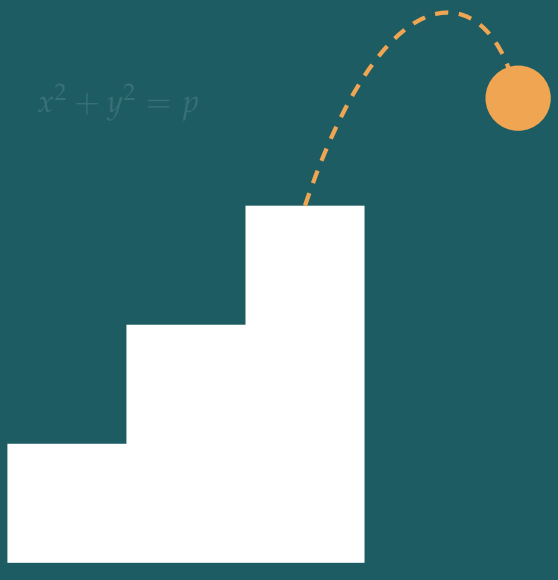
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2005 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

2005 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2005 年详细解答

2005 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 数字和的计数 纯数学 3

第 2 题 抛物线的切线与法线 纯数学 6

第 3 题 分式方程 纯数学 10

第 4 题 倍角公式 纯数学 13

第 5 题 幂的积分 纯数学 17

第 6 题 阿波罗尼奥斯圆 纯数学 20

第 7 题 裂项相消乘积 纯数学 24

第 8 题 一阶微分方程 纯数学 27

第 9 题 刚体平衡 力学 31

第 10 题 碰撞与恢复系数 力学 34

第 11 题 匀变量运动学 力学 38

第 12 题 温氏图与界限 概率与统计 42

第 13 题 对称分布 概率与统计 45

第 14 题 混合型随机变量 概率与统计 48

STEP II

第 1 题 多项式与驻点 纯数学 52

第 2 题 欧拉 φ 函数 纯数学 55

第 3 题 分部积分法 纯数学 59

第 4 题 反正切恒等式 纯数学 63

第 5 题 内切圆与外接圆 纯数学 67

第 6 题 二项式级数的和 纯数学 71

第 7 题 向量与圆周运动 纯数学 74

第 8 题 可分离变量微分方程 纯数学 77

第 9 题 斜面上的摩擦力 力学 . . . 82
第 10 题 碰撞的抛体 力学 . . . 86
第 11 题 斜面上的摩擦力 力学 . . . 89
第 12 题 贝叶斯定理 概率与统计 . . . 93
第 13 题 泊松分布与二项分布 概率与统计 . . . 97
第 14 题 混合分布 概率与统计 . . . 100

STEP III

第 1 题 三角方程与图像 纯数学 . . . 104
第 2 题 可分离变量微分方程 纯数学 . . . 109
第 3 题 二次函数的复合 纯数学 . . . 113
第 4 题 非线性递推关系 纯数学 . . . 117
第 5 题 抛物线的公切线 纯数学 . . . 120
第 6 题 解三次方程 纯数学 . . . 124
第 7 题 代换积分法 纯数学 . . . 128
第 8 题 复数与圆 纯数学 . . . 132
第 9 题 碰撞与恢复系数 力学 . . . 136
第 10 题 弹性带与摩擦力 力学 . . . 139
第 11 题 刚体平衡 力学 . . . 143
第 12 题 无偏估计量的组合 概率与统计 . . . 147
第 13 题 几何分布 概率与统计 . . . 150
第 14 题 连续随机变量 概率与统计 . . . 153

详细解答

STEP I 2005，第 1 题

纯数学 · 数字和的计数

> **STEP I 2005，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/05-s1-q1

关键思路

五位数的数字和（digit sum）最大可能是 $9 \times 5 = 45$，而题目要求的两个数字和都很接近这个值。因此，与其直接数各个数字本身，不如数每个数字比 9 少了多少。这些差额（shortfall）都是很小的非负数，总和又是一个固定的小数目，数起来十分容易：既可以逐一列出各种情况，也可以用“星与棒”法（stars and bars）。

解答

把这个数写成 $d_1d_2d_3d_4d_5$，其中 $d_1 \in \{1,\dots,9\}$（首位不能是零），而 $d_2,\dots,d_5 \in \{0,\dots,9\}$。设

$$e_i = 9 - d_i \qquad (i = 1,\dots,5),$$

即第 i 个数字的*差额*。每个 e_i 都是整数，且 $0 \le e_i \le 9$（并且 $e_1 \le 8$，因为 $d_1 \ge 1$），而

$$d_1 + d_2 + d_3 + d_4 + d_5 = 45 - (e_1 + e_2 + e_3 + e_4 + e_5).$$

不同的数给出不同的序列 $(e_1,\dots,e_5)$，反之亦然，所以数算数字和为 S 的数，等同于数算总和为 $45 - S$ 的有序差额序列。

第 (i) 部分

数字和为 43，即差额的总和是 $45 - 43 = 2$。把 2 分拆成非负部分只有两种方法：

- **一个数字少了 2，其余少了 0**。一个数字是 7，其余四个是 9。这个 7 可以放在 5 个位置中的任何一个：共 5 个数（$79999, 97999, \dots, 99997$）。
- **两个数字各少了 1，其余少了 0**。两个数字是 8，其余三个是 9。我们从 5 个位置中选

出 2 个来放 8：共 $\binom{5}{2} = 10$ 个数（88999, 89899, . . . , 99988）。

每个数字都至少是 7，所以首位数字绝不会是 0，以上各数确实都是五位数。这两种情况互不重叠，并且涵盖了所有可能，所以共有

$$5 + 10 = 15$$

个五位数的数字和为 43，正如题目所求。

第 (ii) 部分

现在差额的总和必须是 $45 - 39 = 6$。由于 $6 \leq 8$，没有任何一个差额会超出其上限：$e_1 \leq 6 \leq 8$ 保证 $d_1 \geq 3$，而 $e_i \leq 6 \leq 9$ 保证其余数字都在 3 与 9 之间。所以我们只需求出下式的有序解的数目：

$$e_1 + e_2 + e_3 + e_4 + e_5 = 6, \qquad e_i \geq 0 \text{ 且为整数.}$$

想一想 怎样才可以不用逐一列出而数出这些解?

想像把 6 颗星排成一行，用 4 根棒把它们分成 5 组：

$$\star\star \,|\, |\, \star \,|\, \star\star\star \,|\quad \text{代表} \quad (e_1, e_2, e_3, e_4, e_5) = (2, 0, 1, 3, 0).$$

6 颗星和 4 根棒的每一种排列都恰好给出一个解，而每一个解也恰好来自一种排列。只要选定 10 个位置中哪 4 个放棒，排列便随之确定，所以解的数目是

$$\binom{10}{4} = 210.$$

因此，共有 **210** 个五位数的数字和为 39。

如果你宁愿像第 (i) 部分那样分情况来数，可以列出把 6 分拆成最多五个正数部分的所有方法，再数出每种方法的排列数目：

差额	数的个数	差额	数的个数
6	5	3, 2, 1	$5 \cdot 4 \cdot 3 = 60$
5, 1	$5 \cdot 4 = 20$	3, 1, 1, 1	$5 \cdot 4 = 20$
4, 2	$5 \cdot 4 = 20$	2, 2, 2	$\binom{5}{2} = 10$
4, 1, 1	$5 \cdot \binom{4}{2} = 30$	2, 2, 1, 1	$\binom{5}{2}\binom{3}{2} = 30$
3, 3	$\binom{5}{2} = 10$	2, 1, 1, 1, 1	5

总数是 $5 + 20 + 20 + 30 + 10 + 60 + 20 + 10 + 30 + 5 = 210$，与星与棒法的结果一致。(6 的分拆（partition）1, 1, 1, 1, 1, 1 需要六个数字，所以不会出现。)

验算

- 星与棒法可重现第 (i) 部分的结果：差额总和为 2 时，得 $\binom{6}{4} = 15$。
- 极端情况同样成立：差额总和为 0 时，得 $\binom{4}{4} = 1$ 个数（99999）；差额总和为 1 时，得 $\binom{5}{4} = 5$。
- 用电脑搜索全部 90 000 个五位数，结果恰好有 15 个数的数字和为 43，210 个数的数字和为 39，而各种情况的分布亦与上表相符。

常见错误

- 把 $\{8,8,9,9,9\}$ 当作一个数，或当作 5 个数：数字的次序是有分别的，共有 $\binom{5}{2} = 10$ 种排列。
- 忘记检查首位数字。本题的首位不可能是 0，因为每个数字都至少是 3；但如果数字和较小（例如 20），星与棒法便会多算，这时必须剔除 $d_1 = 0$ 或有数字大于 9 的情况。

STEP 之外

把 n 写成 k 个非负整数的有序和，方法的数目是 $\binom{n+k-1}{k-1}$，这也是 $(1 + x + x^2 + \cdots)^k$ 中 x^n 的系数（coefficient）。若限制每个数字最大为 9，上式便要换成 $(1 + x + \cdots + x^9)^k$，而截断所带来的影响可用容斥原理（inclusion–exclusion principle）处理。这些生成函数（generating function）方法是大学一年级组合数学（combinatorics）的基本内容；Siklos 的 *Advanced Problems in Mathematics* 和 Davenport 的 *The Higher Arithmetic* 都是继续探索数字与计数问题的好去处。

详细解答

STEP I 2005，第 2 题

纯数学 · 抛物线的切线与法线

STEP I 2005，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/05-s1-q2

关键思路

把抛物线（parabola）$y^2 = 4x$ 用 $(t^2, 2t)$ 参数化（parametrise）之后，每条切线（tangent）和法线（normal）都是一条关于 x 和 y 的简短线性方程，所以求交点不过是解两条联立线性方程。$(1, 0)$ 是焦点（focus），它位于 PQ 上这个条件，会化为 p 与 q 之间一条简洁的关系式，而正正是这条关系式令 P 和 Q 处的切线互相垂直。得到这一点之后，那个长方形几乎是自然而然地出现。

解答

切线与点 R

在点 $(t^2, 2t)$，我们有 $\frac{\mathrm{d}x}{\mathrm{d}t} = 2t$ 和 $\frac{\mathrm{d}y}{\mathrm{d}t} = 2$，所以 C 在该点的斜率（gradient）是 $\frac{2}{2t} = \frac{1}{t}$（这没有问题，因为 $t \neq 0$）。因此 P 处的切线是

$$y - 2p = \frac{1}{p}(x - p^2) \quad \Longleftrightarrow \quad py = x + p^2,$$

同理，Q 处的切线是 $qy = x + q^2$。两式相减，

$$(p - q)\, y = p^2 - q^2 = (p - q)(p + q),$$

由于 $p \neq q$，我们可以除以 $p - q$，得到 $y = p + q$。于是 $x = py - p^2 = p(p + q) - p^2 = pq$。所以

$$R = (pq,\ p + q),$$

正如题目所求。

法线与点 S

$(t^2, 2t)$ 处的法线与切线垂直，所以它的斜率是 $-t$。P 处的法线是

$$y - 2p = -p\,(x - p^2) \quad \Longleftrightarrow \quad px + y = 2p + p^3,$$

而 Q 处的法线是 $qx + y = 2q + q^3$。两式相减，

$$(p - q)\,x = 2(p - q) + (p^3 - q^3) = (p - q)\big(2 + p^2 + pq + q^2\big),$$

所以，再次除以 $p - q \neq 0$，

$$x = p^2 + pq + q^2 + 2, \qquad y = 2p + p^3 - px = -p^2q - pq^2 = -pq\,(p + q).$$

以上结果对任何 p 和 q 都成立；现在我们运用那个额外条件。

运用 $(1, 0)$ 位于 PQ 上这个条件

弦（chord）PQ 的斜率是

$$\frac{2p - 2q}{p^2 - q^2} = \frac{2}{p + q}$$

（当 $p + q \neq 0$ 时），所以它的方程是 $y - 2p = \dfrac{2}{p + q}(x - p^2)$，化简后得

$$(p + q)\,y = 2x + 2pq.$$

这个形式同时涵盖 $p + q = 0$ 的情况，这时它给出通过 P 和 Q 的铅垂线 $x = -pq = p^2$。代入 $(x, y) = (1, 0)$ 得 $0 = 2 + 2pq$，所以

$$pq = -1.$$

代入 S 的坐标：

$$S = \big(p^2 + q^2 + 2 + pq,\ -pq(p + q)\big) = \big(p^2 + q^2 + 1,\ p + q\big),$$

正如题目所求。同时留意，现在 $R = (-1,\ p + q)$，它位于直线 $x = -1$ 上，亦即准线（directrix）上。

$PSQR$ 是长方形

> ? **想一想** 要证明四个给定的点构成一个长方形，最干净利落的方法是甚么？

一条不错的路线是：先证明它是平行四边形（parallelogram）（对角线互相平分），再证明其中一只角是直角。

平行四边形。在 $PSQR$ 中，对角线是 PQ 和 SR。利用 $pq = -1$，

$$R + S = \big(pq + p^2 + q^2 + 1,\ 2(p+q)\big) = \big(p^2 + q^2,\ 2p + 2q\big) = P + Q,$$

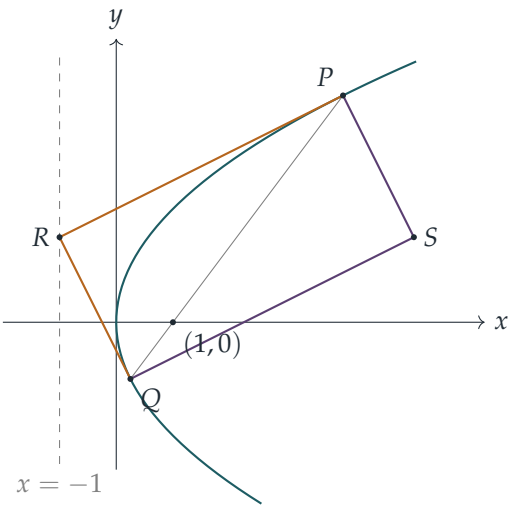
所以 RS 的中点等于 PQ 的中点。对角线互相平分，所以 $PSQR$ 是平行四边形。它是一个真正的平行四边形，而不是退化（degenerate）的：P、R 和 Q 并不共线（collinear），因为 R 同时位于两条切线上，而抛物线上两个不同点的切线只相交一次，并且任何一条都不会通过另一个切点。

一只直角。边 RP 和 RQ 分别沿着 P 和 Q 处的切线，斜率为 $\frac{1}{p}$ 和 $\frac{1}{q}$。它们的乘积是

$$\frac{1}{p} \cdot \frac{1}{q} = \frac{1}{pq} = -1,$$

所以两条切线互相垂直，角 PRQ 是直角。

有一只角是直角的平行四边形就是长方形，所以 $PSQR$ 是长方形。（等价的说法是：P 和 Q 处的角都是直角，因为每条法线都与自己的切线垂直；再加上 R 处的直角，第四只角也必定是直角。）



图中所示为 $p = 2$、$q = -\frac{1}{2}$ 的情况：切线为琥珀色，法线为梅子色，焦点弦 PQ 为灰色。

验算

取 $p = 2$、$q = -\frac{1}{2}$，于是 $pq = -1$。这时 $P = (4,4)$ 而 $Q = (\frac{1}{4}, -1)$，通过这两点的直线 $y - 4 = \frac{4}{3}(x-4)$ 确实通过 $(1,0)$。由公式得 $R = (-1, \frac{3}{2})$ 和 $S = (\frac{21}{4}, \frac{3}{2})$。直接验证：$\overrightarrow{RP} = (5, \frac{5}{2})$ 与 $\overrightarrow{QS} = (5, \frac{5}{2})$ 相等，而 $\overrightarrow{RQ} = (\frac{5}{4}, -\frac{5}{2})$ 与 $\overrightarrow{RP}$ 的点积（dot product）是 $\frac{25}{4} - \frac{25}{4} = 0$。所以 $PSQR$ 的确是长方形。我们亦以精确分数，就约二千组随机选取的 p 和 q 核对了所有坐标公式。

常见错误

- 除以 $p-q$ 而没有说明为何可以这样做；题目给出 $p \neq q$，正是为了让你可以这样做。
- 只证明了某一个角上的两条邻边互相垂直，单凭这一点并不足以构成长方形：你需要的是平行四边形加上一只直角，或者三个角都是直角。
- 忘记在 S 中使用 $pq=-1$：一般情况下法线的交点是 $\left(p^2+pq+q^2+2,\ -pq(p+q)\right)$，只有在焦点弦上它才会化简成题目给出的形式。

STEP 之外

通过焦点的弦称为*焦点弦*（focal chord），而本题证明了关于它的两个经典事实：弦两端的切线互相垂直，并且相交于准线上。两者都是圆锥曲线（conics）的极点与极线（poles and polars）理论的特殊情况，在这套理论中，准线就是焦点的极线；射影几何（projective geometry）令这类结果几乎不证自明。若想循这个方向探索处理圆锥曲线的参数方法，Siklos 的 *Advanced Problems in Mathematics*（数学进阶难题集）是一个好去处。

详细解答

STEP I 2005，第 3 题

纯数学 · 分式方程

STEP I 2005，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/05-s1-q3

关键思路

把等式两边同乘 $(x-a)(x-b)$，将每条方程化为多项式方程（polynomial equation），然后数算它的实根（real root）数目。需要小心的地方是：去分母原则上有可能多出原方程不容许的“解” $x=a$ 或 $x=b$，所以我们必须验证这种情况绝不会发生。在第 (ii) 部分，二次方程有重根（repeated root）当且仅当其判别式（discriminant）为零，由此便得出所求的条件。

解答

第 (i) 部分

当 $x \neq a, b$ 时，我们可以乘以 $(x-a)(x-b)$，而方程等价于

$$\begin{aligned} x(x-b)+x(x-a) &= (x-a)(x-b) \\ 2x^2-(a+b)x &= x^2-(a+b)x+ab \\ x^2 &= ab. \end{aligned}$$

由于 a 与 b 同号，$ab>0$，所以这方程有两个相异实根 $x=\sqrt{ab}$ 和 $x=-\sqrt{ab}$。

我们仍须确认这两个根是容许的值。假如 $x=a$ 是根，则 $a^2=ab$，而由于 $a\neq 0$，这给出 $a=b$，与事实不符。同理，$x=b$ 也不是根。因此两个根都是真正的解，方程恰有两个相异实数解 $x=\pm\sqrt{ab}$。

第 (ii) 部分

同样，当 $x \neq a, b$ 时乘以 $(x-a)(x-b)$：

$$2x^2 - (a+b)x = (1+c)\big(x^2 - (a+b)x + ab\big),$$

整理后得

$$q(x) \equiv (1-c)x^2 + c(a+b)x - (1+c)ab = 0.$$

因为 $c \neq 1$，首项系数（leading coefficient）不为零，所以这确实是一条二次方程。

> ? **想一想** q 的其中一个根会不会是 a 或 b，以致必须舍去?

代入看看：

$$\begin{aligned} q(a) &= (1-c)a^2 + c(a+b)a - (1+c)ab = a^2 - ab = a(a-b), \\ q(b) &= b^2 - ab = b(b-a) \end{aligned}$$

（含 c 的项互相抵消）。由于 a、b 皆不为零且互不相等，两者都不为零。所以 a 和 b 永远不是 q 的根。这表示 q 的每个实根都是原方程的解，反之亦然。

q 的判别式为

$$\begin{aligned} \Delta &= c^2(a+b)^2 + 4(1-c)(1+c)ab \\ &= c^2\left[(a+b)^2 - 4ab\right] + 4ab \\ &= c^2(a-b)^2 + 4ab. \end{aligned}$$

若 $c^2 = -\dfrac{4ab}{(a-b)^2}$（而 $a \neq b$，所以可以除以 $(a-b)^2$），则 $\Delta = 0$。这时二次方程恰有一个（重）实根

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

而根据上面的验证，这个根不是 a 或 b。所以方程恰有一个实数解，正如所求。（事实上逆命题（converse）也成立。若 $\Delta > 0$ 则有两个解，若 $\Delta < 0$ 则无解，所以这个条件也是必要的。）

改写条件。 通分后，

$$1 - \left(\frac{a+b}{a-b}\right)^2 = \frac{(a-b)^2 - (a+b)^2}{(a-b)^2} = \frac{-4ab}{(a-b)^2},$$

所以条件就是 $c^2 = 1 - \left(\dfrac{a+b}{a-b}\right)^2$，正如所求。

推导 $0 < c^2 \le 1$。平方数 $\left(\frac{a+b}{a-b}\right)^2$ 为非负，所以由第二种形式可得 $c^2 \le 1$。第一种形式给出 $c^2 = -4ab/(a-b)^2$。由于 $ab \neq 0$，这个值不为零；而对于实数 c 有 $c^2 \ge 0$，所以 $c^2 > 0$。（这也告诉我们 $ab < 0$：a 与 b 异号，这与第 (i) 部分吻合。）因此，条件只有在 $0 < c^2 \le 1$ 时才可能成立。

上限值是可以取到的。$c^2 = 1$ 当且仅当 $a + b = 0$，这时 $c = -1$（因为 $c \neq 1$）。

验算

- 取 $a = 4$，$b = -1$。则 $c^2 = 16/25$；选 $c = \frac{4}{5}$。重根为 $x = -\frac{4}{5} \cdot 3/(2 \cdot \frac{1}{5}) = -6$，而确实有 $\frac{-6}{-10} + \frac{-6}{-5} = 0.6 + 1.2 = 1.8 = 1 + c$。
- 极端情况 $a = 1$，$b = -1$，$c = -1$ 给出 $\frac{x}{x-1} + \frac{x}{x+1} = 0$，即 $2x^2 = 0$：只有一个解 $x = 0$。
- 在第 (i) 部分，$a = 1$，$b = 4$ 给出 $x = \pm 2$：$\frac{2}{1} + \frac{2}{-2} = 1$ 及 $\frac{-2}{-3} + \frac{-2}{-6} = 1$。

常见错误

- 去分母之后忘记检查所得的根不是 a 或 b。在本题中这一步结果并无影响，但仍必须加以证明。
- 展开判别式时没有把含 c^2 的项归并在一起。写出 $(a+b)^2 - 4ab = (a-b)^2$，条件才会干净利落地得出。
- 声称 $c^2 \ge 0$ 而不是 $c^2 > 0$。严格不等式（strict inequality）来自 $ab \neq 0$。

STEP 之外

等式左边是一个有理函数（rational function），而“它的图像与水平线 $y = 1 + c$ 相交多少次?”其实是一个关于其值域（range）的问题。这里找到的条件，挑出了该函数的转向点（turning point）的值，也就是水平线刚好与图像相切的高度。追问多项式何时有重根，会引出判别式，以及更一般的结式（resultant），这些都会在大学一年级的代数课程中出现。想做更多这类风格的题目，可参考 Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）。

详细解答

STEP I 2005，第 4 题

纯数学 · 倍角公式

> **STEP I 2005，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/05-s1-q4

关键思路

两部分都与倍角（multiple angles）有关。在 (i) 中，唯一真正要判断的是 $\sin\theta$ 的正负号，而题目给定的范围已把它确定下来；之后只需运用二倍角公式（double-angle formula）和三倍角公式（triple-angle formula）便可。在 (ii) 中，我们把 $\tan 3\theta$ 写成 $\tan(2\theta+\theta)$ 来推导，然后“由此”（hence）一步便化为关于 $t=\tan\theta$ 的三次方程（cubic）。STEP 题目中的三次方程几乎总有一个可以看出来的根（root），而 θ 的范围则告诉我们应保留哪一个根。

解答

第 (i) 部分

由于 $\cos^2\theta+\sin^2\theta=1$，

$$\sin^2\theta = 1-\tfrac{9}{25} = \tfrac{16}{25}, \qquad \text{所以} \qquad \sin\theta = \pm\tfrac{4}{5}.$$

当 $\frac{3\pi}{2}\le\theta\le 2\pi$ 时，θ 位于第四象限（fourth quadrant），该处 $\sin\theta\le 0$。所以 $\sin\theta=-\frac{4}{5}$，而

$$\sin 2\theta = 2\sin\theta\cos\theta = 2\left(-\tfrac{4}{5}\right)\left(\tfrac{3}{5}\right) = -\frac{24}{25},$$

正是所求。

要求 $\cos 3\theta$，我们需要三倍角公式。记 $c = \cos\theta$ 及 $s = \sin\theta$，

$$\begin{aligned}\cos 3\theta = \cos(2\theta+\theta) &= \cos 2\theta\cos\theta - \sin 2\theta\sin\theta \\ &= (2c^2-1)c - 2s^2c \\ &= (2c^2-1)c - 2(1-c^2)c \\ &= 4c^3 - 3c.\end{aligned}$$

代入 $c = \frac{3}{5}$，

$$\cos 3\theta = 4\cdot\frac{27}{125} - \frac{9}{5} = \frac{108-225}{125} = -\frac{117}{125}.$$

（这里 $\sin\theta$ 的正负号并无影响，因为 $\cos 3\theta$ 只取决于 $\cos\theta$。）

第 (ii) 部分

恒等式 (identity)。记 $t = \tan\theta$。由二倍角公式，$\tan 2\theta = \dfrac{2t}{1-t^2}$，再由和角公式（addition formula），

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

把分子和分母同乘以 $1-t^2$，

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

这就是所求的恒等式。以上推导需要 $\tan 2\theta$ 存在，即 $t^2 \neq 1$；当 $t = \pm 1$ 时，我们可以直接验证：例如取 $\theta = \frac{\pi}{4}$，右方为 $\frac{3-1}{1-3} = -1 = \tan\frac{3\pi}{4}$，而 $t = -1$ 的情况相同，只是正负号互换。

解 $\tan 3\theta = \frac{11}{2}$。由恒等式，

$$\frac{3t-t^3}{1-3t^2} = \frac{11}{2} \quad\Longrightarrow\quad 6t - 2t^3 = 11 - 33t^2 \quad\Longrightarrow\quad 2t^3 - 33t^2 - 6t + 11 = 0.$$

这里交叉相乘并无问题，因为 $1-3t^2 \neq 0$：假如它等于零，$\tan 3\theta$ 便没有定义，而不会是 $\frac{11}{2}$。

> **想一想** 怎样找出 $2t^3 - 33t^2 - 6t + 11$ 的一个根?

任何写成最简分数 p/q 的有理根（rational root），其 p 必整除 11，q 必整除 2，所以我们只需试 $\pm 1, \pm 11, \pm\frac{1}{2}, \pm\frac{11}{2}$。试 $t = \frac{1}{2}$：

$$2\cdot\tfrac{1}{8} - 33\cdot\tfrac{1}{4} - 3 + 11 = \tfrac{1}{4} - \tfrac{33}{4} + 8 = 0.$$

所以 $2t-1$ 是一个因式（factor），比较系数可得

$$2t^3-33t^2-6t+11=(2t-1)(t^2-16t-11).$$

该二次式的根为 $t=8\pm\sqrt{64+11}=8\pm5\sqrt{3}$。所以三个候选值是

$$t=\tfrac{1}{2},\qquad t=8+5\sqrt{3}\approx16.66,\qquad t=8-5\sqrt{3}\approx-0.66.$$

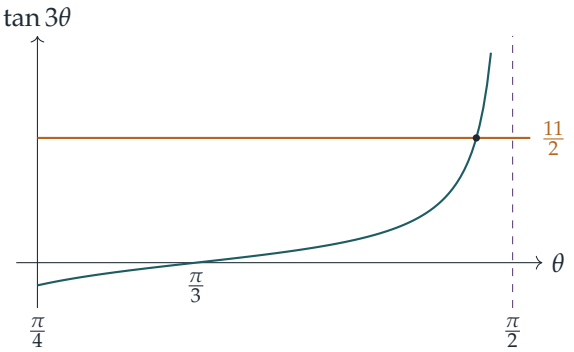
它们对应 $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ 内的三个角，这些角的三倍的正切都是 $\frac{11}{2}$。

现在运用题目给定的范围。由于 tan 在 $\left[\frac{\pi}{4},\frac{\pi}{2}\right)$ 上递增，且 $\tan\frac{\pi}{4}=1$，我们需要 $t\geq1$（而 $\theta=\frac{\pi}{2}$ 本身须排除，因为 $\tan\theta$ 在该处没有定义）。只有一个候选值符合要求：

$$\tan\theta=8+5\sqrt{3}.$$

另外两个确实是三次方程的根，但所属的角在范围之外：$\frac{1}{2}<1$ 而 $8-5\sqrt{3}<0$。

图像亦确认恰好只有一个解：当 θ 由 $\frac{\pi}{4}$ 变到 $\frac{\pi}{2}$ 时，3θ 由 $\frac{3\pi}{4}$ 变到 $\frac{3\pi}{2}$，而在该区间上 $\tan3\theta$ 由 -1 连续递增至 $+\infty$，所以它恰好一次穿过 $\frac{11}{2}$。



验算

- 在 (i) 中，$\theta=2\pi-\arccos\frac{3}{5}\approx5.3559$，用计算机可得 $\sin2\theta=-0.96$ 及 $\cos3\theta=-0.936=-\frac{117}{125}$。
- 在 (ii) 中，$\theta=\arctan(8+5\sqrt{3})\approx1.5108$，介乎 $\frac{\pi}{4}\approx0.785$ 与 $\frac{\pi}{2}\approx1.571$ 之间，而 $\tan3\theta\approx5.5000$。
- 快速检验恒等式：当 $\theta=\frac{\pi}{12}$ 时，$t=2-\sqrt{3}$，右方化简后得 $1=\tan\frac{\pi}{4}$。

常见错误

- 在 (i) 中取 $\sin\theta=+\frac{4}{5}$，以致 $\sin2\theta$ 的正负号出错。

▶ 在 (ii) 中因为 $t = \frac{1}{2}$ 是自己看出来的根，便就此停下。它确是三次方程的根，但 $\arctan\frac{1}{2} < \frac{\pi}{4}$。

STEP 之外

这个三次方程有三个根，是因为 $\tan 3\theta = \frac{11}{2}$ 在 $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ 内有三个 θ 的解，彼此相隔 $\frac{\pi}{3}$。同一机制反过来运作，便是三次方程的经典三角解法；更一般而言，棣美弗定理（de Moivre's theorem）把 $\tan n\theta$ 表示为 $\tan\theta$ 的两个多项式（polynomial）之比，而这些多项式的根正是等距分布的角的正切。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）说明了为何倍角公式会如此自然地从 $(\cos\theta + \mathrm{i}\sin\theta)^n$ 中产生。

详细解答

STEP I 2005，第 5 题

纯数学 · 幂的积分

STEP I 2005，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/05-s1-q5

关键思路

$x+1$ 的幂可以用反向链式法则（reverse chain rule）来积分：$\int (x+1)^n \,\mathrm{d}x = \frac{(x+1)^{n+1}}{n+1}$，但有一个例外：当 $n=-1$ 时，答案是对数（logarithm）。因此，本题所有的分类讨论都源于同一个问题："指数何时会变成 -1?"在 (ii) 中，多出来的因子 x 可以通过写成 $x=(x+1)-1$ 来消去，剩下两个单纯的幂，所以 m 有两个例外值。

解答

第 (i) 部分

若 $k \neq 0$，指数 $k-1$ 不等于 -1，所以

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

若 $k=0$，被积函数（integrand）是 $\frac{1}{x+1}$，而由于在 $[0,1]$ 上 $x+1>0$，

$$\int_0^1 \frac{1}{x+1} \,\mathrm{d}x = \Big[\ln(x+1)\Big]_0^1 = \ln 2.$$

? **想一想** 为甚么 $k \neq 0$ 时的答案会与 $k=0$ 时的答案有关?

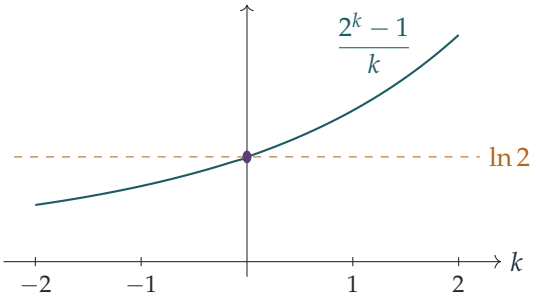
因为当 k 稍微偏离 0 时，被积函数几乎没有变化。对于 $0 \leq x \leq 1$，

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

这是因为 $1 \le x+1 \le 2$ 意味着 $(x+1)^k$ 介乎 1 与 2^k 之间。在长度为 1 的区间上积分，两个积分之差至多为 $|2^k-1|$，而当 $k \to 0$ 时，此值趋向 0。因此，对于 $k \approx 0$（且 $k \neq 0$），

$$\frac{2^k-1}{k} = \int_0^1 (x+1)^{k-1}\,\mathrm{d}x \approx \int_0^1 (x+1)^{-1}\,\mathrm{d}x = \ln 2,$$

正如所求。换句话说，$k \neq 0$ 时的公式在 $k=0$ 处有一个“洞”，而 $k=0$ 时的积分恰好把它填上。



第 (ii) 部分

写成 $x = (x+1)-1$，于是

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

每一项都是 $x+1$ 的幂，而对它积分时，恰好在其指数为 -1 时才会出现对数。这发生于 $m+1=-1$ 或 $m=-1$，所以要分的情况是 $m=-1$、$m=-2$，以及其余所有情况。

一般情况， $m \neq -1, -2$。

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

通分之后，分子为

$$(m+1)\left(2^{m+2}-1\right) - (m+2)\left(2^{m+1}-1\right) = 2^{m+1}\left(2(m+1)-(m+2)\right)+1 = m\,2^{m+1}+1,$$

所以

$$\int_0^1 x(x+1)^m\,\mathrm{d}x = \frac{m\,2^{m+1}+1}{(m+1)(m+2)}, \qquad m \neq -1, -2.$$

情况 $m=-1$。

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

情况 $m=-2$。

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

想一想 一般公式在 $m=-1$ 和 $m=-2$ 处会出现 $0/0$。这是否矛盾?

