \ge 1$ 的各个值。

平均值

$\sum k\,p_k$ 中 $k = 0$ 的一项是零，所以

$$\sum_{k=1}^{\infty} k\,p_k = \sum_{k=0}^{\infty} k\,p_k = \lambda,$$

即 $\mathrm{Po}(\lambda)$ 的平均值。（直接计算：$\sum_{k\geq 1} k\lambda^k \mathrm{e}^{-\lambda}/k! = \lambda \sum_{k\geq 1} \lambda^{k-1}\mathrm{e}^{-\lambda}/(k-1)! = \lambda$。）因此

$$\mu = \mathrm{E}(X) = A\lambda = \frac{\lambda}{1-\mathrm{e}^{-\lambda}}.$$

方差

同样地，$\sum k^2 p_k$ 中 $k=0$ 的一项是零，而对于 $\mathrm{Po}(\lambda)$，我们有 $\mathrm{E}(Y^2) = \mathrm{Var}(Y) + (\mathrm{E}\,Y)^2 = \lambda + \lambda^2$。如果你不想直接引用这个结果，可以写 $k^2 = k(k-1) + k$：

$$\sum_{k\geq 2} k(k-1)\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} = \lambda^2 \sum_{k\geq 2} \frac{\lambda^{k-2}\mathrm{e}^{-\lambda}}{(k-2)!} = \lambda^2, \qquad \text{所以} \qquad \sum_{k\geq 1} k^2 p_k = \lambda^2 + \lambda.$$

于是

$$\begin{aligned} \mathrm{E}(X^2) &= A\lambda(1+\lambda) = \mu(1+\lambda), \\ \mathrm{Var}(X) &= \mathrm{E}(X^2) - \mu^2 = \mu(1+\lambda) - \mu^2 = \mu(1-\mu+\lambda), \end{aligned}$$

这正是题目要求的结果。

不等式

> ? **想一想** 严格不等式 $\mu < 1+\lambda$ 可以从哪里来?

从方差而来。X 取 $1, 2, 3, \ldots$ 中每一个值的概率都是正数，所以它不是常数，而一个非常数的随机变量（random variable）的方差必定严格大于零（因为 $\mathrm{Var}(X) = \mathrm{E}\big((X-\mu)^2\big)$，而 $(X-\mu)^2 > 0$ 的概率是正数）。所以

$$\mu(1-\mu+\lambda) > 0.$$

由于 $\mu > 0$(它是一个取正值的随机变量的平均值),我们可以两边除以 μ,得到 $1-\mu+\lambda > 0$，即 $\mu < 1+\lambda$。

至于下界，$\mu = A\lambda$，其中 $A > 1$ 而 $\lambda > 0$，所以 $\mu > \lambda$。两者合起来，

$$\lambda < \mu < 1+\lambda.$$

$\lambda = 100$ 时的正态近似

e^{-100} 小得惊人（大约是 4×10^{-44}），所以在任何合理的准确度下

$$\mu = \frac{100}{1-\mathrm{e}^{-100}} \approx 100, \qquad \mathrm{Var}(X) = \mu(101-\mu) \approx 100\times 1 = 100.$$

所以我们用 $Y \sim \mathrm{N}(100, 10^2)$ 来近似 X，这就是正态近似（normal approximation）。由于 X 只取整数值，我们要使用连续性校正（continuity correction）：事件 $X = 100$ 对应

$99.5 < Y < 100.5$。以 $Z = (Y - 100)/10$ 标准化，

$$\begin{aligned} \mathrm{P}(X = 100) &\approx \mathrm{P}(-0.05 < Z < 0.05) \\ &= 2\Phi(0.05) - 1 \\ &\approx 2(0.51994) - 1 = 0.0399, \end{aligned}$$

所以 $\mathrm{P}(X = 100) \approx 0.04$（准确至小数点后 2 位）。

验算

- 当 $\lambda \to 0$ 时，$\mu = \lambda/(1 - \mathrm{e}^{-\lambda}) \to 1$，而方差 $\to 0$：几乎所有概率都集中在 $k = 1$，这是理所当然的。这也与不等式 $\lambda < \mu < 1 + \lambda$ 把 μ 夹向 1 的情况吻合。
- 对许多随机选取的 λ 以数值方法求级数的和，所得结果与 A、μ 和 $\mu(1 - \mu + \lambda)$ 一致，准确至机器精度。
- 当 $\lambda = 100$ 时，准确的概率是 $\dfrac{100^{100}\mathrm{e}^{-100}}{100!\,(1 - \mathrm{e}^{-100})} \approx 0.03986$，而正态近似给出 0.03988。两者四舍五入后都是 0.04。

常见错误

- 忘记连续性校正，算出 $\mathrm{P}(Y = 100) = 0$，或者未加解释就使用密度函数的值。
- 只证明了 $\mu \leq 1 + \lambda$：要得到严格不等式，必须指出 X 不是常数。
- 除以 μ 时没有说明 $\mu > 0$。

STEP 之外

X 服从*零截尾泊松*（zero-truncated Poisson）分布，统计学家会在无法观察到计数为零的情况下使用它（例如经过检查站的汽车的乘载人数：没有一辆车的乘载人数是零）。正态近似之所以行得通，是因为一个 $\mathrm{Po}(100)$ 变量是 100 个独立的 $\mathrm{Po}(1)$ 变量之和，所以中心极限定理（central limit theorem）适用；在这个尺度下，截尾的影响根本看不出来。Grimmett and Welsh, *Probability: An Introduction*（概率论入门教材）讲解了母函数（generating functions）和中心极限定理，它们正是这道题前后两部分背后的理论。

详细解答

STEP II 2003，第 14 题

概率与统计 · 假设检验的错误

STEP II 2003，第 14 题

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

step.maths.org/questions/03-s2-q14

关键思路

掷出第一个 6 所需的投掷次数服从几何分布（geometric distribution），而它的尾部概率形式极为简洁：事件“需要多于 M 次投掷”其实就是“首 M 次投掷全部失败”。因此两个错误概率都纯粹是 M 次幂，而要在两者之间消去 M，只需取对数（logarithm）即可。

解答

两个错误概率

把各次投掷视为互相独立（independent），每次掷出 6 的概率同为 p。设 X 为直至首次掷出 6 为止（包括该次）的投掷次数，则

$$X > M \iff \text{首 } M \text{ 次投掷中没有一次是 6}, \qquad \text{所以} \qquad \mathrm{P}(X > M) = (1-p)^M,$$

这里用到了各次投掷的独立性。（毋须求几何级数（geometric series）$\sum p(1-p)^{k-1}$ 的和：直接考虑事件本身更快捷。）

这个检验恰好在 $X > M$ 时接受 H_1。因此 α，即实际上 $p = B$ 时接受 H_1 的概率，是

$$\alpha = (1-B)^M,$$

而 β，即实际上 $p = A$ 时接受 H_0（亦即 $X \le M$）的概率，是

$$\beta = 1 - \mathrm{P}(X > M \mid p = A) = 1 - (1-A)^M.$$

消去 M

我们想把 $(1-A)^M$ 写成 $\alpha=(1-B)^M$ 的幂。由于 $0<B<1$，$\ln(1-B)$ 有定义且不等于零，所以对 α 取对数可得

$$M=\frac{\ln\alpha}{\ln(1-B)}.$$

于是

$$(1-A)^M=\mathrm{e}^{M\ln(1-A)}=\exp\left(\frac{\ln(1-A)}{\ln(1-B)}\ln\alpha\right)=\alpha^K,\qquad \text{其中}\quad K=\frac{\ln(1-A)}{\ln(1-B)}.$$

因此

$$\beta=1-\alpha^K,\qquad K=\frac{\ln(1-A)}{\ln(1-B)},$$

而 K 只取决于 A 和 B，与 M 无关，正如题目所求。

? **想一想** 关于 K 的大小，我们可以说些甚么？

由于 $0<A<B<1$，我们有 $1>1-A>1-B>0$，所以两个对数都是负数，且 $\ln(1-B)<\ln(1-A)<0$。两者相除，可知 K 是正数，并且

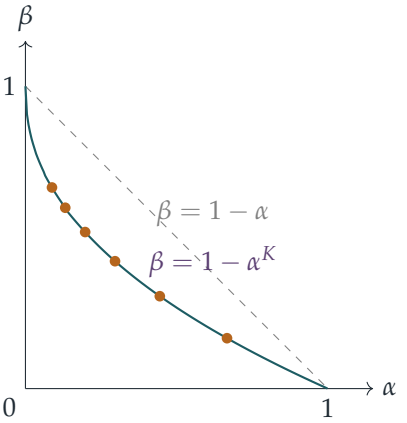
$$0<K<1.$$

图像的形状正是由此决定。

β 对 α 的图像

当 M 依次取 $1,2,3,\ldots$ 时，$\alpha=(1-B)^M$ 在 $(0,1)$ 内取值，并递减趋向 0；严格来说，图像是一列点，全部位于曲线 $\beta=1-\alpha^K$，$0\le\alpha\le1$ 之上。对于这条曲线：

- 它通过 $(0,1)$ 和 $(1,0)$；
- $\dfrac{\mathrm{d}\beta}{\mathrm{d}\alpha}=-K\alpha^{K-1}<0$，所以 β 随 α 增加而减少：令其中一种错误较难发生，另一种错误就会较易发生；
- 由于 $K-1<0$，当 $\alpha\to0^+$ 时斜率（gradient）趋向 $-\infty$（曲线从 $(0,1)$ 垂直出发），而在 $\alpha=1$ 处斜率等于 $-K$，介乎 -1 与 0 之间；
- $\dfrac{\mathrm{d}^2\beta}{\mathrm{d}\alpha^2}=K(1-K)\alpha^{K-2}>0$，所以曲线是凸的（convex）；
- 对于 $0<\alpha<1$ 和 $0<K<1$，我们有 $\alpha^K>\alpha$，所以 $\beta<1-\alpha$：曲线位于连接 $(0,1)$ 与 $(1,0)$ 的直线下方。



图中的曲线是按 $A = \frac{1}{6}$、$B = \frac{1}{3}$（所以 $K \approx 0.45$）绘画的，并标出了 $M = 1, \ldots, 6$ 所对应的点；M 增加时，点会向左上方移动。

验算

- 取 $A = \frac{1}{6}$、$B = \frac{1}{3}$、$M = 4$。直接计算，$\alpha = (\frac{2}{3})^4 \approx 0.1975$，$\beta = 1 - (\frac{5}{6})^4 \approx 0.5177$。以 $K = \ln \frac{5}{6} / \ln \frac{2}{3} \approx 0.4497$ 计算，$1 - \alpha^K \approx 0.5177$，结果相同。把这个检验模拟 2×10^5 次，所得结果为 0.198 和 0.518。
- 极端情况的表现亦合乎常理：若 A 接近 B，则 $K \approx 1$ 而 $\beta \approx 1 - \alpha$，这个检验比起不理会投掷结果、随机作决定好不了多少；若 A 极小，则 $K \approx 0$，对于任何合理的 α，都可以令 β 变得很小。

常见错误

- 混淆两个假设（hypothesis）：这里 H_0 是较大的值 B，所以等了很久才掷出 6，是不利于 H_0 的证据。
- 把 K 上下颠倒，写成 $\ln(1 - B) / \ln(1 - A)$，其值大于 1；检查 $\beta < 1 - \alpha$（任何合理的检验都胜过乱猜）便可发现这个错误。

STEP 之外

β 对 α 的曲线，就是两类错误之间的取舍曲线（trade-off curve），与统计学和机器学习中广泛使用的 ROC 曲线（ROC curve）关系密切。奈曼–皮尔逊引理（Neyman–Pearson lemma）表明，对于像这样的两个简单假设（simple hypotheses），基于似然比（likelihood

ratio）的检验能就每个 α 给出最小可能的 β；这里似然比对 X 是单调的（monotonic），所以“当 X 很大时拒绝 H_0”这个检验其实是最优的。想进一步了解这些概念，可先阅读 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）打好基础，然后再读任何一门数理统计（mathematical statistics）的入门课程。

详细解答

STEP III 2003，第 1 题

纯数学 · 反双曲函数积分法

STEP III 2003，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/03-s3-q1

关键思路

头两个结果是细心运用链式法则（chain rule）的计算。它们真正有用之处，是提供了一个范本：每个反函数（inverse function）的导数（derivative）都具有 $\dfrac{\text{常数}}{(\text{一次式})\sqrt{\text{一次式}}}$ 的形式，而这正正是两个被积函数（integrand）的形式。因此，对每个积分，我们选取 a 和 b 去配合这个范本。根号外的括号告诉我们该用哪一条公式。若它是 $x+b$，即常数较大的那个，便用 arcsin；若它是 $x+a$，即常数较小的那个，便用 arcosh。

解答

两个导数

设 $y=\dfrac{x+a}{x+b}$。由商法则（quotient rule），

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

由于 $x+a>0$ 及 $x+b>0$，可知 $y>0$；又由于 $b>a$，分子小于分母，所以 $0<y<1$，而 $\arcsin y$ 在该处可微（differentiable）。接着，

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

这里用了平方差（difference of two squares）公式。由于 $x+b>0$，$(x+b)^2$ 的平方根就是 $x+b$ 本身，所以

$$\sqrt{1-y^2}=\frac{\sqrt{b-a}\,\sqrt{a+b+2x}}{x+b}.$$

由链式法则，

$$\frac{\mathrm{d}}{\mathrm{d}x}\arcsin\left(\frac{x+a}{x+b}\right)=\frac{1}{\sqrt{1-y^2}}\cdot\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{x+b}{\sqrt{b-a}\sqrt{a+b+2x}}\cdot\frac{b-a}{(x+b)^2}$$
$$=\frac{\sqrt{b-a}}{(x+b)\sqrt{a+b+2x}},$$

因为 $(b-a)/\sqrt{b-a}=\sqrt{b-a}$。这就是所求的结果。

至于 arcosh，设 $z=\dfrac{x+b}{x+a}$。这里 $z>1$，因为 $x+b>x+a>0$。于是

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

由于 $x+a>0$，$\sqrt{z^2-1}=\sqrt{b-a}\sqrt{a+b+2x}\,/(x+a)$，因此

$$\frac{\mathrm{d}}{\mathrm{d}x}\operatorname{arcosh}\left(\frac{x+b}{x+a}\right)=\frac{x+a}{\sqrt{b-a}\sqrt{a+b+2x}}\cdot\left(-\frac{b-a}{(x+a)^2}\right)=-\frac{\sqrt{b-a}}{(x+a)\sqrt{a+b+2x}}.$$

两个结果看来几乎一样。分别只在于正负号，以及根号外是哪一个括号。

想一想 arcsin 的计算实际上用到了哪些假设?

我们需要 $x+b>0$ 来化简 $\sqrt{(x+b)^2}$，也需要 $1-y^2>0$。从因式分解后的形式可见，在 $b>a$ 的前提下，$1-y^2>0$ 恰好在 $a+b+2x>0$ 时成立。我们从没有用过 $x+a>0$ 本身。所以，只要 $x+b>0$ 及 $a+b+2x>0$，arcsin 的公式便成立，即使 $x+a$ 是负数亦然，此时 y 位于 $(-1,0]$ 之内。这一点在 (ii) 中至关重要。

第 (i) 部分

根号外的括号是 $x+1$，所以我们尝试把它与 arcosh 公式中的 $x+a$ 配对。写成 $\sqrt{x+3}=\sqrt{2x+6}\,/\sqrt{2}$，于是除了 $a=1$ 之外，还需要 $a+b=6$。由此得 $b=5$，故 $b-a=4$。取这些数值时，假设 $x+a>0$、$x+b>0$ 及 $b>a$ 在 $x>-1$ 时全部成立，而且

$$\frac{\mathrm{d}}{\mathrm{d}x}\operatorname{arcosh}\left(\frac{x+5}{x+1}\right)=-\frac{2}{(x+1)\sqrt{2x+6}}=-\frac{\sqrt{2}}{(x+1)\sqrt{x+3}}.$$

两边除以 $-\sqrt{2}$，

$$\int\frac{1}{(x+1)\sqrt{x+3}}\,\mathrm{d}x=-\frac{1}{\sqrt{2}}\operatorname{arcosh}\left(\frac{x+5}{x+1}\right)+C.$$

第 (ii) 部分

现在根号外的括号是 $x+3$，所以我们把它与 arcsin 公式中的 $x+b$ 配对：$b=3$。根号内需要 $a+b+2x=2(x+1)$，所以 $a+b=2$，$a=-1$。于是 $b-a=4$，并且

$$\frac{\mathrm{d}}{\mathrm{d}x}\arcsin\left(\frac{x-1}{x+3}\right)=\frac{2}{(x+3)\sqrt{2x+2}}=\frac{\sqrt{2}}{(x+3)\sqrt{x+1}}.$$

想一想 当 $a = -1$ 时，假设 $x + a > 0$ 要求 $x > 1$。我们的答案在整个范围 $x > -1$ 上仍然成立吗?