不是：这其实是第 (i) 部分的重演。在 $m=-1$ 处，分子为 $-1\cdot 2^0+1=0$，而在 $m=-2$ 处，分子为 $-2\cdot 2^{-1}+1=0$，所以公式在这两处只是没有定义，而不是错误；而且与第 (i) 部分一样，它的极限（limit）值正是特殊情况的答案（$1-\ln 2\approx 0.307$ 和 $\ln 2-\frac{1}{2}\approx 0.193$）。

验算

- $m=0$：$\int_0^1 x\,\mathrm{d}x=\frac{1}{2}$，而公式给出 $\frac{0+1}{1\cdot 2}=\frac{1}{2}$。至于 $m=1$：$\int_0^1(x^2+x)\,\mathrm{d}x=\frac{1}{3}+\frac{1}{2}=\frac{5}{6}$，而公式给出 $\frac{4+1}{6}$。
- 数值上，$k=10^{-4}$ 给出 $\frac{2^k-1}{k}\approx 0.693171$，对比 $\ln 2\approx 0.693147$。
- 对所有积分使用辛普森法则（Simpson's rule），取多个 k 和 m 的值（包括 $m=-1,-2$），结果与公式吻合至 10^{-9}。把 $m=-1\pm 10^{-5}$ 代入一般公式，得 $0.30685\ldots$，与 $1-\ln 2$ 相符。

常见错误

- 在 $k=0$ 时写出 $\frac{(x+1)^0}{0}$，或没有留意到 (ii) 中 m 有两个例外值，而不是一个。
- 在 (ii) 中使用分部积分法（integration by parts），却仍然除以 $m+1$，没有察觉这在 $m=-1$ 时并不成立。
- 在 $m=-2$ 的情况中，弄错 $\int(x+1)^{-2}\,\mathrm{d}x=-(x+1)^{-1}$ 的正负号。

STEP 之外

(i) 中的推论说明 $\lim_{k\to 0}\frac{2^k-1}{k}=\ln 2$，这正是 2^k 在 $k=0$ 处的导数（derivative）。当中用到的一般事实是：若被积函数一致收敛（converge uniformly），它们的积分也会收敛。这是大学一年级分析课程的一条核心定理，而缺少一致性时，结论便不成立。同样的想法也解释了为何当 $n\to -1$ 时，$\int x^n\,\mathrm{d}x$ 会“变成”$\ln x$。Abbott 的 *Understanding Analysis* 对一致收敛与积分有仔细的论述。

详细解答

STEP I 2005，第 6 题

纯数学 · 阿波罗尼奥斯圆

> **STEP I 2005，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/05-s1-q6

关键思路

像 $AP = 2BP$ 这样的条件含有平方根，所以先把它平方：距离的平方是多项式（polynomial），而且两边 $x^2 + y^2$ 项的系数（coefficient）相同，这正是轨迹为圆的原因。在第 (ii) 部分，我们对 x-axis 上任意一对点作同样的处理；而两条形如 $x^2 + y^2 + px + q = 0$ 的方程描述同一个圆，当且仅当两者的 p 和 q 分别相同。比较系数（comparing coefficients）可得 k^2 的两个表达式，再消去 k 便得出两个结果。

解答

第 (i) 部分

两个长度都是非负的，所以 $AP = 2BP$ 等价于 $AP^2 = 4BP^2$：

$$(x-5)^2 + (y-16)^2 = 4\left[(x+4)^2 + (y-4)^2\right].$$

展开得

$$x^2 - 10x + y^2 - 32y + 281 = 4x^2 + 32x + 4y^2 - 32y + 128.$$

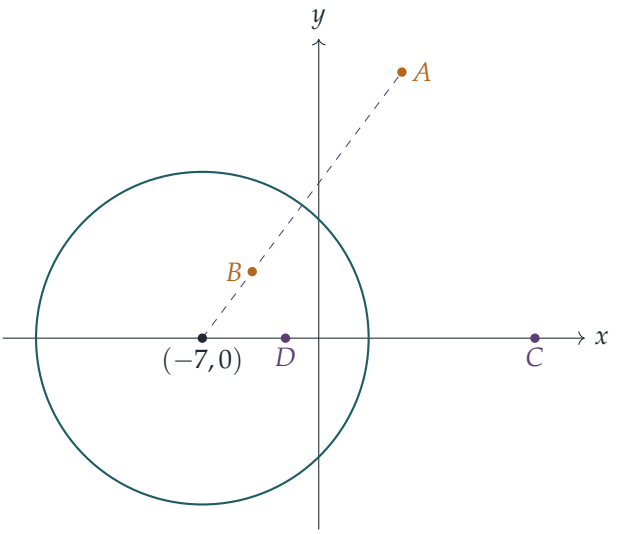
含 y 的一次项互相抵消，把所有项移到右边，得

$$3x^2 + 42x + 3y^2 - 153 = 0, \qquad \text{即} \qquad x^2 + 14x + y^2 - 51 = 0.$$

配方（completing the square），$x^2 + 14x = (x+7)^2 - 49$，所以

$$(x+7)^2 + y^2 = 100,$$

这正是所要证明的。轨迹是圆心为 $(-7, 0)$、半径为 10 的圆。



（点 $C=(13,0)$ 和 $D=(-2,0)$ 是第 (ii) 部分其中一对可能的点，此时 $k=2$。留意 A、B 与圆心同样共线（collinear）。）

第 (ii) 部分

与之前一样，$QC=kQD$ 等价于 $QC^2=k^2QD^2$。设 $Q=(x,y)$，则

$$(x-a)^2+y^2=k^2\left[(x-b)^2+y^2\right].$$

把各项移到右边并整理，

$$(k^2-1)(x^2+y^2)-2(k^2b-a)x+(k^2b^2-a^2)=0.$$

由于 $k>1$，我们有 $k^2-1\neq 0$，故可以除以它：

$$x^2+y^2-\frac{2(k^2b-a)}{k^2-1}x+\frac{k^2b^2-a^2}{k^2-1}=0.$$

? **想一想** 为甚么我们可以与 $x^2+14x+y^2-51=0$ 比较系数?

方程 $x^2+y^2+px+q=0$ 可以改写成 $(x+\frac{1}{2}p)^2+y^2=\frac{1}{4}p^2-q$。如果它是圆心为 $(-7,0)$、半径为 10 的圆，那么圆心迫使 $p=14$，半径迫使 $\frac{1}{4}p^2-q=100$，所以 $q=-51$。一个圆只有一个圆心和一个半径，所以两条方程必须有相同的 p 和相同的 q。

比较 x 的系数：

$$\frac{k^2b-a}{k^2-1}=-7 \implies k^2b-a=-7k^2+7 \implies k^2(b+7)=a+7.$$

比较常数项：

$$\frac{k^2b^2-a^2}{k^2-1}=-51 \implies k^2b^2-a^2=-51k^2+51 \implies k^2(b^2+51)=a^2+51.$$

现在 $b^2+51>0$，而且 $b+7\neq 0$：假如 $b=-7$，第一条方程便会给出 $a=-7=b$，而这是题目排除了的情况。因此我们可以分别除以这两个量，得到 k^2 的两个表达式：

$$\frac{a+7}{b+7}=k^2=\frac{a^2+51}{b^2+51},$$

这就是第一个结果。

至于第二个结果，交叉相乘：

$$(a+7)(b^2+51)=(b+7)(a^2+51).$$

展开后，常数 357 在两边互相抵消，余下

$$ab^2+51a+7b^2=a^2b+51b+7a^2.$$

把各项分组，使每一组都出现 $a-b$：

$$\begin{aligned}0&=(a^2b-ab^2)-51(a-b)+7(a^2-b^2)\\&=(a-b)\big[ab-51+7(a+b)\big].\end{aligned}$$

由于 $a\neq b$，我们可以除以 $a-b$，所以

$$ab+7a+7b=51,\quad \text{所以}\quad ab+7a+7b+49=100,\quad \text{即}\quad (a+7)(b+7)=100,$$

这正是所要证明的。

验算

- 试取 $b=-2$：这时 $b+7=5$，所以 $a+7=20$，$a=13$，得出 $k^2=20/5=4$。点 $(3,0)$ 在圆上，而 $QC=10=2\times 5=2\,QD$。点 $(-17,0)$ 也在圆上，此时 $QC=30=2\times 15$。
- 在第 (i) 部分，点 $(3,0)$ 给出 $AP=\sqrt{4+256}=\sqrt{260}$ 和 $BP=\sqrt{49+16}=\sqrt{65}$，而确实有 $\sqrt{260}=2\sqrt{65}$。
- 我们亦以数值方法验证过：对于许多满足 $(a+7)(b+7)=100$ 的点对，圆上随机选取的点都满足 $QC=kQD$；而且每一对满足第一个结果且 $a\neq b$ 的数，也都满足第二个结果。

常见错误

- 未除以 k^2-1 便比较系数：两条方程只需相差一个整体的常数倍便已描述同一个圆。
- 除以 $b+7$ 之前没有检查它是否非零。
- 遗漏因子 $a-b$：它所对应的正是题目排除了的平凡解 $a=b$。

STEP 之外

与两个定点的距离成固定比 $k \neq 1$ 的点的轨迹（locus），称为阿波罗尼奥斯圆（*circle of Apollonius*）。结果 $(a+7)(b+7)=100$ 说明 C 和 D 位于圆心的同一侧，并且 $MC \cdot MD = r^2$，其中 M 是圆心，r 是半径：C 和 D 是关于这个圆的反演点（*inverse points*）。（A 和 B 也是如此：$MA=20$，$MB=5$。）圆的反演（inversion）是几何和复分析（complex analysis）中十分有力的工具，在复分析中它就是映射 $z \mapsto 1/\bar{z}$；Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）对此有精彩的阐述。

详细解答

STEP I 2005，第 7 题

纯数学 · 裂项相消乘积

> **STEP I 2005，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/05-s1-q7

关键思路

三个乘积全部都可以裂项相消（telescope）：把每个因子写成比 $g(r+1)/g(r)$ 的形式，或写成若干个这类比的乘积，几乎所有项便会互相约去，只剩下首尾两个值。第 (i) 及第 (ii) 部分是热身，用来展示这个机制。第 (iii) 部分的工夫，在于看出那个三角因子其实暗地里正是这样的一个比。以 $\sin\frac{(2r-1)\pi}{n}$ 为公分母通分后，分子就变成一个复角（compound angle）的正弦。

解答

第 (i) 部分

把乘积逐项写出，

$$\prod_{r=1}^{n}\frac{r+1}{r}=\frac{2}{1}\times\frac{3}{2}\times\frac{4}{3}\times\cdots\times\frac{n+1}{n}.$$

每个分子都与下一个因子的分母约去，所以只剩下最后一个分子和第一个分母：

$$\prod_{r=1}^{n}\frac{r+1}{r}=\frac{n+1}{1}=n+1.$$

第 (ii) 部分

把分子因式分解（factorise），$r^2-1=(r-1)(r+1)$，并把每个因子拆成两个：

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

由于乘积的乘积就是各因子的乘积，我们可以把这两组因子分开处理（这里 $n \geq 2$）：

$$\prod_{r=2}^{n} \frac{r-1}{r} = \frac{1}{2} \times \frac{2}{3} \times \cdots \times \frac{n-1}{n} = \frac{1}{n},$$
$$\prod_{r=2}^{n} \frac{r+1}{r} = \frac{3}{2} \times \frac{4}{3} \times \cdots \times \frac{n+1}{n} = \frac{n+1}{2}.$$

两者都与第 (i) 部分一模一样地裂项相消。相乘得

$$\prod_{r=2}^{n} \frac{r^2-1}{r^2} = \frac{1}{n} \times \frac{n+1}{2} = \frac{n+1}{2n}.$$

第 (iii) 部分

设 $\alpha = \dfrac{2\pi}{n}$ 及 $\beta_r = \dfrac{(2r-1)\pi}{n}$，则第 r 个因子为 $\cos\alpha + \sin\alpha \cot\beta_r$。

? **想一想** $\cos\alpha + \sin\alpha \cot\beta$ 怎样才能写成两个相似的量之比?

利用 $\cot\beta = \cos\beta / \sin\beta$ 并通分。这时分子正是一个复角的展开式：

$$\cos\alpha + \sin\alpha \frac{\cos\beta}{\sin\beta} = \frac{\sin\beta\cos\alpha + \cos\beta\sin\alpha}{\sin\beta} = \frac{\sin(\alpha+\beta)}{\sin\beta}.$$

这一步需要 $\sin\beta_r \neq 0$，而“n 是偶数”这个条件正是在这里用上。只有当 $(2r-1)/n$ 是整数时，才会有 $\sin\beta_r = 0$。但 $2r-1$ 是奇数而 n 是偶数，奇数不可能是 n 的倍数。因此每个因子都有定义，而下文的每个分母都不为零。

现在 $\alpha + \beta_r = \dfrac{2\pi + (2r-1)\pi}{n} = \dfrac{(2r+1)\pi}{n} = \beta_{r+1}$，所以

$$\cos\alpha + \sin\alpha \cot\beta_r = \frac{\sin\beta_{r+1}}{\sin\beta_r}.$$

这正好是第 (i) 部分的形式，其中 $g(r) = \sin\beta_r$。乘积裂项相消：

$$\prod_{r=1}^{n} \frac{\sin\beta_{r+1}}{\sin\beta_r} = \frac{\sin\beta_2}{\sin\beta_1} \times \frac{\sin\beta_3}{\sin\beta_2} \times \cdots \times \frac{\sin\beta_{n+1}}{\sin\beta_n}$$
$$= \frac{\sin\beta_{n+1}}{\sin\beta_1} = \frac{\sin\left(\frac{(2n+1)\pi}{n}\right)}{\sin\frac{\pi}{n}} = \frac{\sin\left(2\pi + \frac{\pi}{n}\right)}{\sin\frac{\pi}{n}}.$$

由于 sin 的周期（period）为 2π，分子等于分母，故

$$\prod_{r=1}^{n} \left(\cos\tfrac{2\pi}{n} + \sin\tfrac{2\pi}{n} \cot\tfrac{(2r-1)\pi}{n}\right) = 1.$$

验算

- 第 (i) 部分，取 $n = 3$：$2 \cdot \frac{3}{2} \cdot \frac{4}{3} = 4 = n + 1$。
- 第 (ii) 部分，取 $n = 2$：唯一的因子是 $\frac{3}{4}$，而 $\frac{n+1}{2n} = \frac{3}{4}$。当 $n \to \infty$ 时，乘积趋向 $\frac{1}{2}$，这是一个著名的无穷乘积（infinite product）。
- 第 (iii) 部分，取 $n = 4$：$\cos \frac{\pi}{2} = 0$ 及 $\sin \frac{\pi}{2} = 1$，所以各因子为 $\cot \frac{\pi}{4}$、$\cot \frac{3\pi}{4}$、$\cot \frac{5\pi}{4}$ 和 $\cot \frac{7\pi}{4}$，即 1、-1、1 和 -1，乘积为 1。我们亦以数值计算验证过，对 60 以内的每个偶数 n，乘积都等于 1（误差在 10^{-8} 以内）。
- 若 n 是奇数，$r = \frac{n+1}{2}$ 的那个因子涉及 $\cot \pi$，而它没有定义。这就解释了题目为何要有这个条件。

常见错误

- 在第 (ii) 部分，由 $r = 1$ 开始相消：下限是 2，所以最先的两个因子是 $\frac{1}{2}$ 和 $\frac{3}{2}$。
- 在第 (iii) 部分，停在 $\sin \frac{(2n+1)\pi}{n} \big/ \sin \frac{\pi}{n}$ 就算，没有利用周期性把它化简为 1。
- 没有解释“n 是偶数”用在哪里。正是这个条件，令每个余切（cotangent）以及每个分母都有妥善的定义。

STEP 之外

乘积的裂项相消，与你在部分分式（partial fractions）中遇到的裂项求和，是乘法与加法上的一对近亲。取对数（logarithm）就能把其中一个变成另一个。第 (ii) 部分是无穷乘积 $\prod_{r \geq 2}(1 - 1/r^2) = \frac{1}{2}$ 的有限版本。像这样的无穷乘积，以及欧拉（Euler）乘积

$$\sin \pi x = \pi x \prod_{r \geq 1} \left(1 - \frac{x^2}{r^2}\right),$$

要到大学第一门分析（analysis）课程才会严谨地研究；在那里，乘积的收敛（convergence）是通过对数之和的收敛来定义的。Abbott 的 *Understanding Analysis* 是建立所需收敛概念的好起点。

详细解答

STEP I 2005，第 8 题

纯数学 · 一阶微分方程

STEP I 2005，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/05-s1-q8

关键思路

开头的恒等式(identity)其实是一道配方,教我们辨认形如 $2xy\,\frac{\mathrm{d}y}{\mathrm{d}x} = ky^2 + (\text{只含}x\text{ 的式子})$ 的方程。方程一旦写成这个形状，代换（substitution）$y^2 = x^k f(x)$ 便会把 y^2 项消去，剩下 $f'(x)$ 等于一个 x 的函数，直接积分即可。真正要花工夫的，只有选取 k，以及在第 (ii) 部分把方程两边同除，使左边变成 $2xy\,\frac{\mathrm{d}y}{\mathrm{d}x}$。

解答

恒等式

把 $y^2 = x^k f(x)$ 对 x 作隐函数求导（implicit differentiation），右边用积法则（product rule）：

$$2y\frac{\mathrm{d}y}{\mathrm{d}x} = kx^{k-1}f(x) + x^k\frac{\mathrm{d}f}{\mathrm{d}x}.$$

两边乘以 x，

$$2xy\frac{\mathrm{d}y}{\mathrm{d}x} = k\,x^k f(x) + x^{k+1}\frac{\mathrm{d}f}{\mathrm{d}x} = ky^2 + x^{k+1}\frac{\mathrm{d}f}{\mathrm{d}x},$$

因为 $x^k f(x) = y^2$。这就是所求的结果。

第 (i) 部分

取 $k = 1$，写成 $y^2 = xf(x)$，恒等式便给出 $2xy\,\frac{\mathrm{d}y}{\mathrm{d}x} = y^2 + x^2 f'(x)$。与题目给出的方程比较，$y^2$ 项互相吻合，所以方程成立当且仅当

$$x^2 f'(x) = x^2 - 1, \qquad \text{即} \qquad f'(x) = 1 - \frac{1}{x^2} \quad (x \neq 0).$$

积分得 $f(x) = x + \dfrac{1}{x} + c$，所以

$$y^2 = xf(x) = x^2 + cx + 1.$$

由条件“当 $x = 1$ 时 $y = 2$”得 $4 = 1 + c + 1$，所以 $c = 2$，而

$$y^2 = x^2 + 2x + 1 = (x+1)^2.$$

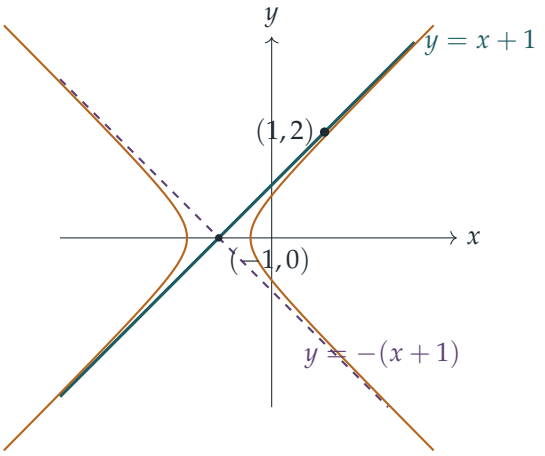
因此 $y = x + 1$ 或 $y = -(x+1)$；只有 $y = x + 1$ 通过 $(1, 2)$，所以解是

$$y = x + 1.$$

（代回原式：$2x(x+1)\cdot 1 = 2x^2 + 2x$，而 $(x+1)^2 + x^2 - 1 = 2x^2 + 2x$，所以它对每一个 x 都满足方程，包括我们曾经用来作除数的 $x = 0$。）

想一想 答案只不过是 $y = x + 1$，那么“从几何角度描述”究竟想我们写甚么？

解曲线是一条直线：它的斜率(gradient)为 1，通过 $(-1, 0)$ 和 $(0, 1)$，与 x-axis 所在的横轴成 45° 角。值得一提的是，这个结果其实有点出人意表。通解(general solution) $y^2 - x^2 - cx = 1$ 是一族圆锥曲线（conics）；配方（completing the square）后得 $y^2 - (x + \frac{c}{2})^2 = 1 - \frac{c^2}{4}$，这是一条等轴双曲线（rectangular hyperbola），中心在 $(-\frac{c}{2}, 0)$，渐近线（asymptotes）为 $y = \pm(x + \frac{c}{2})$。初始条件（initial condition）选出了 $c = 2$，这正是右边等于零的情况，双曲线退化（degenerate）成它自己的一对渐近线 $y = \pm(x+1)$。我们的解就是这对直线中的 $y = x + 1$ 那一支。



实线是所求的解；虚线是 $y^2 = (x+1)^2$ 的另一半。琥珀色的曲线勾画出这族曲线中非退化成员的形状：一条等轴双曲线，以类似这样的一对直线为渐近线，并逐渐向它们靠近。

第 (ii) 部分

要使用恒等式，左边必须是 $2xy\,\frac{\mathrm{d}y}{\mathrm{d}x}$，所以把方程除以 x（$\ln x$ 项已限定 $x>0$，所以这一步并无问题）：

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

y^2 的系数是 -1，所以取 $k=-1$，即 $y^2 = f(x)/x$。这时 $x^{k+1} = x^0 = 1$，恒等式变成 $2xy\,\frac{\mathrm{d}y}{\mathrm{d}x} = -y^2 + f'(x)$。比较两式，

$$f'(x) = \frac{2\ln x}{x}.$$

由于 $\frac{\mathrm{d}}{\mathrm{d}x}(\ln x)^2 = \frac{2\ln x}{x}$，我们得到 $f(x) = (\ln x)^2 + c$，于是

$$y^2 = \frac{(\ln x)^2 + c}{x}.$$

当 $x=1$ 时 $y=1$：$1 = (0+c)/1$，所以 $c=1$。这里取正平方根，因为在 $x=1$ 处 $y=1>0$，而且 y 永不为零（分子 $1+(\ln x)^2$ 最少是 1），所以解是

$$y = \sqrt{\frac{1+(\ln x)^2}{x}}, \qquad x>0.$$

验算

- 在 (i) 中，直线 $y=x+1$ 可以如上文般直接代入方程验证；由 $(1,2)$ 出发对方程作数值积分（龙格—库塔法，Runge–Kutta），在 $x=2$ 得 $y=3.000$，在 $x=5$ 得 $y=6.000$。
- 在 (ii) 中，$y^2 = (1+(\ln x)^2)/x$ 给出 $y(3)\approx 0.8577$ 和 $y(0.2)\approx 4.2369$；由 $(1,1)$ 出发对微分方程作数值积分，结果在小数点后多个位都吻合。
- 恒等式本身：试取 $k=0$，它便简化成 $2xy\,y' = xf'$，这只不过是 $(y^2)' = f'$ 乘以 x。

常见错误

- 在 (ii) 中出于习惯而用了 $k=1$。除以 x 之后，y^2 项的正负号才是决定 $k=-1$ 的关键。
- 在 (ii) 中忘记除以 x，结果尝试拿 $2x^2y\,y'$ 去配对恒等式。
- 把 (i) 的答案停留在 $y^2=(x+1)^2$，没有选出通过 $(1,2)$ 的那一支。

STEP 之外

这里的代换其实是一个标准技巧的变装版本：设 $u = y^2$，便把 $2xy\,y' = ky^2 + g(x)$ 化成线性（linear）方程 $xu' - ku = g(x)$，而 x^{-k} 就是它的积分因子（integrating factor）。经幂代换后变成线性的方程称为伯努利方程（Bernoulli equations），是大学常微分方程（ODE）课程中最早遇到的非线性方程之一。Boyce 与 DiPrima 的 *Elementary Differential Equations*（一本常微分方程入门教科书）把它们与积分因子和恰当方程（exact equations）放在一起讨论。

详细解答

STEP I 2005，第 9 题

力学 · 刚体平衡

STEP I 2005，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/05-s1-q9

关键思路

整道题的关键在于一个未知量：这根非均匀杆的质心（centre of mass）G 位于何处。头两个装置各自给出以该位置表示的 T，把两者等同起来，便可同时定出 G 和 T。之后，倾斜的装置就是一道标准的刚体（rigid body）问题：先对未知接触力所作用的那一点取力矩（moments），然后分解（resolve）各力。

解答

我们把杆看作刚体，其全部重量 W 作用于质心 G，并把绳看作轻而不可伸长（light and inextensible）。设 $AG = x$，则 $GB = 3l - x$。

求 G 的位置并证明 $5T = 2W$

*两端各系一绳。*对 B 取力矩，可消去 B 处未知的张力（tension）：

$$T \cdot 3l = W(3l - x).$$

*支点距离 A 为 l，绳系于 B。*设 P 为支点（pivot），则 $PB = 2l$。对 P 取力矩，可消去未知的支点反作用力。B 处的绳要把杆拉住，G 必须位于 P 靠 B 的一侧，故 $PG = x - l$，而

$$T \cdot 2l = W(x - l).$$

消去 T（把第一条方程乘以 2，第二条乘以 3，然后比较）：

$$2W(3l - x) = 3W(x - l) \quad \Longrightarrow \quad 9l = 5x \quad \Longrightarrow \quad x = \tfrac{9}{5}l.$$

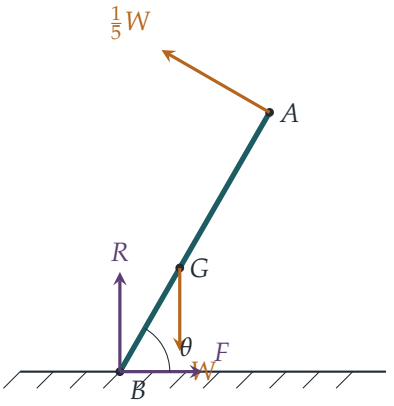
于是由第二条方程得 $2lT = W \cdot \frac{4}{5}l$，故 $T = \frac{2}{5}W$，即

$$5T = 2W,$$

正如所求。留意 $x = \frac{9}{5}l > l$，这与我们假设 G 在支点之外是一致的。因此 G 与 B 的距离为 $GB = 3l - \frac{9}{5}l = \frac{6}{5}l$。

杆搁在地面上

现在 B 在粗糙的地面上，杆与水平成 θ 角，而 A 处的绳垂直于杆，张力为 $\frac{1}{2}T = \frac{1}{5}W$。在 B 处有向上的法向反作用力（normal reaction）R 和沿地面的摩擦力（friction）F。



对 B 取力矩。这样可消去 R 和 F。绳的拉力垂直于杆，作用点与 B 相距 $3l$，故其力矩为 $\frac{1}{5}W \cdot 3l$。重量沿杆相距 $\frac{6}{5}l$ 之处铅垂作用，故其作用线与 B 的垂直距离为 $\frac{6}{5}l\cos\theta$。因此

$$\tfrac{1}{5}W \cdot 3l = W \cdot \tfrac{6}{5}l\cos\theta \quad\Longrightarrow\quad \cos\theta = \tfrac{1}{2},$$

所以 $\theta = 60°$（即 $\frac{\pi}{3}$）。

想一想 绳的张力的各个分量指向哪一方?

杆与水平成 θ 角，而绳与杆成直角，所以绳与铅垂线成 θ 角。因此其张力 $\frac{1}{5}W$ 的铅垂分量为 $\frac{1}{5}W\cos\theta$（向上），水平分量为 $\frac{1}{5}W\sin\theta$，指向越过 B 的后方，即背离 A 所在的一侧。摩擦力必须抵销这个水平拉力，所以它在 B 处指向 A 所在的一侧，如图所示。

沿铅垂方向分解：

$$R + \tfrac{1}{5}W\cos 60° = W \quad\Longrightarrow\quad R = W - \tfrac{1}{10}W = \tfrac{9}{10}W.$$

沿水平方向分解：

$$F = \tfrac{1}{5}W\sin 60° = \tfrac{\sqrt{3}}{10}W.$$

只要 $F \le \mu R$，杆便不会滑动，即

$$\mu \ge \frac{F}{R} = \frac{\sqrt{3}/10}{9/10} = \frac{\sqrt{3}}{9}.$$

所以能防止滑动的最小摩擦系数（coefficient of friction）为 $\mu = \dfrac{\sqrt{3}}{9} = \dfrac{1}{3\sqrt{3}} \approx 0.192$（等号成立时，杆处于极限平衡（limiting equilibrium））。

验算

- 取 $x = \frac{9}{5}l$，双绳装置给出 $T = W \cdot \frac{6/5}{3} = \frac{2}{5}W$，而支点装置给出 $T = W \cdot \frac{4/5}{2} = \frac{2}{5}W$：两者相同，理应如此。
- 在倾斜的位置，再对 A 取力矩（取 $W = l = 1$，$\theta = 60^\circ$，$R = 0.9$，$F \approx 0.1732$），所得的合力矩同样为零，这一并验证了 R 和 F。
- 数值上 $\mu_{\min} = \sqrt{3}/9 \approx 0.19245$，这已由电脑建立各力的向量并检查全部三条平衡方程加以确认。

常见错误

- 以 $\frac{6}{5}l\sin\theta$ 作为重量的力臂。由 B 至重量作用线的水平距离是 $\frac{6}{5}l\cos\theta$。
- 忘记垂直于杆的绳有铅垂分量，而取 $R = W$。这样会得出 $\mu = \sqrt{3}/10$，数值偏小。
- 假设杆是均匀的，以为 G 在中点。第一部分的全部目的正是要找出 G 的位置。

STEP 之外

利用平衡时的量度来确定质心的位置，是一个强大想法的静力学版本：一个力系对某一点的力矩决定了该力系，只差一个合力。到了大学，这会发展成刚体静力学以至动力学的研究，其中质心和转动惯量（moment of inertia）就是支配物体运动的两个数。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics* 讲述力矩（torque）与刚体，并附有大量这类风格而难度更高的题目。

详细解答

STEP I 2005，第 10 题

力学 · 碰撞与恢复系数

STEP I 2005，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：

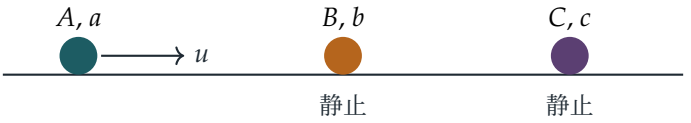
step.maths.org/questions/05-s1-q10

关键思路

每一次碰撞都是两个光滑质点（particle）之间的直接碰撞（direct impact），因此可得两条方程：沿直线方向的动量守恒（conservation of momentum），以及牛顿实验定律（Newton's experimental law）（分离速率 $= e\times$ 接近速率）。只要对一般的一对质点解一次这两条方程，所得公式便可在第二次碰撞中重复使用。至于各个特殊情况，真正的问题是：两次碰撞之后，事情是否就此结束？题目要求的是最终速度，所以必须检查没有任何质点会追上前面的那一个。

解答

我们把 A、B、C 视为在光滑平面上沿同一直线运动的质点，因此没有水平方向的外力作用，每次碰撞中沿直线方向的动量都守恒。取 u 的方向为正。设 p 为 A 与 B 碰撞后的速度，q 为 B 与 C 碰撞后的速度。



第一次碰撞：求 v

碰撞前 A 的速度为 u，而 B 静止；碰撞后两者的速度分别为 p 和 v。由动量守恒和牛顿定律得

$$au = ap + bv, \qquad v - p = eu.$$

把 $p = v - eu$ 代入第一条方程，得 $au = av - aeu + bv$，所以

$$v = \frac{a(1+e)u}{a+b}, \qquad p = v - eu = \frac{(a-eb)u}{a+b}.$$

留意 $v > 0$，所以 B 确实朝着 C 运动。

第二次碰撞：求 w

这与刚才的情况相同，只是换成 B（质量 b，速率 v）撞向 C（质量 c，静止）。因此可以重用上面的公式，作代换 $a \to b$、$b \to c$、$u \to v$：

$$w = \frac{b(1+e)v}{b+c}, \qquad q = \frac{(b-ec)v}{b+c}.$$

代入 v 的表达式，

$$w = \frac{b(1+e)}{b+c} \cdot \frac{a(1+e)u}{a+b} = \frac{abu(1+e)^2}{(a+b)(b+c)},$$

正如所求。

这两次碰撞之后，各质点的速度为 p、q、w。只有当 $p > q$（A 追上 B）或 $q > w$（B 追上 C）时，才会再有碰撞；由牛顿定律，$q - w = -ev < 0$，所以后者永不发生，我们只需比较 p 和 q。

> ? **想一想** 为甚么不能在两次碰撞之后就停下来，直接写出 p、q、w 作为答案?

因为题目问的是最终速度。假如之后 A 仍然比 B 快，它便会再次撞上 B，一切都会随之改变。所以在每种情况下，我们都必须检查 $p \le q$。

第 (i) 部分

这里 $a = eb$ 及 $b = ec$（其中 $e > 0$，因为质量为正数）。于是 $a - eb = 0$ 及 $b - ec = 0$，所以

$$p = 0, \qquad q = 0.$$

至于 w，利用 $a + b = b(1+e)$ 及 $b + c = b + b/e = b(1+e)/e$：

$$w = \frac{eb \cdot b\,u(1+e)^2}{b(1+e) \cdot b(1+e)/e} = e^2 u.$$

A 和 B 都静止，而 C 向前离去，所以不会再有碰撞。最终速度为

$$A:\ 0, \qquad B:\ 0, \qquad C:\ e^2 u.$$

每次碰撞都把全部动量传给下一个质点，就像一个牛顿摆（Newton's cradle），只是后面的质点较轻。

第 (ii) 部分

现在 $b = ea$ 及 $c = eb = e^2a$。于是

$$v = \frac{a(1+e)u}{a(1+e)} = u, \qquad p = \frac{(a-e^2a)u}{a(1+e)} = (1-e)u,$$

当中用了 $1-e^2 = (1-e)(1+e)$。在第二次碰撞中，质量 b、c 之比与 a、b 之比相同，而 B 以速率 u 到达，所以它完全是第一次碰撞的翻版：

$$w = u, \qquad q = (1-e)u.$$

（直接计算：$w = \dfrac{a \cdot ea \cdot u(1+e)^2}{a(1+e) \cdot ea(1+e)} = u$。）

现在 $p = q$，所以 A 永远追不上 B；而当 $e > 0$ 时 $q = (1-e)u < u = w$，所以 B 永远追不上 C。最终速度为

$$A:\ (1-e)u, \qquad B:\ (1-e)u, \qquad C:\ u,$$

全部沿 A 原来的运动方向。

验算

- (ii) 的动量：$a(1-e)u + ea(1-e)u + e^2a\,u = au$。(i) 的动量：$c \cdot e^2u = e^2c\,u = au$。
- 在 (ii) 中取 $e = 1$（完全弹性，perfectly elastic），A 和 B 最终静止，而 C 以 u 离去；动能（kinetic energy）守恒：$\frac{1}{2}au^2 = \frac{1}{2}cu^2$ 需要 $c = a$，这正是 $c = e^2a$ 在 $e = 1$ 时的情况。
- 一个数值例子：取 $e = \frac{1}{2}$，质量为 $1, 2, 4$ 及 $u = 1$（第 (i) 部分），最终速度为 $0, 0, \frac{1}{4}$；质量为 $4, 2, 1$（第 (ii) 部分）则得 $\frac{1}{2}, \frac{1}{2}, 1$。我们以精确算术，对数以千计随机选取的质量和 e 值，并模拟其后可能出现的任何碰撞，验证了这些结果以及 v 和 w 的一般公式。

常见错误

- 写牛顿定律时把速度的次序弄反：正确的是碰撞后的（前方质点的速率 − 后方质点的速率）$= e \times$ 碰撞前的（后方 − 前方）。
- 直接写出两次碰撞后的速度，而没有检查 A 会否再次追上 B。

STEP 之外

连串碰撞是描述动量和能量如何在物料中传递的最简单模型，而第 (i) 部分的质量比经过特意选取，使全部动量最终都落在最后一个质点上。若把质量的次序倒过来，重的在

后、轻的在前，并且 e 接近 1，最后一个质点便会比第一个质点运动得更快：这正是那个玩具背后的原理——把一个小球叠在一个大球上，让两者一同落下，小球便会高高弹上半空。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）以质心参考系（centre-of-mass frame）处理碰撞，令 v 和 w 这类公式变得几乎一目了然。

详细解答

STEP I 2005，第 11 题

力学 · 向量运动学

STEP I 2005，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/05-s1-q11

关键思路

这道题由始至终都在考两个向量(vector)$\mathbf{r}$ 与 $\mathbf{v} = \dot{\mathbf{r}}$ 之间的相对位置。垂直(perpendicular)即纯量积（scalar product）$\mathbf{r} \cdot \mathbf{v}$ 等于零；平行（parallel）即“交叉”组合 $xv_y - yv_x$ 等于零。只要写出 $\sin 2t = 2\sin t \cos t$ 和 $\cos 2t = 1 - 2\sin^2 t$，这两个式子都可以化简得十分整齐。至于求距离和绘画草图，把所有式子都用 $\cos t$ 表示，或者干脆消去 t，会方便得多。

解答

设 $\mathbf{r} = x\,\mathbf{i} + y\,\mathbf{j}$，其中

$$x = \sin 2t = 2\sin t \cos t, \qquad y = 2\cos t, \qquad 0 \le t < 2\pi.$$

求导后，得速度（velocity）为

$$\mathbf{v} = \dot{\mathbf{r}} = (2\cos 2t)\,\mathbf{i} - (2\sin t)\,\mathbf{j}.$$

第 (i) 部分

当 $x = 0$ 且 $y = 0$ 时，质点（particle）位于原点。条件 $y = 0$ 即 $\cos t = 0$，所以在给定范围内 $t = \frac{\pi}{2}$ 或 $t = \frac{3\pi}{2}$。在这两个时刻，$x = 2\sin t \cos t = 0$ 自动成立。因此质点恰好在 $t = \frac{\pi}{2}$ 和 $t = \frac{3\pi}{2}$ 这两个时刻位于原点，其他时刻都不在原点，这正是要证明的结果。

第 (ii) 部分

当 $\mathbf{r}\cdot\mathbf{v}=0$ 时，速度与位移（displacement）垂直（两个向量都须非零，“垂直”才有意义）。现在

$$\begin{aligned}\mathbf{r}\cdot\mathbf{v} &= 2\sin 2t\cos 2t - 4\sin t\cos t\\ &= 2\sin 2t\cos 2t - 2\sin 2t\\ &= 2\sin 2t\,(\cos 2t-1)\\ &= -4\sin 2t\,\sin^2 t,\end{aligned}$$

其中用了 $\cos 2t-1=-2\sin^2 t$。当 $\sin 2t=0$ 或 $\sin t=0$ 时，此式等于零，即

$$t=0,\ \tfrac{\pi}{2},\ \pi,\ \tfrac{3\pi}{2}.$$

> **想一想** $t=\frac{\pi}{2}$ 和 $t=\frac{3\pi}{2}$ 真的应该列入答案吗?

不应该。由第 (i) 部分可知，在这两个时刻 $\mathbf{r}=\mathbf{0}$，而零向量没有方向，所以说它与任何向量垂直都没有实质意义；纯量积等于零，只是出于一个平凡的原因。在 $t=0$ 时，$\mathbf{r}=2\mathbf{j}$ 而 $\mathbf{v}=2\mathbf{i}$；在 $t=\pi$ 时，$\mathbf{r}=-2\mathbf{j}$ 而 $\mathbf{v}=2\mathbf{i}$：这些才是真正互相垂直的非零向量。因此，速度与位移垂直的时刻是

$$t=0 \quad \text{和} \quad t=\pi.$$

第 (iii) 部分

首先，$\mathbf{v}$ 永不为零：若 $\mathbf{v}=\mathbf{0}$，则必须 $\sin t=0$，但这时 $\cos 2t=1-2\sin^2 t=1\neq 0$。

两个非零向量 (x,y) 与 (v_x,v_y) 平行，当且仅当 $xv_y-yv_x=0$。这里

$$\begin{aligned}xv_y-yv_x &= \sin 2t\,(-2\sin t)-2\cos t\,(2\cos 2t)\\ &= -4\sin^2 t\cos t-4\cos t\,(1-2\sin^2 t)\\ &= -4\cos t\,(1-\sin^2 t)\\ &= -4\cos^3 t.\end{aligned}$$

此式只在 $\cos t=0$ 时等于零，而由第 (i) 部分可知，这只在质点位于原点时才会发生。所以只要质点不在原点，就有 $\cos t\neq 0$，这个组合不等于零，速度亦不与位移平行。证明至此完成。

第 (iv) 部分

设 $c=\cos^2 t$，则 $0\le c\le 1$。于是

$$|\mathbf{r}|^2=\sin^2 2t+4\cos^2 t=4\sin^2 t\cos^2 t+4\cos^2 t=4c(1-c)+4c=4c(2-c).$$

二次式 $4c(2-c)$ 的顶点在 $c=1$，并在 $[0,1]$ 上递增，所以它在这个区间上的最大值在 $c=1$ 处取得，这时 $|\mathbf{r}|^2=4$。因此质点与原点的最大距离是

$$|\mathbf{r}|_{\max}=2,$$

在 $\cos^2 t=1$ 时达到，即 $t=0$ 和 $t=\pi$，对应的点是 $(0,\pm2)$。这正好是第 (ii) 部分求得的时刻，令人安心：距离达到最大值时，速度的径向分量（radial component）必定为零。

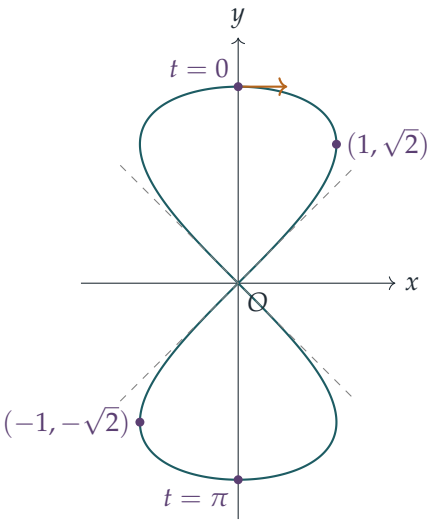
至于草图，先消去 t。由于 $x=y\sin t$ 及 $\sin^2 t=1-\cos^2 t=1-\frac{1}{4}y^2$，

$$x^2=y^2\Big(1-\frac{y^2}{4}\Big),\qquad \text{即}\qquad 4x^2=y^2(4-y^2).$$

草图上应显示以下特征：

- 曲线对两条坐标轴都对称，且 $-2\le y\le 2$。
- $|x|=|\sin 2t|\le 1$，当 $y=\pm\sqrt{2}$ 时 $|x|=1$。
- 在原点附近 $4x^2\approx 4y^2$，所以曲线在该处沿直线 $y=\pm x$ 自相交。这与 $t=\frac{\pi}{2}$ 时的速度 $-2\mathbf{i}-2\mathbf{j}$ 及 $t=\frac{3\pi}{2}$ 时的速度 $-2\mathbf{i}+2\mathbf{j}$ 吻合。
- 在 $(0,\pm2)$ 处路径是水平的，正如第 (ii) 部分所示。

随着 t 增加：质点从 $(0,2)$ 出发，沿 $+\mathbf{i}$ 方向运动，顺着上方环圈的右边下行至原点，再绕下方环圈一周（先走左边）回到原点，最后沿上方环圈的左边上行。整条路径是一个 8 字形。



验算

- 在 $t=\frac{\pi}{4}$ 时：$\mathbf{r}=\mathbf{i}+\sqrt{2}\,\mathbf{j}$，满足 $4x^2=4=y^2(4-y^2)=2\cdot 2$，而 $|\mathbf{r}|=\sqrt{3}<2$，与 $4c(2-c)=4\cdot\frac{1}{2}\cdot\frac{3}{2}=3$ 一致。
- 在 $[0,2\pi)$ 上对 t 作细密取样，得最大距离为 2.000，距离接近零的情况只出现在 $t\approx 1.57$ 和 4.71，并且在随机选取的时刻验证了恒等式 $\mathbf{r}\cdot\mathbf{v}=-4\sin 2t\sin^2 t$ 和 $xv_y-yv_x=-4\cos^3 t$。最小速率约为 1.32，所以 $\mathbf{v}$ 永不为零。

常见错误

- 在第 (ii) 部分把 $t=\frac{\pi}{2}$ 和 $\frac{3\pi}{2}$ 列入答案：纯量积在这两个时刻等于零，只是因为 $\mathbf{r}=\mathbf{0}$。
- 在第 (iii) 部分忘记验证 $\mathbf{v}$ 永不为零；零速度会与任何向量“平行”。
- 把两个环圈画成左右并排：这个 8 字是直立的，两个环圈分别在原点的上方和下方，因为 y 的范围是 $[-2,2]$，而 x 的范围只有 $[-1,1]$。

STEP 之外

$\mathbf{r}\cdot\mathbf{v}=\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{r}|^2$ 和 xv_y-yv_x 这两个量，分别是运动的径向部分和（每单位质量计）对原点的角动量（angular momentum）；在极坐标（polar coordinates）中，后者就是 $r^2\dot{\theta}$。第 (iii) 部分说明角动量只在原点处为零，所以质点从不沿直线直接趋向或远离 O。这条曲线本身就是 Gerono 双纽线（lemniscate of Gerono），是众多经典 8 字形曲线之一。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）正是以这种思路阐述极坐标和角动量。

详细解答

STEP I 2005，第 12 题

概率与统计 · 温氏图与界限

STEP I 2005，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/05-s1-q12

关键思路

凡是问“哪个范围的数值才不会自相矛盾?”的题目，其实都是关于一幅温氏图（Venn diagram）：图中各个区域的值未知，但必须全部非负。先用尽量少的自由未知数表示每一个区域；这样，每个“此区域 ≥ 0”的条件就变成一个界限，把所有界限合起来便得到所求范围。在第 (ii) 部分，我们用一个计数论证（把三个单项概率相加时，每名巫师被数了多少次），毋须知道个别区域的值，便可求出“恰好两件”的概率。

解答

第 (i) 部分

设 S 为哈比人抽烟斗的事件（event），H 为他戴帽子的事件。令 $t = \mathrm{P}(S \cap H)$。温氏图的四个区域是

$$\mathrm{P}(S \cap H) = t, \qquad \mathrm{P}(S \cap H') = 0.7 - t,$$
$$\mathrm{P}(S' \cap H) = 0.4 - t, \qquad \mathrm{P}(S' \cap H') = 1 - (0.7 + 0.4 - t) = t - 0.1.$$

只要所选的 t 令四者皆为非负，就得到一个真正的概率分布（probability distribution），反之亦然。真正起约束作用的条件是 $t \geq 0.1$（来自最后一个区域）和 $t \leq 0.4$（来自第三个区域）；其余条件随之成立。所以 $0.1 \leq t \leq 0.4$，而由于 $p = 0.7 - t$，

$$0.3 \leq p \leq 0.6.$$

两个端点都可以取到：当抽烟斗而又戴帽子的哈比人尽量少时（每个哈比人至少做两件事的其中一件），$p = 0.6$；当每个戴帽子的哈比人都抽烟斗时，$p = 0.3$。

第 (ii) 部分：恰好两件

设 H、C、R 分别为巫师戴帽子、披斗篷和戴戒指的事件。设 a、b 分别为巫师在这三件物品中恰好穿戴一件和恰好穿戴两件的概率。已知一件也没有的概率是 0.05，三件齐备的概率是 0.1，所以

$$0.05 + a + b + 0.1 = 1 \quad \Longrightarrow \quad a + b = 0.85.$$

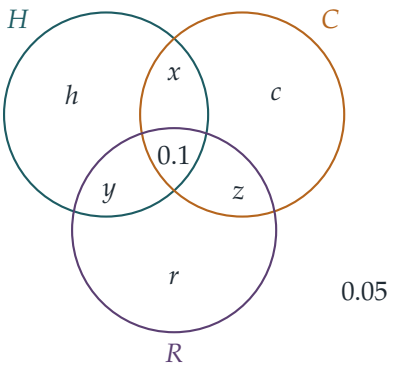
现在把三个已知概率相加：$\mathrm{P}(H) + \mathrm{P}(C) + \mathrm{P}(R) = 0.7 + 0.8 + 0.4 = 1.9$。在这个和之中，恰好穿戴 k 件物品的巫师被数了 k 次，所以

$$a + 2b + 3(0.1) = 1.9 \quad \Longrightarrow \quad a + 2b = 1.6.$$

两式相减，得 $b = 0.75$（而 $a = 0.1$）。巫师恰好穿戴其中两件物品的概率是 $\frac{3}{4}$。

第 (ii) 部分：q 的范围

如图所示，为圆内的七个区域标上记号：h、c、r 代表恰好一件，x、y、z 代表恰好两件，中央是 0.1，圆外是 0.05。



由 H、C、R 各自的总和，以及第 (ii) 部分至今所得的结果，

$$h + x + y = 0.6, \qquad c + x + z = 0.7,$$
$$r + y + z = 0.3, \qquad x + y + z = 0.75.$$

利用最后一条方程，从首三条方程中各消去一对未知数，

$$h = z - 0.15, \qquad c = y - 0.05, \qquad r = x - 0.45.$$

（验算一下：$h + c + r = 0.75 - 0.65 = 0.1 = a$。）所以整个分布完全由满足 $x + y + z = 0.75$ 的 x、y、z 决定，而它是一个真正的分布，当且仅当六个区域全部非负：

$$x \geq 0.45, \qquad y \geq 0.05, \qquad z \geq 0.15.$$

想一想 x 最大可以是多少?

由于 $x = 0.75 - y - z$，而 $y + z \geq 0.05 + 0.15 = 0.2$，所以 $x \leq 0.55$，等号在 $y = 0.05$、$z = 0.15$ 时成立。最小值 $x = 0.45$ 也可以取到，例如取 $y = 0.1$、$z = 0.2$。两者之间的每一个值同样可以取到，因为我们可以连续地改变 y 和 z。因此

$$0.45 \leq x \leq 0.55.$$

最后，披着斗篷时戴帽子但不戴戒指，正好就是区域 x，所以

$$q = \mathrm{P}(H \cap R' \mid C) = \frac{\mathrm{P}(H \cap C \cap R')}{\mathrm{P}(C)} = \frac{x}{0.8}, \qquad \text{由此得} \qquad \frac{9}{16} \leq q \leq \frac{11}{16}.$$

验算

- 在极端情况 $x = 0.55$、$y = 0.05$、$z = 0.15$ 下，我们得到 $h = 0$、$c = 0$、$r = 0.1$；八个区域之和为 1，并给出 $\mathrm{P}(H) = 0.7$、$\mathrm{P}(C) = 0.8$、$\mathrm{P}(R) = 0.4$，符合要求。
- 我们以 0.01 为步长，列举了八个区域所有与数据相符的赋值：每一个都满足 $x + y + z = 0.75$，而 q 的取值由 0.5625 至 0.6875。同样地，在 (i) 中，p 的取值由 0.3 至 0.6。

常见错误

- 在 (i) 中，忘记了“两者皆非”这个区域，因而得到下界 $t \geq 0$ 而不是 $t \geq 0.1$，结果得出错误的上限 $p \leq 0.7$。
- 在 (ii) 中，除以 1 而不是除以 $\mathrm{P}(C) = 0.8$：“已知他披着斗篷”这个条件会改变分母。
- 把 H、C、R 当作独立（independent）事件。题目从没有说它们独立，而事实上 $\mathrm{P}(H \cap C \cap R) = 0.1 \neq 0.7 \times 0.8 \times 0.4$。

STEP 之外

某个量是未知概率的线性函数，而这些概率须满足若干线性方程及非负条件，要求这个量的取值范围，其实就是一个小型的*线性规划*（linear programme）问题：极值总是出现在可行域（feasible region）的角点，这正是我们的极值都来自把某些区域设为零的原因。同样的想法可得出第 (i) 部分所用的弗雷歇（Fréchet）界限 $\max(0, \mathrm{P}(A) + \mathrm{P}(B) - 1) \leq \mathrm{P}(A \cap B) \leq \min(\mathrm{P}(A), \mathrm{P}(B))$。想进一步了解容斥原理（inclusion–exclusion）和条件概率（conditional probability），可阅读 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门教材）。

详细解答

STEP I 2005，第 13 题

概率与统计 • 对称分布

STEP I 2005，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/05-s1-q13

关键思路

题目除了给出两个累积概率（cumulative probability）和对称性（symmetry）之外，对分布的形状只字未提，所以一切都要由这些资料推导出来。关于 μ 对称，还附带送来两个事实：μ 的两侧各占一半概率，而且 $\mu - k\sigma$ 以下的尾部与 $\mu + k\sigma$ 以上的尾部概率相同。第 (ii) 部分的诀窍，是留意到题中提及的每一个容量（495、500、505 ml），对其中一种牛奶而言，都正好落在 μ、$\mu \pm \frac{1}{2}\sigma$ 或 $\mu + \sigma$ 的位置。

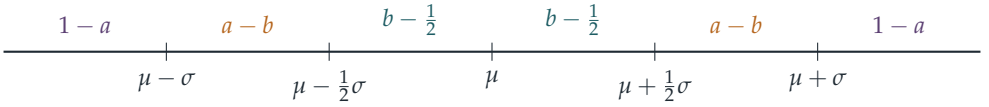
解答

全题把 X 视为连续（continuous）随机变量（它描述的是容量），因此单一数值的概率为零，写 $<$ 还是 $\leq$ 并无分别。

由关于 μ 的对称性，可得 $\mathrm{P}(X \leq \mu) = \frac{1}{2}$，并且对任意 k 都有 $\mathrm{P}(X \leq \mu - k\sigma) = \mathrm{P}(X \geq \mu + k\sigma)$。特别地，

$$\mathrm{P}\left(X \leq \mu - \tfrac{1}{2}\sigma\right) = 1 - b, \qquad \mathrm{P}\left(X \leq \mu - \sigma\right) = 1 - a.$$

不妨以 σ 为单位，从 μ 起量度，把每一段区间的概率记下来：



例如 $\mathrm{P}(\mu < X \leq \mu + \frac{1}{2}\sigma) = b - \frac{1}{2}$，而 $\mathrm{P}(\mu + \frac{1}{2}\sigma < X \leq \mu + \sigma) = a - b$；左边各段的概率则由对称性得出。

第 (i) 部分

把累积概率相减，

$$\mathrm{P}\big(\mu-\tfrac{1}{2}\sigma\le X\le\mu+\sigma\big)=\mathrm{P}(X\le\mu+\sigma)-\mathrm{P}\big(X<\mu-\tfrac{1}{2}\sigma\big)=a-(1-b)=a+b-1.$$

至于条件概率（conditional probability），分子所对应的事件是 $\{\mu-\frac{1}{2}\sigma\le X\le\mu+\frac{1}{2}\sigma\}$，其概率为 $b-(1-b)=2b-1$，而 $\mathrm{P}(X\ge\mu-\frac{1}{2}\sigma)=1-(1-b)=b$。因此

$$\mathrm{P}\big(X\le\mu+\tfrac{1}{2}\sigma\mid X\ge\mu-\tfrac{1}{2}\sigma\big)=\frac{2b-1}{b}.$$

第 (ii)(a) 部分

设 S 为该盒牛奶是脱脂奶的事件，F 为该盒牛奶是全脂奶的事件，则 $\mathrm{P}(S)=0.6$ 及 $\mathrm{P}(F)=0.4$。设 V 为以 ml 计的容量。

想一想 以 σ 为单位，500 ml 和 505 ml 在每种牛奶的分布中各落在甚么位置？

对脱脂奶（$\mu=500$, $\sigma=10$）而言，500 是 μ，505 是 $\mu+\frac{1}{2}\sigma$。对全脂奶（$\mu=495$, $\sigma=10$）而言，500 是 $\mu+\frac{1}{2}\sigma$，505 是 $\mu+\sigma$。从上图各段读出：

$$\mathrm{P}(V<505\mid S)=b,\qquad \mathrm{P}(500<V<505\mid S)=b-\tfrac{1}{2},$$
$$\mathrm{P}(V<505\mid F)=a,\qquad \mathrm{P}(500<V<505\mid F)=a-b.$$

根据全概率公式（law of total probability），

$$\mathrm{P}(V<505)=0.6b+0.4a,$$
$$\mathrm{P}(500<V<505)=0.6\big(b-\tfrac{1}{2}\big)+0.4(a-b)=0.4a+0.2b-0.3.$$

“多于 500 ml 而少于 505 ml”这个事件正是我们所需的交集，所以

$$\mathrm{P}(V>500\mid V<505)=\frac{0.4a+0.2b-0.3}{0.4a+0.6b}=\frac{4a+2b-3}{4a+6b}.$$

第 (ii)(b) 部分

随机抽取的牛奶盒数量非常大时，观察所得的比例可用来估计相应的概率。第一项资料表示 $\mathrm{P}(V\le505)=0.7$，按 (a) 部分的计算，即

$$0.6b+0.4a=0.7,\quad\text{即}\quad 4a+6b=7. \tag{1}$$

想一想 对每种牛奶而言，$\mathrm{P}(V\ge495)$ 是多少？

对脱脂奶而言，495 是 $\mu - \frac{1}{2}\sigma$，所以 $\mathrm{P}(V \geq 495 \mid S) = b$（这正是第 (i) 部分的分母）。对全脂奶而言，495 是 μ，所以由对称性得 $\mathrm{P}(V \geq 495 \mid F) = \frac{1}{2}$。根据贝叶斯定理（Bayes' theorem），

$$\mathrm{P}(F \mid V \geq 495) = \frac{0.4 \times \frac{1}{2}}{0.6b + 0.4 \times \frac{1}{2}} = \frac{0.2}{0.6b + 0.2}.$$

令此式等于 $\frac{1}{3}$，得 $0.6b + 0.2 = 0.6$，所以 $b = \frac{2}{3}$。于是由 (1) 得 $4a = 7 - 4 = 3$，所以

$$a \approx \tfrac{3}{4}, \qquad b \approx \tfrac{2}{3}.$$

这些只是估计值，因为观察所得的比例只是真实概率的近似。

验算

- 估计值合情合理：$\frac{1}{2} < b < a < 1$，这是必然的，因为 $\mu < \mu + \frac{1}{2}\sigma < \mu + \sigma$。
- 代回检验：$0.6 \cdot \frac{2}{3} + 0.4 \cdot \frac{3}{4} = 0.4 + 0.3 = 0.7$，以及 $0.2/(0.4 + 0.2) = \frac{1}{3}$。
- 若分布为正态分布（Normal distribution），则 $a \approx 0.841$ 及 $b \approx 0.691$；以这些数值计算，第 (i) 部分得出 0.533 和 0.553。以（非正态的）拉普拉斯分布（Laplace distribution）作模拟，样本数目超过 10^5，结果与第 (i) 及 (ii)(a) 部分的所有公式吻合；例如 (ii)(a) 的模拟结果为 0.2518，而公式给出 0.2515。

常见错误

- 把 $\mathrm{P}(X \leq \mu - \frac{1}{2}\sigma) = b$ 当作正确，而其实应是 $1 - b$：对称性反射的是*尾部*，而不是累积概率。
- 忘记两种牛奶的平均值不同，所以同一个容量（例如 500 ml）对两种牛奶而言，与平均值相距的标准差（standard deviation）数目并不相同。
- 对两种牛奶都用 $\mathrm{P}(V \geq 495) = b$；对全脂奶而言，495 ml 正是平均值。

STEP 之外

这道题的要点在于：即使对分布一无所知，单凭对称性和全概率公式也可以走得很远；这里由两个总体混合而成的情况，是一个简单的*混合模型*（mixture model），而第 (ii)(b) 部分则是根据数据估计其参数的入门版本。另一个同样不依赖分布形状的相关结果是切比雪夫（Chebyshev）不等式，它只用方差（variance）便能给出 $\mathrm{P}(|X - \mu| \geq k\sigma)$ 的上界。可参阅 Blitzstein and Hwang, *Introduction to Probability*（概率论入门），或 Grimmett and Welsh, *Probability: An Introduction*（概率论导引）。

详细解答

STEP I 2005，第 14 题

概率与统计 · 混合型随机变量

STEP I 2005，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/05-s1-q14

关键思路

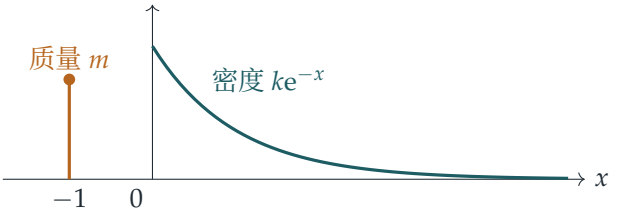
X 是一个混合型随机变量（mixed random variable）：有一团概率集中在 -1 这一点上，其余则连续地分布在 $[0,\infty)$ 上。处理的方法是把每一项计算都拆成两部分：来自 $X=-1$ 的离散项，加上在 $[0,\infty)$ 上对密度所作的积分。只要留意到连续部分不过是一个乘了常数的指数分布（exponential distribution），大部分积分便都是我们早已认识的了。

解答

把题目给出的累积概率(cumulative probability)求导,可知 X 的连续部分的密度(density)为

$$f(x)=\frac{\mathrm{d}}{\mathrm{d}x}k(1-\mathrm{e}^{-x})=k\mathrm{e}^{-x},\qquad x\geq 0.$$

所以 X 的样子是这样的：在 -1 处有一个大小为 m 的点质量（point mass），在 $[0,\infty)$ 上则是曲线 $k\mathrm{e}^{-x}$。



第 (i) 部分

总概率为 1。令 $x \to \infty$，得 $\mathrm{P}(X \geq 0) = k(1-0) = k$，所以

$$m + k = 1, \qquad \text{即} \qquad k = 1 - m.$$

（由于 $m \geq 0$ 是一个概率，而且 $m < \frac{1}{2}$，我们有 $k > \frac{1}{2} > 0$，所以密度确实是正的。）

第 (ii) 部分

拆成两部分，

$$\begin{aligned}\mathrm{E}(X) &= (-1)\cdot m + \int_0^\infty x\,(1-m)\mathrm{e}^{-x}\,\mathrm{d}x \\ &= -m + (1-m)\cdot 1 = 1 - 2m,\end{aligned}$$

其中用到 $\int_0^\infty x\mathrm{e}^{-x}\,\mathrm{d}x = 1$（可用分部积分法（integration by parts），或认出这是参数为 1 的指数分布的平均值）。这正是所要求的结果。

第 (iii) 部分

方差（*variance*）。同样地，利用 $\int_0^\infty x^2\mathrm{e}^{-x}\,\mathrm{d}x = 2$（分部积分两次），

$$\mathrm{E}(X^2) = (-1)^2\cdot m + (1-m)\cdot 2 = 2 - m.$$

因此

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2 = 2 - m - (1-2m)^2 \\ &= 1 + 3m - 4m^2 = (1-m)(1+4m).\end{aligned}$$

想一想 中位数（median）可以是 -1，或者落在 -1 与 0 之间吗?