成立。正如上文所见, arcsin 的公式只需要 $x + b > 0$ 及 $a + b + 2x > 0$。在这里, 这两个条件就是 $x + 3 > 0$ 及 $2x + 2 > 0$, 对所有 $x > -1$ 都成立。在这个范围内, $(x-1)/(x+3)$ 由 -1 向上趋近 1，所以 arcsin 有定义。因此，对 $x > -1$,

$$\int \frac{1}{(x+3)\sqrt{x+1}}\,\mathrm{d}x = \frac{1}{\sqrt{2}}\arcsin\left(\frac{x-1}{x+3}\right) + C.$$

另一做法：代换

两个积分也可以用代换（substitution）$u = \sqrt{x+c}$ 解决，其中 $x + c$ 是根号下的式子，因为这时 $\mathrm{d}x = 2u\,\mathrm{d}u$。在 (ii) 中，取 $u = \sqrt{x+1}$，积分变成 $\int \frac{2}{u^2+2}\,\mathrm{d}u = \sqrt{2}\arctan\left(\sqrt{(x+1)/2}\right) + C$。这与我们的答案只相差常数 $-\pi/(2\sqrt{2})$。在 (i) 中，取 $u = \sqrt{x+3}$，得到 $\int \frac{2}{u^2-2}\,\mathrm{d}u = \frac{1}{\sqrt{2}}\ln\left|\frac{\sqrt{x+3}-\sqrt{2}}{\sqrt{x+3}+\sqrt{2}}\right| + C$。这正是以对数（logarithm）写出的 arcosh 答案。

验算

- 把答案求导还原。例如，在 $x = 0$ 处，$\frac{1}{\sqrt{2}}\arcsin\frac{x-1}{x+3}$ 的导数应为 $\frac{1}{3}$，而上面的公式给出 $\frac{\sqrt{2}}{\sqrt{2}\cdot 3\cdot 1} = \frac{1}{3}$。
- 数值上，用辛普森法则（Simpson's rule）可得 $\int_0^1 \frac{\mathrm{d}x}{(x+1)\sqrt{x+3}} \approx 0.374543$ 及 $\int_0^1 \frac{\mathrm{d}x}{(x+3)\sqrt{x+1}} \approx 0.240301$。两个数值与原函数（antiderivative）算出的结果吻合，误差小于 10^{-12}。在下限低至 $x = -0.9$ 的区间上，结果同样吻合，而这些区间已伸入 $-1 < x < 1$ 的范围，即 (ii) 中 $x + a < 0$ 之处。
- 正负号：(i) 的被积函数是正的，所以它的积分应该递增。我们的答案确是如此，因为 $\frac{x+5}{x+1}$ 随 x 增加而递减，所以它的 arcosh 的负值递增。

常见错误

- 把括号配错公式：在 (i) 中，根号外的 $x + 1$ 是常数较小的那个，所以它必定是 $x + a$，即要用 arcosh，而不是 arcsin。
- 忘记了把 $\sqrt{x+3}$ 写成 $\sqrt{2x+6}/\sqrt{2}$ 时所产生的因子 $\sqrt{2}$。
- 漏掉 arcosh 导数中的负号，这个负号来自 $\mathrm{d}z/\mathrm{d}x < 0$。

STEP 之外

(ii) 的两个答案，一个用 arcsin，一个用 arctan，两者相差一个常数。这是必然的，因为在一个区间上导数相同的两个函数只相差一个常数，这是中值定理（mean value theorem）的推论。若集合由两段分开的部分组成，各段上的常数便可以不同。(i) 中出现对数也并非偶然：反双曲函数（inverse hyperbolic functions）其实是乔装了的对数，这正是 $1/\sqrt{\text{二次式}}$ 的积分不是得出反三角函数（inverse trigonometric functions）便是得出对数的原因。Abbott 的 *Understanding Analysis* 严谨地证明了中值定理，而 Siklos 的 *Advanced Problems in Mathematics* 则收录了更多这类积分。

详细解答

STEP III 2003，第 2 题

纯数学 · 二项式级数

STEP III 2003，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/03-s3-q2

关键思路

中心二项式系数（central binomial coefficient）$\binom{2r}{r}$ 就藏在负半整数幂的二项级数（binomial series）之中。只要用开首的恒等式把 $(2r)!$ 拆成奇数因子和偶数因子，便会发现 $(1-p)^{-1/2}$ 中 p^r 的系数恰好是 $\binom{2r}{r}/4^r$。于是每个和其实都是乔装了的二项级数，余下的工作只是选取 p（以及在 (ii) 中选对幂次）。

解答

恒等式

把 $(2r)!$ 的因子分成奇数和偶数两组：

$$(2r)! = \big[1 \times 3 \times \cdots \times (2r-1)\big] \times \big[2 \times 4 \times \cdots \times 2r\big].$$

r 个偶数因子各自是 $1, 2, \ldots, r$ 其中一个数的 2 倍，所以第二个括号等于 $2^r\, r!$。因此

$$\binom{2r}{r} = \frac{(2r)!}{r!\,r!} = \frac{1 \times 3 \times \cdots \times (2r-1) \times 2^r\, r!}{r!\,r!} = \frac{1 \times 3 \times \cdots \times (2r-1)}{r!} \times 2^r,$$

这正是所要证明的结果。

第 (i) 部分

当 $|p|<1$ 时，由二项级数可得

$$\begin{aligned}(1-p)^{-\frac{1}{2}} &= 1+\left(-\tfrac{1}{2}\right)(-p)+\frac{\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)}{2!}(-p)^2\\ &\quad+\frac{\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)}{3!}(-p)^3+\cdots\\ &= 1+\tfrac{1}{2}p+\tfrac{3}{8}p^2+\tfrac{5}{16}p^3+\cdots.\end{aligned}$$

通项也依循同样的规律。在 p^r 的系数中，因子 $-\frac{1}{2},-\frac{3}{2},\ldots$ 带来 r 个负号，$(-p)^r$ 又带来 r 个，两者互相抵消，剩下

$$\frac{\frac{1}{2}\times\frac{3}{2}\times\cdots\times\frac{2r-1}{2}}{r!}\,p^r=\frac{1\times3\times\cdots\times(2r-1)}{2^r\,r!}\,p^r.$$

想一想 这与 $\binom{2r}{r}$ 有甚么关系?

由恒等式可知 $\dfrac{1\times3\times\cdots\times(2r-1)}{r!}=\dfrac{\binom{2r}{r}}{2^r}$。所以 p^r 的系数是 $\binom{2r}{r}/4^r$；当 $r=0$ 时这也成立，因为两边都等于 1。因此

$$(1-p)^{-\frac{1}{2}}=\sum_{r=0}^{\infty}\binom{2r}{r}\left(\frac{p}{4}\right)^r\qquad(|p|<1).$$

我们要求的和中 $(p/4)^r=1/8^r$，所以取 $p=\frac{1}{2}$，这个值在收敛区间（interval of convergence）之内。于是

$$\sum_{r=0}^{\infty}\frac{\binom{2r}{r}}{8^r}=\left(1-\tfrac{1}{2}\right)^{-\frac{1}{2}}=\left(\tfrac{1}{2}\right)^{-\frac{1}{2}}=\sqrt{2}.$$

第 (ii) 部分

多出来的因子 $2r+1$ 提示我们把幂次再推前一步，考虑 $(1-p)^{-3/2}$。当 $|p|<1$ 时，其二项级数中 p^r 的系数是（正负号如前一样互相抵消）

$$\begin{aligned}\frac{\frac{3}{2}\times\frac{5}{2}\times\cdots\times\frac{2r+1}{2}}{r!} &= \frac{3\times5\times\cdots\times(2r+1)}{2^r\,r!}\\ &= (2r+1)\,\frac{1\times3\times\cdots\times(2r-1)}{2^r\,r!}=\frac{(2r+1)\binom{2r}{r}}{4^r},\end{aligned}$$

这是因为 $3\times5\times\cdots\times(2r+1)=(2r+1)\times\left[1\times3\times\cdots\times(2r-1)\right]$，然后再像 (i) 那样运用恒等式。（当 $r=0$ 时，首尾两端都等于 1。）所以

$$(1-p)^{-\frac{3}{2}}=\sum_{r=0}^{\infty}(2r+1)\binom{2r}{r}\left(\frac{p}{4}\right)^r\qquad(|p|<1).$$

现在需要 $p/4=\frac{1}{5}$，即 $p=\frac{4}{5}$，同样满足 $|p|<1$。因此

$$\sum_{r=0}^{\infty}\frac{(2r+1)\binom{2r}{r}}{5^r}=\left(1-\tfrac{4}{5}\right)^{-\frac{3}{2}}=\left(\tfrac{1}{5}\right)^{-\frac{3}{2}}=5^{\frac{3}{2}}=(\sqrt{5})^3,$$

这正是所要证明的结果。

另一做法：当 $|x|<\frac{1}{4}$ 时，记 $f(x)=\sum\binom{2r}{r}x^r=(1-4x)^{-1/2}$。在收敛区间内对幂级数（power series）逐项求导，得 $\sum r\binom{2r}{r}x^r=xf'(x)=2x(1-4x)^{-3/2}$，所以

$$\sum(2r+1)\binom{2r}{r}x^r=4x(1-4x)^{-\frac{3}{2}}+(1-4x)^{-\frac{1}{2}}=(1-4x)^{-\frac{3}{2}},$$

代入 $x=\frac{1}{5}$ 便得到相同的答案。

验算

- (i) 的首几项是 $1+\frac{2}{8}+\frac{6}{64}+\frac{20}{512}\approx 1.371$，已经接近 $\sqrt{2}\approx 1.414$；把 200 项加起来得 $1.41421356\ldots$。
- (ii) 的级数收敛得较慢（相邻项之比趋向 $\frac{4}{5}$），但 400 项之和是 $11.18033988\ldots$，与 $5\sqrt{5}\approx 11.1803$ 吻合。
- 当 $r=0,1,2,3$ 时，系数 $\binom{2r}{r}/4^r$ 分别是 $1,\frac{1}{2},\frac{3}{8},\frac{5}{16}$，与直接求得的四项一致。

常见错误

- 弄错正负号：系数中的 r 个负因子与 $(-p)^r$ 所带来的 $(-1)^r$ 必定互相抵消，所以每个系数都是正数。
- 忘记检查所选的 p 是否满足 $|p|<1$；若不满足，级数未必收敛至该函数。
- 在 (ii) 中换一个 p 再用 $(1-p)^{-1/2}$：单凭这个级数，不可能得出因子 $2r+1$。

STEP 之外

函数 $\sum\binom{2r}{r}x^r=(1-4x)^{-1/2}$ 是中心二项式系数的生成函数(generating function),是组合数学(combinatorics)中的标准工具;例如把它平方便得 $\sum_r 4^r x^r$,从而证明 $\sum_k\binom{2k}{k}\binom{2r-2k}{r-k}=4^r$。逐项求导为何成立，以及幂级数确切在哪里收敛，都是大学一年级分析学（analysis）的课题。关于幂级数，可参阅 Abbott 的 *Understanding Analysis*；想做更多与级数有关的 STEP 练习，可参阅 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

STEP III 2003，第 3 题

纯数学 · 曲线描绘

> **STEP III 2003，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/03-s3-q3

关键思路

设 $u=(1+x)^m$ 及 $v=(1-x)^m$。于是 $\mathrm{f}=(u-v)/(u+v)$，而商法则（quotient rule）的结果会化简得非常漂亮，因为 $u'v-uv'$ 有公因式 $(1-x^2)^{m-1}$。这个导数（derivative）永不为负，单凭这一点，我们几乎已掌握 $y=\mathrm{f}(x)$ 的整个形状。至于倒数（reciprocal）曲线，则藏着一种对称（symmetry）：把 x 换成 $1/x$，f 就变成 $1/\mathrm{f}$。

解答

分母永不为零

假设对某个实数 x 有 $(1+x)^m+(1-x)^m=0$。

若 m 是偶数，两项皆为非负，故两项必须同时为零。这就需要同时有 $x=-1$ 和 $x=1$，并不可能。

若 m 是奇数，则 $(1+x)^m=-(1-x)^m=(x-1)^m$。当 m 是奇数时，映射 $t\mapsto t^m$ 在 $\mathbb{R}$ 上严格递增（strictly increasing），因此是一对一（one-to-one）的，于是 $1+x=x-1$，即 $1=-1$。这是矛盾（contradiction）。

所以对每个实数 x 都有 $(1+x)^m+(1-x)^m\neq 0$，而 f 在整个 $\mathbb{R}$ 上有定义并且可微（differentiable）。

导数

设 $u=(1+x)^m$ 及 $v=(1-x)^m$，则 $u'=m(1+x)^{m-1}$ 及 $v'=-m(1-x)^{m-1}$。根据商法则，

$$\mathrm{f}'(x)=\frac{(u'-v')(u+v)-(u-v)(u'+v')}{(u+v)^2}=\frac{2(u'v-uv')}{(u+v)^2}.$$

分子可以化简：

$$\begin{aligned}u'v-uv' &= m(1+x)^{m-1}(1-x)^m+m(1+x)^m(1-x)^{m-1}\\ &= m(1+x)^{m-1}(1-x)^{m-1}\big[(1-x)+(1+x)\big]\\ &= 2m(1-x^2)^{m-1}.\end{aligned}$$

因此

$$\mathrm{f}'(x)=\frac{4m\,(1-x^2)^{m-1}}{\big[(1+x)^m+(1-x)^m\big]^2}.$$

$m=5$ 的情况：曲线 $y=\mathrm{f}(x)$

用二项式定理（binomial theorem）展开，分子中的偶次幂互相抵消，分母中的奇次幂互相抵消：

$$\mathrm{f}(x)=\frac{2(5x+10x^3+x^5)}{2(1+10x^2+5x^4)}=\frac{x(x^4+10x^2+5)}{5x^4+10x^2+1},\qquad \mathrm{f}'(x)=\frac{20(1-x^2)^4}{\big[(1+x)^5+(1-x)^5\big]^2}.$$

现在逐一读出曲线的特征。

- 对称性。$\mathrm{f}(-x)=-\mathrm{f}(x)$，所以曲线绕原点有旋转对称（rotational symmetry）。曲线通过 $(0,0)$，该处的斜率（gradient）为 $\mathrm{f}'(0)=20/2^2=5$。
- 单调性（*monotonic*）。对所有 x 都有 $\mathrm{f}'(x)\geq 0$，等号只在 $x=\pm1$ 时成立。所以 f 处处递增。在 $x=\pm1$ 处曲线有驻点（stationary point），而由于 $(1-x^2)^4$ 不变号，这些驻点是拐点（point of inflection），而非转向点（turning point）。它们位于 $(1,1)$ 和 $(-1,-1)$：$(1-x)^5$ 或 $(1+x)^5$ 一项变为零，余下 $\mathrm{f}=\pm1$。
- $|x|$ 很大时。$\mathrm{f}(x)\approx x/5$。更准确地说，

  $$\mathrm{f}(x)-\frac{x}{5}=\frac{x(40x^2+24)}{5(5x^4+10x^2+1)},$$

  当 $|x|\to\infty$ 时，上式趋向 0。所以 $y=x/5$ 是斜渐近线（oblique asymptote）。曲线在 $x>0$ 时位于渐近线之上，在 $x<0$ 时位于渐近线之下。

? **想一想** $\mathrm{f}(x)$ 的绝对值在哪里大于 1，又在哪里小于 1?