中位数。不可以：X 从不在 $(-1,0)$ 内取值，而且 $\mathrm{P}(X \leq -1) = m < \frac{1}{2}$，所以单靠点质量并不足以累积到一半的概率。条件 $m < \frac{1}{2}$ 正正是用在这里。因此中位数 M 满足 $M \geq 0$ 及 $\mathrm{P}(X \leq M) = \frac{1}{2}$：

$$\begin{aligned} m + (1-m)(1-\mathrm{e}^{-M}) &= \tfrac{1}{2} \\ 1 - (1-m)\mathrm{e}^{-M} &= \tfrac{1}{2} \\ \mathrm{e}^{-M} &= \frac{1}{2(1-m)}.\end{aligned}$$

于是

$$M = \ln\big(2(1-m)\big) = \ln(2-2m).$$

作为一致性检查，由 $0 \le m < \frac{1}{2}$ 得 $1 < 2 - 2m \le 2$，所以 $M > 0$，与应有的情况相符。由于分布在 $[0, \infty)$ 上是连续的，这个 M 同时满足 $\mathrm{P}(X \ge M) = \frac{1}{2}$，所以中位数是毫不含糊的。

第 (iv) 部分

当 $X = -1$ 时，$|X|^{1/2} = 1$；当 $X \ge 0$ 时，$|X|^{1/2} = \sqrt{X}$。所以

$$\mathrm{E}\big(|X|^{1/2}\big) = 1 \cdot m + (1 - m) \int_0^\infty \sqrt{x}\, \mathrm{e}^{-x}\, \mathrm{d}x.$$

想一想 怎样把 $\int_0^\infty \sqrt{x}\, \mathrm{e}^{-x}\, \mathrm{d}x$ 化成题目给出的那个积分？

题目给出的积分含有 e^{-y^2}，这提示我们代入 $x = y^2$（其中 $y \ge 0$），于是 $\sqrt{x} = y$ 及 $\mathrm{d}x = 2y\, \mathrm{d}y$；积分上下限 0 与 ∞ 保持不变：

$$\int_0^\infty \sqrt{x}\, \mathrm{e}^{-x}\, \mathrm{d}x = \int_0^\infty y\, \mathrm{e}^{-y^2}\, 2y\, \mathrm{d}y = 2 \int_0^\infty y^2 \mathrm{e}^{-y^2}\, \mathrm{d}y = 2 \cdot \tfrac{1}{4}\sqrt{\pi} = \tfrac{1}{2}\sqrt{\pi}.$$

因此

$$\mathrm{E}\big(|X|^{1/2}\big) = m + \tfrac{1}{2}\sqrt{\pi}\,(1 - m).$$

验算

- 当 $m = 0$ 时，X 恰好是参数为 1 的指数分布：我们的公式给出平均值 1、方差 1、中位数 $\ln 2$ 及 $\mathrm{E}(\sqrt{X}) = \frac{1}{2}\sqrt{\pi} = \Gamma(\frac{3}{2})$，全部都是标准数值。
- 当 $m \to \frac{1}{2}$ 时，中位数趋向 $\ln 1 = 0$，这是合理的：一半的概率位于 -1，所以中位数会滑落到连续部分的最底端。
- 取 $m = 0.3$ 并抽取 4×10^5 个样本作模拟，得出平均值 0.402（公式为 0.4）、方差 1.551（公式为 1.54）、中位数 0.336（公式为 $\ln 1.4 \approx 0.336$）及 $\mathrm{E}(|X|^{1/2}) \approx 0.922$（公式为 0.920）。

常见错误

- 在 $\mathrm{E}(X^2)$ 或 $\mathrm{E}(|X|^{1/2})$ 中处理点质量时出错：它的贡献是 $+m$ 而不是 $-m$，因为 $(-1)^2 = |-1|^{1/2} = 1$；漏掉这一项同样是错的。
- 解 $(1 - m)(1 - \mathrm{e}^{-M}) = \frac{1}{2}$ 来求中位数，这样便忽略了在 -1 处已经累积的概率 m。
- 换元时没有把 $\mathrm{d}x$ 换成 $2y\, \mathrm{d}y$，结果得出 $\frac{1}{4}\sqrt{\pi}$ 而不是 $\frac{1}{2}\sqrt{\pi}$。

STEP 之外

同时包含点质量和密度的分布在实际应用中十分常见（一宗保险索偿有某个概率为零，否则便是连续的）。在大学里，这类分布会用统一的方法处理：把期望值（expectation）写成对分布函数（distribution function）的积分，$\mathrm{E}(g(X)) = \int g(x)\,\mathrm{d}F(x)$，一次过涵盖离散、连续和混合三种情况。第 (iv) 部分的积分就是伽玛函数（Gamma function）的值 $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）是很好的下一步读物。

详细解答

STEP II 2005，第 1 题

纯数学 · 多项式与驻点

STEP II 2005，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/05-s2-q1

关键思路

用积法则（product rule）对任何形如 $P(x)\mathrm{e}^{-x^2}$ 的式子求导，会得到 $\big(P'(x)-2xP(x)\big)\mathrm{e}^{-x^2}$；由于 e^{-x^2} 永不为零，驻点（stationary point）正好就是多项式（polynomial）$P'(x)-2xP(x)$ 的实根。于是第二部分变成一个逆向问题：先决定我们希望 $P'-2xP$ 是哪一个多项式，然后通过比较系数（comparing coefficients）倒推出 P。

解答

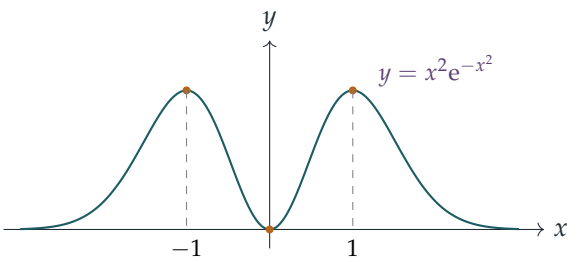
第一部分

由积法则，

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(x^2\mathrm{e}^{-x^2}\right)=2x\,\mathrm{e}^{-x^2}+x^2\cdot(-2x)\,\mathrm{e}^{-x^2}=2x\big(1-x^2\big)\mathrm{e}^{-x^2}.$$

由于对所有 x 都有 $\mathrm{e}^{-x^2}>0$，上式为零当且仅当 $x(1-x^2)=0$，即

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



图像在原点有一个极小点，并在 $x=\pm1$ 处有高度为 e^{-1} 的极大点。

第二部分

对任何多项式 P,

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

所以，同样因为 $\mathrm{e}^{-x^2} > 0$，导数（derivative）为零之处正好就是 $P'(x) - 2xP(x) = 0$ 之处。第一部分正符合这个模式：$P(x) = x^2$ 给出 $P' - 2xP = -2x(x^2 - 1)$。因此我们的目标是

$$P'(x) - 2xP(x) = -2x\left(x^2 - a^2\right)\left(x^2 - b^2\right), \qquad (*)$$

它的实根只有 0、$\pm a$ 和 $\pm b$。这是五个相异的值，因为 a 和 b 是相异的正数。

想一想 P 的次数应该是多少？为甚么它应该只含偶数次幂？

若 P 的次数为 n, 首项系数 (leading coefficient) 为 c, 则 P' 的次数为 $n-1$, 所以 $P' - 2xP$ 的次数为 $n+1$，首项系数为 $-2c$。$(*)$ 的右方次数为 5，首项系数为 -2，所以我们需要 $n = 4$ 和 $c = 1$。此外，右方是一个奇函数（odd function）；若 P 是偶函数（even function），则 P' 和 $2xP$ 都是奇函数，所以很自然会尝试

$$P(x) = x^4 + \beta x^2 + \gamma.$$

于是 $P'(x) = 4x^3 + 2\beta x$，而

$$\begin{aligned} P'(x) - 2xP(x) &= 4x^3 + 2\beta x - 2x^5 - 2\beta x^3 - 2\gamma x \\ &= -2x^5 + (4 - 2\beta)x^3 + (2\beta - 2\gamma)x, \end{aligned}$$

另一方面，$(*)$ 的右方展开后是

$$-2x^5 + 2\left(a^2 + b^2\right)x^3 - 2a^2b^2x.$$

比较 x^3 和 x 的系数:

$$\begin{aligned} 4 - 2\beta = 2\left(a^2 + b^2\right) \quad &\Longrightarrow \quad \beta = 2 - a^2 - b^2, \\ 2\beta - 2\gamma = -2a^2b^2 \quad &\Longrightarrow \quad \gamma = \beta + a^2b^2 = a^2b^2 - a^2 - b^2 + 2. \end{aligned}$$

所以其中一个合适的多项式是

$$\boxed{P(x) = x^4 + \left(2 - a^2 - b^2\right)x^2 + a^2b^2 - a^2 - b^2 + 2.}$$

这样选取后，$(*)$ 恒成立，所以

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(P(x)\mathrm{e}^{-x^2}\right) = -2x\left(x^2 - a^2\right)\left(x^2 - b^2\right)\mathrm{e}^{-x^2},$$

它在 $x = 0$、$x = \pm a$ 和 $x = \pm b$ 时为零，而且由于 $\mathrm{e}^{-x^2} \neq 0$，在其他 x 值都不为零。P 的任何非零常数倍同样可行。

验算

- 取 $a=1$、$b=2$：这时 $P(x)=x^4-3x^2+1$，而 $P'-2xP=4x^3-6x-2x^5+6x^3-2x=-2x^5+10x^3-8x=-2x(x^2-1)(x^2-4)$，符合要求。
- 数值上，当 $a=0.7$、$b=1.9$ 时，$P(x)\mathrm{e}^{-x^2}$ 的导数在 $[-5,5]$ 上只在 $x=0,\pm0.7,\pm1.9$ 附近变号；这条公式亦已用数千组随机的 (a,b,x) 验证过。
- 粗略地令 $b\to 0$：$P\to x^4+(2-a^2)x^2+2-a^2$，导数变成 $-2x^3(x^2-a^2)\mathrm{e}^{-x^2}$，位于 0 和 $\pm b$ 的根合而为一。

常见错误

- 漏掉链式法则（chain rule）带来的因子：e^{-x^2} 的导数是 $-2x\,\mathrm{e}^{-x^2}$，而不是 e^{-x^2}。
- 为了配合五个根而猜 P 的次数是 5：对 $P\mathrm{e}^{-x^2}$ 求导会令次数增加一，所以 P 的次数是 4。
- 没有说明为何没有其他驻点：这需要用到 $\mathrm{e}^{-x^2}\neq 0$，以及该五次多项式只有这些根这个事实。

STEP 之外

映射 $P\mapsto P'-2xP$ 是多项式上的一个线性算子（linear operator），而解 $(*)$ 就是解一条一阶线性微分方程（first-order linear differential equation），其积分因子（integrating factor）为 e^{-x^2}。形如 $H(x)\mathrm{e}^{-x^2}$（其中 H 为多项式）的函数引出埃尔米特多项式（Hermite polynomials），它们出现于概率（正态分布，normal distribution）和量子谐振子（quantum harmonic oscillator）之中。Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科书）涵盖积分因子和埃尔米特方程（Hermite's equation）。

详细解答

STEP II 2005，第 2 题

纯数学 · 欧拉 φ 函数

STEP II 2005，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/05-s2-q2

关键思路

一切都源于把 N 写成质因数分解（prime factorisation）的形式。每个因子 $\left(1-\frac{1}{p}\right)=\frac{p-1}{p}$ 的分母都有一个 p，而 N 必定含有至少一个 p 可以把它约去。一旦把 $\mathrm{f}(N)$ 写成各质数幂（prime power）的乘积，便能清楚看出 $\mathrm{f}(m)\mathrm{f}(n)$ 与 $\mathrm{f}(mn)$ 何时相等：关键在于 m 与 n 有没有公共的质因数。

解答

第 (i) 部分

(a) 整除 $12=2^2\cdot 3$ 的质数是 2 和 3，所以

$$\mathrm{f}(12)=12\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)=12\cdot\frac{1}{2}\cdot\frac{2}{3}=4.$$

整除 $180=2^2\cdot 3^2\cdot 5$ 的质数是 2、3 和 5，所以

$$\mathrm{f}(180)=180\cdot\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{4}{5}=48.$$

(b) 当 $N=1$ 时没有质因数，乘积是空积（empty product），故 $\mathrm{f}(1)=1$。当 $N\geq 2$ 时，写

$$N=p_1^{a_1}p_2^{a_2}\cdots p_k^{a_k},$$

其中 p_i 是整除 N 的相异质数，而每个 $a_i\geq 1$。于是

$$\begin{aligned}\mathrm{f}(N)&=p_1^{a_1}\cdots p_k^{a_k}\cdot\frac{p_1-1}{p_1}\cdots\frac{p_k-1}{p_k}\\&=p_1^{a_1-1}(p_1-1)\,p_2^{a_2-1}(p_2-1)\cdots p_k^{a_k-1}(p_k-1).\end{aligned}$$

由于每个 $a_i \geq 1$，每个指数 $a_i - 1$ 都是非负整数，所以每个因子 $p_i^{a_i-1}(p_i - 1)$ 都是正整数。因此对所有 N，$\mathrm{f}(N)$ 都是整数。本题余下部分都会用到这条公式：

$$\mathrm{f}(N) = \prod_{i=1}^{k} p_i^{a_i-1}(p_i - 1), \qquad (*)$$

第 (ii) 部分

(a) 这是错的。取 $m = n = 2$，则 $\mathrm{f}(2) = 2 \cdot \frac{1}{2} = 1$，所以 $\mathrm{f}(2)\mathrm{f}(2) = 1$，但 $\mathrm{f}(4) = 4 \cdot \frac{1}{2} = 2$。问题在于因子 $\left(1 - \frac{1}{2}\right)$ 在左边出现了两次，在右边却只出现一次，因为 2 作为 4 的质因数只计算一次。

(b) 这是对的。若 p 和 q 是相异质数，则 $\mathrm{f}(p) = p\left(1 - \frac{1}{p}\right) = p - 1$，同理 $\mathrm{f}(q) = q - 1$。整除 pq 的质数恰好是 p 和 q，而且两者不同，所以

$$\mathrm{f}(pq) = pq\left(1 - \frac{1}{p}\right)\left(1 - \frac{1}{q}\right) = (p-1)(q-1) = \mathrm{f}(p)\mathrm{f}(q).$$

(c) 这个命题是说：$\mathrm{f}(p)\mathrm{f}(q) = \mathrm{f}(pq)$ *只有在* p 和 q 是相异质数时才可能成立；换言之，这条等式迫使 p 和 q 必须是相异质数。这是错的。取 $p = 3$ 和 $q = 4$；等式要有意义，两者都不必是质数。于是

$$\mathrm{f}(3)\mathrm{f}(4) = 2 \cdot 2 = 4 = \mathrm{f}(12),$$

这里用了第 (i)(a) 部分的结果，但 4 不是质数。（更简单的例子是 $p = 1$，因为对任何 q 都有 $\mathrm{f}(1)\mathrm{f}(q) = \mathrm{f}(q)$。）

> ? **想一想** 如果不是质数这个性质，那么真正令等式成立的是甚么？

真正重要的是 p 和 q 没有公共的质因数。若 $\gcd(m, n) = 1$，则 m 与 n 各自所含的质数幂互不相同，所以 mn 的分解式不过是把两个分解式并排写在一起，而由 $(*)$ 可得 $\mathrm{f}(mn) = \mathrm{f}(m)\mathrm{f}(n)$。若有某个质数 r 同时整除两者，则因子 $\left(1 - \frac{1}{r}\right)$ 在 $\mathrm{f}(m)\mathrm{f}(n)$ 中出现两次，在 $\mathrm{f}(mn)$ 中却只出现一次，所以 $\mathrm{f}(m)\mathrm{f}(n) < \mathrm{f}(mn)$。因此等式恰好在 $\gcd(m, n) = 1$ 时成立，而 3、4 只是其中一对互质（coprime）但并非两个都是质数的数。

假如你把第 (ii) 部分中的 p 和 q 一律理解为质数，那么 (c) 问的只是 $p = q$ 的情况会否出现，而答案是不会：$\mathrm{f}(p)^2 = (p-1)^2$，而 $\mathrm{f}(p^2) = p(p-1)$，由于 $p - 1 \neq p$，两者永不相等。在这种理解下，(c) 是对的。不过，配合上面反例（counterexample）的一般理解，才是较自然的读法。

第 (iii) 部分

由 $(*)$，质数幂满足

$$\mathrm{f}(p^m) = p^{m-1}(p - 1).$$

把目标数值分解，得 $146410 = 10 \times 14641 = 2 \cdot 5 \cdot 11^4$，因为 $14641 = 121^2 = 11^4$。这已经是 $p^{m-1}(p-1)$ 的形式，其中 $p = 11$：$11^4 \cdot 10 = 11^4(11-1)$，所以 $m - 1 = 4$。因此

$$p = 11, \quad m = 5: \qquad \mathrm{f}(11^5) = 11^4 \cdot 10 = 146410.$$

事实上这是唯一的答案。若 $m \geq 2$，则 p 整除 146410，所以 $p \in \{2, 5, 11\}$。取 $p = 2$ 会得到 $2^{m-1} = 146410$，这不可能，因为 146410 有奇质因数。取 $p = 5$ 会得到 $4 \cdot 5^{m-1} = 146410$，这也不可能，因为 146410 不能被 4 整除。余下只有 $p = 11$，由此得 $m = 5$。若 $m = 1$，则需要 $p = 146411$ 是质数，但 $146411 = 41 \times 3571$。

验算

- 这个函数就是欧拉函数（Euler's totient）：$\mathrm{f}(N)$ 计算 $\{1, \ldots, N\}$ 中与 N 互质的整数的个数。例如当 $N = 12$ 时，这样的数有四个，就是 $1, 5, 7, 11$。经电脑验证，对每个 $N < 2000$，乘积公式都与这个个数吻合，而且结果是整数。
- 题目中的示例给出 $\mathrm{f}(80) = 80 \cdot \frac{1}{2} \cdot \frac{4}{5} = 32 = 2^3(2-1) \cdot 5^0(5-1)$，与 $(*)$ 吻合。
- 电脑测试了 20 000 对随机选取的 $m, n \leq 400$，发现 $\mathrm{f}(m)\mathrm{f}(n) = \mathrm{f}(mn)$ 恰好在 $\gcd(m, n) = 1$ 时成立。另外，搜索所有质数 $p \leq 146411$ 后，发现 $(p, m) = (11, 5)$ 是 (iii) 的唯一解。

常见错误

- 在乘积中把同一个质数计算多于一次：对 $180 = 2^2 \cdot 3^2 \cdot 5$ 而言，因子有三个，不是五个。
- 把“仅当”（only if）当作“若”（if）。(b) 部分是“若”的方向，而 (c) 部分是它的逆命题（converse），两者是不同的命题。
- 在 (iii) 中看不出 $14641 = 11^4$。留意到 $11^2 = 121$ 和 $121^2 = 14641$，或者认出帕斯卡三角形（Pascal's triangle）中 $1, 4, 6, 4, 1$ 这一行，便能很快找到。

STEP 之外

这个函数就是欧拉函数 φ，而第 (ii) 部分的结果说明它是积性（multiplicative）的：只要 $\gcd(m, n) = 1$，便有 $\varphi(mn) = \varphi(m)\varphi(n)$，但一般情况下则不然。它出现在欧拉定理（Euler's theorem）之中：若 a 与 n 互质，则 $a^{\varphi(n)} \equiv 1 \pmod{n}$；这条定理推广了费马小定理（Fermat's little theorem），也是 RSA 加密法（RSA encryption）的基础。在群论

（group theory）中，$\varphi(n)$ 是模 n 单位群（group of units）的阶（order）。Davenport 的 *The Higher Arithmetic*（一本经典的数论入门书）把这一切阐述得十分精彩。

详细解答

STEP II 2005，第 3 题

纯数学 · 分部积分法

STEP II 2005，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/05-s2-q3

关键思路

先求导：$\sin x - x\cos x$ 的导数（derivative）会化简成一个正负号一目了然的式子，草图和 y 的上下界便一并得出。两个积分都是常规的分部积分法（integration by parts），平方项则借助二倍角公式（double-angle formulae）处理。最后的不等式来自两者的比较：介乎 0 与 1 之间的数，不会小于它自己的平方。

解答

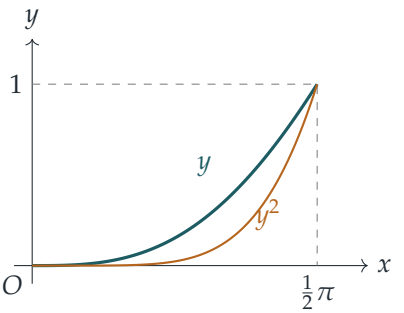
草图与上下界

用积法则（product rule）求导，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \cos x - (\cos x - x\sin x) = x\sin x.$$

当 $0 < x \leq \frac{1}{2}\pi$ 时，x 和 $\sin x$ 均为正数，所以 y 在 $[0, \frac{1}{2}\pi]$ 上严格递增（strictly increasing）。因此它的最小值是 $y(0) = 0 - 0 = 0$，最大值是 $y(\frac{1}{2}\pi) = 1 - \frac{1}{2}\pi \cdot 0 = 1$。故此 $0 \leq y \leq 1$，而且事实上当 $0 < x < \frac{1}{2}\pi$ 时有 $0 < y < 1$。

至于曲线的形状，在 $(0, \frac{1}{2}\pi]$ 上 $\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2} = \sin x + x\cos x > 0$，所以整条曲线都向上弯。在原点处斜率（gradient）为 0。利用马克劳林级数（Maclaurin series），$y = \left(x - \frac{1}{6}x^3\right) - x\left(1 - \frac{1}{2}x^2\right) + \cdots = \frac{1}{3}x^3 + \cdots$，所以曲线离开原点时非常平坦，形状像三次曲线。在 $x = \frac{1}{2}\pi$ 处斜率为 $\frac{1}{2}\pi$。



图中也画出了琥珀色的曲线 y^2，因为它是最后一部分的关键。

第 (i) 部分

我们需要求 $\int_0^{\pi/2} x\cos x\,\mathrm{d}x$。取 $u=x$、$\mathrm{d}v=\cos x\,\mathrm{d}x$ 作分部积分，

$$\int_0^{\pi/2} x\cos x\,\mathrm{d}x=\Big[x\sin x\Big]_0^{\pi/2}-\int_0^{\pi/2}\sin x\,\mathrm{d}x=\tfrac{1}{2}\pi-1.$$

由于 $\int_0^{\pi/2}\sin x\,\mathrm{d}x=1$，

$$\int_0^{\pi/2} y\,\mathrm{d}x=1-\left(\tfrac{1}{2}\pi-1\right)=2-\frac{\pi}{2},$$

正是所求。

第 (ii) 部分

展开，

$$y^2=\sin^2 x-2x\sin x\cos x+x^2\cos^2 x=\sin^2 x-x\sin 2x+\tfrac{1}{2}x^2+\tfrac{1}{2}x^2\cos 2x,$$

其中用到 $2\sin x\cos x=\sin 2x$ 和 $\cos^2 x=\frac{1}{2}(1+\cos 2x)$。我们逐一处理这四项。

第一，$\sin^2 x=\frac{1}{2}(1-\cos 2x)$，而 $\sin 2x$ 在上下限处均为零，所以

$$\int_0^{\pi/2}\sin^2 x\,\mathrm{d}x=\Big[\tfrac{1}{2}x-\tfrac{1}{4}\sin 2x\Big]_0^{\pi/2}=\frac{\pi}{4}.$$

其次，取 $u=x$、$\mathrm{d}v=\sin 2x\,\mathrm{d}x$ 作分部积分，并利用 $\cos\pi=-1$，

$$\int_0^{\pi/2} x\sin 2x\,\mathrm{d}x=\Big[-\tfrac{1}{2}x\cos 2x\Big]_0^{\pi/2}+\tfrac{1}{2}\int_0^{\pi/2}\cos 2x\,\mathrm{d}x=\frac{\pi}{4}+0=\frac{\pi}{4}.$$

第三，$\int_0^{\pi/2}\frac{1}{2}x^2\,\mathrm{d}x=\frac{1}{6}\left(\frac{1}{2}\pi\right)^3=\dfrac{\pi^3}{48}$。

最后，取 $u = x^2$、$\mathrm{d}v = \cos 2x\,\mathrm{d}x$ 作分部积分；边界项 $\left[\frac{1}{2}x^2 \sin 2x\right]_0^{\pi/2}$ 为零，而余下的积分正是刚才求得的那一个：

$$\int_0^{\pi/2} x^2 \cos 2x\,\mathrm{d}x = 0 - \int_0^{\pi/2} x \sin 2x\,\mathrm{d}x = -\frac{\pi}{4},$$

所以最后一项的贡献是 $\frac{1}{2}\cdot\left(-\frac{1}{4}\pi\right) = -\frac{1}{8}\pi$。

把四项相加，

$$\int_0^{\pi/2} y^2\,\mathrm{d}x = \frac{\pi}{4} - \frac{\pi}{4} + \frac{\pi^3}{48} - \frac{\pi}{8} = \frac{\pi^3}{48} - \frac{\pi}{8},$$

正是所求。

推论

想一想 两个积分怎样能够得出一个关于 π 的不等式?

比较两个被积函数(integrand)。由于 $0 \le y \le 1$,在整个区间上都有 $y - y^2 = y(1-y) \ge 0$。更进一步，我们已证明当 $0 < x < \frac{1}{2}\pi$ 时 $0 < y < 1$，所以在该处 $y(1-y) > 0$。一个非负的连续函数（continuous function），若在某个开区间上严格为正，它的积分必定严格为正。因此

$$\int_0^{\pi/2} y^2\,\mathrm{d}x < \int_0^{\pi/2} y\,\mathrm{d}x, \qquad \text{即} \qquad \frac{\pi^3}{48} - \frac{\pi}{8} < 2 - \frac{\pi}{2}.$$

两边乘以 48（它是正数，所以不等号方向不变），

$$\pi^3 - 6\pi < 96 - 24\pi, \qquad \text{所以} \qquad \pi^3 + 18\pi < 96.$$

这就是所要求的结果。

验算

- 以数值计算，$2 - \frac{1}{2}\pi \approx 0.4292$，而 $\frac{1}{48}\pi^3 - \frac{1}{8}\pi \approx 0.2533$；对原来的被积函数使用辛普森法则（Simpson's rule），所得数值准确至小数点后十二位都相同。两者都介乎 0 与 $\frac{1}{2}\pi \approx 1.571$ 之间，这是必然的，因为 $0 \le y \le 1$。
- 第二个积分比第一个小，正如图中 y^2 位于 y 之下所显示的那样。
- 取 $\pi \approx 3.1416$，$\pi^3 + 18\pi \approx 31.006 + 56.549 = 87.55$，远低于 96。

常见错误

- 在 $\int x \sin 2x\,\mathrm{d}x$ 中弄错边界项的正负号：在 $x = \frac{1}{2}\pi$ 处，$\cos 2x = \cos \pi = -1$，所以 $-\frac{1}{2}x\cos 2x$ 等于 $+\frac{1}{4}\pi$。

- 在最后一步只得出 $\leq$。要得到严格不等式，需要 $y(1-y) > 0$ 在某处成立，而不仅是 ≥ 0。
- 忘记说明为何 $y \leq 1$：只知道 $y(\frac{1}{2}\pi) = 1$ 并不足够，还须用到 y 递增这一事实。

STEP 之外

最后一步是一个重要想法的小小例子：函数之间的不等式，经积分后会变成数与数之间的不等式，而只要巧妙地选取函数，便可以把 π 这类常数困在某个范围之内。同样的策略可以由 $\int_0^1 \frac{x^4(1-x)^4}{1+x^2}\,\mathrm{d}x > 0$ 得出 $\frac{22}{7} > \pi$ 这个著名的证明，而更有分量的例子，是尼云（Niven）对 π 是无理数（irrational）的证明。在大学第一门分析课程中，你会严格地证明：一个非负的连续函数，若在某处为正，它的积分必为正。Abbott 的 *Understanding Analysis*（《理解分析》）是一本平易近人的书，可以看到这个证明怎样严谨地完成。

详细解答

STEP II 2005，第 4 题

纯数学 · 反正切恒等式

STEP II 2005，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/05-s2-q4

关键思路

处理反正切（inverse tangent）之和，方法是对等式两边取正切，再运用 $\tan(A+B)=\dfrac{\tan A+\tan B}{1-\tan A\tan B}$。条件 $bc=a^2+1$ 恰好令分母可以因式分解（factorise），并与分子约去。其后的问题就在于结构：第一个结果是一条把一个反正切拆开成两个的法则，而四项恒等式（identity）就是把这条法则用两次，在树的每一层各用一次。至于数值部分，则要倒推回去，找出是哪一种拆法产生了那四个数。

解答

第一个恒等式

设 $A=\arctan\frac{1}{a+b}$ 及 $B=\arctan\frac{1}{a+c}$。由于 $a,b,c>0$，$\frac{1}{a+b}$ 与 $\frac{1}{a+c}$ 均为正数，所以 $A,B\in(0,\frac{\pi}{2})$，而 $A+B\in(0,\pi)$。利用 $bc=a^2+1$，

$$\begin{aligned}(a+b)(a+c)-1&=a^2+ab+ac+bc-1\\&=2a^2+ab+ac=a(2a+b+c).\end{aligned}$$

因此，根据复角公式（addition formula），

$$\begin{aligned}\tan(A+B)&=\frac{\dfrac{1}{a+b}+\dfrac{1}{a+c}}{1-\dfrac{1}{(a+b)(a+c)}}=\frac{(a+c)+(a+b)}{(a+b)(a+c)-1}\\&=\frac{2a+b+c}{a(2a+b+c)}=\frac{1}{a},\end{aligned}$$

其中因为 $2a+b+c>0$，所以可以约去。

想一想 由 $\tan(A+B)=\frac{1}{a}$ 真的可以得出 $A+B=\arctan\frac{1}{a}$ 吗?

不是必然的: tan 的周期(period)为 π, 所以我们需要 $A+B$ 位于 $(-\frac{\pi}{2},\frac{\pi}{2})$ 之内, 即 arctan 的值域(range)。我们已知 $A+B\in(0,\pi)$。此外 $(a+b)(a+c)-1=a(2a+b+c)>0$, 所以 $\tan A\tan B<1$, 复角公式中的分母为正数; 加上分子为正数, 便得 $\tan(A+B)>0$, 这就排除了 $A+B\in[\frac{\pi}{2},\pi)$ 的可能。(事实上, $\tan(A+B)$ 在 $\frac{\pi}{2}$ 处无定义, 而在 $(\frac{\pi}{2},\pi)$ 上为负数。) 所以 $A+B\in(0,\frac{\pi}{2})$, 而 $\arctan\frac{1}{a}$ 是该区间内唯一一个正切值为 $\frac{1}{a}$ 的角, 故

$$\arctan\frac{1}{a+b}+\arctan\frac{1}{a+c}=\arctan\frac{1}{a}. \qquad \blacksquare$$

四项恒等式

把第一个恒等式应用三次; 每次所涉及的三个数均为正数。

- 取 $a=p+q$、$b=s$、$c=t$: 条件为 $st=(p+q)^2+1$, 这是已知的, 所以

$$\arctan\frac{1}{p+q+s}+\arctan\frac{1}{p+q+t}=\arctan\frac{1}{p+q}.$$

- 取 $a=p+r$、$b=u$、$c=v$: 条件 $uv=(p+r)^2+1$ 是已知的, 所以

$$\arctan\frac{1}{p+r+u}+\arctan\frac{1}{p+r+v}=\arctan\frac{1}{p+r}.$$

- 取 $a=p$、$b=q$、$c=r$: 条件 $qr=p^2+1$ 是已知的, 所以

$$\arctan\frac{1}{p+q}+\arctan\frac{1}{p+r}=\arctan\frac{1}{p}.$$

把首两式相加, 再利用第三式, 可知四个反正切之和为 $\arctan\frac{1}{p}$, 正是所求。$\blacksquare$

数值恒等式

我们需要 $p=7$, 所以 $qr=50$。与其猜测 q 和 r, 不如由那四个数倒推回去。若 13 和 21 是某次使用第一个恒等式时的一对 $a+b$、$a+c$, 则 $b=13-a$、$c=21-a$, 而 $bc=a^2+1$ 给出

$$(13-a)(21-a)=a^2+1 \iff 273-34a=1 \iff a=8.$$

同样地, 若 82 和 187 是一对, 则

$$(82-a)(187-a)=a^2+1 \iff 15334-269a=1 \iff a=57.$$

于是尝试取 $p+q=8$ 及 $p+r=57$, 即 $q=1$、$r=50$; 然后 $s=13-8=5$、$t=21-8=13$、$u=82-57=25$、$v=187-57=130$。七个数全为正数, 而三个条

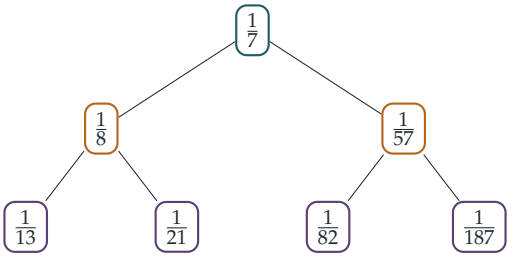
件均成立：

$$qr = 50 = 7^2 + 1, \qquad st = 65 = 8^2 + 1, \qquad uv = 3250 = 57^2 + 1.$$

取这些值时，$p+q+s = 13$、$p+q+t = 21$、$p+r+u = 82$ 及 $p+r+v = 187$，所以由四项恒等式可得

$$\arctan\frac{1}{13} + \arctan\frac{1}{21} + \arctan\frac{1}{82} + \arctan\frac{1}{187} = \arctan\frac{1}{7}. \qquad \blacksquare$$

下面的树显示 $\arctan\frac{1}{7}$ 如何拆开。



每个节点的反正切都等于其子节点的反正切之和：$1 \cdot 50 = 7^2 + 1$，$5 \cdot 13 = 8^2 + 1$，$25 \cdot 130 = 57^2 + 1$。

验算

- 以数值计算，最后那个恒等式的两边均为 0.1418970546…。
- 以分数作精确计算：对于 $\frac{1}{13}, \frac{1}{21}$，$\tan(A+B)$ 为 $\frac{34}{272} = \frac{1}{8}$；对于 $\frac{1}{82}, \frac{1}{187}$，其值为 $\frac{269}{15333} = \frac{1}{57}$；而对于 $\frac{1}{8}, \frac{1}{57}$，其值为 $\frac{65}{455} = \frac{1}{7}$。
- 第一个恒等式的一个特殊情况：取 $a = 1$、$b = 1$、$c = 2$，便得著名的 $\arctan\frac{1}{2} + \arctan\frac{1}{3} = \arctan 1 = \frac{\pi}{4}$。

常见错误

- 由 $\tan(A+B) = \frac{1}{a}$ 直接断定 $A + B = \arctan\frac{1}{a}$，而没有检查 $A+B$ 是否位于 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 之内。
- 盲目地逐一尝试 50、65 和 3250 的各种因数分解；其实由 13, 21 这一对倒推回去，一行便可求得 $a = 8$。

STEP 之外

这类恒等式正是梅钦类公式（Machin-type formulae）背后的原理，例如 $\frac{\pi}{4} = 4\arctan\frac{1}{5} - \arctan\frac{1}{239}$；数百年来，人们一直用这些公式以人手计算 π，因为当 x 很小时，$\arctan x$ 的级数（series）收敛（converge）得很快。要看清这些恒等式，最简洁的途径是通过复数（complex numbers）：$\arctan\frac{1}{n}$ 是 $n+\mathrm{i}$ 的辐角（argument），所以反正切相加就对应于高斯整数（Gaussian integers）相乘，而条件 $bc = a^2+1$ 其实是 $a^2+1 = (a+\mathrm{i})(a-\mathrm{i})$ 的一个因式分解，只是换了个模样。Needham 的 *Visual Complex Analysis* 深入阐述了这种从几何角度看乘法的观点，而 Davenport 的 *The Higher Arithmetic* 则讨论了二平方和（sums of two squares）。

详细解答

STEP II 2005，第 5 题

纯数学 · 内切圆与外接圆

STEP II 2005，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/05-s2-q5

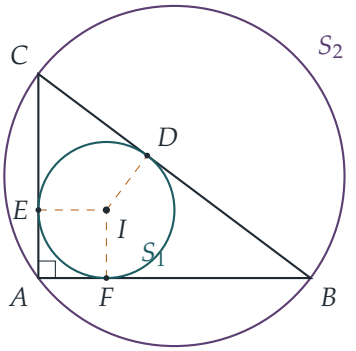
关键思路

由圆外一点向圆所作的两条切线（tangent）长度相等。把这个性质用于每个顶点，三角形的每条边便分成两段切线长，而在直角那个顶点，两段切线长刚好都是 r。这样几乎立刻就得到 r 的公式。至于面积比，直角告诉我们斜边（hypotenuse）是 S_2 的直径，而毕氏定理（Pythagoras′ theorem）让我们把一切只用 $b+c$ 和 a 表示，这就是 q 出现的原因。

解答

内切圆半径

设 S_1 的圆心为 I，并分别与 BC、CA、AB 相切于 D、E、F。



半径与切线互相垂直，所以 $IE \perp AC$ 及 $IF \perp AB$。连同 A 处的直角，可知 $AEIF$ 是长方形，而且它的一对邻边 $IE = IF = r$，所以它是正方形，并且

$$AE = AF = r.$$

由圆外一点所作的切线长度相等，所以 $BF = BD$ 及 $CE = CD$。因此

$$c = AF + FB = r + BD, \qquad b = AE + EC = r + CD, \qquad a = BD + DC.$$

把首两式相加，再减去第三式，

$$b + c - a = 2r + BD + CD - (BD + DC) = 2r,$$

这正是所需的结果。

面积比

由于 $\angle BAC = 90°$，根据半圆上的圆周角定理（angle in a semicircle theorem）的逆定理，BC 是外接圆（circumcircle）S_2 的直径。所以 S_2 的半径是 $\frac{1}{2}a$，面积是 $\frac{1}{4}\pi a^2$。

S_1 与三角形之间的区域，就是三角形除去 S_1 的圆盘后余下的部分，所以它的面积是 $\frac{1}{2}bc - \pi r^2$。因此

$$R = \frac{\frac{1}{2}bc - \pi r^2}{\frac{1}{4}\pi a^2}, \qquad \text{所以} \qquad \pi R = \frac{2bc - 4\pi r^2}{a^2} = \frac{2bc}{a^2} - \pi\,\frac{(b + c - a)^2}{a^2},$$

这里用了 $2r = b + c - a$。

想一想 怎样把 $2bc/a^2$ 化成只含 q 的式子?

由毕氏定理得 $b^2 + c^2 = a^2$，所以 $2bc = (b + c)^2 - (b^2 + c^2) = (b + c)^2 - a^2$。除以 a^2，

$$\frac{2bc}{a^2} = q^2 - 1, \qquad \frac{(b + c - a)^2}{a^2} = \left(\frac{b + c}{a} - 1\right)^2 = (q - 1)^2.$$

因此

$$\begin{aligned} \pi R &= q^2 - 1 - \pi(q^2 - 2q + 1) \\ &= -(\pi - 1)q^2 + 2\pi q - (\pi + 1), \end{aligned}$$

这就是所求的结果。

R 的上界

右边是 q 的二次式，首项系数 $-(\pi-1)$ 为负数，所以我们用配方法（completing the square）：

$$\begin{aligned}\pi R &= -(\pi-1)\left(q^2 - \frac{2\pi}{\pi-1}\,q\right) - (\pi+1)\\ &= -(\pi-1)\left(q - \frac{\pi}{\pi-1}\right)^2 + \frac{\pi^2}{\pi-1} - (\pi+1)\\ &= -(\pi-1)\left(q - \frac{\pi}{\pi-1}\right)^2 + \frac{\pi^2-(\pi^2-1)}{\pi-1}\\ &= \frac{1}{\pi-1} - (\pi-1)\left(q - \frac{\pi}{\pi-1}\right)^2.\end{aligned}$$

由于 $\pi-1>0$，平方项是被减去的，所以对任何实数 q 都有 $\pi R \le \dfrac{1}{\pi-1}$；再除以 $\pi>0$，

$$R \le \frac{1}{\pi(\pi-1)}.$$

这就是题目要求我们推导的结论。

值得留意的是，这个上界其实并不能达到。等号成立需要 $q=\pi/(\pi-1)\approx 1.467$，但对直角三角形而言，$(b+c)^2 = a^2+2bc \le a^2+(b^2+c^2) = 2a^2$（由算术平均—几何平均不等式（AM–GM），$2bc \le b^2+c^2$），所以 $q \le \sqrt{2} \approx 1.414$。$R$ 的最大可能值来自等腰三角形，即 $q=\sqrt{2}$，此时 $R = (2\sqrt{2}\,\pi - 3\pi + 1)/\pi \approx 0.1467$，仅略低于上界 $1/(\pi(\pi-1)) \approx 0.1486$。

验算

对于 3–4–5 三角形（$b=4$，$c=3$，$a=5$），公式给出 $r=1$；而面积 6 除以半周长（semi-perimeter）6 的确是 1，这正是标准公式 $r=$ 面积$/s$。于是 $R = (6-\pi)/(\frac{25}{4}\pi) \approx 0.14558$，而把 $q=1.4$ 代入二次式也得到相同的值。随机测试了 10^5 个直角三角形，每一次都验证了 πR 的恒等式，而且没有发现 R 超过 0.14674，与上界 0.14863 吻合。

常见错误

- 忘记 S_2 的半径是 $\frac{1}{2}a$ 而不是 a：斜边是直径。
- 试图直接展开 $2bc/a^2$，而没有利用 $b^2+c^2=a^2$ 把它写成 $2bc=(b+c)^2-a^2$。
- 没有检查 $q=\pi/(\pi-1)$ 是否可以取到，就声称上界可以达到；其实并不能达到，不过题目只要求证明不等式。

STEP 之外

一般三角形的内切圆（incircle）与外接圆，由欧拉（Euler）公式 $d^2 = R_2(R_2 - 2r)$ 联系起来，其中 d 是两圆圆心之间的距离，R_2 是外接圆半径（circumradius）；由此可得 $R_2 \geq 2r$，而等号只在等边三角形时成立。这类结果，以及在一族图形中求某个比值的最优值这种想法，都通往几何不等式（geometric inequalities）的研究。Coxeter 与 Greitzer 合著的 *Geometry Revisited*（《几何学的再探索》）是很优美的下一步读物，而 Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》，免费，采用 CC BY 授权）则有更多 STEP 风格的练习。

详细解答

STEP II 2005，第 6 题

纯数学 · 二项式级数的和

STEP II 2005，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/05-s2-q6

关键思路

这里每一个和都是某个幂级数（power series）在特定的 x 处的取值，所以要做的就是认出它来自哪一个函数。在第 (i) 部分，窍门是把 n 和 n^2 改写成 1、$n+1$ 和 $(n+1)(n+2)$ 的组合，而这些正是我们刚刚写下的系数。在第 (ii) 部分，只要留意到 $(1-x)^{-1/2}$ 和 $(1-x)^{-3/2}$ 的系数之间的关系，同样的想法就行得通。

解答

第 (i) 部分

根据负整数指数的二项式定理（binomial theorem）（当 $|x|<1$ 时成立），对于 $n \geq 0$，x^n 的系数（coefficient）为

$$(1-x)^{-1}: \quad 1, \qquad (1-x)^{-2}: \quad n+1, \qquad (1-x)^{-3}: \quad \tfrac{1}{2}(n+1)(n+2).$$

例如，$(1-x)^{-3}$ 中的系数是 $\dfrac{(-3)(-4)\cdots(-3-n+1)}{n!}(-1)^n = \dfrac{3\cdot 4\cdots(n+2)}{n!} = \dfrac{(n+1)(n+2)}{2}$。

第一个和。由于 $n=(n+1)-1$，当 $|x|<1$ 时

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

（把和拆开是可以的，因为右边两个级数都收敛（converge）。）$n=0$ 的一项为零，所以代入 $x=\frac{1}{2}$，

$$\sum_{n=1}^{\infty} n\,2^{-n} = 4-2 = 2.$$

第二个和。我们想把 n^2 写成 $(n+1)(n+2)$、$n+1$ 和 1 的组合。由于 $(n+1)(n+2) = n^2+3n+2$，

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

因此，当 $|x|<1$ 时，

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

其中出现因子 2，是因为 $(1-x)^{-3}$ 中的系数是 $(n+1)(n+2)$ 的一半。同样，$n=0$ 的一项为零，而在 $x=\frac{1}{2}$ 处

$$\sum_{n=1}^{\infty} n^2 2^{-n} = 2\cdot 8 - 3\cdot 4 + 2 = 6.$$

第 (ii) 部分

根据指数为 $-\frac{1}{2}$ 的二项式定理，$(1-x)^{-1/2}$ 中 x^n 的系数是

$$\frac{(-\frac{1}{2})(-\frac{3}{2})\cdots(-\frac{2n-1}{2})}{n!}(-1)^n = \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2^n\, n!}.$$

想一想 怎样把奇数的连乘积化成阶乘（factorial）？

把缺少的偶数补上：$(2n)! = \big(1\cdot 3\cdots(2n-1)\big)\big(2\cdot 4\cdots 2n\big)$，而 $2\cdot 4\cdots 2n = 2^n n!$。所以 $1\cdot 3\cdots(2n-1) = \dfrac{(2n)!}{2^n n!}$，而系数就是

$$a_n = \frac{(2n)!}{2^n n!\cdot 2^n n!} = \frac{(2n)!}{(n!)^2}\frac{1}{2^{2n}}.$$

这对 $n=0$ 也成立（空积，$a_0=1$），所以当 $|x|<1$ 时

$$(1-x)^{-1/2} = \sum_{n=0}^{\infty} \frac{(2n)!}{(n!)^2}\frac{x^n}{2^{2n}},$$

正如所求。

第一个和。题目所求的和是 $\sum a_n x^n$ 在 $x=\frac{1}{3}$ 处的值，而这一点位于展开式成立的区间之内。所以它等于

$$\left(1-\tfrac{1}{3}\right)^{-1/2} = \sqrt{\tfrac{3}{2}} = \frac{\sqrt{6}}{2}.$$

第二个和。现在我们需要 $\sum n a_n x^n$。按照第 (i) 部分的思路，把 na_n 与一个我们已知的级数联系起来。与上面完全一样，$(1-x)^{-3/2}$ 中 x^m 的系数是

$$b_m = \frac{3\cdot 5\cdots(2m+1)}{2^m\, m!} = (2m+1)\, a_m.$$

由 a_n 的公式，相邻两个系数之比是 $a_n/a_{n-1} = \dfrac{(2n)(2n-1)}{n^2\cdot 4} = \dfrac{2n-1}{2n}$，所以对于 $n \geq 1$

$$n\,a_n = \tfrac{1}{2}(2n-1)\,a_{n-1} = \tfrac{1}{2}\,b_{n-1}.$$

因此，当 $|x| < 1$ 时，

$$\sum_{n=1}^{\infty} n a_n x^n = \frac{x}{2}\sum_{n=1}^{\infty} b_{n-1}x^{n-1} = \frac{x}{2}(1-x)^{-3/2}.$$

（等价的做法是：把 $(1-x)^{-1/2}$ 的级数逐项求导（differentiate term by term），这在它的收敛区间（interval of convergence）内是容许的，然后乘以 x。）在 $x = \frac{1}{3}$ 处：

$$\sum_{n=1}^{\infty} \frac{n\,(2n)!}{(n!)^2\,2^{2n}\,3^n} = \frac{1}{6}\left(\frac{3}{2}\right)^{3/2} = \frac{1}{6}\cdot\frac{3}{2}\sqrt{\frac{3}{2}} = \frac{\sqrt{6}}{8}.$$

验算

- $\sum n2^{-n}$ 的首几项是 $\frac{1}{2}+\frac{1}{2}+\frac{3}{8}+\frac{1}{4}+\cdots$，加到第四项已经是 1.625，并慢慢向 2 靠近；$\sum n^2 2^{-n}$ 加到 200 项的部分和（partial sum）与 6 吻合至小数点后很多个位。
- 数值上，$\sqrt{6}/2 \approx 1.224745$ 而 $\sqrt{6}/8 \approx 0.306186$；第 (ii) 部分两个级数的部分和与这两个值吻合至小数点后 12 个位。
- 对通项作一个简单的合理性检查：$n = 1$ 给出 $a_1 = \frac{2}{4} = \frac{1}{2}$，正是 $(1-x)^{-1/2} = 1 + \frac{1}{2}x + \frac{3}{8}x^2 + \cdots$ 中那个熟悉的系数。

常见错误

- 忘记了 $(1-x)^{-3}$ 的系数中的 $\frac{1}{2}$，结果得出 $8 - 12 + 2 = -2$ 作为 $\sum n^2 2^{-n}$ 的值，而正项之和不可能是负数。
- 把 $(1-x)^{-2}$ 的级数的系数写成 n 而不是 $n+1$：这个级数是 $1 + 2x + 3x^2 + \cdots$。
- 在最后一个和中，漏掉一个因子 x：$\sum n a_n x^{n-1}$ 是导数，但题目所求的和是 $\sum n a_n x^n$。

STEP 之外

这里用到的手法（把级数拆成几部分、逐项求导）之所以合理，是因为幂级数在其收敛半径（radius of convergence）之内绝对收敛（converges absolutely），并且可以在那里逐项求导。严谨地证明这一点，是大学第一门分析（analysis）课程的标准内容。系数 $\binom{2n}{n}/4^n$ 也会在概率中出现，它就是简单随机漫步（simple random walk）在走了 $2n$ 步之后回到起点的概率。关于分析方面的内容，可参阅 Abbott 的 *Understanding Analysis*（一本分析入门书）。

详细解答

STEP II 2005，第 7 题

纯数学 · 向量与圆周运动

> **STEP II 2005，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/05-s2-q7

关键思路

每个位置向量（position vector）都具有 cos(角) $\mathbf{a}$ + sin(角) $\mathbf{b}$ 的形式；当 $\mathbf{a}$ 与 $\mathbf{b}$ 互相垂直且长度相等时，它在 $\mathbf{a}$ 与 $\mathbf{b}$ 所在的平面上描出一个圆。一旦知道 $|OP|$ 和 $|OQ|$ 是常数，角 POQ 便可直接由纯量积（scalar product）求得，再用积化和差公式（product-to-sum formulae）把它整理成单一个余弦。最后一部分就变成在两个完整周期内关于一个余弦的不等式。

解答

第 (i) 部分

以 $\mathbf{p}$ 和 $\mathbf{q}$ 表示 P 和 Q 的位置向量。

对于 P，$\mathbf{p} = \cos t\,\mathbf{i} + \sin t\,\mathbf{j}$，所以 $|\mathbf{p}| = 1$，而且 $\mathbf{p}$ 位于 xy 平面上。因此 P 在平面 $z = 0$ 上沿着以 O 为圆心、半径为 1 的圆运动；从上方（即从正 z 轴方向）看是逆时针方向，角速率（angular speed）恒为 1（所以速率为 1）。它由 $(1, 0, 0)$ 出发，当 t 由 0 变到 2π 时刚好转一圈。

对于 Q，设 $s = t + \frac{1}{4}\pi$，并设

$$\mathbf{a} = \tfrac{3}{2}\mathbf{i} + \tfrac{3\sqrt{3}}{2}\mathbf{k}, \qquad \mathbf{b} = 3\mathbf{j}, \qquad \text{于是} \qquad \mathbf{q} = \cos s\,\mathbf{a} + \sin s\,\mathbf{b}.$$

现在 $|\mathbf{a}|^2 = \frac{9}{4} + \frac{27}{4} = 9$，$|\mathbf{b}|^2 = 9$，且 $\mathbf{a} \cdot \mathbf{b} = 0$。因此

$$|\mathbf{q}|^2 = 9\cos^2 s + 9\sin^2 s + 2\cos s \sin s\,(\mathbf{a} \cdot \mathbf{b}) = 9,$$

而且 $\mathbf{q}$ 始终位于通过 O、由 $\mathbf{a}$ 和 $\mathbf{j}$ 张成的平面上。该平面包含 y 轴和方向 $\mathbf{i} + \sqrt{3}\,\mathbf{k}$，所以它的方程是 $z = \sqrt{3}\,x$；它的法线（normal）$-\sqrt{3}\,\mathbf{i} + \mathbf{k}$ 与 $\mathbf{k}$ 的夹角为 $\arccos \frac{1}{2} = \frac{1}{3}\pi$。

所以 Q 在平面 $z = \sqrt{3}\,x$ 上沿着以 O 为圆心、半径为 3 的圆运动；这个平面包含 y 轴，并与 P 的圆所在的平面成 60° 倾角。它的角速率同样是 1（速率为 3），也是转一圈，由 $\mathbf{a}$ 的方向转向 $\mathbf{j}$，而它的相位（phase）始终比 P 的领先 $\frac{1}{4}\pi$：在 $t = 0$ 时，它位于与 $\mathbf{a}$ 成 $\frac{1}{4}\pi$ 角之处，即 $\frac{3}{2\sqrt{2}}(1,\, 2,\, \sqrt{3})$。

第 (ii) 部分

由于 $|\mathbf{p}| = 1$ 及 $|\mathbf{q}| = 3$，

$$\cos\theta = \frac{\mathbf{p}\cdot\mathbf{q}}{3} = \frac{\frac{3}{2}\cos t\cos s + 3\sin t\sin s}{3} = \tfrac{1}{2}\cos t\cos s + \sin t\sin s,$$

其中 $s = t + \frac{1}{4}\pi$。利用积化和差公式，并注意 $t + s = 2t + \frac{1}{4}\pi$ 及 $s - t = \frac{1}{4}\pi$，可得

$$\cos t\cos s = \tfrac{1}{2}\big[\cos(2t + \tfrac{1}{4}\pi) + \cos\tfrac{1}{4}\pi\big],$$
$$\sin t\sin s = \tfrac{1}{2}\big[\cos\tfrac{1}{4}\pi - \cos(2t + \tfrac{1}{4}\pi)\big].$$

记 $C = \cos(2t + \frac{1}{4}\pi)$，并利用 $\cos\frac{1}{4}\pi = \frac{\sqrt{2}}{2}$，

$$\begin{aligned}\cos\theta &= \tfrac{1}{4}\big(C + \tfrac{\sqrt{2}}{2}\big) + \tfrac{1}{2}\big(\tfrac{\sqrt{2}}{2} - C\big)\\ &= \tfrac{\sqrt{2}}{8} + \tfrac{\sqrt{2}}{4} - \tfrac{1}{4}C = \tfrac{3\sqrt{2}}{8} - \tfrac{1}{4}\cos(2t + \tfrac{1}{4}\pi),\end{aligned}$$

这正是所要证明的结果。

第 (iii) 部分

> **想一想** 怎样把 $\theta \geq \frac{1}{4}\pi$ 转化为关于 $\cos\theta$ 的条件?

余弦在 $[0, \pi]$ 上严格递减（strictly decreasing），而 θ 和 $\frac{1}{4}\pi$ 都在这个区间内，所以

$$\theta \geq \tfrac{1}{4}\pi \iff \cos\theta \leq \cos\tfrac{1}{4}\pi = \tfrac{\sqrt{2}}{2} = \tfrac{4\sqrt{2}}{8}.$$

由第 (ii) 部分，这等价于

$$\tfrac{3\sqrt{2}}{8} - \tfrac{1}{4}\cos(2t + \tfrac{1}{4}\pi) \leq \tfrac{4\sqrt{2}}{8} \iff \cos(2t + \tfrac{1}{4}\pi) \geq -\tfrac{\sqrt{2}}{2}.$$

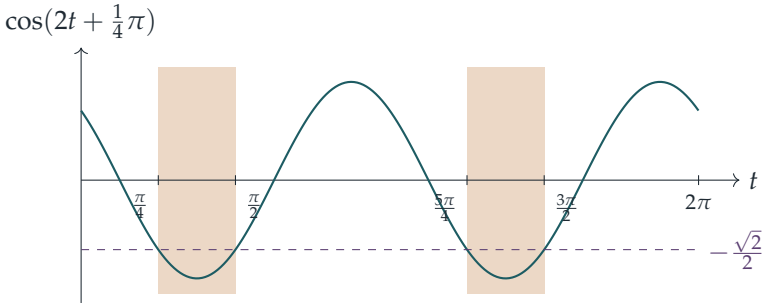
设 $u = 2t + \frac{1}{4}\pi$。当 t 取遍 $[0, 2\pi]$ 时，u 取遍 $[\frac{1}{4}\pi, \frac{17}{4}\pi]$，这个区间的长度是 4π：刚好是 $\cos u$ 的两个周期。在任何一个周期内，$\cos u < -\frac{\sqrt{2}}{2}$ 恰好在一个长度为 $\frac{1}{2}\pi$ 的开区间上成立（即 $\frac{3}{4}\pi$ 与 $\frac{5}{4}\pi$ 之间，模 2π）。在这里，不符合条件的值是

$$u \in (\tfrac{3}{4}\pi, \tfrac{5}{4}\pi) \cup (\tfrac{11}{4}\pi, \tfrac{13}{4}\pi) \iff t \in (\tfrac{1}{4}\pi, \tfrac{1}{2}\pi) \cup (\tfrac{5}{4}\pi, \tfrac{3}{2}\pi),$$

以 t 计算，每段的长度为 $\frac{1}{4}\pi$。所以 $\theta < \frac{1}{4}\pi$ 的总时间是 $\frac{1}{2}\pi$，而 $\theta \geq \frac{1}{4}\pi$ 的总时间是

$$2\pi - \tfrac{1}{2}\pi = \tfrac{3}{2}\pi,$$

这正是所要证明的结果。



阴影部分是曲线降到 $-\frac{\sqrt{2}}{2}$ 以下的地方，也就是 $\theta < \frac{1}{4}\pi$ 的地方。

验算

- 在 $t = 0$ 时：$\mathbf{p} = \mathbf{i}$ 及 $\mathbf{q} = \frac{3}{2\sqrt{2}}(\mathbf{i} + 2\mathbf{j} + \sqrt{3}\,\mathbf{k})$，所以直接计算得 $\cos\theta = \frac{3}{2\sqrt{2}} \cdot \frac{1}{3} = \frac{\sqrt{2}}{4}$。公式则给出 $\frac{3\sqrt{2}}{8} - \frac{1}{4} \cdot \frac{\sqrt{2}}{2} = \frac{\sqrt{2}}{4}$。两者吻合。
- $\cos\theta$ 介乎 $\frac{3\sqrt{2}}{8} - \frac{1}{4} \approx 0.280$ 与 $\frac{3\sqrt{2}}{8} + \frac{1}{4} \approx 0.780$ 之间，所以 θ 永远不会是 0。这是合理的：$\mathbf{p}$ 与 $\mathbf{q}$ 只可能沿着两个平面相交的 y 轴平行，但 P 在 $t = \frac{1}{2}\pi, \frac{3}{2}\pi$ 时位于该轴上，而 Q 则在 $t = \frac{1}{4}\pi, \frac{5}{4}\pi$ 时位于该轴上。
- 在 $[0, 2\pi]$ 内对 t 取二百万个样本点，并直接由向量计算 θ，得出 $\theta \geq \frac{1}{4}\pi$ 的总时间为 4.712389，与 $\frac{3}{2}\pi \approx 4.712389$ 相符。

常见错误

- 忘记除以 $|\mathbf{p}||\mathbf{q}| = 3$，或因为 $\mathbf{i}$ 和 $\mathbf{j}$ 的系数不同而以为 $|\mathbf{q}|$ 会变化。
- 把不等号的方向弄反：$\theta \geq \frac{1}{4}\pi$ 表示 $\cos\theta \leq \frac{\sqrt{2}}{2}$，因为余弦在 $[0, \pi]$ 上递减。
- 只计算了 $\cos(2t + \frac{1}{4}\pi)$ 的一个周期：$2t$ 表示当 t 转一圈时，它转了两圈。

STEP 之外

任何以原点为圆心的圆都可以写成 $\cos s\,\mathbf{a} + \sin s\,\mathbf{b}$，其中 $\mathbf{a} \perp \mathbf{b}$ 且 $|\mathbf{a}| = |\mathbf{b}|$；如果去掉这些条件，便会得到一个椭圆（ellipse），而一般来说，得到的是单位圆在线性映射（linear map）下的像。以这种方式理解曲线，并用纯量积量度运动中的向量之间的夹角，正是大学一年级所学的线性代数（linear algebra）和向量微积分（vector calculus）的起点。Strang 的 *Introduction to Linear Algebra*（线性代数入门）把这些概念阐述得十分清晰。

详细解答

STEP II 2005，第 8 题

纯数学 · 可分离变量微分方程

STEP II 2005，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/05-s2-q8

关键思路

这条方程是可分离变量（separable）的：与 y 有关的部分只有 y^2 这一个因子，所以除以它之后便得到 $\frac{1}{y^2}\frac{\mathrm{d}y}{\mathrm{d}x}$，也就是 $-1/y$ 的导数。对 x 的积分可用代换（substitution）$u = 1 + x^2$ 来处理，因为奇数次幂 x^3 正好提供了 u 所需要的 $x\,\mathrm{d}x$。至于第二条曲线，留意右边是 z^3 而不是 y^2：这时应该考虑的量是 $1/z^2$，而它原来与 $1/y$ 满足完全相同的方程。

解答

求 y

在 C 上有 $y(0) = 1 > 0$，而当 $x \geq 0$ 时方程的右边永不为负，所以 y 永不递减。因此 $y \geq 1$ 恒成立，我们可以除以 y^2：

$$\frac{1}{y^2}\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{x^3}{(1+x^2)^{5/2}} \quad\Longrightarrow\quad -\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{y}\right) = \frac{x^3}{(1+x^2)^{5/2}}.$$

计算积分时，设 $u = 1 + x^2$，于是 $\mathrm{d}u = 2x\,\mathrm{d}x$，而 $x^3\,\mathrm{d}x = x^2 \cdot x\,\mathrm{d}x = \frac{1}{2}(u-1)\,\mathrm{d}u$：

$$\begin{aligned}\int \frac{x^3}{(1+x^2)^{5/2}}\,\mathrm{d}x &= \frac{1}{2}\int (u^{-3/2} - u^{-5/2})\,\mathrm{d}u = -u^{-1/2} + \tfrac{1}{3}u^{-3/2} + \text{常数} \\ &= -\frac{3u-1}{3u^{3/2}} + \text{常数} = -\frac{2+3x^2}{3(1+x^2)^{3/2}} + \text{常数}.\end{aligned}$$

因此

$$\frac{1}{y} = \frac{2+3x^2}{3(1+x^2)^{3/2}} + k$$

其中 k 为某个常数。在 $x=0$ 处，上式变成 $1=\frac{2}{3}+k$，所以 $k=\frac{1}{3}$，这就是所要求的结果：

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

x 很大时

把主导的幂次提出来：从分子提出 $3x^2$，从 $(1+x^2)^{3/2}$ 提出 x^3：

$$\frac{2+3x^2}{3(1+x^2)^{3/2}}=\frac{3x^2\left(1+\frac{2}{3x^2}\right)}{3x^3\left(1+\frac{1}{x^2}\right)^{3/2}}=\frac{1}{x}\left(1+\frac{2}{3x^2}\right)\left(1-\frac{3}{2x^2}+\cdots\right)=\frac{1}{x}+O\left(\frac{1}{x^3}\right),$$

这里用了二项展开（binomial expansion）；由于 $1/x^2<1$ 对很大的 x 成立，展开式有效。所以

$$\frac{1}{y}=\frac{1}{3}+\frac{1}{x}+O\left(\frac{1}{x^3}\right)=\frac{1}{3}\left(1+\frac{3}{x}+O\left(\frac{1}{x^3}\right)\right),$$

再取倒数，并利用 $(1+t)^{-1}=1-t+O(t^2)$，得

$$y=3\left(1-\frac{3}{x}+O\left(\frac{1}{x^2}\right)\right)\approx 3-\frac{9}{x},$$

正如题目所求。

> ? **想一想** 为甚么可以舍去 $\frac{2}{3x^2}$ 和 $-\frac{3}{2x^2}$ 这两项，却不能舍去 $\frac{1}{x}$？

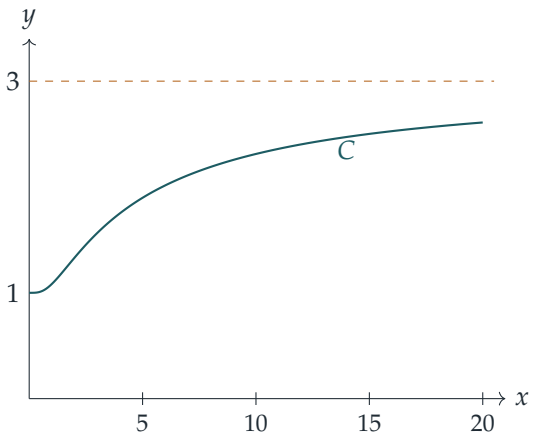
目标 $3-\frac{9}{x}$ 准确至 $\frac{1}{x}$ 阶。在 $1/y$ 中，各修正项比 $\frac{1}{x}$ 还要小一个 $\frac{1}{x^2}$ 的因子，所以它们只会在 $\frac{1}{x^2}$ 阶或更高阶影响 y。而 $\frac{1}{x}$ 这一项本身，正是产生 $-\frac{9}{x}$ 的来源。

C 的略图

现在把曲线的特征整理出来。

- $y(0)=1$,而且 $\mathrm{d}y/\mathrm{d}x=0$ 于 $x=0$ 处成立。在原点附近 $\mathrm{d}y/\mathrm{d}x\approx x^3$,所以 $y\approx 1+\frac{1}{4}x^4$：曲线起初非常平坦。
- 当 $x>0$ 时斜率严格为正，所以 y 递增。
- $1/y$ 的第一项为正，所以 $1/y>\frac{1}{3}$，由此得 $1\le y<3$ 对所有 x 成立。曲线永不会趋向无穷大。
- 当 $x\to\infty$ 时，$y\to 3$ 从下方趋近，形如 $3-\frac{9}{x}$。这个趋近过程很慢：$y(10)\approx 2.31$。

由于曲线起初向上凹,其后渐趋平缓并靠近 $y=3$,它必定有一个拐点(point of inflection)。



z 的曲线

只要 $z \neq 0$，我们就可以除以 z^3。由于 $\dfrac{\mathrm{d}}{\mathrm{d}x}\left(z^{-2}\right) = -2z^{-3}\dfrac{\mathrm{d}z}{\mathrm{d}x}$，

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{z^2}\right) = -\frac{2}{z^3}\cdot\frac{x^3z^3}{2(1+x^2)^{5/2}} = -\frac{x^3}{(1+x^2)^{5/2}},$$

这正是 $1/y$ 所满足的方程。在两条曲线上，$1/z^2 = 1$ 于 $x = 0$ 处成立，与 $1/y$ 的值相同。所以

$$\frac{1}{z^2} = \frac{1}{y}, \qquad \text{即} \qquad z^2 = y.$$

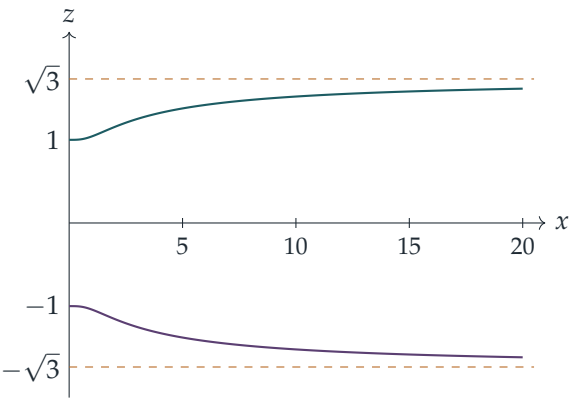
这同时确认了 z 永不会到达 0，从而说明刚才的除法是合理的。只要 $z \neq 0$，便有 $z^2 = y \geq 1$，所以 z 不可能接近 0。根据连续性，z 会一直保持起始时的正负号，于是两条曲线分别是

$$z = \sqrt{y} \quad (z(0) = 1), \qquad z = -\sqrt{y} \quad (z(0) = -1).$$

上方的曲线由 1 上升并趋向 $\sqrt{3}$，同样是起初很平坦（这次是 $z \approx 1 + \frac{1}{8}x^4$），而当 x 很大时

$$z = \sqrt{3}\left(1 - \frac{3}{x} + \cdots\right)^{1/2} \approx \sqrt{3} - \frac{3\sqrt{3}}{2x}.$$

下方的曲线是它对 x 轴的反射：它由 -1 下降，趋向渐近线（asymptote）$z = -\sqrt{3}$。



想一想 为甚么 $z(0) = -1$ 的那条曲线，不只是把另一条曲线平移或作其他变换而得的复制品?