由于 f 递增且 $f(\pm1)=\pm1$，当 $|x|<1$ 时有 $|f(x)|<1$，当 $|x|>1$ 时有 $|f(x)|>1$。这一点对倒数曲线很重要。（当 $|x|<1$ 时，代入 $x=\tanh t$ 可得 $(1+x)/(1-x)=e^{2t}$，再把 f 的分子和分母同除以 $(1-x)^5$，便得 $f(\tanh t)=\tanh 5t$。这是理解曲线在 -1 与 1 之间呈 S 形的另一个角度。）

曲线 $y=1/f(x)$

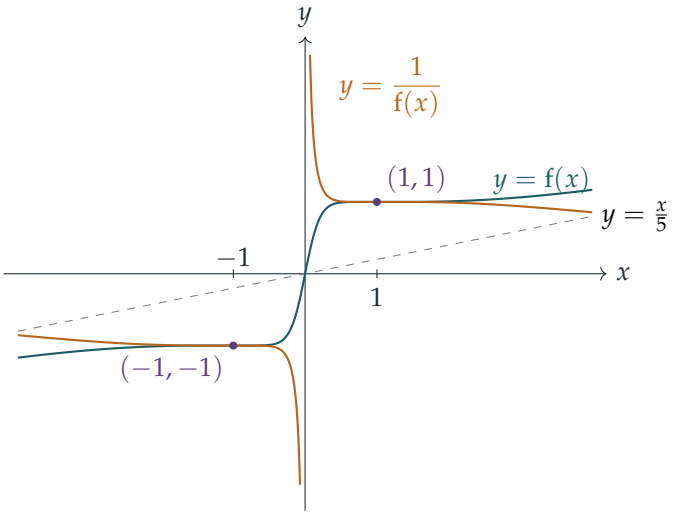
把 x 换成 $1/x$，再把分子和分母同乘以 x^5，可得

$$f(1/x)=\frac{(x+1)^5-(x-1)^5}{(x+1)^5+(x-1)^5}=\frac{(1+x)^5+(1-x)^5}{(1+x)^5-(1-x)^5}=\frac{1}{f(x)},$$

当中用了 $(x-1)^5=-(1-x)^5$。所以 $y=1/f(x)$ 就是曲线 $y=f(1/x)$，它的特征可由 f 的特征推出：

- 它是奇函数（odd function），在 $x=0$ 处没有定义，而这是 f 唯一的零点。在 0 附近，$1/f(x)\approx 1/(5x)$，所以 y 轴是铅垂渐近线（vertical asymptote），当 $x\to 0^+$ 时 $y\to+\infty$，当 $x\to 0^-$ 时 $y\to-\infty$。
- 它的导数是 $-f'(x)/f(x)^2\le 0$，所以它在 $x>0$ 上递减，在 $x<0$ 上亦递减。它在 $(1,1)$ 和 $(-1,-1)$ 有属于驻点的拐点。
- 当 $|x|\to\infty$ 时，$1/f(x)\approx 5/x\to 0$，所以 x 轴是渐近线。曲线在 $x>0$ 时位于其上，在 $x<0$ 时位于其下。

两条曲线只在 $f(x)^2=1$ 之处相遇，即 $(1,1)$ 和 $(-1,-1)$。两条曲线在该处都是水平的，所以它们是相切而非相交穿过。



验算

- 取 $m = 1$：公式给出 $\mathrm{f}'(x) = 4/2^2 = 1$，而事实上 $\mathrm{f}(x) = 2x/2 = x$。
- 以数值计算，$\mathrm{f}(1000)/1000 \approx 0.2000016$，与渐近线 $y = x/5$ 吻合。对 $m = 1, \ldots, 8$，在随机选取的点上，以中心差分（central difference）求得的导数与 f' 的公式一致。
- $\mathrm{f}(\tanh t) = \tanh 5t$ 和 $\mathrm{f}(1/x) = 1/\mathrm{f}(x)$ 在随机选取的点上均准确至 10^{-9}。

常见错误

- 把 $x = \pm 1$ 称为极大点或极小点。导数在该处触及零但不变号，所以这些是属于驻点的拐点。
- 遗漏斜渐近线：f 是五次多项式除以四次多项式，所以它的增长有如 $x/5$，而不会趋于平缓。
- 把 $y = 1/\mathrm{f}(x)$ 画成在 $(\pm 1, \pm 1)$ 穿过 $y = \mathrm{f}(x)$。两条曲线在该处都是平的，所以它们只是相切。

STEP 之外

代换 $x = \tanh t$ 显示，f 正是以 $\tanh t$ 表示 $\tanh 5t$ 的公式，也就是 $\tan 5\theta$ 的倍角公式（multiple-angle formula）的双曲（hyperbolic）版本。由此观之，f 是一个有理映射（rational map），它共轭（conjugate）于乘以 5 的运算。这类映射，以及它们在迭代（iteration）下的表现，正是复动力系统（complex dynamics）的起点之一。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）是一本优美的书，可从中看到这类有理函数如何作用于整个复平面。

详细解答

STEP III 2003，第 4 题

纯数学 · 参数曲线的切线

> **STEP III 2003，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/03-s3-q4

关键思路

直线与这条曲线相交之处，正是某个以参数为变量的三次式（cubic）等于零之处。切线（tangent）会在 t 处产生一个二重根（double root），所以由根之和便可立即求出第三个根 T。T 的公式看起来很像 tan 的倍角公式（double-angle formula），这提示我们设 $t = \tan\theta$：映射 $t \mapsto T$ 于是变成角度上的一个简单线性映射，要把它迭代（iterate）便很容易。

解答

T 的公式

由于 $y = t + t^3$ 对 t 严格递增（strictly increasing），不同的参数（parameter）给出不同的点，所以曲线上的点与参数的值一一对应（one-to-one）。

求导可得 $\dfrac{\mathrm{d}x}{\mathrm{d}t} = 2t$ 和 $\dfrac{\mathrm{d}y}{\mathrm{d}t} = 1 + 3t^2$，所以当 $t \neq 0$ 时，在 t 处的切线的斜率（gradient）为

$$m = \frac{1 + 3t^2}{2t}.$$

（因为 $t \neq 0$，切线不是垂直的。）切线的方程是 $y - (t + t^3) = m(x - t^2)$。参数为 s 的点在这条切线上，当且仅当

$$s + s^3 - (t + t^3) = m(s^2 - t^2), \qquad \text{即} \qquad f(s) \equiv s^3 - ms^2 + s - \left(t + t^3 - mt^2\right) = 0.$$

显然 $f(t) = 0$，而根据 m 的选取，$f'(t) = 3t^2 - 2mt + 1 = 0$。所以 $s = t$ 是这个三次式的（至少）二重根，而各根为 t, t, T，其中 T 是某个实数。$s^3 - ms^2 + \ldots$ 的根之和是 m，

所以

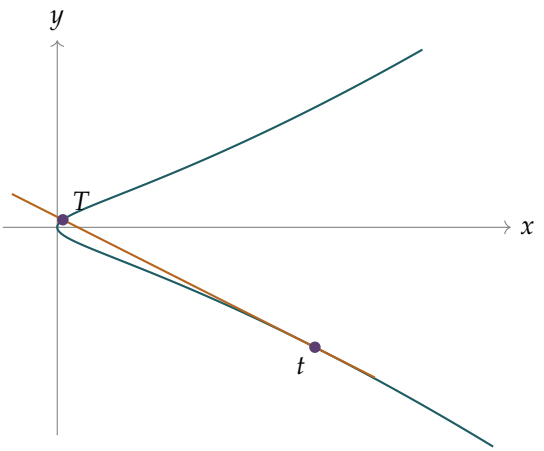
$$2t + T = \frac{1+3t^2}{2t}$$
$$T = \frac{1+3t^2-4t^2}{2t} = \frac{1-t^2}{2t}.$$

这个根 T 是该直线与曲线唯一可能相交的另一处。

只有当 $T \neq t$ 时，切线才会与曲线再次相交。现在

$$T = t \iff 1 - t^2 = 2t^2 \iff 3t^2 = 1,$$

所以我们需要 $3t^2 \neq 1$。（当 $3t^2 = 1$ 时，t 是三重根，切线只与曲线相接一次：这些正是拐点（points of inflection）。）这就证明了两个结果。



以 $t_0 = \tan\frac{7}{18}\pi$ 开始的数列

设 $t = \tan\theta$。由倍角公式，

$$T = \frac{1-\tan^2\theta}{2\tan\theta} = \frac{1}{\tan 2\theta} = \cot 2\theta = \tan\left(\frac{\pi}{2} - 2\theta\right).$$

所以若 $t_i = \tan\theta_i$，我们可以取 $\theta_{i+1} = \frac{\pi}{2} - 2\theta_i$。由于 tan 的周期（period）为 π，而且在任何长度为 π 的区间上（除去 $\frac{\pi}{2}$ 的奇数倍）都是一对一的，两个参数相等，当且仅当它们的角相差 π 的倍数。

? **想一想** 在迭代之前，我们必须检查每个 t_i 的甚么性质，才能确保下一个点存在?

这个构作在每一步都需要 $t_i \neq 0$ 和 $3t_i^2 \neq 1$，即 θ_i 不是 $\frac{\pi}{2}$ 的倍数，也不是 $\pm\frac{\pi}{6}$ 加上 π 的倍数。以 $\frac{\pi}{18}$ 的倍数计，我们必须避开模（modulo）18 下的 $0, \pm 3, \pm 9$。

由 $\theta_0 = \frac{7\pi}{18}$ 开始：

$$\begin{aligned}\theta_1 &= \frac{9\pi}{18} - \frac{14\pi}{18} = -\frac{5\pi}{18},\\ \theta_2 &= \frac{9\pi}{18} + \frac{10\pi}{18} = \frac{19\pi}{18} \equiv \frac{\pi}{18} \pmod{\pi},\\ \theta_3 &= \frac{9\pi}{18} - \frac{2\pi}{18} = \frac{7\pi}{18}.\end{aligned}$$

$7, -5, 1$ 都不在禁用之列，所以每一步都成立。因此 $t_3 = \tan\frac{7\pi}{18} = t_0$，即 $P_3 = P_0$。但 $\theta_1 - \theta_0 = -\frac{12\pi}{18}$ 不是 π 的倍数，所以 $t_1 = \tan(-\frac{5\pi}{18}) < 0 < t_0$，而 $P_1 \neq P_0$。这三点的参数分别是 $\tan 70^\circ$、$-\tan 50^\circ$ 和 $\tan 10^\circ$。

第二个起始值

设 $g(\theta) = \frac{\pi}{2} - 2\theta$。迭代可得

$$g^2(\theta) = -\frac{\pi}{2} + 4\theta, \qquad g^3(\theta) = \frac{3\pi}{2} - 8\theta.$$

所以 $P_3 = P_0$ 当且仅当 $\frac{3\pi}{2} - 8\theta \equiv \theta \pmod{\pi}$，即 $9\theta \equiv \frac{3\pi}{2} \pmod{\pi}$，由此得

$$\theta = \frac{\pi}{6} + \frac{k\pi}{9} = \frac{(3+2k)\pi}{18}, \qquad k \in \mathbb{Z}.$$

同理，$P_1 = P_0$ 当且仅当 $3\theta \equiv \frac{\pi}{2} \pmod{\pi}$，即 $\theta = \frac{(3+6j)\pi}{18}$。这些正是 $3 + 2k \equiv 3, 9, 15 \pmod{18}$ 的情况：它们给出 $t = \pm\frac{1}{\sqrt{3}}$，或者根本没有有限的 t，本来就已被排除。

在模 π 之下，余下的是 $\theta \in \{\frac{\pi}{18}, \frac{5\pi}{18}, \frac{7\pi}{18}, \frac{11\pi}{18}, \frac{13\pi}{18}, \frac{17\pi}{18}\}$。要 $t_0 > 0$，我们需要 $0 < \theta < \frac{\pi}{2}$，由此得 $\tan\frac{\pi}{18}$、$\tan\frac{5\pi}{18}$ 和题目给出的 $\tan\frac{7\pi}{18}$。$\tan\frac{\pi}{18}$ 这个值只是上述循环（cycle）中的 P_2。给出一个真正全新循环的值是

$$t_0 = \tan\frac{5\pi}{18} \ (= \tan 50^\circ),$$

其中 $\theta_0 = \frac{5\pi}{18}$，$\theta_1 = -\frac{\pi}{18}$，$\theta_2 = \frac{11\pi}{18}$，$\theta_3 = \frac{9\pi}{18} - \frac{22\pi}{18} = -\frac{13\pi}{18} \equiv \frac{5\pi}{18}$。$5, -1, 11$ 都不在禁用之列，而且 $t_1 = -\tan 10^\circ \neq t_0$。$\tan\frac{5\pi}{18}$ 或 $\tan\frac{\pi}{18}$ 都可以作为本题的答案。

> ? **想一想** 为甚么可以“模 π”地处理角度?

因为我们唯一用到的就是 $t = \tan\theta$，而 $\tan(\theta + \pi) = \tan\theta$。把 θ_i 加上 π，$g(\theta_i)$ 便改变 2π，这并不改变 t_{i+1}。所以整个参数数列只取决于 θ_0 模 π 的值。

验算

▶ 取 $t = -1$ 可得 $T = 0$：在 $(1, -2)$ 处的切线斜率为 -2，而它确实通过原点。

- 以数值计算，$t_0 = \tan 70^\circ \approx 2.747477$ 给出 $t_1 \approx -1.191754$、$t_2 \approx 0.176327$ 和 $t_3 \approx 2.747477$。同样地，$\tan 50^\circ \approx 1.191754 \to -0.176327 \to -2.747477 \to 1.191754$。
- 随机选取 t，以数值方法解三次方程求另一个交点，结果与 $T = (1 - t^2)/(2t)$ 吻合。

常见错误

- 先求弦（chord）的斜率再令两者相等，却忘记约去因子 $T - t$。利用关于根之和的韦达公式（Vieta's formula），便可完全避开这个问题。
- 直接用小数对 t 迭代这个映射，结果看不出精确的循环。正是代换 $t = \tan\theta$ 才令 $P_3 = P_0$ 精确成立。
- 把 $\tan\frac{\pi}{6}$ 当作第二个值：这时 $T = t$，所以根本没有下一个点。

STEP 之外

弦切构作法（chord-and-tangent construction）是椭圆曲线（elliptic curves）理论的核心：在该理论中，同一条“第三个交点”规则把三次曲线上的点变成一个群（group）。这里的曲线 $y^2 = x(1+x)^2$ 是一条奇异三次曲线（singular cubic）（它在 $(-1, 0)$ 有一个孤立点），而代换 $t = \tan\theta$ 表明，切线映射其实是乔装了的角度加倍。这就是为甚么这些循环来自把圆分成九等份。Silverman 与 Tate 合著的 *Rational Points on Elliptic Curves*（《椭圆曲线上的有理点》）以本科程度把这个课题阐述得十分精彩。

详细解答

STEP III 2003，第 5 题

纯数学 · 三次方程的根

STEP III 2003，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/03-s3-q5

关键思路

从图像入手数算根（root）的个数。首项系数（leading coefficient）为正的二次函数，其图像与 x 轴相交两次，当且仅当其极小值位于该轴下方。首项系数为正的三次函数，其图像与该轴相交三次，当且仅当其极大值在轴的上方而极小值在轴的下方。要令所有根皆为正数，就要看转向点（turning point）位于 y 轴的哪一侧，以及曲线与 y 轴相交处的正负号。最后一条方程其实是经过平移的三次曲线的一个乔装例子：配立方（complete the cube）便可把它化成该形式。

解答

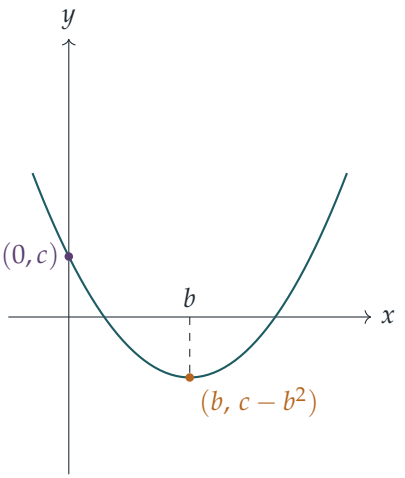
二次函数

配方（completing the square）得

$$y = x^2 - 2bx + c = (x - b)^2 + c - b^2,$$

所以转向点（极小点）位于 $(b,\ c - b^2)$。

当有两个相异实根时，曲线是一条 $\cup$ 形的抛物线（parabola），对称于 $x = b$，其顶点（vertex）位于 x 轴下方。它与 y 轴相交于 $(0, c)$。



两个相异实根。抛物线的首项系数为正，所以当 $x \to \pm\infty$ 时 $y \to +\infty$，而它的最小值是 $c - b^2$。若 $c - b^2 < 0$，曲线会降至轴的下方，然后在 $x = b$ 的两侧各自回升，因此与轴相交两次。若 $c - b^2 = 0$，曲线仅与轴相切（重根（repeated root））；若 $c - b^2 > 0$，则曲线与轴永不相遇。所以条件是

$$b^2 > c.$$

两个相异正根。我们断言条件是

$$b^2 > c, \qquad b > 0, \qquad c > 0.$$

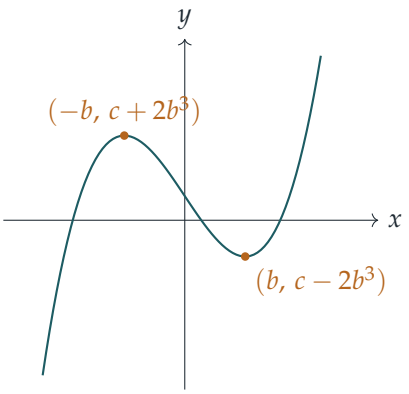
必要性（necessary）：两根 α, β 满足 $\alpha + \beta = 2b$ 及 $\alpha\beta = c$，若两根皆为正数，则这两个量亦皆为正数。**充分性（sufficient）：**若 $b^2 > c$，则有两个根，分别位于 $x = b$ 的两侧。由于 $b > 0$，右边的根是正数。左边的根位于曲线递减的区间内；曲线在 $x = 0$ 处为正（高度为 $c > 0$），在 $x = b > 0$ 处为负，所以这个根严格介乎 0 与 b 之间，因而也是正数。

三次函数

对于 $y = x^3 - 3b^2x + c$，我们有 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = 3x^2 - 3b^2 = 3(x - b)(x + b)$，它在 $x = \pm b$ 时等于零。代入得

$$x = -b:\ y = -b^3 + 3b^3 + c = c + 2b^3, \qquad x = b:\ y = b^3 - 3b^3 + c = c - 2b^3.$$

由于 $b > 0$，转向点分别是位于 $(-b,\ c + 2b^3)$ 的极大点和位于 $(b,\ c - 2b^3)$ 的极小点：导数（derivative）在 $x < -b$ 时为正，在两者之间为负，在 $x > b$ 时为正。



曲线由 $-\infty$ 上升至极大点，下降至极小点，然后上升至 $+\infty$，而且在这三段上各自严格单调（strictly monotonic）。所以曲线在每一段上各有一个根，从而共有三个相异的根，当且仅当极大值在轴的上方而极小值在轴的下方：

$$c+2b^3>0 \text{ 及 } c-2b^3<0, \qquad \text{即} \qquad -2b^3<c<2b^3.$$

若其中一个转向点的函数值为零，便会得到重根；若两者同为严格的正数或同为严格的负数，则只有一个实根。所以 $|c|<2b^3$ 是充分必要条件。

平移后的三次曲线的三个相异正根

曲线 $y=(x-a)^3-3b^2(x-a)+c$ 是把上一条曲线向右平移（translate）a 而得（我们仍假设 $b>0$）。所以它有三个相异实根，当且仅当 $|c|<2b^3$；这时它的极大点在 $x=a-b$，极小点在 $x=a+b$，而三个根 $x_1<x_2<x_3$ 满足

$$x_1<a-b<x_2<a+b<x_3.$$

三个根全为正数，当且仅当最小的根 x_1 为正数。记

$$y(0)=(-a)^3-3b^2(-a)+c=c-a^3+3ab^2$$

为曲线与 y 轴相交处的高度。我们断言条件是

$$|c|<2b^3, \qquad a>b, \qquad c<a^3-3ab^2.$$

想一想 为甚么单凭“$y(0)<0$”并不足以保证三个根全都在原点的右方?

因为 y 在 x_2 与 x_3 之间同样是负数。在 $x=0$ 处取负值，有可能表示原点位于中间的根与最大的根之间。我们还须知道 0 位于曲线第一段上升的部分，即极大点的左方。这正是 $a>b$ 的作用。

必要性。假设有三个相异正根。那么由上一部分可知 $|c| < 2b^3$。此外 $0 < x_1 < a - b$，所以 $a > b$。最后，y 在 $(-\infty, a-b]$ 上严格递增，而 $y(x_1) = 0$，所以 $y(0) < 0$，因为 $0 < x_1$。由此得 $c < a^3 - 3ab^2$。

充分性。假设三个条件全部成立。这时有三个相异实根，而且 $0 < a - b$，所以 0 位于 $(-\infty, a-b]$ 内，而 y 在此区间上严格递增。由于 $y(0) < 0 = y(x_1)$，我们得到 $0 < x_1$，因此三个根全为正数。

所以该方程有三个相异正根，当且仅当

$$a > b \quad \text{及} \quad -2b^3 < c < \min\bigl(2b^3,\ a^3 - 3ab^2\bigr).$$

（由于当 $a > b > 0$ 时 $a^3 - 3ab^2 + 2b^3 = (a-b)^2(a+2b) > 0$，上述对 c 的上下界并无矛盾。）

最后的方程

先除以 2，然后配立方。$(x-a)^3$ 中的 x^2 项是 $-3ax^2$，所以我们需要 $3a = \frac{9}{2}$，即 $a = \frac{3}{2}$。设 $x = t + \frac{3}{2}$，

$$\begin{aligned} x^3 - \tfrac{9}{2}x^2 + \tfrac{7}{2}x - \tfrac{1}{2} &= t^3 + \bigl(\tfrac{27}{4} - \tfrac{27}{2} + \tfrac{7}{2}\bigr)t + \bigl(\tfrac{27}{8} - \tfrac{81}{8} + \tfrac{21}{4} - \tfrac{1}{2}\bigr) \\ &= t^3 - \tfrac{13}{4}t - 2, \end{aligned}$$

其中 t^2 项因 a 的选取而互相抵消。所以

$$2x^3 - 9x^2 + 7x - 1 = 0 \iff \bigl(x - \tfrac{3}{2}\bigr)^3 - 3b^2\bigl(x - \tfrac{3}{2}\bigr) + c = 0, \qquad a = \tfrac{3}{2},\ b^2 = \tfrac{13}{12},\ c = -2,$$

其中 $b = \sqrt{13/12} > 0$。现在逐一检查三个条件。

- $|c| < 2b^3$：这即是说 $2 < 2b^3$，亦即 $b^2 > 1$，而这是成立的，因为 $b^2 = \frac{13}{12} > 1$。
- $a > b$：两者皆为正数，而且 $a^2 = \frac{9}{4} = \frac{27}{12} > \frac{13}{12} = b^2$。
- $c < a^3 - 3ab^2$：这里 $a^3 - 3ab^2 = a(a^2 - 3b^2) = \frac{3}{2}\bigl(\frac{9}{4} - \frac{13}{4}\bigr) = -\frac{3}{2}$，而 $-2 < -\frac{3}{2}$。

三个条件全部成立，所以 $2x^3 - 9x^2 + 7x - 1 = 0$ 有三个相异正根，正如所求。

验算

- 正负号的变化可直接确认最后的答案。设 $p(x) = 2x^3 - 9x^2 + 7x - 1$，我们有 $p(0) = -1$、$p(\frac{1}{2}) = \frac{1}{2}$、$p(1) = -1$、$p(3) = -7$ 及 $p(4) = 11$。所以在 $(0, \frac{1}{2})$、$(\frac{1}{2}, 1)$ 及 $(3, 4)$ 内各有一个根。它们的数值分别是 0.185、0.760 及 3.555。
- 数值 $y(0) = c - a^3 + 3ab^2 = -2 + \frac{3}{2} = -\frac{1}{2}$ 与首一（monic）三次多项式 $x^3 - \frac{9}{2}x^2 + \frac{7}{2}x - \frac{1}{2}$ 的常数项吻合，这是理所当然的。

- 以数千组随机选取的 (a,b,c) 作测试，结果每一次都与上述条件相符。

常见错误

- 只用 $y(0)<0$（或只用 $a>b$）来判断根是否为正。两者缺一不可：当原点位于较大的两个根之间时，$y(0)<0$ 同样成立。
- 忘记了 $|c|\le 2b^3$ 中等号成立时会得到重根，因此不算“三个相异的根”。
- 把平移的方向弄反：以 $x-a$ 表示的曲线，是把原来的曲线向*右*移动 a。

STEP 之外

任何三次多项式都可以像这里一样，借着平移化成“缺项”（depressed）形式 t^3+pt+q。这时它有三个相异实根，当且仅当 $4p^3+27q^2<0$；取 $p=-3b^2$、$q=c$，这就是上文求得的条件 $c^2<4b^6$。$-(4p^3+27q^2)$ 这个量称为*判别式*（discriminant），它是三次方程如何得以求解（卡尔达诺（Cardano）公式）这段故事的开端，最终更引出伽罗瓦理论（Galois theory）。Allenby 的 *Rings, Fields and Groups*（环、域与群）是迈向该理论的理想下一步。

详细解答

STEP III 2003，第 6 题

纯数学 · 三角函数裂项相消求和

STEP III 2003，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/03-s3-q6

关键思路

各项角度以相等间隔递增的余弦之和，照原样处理并不方便；但只要把每一项都乘以 $2\sin\frac{1}{2}\theta$，每个余弦就会变成两个正弦之*差*，整个和式随即可以裂项相消（telescoping）。余下的是两个正弦之差，再用和差化积公式（factor formula）便可化为乘积。令一个乘积等于零，是很容易处理的事。唯一要留神的是：我们乘上了 $\sin\frac{1}{2}\theta$，而它本身也可能等于零。

解答

恒等式

利用 $\sin(A+B)-\sin(A-B)=2\cos A\sin B$，取 $A=r\theta$ 及 $B=\frac{1}{2}\theta$：

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

这就是所需的结果。

级数求和

记

$$S(\theta)=\sum_{r=a}^{b-1}\cos r\theta.$$

乘以 $2\sin\frac{1}{2}\theta$，并对由 a 至 $b-1$ 的每个 r 使用上述恒等式（identity），

$$\begin{aligned} 2\sin\tfrac{1}{2}\theta\, S(\theta) &= \sum_{r=a}^{b-1}\Big[\sin\big(r+\tfrac{1}{2}\big)\theta - \sin\big(r-\tfrac{1}{2}\big)\theta\Big] \\ &= \sin\big(b-\tfrac{1}{2}\big)\theta - \sin\big(a-\tfrac{1}{2}\big)\theta, \end{aligned}$$

因为每个 $\sin\big(r+\frac{1}{2}\big)\theta$ 都与下一项的 $-\sin\big((r+1)-\frac{1}{2}\big)\theta$ 互相抵消，最后只剩下末项的上端和首项的下端。

现在使用 $\sin P - \sin Q = 2\cos\frac{P+Q}{2}\sin\frac{P-Q}{2}$，其中 $\frac{P+Q}{2} = \frac{1}{2}(a+b-1)\theta$ 及 $\frac{P-Q}{2} = \frac{1}{2}(b-a)\theta$：

$$\sin\tfrac{1}{2}\theta\, S(\theta) = \cos\tfrac{1}{2}(a+b-1)\theta\ \sin\tfrac{1}{2}(b-a)\theta. \qquad (*)$$

解 $S(\theta) = 0$

> **想一想** 我们可以直接把 $(*)$ 除以 $\sin\frac{1}{2}\theta$ 吗?

只有当 $\sin\frac{1}{2}\theta \neq 0$，即 θ 不是 2π 的倍数时才可以。所以我们先处理这些值。若 $\theta = 2m\pi$，其中 m 为整数，则每一项 $\cos r\theta$ 都等于 1，故 $S(\theta) = b-a \geq 2$。这些值永远不是解。

对于其余所有 θ，我们有 $\sin\frac{1}{2}\theta \neq 0$，而由 $(*)$ 可得

$$S(\theta) = 0 \iff \cos\tfrac{1}{2}(a+b-1)\theta = 0 \ \text{ 或 } \ \sin\tfrac{1}{2}(b-a)\theta = 0.$$

- $\cos\frac{1}{2}(a+b-1)\theta = 0$ 当且仅当 $\frac{1}{2}(a+b-1)\theta = \big(k+\frac{1}{2}\big)\pi$，即

$$\theta = \frac{(2k+1)\pi}{a+b-1}, \qquad k \in \mathbb{Z}.$$

  （注意 $a+b-1 \geq 3$，所以我们并没有除以零。）这些值全都不是 2π 的倍数：否则须有 $2k+1 = 2m(a+b-1)$，即一个奇数等于一个偶数。所以它们全部都是解。

- $\sin\frac{1}{2}(b-a)\theta = 0$ 当且仅当 $\frac{1}{2}(b-a)\theta = k\pi$，即

$$\theta = \frac{2k\pi}{b-a}, \qquad k \in \mathbb{Z}.$$

  这个值是 2π 的倍数，当且仅当 $b-a$ 整除 k，而我们已经知道这些值并不是解。所以必须把它们剔除。

> **想一想** 在 2π 的倍数处，$(*)$ 看来是成立的，为甚么仍要把它们从第二族解中剔除?

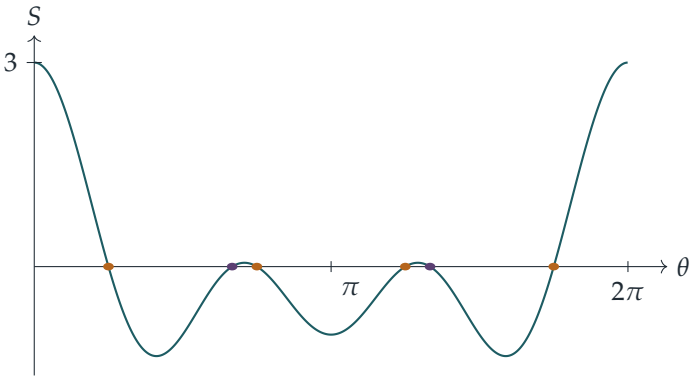
在 $\theta = 2m\pi$ 处，不论 S 取甚么值，$(*)$ 的两边都是零，所以 $(*)$ 在这些位置完全没有告诉我们关于 S 的任何资料。这只是全式乘以 $\sin\frac{1}{2}\theta$ 所造成的假象。由于 $b-a \geq 2$，第二族之中仍然有真正的解，例如 $\theta = 2\pi/(b-a)$。

综合以上所得，全部的解正好是

$$\theta = \frac{(2k+1)\pi}{a+b-1}\ (k \in \mathbb{Z}) \qquad 或 \qquad \theta = \frac{2k\pi}{b-a}\ (k \in \mathbb{Z},\ b-a \nmid k).$$

这两族解可能有重叠。这并无大碍，因为同时属于两族的值仍然只算一个解。

举例来说，当 $a = 1$ 及 $b = 4$ 时，方程是 $\cos\theta + \cos 2\theta + \cos 3\theta = 0$。在 $(0, 2\pi)$ 内，它的根有来自第一族的 $\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}$，以及来自第二族的 $\frac{2\pi}{3}, \frac{4\pi}{3}$。注意 $\theta = 0$ 和 $\theta = 2\pi$ 并不包括在内，该处 $S = 3$。



曲线 $S = \cos\theta + \cos 2\theta + \cos 3\theta$：琥珀色的根来自余弦因子，紫红色的根来自正弦因子。

验算

- 取 $a = 1$、$b = 4$：在 $\theta = \frac{\pi}{4}$ 处，$\frac{\sqrt{2}}{2} + 0 - \frac{\sqrt{2}}{2} = 0$；在 $\theta = \frac{2\pi}{3}$ 处，$-\frac{1}{2} - \frac{1}{2} + 1 = 0$。
- 以数值方法检验，对多组 a, b 随机选取 θ，$(*)$ 均成立。对所有 $1 \le a \le 6$、$a+2 \le b \le 11$，在 $(0, 6\pi)$ 上对 S 作细密扫描，找到的根与公式所给出的完全一致，再无其他。例如当 $a = 1$、$b = 3$ 时，根 $\theta = 13\pi/3 \approx 13.6136$ 来自第一族。