这条方程对 z 而言是奇的：把 z 换成 $-z$，两边都会乘上因子 -1。所以若 $z(x)$ 是解，则 $-z(x)$ 也是解，而由 -1 出发的解就是由 $+1$ 出发的解的反射。这与 $z = \pm\sqrt{y}$ 吻合。

验算

- 对答案求导：由 $1/y = u^{-1/2} - \frac{1}{3}u^{-3/2} + \frac{1}{3}$，可得 $\mathrm{d}(1/y)/\mathrm{d}x = 2x\left(-\frac{1}{2}u^{-3/2} + \frac{1}{2}u^{-5/2}\right) = -x^3u^{-5/2}$，正合所需。
- 以数值方法（龙格－库塔法，Runge–Kutta）解原方程，得 $y(1) = 1.08391$ 和 $y(5) = 1.89777$，与公式吻合至 10 位或以上有效数字。以数值方法解 z 的方程，得 $z(5) = \pm 1.37759 = \pm\sqrt{y(5)}$。
- 在 $x = 100$ 处，公式给出 $y = 2.9126$，而 $3 - \frac{9}{x} = 2.91$。两者相差约 $27/x^2$，与 $O(1/x^2)$ 的误差一致。

常见错误

- 漏掉负号：$\int y^{-2}\,\mathrm{d}y = -y^{-1}$。
- 断定 $y \to 3$ 的速度很快，理由是 $3 - \frac{9}{x}$“看来很接近”3。其实在 $x = 10$ 处，曲线仍只到达约 2.3。
- 两条曲线都写成 $z = \sqrt{y}$，或者没有解释为甚么 z 不能变号。

STEP 之外

两条方程都具有 $\mathrm{d}y/\mathrm{d}x = f(x)\,y^n$ 的形式，这是 伯努利方程（*Bernoulli equation*）的特殊情况；代换 $w = y^{1-n}$ 总能把这类方程化为线性方程。我们用的正是这个代换：$w = 1/y$ 和 $w = 1/z^2$。“方程相同，初值相同，所以函数相同”这一步，是一阶方程的存在唯一性定理（existence and uniqueness theorem，即 Picard–Lindelöf 定理）的一个例子。Boyce 与 DiPrima 合著的 *Elementary Differential Equations*（初等微分方程教科书）对两者都有论述。

详细解答

STEP II 2005，第 9 题

力学 · 斜面上的摩擦力

STEP II 2005，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/05-s2-q9

关键思路

把两个质点（particle）分开处理。A 没有受到外加的力，所以由它的方程便可确定绳子必须提供的张力（tension）。所有可供选择的余地，都在于 P 作用于 B 的方向。把 P 倾斜会改变 B 所受的法向反作用力（normal reaction），从而改变 B 所受的摩擦力（friction）。令 P 偏离斜面，会浪费它沿斜面方向的部分分量，却同时改变了摩擦力；最佳的倾斜角度正好令这两种效应取得平衡。最后问题归结为求形如 $a\cos\theta + b\sin\theta$ 的式子的最大值。

解答

我们把 A 和 B 视为质点，并把绳子视为轻而不可伸长（light and inextensible），且沿 ℓ 放置。我们取 P 作用于包含 ℓ 的铅垂平面内。P 横过斜面的分量毫无帮助。注意

$$\sin 30^\circ = \tfrac{1}{2}, \qquad \cos 30^\circ = \tfrac{\sqrt{3}}{2}, \qquad \tan 30^\circ = \tfrac{1}{\sqrt{3}} = \tfrac{1}{3}\sqrt{3}.$$

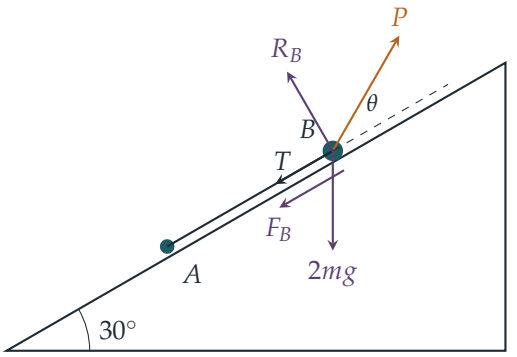
所以 $\mu_B = \tan 30^\circ$：若只有 B 本身，它恰好处于即将滑动的状态。另外 $\mu_A = \frac{1}{2}\tan 30^\circ$，所以若只有 A 本身，它会滑下。

质点 A。沿垂直于斜面的方向分解，得 $R_A = mg\cos 30^\circ = \frac{\sqrt{3}}{2}mg$，所以 A 所受的最大摩擦力为

$$\mu_A R_A = \tfrac{\sqrt{3}}{6}\cdot\tfrac{\sqrt{3}}{2}mg = \tfrac{1}{4}mg.$$

重量沿斜面向下的分量为 $mg\sin 30^\circ = \frac{1}{2}mg$。

质点 B。设 P 与 ℓ 向上的方向成角 θ，以离开斜面的一方为正。（θ 为负表示 P 把质点压向斜面。）设 T 为张力。



图中所示的摩擦力 F_B 沿斜面向下，即第 (i) 部分的情况。沿垂直于斜面的方向，

$$R_B + P\sin\theta = 2mg\cos 30^\circ \quad\Longrightarrow\quad R_B = \sqrt{3}\,mg - P\sin\theta.$$

所以 B 处可提供的最大摩擦力为

$$\mu_B R_B = \tfrac{1}{\sqrt{3}}\left(\sqrt{3}\,mg - P\sin\theta\right) = mg - \tfrac{1}{\sqrt{3}}P\sin\theta.$$

B 的重量沿斜面向下的分量为 $2mg\sin 30^\circ = mg$。

第 (i) 部分

要令两个质点沿 ℓ 向上移动，每个质点所受的摩擦力都沿斜面向下。当 P 取最小值时，两者都处于即将移动的状态，所以两者所受的摩擦力都是极限摩擦力（limiting friction）。

要令 A 向上移动，绳子必须克服重量的分量和摩擦力：

$$T = \tfrac{1}{2}mg + \tfrac{1}{4}mg = \tfrac{3}{4}mg.$$

对于 B，沿斜面方向，P 的分量必须克服 T、重量的分量和极限摩擦力：

$$P\cos\theta = \tfrac{3}{4}mg + mg + \left(mg - \tfrac{1}{\sqrt{3}}P\sin\theta\right),$$

$$P\left(\cos\theta + \tfrac{1}{\sqrt{3}}\sin\theta\right) = \tfrac{11}{4}mg.$$

? **想一想** 把 P 倾斜离开斜面，明明会减少 P 沿斜面的分量，为甚么反而有帮助?

把 B 稍微向上提，会减少 R_B，从而减少 P 需要克服的摩擦力。当倾斜角 θ 很小时，$\cos\theta$ 只按 $1 - \frac{1}{2}\theta^2$ 的方式下降，而 $\sin\theta$ 则按 θ 的方式增长。所以起初是摩擦力减少所带来的好处占优。要求出最佳角度，可写成

$$\cos\theta + \tfrac{1}{\sqrt{3}}\sin\theta = \tfrac{2}{\sqrt{3}}\left(\tfrac{\sqrt{3}}{2}\cos\theta + \tfrac{1}{2}\sin\theta\right) = \tfrac{2}{\sqrt{3}}\cos(\theta - 30^\circ).$$

此式最大为 $\frac{2}{\sqrt{3}}$，而等号恰好在 $\theta = 30^\circ$ 时成立。由于 P 乘以这个因子必须等于 $\frac{11}{4}mg$，所以当这个因子最大时 P 最小：

$$P_{\min} = \tfrac{11}{4}mg \cdot \tfrac{\sqrt{3}}{2} = \tfrac{11}{8}\sqrt{3}\,mg,$$

这正是所要求的结果。我们必须检查 B 仍然留在斜面上：

$$R_B = \sqrt{3}\,mg - \tfrac{11}{8}\sqrt{3}\,mg \cdot \tfrac{1}{2} = \tfrac{5}{16}\sqrt{3}\,mg > 0.$$

所以 P 作用于通过 ℓ 的铅垂平面内，与 ℓ 成 30°，并偏向离开斜面的一方。换言之，它**与水平成** 60° **向上**，把质点沿斜面向上拉。（留意与斜面所成的最佳角度是 $\arctan \mu_B$，即 B 的摩擦角（angle of friction）。）

第 (ii) 部分

现在 P 只需阻止系统向下滑，所以每个质点所受的摩擦力都沿斜面向上。若没有 P，A 不能独自保持静止，因为它的重量分量 $\frac{1}{2}mg$ 大于最大摩擦力 $\frac{1}{4}mg$。所以绳子必须拉住 A：

$$T \geq \tfrac{1}{2}mg - \tfrac{1}{4}mg = \tfrac{1}{4}mg.$$

特别地，绳子必须拉紧。要令 B 不向下滑，沿斜面向上的力必须足以平衡 T 和重量的分量：

$$P\cos\theta + \left(mg - \tfrac{1}{\sqrt{3}}P\sin\theta\right) \geq T + mg.$$

左边与 T 无关，所以最小的 P 对应最小的张力 $T = \frac{1}{4}mg$。于是

$$P\left(\cos\theta - \tfrac{1}{\sqrt{3}}\sin\theta\right) \geq \tfrac{1}{4}mg.$$

而 $\cos\theta - \frac{1}{\sqrt{3}}\sin\theta = \frac{2}{\sqrt{3}}\cos(\theta + 30^\circ) \leq \frac{2}{\sqrt{3}}$，等号在 $\theta = -30^\circ$ 时成立。这一次，最佳方向是把 B 压向斜面，这样会增大 R_B，从而增大有助稳住 B 的摩擦力。因此力的最小量值为

$$P_{\min} = \tfrac{1}{4}mg \cdot \tfrac{\sqrt{3}}{2} = \tfrac{1}{8}\sqrt{3}\,mg.$$

它的方向在 ℓ 向上方向之下 30°。由于 ℓ 以 30° 向上倾斜，这个方向正是水平方向：即沿水平方向把 B 推向斜面。取这个值时，A 和 B 都处于极限平衡（limiting equilibrium），而 $T = \frac{1}{4}mg$。

验算

- 我们取 $m = g = 1$ 作数值检验。在铅垂平面内扫描 P 的每一个方向，并以二分法（bisection）求出每个方向所需的最小 P。在 (i) 中，最小值为 $2.3816 \approx \frac{11}{8}\sqrt{3}$，出现于与斜面成 30.0° 之处。在 (ii) 中，最小值为 $0.2165 \approx \frac{1}{8}\sqrt{3}$，出现于 −30.0°，即水平方向。我们亦确认了在 (ii) 的最小值处确实存在完整的平衡状态。

- 就 (i) 而言，当 P 沿斜面方向（$\theta = 0$）时，把各项加起来：所需的 $P = \frac{1}{2}mg + \frac{1}{4}mg + mg + mg = \frac{11}{4}mg \approx 2.75mg$。这比 $2.38mg$ 大，正如所料。
- 就 (ii) 而言，当 P 沿斜面方向时，系统需要 $\frac{3}{2}mg - \frac{1}{4}mg - mg = \frac{1}{4}mg$。水平推力的效果更佳，只需 $\frac{\sqrt{3}}{8}mg \approx 0.217mg$。

常见错误

- 假设 P 沿斜面方向作用。题目问的是最小的量值，所以方向可由你自行选择。
- 忘记 $P\sin\theta$ 会改变 R_B，因而改变 B 所受的极限摩擦力。
- 在 (ii) 中沿用 (i) 的摩擦力方向。当要防止的是向下滑动时，摩擦力沿斜面向上。

STEP 之外

两部分都建基于同一原理：对抗摩擦力时最有效率的力，与运动方向所成的角正是摩擦角 $\lambda = \arctan\mu$。在 (i) 中它偏离斜面，在 (ii) 中则偏向斜面。原因是法向反作用力与极限摩擦力可合成为单一个“总反作用力”（total reaction），它与法线成角 λ。以这个总反作用力来思考，许多摩擦力问题都可化为力三角形（triangle of forces）的几何问题。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门）以大学一年级的程度，仔细地处理了摩擦力和受约束运动。

详细解答

STEP II 2005，第 10 题

力学 • 相撞的抛体

STEP II 2005，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/05-s2-q10

关键思路

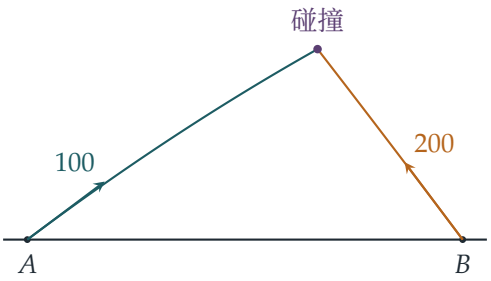
两枚导弹都是抛体（projectile），所以我们可以写出每一枚在任何时刻的位置。碰撞（collision）的意思是两者在同一瞬间位置相同，由此得到两条方程（水平和垂直方向各一），含两个未知数：碰撞时间和 T。水平方向的方程是线性的，可以用来消去碰撞时间，垂直方向的方程随之变成 T 的二次方程（quadratic）。之后我们要判断哪一个根在物理上合理。

解答

建立模型

我们把两枚导弹都视为质点（particle），只受重力作用而自由运动，不计空气阻力。取 A 为原点，x 轴沿 AB 方向（所以 B 位于 $x = 180$），y 轴垂直向上。时间 t 以秒为单位，由第一枚导弹发射的一刻起计。

对第一枚导弹，由 $\sin\alpha = \frac{3}{5}$ 得 $\cos\alpha = \frac{4}{5}$，所以它的初速度分量是 $100 \cdot \frac{4}{5} = 80$ 和 $100 \cdot \frac{3}{5} = 60$。对第二枚导弹，由 $\sin\beta = \frac{4}{5}$ 得 $\cos\beta = \frac{3}{5}$。它朝着 A 飞行，所以它的分量是 $-200 \cdot \frac{3}{5} = -120$ 和 $200 \cdot \frac{4}{5} = 160$。



设 $s = t - T$ 为第二枚导弹发射后所经过的时间。那么，当两者都在飞行时，

$$\begin{aligned} &\text{第一枚导弹：} \quad x = 80t, && y = 60t - 5t^2, \\ &\text{第二枚导弹：} \quad x = 180 - 120s, && y = 160s - 5s^2. \end{aligned}$$

碰撞

碰撞时两者的 x 坐标相同：

$$80t = 180 - 120(t - T) \quad \Longrightarrow \quad 200t = 180 + 120T \quad \Longrightarrow \quad t = 0.9 + 0.6T,$$

因此 $s = t - T = 0.9 - 0.4T$。

y 坐标亦相同：$60t - 5t^2 = 160s - 5s^2$，即

$$60t - 160s = 5(t^2 - s^2) = 5(t - s)(t + s) = 5T(t + s),$$

这里用了 $t - s = T$。代入 t 和 s 的表达式：

$$\begin{aligned} 60t - 160s &= 60(0.9 + 0.6T) - 160(0.9 - 0.4T) = 100T - 90, \\ 5T(t + s) &= 5T(1.8 + 0.2T) = 9T + T^2. \end{aligned}$$

令两式相等，

$$T^2 - 91T + 90 = 0 \quad \Longrightarrow \quad (T - 1)(T - 90) = 0,$$

所以 $T = 1$ 或 $T = 90$。

想一想 两个根都满足方程。为甚么 $T = 90$ 不可能是答案?

当 $T = 90$ 时，我们得到 $s = 0.9 - 36 = -35.1$，即是说这次“碰撞”会在第二枚导弹发射之前 35.1 秒发生。（碰撞点亦远低于地面：这些方程所描述的，是延伸至实际飞行范围以外的抛物线。）所以这个根没有物理意义。况且第一枚导弹只在空中停留 $2 \cdot 60/10 = 12$ 秒，早就已经落地了。

当 $T = 1$ 时，我们得到 $t = 1.5$ 和 $s = 0.5 > 0$，所以两枚导弹都在飞行中。高度分别是 $60(1.5) - 5(1.5)^2 = 78.75$ 和 $160(0.5) - 5(0.5)^2 = 78.75$，两者皆为正数，而且两枚导弹

都位于 $x = 120$。因此

$$T = 1 \text{ 秒}.$$

验算

把 $T = 1$ 代回去：在 $t = 1.5\,\mathrm{s}$ 时，两枚导弹都位于 $(120, 78.75)$ 米，这一点我们已用精确分数核实。以发射延迟 $1\,\mathrm{s}$ 用数值方法逐步计算两者的运动，两枚导弹之间的距离在 $t = 1.5\,\mathrm{s}$ 时降至几乎为零。再作一个合理性检查：碰撞发生在 $x = 120$，介乎 A 与 B 之间，而第一枚导弹仍在上升，因为它在该处的垂直速度是 $60 - 15 = 45\,\mathrm{m\,s^{-1}}$。

常见错误

- 两枚导弹用了同一个时间变量。第二枚导弹的飞行时间是 $t - T$，而不是 t。
- 忘记第二枚导弹是沿 x 轴的负方向运动。
- 保留 $T = 90$，而没有检查是否 $t - T > 0$。

STEP 之外

这个问题有一个更巧妙的看法。两枚导弹的加速度同为 g，所以一旦两者都在飞行，它们的相对速度（relative velocity）便是常数，各自看到对方沿直线运动。碰撞发生的充要条件，是这个相对速度的方向正好沿着时刻 T 时连接两者的直线，把这个条件写出来便得到同一条二次方程。像这样转换到自由下落的参考系（frame），正是通往爱因斯坦（Einstein）等效原理（equivalence principle）的第一步。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对相对运动和非惯性参考系（non-inertial frame）有严谨的论述。

详细解答

STEP II 2005，第 11 题

力学 · 斜面上的摩擦力

STEP II 2005，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：

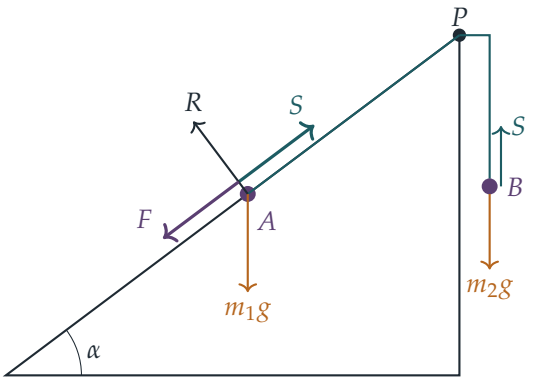
step.maths.org/questions/05-s2-q11

关键思路

由于 $\tan\alpha = \frac{3}{4}$，我们得到一个 3–4–5 三角形，所以 $\sin\alpha = \frac{3}{5}$ 及 $\cos\alpha = \frac{4}{5}$，而作用于 A 的每一个力都成为 $m_1 g$ 的简洁倍数。A 的运动于是分成三个阶段，每个阶段的加速度（acceleration）都是常数：被绳子拉上斜面、自由地向上滑行，以及向下滑回。摩擦力（friction）的方向在最后两个阶段之间改变，而这正是本题暗藏玄机的地方。

解答

我们把 A 和 B 视为质点（particle），绳子视为轻而不可伸长（light and inextensible），滑轮（pulley）视为光滑，因此两边的张力（tension）S 相同。设 $\alpha = \arctan\frac{3}{4}$，使得 $\sin\alpha = \frac{3}{5}$ 及 $\cos\alpha = \frac{4}{5}$。



绳子拉紧时的加速度

在垂直于斜面的方向上，A 没有加速度，所以

$$R = m_1 g\cos\alpha = \tfrac{4}{5}m_1 g, \qquad \text{而极限摩擦力（limiting friction）为} \qquad \mu R = \tfrac{1}{2}\cdot\tfrac{4}{5}m_1 g = \tfrac{2}{5}m_1 g.$$

首先我们检查系统确实会运动。假如 B 下降，A 便会沿斜面向上移动，阻碍它的是其重量的分量 $\frac{3}{5}m_1 g$ 以及最多为 $\frac{2}{5}m_1 g$ 的摩擦力：合共最多为 $m_1 g$。向另一方拉动的是 B 的重量 $m_2 g$，而 $m_2 > m_1$，所以摩擦力无法令系统保持静止：B 下降，A 沿斜面向上移动，而作用于 A 的摩擦力达到极限值，方向沿斜面向下。

设 a 为两者共同的加速度。对 A（沿斜面向上）和对 B（向下）应用牛顿第二定律（Newton's second law），得

$$\begin{aligned} m_1 a &= S - \tfrac{3}{5}m_1 g - \tfrac{2}{5}m_1 g = S - m_1 g, \\ m_2 a &= m_2 g - S. \end{aligned}$$

两式相加，$(m_1 + m_2)a = (m_2 - m_1)g$，所以

$$a = \frac{m_2 - m_1}{m_2 + m_1}\, g,$$

正如所求。为方便起见，我们写成 $a = kg$，其中 $k = \dfrac{m_2 - m_1}{m_2 + m_1}$ 且 $0 < k < 1$。

到达最高点的时间

在时间 T，A 正以速率 $u = aT = kgT$ 沿斜面向上移动，并已走了 $\frac{1}{2}kgT^2$ 的距离。

绳子断开时，S 消失。A 仍在沿斜面向上移动，所以摩擦力仍然沿斜面向下，而沿斜面向下的力合共为 $\frac{3}{5}m_1 g + \frac{2}{5}m_1 g = m_1 g$。因此 A 的减速度恰好是 g，并在再过一段时间

$$\frac{u}{g} = kT.$$

之后静止下来。故此 A 在释放后的时间

$$T + kT = (1+k)T = \frac{2m_2}{m_1 + m_2}\, T$$

到达最高点。（题目告诉我们 A 永远不会到达滑轮，所以这个阶段不会被提早中断。）在这个阶段中，A 再走了 $\dfrac{u^2}{2g} = \frac{1}{2}k^2 gT^2$，所以它由释放点起沿斜面向上所走的总距离为

$$d = \tfrac{1}{2}kgT^2 + \tfrac{1}{2}k^2 gT^2 = \tfrac{1}{2}k(1+k)gT^2.$$

? **想一想** A 停下来之后，会留在原处吗?

只有当摩擦力足以支撑它时，它才会留在原处。重量沿斜面向下的分量是 $\frac{3}{5}m_1g$，而摩擦力最多只能提供 $\frac{2}{5}m_1g$。换句话说，$\tan\alpha = \frac{3}{4} > \frac{1}{2} = \mu$。所以 A 会向下滑回，而这时摩擦力沿斜面向上。沿斜面向下的净力为 $\frac{3}{5}m_1g - \frac{2}{5}m_1g = \frac{1}{5}m_1g$，所以 A 以 $\frac{1}{5}g$ 的加速度沿斜面向下加速，比它上行时的减速度缓和得多。

质量之比

A 由最高点从静止出发，需时 t_d 滑回距离 d 而返抵释放点，其中

$$\tfrac{1}{2}\cdot\tfrac{1}{5}g\,t_d^2 = d = \tfrac{1}{2}k(1+k)gT^2, \qquad \text{所以} \qquad t_d^2 = 5k(1+k)T^2.$$

题目告诉我们 t_d 等于到达最高点的时间 $(1+k)T$。两边平方，这表示

$$(1+k)^2T^2 = 5k(1+k)T^2.$$

两边除以正数 $(1+k)T^2$，得 $1+k=5k$，所以 $k=\frac{1}{4}$。留意 T 已经消去：这就是为何无论绳子何时断开，这个等式都成立，而它亦唯一地确定了 k 的值。于是

$$\frac{m_2-m_1}{m_2+m_1} = \frac{1}{4} \iff 4m_2-4m_1 = m_2+m_1 \iff 3m_2 = 5m_1,$$

所以

$$m_1 : m_2 = 3 : 5.$$

验算

- 极限情况：若 $m_2 \to m_1$，则 $k\to 0$，而到达最高点的时间趋向 T（几乎没有速率可以失去）；若 $m_2 \gg m_1$，则 $k\to 1$，而 A 靠惯性滑行的时间与它被拉动的时间一样长，这正如我们所料，因为它达到速率 gT 后以 g 减速。
- 取 $m_1=3$、$m_2=5$ 及 $T=1\,\mathrm{s}$，我们有 $k=\frac{1}{4}$；到达最高点的时间是 $1.25\,\mathrm{s}$，$d=\frac{5}{32}g$，而下滑需时 $\sqrt{2d/(g/5)} = \sqrt{25/16} = 1.25\,\mathrm{s}$。对全部三个阶段逐步进行数值模拟，在多个 T 值之下，两个时间都得出 $1.250\,\mathrm{s}$。

常见错误

- 忘记了 A 开始滑回时摩擦力会反向，因而以 g（或 $\frac{3}{5}g$）而非 $\frac{1}{5}g$ 作为下滑的加速度。
- 只由绳子断开的位置起计算下滑的距离，而不是由释放点起计算。
- 没有检查 A 是否真的会滑回：这需要 $\tan\alpha > \mu$。

STEP 之外

摩擦力是取决于运动状态的力的经典例子：它的方向随速度而转换，而当物体静止时，它的大小有一个上限。这使得运动成为分段的，每个阶段各有不同的运动方程（equation of motion），而同一个想法亦支配着黏滑（stick–slip）现象，例如吱吱作响的门铰和地震断层。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）以大学一年级物理所要求的严谨程度，处理摩擦力和受约束的运动。

详细解答

STEP II 2005，第 12 题

概率与统计 • 贝叶斯定理

STEP II 2005，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/05-s2-q12

关键思路

这是以公平硬币作为先验（prior）的贝叶斯定理（Bayes' theorem）。由于硬币是公平的，在已知回答的条件下，掷出正面的概率为 $\frac{1}{2}$，当且仅当无论硬币掷出正面还是反面，得到这些回答的可能性都相等。因此在每一部分，我们只需比较两个似然（likelihood）$\mathrm{P}(\text{回答} \mid H)$ 和 $\mathrm{P}(\text{回答} \mid T)$，并以哪一位孪生姊妹在线作为条件来计算它们。

解答

设 H 和 T 分别为硬币掷出正面和反面的事件，故 $\mathrm{P}(H) = \mathrm{P}(T) = \frac{1}{2}$。设 A 为 Anna 在线的事件，B 为 Bella 在线的事件，其中 $\mathrm{P}(A) = p$，$\mathrm{P}(B) = q = 1 - p$。谁在线与硬币毫无关系，所以 A 和 B 均与 H 独立（independent）。在第 (ii) 和第 (iii) 部分，我们假设硬币只掷一次，而两封电子邮件都由同一位孪生姊妹回复，即我们发信时在线的那一位。（另一种理解方式见文末的备注。）

记 $t = ap + bq$，即单独一个回答属实的整体概率。

对于任何一组回答纪录 E，由贝叶斯定理及 $\mathrm{P}(H) = \mathrm{P}(T)$ 可得

$$\mathrm{P}(H \mid E) = \frac{\frac{1}{2}\mathrm{P}(E \mid H)}{\frac{1}{2}\mathrm{P}(E \mid H) + \frac{1}{2}\mathrm{P}(E \mid T)} = \frac{\mathrm{P}(E \mid H)}{\mathrm{P}(E \mid H) + \mathrm{P}(E \mid T)}.$$

此式等于 $\frac{1}{2}$，当且仅当 $\mathrm{P}(E \mid H) = \mathrm{P}(E \mid T)$。单凭这一点观察，便已完成了大部分工作。

第 (i) 部分

设 Y 为回答是“是”的事件。若硬币掷出正面，“是”便是真话。以谁在线作为条件，

$$\mathrm{P}(Y \mid H) = pa + qb = t.$$

若硬币掷出反面，“是”便是谎话，故

$$\mathrm{P}(Y \mid T) = p(1-a) + q(1-b) = (p+q) - (ap+bq) = 1 - t.$$

因此

$$\mathrm{P}(H \mid Y) = \frac{t}{t + (1-t)} = t = ap + bq.$$

所以 $\mathrm{P}(H \mid Y) = \frac{1}{2}$ 当且仅当 $ap + bq = \frac{1}{2}$，亦即当且仅当 $2(ap+bq) = 1$，正如所求。

第 (ii) 部分

设 E 为先答“是”、后答“否”的事件。假设 Anna 在线。她的两个回答互相独立，而且无论硬币掷出哪一面，两个回答中都恰好有一个是真话：

$$\begin{aligned} \mathrm{P}(E \mid H \cap A) &= a(1-a) && \text{（如实答“是”，然后说谎），} \\ \mathrm{P}(E \mid T \cap A) &= (1-a)a && \text{（说谎，然后如实答“否”）．} \end{aligned}$$

两者相等。同理，$\mathrm{P}(E \mid H \cap B) = b(1-b) = \mathrm{P}(E \mid T \cap B)$。因此

$$\mathrm{P}(E \mid H) = pa(1-a) + qb(1-b) = \mathrm{P}(E \mid T),$$

再由上面的观察可知 $\mathrm{P}(H \mid E) = \frac{1}{2}$，不论 p、a 和 b 取何值。

> ? **想一想** 为甚么这里完全不需要任何条件，就恰好得到 $\frac{1}{2}$?