常见错误

- 未经检查便把 $(*)$ 除以 $\sin\frac{1}{2}\theta$，因而把 $\theta = 2m\pi$ 也算作解；在这些位置 $S = b - a \neq 0$。
- 弄错裂项相消后剩下的首末两项：它们是 $\sin\left(b - \frac{1}{2}\right)\theta$ 和 $\sin\left(a - \frac{1}{2}\right)\theta$，因为最后一个余弦是 $\cos(b-1)\theta$。
- 只给出某一区间（例如 $[0, 2\pi)$）内的解，但题目要求的是全部的解。

STEP 之外

$\sum \cos r\theta$ 的闭合形式（closed form），其实是以 $e^{i\theta}$ 为公比的等比级数（geometric series）的实部。取 $a = -n$ 及 $b = n+1$，便得到狄利克雷核（Dirichlet kernel）$\sin\left(n+\frac{1}{2}\right)\theta / \sin\frac{1}{2}\theta$，它是傅里叶级数（Fourier series）理论以及其收敛（converge）问题的核心。想明白为何这类三角和一旦改用 $e^{i\theta}$ 来处理就会变得简单，Needham 的 *Visual Complex Analysis* 是很好的入门读物。

详细解答

STEP III 2003，第 7 题

纯数学 · 西姆松线

> **STEP III 2003，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/03-s3-q7

关键思路

要找垂足（foot of the perpendicular），可从 P 出发，沿直线的法线（normal）方向一直走到碰上该直线为止，所以写 $N = P + \lambda(\text{法向量})$，再解出 λ。至于第二部分，关于 s 和 t 的条件是两个斜率（gradient）之积等于 -1：它说明 PA 与 PB 互相垂直。把它化成圆的方程后，正好令共线（collinear）的验证迎刃而解。

解答

N 的坐标

直线 AB 的方程为 $\dfrac{x}{a} + \dfrac{y}{b} = 1$，即

$$bx + ay = ab.$$

这条直线的一个法向量（normal vector）是 (b, a)，所以由 P 所作垂线的垂足 N 具有以下形式：

$$N = (s + \lambda b,\ t + \lambda a)$$

其中 λ 为某个数。由于 N 在 AB 上，

$$b(s + \lambda b) + a(t + \lambda a) = ab \quad \Longrightarrow \quad \lambda = \frac{ab - bs - at}{a^2 + b^2}.$$

因此

$$x = s + \lambda b = \frac{s(a^2 + b^2) + ab^2 - b^2 s - abt}{a^2 + b^2} = \frac{ab^2 - abt + a^2 s}{a^2 + b^2} = \frac{ab^2 - a(bt - as)}{a^2 + b^2},$$

$$y = t + \lambda a = \frac{t(a^2 + b^2) + a^2 b - abs - a^2 t}{a^2 + b^2} = \frac{a^2 b + b^2 t - abs}{a^2 + b^2} = \frac{a^2 b + b(bt - as)}{a^2 + b^2},$$

这正是所需的结果。

N 在 XY 上

首先看看这个条件说的是甚么。要令它有意义，必须有 $s \neq 0$ 及 $s \neq a$；又因为乘积是 -1 而不是 0，所以还有 $t \neq 0$ 及 $t \neq b$。展开后得

$$t(t-b) = -s(s-a) \quad \Longleftrightarrow \quad s^2 + t^2 = as + bt. \tag{$*$}$$

现在 $X = (s, 0)$，$Y = (0, t)$。由于 $s \neq 0$ 及 $t \neq 0$，这是两个不在原点的相异点，而直线 XY 为

$$\frac{x}{s} + \frac{y}{t} = 1, \qquad \text{即} \qquad tx + sy = st.$$

所以我们要证明 N 的坐标满足 $tx + sy = st$。

> **想一想** 我们应该直接代入然后听天由命，还是有更整洁的方法去整理这些代数运算？

为求简洁，设 $k = bt - as$。利用上面的公式，

$$\begin{aligned}(a^2+b^2)(tx+sy) &= t(ab^2 - ak) + s(a^2b + bk) \\ &= ab^2t + a^2bs + k(bs - at).\end{aligned}$$

最后那个乘积展开后十分整齐：

$$k(bs-at) = (bt-as)(bs-at) = b^2st - abt^2 - abs^2 + a^2st = (a^2+b^2)st - ab(s^2+t^2).$$

因此

$$(a^2+b^2)(tx+sy) = (a^2+b^2)st + ab\big(as + bt - s^2 - t^2\big).$$

这对任何 P 都成立。根据 $(*)$，括号内的式子等于零，所以 $tx + sy = st$，即 N 在直线 XY 上。

几何意义

PB 的斜率是 $\dfrac{t-b}{s}$，PA 的斜率是 $\dfrac{t}{s-a}$，所以该条件说的是 $PA \perp PB$：角 APB 是直角。等价地，$(*)$ 可写成

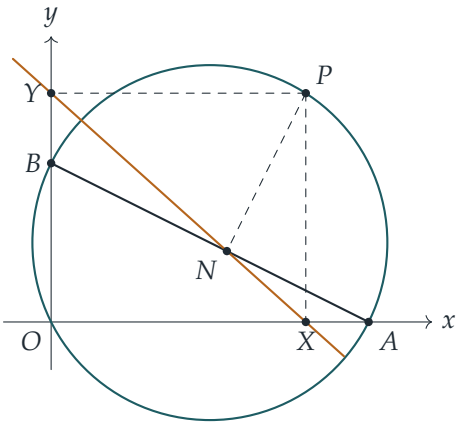
$$\left(s - \tfrac{a}{2}\right)^2 + \left(t - \tfrac{b}{2}\right)^2 = \frac{a^2+b^2}{4},$$

所以 P 在以 AB 为直径的圆上。由于角 AOB 也是直角，这个圆通过 O：它就是三角形 OAB 的外接圆（*circumcircle*）。

X、Y 和 N 三点，分别是由 P 到三角形三边 OA、OB 和 AB 所作垂线的垂足。所以这个结果说的是：

若 P 在一个三角形的外接圆上，则由 P 到三边（有需要时延长）所作垂线的垂足共线。

这就是西姆松线（*Simson line*）定理，这里证明的是直角三角形的情况，而直角顶点放在原点。



取 $a = 4$，$b = 2$：P 在通过 O、A、B 的圆上；直线 XY（琥珀色）通过 N。

验算

- 取 $P = O$，即 $s = t = 0$：公式给出 $N = \left(\frac{ab^2}{a^2+b^2}, \frac{a^2b}{a^2+b^2}\right)$，它既在 $bx + ay = ab$ 上，也在通过 O 且垂直于 AB 的直线 $y = \frac{a}{b}x$ 上，与预期相符。
- 取 $a = 4$，$b = 2$，以及圆上的点 $P = (2 + \sqrt{5}\cos 1,\ 1 + \sqrt{5}\sin 1) \approx (3.208, 2.882)$，公式给出 $N \approx (2.214, 0.893)$，而 $\frac{2.214}{3.208} + \frac{0.893}{2.882} \approx 0.690 + 0.310 = 1$，所以 N 在 XY 上。
- 恒等式 $(a^2 + b^2)(tx + sy - st) = ab(as + bt - s^2 - t^2)$ 同时给出逆命题（converse）：若 P 不在坐标轴上，则仅当 P 在圆上时，N 才在 XY 上。

常见错误

- 把 (a, b) 当作 AB 的法向量。AB 的方向是 $(-a, b)$，所以法向量是 (b, a)。
- 一次过把所有东西展开，结果淹没在算式之中。正如题目给出的答案所提示，把 $bt - as$ 当作一个符号来处理，代数运算只需寥寥数行。
- 写到“角 APB 是直角”便停笔。题目想要的几何意义是西姆松线：X、Y、N 是到三角形 OAB 三边所作垂线的垂足。

STEP 之外

西姆松线对任何三角形都成立，其综合几何证明（synthetic proof）用到圆内接四边形（cyclic quadrilateral）和同弓形内的圆周角相等。它的逆命题也成立，而当 P 沿外接圆移动时，各条西姆松线的包络（envelope）是一条三尖瓣线（deltoid，即施泰纳内摆线，Steiner's hypocycloid）。坐标法和复数法令这类结果变得按部就班便可推出；Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）展示了复数如何优雅地处理圆的几何，而 Coxeter 与 Greitzer 的 *Geometry Revisited*（《几何学的新探索》）则涵盖西姆松线以及更多经典的三角形几何。

详细解答

STEP III 2003，第 8 题

纯数学 · 一阶微分方程

STEP III 2003，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/03-s3-q8

关键思路

第 (i) 部分是隐函数求导（implicit differentiation），但它其实是一个技巧的示范：曲线族 $(y+ax)^n(y+bx)=c$ 的斜率（gradient）是 x 和 y 的两个线性式之比。第 (ii) 部分则把这个技巧倒过来用。我们对一般的曲线族求导，读出以 a、b 和 n 表示的斜率，再选取这些数值，使它与题目给出的方程吻合。一旦吻合，这个式子沿每一条解曲线（solution curve）都是常数，而这个常数关系就是通解（general solution）。

解答

第 (i) 部分

利用乘积法则（product rule）和链式法则（chain rule），把 $(y+2x)^3(y-4x)=c$ 对 x 求导，并以 y' 表示 $\frac{\mathrm{d}y}{\mathrm{d}x}$：

$$3(y+2x)^2(y'+2)(y-4x)+(y+2x)^3(y'-4)=0.$$

在 $y+2x\neq 0$ 的点上，我们可以除以 $(y+2x)^2$：

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

把含 y' 的项合并，

$$y'\left[3(y-4x)+(y+2x)\right]=-6(y-4x)+4(y+2x),$$
$$y'(4y-10x)=32x-2y,$$

再把分子和分母同除以 2，

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{16x-y}{2y-5x}.$$

这正是所求。（在直线 $y=-2x$ 上，曲线的 $c=0$，而公式在那里仍然给出正确的斜率 -2。）

第 (ii) 部分

设 $F=(y+ax)^n(y+bx)$，其中 y 是 x 的函数。与第 (i) 部分完全一样，

$$\begin{aligned}\frac{\mathrm{d}F}{\mathrm{d}x}&=n(y+ax)^{n-1}(y'+a)(y+bx)+(y+ax)^n(y'+b)\\&=(y+ax)^{n-1}\Big[y'\big((n+1)y+(nb+a)x\big)+(na+b)y+(n+1)ab\,x\Big].\end{aligned}$$

因此，在 $y=-ax$ 以外，F 沿一条曲线为常数，当且仅当

$$y'=-\frac{(na+b)y+(n+1)ab\,x}{(n+1)y+(nb+a)x}. \qquad (*)$$

（第 (i) 部分就是 $a=2$、$b=-4$、$n=3$ 的情况，得出 $(32x-2y)/(4y-10x)$。）

> ? **想一想** 一个分数可以有很多种写法，那么我们怎样把 $(*)$ 与 $\dfrac{10x-4y}{3x-y}$ 配对起来?

分子和分母同乘一个数，分数的值不变，所以我们只能在相差一个公因子的意义下把两者配对。把目标写成 $-(4y-10x)/(y-3x)$，再把分子和分母同乘 $k=n+1$，使两个分母中 y 的系数相同。于是我们需要

$$na+b=-4(n+1),\qquad (n+1)ab=10(n+1),\qquad nb+a=-3(n+1).$$

由于 $n+1\neq 0$（若 $n=-1$，$(*)$ 的分母便没有 y 项），中间的方程给出 $ab=10$。把首尾两个方程相加，得 $(n+1)(a+b)=-7(n+1)$，所以 $a+b=-7$。因此 a 和 b 是 $t^2+7t+10=0$ 的根，即 -2 和 -5。把首尾两个方程相减，得

$$(n-1)(a-b)=-(n+1).$$

- 取 $a=-5$、$b=-2$：$-3(n-1)=-(n+1)$，所以 $n=2$。
- 取 $a=-2$、$b=-5$：$3(n-1)=-(n+1)$，所以 $n=\frac{1}{2}$。

两种选择描述的是同一批曲线，因为把 $(y-2x)^{1/2}(y-5x)=c$ 两边平方便得 $(y-5x)^2(y-2x)=c^2$。我们取 $n=2$，这样可以避开平方根。

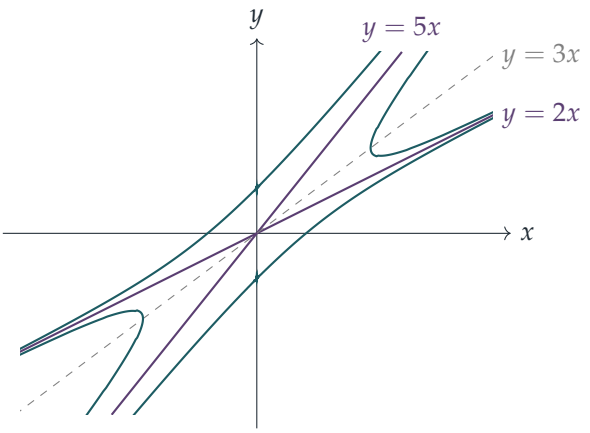
为稳妥起见，我们直接验证。设 $G=(y-5x)^2(y-2x)$，

$$\begin{aligned}\frac{\mathrm{d}G}{\mathrm{d}x}&=(y-5x)\big[2(y'-5)(y-2x)+(y-5x)(y'-2)\big]\\&=(y-5x)\big[3y'(y-3x)-12y+30x\big]\\&=3(y-5x)\big[y'(y-3x)-(4y-10x)\big].\end{aligned}$$

若 y 满足该微分方程（differential equation），则 $y'(y-3x)=4y-10x$，所以方括号内的式子等于零，$\mathrm{d}G/\mathrm{d}x=0$。因此 G 在每一条解曲线上都是常数，通解为

$$\boxed{(y-5x)^2(y-2x)=C}$$

其中 C 为任意常数。$C=0$ 给出直线 $y=5x$ 和 $y=2x$。两者确实都是解：在 $y=2x$ 上，右边是 $2x/x=2$；在 $y=5x$ 上，右边是 $(-10x)/(-2x)=5$。所有曲线在与直线 $y=3x$ 相交之处都有铅垂切线，而方程在该直线上没有定义。



曲线 $(y-5x)^2(y-2x)=\pm1$（蓝绿色）和 $C=0$ 的直线解（紫红色）。位于两条直线之间的分支，在与虚线 $y=3x$ 相交之处有铅垂切线。

第 (ii) 部分的另一种解法

右边只取决于 y/x，所以代换 $y=vx$ 同样可行。它给出

$$x\frac{\mathrm{d}v}{\mathrm{d}x}=\frac{(v-2)(v-5)}{3-v},$$

分离变量（separating the variables）并运用部分分式（partial fractions），可得 $\ln|x|=-\frac{1}{3}\ln|v-2|-\frac{2}{3}\ln|v-5|+$ 常数，整理后得到相同的答案。题目所用的方法则无须用到部分分式。

验算

- 第 (i) 部分本身就是对 $(*)$ 的一次验算：把 $a=2$、$b=-4$、$n=3$ 代入，便重现该处求得的斜率。
- 在点 $(1,0)$，方程给出斜率 $10/3$。在该点对 $(y-5x)^2(y-2x)=C$ 作隐函数求导，得 $(-5)\big[2(y'-5)(-2)+(-5)(y'-2)\big]=0$，所以 $-9y'+30=0$，$y'=10/3$。

- 数值方面，我们用龙格－库塔法（Runge–Kutta method）从五十个随机起点逐步求解该微分方程。沿每一条路径，$(y-5x)^2(y-2x)$ 的值都保持不变，误差约为 10^{-13}。

常见错误

- 不考虑公因子，直接把 $(*)$ 的系数与 $10x-4y$ 和 $3x-y$ 配对，结果得出互相矛盾的方程。
- 把答案留作 $(y-2x)^{1/2}(y-5x)=c$，而没有留意到它其实是同一个曲线族；而且除非小心处理正负号，否则还会漏掉 $y<2x$ 的解。
- 忘记 $C=0$ 是容许的，它给出直线解 $y=2x$ 和 $y=5x$。

STEP 之外

这里的方程是*齐次*（homogeneous）的：右边只取决于 y/x。寻找一个沿着解保持不变的函数 $G(x,y)$，正是*恰当方程*（exact equations）和*积分因子*（integrating factors）背后的想法。在力学中，同样的想法就是寻找能量之类的守恒量（conserved quantities）。Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程）在开首几章便涵盖了这两个课题。

详细解答

STEP III 2003，第 9 题

力学 • 平衡的稳定性

STEP III 2003，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/03-s3-q9

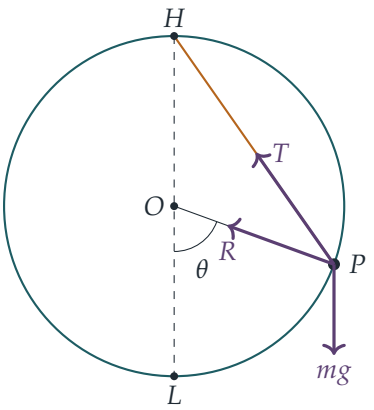
关键思路

金属线是光滑的，所以它的反作用力（reaction）总是沿半径方向，并不作功。因此一切都发生在切向（tangential）方向：把张力（tension）和重量沿切线分解，便得到一条关于 $\ddot{\theta}$ 的方程，平衡（equilibrium）位置和微小振动（small oscillations）都可由它求出。等腰三角形 OHP 的几何性质提供了所需的全部角度。

解答

建立模型

把 P 看成光滑金属线上的质点（particle），所以金属线的反作用力 R 沿 PO 作用。绳子是轻绳，并服从胡克定律（Hooke's law）。取 $0 < \theta < \pi$，即 P 位于垂直直径 LH 的其中一侧。



由于 $OH = OP = a$，三角形 OHP 是等腰三角形，且 $\angle HOP = \pi - \theta$，所以它的两个底角为

$$\angle OHP = \angle OPH = \tfrac{1}{2}\theta, \qquad HP = 2a\cos\tfrac{1}{2}\theta.$$

当 $HP > a$，即 $\cos\frac{1}{2}\theta > \frac{1}{2}$ 时，绳子处于绷紧（taut）状态，这时由胡克定律得

$$T = \frac{\alpha mg}{a}\left(2a\cos\tfrac{1}{2}\theta - a\right) = \alpha mg\left(2\cos\tfrac{1}{2}\theta - 1\right).$$

现在沿 P 点的切线分解，取 θ 增加的方向（即朝向 H）为正。绳子与 PO 成 $\frac{1}{2}\theta$ 角，所以它与这条切线成 $\frac{1}{2}\pi - \frac{1}{2}\theta$ 角，其切向分量为 $T\sin\frac{1}{2}\theta$。重量的切向分量为 $-mg\sin\theta$，而 R 没有切向分量。切向加速度为 $a\ddot{\theta}$，故

$$\begin{aligned} ma\ddot{\theta} &= T\sin\tfrac{1}{2}\theta - mg\sin\theta \\ &= mg\left[\alpha\left(2\cos\tfrac{1}{2}\theta - 1\right)\sin\tfrac{1}{2}\theta - 2\sin\tfrac{1}{2}\theta\cos\tfrac{1}{2}\theta\right], \end{aligned}$$

所以，只要绳子绷紧，

$$\ddot{\theta} = \frac{g}{a}\sin\tfrac{1}{2}\theta\left[2(\alpha-1)\cos\tfrac{1}{2}\theta - \alpha\right]. \qquad (*)$$

（把能量 $\frac{1}{2}ma^2\dot{\theta}^2 - mga\cos\theta + \frac{1}{2}\alpha mga(2\cos\frac{1}{2}\theta - 1)^2$ 求导，也会得到同一条方程。）

平衡位置

平衡时 $\ddot{\theta} = 0$。当 $0 < \theta < \pi$ 时，$\sin\frac{1}{2}\theta \neq 0$，所以我们需要

$$\cos\tfrac{1}{2}\theta = \frac{\alpha}{2(\alpha-1)},$$

由于 $\alpha > 1$，这个式子是有意义的。要由此得到 $0 < \theta < \pi$，我们需要 $0 < \cos\frac{1}{2}\theta < 1$。这个值是正数，而

$$\frac{\alpha}{2(\alpha-1)} < 1 \iff \alpha < 2\alpha - 2 \iff \alpha > 2.$$

我们还需要绳子在该处绷紧，因为 $(*)$ 是在这个假设下得出的：$\dfrac{\alpha}{2(\alpha-1)} > \dfrac{1}{2} \iff \alpha > \alpha - 1$，这恒成立。所以若 $\alpha > 2$，便存在一个满足 $0 < \theta < \pi$ 的平衡位置，正如题目所求。

? **想一想** 假如绳子在平衡位置是松弛（slack）的，会怎样?

那样就只有重力有切向分量，而 $-mg\sin\theta = 0$ 在 $0 < \theta < \pi$ 内无解。这就是为甚么我们求得的平衡位置自动满足 $\cos\frac{1}{2}\theta > \frac{1}{2}$ 这一点如此重要。

当 $\alpha = 2 + \sqrt{2}$ 时的微小振动

这时

$$\cos\tfrac{1}{2}\theta = \frac{2+\sqrt{2}}{2(1+\sqrt{2})} = \frac{\sqrt{2}(\sqrt{2}+1)}{2(\sqrt{2}+1)} = \frac{1}{\sqrt{2}},$$

所以平衡位置在 $\frac{1}{2}\theta = \frac{1}{4}\pi$，即 $\theta = \frac{1}{2}\pi$，此时 P 与 O 处于同一水平。设 $\theta = \frac{1}{2}\pi + \phi$，则 $\ddot{\theta} = \ddot{\phi}$。把 $(*)$ 写成 $\ddot{\theta} = \frac{g}{a} f(\theta)$，其中 $f(\frac{1}{2}\pi) = 0$。当 ϕ 很小时，

$$f(\tfrac{1}{2}\pi + \phi) \approx f'(\tfrac{1}{2}\pi)\,\phi.$$

由于 $(*)$ 中的方括号在 $\theta = \frac{1}{2}\pi$ 时等于零，使用乘积法则（product rule）后只剩下一项：

$$f'(\tfrac{1}{2}\pi) = \sin\tfrac{1}{4}\pi \cdot \left[-2(\alpha-1)\cdot\tfrac{1}{2}\sin\tfrac{1}{4}\pi\right] = -(\alpha-1)\sin^2\tfrac{1}{4}\pi = -\frac{\alpha-1}{2} = -\frac{\sqrt{2}+1}{2}.$$

因此，当 ϕ 很小时，

$$\ddot{\phi} \approx -\frac{g(\sqrt{2}+1)}{2a}\,\phi$$

正如题目所求。

最后部分

ϕ 的系数是负数，所以质点受到微小位移后，会围绕 $\theta = \frac{1}{2}\pi$ 作（近似的）简谐运动（simple harmonic motion）：这个平衡是稳定（stable）的。角频率（angular frequency）为 $\omega = \sqrt{g(\sqrt{2}+1)/(2a)}$，所以周期（period）为

$$\frac{2\pi}{\omega} = 2\pi\sqrt{\frac{2a}{(\sqrt{2}+1)g}} = 2\pi\sqrt{\frac{2(\sqrt{2}-1)a}{g}},$$

而振幅（amplitude）会一直保持很小，因为能量守恒（金属线光滑，绳子具弹性，所以没有任何东西耗散能量）。

验算

- 当 $\alpha = 2$ 时，平衡位置的公式给出 $\cos\frac{1}{2}\theta = 1$，即 $\theta = 0$：平衡位置与最低点重合，这与条件 $\alpha > 2$ 吻合。
- 当 $\alpha \to \infty$ 时，$\cos\frac{1}{2}\theta \to \frac{1}{2}$，所以 $\theta \to \frac{2}{3}\pi$ 而 $HP \to a$：非常硬的绳子几乎不会伸长。
- 取 $a = 1.3\,\mathrm{m}$ 和 $g = 9.81\,\mathrm{m\,s^{-2}}$，由 $\phi = 0.01$ 开始对完整的方程 $(*)$ 作数值逐步计算，得到周期约为 $2.0819\,\mathrm{s}$，与 $2\pi\sqrt{2(\sqrt{2}-1)a/g} \approx 2.0818\,\mathrm{s}$ 相符。

常见错误

- 使用 $HP = 2a\sin\frac{1}{2}\theta$：那是弦 LP 的长度，而不是 HP。三角形 OHP 中位于 O 的角是 $\pi - \theta$。
- 忘记检查绳子在平衡位置是否绷紧，因为松弛的绳子没有张力。

- 沿径向而不是沿切向分解：径向方程会引入未知的反作用力 R，而且完全没有提供关于 $\ddot{\theta}$ 的信息。

STEP 之外

在平衡位置附近作线性化（linearising），$\ddot{\phi} \approx f'(\theta_0)\,\phi$，是研究任何动力系统（dynamical system）稳定性的标准第一步：$f'(\theta_0)$ 的正负号（等价地，就势能（potential energy）而言，是 $V''(\theta_0)$ 的正负号）决定了系统是振动还是一去不返。在大学一年级的课程中，这会发展成非线性常微分方程（nonlinear ODEs）的相平面（phase plane）分析。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）通过势能来处理稳定平衡附近的微小振动，讲解得十分清楚。

详细解答

STEP III 2003，第 10 题

力学 • 变加速度

STEP III 2003，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/03-s3-q10

关键思路

加速度 $kx\dot{x}$ 是一个对时间的恰当导数（exact derivative）：它就是 $\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}kx^2\right)$。因此我们可以立即积分一次，得到一阶方程（first-order equation）$\dot{x} = \frac{1}{2}kx^2 + C$，而这个方程可以分离变量。常数 C 的正负决定一切：常数为正时得到反正切（arctangent）（解会爆破），常数为负时得到部分分式（partial fractions）和对数（解会稳定下来）。

解答

我们把质点看成在 x 轴上运动，并且 $\ddot{x} = kx\dot{x}$。由于 $\frac{\mathrm{d}}{\mathrm{d}t}(x^2) = 2x\dot{x}$，运动方程为

$$\frac{\mathrm{d}\dot{x}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\tfrac{1}{2}kx^2\right), \qquad \text{所以} \qquad \dot{x} = \tfrac{1}{2}kx^2 + C$$

其中 C 为常数。当 $t = 0$ 时，$x = d$ 且 $\dot{x} = U$，所以 $C = U - \frac{1}{2}kd^2$，于是

$$\dot{x} = \tfrac{1}{2}k\left(x^2 - d^2\right) + U. \qquad (*)$$

第 (i) 部分

当 $U = kd^2$ 时，方程 $(*)$ 变成

$$\dot{x} = \tfrac{1}{2}k\left(x^2 + d^2\right).$$

右边恒为正，所以我们可以分离变量（separate variables）：

$$\int_d^x \frac{\mathrm{d}u}{u^2 + d^2} = \int_0^t \tfrac{1}{2}k\,\mathrm{d}s \quad \Longrightarrow \quad \frac{1}{d}\left[\arctan\frac{x}{d} - \arctan 1\right] = \tfrac{1}{2}kt.$$

由于 $\arctan 1 = \frac{\pi}{4}$，由此得 $\arctan(x/d) = \frac{1}{2}dkt + \frac{\pi}{4}$，所以

$$x = d\tan\left(\frac{dkt}{2} + \frac{\pi}{4}\right).$$

只要 tan 的辐角保持小于 $\frac{\pi}{2}$，这个式子就成立（反正切的取值范围是 $(-\frac{\pi}{2}, \frac{\pi}{2})$）。当 t 由 0 开始增加时，辐角由 $\frac{\pi}{4}$ 开始增加，并在以下时刻达到 $\frac{\pi}{2}$：

$$\frac{dkt}{2} = \frac{\pi}{4}, \qquad \text{即} \qquad t = \frac{\pi}{2dk}.$$

由于 $\tan\theta \to +\infty$ 当 $\theta \to \frac{\pi}{2}$ 从下方趋近时成立，而且 $d > 0$，所以当 $t \to \frac{\pi}{2dk}$ 时，$x \to \infty$，这正是所要证明的。质点在有限时间内到达无穷远。

第 (ii) 部分

现在 $U < 0$，所以 $C = U - \frac{1}{2}kd^2 < 0$。设 $C = -\frac{1}{2}ka^2$，其中 $a > 0$，即

$$a = \sqrt{d^2 - \frac{2U}{k}}, \qquad \text{使得} \qquad \dot{x} = \tfrac{1}{2}k\left(x^2 - a^2\right).$$

因为 $U < 0$ 且 $k > 0$，所以 $a > d > 0$。

> **想一想** 部分分式所得的对数，应该用哪一个：$\ln\frac{x-a}{x+a}$ 还是 $\ln\frac{a-x}{a+x}$？

这取决于 x 落在哪个范围。常数解 $x = \pm a$ 正是 $\dot{x} = 0$ 之处。起初 $x = d$ 严格地位于 $-a$ 与 a 之间，而由该处出发的解不可能越过常数解 $x = \pm a$（这类一阶方程的两个解不可能相遇），所以在任何时刻都有 $-a < x < a$。在这个范围内 $\dot{x} < 0$，所以 x 递减，而 $a - x$ 和 $a + x$ 这两个量皆为正。

利用 $\dfrac{1}{x^2 - a^2} = \dfrac{1}{2a}\left(\dfrac{1}{x-a} - \dfrac{1}{x+a}\right)$ 并分离变量，

$$\frac{1}{2a}\left[\ln\frac{a-u}{a+u}\right]_d^x = \tfrac{1}{2}kt \quad \Longrightarrow \quad \frac{a-x}{a+x} = A\,\mathrm{e}^{akt}, \qquad \text{其中} A = \frac{a-d}{a+d}.$$

注意 $0 < A < 1$，因为 $a > d > 0$。解出 x，

$$x = a\,\frac{1 - A\mathrm{e}^{akt}}{1 + A\mathrm{e}^{akt}}, \qquad a = \sqrt{d^2 - \frac{2U}{k}}, \quad A = \frac{a-d}{a+d}.$$

这个式子对所有 $t \geq 0$ 都有定义（分母为正）。等价地，由于 $\ln A = -2\,\mathrm{artanh}(d/a)$，它可以写成 $x = a\tanh\left(\mathrm{artanh}(d/a) - \frac{1}{2}akt\right)$。

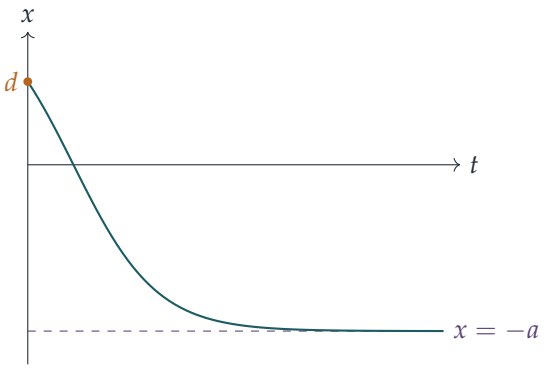
当 $t \to \infty$ 时，$A\mathrm{e}^{akt} \to \infty$，因为 $A > 0$ 且 $ak > 0$。把分子和分母同时除以 $A\mathrm{e}^{akt}$，

$$x = a\,\frac{A^{-1}\mathrm{e}^{-akt} - 1}{A^{-1}\mathrm{e}^{-akt} + 1} \longrightarrow -a.$$

所以 x 趋向极限（limit）

$$-a = -\sqrt{d^2 - \frac{2U}{k}}\,.$$

（这个极限也与 k 有关；它不能只用 d 和 U 来表示，因为固定 d 和 U 而改变 k，极限也会随之改变。）



图像（以 $d=1$、$a=2$、$k=1$ 绘画）显示质点由 d 出发，初速度为负，经过原点后越来越慢地趋近 $-a$。

验算

- 在第 (i) 部分，$t=0$ 给出 $x=d\tan\frac{\pi}{4}=d$，而求导得 $\dot{x}=\frac{1}{2}kd^2\sec^2\frac{\pi}{4}=kd^2=U$。
- 在第 (ii) 部分，$t=0$ 给出 $x=a\frac{1-A}{1+A}=d$。当 $U\to 0^-$ 时，$a\to d$ 且 $A\to 0$，所以 x 一直停留在 d：这是正确的，因为 $x=d$、$\dot{x}=0$ 会令加速度为零。
- 数值上，以龙格－库塔法（Runge–Kutta method）对随机选取的 k、d 和 U 逐步求解 $\ddot{x}=kx\dot{x}$，结果与两条公式吻合，误差小于 10^{-6}。例如，取 $k=1$、$d=1$、$U=-1.5$ 时有 $a=2$，而数值计算所得的 x 稳定于 -2.000000。