回答互相矛盾，意味着无论硬币掷出哪一面，这位孪生姊妹都是说了一次真话、说了一次谎话。这个组合在正面和反面的情况下出现的可能性相等，所以它不带有任何关于硬币的信息，我们得到的仍然是先验概率 $\frac{1}{2}$。

第 (iii) 部分

现在设 E 为两次都答“是”的事件。已知掷出正面时，两个回答都是真话；已知掷出反面时，两个回答都是谎话。以谁在线作为条件，

$$\begin{aligned} \mathrm{P}(E \mid H) &= pa^2 + qb^2, \\ \mathrm{P}(E \mid T) &= p(1-a)^2 + q(1-b)^2. \end{aligned}$$

两式相减，并利用 $(1-a)^2 = 1-2a+a^2$，

$$\begin{aligned}\mathrm{P}(E \mid H) - \mathrm{P}(E \mid T) &= p\big(a^2 - (1-a)^2\big) + q\big(b^2 - (1-b)^2\big)\\ &= p(2a-1) + q(2b-1)\\ &= 2(ap+bq) - (p+q)\\ &= 2(ap+bq) - 1.\end{aligned}$$

所以若 $2(ap+bq) = 1$，两个似然便相等，而 $\mathrm{P}(H \mid E) = \frac{1}{2}$，正如所求。

（事实上，这个计算所显示的不止于此：两个似然之差恰好是 $2(ap+bq)-1$，所以 $\mathrm{P}(H \mid E) = \frac{1}{2}$ 成立当且仅当 $2(ap+bq) = 1$，与第 (i) 部分的条件相同。）

备注：另一种理解方式

如果改为每封电子邮件都重新、独立地决定是哪一位孪生姊妹刚好在线，那么每个回答属实的概率都是 t，而且互相独立。这样，在 (ii) 中 $\mathrm{P}(E \mid H) = t(1-t) = \mathrm{P}(E \mid T)$，而在 (iii) 中 $\mathrm{P}(E \mid H) = t^2$ 及 $\mathrm{P}(E \mid T) = (1-t)^2$，两者在 $t = \frac{1}{2}$ 时相等。结论并无分别。

验算

- 若两位孪生姊妹永远说真话（$a = b = 1$），则 $t = 1 \neq \frac{1}{2}$，而在 (i) 中，一个“是”的回答令正面成为必然：$\mathrm{P}(H \mid Y) = t = 1$。这时 (ii) 中互相矛盾的回答不可能出现，这与 $a(1-a) = b(1-b) = 0$ 一致。
- 取 $p = 0.3$、$a = 0.8$ 及 $b = \frac{0.5-0.24}{0.7} \approx 0.371$（故 $2(ap+bq) = 1$），在同一位孪生姊妹回复的模型中模拟 2×10^5 轮，得出 $\mathrm{P}(H \mid \text{是、是}) \approx 0.496$。以分数对数千组随机的 (p, a, b) 作精确枚举，确认了全部三个结果在两种理解方式下均成立。
- 若 $p = \frac{1}{2}$ 及 $a = b = 0.9$，条件不成立，而两个“是”的回答给出 $\mathrm{P}(H \mid E) = \frac{81}{82}$，与 $\frac{1}{2}$ 相去甚远，可见 (iii) 中的条件确实重要。

常见错误

- 未加说明便写下 $\mathrm{P}(Y \mid T) = 1 - \mathrm{P}(Y \mid H)$。此式在这里碰巧成立，但那只是因为在反面的情况下答“是”属于说谎；作答时应说明这一点。
- 在 (iii) 中，把两个回答当作各自以概率 t 属实的独立回答，同时却又假设两个回答出自同一位孪生姊妹。在已知是哪一位孪生姊妹的条件下，两个回答是独立的，但在无条件的情况下并不独立，因为两者都取决于谁在线。

STEP 之外

这里贯穿始终的原理只有一个：后验（posterior）赔率（odds）等于先验赔率乘以似然比（likelihood ratio）；在两个假设下出现的可能性相等的证据，不会告诉你任何事情。同一位孪生姊妹回复的模型是一个简单的混合模型（*mixture*）：在已知某个隐变量（hidden variable, 即谁在线）的条件下, 各个回答互相独立, 但整体而言却是相依的, 这个想法正是隐变量模型和贝叶斯模型的核心。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门）正是以这种精神，对条件概率和贝叶斯法则作了十分出色的阐述。

详细解答

STEP II 2005，第 13 题

概率与统计 · 泊松分布与二项分布

STEP II 2005，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/05-s2-q13

关键思路

每一页抽样页要么错误少于两个，要么不是，而且各页互相独立，所以 Y 数算的是 n 次独立试验（independent trials）中成功的次数：它服从二项分布（binomial distribution）。其后的一切都在探讨 λ 很小时会发生甚么事，而所需的工具就是 $\mathrm{e}^{-\lambda}$（或 $\ln(1+\lambda)$）的麦克劳林级数（Maclaurin series），只保留第一个没有被抵消的项。

解答

Y 的分布

设 X 为某一页上的错误数目，则 $X \sim \mathrm{Po}(\lambda)$。若 $X \le 1$，便称该页为良好页。于是

$$\mathrm{P}(\text{良好}) = \mathrm{P}(X=0) + \mathrm{P}(X=1) = \mathrm{e}^{-\lambda} + \lambda\mathrm{e}^{-\lambda} = (1+\lambda)\mathrm{e}^{-\lambda} = p.$$

样本由 n 页不同的页组成。不同页上的错误数目互相独立，而且全都服从同一个 $\mathrm{Po}(\lambda)$ 分布，所以无论选中哪 n 页，“第 j 页是良好页”这些事件都互相独立，各自的概率均为 p。（条件 $n \ll N$ 只是让我们放心，有足够多的页可供抽样；独立性并不需要这个条件，它来自模型本身。）

所以 Y 是 n 次独立试验中成功的次数，每次成功的概率为 p。某一组指定的 k 页良好页和 $n-k$ 页不良页出现的概率是 $p^k q^{n-k}$，而选出哪 k 页为良好页共有 $\binom{n}{k}$ 种方法，所以

$$\mathrm{P}(Y=k) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, \ldots, n,$$

其中 $p = (1+\lambda)\mathrm{e}^{-\lambda}$ 及 $q = 1-p$，正如题目所求。换言之，$Y \sim \mathrm{B}(n, p)$。

第 (i) 部分

展开 $\mathrm{e}^{-\lambda} = 1 - \lambda + \frac{1}{2}\lambda^2 - \frac{1}{6}\lambda^3 + \cdots$，再乘以 $1+\lambda$：

$$\begin{aligned} p = (1+\lambda)\mathrm{e}^{-\lambda} &= \left(1 - \lambda + \tfrac{1}{2}\lambda^2 - \tfrac{1}{6}\lambda^3 + \cdots\right) + \left(\lambda - \lambda^2 + \tfrac{1}{2}\lambda^3 - \cdots\right) \\ &= 1 - \tfrac{1}{2}\lambda^2 + \tfrac{1}{3}\lambda^3 - \cdots . \end{aligned}$$

λ 的一次项互相抵消，所以当 λ 很小时，

$$q = 1 - p = \tfrac{1}{2}\lambda^2 - \tfrac{1}{3}\lambda^3 + \cdots \approx \tfrac{1}{2}\lambda^2$$

因为被略去的各项比 $\frac{1}{2}\lambda^2$ 还要小，相差一个 λ 阶的因子。

> **想一想** 为甚么不是 $q \approx \lambda$，像 $\mathrm{e}^{-\lambda}$ 的首项所暗示的那样?

因为来自 $\mathrm{e}^{-\lambda}$ 的 $-\lambda$ 刚好被乘以 $1+\lambda$ 所产生的 $+\lambda$ 完全抵消。一页只有在至少有两个错误时才是不良页，而 $\mathrm{P}(X \geq 2) \approx \mathrm{P}(X = 2) = \frac{1}{2}\lambda^2\mathrm{e}^{-\lambda} \approx \frac{1}{2}\lambda^2$。这正好从另一个角度独立地印证了 (i)。

第 (ii) 部分

$\mathrm{P}(Y = n) = p^n$，而我们要找满足 $p^n \geq 1 - \lambda$ 的最大 n。两边都是正数（对于 $0 < \lambda < 1$），而 ln 是递增函数，所以这等价于

$$n \ln p \geq \ln(1 - \lambda).$$

现在 $\ln p = \ln(1+\lambda) - \lambda = \left(\lambda - \frac{1}{2}\lambda^2 + \cdots\right) - \lambda \approx -\frac{1}{2}\lambda^2$，而 $\ln(1-\lambda) \approx -\lambda$。由于 $\ln p < 0$，除以它时不等式（inequality）的方向会倒转：

$$n \leq \frac{\ln(1-\lambda)}{\ln p} \approx \frac{-\lambda}{-\frac{1}{2}\lambda^2} = \frac{2}{\lambda}.$$

所以满足条件的最大 n 约为 $2/\lambda$。

等价地，利用 (i)：$p^n = (1-q)^n \approx 1 - nq \approx 1 - \frac{1}{2}n\lambda^2$，而 $1 - \frac{1}{2}n\lambda^2 \geq 1 - \lambda$ 成立，当且仅当 $n \leq 2/\lambda$。$(1-q)^n \approx 1 - nq$ 这一步需要 nq 很小；这里 $nq \approx \frac{2}{\lambda} \cdot \frac{1}{2}\lambda^2 = \lambda$，确实很小，所以这个近似是自洽的。

第 (iii) 部分

由于 $Y > 1$ 蕴含 $Y > 0$，

$$\mathrm{P}(Y > 1 \mid Y > 0) = \frac{\mathrm{P}(Y > 1)}{\mathrm{P}(Y > 0)} = \frac{\mathrm{P}(Y > 0) - \mathrm{P}(Y = 1)}{\mathrm{P}(Y > 0)} = 1 - \frac{npq^{n-1}}{1 - q^n}.$$

（这里 $n \geq 2$ 是有作用的：它令 $Y > 1$ 成为可能发生的事件，并令幂 q^{n-1} 的次数至少与 q 相同。）现在固定 n，并设 λ 很小。那么 $q \approx \frac{1}{2}\lambda^2$ 很小，所以 $p \approx 1$ 及 $1 - q^n \approx 1$，由此得

$$\mathrm{P}(Y > 1 \mid Y > 0) \approx 1 - nq^{n-1} \approx 1 - n\left(\frac{\lambda^2}{2}\right)^{n-1},$$

最后一步用了 (i)。这就是所求的结果。

验算

- 以 $\lambda = 0.8$ 模拟 200,000 个样本，每个样本有 $n = 5$ 页泊松（Poisson）页，所得 $Y = 0, \ldots, 5$ 的相对频数与 $\mathrm{B}(5, p)$ 相符，误差在 0.002 以内。
- 当 $\lambda = 0.01$ 时，$q = 4.967 \times 10^{-5}$，而 $\frac{1}{2}\lambda^2 = 5 \times 10^{-5}$。
- 当 $\lambda = 0.01$ 时，满足 $p^n \geq 0.99$ 的最大 n 的精确值是 202，而 $2/\lambda = 200$。（在每个级数中多保留一项，会得到 $2/\lambda + \frac{7}{3}$，所以当 $2/\lambda$ 增大时，误差仍然有界。）
- 当 $\lambda = 0.01$，$n = 3$ 时：精确值为 $1 - \mathrm{P}(Y > 1 \mid Y > 0) = 7.40 \times 10^{-9}$，而 $3(\lambda^2/2)^2 = 7.5 \times 10^{-9}$。

常见错误

- 把“少于两个错误”理解为只有 $X = 0$，或理解为 $X \leq 2$。
- 把 $(1 + \lambda)\mathrm{e}^{-\lambda}$ 的展开式停在一次项，然后断定 $q \approx 0$：第一个没有被抵消的项其实是二次项。
- 在 (ii) 中除以 $\ln p < 0$ 时，忘记把不等式的方向倒转。

STEP 之外

第 (ii) 部分是渐近分析（asymptotic analysis）的一个例子：条件由乘积 $n\lambda^2$ 所控制，所以 n 的自然尺度是 $1/\lambda$，而像 p^n 这样的量，其表现就如 $\mathrm{e}^{-n\lambda^2/2}$。二项分布的泊松近似，以及概率论中许多“稀有事件”的极限，背后都是同一套推理。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率论入门教材）对这些极限定理有严谨细致的阐述。

详细解答

STEP II 2005，第 14 题

概率与统计 · 混合分布

STEP II 2005，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/05-s2-q14

关键思路

这个密度是两个我们早已熟悉的密度的加权混合：一个是标准正态分布（standard normal distribution），另一个是 $[0,\lambda]$ 上的均匀分布（uniform distribution）。因此，f 的每个积分都可拆成正态部分和均匀部分，而每一部分都可以直接由标准事实（总概率为 1、均值、$\mathrm{E}(X^2)$）读出，完全不必做任何艰深的积分。方差（variance）则由 $\mathrm{E}(X^2)-\mu^2$ 求得。

解答

常数 k、均值与方差

由于 ϕ 和 g 都是概率密度函数（probability density function），两者在整条实数线上的积分都是 1。所以

$$1=\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = k\,[1+\lambda\cdot 1] \quad\Longrightarrow\quad k=\frac{1}{1+\lambda}.$$

至于均值（mean），我们利用两个成分各自的均值：标准正态分布的均值为 0，而 $[0,\lambda]$ 上的均匀密度 g 的均值为 $\frac{1}{2}\lambda$。因此

$$\mu=\int_{-\infty}^{\infty} xf(x)\,\mathrm{d}x = k\left[\int x\phi(x)\,\mathrm{d}x+\lambda\int xg(x)\,\mathrm{d}x\right]=\frac{1}{1+\lambda}\left[0+\lambda\cdot\frac{\lambda}{2}\right]=\frac{\lambda^2}{2(1+\lambda)}.$$

同样地，标准正态分布满足 $\int x^2\phi(x)\,\mathrm{d}x=1$（由于其均值为 0，这就是它的方差），而

$$\int x^2 g(x)\,\mathrm{d}x=\frac{1}{\lambda}\int_0^{\lambda} x^2\,\mathrm{d}x=\frac{\lambda^2}{3}.$$

所以

$$\mathrm{E}(X^2)=\frac{1}{1+\lambda}\left[1+\lambda\cdot\frac{\lambda^2}{3}\right]=\frac{3+\lambda^3}{3(1+\lambda)}.$$

想一想 为甚么 $\mathrm{E}(X^2)$ 可以这样拆开，方差却不可以?

期望值（expectation）是积分，而积分是线性的，所以 $\mathrm{E}(X^2)$ 是各成分二阶矩（second moment）的加权组合，权重与 f 由各成分密度组合而成时所用的完全相同。方差并非线性（它涉及 μ^2），所以我们必须先把二阶矩组合起来，然后才减去 μ^2。这样做便得

$$\begin{aligned}\sigma^2 = \mathrm{E}(X^2) - \mu^2 &= \frac{3+\lambda^3}{3(1+\lambda)} - \frac{\lambda^4}{4(1+\lambda)^2} \\ &= \frac{4(3+\lambda^3)(1+\lambda) - 3\lambda^4}{12(1+\lambda)^2} \\ &= \frac{12+12\lambda+4\lambda^3+4\lambda^4-3\lambda^4}{12(1+\lambda)^2} = \frac{\lambda^4+4\lambda^3+12\lambda+12}{12(1+\lambda)^2},\end{aligned}$$

这正是所要求的结果。

$\lambda = 2$ 的情况

此时 $k = \frac{1}{3}$，而在 $[0,2]$ 上 $\lambda g(x) = 1$，所以

$$f(x) = \begin{cases} \frac{1}{3}\phi(x) & x < 0 \text{ 或} x > 2, \\ \frac{1}{3}\big(\phi(x)+1\big) & 0 \le x \le 2. \end{cases}$$

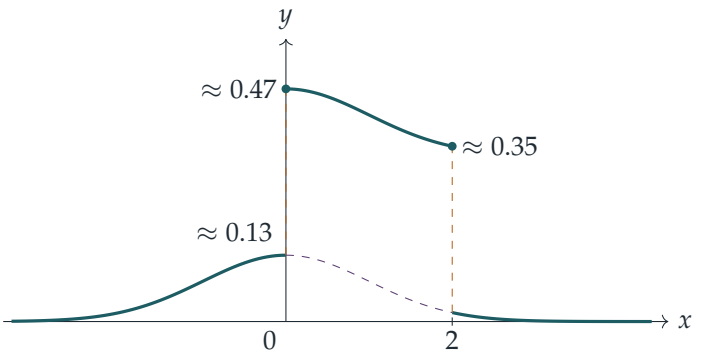
此外

$$\mu = \frac{4}{6} = \frac{2}{3}, \qquad \sigma^2 = \frac{16+32+24+12}{12\cdot 9} = \frac{84}{108} = \frac{7}{9}, \qquad \sigma = \frac{\sqrt{7}}{3}.$$

第 (i) 部分

这条曲线是标准正态钟形曲线的三分之一，并在区间 $[0,2]$ 上向上提升了 $\frac{1}{3}$。所以它在 $x = 0$ 处向上跳跃，在 $x = 2$ 处向下跳跃，跳跃的高度都是 $\frac{1}{3}$。利用 $\phi(0) = 1/\sqrt{2\pi} \approx 0.399$ 和 $\phi(2) \approx 0.054$：

- 在 0 的左边紧贴处，$f \approx 0.133$；在 0 的右边紧贴处，$f \approx 0.466$（最大值）；
- 在 2 的左边紧贴处，$f \approx 0.351$；在 2 的右边紧贴处，$f \approx 0.018$；
- 在 $[0,2]$ 上曲线递减，形状与正态曲线一致；
- 当 $x \to \pm\infty$ 时 $f \to 0$，而且处处 $f > 0$。



$[0,2]$ 上的紫红色虚线是 $\frac{1}{3}\phi(x)$，画出来以作比较；实线则是 $y=f(x)$。（纵轴的比例已被拉长。）

第 (ii) 部分

设 $F(x)=\mathrm{P}(X\le x)=\int_{-\infty}^{x} f(t)\,\mathrm{d}t$。对 $\frac{1}{3}\phi$ 积分得 $\frac{1}{3}\Phi$，而 $[0,2]$ 上额外的 $\frac{1}{3}$ 所贡献的，是 $\frac{1}{3}$ 乘以 $[0,2]$ 中位于 x 左边那一段的长度。所以

$$F(x)=\begin{cases}\frac{1}{3}\Phi(x) & x<0,\\ \frac{1}{3}\big(\Phi(x)+x\big) & 0\le x\le 2,\\ \frac{1}{3}\big(\Phi(x)+2\big) & x>2.\end{cases}$$

各段在 $x=0$ 和 $x=2$ 处互相吻合，这是必然的（没有点质量（point mass）时，累积分布函数（CDF）是连续的），而且当 $x\to\infty$ 时 $F(x)\to\frac{1}{3}(1+2)=1$。

第 (iii) 部分

这里 $\mu+2\sigma=\frac{2}{3}+\frac{2}{3}\sqrt{7}\approx 2.43$，大于 2，所以在该处要用 F 的最后一个分支。由于 X 是连续的，

$$\begin{aligned}\mathrm{P}(0<X<\mu+2\sigma) &= F\big(\tfrac{2}{3}+\tfrac{2}{3}\sqrt{7}\big)-F(0)\\ &= \tfrac{1}{3}\big(\Phi(\tfrac{2}{3}+\tfrac{2}{3}\sqrt{7})+2\big)-\tfrac{1}{3}\Phi(0)\\ &= \tfrac{1}{3}\big(0.9921+2-0.5\big)=\frac{2.4921}{3}=0.8307.\end{aligned}$$

验算

- 当 $\lambda\to 0$ 时，均匀部分消失：公式给出 $\mu\to 0$ 和 $\sigma^2\to\frac{12}{12}=1$，正是标准正态分布的数值，合乎预期。

- 对于 $\lambda = 2$，全方差定律（law of total variance，见下文）以另一种方法给出相同的答案。各成分的权重为 $\frac{1}{3}, \frac{2}{3}$，方差为 $1, \frac{1}{3}$，均值为 $0, 1$。方差的平均值是 $\frac{1}{3} + \frac{2}{9} = \frac{5}{9}$，均值的方差是 $\frac{2}{3} - \frac{4}{9} = \frac{2}{9}$，总和是 $\frac{7}{9}$。
- 以数值方法对 f 积分（中点法则，在 0 和 λ 处分段），取 $\lambda = 0.5, 1, 2, 3.7$，所得的总概率 1、均值和方差公式全部吻合，误差小于 10^{-6}；对于 $\lambda = 2$，结果是 $\mu = 0.6667$ 和 $\sigma^2 = 0.7778$。
- 直接计算 $\Phi(2.4305)$ 得 0.99246，而非题目所引用的 0.9921；概率则是 0.8308，而非 0.8307。这微小的差异源自题目所给的数值，而我们应该采用题目的数值。

常见错误

- 忘记了因子 $k = \frac{1}{3}$，以致总概率变成 3。
- 把方差算成 $k[1 + \lambda \cdot \frac{\lambda^2}{12}]$，即把各成分的方差组合起来。混合分布的方差不能线性地组合；二阶矩才可以。
- 把 f 画成连续曲线：它在 $x = 0$ 和 $x = \lambda$ 处有跳跃。

STEP 之外

形如 $\sum_i w_i f_i(x)$ 的密度，其中权重 $w_i \geq 0$ 且总和为 1，称为混合分布（mixture distribution）：以概率 w_i 选出成分 i，然后从 f_i 中抽样。这里的权重是 $\frac{1}{1+\lambda}$ 和 $\frac{\lambda}{1+\lambda}$。混合分布的方差其实就是全方差定律的化身：各成分方差的平均值，加上各成分均值的方差。混合分布是统计学和机器学习中聚类（clustering）方法的基础。Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门教材）对条件期望值（conditional expectation）和全方差定律有清晰的讲解。

详细解答

STEP III 2005，第 1 题

纯数学 · 三角方程与图像

STEP III 2005，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/05-s3-q1

关键思路

凡是形如 sin(某式) = cos(另一式) 的方程，都可以利用 $\cos B = \sin\left(\frac{\pi}{2} - B\right)$ 转化为关于两个自变量（argument）的命题。把这个做法用于 $\sin(\sin x) = \cos(\cos x)$，便得出 $\sin x \mp \cos x$ 必须等于 $\frac{\pi}{2}$ 的奇数倍，而 $|\sin x \pm \cos x|$ 的上界表明它根本太小，无法达到这样的值。至于草图，只要想清楚当内层自变量只在一个很短的区间内变动时，外层函数会对它产生甚么作用，便可画出。

解答

方程 $\sin A = \cos B$

由于 $\cos B = \sin\left(\frac{\pi}{2} - B\right)$，我们需要求 $\sin A = \sin C$ 的通解（general solution），其中 $C = \frac{\pi}{2} - B$。正弦函数在每个长度为 2π 的周期（period）内，对 $[-1, 1]$ 中的每个值都恰好取到两次，分别在 C 和 $\pi - C$（当该值为 ± 1 时，两者重合），所以

$$\sin A = \sin C \iff A = C + 2k\pi \text{ 或 } A = \pi - C + 2k\pi \quad \text{对某个整数} k.$$

代入 $C = \frac{\pi}{2} - B$：

$$\begin{aligned} A &= \tfrac{\pi}{2} - B + 2k\pi = (4k+1)\tfrac{\pi}{2} - B, \\ \text{或} \quad A &= \pi - \tfrac{\pi}{2} + B + 2k\pi = (4k+1)\tfrac{\pi}{2} + B. \end{aligned}$$

因此 $\sin A = \cos B$ 当且仅当存在整数 n 使得 $A = (4n+1)\frac{\pi}{2} \pm B$。两个方向都已涵盖，因为上面每一步都是“当且仅当”（if and only if）。（若要直接验证“若”的方向：$\sin\left((4n+1)\frac{\pi}{2} \pm B\right) = \sin\left(\frac{\pi}{2} \pm B\right) = \cos B$，这里用到周期为 2π。）

上界与方程 $\sin(\sin x) = \cos(\cos x)$

利用复角公式（compound angle formula），

$$\sin x \pm \cos x = \sqrt{2}\left(\sin x \cos \tfrac{\pi}{4} \pm \cos x \sin \tfrac{\pi}{4}\right) = \sqrt{2}\sin\left(x \pm \tfrac{\pi}{4}\right),$$

又因为对任何 θ 都有 $|\sin\theta| \le 1$，所以对所有 x 都有 $|\sin x \pm \cos x| \le \sqrt{2}$，这正是所要证明的。

现在假设对某个 x 有 $\sin(\sin x) = \cos(\cos x)$。在第一个结果中取 $A = \sin x$ 和 $B = \cos x$，可知存在整数 n 使得

$$\sin x = (4n+1)\tfrac{\pi}{2} \pm \cos x, \qquad \text{即} \qquad \sin x \mp \cos x = (4n+1)\tfrac{\pi}{2}.$$

想一想 为甚么右边不可能很小?

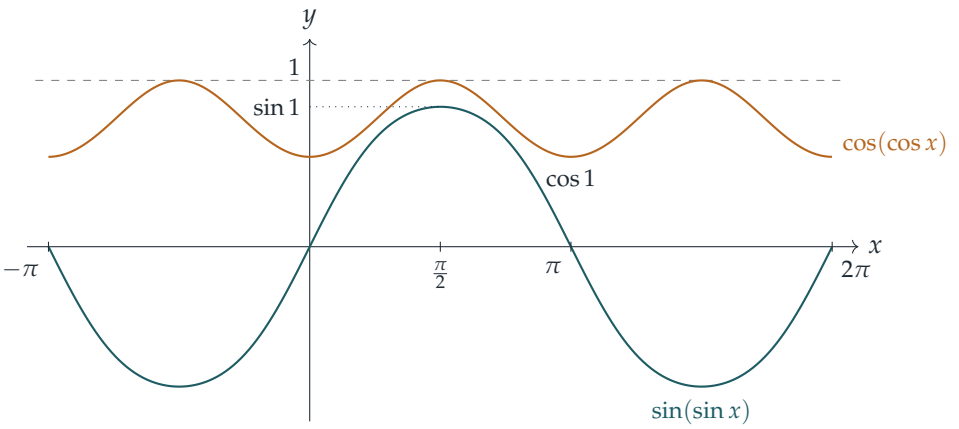
因为 $4n+1$ 是奇数，它永远不等于 0，所以 $|4n+1| \ge 1$，右边的绝对值（modulus）最少是 $\frac{\pi}{2} \approx 1.571$。而左边的绝对值最多是 $\sqrt{2} \approx 1.414$。由于 $\sqrt{2} < \frac{\pi}{2}$ （两边平方：$2 < \frac{\pi^2}{4}$，因为 $\pi^2 > 9 > 8$），这是不可能的。因此 $\sin(\sin x) = \cos(\cos x)$ 无解。

绘画 $y = \sin(\sin x)$ **和** $y = \cos(\cos x)$ **的草图**

$y = \sin(\sin x)$。内层函数 $\sin x$ 的取值范围是 $[-1,1]$，而在这个区间上 sin 是递增的（因为 $[-1,1] \subset [-\frac{\pi}{2}, \frac{\pi}{2}]$）。所以 $\sin(\sin x)$ 的升降与 $\sin x$ 完全一致：它的周期为 2π，是奇函数（odd function），在 $x = n\pi$ 处为零，在 $x = \frac{\pi}{2} + 2n\pi$ 处取极大值 $\sin 1 \approx 0.841$，在 $x = -\frac{\pi}{2} + 2n\pi$ 处取极小值 $-\sin 1$。它的斜率（gradient）$\cos x \cos(\sin x)$ 在原点等于 1。它看起来像一条稍为压扁了的正弦曲线。

$y = \cos(\cos x)$。由于 $\cos(-u) = \cos u$ 及 $\cos(x+\pi) = -\cos x$，我们有 $\cos(\cos(x+\pi)) = \cos(\cos x)$：这个图像的周期为 π，而且是偶函数（even function）。在 $[-1,1]$ 上，$\cos u$ 在 $u = 0$ 处最大（$= 1$），在 $u = \pm 1$ 处最小（$= \cos 1 \approx 0.540$）。所以 $\cos(\cos x)$ 在 $\cos x = 0$ 之处，即 $x = \frac{\pi}{2} + n\pi$，取极大值 1；在 $\cos x = \pm 1$ 之处，即 $x = n\pi$，取极小值 $\cos 1$。它永远介乎 0.54 与 1 之间。

根据上一部分，两条曲线永不相交；又因为 $\cos(\cos 0) = \cos 1 > 0 = \sin(\sin 0)$，由连续性（continuity）可知余弦曲线处处严格位于正弦曲线之上。两者在 $x \approx 0.88$ 附近最为接近（相距约 0.107），那里 $\sin(\sin x)$ 几乎到达顶峰，而 $\cos(\cos x)$ 仍在上升。



绘画 $y = \sin(2\sin x)$ 的草图

这个函数同样以 2π 为周期，而且是奇函数，零点在 $x = n\pi$，所以只需弄清楚 $0 \le x \le \pi$ 的情况；在这个区间上，图像关于 $x = \frac{\pi}{2}$ 对称（因为 $\sin(\pi - x) = \sin x$）。

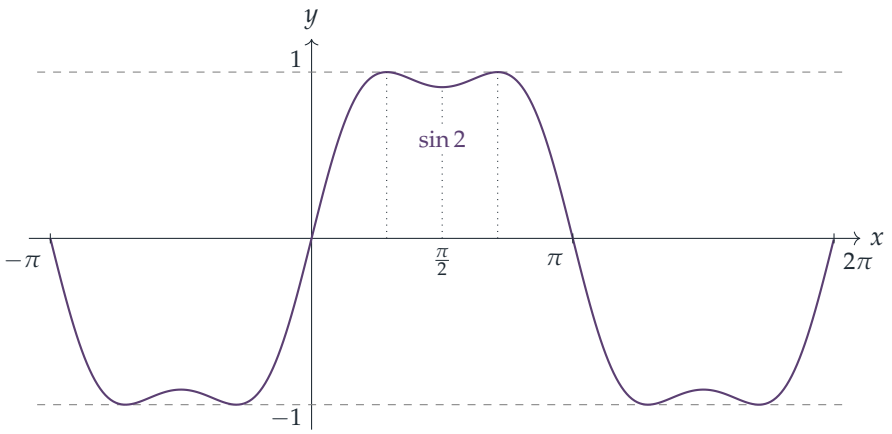
现在的分别在于，内层自变量 $u = 2\sin x$ 在这个区间上的取值范围是 $[0, 2]$，而 $2 > \frac{\pi}{2}$。所以当 x 由 0 增加到 $\frac{\pi}{2}$ 时：

- u 由 0 上升到 $\frac{\pi}{2}$，同时 $\sin x$ 上升到 $\frac{\pi}{4}$，此时 $x = \arcsin\frac{\pi}{4} \approx 0.903$；期间 $y = \sin u$ 由 0 上升到它的极大值 1；
- u 继续由 $\frac{\pi}{2}$ 增加到 2，在这一段 $\sin u$ 是递减的，所以 y 回落，在 $x = \frac{\pi}{2}$ 处降至 $\sin 2 \approx 0.909$。

根据对称性，y 在 $x = \pi - \arcsin\frac{\pi}{4} \approx 2.238$ 处再次升至 1，然后在 $x = \pi$ 处降至 0。用导数（derivative）验证，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 2\cos x \cos(2\sin x),$$

它等于零，恰好是当 $\cos x = 0$（在 $\frac{\pi}{2}$ 处的局部极小值）或 $2\sin x = \pm\frac{\pi}{2}$（值为 1 的极大值）之时；在原点的斜率为 2。在 $[\pi, 2\pi]$ 上，图像是上述情况的反射：两个值为 -1 的极小值，中间在 $x = \frac{3\pi}{2}$ 处有一个值为 $-\sin 2$ 的局部极大值。正弦曲线的每一个峰都变成了带有浅浅凹陷的驼峰。



验算

- 取 $B = 0.3$ 和 $n = 1$：$A = \frac{5\pi}{2} - 0.3$ 给出 $\sin A = \cos 0.3 \approx 0.9553$，与预期相符。
- 在一个周期内用细密的格点计算 $\cos(\cos x) - \sin(\sin x)$，得到最小值约为 0.107（在 $x \approx 0.878$ 附近），从不为零，这印证了两条曲线并不相交。
- 对于 $\sin(2\sin x)$，在 $[0, \pi]$ 上作数值搜索，恰好找到三个转向点（turning points）：在 $x \approx 0.903$ 和 2.238 处的极大值 1，以及在 $\frac{\pi}{2}$ 处的极小值 $\sin 2 \approx 0.909$。

常见错误

- 只写出 $A = \frac{\pi}{2} - B + 2n\pi$，而遗漏了第二组解 $A = \frac{\pi}{2} + B + 2n\pi$，这组解来自 $\sin(\pi - C) = \sin C$。
- 忘记了 $n = 0$ 也是容许的，只检查 $|4n + 1| \geq 5$ 的情况：最关键的比较，正是与 $\frac{\pi}{2}$ 本身的比较。
- 把 $\sin(2\sin x)$ 画成一条高度为 1 的普通正弦曲线：由于 $2 > \frac{\pi}{2}$，内层自变量会越过正弦函数的顶峰，因而在 $x = \frac{\pi}{2}$ 处出现凹陷。

STEP 之外

这里的草图全都关乎函数的复合（composition）：$f(g(x))$ 的形状取决于 g 的值域（range）覆盖了 f 的图像的哪一部分。把这个想法反复迭代（$\sin(\sin(\sin x))$，如此类推），便会引向不动点（fixed points）和迭代映射（iterated maps）的动力学，当中要问的是 $f(f(\cdots f(x)))$ 何时会稳定下来。至于那个不等式论证，即比较等式两边各自可能的取值范围，则是分析

学（analysis）中随处可见的一种技巧的简单例子；Abbott 的 *Understanding Analysis* 是了解如何把这类论证严格化的好去处。

详细解答

STEP III 2005，第 2 题

纯数学 · 可分离变量微分方程

STEP III 2005，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/05-s3-q2

关键思路

这条微分方程（differential equation）是可分离变量（separable）的，积分后得到的对数可以合并成题目所述的形式。至于距离的问题，应求距离的平方 $D = x^2 + y^2$ 的最小值，这样便可避开平方根。在曲线上，y^2 是 x 的显函数，故 D 变成只含 x 的函数，而其导数可以俐落地因式分解。会出现哪些驻点（stationary point），取决于 c 与 a^2 的大小关系。

解答

通解

分离变量（设 $y \neq 0$），

$$\int \frac{1}{y}\,\mathrm{d}y = -\int \frac{x}{x^2+a^2}\,\mathrm{d}x \quad \Longrightarrow \quad \ln|y| = -\tfrac{1}{2}\ln(x^2+a^2) + k,$$

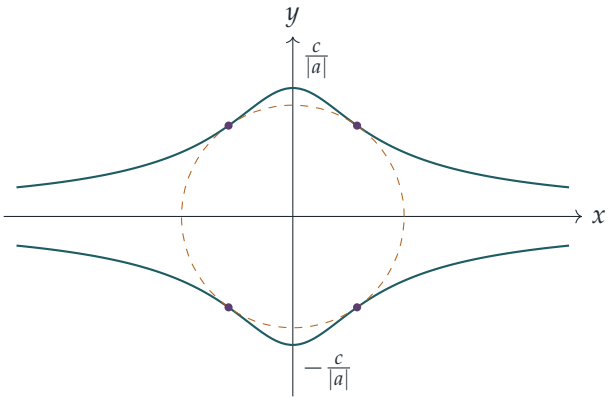
其中 k 为常数。注意 $x^2 + a^2 > 0$ 对任何 x 均成立，因为 $a \neq 0$，所以右边的对数处处有定义。两边乘以二再取指数，

$$y^2 = \frac{\mathrm{e}^{2k}}{x^2+a^2}, \qquad \text{故} \qquad y^2(x^2+a^2) = c^2,$$

其中 $c^2 = \mathrm{e}^{2k}$。解 $y = 0$ 在我们除以 y 时遗失了，它正是 $c = 0$ 的情况，所以 c 确实是任意常数。明确地写出来，通解（general solution）是 $y = c/\sqrt{x^2+a^2}$，其中 c 可正可负。由于曲线的方程中只出现 c^2，以下我们取 $c > 0$。

绘画曲线

曲线 $y^2(x^2+a^2)=c^2$ 由 $y=\pm c/\sqrt{x^2+a^2}$ 两支组成。它对两条坐标轴都对称（方程中只出现 x^2 和 y^2），而且永不与 x 轴相交。在上面的一支，y 在 x^2+a^2 最小时最大，于是在 $x=0$ 处取得极大值 $c/|a|$。当 $|x|\to\infty$ 时，我们有 $y\approx\pm c/|x|\to 0$，所以 x 轴是渐近线（asymptote）。下图所示为 $a=1$、$c=2$ 的情况，并画出了我们将在第 (ii) 部分求得的圆。



x^2+y^2 的导数

设 $D=x^2+y^2$。先利用微分方程，再代入 $y^2=c^2/(x^2+a^2)$，

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

此式在两支上都成立，因为它只涉及 y^2。用积法则（product rule）再求导一次，并利用 $\frac{\mathrm{d}}{\mathrm{d}x}(x^2+a^2)^{-2}=-4x(x^2+a^2)^{-3}$，

$$\frac{\mathrm{d}^2D}{\mathrm{d}x^2}=2\left(1-\frac{c^2}{(x^2+a^2)^2}\right)+2x\cdot\frac{4c^2x}{(x^2+a^2)^3}=2\left(1-\frac{c^2}{(x^2+a^2)^2}\right)+\frac{8c^2x^2}{(x^2+a^2)^3},$$

即为所求。

D 的驻点出现在 $x=0$ 或 $(x^2+a^2)^2=c^2$ 之处。由于 $x^2+a^2>0$ 及 $c>0$，第二个条件即 $x^2+a^2=c$，亦即 $x^2=c-a^2$。

第 (i) 部分

设 $0<c<a^2$。对任何 x，我们有 $x^2+a^2\geq a^2>c$，所以 $(x^2+a^2)^2>c^2$，而 $\mathrm{d}D/\mathrm{d}x$ 中的括号严格为正。因此 $\mathrm{d}D/\mathrm{d}x$ 的正负号与 x 相同：D 在 $x<0$ 时递减，在 $x>0$ 时递增。所以 D 在 $x=0$ 处取得全局最小值（global minimum），而且别无他处。（与此吻合的是，该处的二阶导数为 $2(1-c^2/a^4)>0$。）

在 $x=0$ 处，$y^2=c^2/a^2$，所以最接近原点的点是

$$\left(0,\ \pm\frac{c}{a}\right),$$

与原点的距离为 $c/|a|$。这就是所要求的结果。

第 (ii) 部分

现在设 $c>a^2$。这时有三个驻点：$x=0$ 和 $x=\pm\sqrt{c-a^2}$。

> **想一想** $x=0$ 仍然是最近的点吗?