常见错误

- 使用 $\ln\frac{x-a}{x+a}$ 而不加绝对值：当 $|x|<a$ 时，这是负数的对数。
- 求极限时取了正平方根：x 由 d 开始递减，所以极限是 $-a$，而不是 $+a$。
- 把加速度写成 $v\frac{\mathrm{d}v}{\mathrm{d}x}$，得到 $v\frac{\mathrm{d}v}{\mathrm{d}x}=kxv$，然后除以 v 而没有指出 $v\neq 0$。这在本题行得通，但直接对 $\frac{\mathrm{d}}{\mathrm{d}t}$ 积分便可避开这个问题。

STEP 之外

第 (i) 部分是*有限时间爆破*（finite-time blow-up）的一个小例子：对于方程 $\dot{x}=f(x)$，若 f 的增长快于线性（这里是二次增长），解便可能在有限时间内趋向无穷，所以解不一定对所有 t 都存在。第 (ii) 部分展示了另一种情况：常数解 $x=\pm a$ 是平衡点（equilibria），$-a$ 是稳定的而 $+a$ 是不稳定的，而且相线（phase line）图在未作任何积分之前已能预测

极限。这两个概念，连同阻止解互相穿越的唯一性定理（uniqueness theorem），在 Boyce and DiPrima, *Elementary Differential Equations*（常微分方程入门教科书）中都有论述。

详细解答

STEP III 2003，第 11 题

力学 · 拦截运动学

> **STEP III 2003，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/03-s3-q11

关键思路

暂且放下运动过程，只考虑碰撞会发生的那一点。在那一刻，A、B 与碰撞点构成一个三角形，三边分别是 d、P 走过的距离和 Q 走过的距离，而 B 处的角由方位角（bearing）决定。于是余弦定理（cosine rule）给出一条方程，再把两段走过的距离都用时间表示，便得到一条二次方程（quadratic）。能否发生碰撞，就变成这条二次方程有没有正根（positive root）的问题。

解答

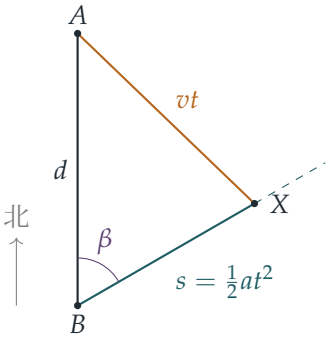
我们把 P 和 Q 视为在水平面上运动的质点（particle）。方位角由正北起顺时针量度，且 $v > 0$。

碰撞方程

假设 Q 在时间 $t > 0$ 于点 X 击中 P。到那时，P 已由静止出发，以恒定加速度 a 沿一条固定直线运动，所以

$$BX = s = \tfrac{1}{2}at^2,$$

而 Q 以恒定速率沿直线运动，所以 $AX = vt$。



由 B 到 A 的方向是正北，而由 B 到 X 的方向的方位角为 β，所以角 ABX 等于 β（若 $\beta > 180°$，则为 $360° - \beta$，两者的余弦相同）。在三角形 ABX 中应用余弦定理，得

$$v^2t^2 = d^2 + s^2 - 2ds\cos\beta.$$

（若 $\beta = 0$ 或 $180°$，这个“三角形”会退化成一条线段，但方程依然成立：它变成 $(vt)^2 = (d \mp s)^2$。）由于 $t^2 = 2s/a$，左边等于 $2v^2s/a$，整理后得

$$s^2 - 2\left(d\cos\beta + \frac{v^2}{a}\right)s + d^2 = 0. \qquad (*)$$

反过来，若 $(*)$ 有一个根 $s > 0$，设 $t = \sqrt{2s/a}$，并设 X 为 P 在该时刻的位置。那么 $(*)$ 所说的正是 $AX = vt$，而且 $AX > 0$。因此，若把 Q 由 A 朝着 X 射出，它会在时间 t 到达 X，与 P 同时到达，并击中它。故此

$$\text{存在合适的方向} \iff (*)\text{ 有正根}s.$$

必要条件（necessary condition）

设 $k = d\cos\beta + v^2/a$，则 $(*)$ 就是 $s^2 - 2ks + d^2 = 0$。

> **想一想** 这条二次方程当然需要有实根，但这样就足够了吗？关于这些根，我们还要知道甚么？

这些根还必须是正数，因为 s 是走过的距离。两根之积是 $d^2 > 0$，所以两根同号，而两根之和是 $2k$。因此，方程有正根，当且仅当判别式（discriminant）非负且 $k > 0$：

$$k^2 \geq d^2 \quad \text{且} \quad k > 0 \iff k \geq d.$$

（若 $k > 0$，则 $k^2 \geq d^2$ 等同于 $k \geq d$，因为 $d > 0$；而 $k \geq d$ 必然导致 $k > 0$。）

所以，若 Q 能击中 P，则

$$d\cos\beta + \frac{v^2}{a} \geq d, \qquad \text{即} \qquad v^2 \geq ad(1 - \cos\beta),$$

这正是所要求的结果。

在 P 移动距离 d 之前击中

现在假设 $v^2 > ad(1-\cos\beta)$，于是 $k > d > 0$。这时判别式 $4(k^2-d^2)$ 严格为正，$(*)$ 有两个相异的根 $s_1 < s_2$。两根皆为正数，因为两根之和 $2k$ 与两根之积 d^2 都是正数。由于两根相异，且

$$s_1 s_2 = d^2,$$

它们不可能同时不小于 d，也不可能同时不大于 d，所以

$$s_1 < d < s_2.$$

明确地写出来，$s_1 = k - \sqrt{k^2-d^2}$。

根据上文的逆向论证（converse），把 Q 朝着点 X_1 射出，其中该点与 B 相距 s_1 并位于 P 的路径上，则 Q 会在时间 $t_1 = \sqrt{2s_1/a}$ 击中 P，而在那一刻 P 只移动了 $s_1 < d$。所以 Q 可以瞄准某个方向，在 P 移动距离 d 之前击中 P，这就是所要证明的。

验算

- 取 $\beta = 0$：P 笔直地朝 A 走去，所以无论 v 的值是多少，两者必定相遇。条件给出 $v^2 \geq 0$，这恒成立。这时 $(*)$ 变成 $s^2 - 2(d+v^2/a)s + d^2 = 0$，而较小的根小于 d：两者在 P 到达 A 之前相遇，与常理相符。
- 取 $a = d = 1$、$\beta = 90^\circ$（即 P 向正东走），以及 $v^2 = 2 > 1 = ad(1-\cos\beta)$。这时 $(*)$ 是 $s^2 - 4s + 1 = 0$，所以 $s_1 = 2 - \sqrt{3} \approx 0.268 < 1$。在 $t_1 = \sqrt{2s_1}$ 时，我们有 $s_1^2 + 1 = 2t_1^2$，所以碰撞点与 A 的距离确实是 vt_1。我们又对 a、d、β 和 v 的大量随机取值作了数值检验，既确认了在 $s_1 < d$ 处击中，也确认了当 $v^2 < ad(1-\cos\beta)$ 时，P 与 A 的距离始终大于 vt。
- 在临界情况 $v^2 = ad(1-\cos\beta)$ 下，我们得到 $k = d$ 和重根（double root）$s = d$：唯一可能的击中时刻，恰好是 P 走了 d 的时候，这就是最后一部分需要严格不等式的原因。

常见错误

- 只要求判别式非负。这样得到的是 $|k| \geq d$，它同时容许 $k \leq -d$，而这时两根皆为负数，不会发生碰撞。
- 在余弦定理中用错了角。B 处的角是 BA（正北）与 BX（方位角 β）之间的夹角，所以它就是 β 本身，而不是 $180^\circ - \beta$。
- 在最后一部分只证明了必要条件。你还必须证明朝着 X_1 瞄准确实可行，这正是那个逆向论证。

STEP 之外

这是一道追踪与拦截问题。这里目标的路径是预先知道的，而拦截者沿直线运动，所以问题归结为一条关于相遇时间的方程。导弹制导和“追逐曲线”（pursuit curves）则不同：追逐者不断转向目标，结果引出的是微分方程（differential equations）。先定下相遇点，再问谁能及时赶到，这种习惯是大学一年级力学的标准工具。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）对相对运动的运动学（kinematics）有细致的阐述。

详细解答

STEP III 2003，第 12 题

概率与统计 · 指数分布的等候时间

STEP III 2003，第 12 题

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

step.maths.org/questions/03-s3-q12

关键思路

如果任何一段时间内的打扰次数都服从泊松分布（Poisson distribution），那么“下一次打扰在 t 分钟之后才出现”与“在接下来的 t 分钟内完全没有打扰”是同一个事件，这样就把离散分布（discrete distribution）变成连续分布（continuous distribution）：等候时间服从指数分布（exponential distribution）。其后，相继两次打扰之间的各个间隔互相独立（independent），所以寻找一个长间隔的过程，就是一连串独立试验，每次试验不是“成功”（间隔达半小时或以上）就是“失败”。余下的工作交给几何分布（geometric distribution）和条件期望（conditional expectation）便可。

解答

T 的分布

打扰平均每十分钟出现一次，所以发生率是每分钟 $\frac{1}{10}$ 次，而在长 t 分钟的时段内，打扰次数服从平均值为 $\frac{t}{10}$ 的泊松分布。

从刚刚发生的那次打扰开始计时。到下一次打扰的时间 T 不少于 t，当且仅当在接下来的 t 分钟内没有打扰，所以

$$\mathrm{P}(T \geq t) = \mathrm{P}(\text{没有任何打扰，为时} t) = \mathrm{e}^{-t/10}.$$

因此，对于 $t \geq 0$，工作时间少于 t 分钟的概率是

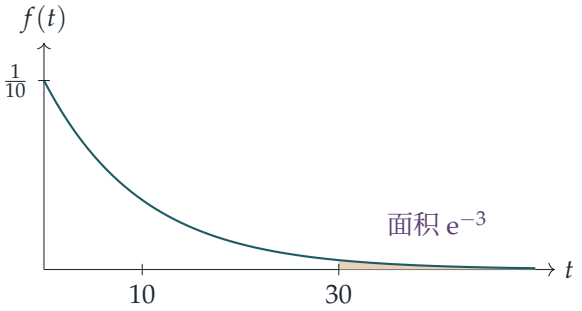
$$\mathrm{P}(T < t) = 1 - \mathrm{e}^{-t/10}.$$

这就是 T 的累积分布函数（cumulative distribution function）$F(t)$（当 $t < 0$ 时 $F(t) = 0$）。

求导，得

$$f(t) = F'(t) = \tfrac{1}{10}\,\mathrm{e}^{-t/10} \quad (t \geq 0), \qquad f(t) = 0 \quad (t < 0).$$

所以 T 服从平均值为 10 分钟的指数分布。



打扰次数的期望值

由于这个过程没有记忆（互不重叠的时段内的泊松计数互相独立），每当一次打扰发生，到下一次打扰的时间都重新开始计算，其分布与 T 相同，并且独立于之前发生的一切。所以相继的间隔 $T_1, T_2, \ldots$ 互相独立，各自的分布都与 T 相同。

如果一个间隔不少于 30 分钟，就称它为长间隔。每个间隔是长间隔的概率为

$$p = \mathrm{P}(T \geq 30) = \mathrm{e}^{-30/10} = \mathrm{e}^{-3},$$

而且各个间隔互相独立。设 N 为第一个长间隔出现之前的打扰次数（刚刚发生的那一次不计算在内）。每个短间隔都以一次打扰告终，而第一个长间隔就是我那不受打扰的半小时，所以 N 就是第一个长间隔之前的短间隔数目。因此

$$\mathrm{P}(N = n) = (1-p)^n p, \qquad n = 0, 1, 2, \ldots$$

想一想 这个几何分布由 0 而不是由 1 开始，我们怎样求它的平均值?

对几何级数（geometric series）$\sum_{n\geq 0} x^n = \frac{1}{1-x}$ 求导，可得当 $|x| < 1$ 时 $\sum_{n\geq 1} n x^{n-1} = \frac{1}{(1-x)^2}$。取 $x = 1-p$，

$$\mathrm{E}(N) = \sum_{n=1}^{\infty} n(1-p)^n p = p(1-p)\cdot\frac{1}{p^2} = \frac{1-p}{p}.$$

（另一个做法：$N+1$ 是直至首次成功为止（包括该次）的试验次数，其平均值为 $\frac{1}{p}$。）代入 $p = \mathrm{e}^{-3}$，

$$\mathrm{E}(N) = \frac{1-\mathrm{e}^{-3}}{\mathrm{e}^{-3}} = \mathrm{e}^3 - 1,$$

正是所要求的结果。

短间隔长度的期望值

我们要求 $\mathrm{E}(T \mid T < 30)$。已知 $T < 30$ 时，T 的条件概率密度函数（conditional density）在 $0 \le t < 30$ 上是 $f(t)/\mathrm{P}(T < 30)$，所以

$$\mathrm{E}(T \mid T < 30) = \frac{1}{1 - \mathrm{e}^{-3}} \int_0^{30} \frac{t}{10}\, \mathrm{e}^{-t/10}\, \mathrm{d}t.$$

用分部积分法（integration by parts），取 $u = t$ 和 $\mathrm{d}v = \frac{1}{10}\mathrm{e}^{-t/10}\,\mathrm{d}t$，于是 $v = -\mathrm{e}^{-t/10}$，

$$\begin{aligned}\int_0^{30} \frac{t}{10}\, \mathrm{e}^{-t/10}\, \mathrm{d}t &= \Big[-t\, \mathrm{e}^{-t/10} \Big]_0^{30} + \int_0^{30} \mathrm{e}^{-t/10}\, \mathrm{d}t \\ &= -30\, \mathrm{e}^{-3} + 10\big(1 - \mathrm{e}^{-3}\big) \\ &= 10 - 40\, \mathrm{e}^{-3}.\end{aligned}$$

因此，把分子和分母同乘以 e^3，

$$\mathrm{E}(T \mid T < 30) = \frac{10 - 40\, \mathrm{e}^{-3}}{1 - \mathrm{e}^{-3}} = \frac{10\, \mathrm{e}^3 - 40}{\mathrm{e}^3 - 1} \text{ 分钟} \approx 8.43 \text{ 分钟}.$$

这个数值略小于 10，是合理的：我们已经把长间隔剔除了。

等候时间的期望值

在不受打扰的半小时开始之前，等候时间就是 N 个短间隔之和。已知 $N = n$ 时，这是 n 个独立的间隔，每个都以短于 30 分钟为条件，所以已知 $N = n$ 时等候时间的期望值是 $n\,\mathrm{E}(T \mid T < 30)$。再对 N 取平均，

$$\mathrm{E}(\text{等候时间}) = \mathrm{E}(N)\, \mathrm{E}(T \mid T < 30) = (\mathrm{e}^3 - 1) \cdot \frac{10\, \mathrm{e}^3 - 40}{\mathrm{e}^3 - 1} = 10\, \mathrm{e}^3 - 40 \text{ 分钟},$$

即大约 161 分钟，或 2 小时 41 分钟。（如果想求直至试卷完成为止的时间，就要加上那半小时本身：$10\, \mathrm{e}^3 - 10 \approx 191$ 分钟。）

验算

- $\int_0^{\infty} \frac{1}{10}\mathrm{e}^{-t/10}\,\mathrm{d}t = 1$，所以 f 确实是一个概率密度函数，而它的平均值是 10，与“平均每十分钟一次”吻合。
- 在没有条件的情况下，$\mathrm{E}(T) = 10$ 应该可以分拆为 $\mathrm{P}(T < 30)\,\mathrm{E}(T \mid T < 30) + \mathrm{P}(T \ge 30)\,\mathrm{E}(T \mid T \ge 30)$。由无记忆性（memorylessness）可知 $\mathrm{E}(T \mid T \ge 30) = 40$，而果然有 $(10 - 40\mathrm{e}^{-3}) + 40\mathrm{e}^{-3} = 10$。
- 把这个过程模拟运行 200 000 次，得出平均打扰次数为 19.1（对照 $\mathrm{e}^3 - 1 \approx 19.09$），平均等候时间为 161.1 分钟（对照 $10\mathrm{e}^3 - 40 \approx 160.86$）。

常见错误

- 使用 $\mathrm{E}(N) = \frac{1}{p} = \mathrm{e}^3$：这样连长间隔也计算在内，但长间隔结束时，那半小时之前并没有再出现打扰。
- 把 $\mathrm{E}(N)$ 乘以无条件平均值 10 来求等候时间。组成等候时间的间隔已知是短间隔，所以它们的平均值是 $\mathrm{E}(T \mid T < 30)$，而不是 10。

STEP 之外

这些打扰构成一个泊松过程（*Poisson process*），而各个间隔是互相独立的指数分布变量这个结果，正是泊松过程的基本性质之一；最后一步 $\mathrm{E}(\text{共}N\text{ 项之和}) = \mathrm{E}(N) \times \mathrm{E}(\text{每项})$ 是一个取条件的论证，与沃尔德恒等式（*Wald's identity*）关系密切。两者合起来，通向更新理论（renewal theory）和排队模型（queueing models），这些都是大学一、二年级概率课程的内容。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）以合适的程度讲解泊松过程和条件期望。

详细解答

STEP III 2003，第 13 题

概率与统计 • 吸收马可夫链

STEP III 2003，第 13 题

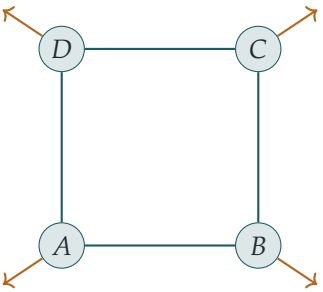
继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/03-s3-q13

关键思路

这是一种“首步分析”（first-step analysis）：以兔子下一步的行动作为条件，把从某个洞室出发的概率，用牠能到达的各个洞室出发的概率来表示。这样便得到四条线性方程，而正方形的对称性（对对角线 AC 作反射，B 与 D 互换）可以减省工夫。最后一部分，我们把同一技巧用于一个修改过的兔穴，当中进入 C 即算作失败，然后应用条件概率（conditional probability）的定义。

解答



通往地面的隧道

在整个解答中，每一步都以概率 $\frac{1}{3}$ 选出，并且与兔子过去的行踪独立（对于一只“发现自己身处”某个洞室的兔子，我们就是这样理解“可能性均等”的）。以 E_A 表示兔子最终经由 A 的隧道离开这一事件（event）。

第 (i) 部分

假设兔子身处 A，并以牠的下一步作为条件。

- 牠有 $\frac{1}{3}$ 的概率立即从 A 的隧道往上走，此时 E_A 已经发生。
- 牠有 $\frac{1}{3}$ 的概率跑到 B。这时牠身处 B，而由于牠往后的行动与牠如何到达该处无关，牠最终从 A 离开的概率是 p_B。
- 牠有 $\frac{1}{3}$ 的概率跑到 D，同理，此时的概率是 p_D。

根据全概率公式（law of total probability），

$$p_A = \tfrac{1}{3}\cdot 1 + \tfrac{1}{3}p_B + \tfrac{1}{3}p_D,$$

正如所求。

第 (ii) 部分

从其余洞室出发，用同样的论证可再得三条方程。唯一的分别是：经由 B、C 或 D 的隧道离开，即表示 E_A 不发生，所以该分支的贡献是 $\frac{1}{3}\cdot 0$：

$$\begin{aligned} p_B &= \tfrac{1}{3}p_A + \tfrac{1}{3}p_C, \\ p_C &= \tfrac{1}{3}p_B + \tfrac{1}{3}p_D, \\ p_D &= \tfrac{1}{3}p_A + \tfrac{1}{3}p_C. \end{aligned}$$

p_B 与 p_D 的方程右边相同，所以 $p_B = p_D$；把这个共同的值记作 b。（这就是对 AC 的反射对称。）于是

$$p_A = \tfrac{1}{3} + \tfrac{2}{3}b, \qquad p_C = \tfrac{2}{3}b,$$

代入 p_B 的方程，得

$$b = \tfrac{1}{3}\left(\tfrac{1}{3} + \tfrac{2}{3}b + \tfrac{2}{3}b\right) = \tfrac{1}{9} + \tfrac{4}{9}b \quad\Longrightarrow\quad \tfrac{5}{9}b = \tfrac{1}{9} \quad\Longrightarrow\quad b = \tfrac{1}{5}.$$

因此

$$p_A = \tfrac{1}{3} + \tfrac{2}{15} = \frac{7}{15}, \qquad \text{并且} \qquad p_B = p_D = \tfrac{1}{5},\ p_C = \tfrac{2}{15}.$$

> **想一想** 我们怎样知道兔子一定会离开，从而确定这些方程真的能把各个概率定下来?

每一步兔子都有 $\frac{1}{3}$ 的概率离开，所以经过 n 步后牠仍在地底的机会是 $(\frac{2}{3})^n \to 0$：牠以概率 1 离开。此外，这四条线性方程只有一个解（我们解方程时并无任何选择余地），所以上面的数值就是所求的概率。

第 (iii) 部分

设 N 为兔子从未到过 C 这一事件。我们要求的是

$$\mathrm{P}(N \mid E_A) = \frac{\mathrm{P}(N \cap E_A)}{\mathrm{P}(E_A)},$$

其中兔子由 A 出发，所以分母是 $p_A = \frac{7}{15}$。

至于分子，设 q_A、q_B、q_D 分别为兔子现时身处 A、B、D（且至今未到过 C）的条件下，从未进入 C 而经由 A 的隧道离开的概率。首步分析的做法与之前完全一样，只是现在走进 C 也算作失败：

$$\begin{aligned} q_A &= \tfrac{1}{3} + \tfrac{1}{3}q_B + \tfrac{1}{3}q_D, \\ q_B &= \tfrac{1}{3}q_A + \tfrac{1}{3}\cdot 0 + \tfrac{1}{3}\cdot 0, \end{aligned}$$

因为兔子从 B 出发，要么跑回 A，要么跑进 C（失败），要么经由错的隧道离开（失败）。根据对称性，$q_D = q_B = \frac{1}{3}q_A$，所以

$$q_A = \tfrac{1}{3} + \tfrac{2}{9}q_A \quad\Longrightarrow\quad \tfrac{7}{9}q_A = \tfrac{1}{3} \quad\Longrightarrow\quad q_A = \tfrac{3}{7}.$$

因此

$$\mathrm{P}(N \mid E_A) = \frac{3/7}{7/15} = \frac{45}{49}.$$

验算

- p_A, p_B, p_C, p_D 这四个值是从各个洞室出发而由 A 离开的概率。根据对称性，它们也分别是从 A 出发而由 A、B、C、D 离开的概率，所以它们的和必须是 1：果然 $\frac{7}{15} + \frac{3}{15} + \frac{2}{15} + \frac{3}{15} = 1$。
- $p_A > p_B = p_D > p_C$，这合乎预期：兔子的起点离 A 越远，从那里离开的可能性就越低。
- (iii) 的答案接近 1：从 A 离开的兔子，通常只走了很短的路程便离开。模拟 4×10^5 只兔子，得出 p_A 为 0.469，条件概率为 0.918，而理论值是 $\frac{7}{15} \approx 0.467$ 和 $\frac{45}{49} \approx 0.918$。

常见错误

- 在 p_B、p_C、p_D 的方程中也加入“立即离开”的 $\frac{1}{3}$ 一项：经由 A 以外任何一条隧道离开，贡献都是 0。
- 以 $q_A = \frac{3}{7}$ 作为 (iii) 的答案。那是 $\mathrm{P}(N \cap E_A)$；题目问的是条件概率，所以我们还要除以 p_A。

STEP 之外

兔子的行动构成一条马可夫链（Markov chain）：牠的下一步只取决于牠现时身处何处。地面是一个吸收态（absorbing state），而 p_A, p_B, p_C, p_D 是吸收概率（absorption probabilities），首步分析总能把它们化为一个线性方程组。同一想法还可以求出被吸收前的期望时间（expected time）、赌徒破产（gambler's ruin）概率等等。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率入门）对马可夫链有很好的初步介绍。

详细解答

STEP III 2003，第 14 题

概率与统计 · 概率生成函数

STEP III 2003，第 14 题

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

step.maths.org/questions/03-s3-q14

关键思路

独立得分之和的概率生成函数（probability generating function，简称 pgf），等于各个得分的 pgf 的乘积。因此，“哪些骰子会给出相同的总点数分布”这个问题，就变成“一个多项式可以怎样分拆成两个因式”的问题。把所有式子分解成不可约（irreducible）的因式，再把这些因式分配给两颗骰子，便可以读出所有可能的重新标号方式。

解答

第一部分

设 X 为一颗标准公平骰子的得分。$1,\dots,6$ 每个点数的概率都是 $\frac{1}{6}$，所以

$$G_X(t) = \mathrm{E}\big(t^X\big) = \tfrac{1}{6}\big(t + t^2 + t^3 + t^4 + t^5 + t^6\big) = \tfrac{1}{6}\,t\,\big(1 + t + t^2 + t^3 + t^4 + t^5\big).$$

要把括号内的式子因式分解，先把各项两两配对，然后再配成平方差：

$$1 + t + t^2 + t^3 + t^4 + t^5 = (1+t) + t^2(1+t) + t^4(1+t) = (1+t)(1+t^2+t^4),$$
$$1 + t^2 + t^4 = (1+t^2)^2 - t^2 = (1+t+t^2)(1-t+t^2).$$

因此

$$G_X(t) = \tfrac{1}{6}\,t\,(1+t)(1-t+t^2)(1+t+t^2).$$

若 X_1 和 X_2 是独立掷出的两颗骰子的得分，则 t^{X_1} 和 t^{X_2} 互相独立（independent），所以 $\mathrm{E}\big(t^{X_1+X_2}\big) = \mathrm{E}\big(t^{X_1}\big)\mathrm{E}\big(t^{X_2}\big)$，而且

$$G_{X_1+X_2}(t) = G_X(t)^2 = \frac{1}{36}\,t^2(1+t)^2(1-t+t^2)^2(1+t+t^2)^2,$$

这正是题目要求的形式。

第二部分

设 Y 和 Z 分别为标号是 $1,2,2,3,3,4$ 和 $1,3,4,5,6,8$ 的两颗骰子的得分。数一数各面的点数，得

$$G_Y(t) = \tfrac{1}{6}\left(t + 2t^2 + 2t^3 + t^4\right),$$
$$G_Z(t) = \tfrac{1}{6}\left(t + t^3 + t^4 + t^5 + t^6 + t^8\right).$$

对于 Y：$1 + 2t + 2t^2 + t^3 = (1+t)(1+t+t^2)$，把右边展开即可验证。所以

$$G_Y(t) = \tfrac{1}{6}\, t\,(1+t)(1+t+t^2).$$

对于 Z, 第一部分中有用的乘积是 $(1+t)(1-t+t^2) = 1+t^3$ 和 $(1+t+t^2)(1-t+t^2) = 1+t^2+t^4$。于是

$$(1+t^3)(1+t^2+t^4) = 1+t^2+t^3+t^4+t^5+t^7,$$

再乘以 t，刚好得到 $t+t^3+t^4+t^5+t^6+t^8$。因此

$$G_Z(t) = \tfrac{1}{6}\, t\,(1+t)(1+t+t^2)(1-t+t^2)^2.$$

两颗骰子独立掷出，故

$$G_{Y+Z}(t) = G_Y(t)G_Z(t) = \frac{1}{36}\, t^2(1+t)^2(1-t+t^2)^2(1+t+t^2)^2 = G_{X_1+X_2}(t).$$

两个 pgf 若作为多项式相等，它们的系数便一一相等，而 t^n 的系数就是总点数为 n 的概率。所以用这两颗骰子掷出每个总点数的概率，都与用两颗标准骰子时相同。

最后部分

一颗标准四面骰子的 pgf 是

$$\tfrac{1}{4}\left(t + t^2 + t^3 + t^4\right) = \tfrac{1}{4}\, t\,(1+t)(1+t^2),$$

所以两颗这样的骰子的总点数，其目标 pgf 是

$$\frac{1}{16}\, t^2(1+t)^2(1+t^2)^2.$$

假设两颗新骰子的 pgf 分别是 $\frac{1}{4}f(t)$ 和 $\frac{1}{4}g(t)$。f 中 t^n 的系数是标号为 n 的面的数目，所以 f 和 g 都是系数为非负整数的多项式，并且 $f(1) = g(1) = 4$。由于每个标号都是正整数，所以 $f(0) = g(0) = 0$。我们需要

$$f(t)\, g(t) = t^2(1+t)^2(1+t^2)^2.$$

想一想 为甚么 f 必定只由 t、$1+t$、$1+t^2$ 这几个因式组成，而不会有其他因式?

因式 t 和 $1+t$ 是一次式，而 $1+t^2$ 没有实根，所以三者在实数范围内都是不可约的。多项式分解成不可约因式的方法是唯一的（不计常数倍），所以右边的任何因式都必定是

$$f(t) = c\,t^k(1+t)^a(1+t^2)^b, \qquad g(t) = \tfrac{1}{c}\,t^{2-k}(1+t)^{2-a}(1+t^2)^{2-b},$$

其中 $0 \le k, a, b \le 2$，而常数 $c > 0$。

现在逐一运用 f 和 g 须符合的条件。

- *常数 c。* f 的最低次项系数是 c，而 g 的是 $1/c$。两者都是面的数目，所以都是正整数，由此必有 $c = 1$。
- *t 的幂次。* $f(0) = 0$ 需要 $k \ge 1$，而 $g(0) = 0$ 需要 $2 - k \ge 1$。所以 $k = 1$：每颗骰子恰好有一个因式 t。
- *四个面。* 由 $f(1) = 2^a \cdot 2^b = 4$ 得 $a + b = 2$，这时 $g(1) = 2^{2-a}2^{2-b} = 4$ 自动成立。

所以 (a, b) 是 $(2,0)$、$(1,1)$ 或 $(0,2)$，而 g 则取与之互补的选择。由于把两颗骰子对调会得出同一对骰子，所以只有两种情况。

- 两者皆为 $(a, b) = (1,1)$：$f = g = t(1+t)(1+t^2) = t + t^2 + t^3 + t^4$，即标准的一对 $1,2,3,4$ 和 $1,2,3,4$。
- $(a, b) = (2,0)$ 和 $(0,2)$：

  $$f(t) = t(1+t)^2 = t + 2t^2 + t^3, \qquad g(t) = t(1+t^2)^2 = t + 2t^3 + t^5,$$

  即标号为 $1,2,2,3$ 和 $1,3,3,5$ 的两颗骰子。

在每种情况中，两个多项式的系数都是非负整数，而且总和为 4，所以两者确实都是骰子；而按照构造方法，它们会给出正确的 pgf。

因此，仅有的标号方式就是标准的一对，以及 $\{1,2,2,3\}$、$\{1,3,3,5\}$ 这一对。

验算

对于那对非标准的四面骰子,把总点数逐一列出:16 个等可能的和是 2, 4, 4, 6, 3, 5, 5, 7, 3, 5, 5, 7, 4, 6, 6, 8, 得出总点数 $2, \ldots, 8$ 的频数为 $1, 2, 3, 4, 3, 2, 1$，与两颗标准骰子完全相同。把 $t = 1$ 代入上面任何一个 pgf 都会得到 1，这是检查每个因式分解的快捷方法。用电脑搜索所有标号由 1 至 7 的四面骰子对，找到的恰好就是上述两对；而 $1,2,2,3,3,4$ 和 $1,3,4,5,6,8$ 这两颗骰子的总点数，亦与标准分布完全吻合（例如两种情况下都有 $\mathrm{P}(\text{总点数} = 7) = \frac{6}{36}$）。

常见错误

- 忘记了常数 c：你需要说明为甚么不可以一颗骰子带有因子 2，而另一颗带有因子 $\frac{1}{2}$。理由就是两颗骰子的最低次项系数都必须是正整数。
- 容许 $k = 0$ 或 $k = 2$。这样其中一颗骰子便会有一面标号为 0，而这并不是正整数。
- 把 $(1 - t + t^2)$ 分解成复数一次因式。在实数范围内它是不可约的，而这正是分配因式时所需要的形式。

STEP 之外

这里的一对六面骰子称为 *Sicherman* 骰子（Sicherman dice），用同样的方法可以证明，它们是唯一一对标号为正整数的非标准六面骰子。这个论证建基于多项式的唯一分解（unique factorisation）；你会在大学第一个环与域（rings and fields）的课程中遇到它，它是算术基本定理（fundamental theorem of arithmetic）的多项式版本。生成函数亦能把许多困难的计数问题化为代数问题。关于这两个课题，可参阅 Allenby 的 *Rings, Fields and Groups*（环、域与群），以及 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）。

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

2003 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。