不是。在 $x=0$ 处，二阶导数为 $2(1-c^2/a^4)$，它是负数，因为由 $c>a^2>0$ 可得 $c^2>a^4$。所以 $x=0$ 现在是 D 的局部极大值（local maximum）。在 $x^2=c-a^2$ 处，我们有 $x^2+a^2=c$，所以括号等于零，而

$$\frac{\mathrm{d}^2D}{\mathrm{d}x^2}=\frac{8c^2(c-a^2)}{c^3}=\frac{8(c-a^2)}{c}>0,$$

得到局部极小值（local minimum）。要知道它们是全局最小值，可以考察 $\mathrm{d}D/\mathrm{d}x$ 的正负号。当 $x>0$ 时，括号在 $x^2<c-a^2$ 时为负，在 $x^2>c-a^2$ 时为正，所以在 $x\geq 0$ 上，函数 D 一直递减至 $x=\sqrt{c-a^2}$，其后递增。因此 D 在 $x\geq 0$ 上的最小值出现在 $x=\sqrt{c-a^2}$，而根据对称性，它在 $x\leq 0$ 上的最小值出现在 $x=-\sqrt{c-a^2}$。

在这些点上，$y^2=c^2/(x^2+a^2)=c^2/c=c$。所以曲线上最接近原点的点是以下四点

$$\left(\pm\sqrt{c-a^2},\ \pm\sqrt{c}\right)$$

（正负号的全部四种组合），与原点的距离为 $\sqrt{2c-a^2}$。

作为覆核，在 $x=0$ 处距离的平方是 c^2/a^2，而

$$\frac{c^2}{a^2}-(2c-a^2)=\frac{(c-a^2)^2}{a^2}>0,$$

所以 y 轴上的点确实较远。

验算

- 把 $y=c/\sqrt{x^2+a^2}$ 代回微分方程，结果成立：$y'=-cx(x^2+a^2)^{-3/2}=-xy/(x^2+a^2)$。
- 取 $a=1$、$c=2$，以穷举方式搜索 $x^2+4/(x^2+1)$ 的值，可找到最小值 $3=2c-a^2$，出现在 $x=\pm1=\pm\sqrt{c-a^2}$，而 $y^2=2=c$。这就是图中的四个圆点，即半径为 $\sqrt{3}$ 的虚线圆与曲线相切之处。
- 当 $c=a^2$ 时，两个答案一致：$\sqrt{c-a^2}=0$ 且 $\sqrt{c}=c/|a|=|a|$，所以四个点合并成 $(0,\pm|a|)$。

常见错误

- 由“$\mathrm{d}D/\mathrm{d}x = 0$ 且 $\mathrm{d}^2D/\mathrm{d}x^2 > 0$”便断定某点的距离最短。这只能说明它是局部极小值；全局的问题要靠 $\mathrm{d}D/\mathrm{d}x$ 的正负号来解决。
- 忘记了下面的一支，以致在第 (ii) 部分只给出两个点而不是四个。
- 写成 $x^2 + a^2 = \pm c$：负号的情况不可能发生，因为 $x^2 + a^2 > 0$。

STEP 之外

在 $y^2(x^2 + a^2) = c^2$ 的条件下求 $x^2 + y^2$ 的最小值，是一个约束优化（constrained optimisation）问题。在大学里，你会学到拉格朗日乘数（Lagrange multipliers），它毋须先解出其中一个变量便能处理这类问题：在最近的点上，曲线的法线（normal）沿半径方向，这正是虚线圆与曲线相切的原因。当 c 越过 a^2 时，原本在 $x = 0$ 的单一最小值一分为二，这是叉式分岔（pitchfork bifurcation）的一个简单例子。想进一步了解可分离变量方程及其解曲线，可参阅 Boyce and DiPrima, *Elementary Differential Equations*（常微分方程入门教科书）。

详细解答

STEP III 2005，第 3 题

纯数学 · 二次函数的复合

STEP III 2005，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/05-s3-q3

关键思路

把两个首一（monic）二次多项式复合（compose），会得到一个首一四次多项式（quartic），而它的系数是互相牵连的：四个参数 p,q,r,s 要产生四个系数 a,b,c,d，但在其中两个系数里，p 和 s 只会以 $2s+p$ 这个组合出现，所以必定有一条关系式要成立。第二种形式其实是同一回事，只是换了个样子：对外层的二次式配方（completing the square），就把 $\mathrm{f}(\mathrm{g}(x))$ 变成（一个平方）$-$（一个常数）。正是这个形式，让我们只用二次公式（quadratic formula）就能解出最后的四次方程。

解答

把二次式复合

把 f 中的 x 换成 $\mathrm{g}(x)$，

$$\begin{aligned}\mathrm{f}(\mathrm{g}(x)) &= (x^2+rx+s)^2 + p(x^2+rx+s) + q \\ &= x^4 + 2rx^3 + (r^2+2s+p)x^2 + r(2s+p)x + (s^2+ps+q).\end{aligned}$$

所以 $x^4+ax^3+bx^2+cx+d = \mathrm{f}(\mathrm{g}(x))$ 对所有 x 成立，当且仅当

$$a = 2r, \qquad b = r^2+(2s+p), \qquad c = r(2s+p), \qquad d = s^2+ps+q.$$

必要性（necessity）。若这样的 p,q,r,s 存在，则 $r = \frac{1}{2}a$ 且 $2s+p = b-r^2 = b-\frac{1}{4}a^2$，所以

$$c = \tfrac{1}{2}a\left(b-\tfrac{1}{4}a^2\right), \qquad \text{即} \qquad a^3-4ab+8c = 0.$$

充分性（sufficiency）。反过来，假设 $a^3-4ab+8c=0$。取 $r=\frac{1}{2}a$，s 可取任意值（例

如 $s=0$），然后令 $p=b-\frac{1}{4}a^2-2s$ 及 $q=d-s^2-ps$。第一、第二和第四条等式按构造自然成立，而第三条就是 $c=\frac{1}{2}a(b-\frac{1}{4}a^2)$，正是我们的条件移项后的样子。留意 d 完全不受限制，因为 q 只在常数项中出现。

所以所求的充要条件（necessary and sufficient condition）是

$$\boxed{a^3-4ab+8c=0}$$

（d 可为任意值）。

$(x^2+vx+w)^2-k$ 这个形式

想一想 像 $\mathrm{g}^2+p\,\mathrm{g}+q$ 这样的和，怎样才能改写成一个平方减去一个常数?

把 $\mathrm{g}(x)$ 当作单一个变量，对 g 配方：

$$\mathrm{f}(\mathrm{g}(x))=\left(\mathrm{g}(x)+\tfrac{1}{2}p\right)^2+q-\tfrac{1}{4}p^2=\left(x^2+rx+s+\tfrac{1}{2}p\right)^2-\left(\tfrac{1}{4}p^2-q\right).$$

现在两个方向都可以证明了。

若条件成立，由第一部分可得 p,q,r,s 使该四次式等于 $\mathrm{f}(\mathrm{g}(x))$，而上面的式子显示它等于 $(x^2+vx+w)^2-k$，其中

$$v=r,\qquad w=s+\tfrac{1}{2}p,\qquad k=\tfrac{1}{4}p^2-q.$$

反过来，$(x^2+vx+w)^2-k$ 本身就是 $\mathrm{f}(\mathrm{g}(x))$，其中 $\mathrm{g}(x)=x^2+vx+w$ 而 $\mathrm{f}(x)=x^2-k$（即 $p=0,q=-k,r=v,s=w$）。根据第一部分的必要性论证，条件 $a^3-4ab+8c=0$ 必定成立。

这样就完成了“当且仅当”（if and only if）的证明。实际运用时，把各个值明确写出来会方便得多：展开得

$$(x^2+vx+w)^2-k=x^4+2vx^3+(v^2+2w)x^2+2vw\,x+(w^2-k),$$

所以我们需要

$$v=\tfrac{1}{2}a,\qquad w=\tfrac{1}{2}\left(b-\tfrac{1}{4}a^2\right)=\tfrac{1}{8}(4b-a^2),\qquad k=w^2-d,$$

而这时 x 的系数是 $2vw=\frac{1}{2}a(b-\frac{1}{4}a^2)$，它恰好在条件成立时等于 c。

解四次方程

这里 $a=-4,b=10,c=-12,d=4$，而且

$$a^3-4ab+8c=-64+160-96=0,$$

所以这个方法适用。由公式得

$$v = -2, \qquad w = \tfrac{1}{8}(40 - 16) = 3, \qquad k = 3^2 - 4 = 5,$$

而确实有 $(x^2 - 2x + 3)^2 - 5 = x^4 - 4x^3 + 10x^2 - 12x + 4$（值得亲手展开一次，以策万全）。方程变成

$$(x^2 - 2x + 3)^2 = 5 \quad \Longleftrightarrow \quad x^2 - 2x + 3 = \sqrt{5} \text{ 或 } x^2 - 2x + 3 = -\sqrt{5}.$$

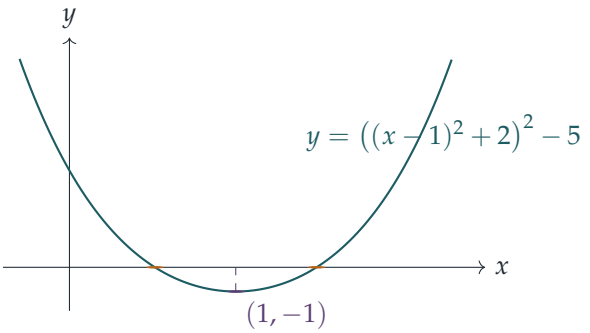
写成 $x^2 - 2x + 3 = (x-1)^2 + 2$：

- $(x-1)^2 = \sqrt{5} - 2$。由于 $\sqrt{5} > 2$，右边是正数，得出两个实根 $x = 1 \pm \sqrt{\sqrt{5} - 2}$。
- $(x-1)^2 = -2 - \sqrt{5} < 0$，得出一对复数根 $x = 1 \pm \mathrm{i}\sqrt{2 + \sqrt{5}}$。

所以方程的根是

$$\boxed{x = 1 \pm \sqrt{\sqrt{5} - 2}, \qquad x = 1 \pm \mathrm{i}\sqrt{2 + \sqrt{5}}}$$

其中只有第一对是实数（如果把题目理解为只求实根，答案就是这两个）。



图像关于 $x = 1$ 对称（它只取决于 $(x-1)^2$），在 $x = 1$ 处取得最小值 $2^2 - 5 = -1$，所以它与横轴恰好相交两次。

验算

- 数值上，$1 \pm \sqrt{\sqrt{5} - 2} \approx 1.4859$ 和 0.5141；把其中任何一个代入 $x^4 - 4x^3 + 10x^2 - 12x + 4$，结果与零相差不超过 10^{-14}，代入 $1 \pm 2.0582\,\mathrm{i}$ 亦然。
- 根之和：四个根加起来是 $4 = -a$，正如所料。
- 试试自己构造一个复合式，例如 $\mathrm{f}(x) = x^2 + x$, $\mathrm{g}(x) = x^2 + 2x$：会得到 $x^4 + 4x^3 + 5x^2 + 2x$，而 $a^3 - 4ab + 8c = 64 - 80 + 16 = 0$。

常见错误

- 对 d 加上条件：常数项总可以用 q（或 k）来配合，所以 d 是自由的。
- 只证明了“当且仅当”的其中一个方向；只要留意到 $(x^2+vx+w)^2-k$ 本身就是 $f(g(x))$ 的形式，一行就能得出逆命题。
- 求出实根后就停笔，没有留意另一个因式会给出一对复数根。

STEP 之外

条件 $a^3-4ab+8c=0$ 有一个简洁的意义：代入 $x=y-\frac{1}{4}a$ 可消去三次项，而这个条件正好表示所得的降次（depressed）四次式没有 y 项，因此它是 y^2 的二次式。这正是费拉里（Ferrari）四次方程一般解法的起点；更广泛地说，它也是“多项式何时能分解为复合式”这个问题的起点，由此可通往伽罗瓦理论（Galois theory）。Allenby 的 *Rings, Fields and Groups*（环、域与群）很适合用来了解哪些多项式方程可用根式求解（solved by radicals）背后的代数。

详细解答

STEP III 2005，第 4 题

纯数学 · 非线性递推关系

STEP III 2005，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/05-s3-q4

关键思路

这个递推关系（recurrence relation）并非线性，但它其实是关于相邻两项之比的。若写成 $u_{n+1} = \rho\, u_n$，递推关系告诉我们下一个比就是 $k - \rho$。因此，这些比在两个和为 k 的数值之间来回交替，其余一切皆由此而来。题目的三个特殊情况，也就化为对这两个比的条件。

解答

归纳法

设 $r = \frac{b}{a}$。把递推关系两边除以 u_{n+1}（当它不等于零时），得

$$\frac{u_{n+2}}{u_{n+1}} = k - \frac{u_{n+1}}{u_n},$$

所以每一个相邻两项之比，都等于 k 减去前一个比。第一个比是 $u_2/u_1 = r$，故下一个比应为

$$c = k - \frac{b}{a} = \frac{ka - b}{a}.$$

注意 $c \neq 0$，因为 $b \neq ka$。另外 $r \neq 0$，因为 $b \neq 0$。这两点合起来保证每一项都不等于零，所以我们永远不会除以零。

设 $P(n)$ 为命题：$u_{2n-1} \neq 0$ 且 $u_{2n} = r\, u_{2n-1}$。$P(1)$ 成立，因为 $u_1 = a \neq 0$ 且 $u_2 = b = ra$。

假设 $P(n)$ 对某个 $n \geq 1$ 成立。那么 $u_{2n} = ru_{2n-1} \neq 0$，而由递推关系得

$$u_{2n+1} = \frac{u_{2n}}{u_{2n-1}}\left(ku_{2n-1} - u_{2n}\right) = r\,(k-r)\,u_{2n-1} = (k-r)\,u_{2n} = c\,u_{2n},$$

这正是题目要求的第二条关系式。因此 $u_{2n+1} \neq 0$，并且

$$u_{2n+2} = \frac{u_{2n+1}}{u_{2n}}\left(ku_{2n} - u_{2n+1}\right) = c\,(k-c)\,u_{2n} = (k-c)\,u_{2n+1} = r\,u_{2n+1},$$

因为 $k - c = r$。这就是 $P(n+1)$。根据数学归纳法（induction），对所有 $n \geq 1$，

$$u_{2n} = \frac{b}{a}\,u_{2n-1}, \qquad u_{2n+1} = c\,u_{2n}, \qquad c = k - \frac{b}{a}.$$

想一想 为甚么要把“$u_{2n-1} \neq 0$”写进归纳命题之中?

递推关系要除以 u_n，所以只有当没有任何一项等于零时，这个数列才有意义。把“不等于零”这一点带进归纳过程，便可以在证明公式的同时，一并证明各项均有定义。$b \neq 0$ 和 $b \neq ka$ 这两个条件，正是在这里用上的。

通项公式

每走两步,先乘以 c,再乘以 $\frac{b}{a}$,所以 $u_{2n+1} = \frac{bc}{a}\,u_{2n-1}$。奇数项构成一个等比数列(geometric progression)，首项为 a，公比为 $\frac{bc}{a}$；而每个偶数项都是它前面那个奇数项的 $\frac{b}{a}$ 倍:

$$u_{2n-1} = a\left(\frac{bc}{a}\right)^{n-1} = \frac{b^{n-1}c^{n-1}}{a^{n-2}}, \qquad u_{2n} = b\left(\frac{bc}{a}\right)^{n-1} = \frac{b^n c^{n-1}}{a^{n-1}}.$$

第 (i) 部分

相邻两项之比依次为 $\frac{b}{a}, c, \frac{b}{a}, c, \ldots$，交替出现，所以数列是等比数列，当且仅当 $c = \frac{b}{a}$，即 $k - \frac{b}{a} = \frac{b}{a}$:

$$ka = 2b.$$

此时公比为 $\frac{b}{a}$。这与 $b \neq ka$ 并无矛盾，因为当 $b \neq 0$ 时，$ka = 2b \neq b$。

第 (ii) 部分

由通项公式可知，对所有 n 都有 $u_{n+2} = \frac{bc}{a}\,u_n$。所以 $u_{n+2} = u_n$ 对所有 n 成立，当且仅当 $\frac{bc}{a} = 1$，即 $b(ka - b) = a^2$:

$$kab = a^2 + b^2.$$

此时数列为 $a, b, a, b, \ldots$。若要求周期 (period) 恰好是 2 (而不是常数数列), 还需要 $a \neq b$。若 $a = b$，条件变成 $k = 2$，而每一项都等于 a。

第 (iii) 部分

同理, $u_{n+4} = \left(\frac{bc}{a}\right)^2 u_n$, 所以 $u_{n+4} = u_n$ 对所有 n 成立, 当且仅当 $\left(\frac{bc}{a}\right)^2 = 1$, 即 $\frac{bc}{a} = \pm 1$。取值 $+1$ 就是第 (ii) 部分的情况，周期为 2（或 1)。若要周期恰好是 4，则需要 $\frac{bc}{a} = -1$,

即 $b(ka-b)=-a^2$：

$$kab=b^2-a^2.$$

此时数列为 $a, b, -a, -b, a, b, \ldots$。由于 $a\neq 0$，我们有 $-a\neq a$，所以它的周期确实是 4，而不会更小。

在 (ii) 和 (iii) 两部分中，条件 $b\neq ka$ 都自动成立。把 $b=ka$ 代入任何一个条件，都会得出 $a^2=0$，而这是不可能的。

验算

试取 $a=1$，$b=2$。当 $k=4$（第 (i) 部分），递推关系给出 $1,2,4,8,16,\ldots$。当 $k=\frac{5}{2}$（第 (ii) 部分），给出 $1,2,1,2,\ldots$。当 $k=\frac{3}{2}$（第 (iii) 部分），给出 $1,2,-1,-2,1,2,\ldots$。我们亦以精确的有理数运算，对大约 2000 组随机选取的 a、b 和 k 执行了这个递推关系。每一组的结果都与通项公式吻合，而且数列是等比数列、周期为 2 或周期为 4，恰好就是上述条件成立的时候。

常见错误

- 递推关系要除以 u_n，却忘记说明为甚么没有任何一项会等于零。
- 把 $\frac{bc}{a}=\pm 1$ 当作 (iii) 的答案。$+1$ 的情况周期是 2。
- 在 (ii) 中，没有留意到 $a=b$、$k=2$ 会得出常数数列。

STEP 之外

代换 $\rho_n=u_{n+1}/u_n$ 把这个非线性递推关系化为一阶映射 $\rho\mapsto k-\rho$。这个映射是一个对合（involution）：连续施行两次便会回到起点，这就是为甚么各个比的周期是 2。通过变量代换来简化非线性的递推关系或微分方程（differential equation），是大学一年级的标准技巧；而研究映射的周期轨道（periodic orbit），正是离散动力系统（discrete dynamical systems）的起点。想找更多同类型的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

STEP III 2005，第 5 题

纯数学 · 抛物线的公切线

STEP III 2005，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/05-s3-q5

关键思路

与其用切点来标记切线(tangent)，不如用斜率(gradient)来标记。对每一个斜率 m，抛物线(parabola) 都恰好有一条该斜率的切线，而第一部分告诉我们它的 y 轴截距 (intercept)。两条抛物线有一条斜率为 m 的公切线 (common tangent)，当且仅当这两个截距相等；于是数公切线的数目，就变成数一条方程的根 m 的数目。这条方程在 $a_1 \neq a_2$ 时是二次方程，在 $a_1 = a_2$ 时是一次方程。

解答

斜率为 m 的切线

$y = ax^2 + bx + c$ 的斜率是 $2ax + b$。由于 $a \neq 0$，它恰好在一点等于 m，即 $x_0 = \dfrac{m-b}{2a}$，而 $y_0 = ax_0^2 + bx_0 + c$。该点的切线是 $y - y_0 = m(x - x_0)$，即 $y - mx = y_0 - mx_0$。利用 $m = 2ax_0 + b$，

$$\begin{aligned} y_0 - mx_0 &= ax_0^2 + bx_0 + c - (2ax_0 + b)x_0 \\ &= c - ax_0^2 = c - a \cdot \frac{(m-b)^2}{4a^2} = c - \frac{(m-b)^2}{4a}, \end{aligned}$$

所以在 P 点的切线是 $y - mx = c - \dfrac{(m-b)^2}{4a}$，正如所求。

有公切线的条件

由第一部分，每条曲线都恰好有一条斜率为 m 的切线，分别是

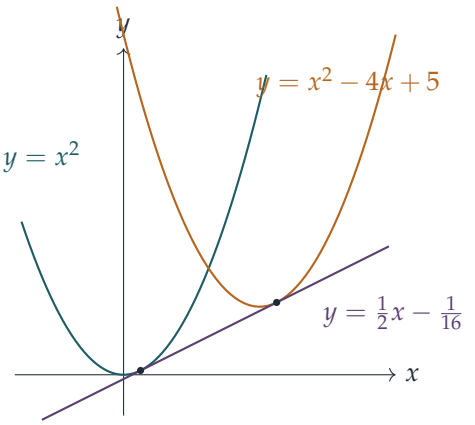
$$y = mx + k_1, \quad k_1 = c_1 - \frac{(m-b_1)^2}{4a_1}, \qquad \text{和} \qquad y = mx + k_2, \quad k_2 = c_2 - \frac{(m-b_2)^2}{4a_2}.$$

两条斜率相同的直线重合，当且仅当它们的截距相等，所以两条曲线有一条斜率为 m 的公切线，当且仅当 $k_1 = k_2$。乘以 $4a_1a_2 \neq 0$（这不会改变等式是否成立），

$$\begin{aligned} 4a_1a_2(k_2 - k_1) &= 4a_1a_2(c_2 - c_1) - a_1(m-b_2)^2 + a_2(m-b_1)^2 \\ &= (a_2 - a_1)m^2 + 2(a_1b_2 - a_2b_1)m + 4a_1a_2(c_2 - c_1) + a_2b_1^2 - a_1b_2^2. \end{aligned}$$

因此，有一条斜率为 m 的公切线，当且仅当这个式子等于零，这就是所求的条件。把该条件的左方记作 $Q(m)$。

由于 $Q(m) = 0$ 的每一个根 m 都恰好给出一条公切线，而不同的根给出不同的直线（它们的斜率不同），所以公切线的数目等于 $Q(m) = 0$ 的相异实根的数目。还要留意，公切线不可能是铅垂的，因为两条曲线都没有铅垂的切线。



两条满足 $a_1 = a_2$ 及 $b_1 \neq b_2$ 的抛物线恰好有一条公切线；图中这条公切线的斜率是 $\frac{1}{2}$。

$a_1 \neq a_2$ 的情况

这时 Q 是一个真正的 m 的二次式，所以它恰好有一个相异实根，当且仅当它的判别式（discriminant）等于零。判别式是

$$\Delta = 4(a_1b_2 - a_2b_1)^2 - 4(a_2 - a_1)\big(4a_1a_2(c_2 - c_1) + a_2b_1^2 - a_1b_2^2\big).$$

不含 c 的各项可以化简得很漂亮：

$$\begin{aligned}&(a_1b_2-a_2b_1)^2-(a_2-a_1)(a_2b_1^2-a_1b_2^2)\\&=\left(a_1^2b_2^2-2a_1a_2b_1b_2+a_2^2b_1^2\right)-\left(a_2^2b_1^2-a_1a_2b_2^2-a_1a_2b_1^2+a_1^2b_2^2\right)\\&=a_1a_2\left(b_1^2-2b_1b_2+b_2^2\right)=a_1a_2(b_1-b_2)^2.\end{aligned}$$

因此

$$\Delta=4a_1a_2\Big[(b_1-b_2)^2-4(a_1-a_2)(c_1-c_2)\Big],$$

而由于 $a_1a_2\neq0$，恰好有一条公切线，当且仅当

$$(b_1-b_2)^2=4(a_1-a_2)(c_1-c_2). \qquad (*)$$

想一想 两条曲线相切（touch）究竟是甚么意思？怎样把它化成一条方程？

两条曲线相切，是指它们在某一点相交，而且在该点的斜率相同。设

$$d(x)=(a_1-a_2)x^2+(b_1-b_2)x+(c_1-c_2),$$

即两个二次式之差。两条曲线在 x_0 相交，即 $d(x_0)=0$；在该处斜率相等，即 $d'(x_0)=0$。所以它们相切，当且仅当 d 有一个根 x_0 满足 $d'(x_0)=0$，也就是有重根（repeated root）。由于 $a_1\neq a_2$，d 是二次式，而它有重根，正正是当它的判别式 $(b_1-b_2)^2-4(a_1-a_2)(c_1-c_2)$ 等于零，这就是条件 $(*)$。

所以，当 $a_1\neq a_2$ 时，两条曲线恰好有一条公切线，当且仅当 $(*)$ 成立，亦即当且仅当它们相切。　　（在这个情况下，唯一的公切线自然就是切点处的切线。）

$a_1=a_2$ 的情况

把共同的值记作 a。Q 的 m^2 项消失，再除以 $a\neq0$，条件变成

$$2(b_2-b_1)m+4a(c_2-c_1)+(b_1-b_2)(b_1+b_2)=0.$$

- 若 $b_1\neq b_2$，这条一次方程恰好有一个根，

  $$m=\frac{b_1+b_2}{2}-\frac{2a(c_2-c_1)}{b_2-b_1},$$

  所以恰好有一条公切线。
- 若 $b_1=b_2$，条件化为 $4a(c_2-c_1)=0$。若 $c_1\neq c_2$，则完全没有公切线（其中一条曲线是另一条沿铅垂方向平移而成，所以它们每个斜率的切线都互相平行而不重合）；若 $c_1=c_2$，则两条曲线完全相同，每一条切线都是公切线，共有无限多条。

因此，当 $a_1=a_2$ 时，两条曲线恰好有一条公切线，当且仅当 $b_1\neq b_2$。

验算

对于 $y = x^2$ 和 $y = -x^2 + 2x - \frac{1}{2}$，两者之差是 $2(x - \frac{1}{2})^2$，所以它们在 $x = \frac{1}{2}$ 相切；这时 $Q(m) = -2m^2 + 4m - 2 = -2(m-1)^2$，只有一个根 $m = 1$，正是 $y = x^2$ 在 $x = \frac{1}{2}$ 的斜率。对于图中的一对曲线 $y = x^2$ 和 $y = x^2 - 4x + 5$，公式给出 $m = -2 + \frac{5}{2} = \frac{1}{2}$，而两条切线的截距都是 $-\frac{1}{16}$。我们还用精确运算，以数千组随机系数验证了恒等式 $Q(m) = 4a_1a_2(k_2 - k_1)$ 和判别式的化简。

常见错误

- 声称 $Q = 0$ 有重根就“表示”两条曲线相切，却没有解释原因：两者的关联要经过另一个二次式 $d(x)$，而把它们连系起来的正是判别式的恒等式。
- 在 $a_1 = a_2$ 的情况下，遗漏了 $b_1 = b_2$、$c_1 = c_2$ 这个子情况（两条曲线完全相同，有无限多条公切线），或者答成“$b_1 \neq b_2$ 或 $c_1 \neq c_2$”。

STEP 之外

不用曲线上的点，而用它的一族切线来描述曲线，这就是射影几何（projective geometry）中对偶性（duality）的想法：圆锥曲线（conic）的切线 $y = mx + k$ 满足 m 与 k 之间自成一格的二次关系，而两条处于一般位置的圆锥曲线有四条公切线，正如它们相交于四点一样。对于两条对称轴为铅垂的抛物线，无穷远线（line at infinity）与两者都相切，并占去了这四条中的两条，所以余下的普通公切线最多只有两条。想做更多同类的题目，可以试试 Siklos 的 *Advanced Problems in Mathematics*（进阶数学难题集）。

详细解答

STEP III 2005，第 6 题

纯数学 · 解三次方程

> **STEP III 2005，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/05-s3-q6

关键思路

恒等式（identity）$4\cosh^3 y - 3\cosh y = \cosh 3y$ 的形状，与“缺项”三次式（depressed cubic）$x^3 - 3bx$ 完全一样，只要把 x 乘上适当的常数便可。因此，没有 x^2 项的三次方程，可以通过与三倍角公式（triple-angle formula）对照来求解，而 arcosh 的对数（logarithm）表达式则把双曲函数（hyperbolic function）形式的答案变回立方根。求得一个根之后，因式分解（factorisation）便给出另外两个根所满足的二次方程，再用单位立方根（cube roots of unity）把它们整理妥当。

解答

双曲函数形式的根

设 $x = 2a\cosh\frac{1}{3}T$。则

$$\begin{aligned} x^3 - 3a^2x &= 8a^3\cosh^3\tfrac{1}{3}T - 6a^3\cosh\tfrac{1}{3}T \\ &= 2a^3\left(4\cosh^3\tfrac{1}{3}T - 3\cosh\tfrac{1}{3}T\right) \\ &= 2a^3\cosh T, \end{aligned}$$

其中用了题目给出的恒等式，并取 $y = \frac{1}{3}T$。所以 $2a\cosh\frac{1}{3}T$ 满足该方程，这正是所要证明的。

$x^3 - 3bx = 2c$ 的一个根

先假设 $c > 0$。由于 $b > 0$，可取 $a = \sqrt{b} > 0$，使得 $3a^2 = 3b$，而我们需要 $2a^3 \cosh T = 2c$，即

$$\cosh T = \frac{c}{b^{3/2}}.$$

这需要 $c/b^{3/2} \geq 1$，而题目的条件正正用在这里：$c^2 \geq b^3$ 且 $c > 0$，故 $c \geq b^{3/2}$。所以可取

$$T = \operatorname{arcosh} \frac{c}{b^{3/2}} = \ln\left(\frac{c}{b^{3/2}} + \sqrt{\frac{c^2}{b^3} - 1} \right) = \ln \frac{c + \sqrt{c^2 - b^3}}{b^{3/2}}.$$

因此

$$e^{T/3} = \frac{(c + \sqrt{c^2 - b^3})^{1/3}}{b^{1/2}} = \frac{u}{\sqrt{b}}, \qquad e^{-T/3} = \frac{\sqrt{b}}{u},$$

而根据第一部分，其中一个根是

$$2\sqrt{b} \cosh \tfrac{1}{3}T = \sqrt{b}\left(\frac{u}{\sqrt{b}} + \frac{\sqrt{b}}{u} \right) = u + \frac{b}{u}.$$

> **想一想** 如果 $c < 0$ 又如何？这时 $c/b^{3/2} \leq -1$，没有任何实数 T 的 cosh 会等于这个值。

把 x 换成 $-x$：方程变成 $x^3 - 3bx = -2c$，其中 $-c > 0$，所以根据刚才处理过的情况，它有根 $u' + b/u'$，其中 $u' = (-c + \sqrt{c^2 - b^3})^{1/3}$，而 $-(u' + b/u')$ 就是原方程的一个根。现在

$$u^3 u'^3 = \left(\sqrt{c^2 - b^3} + c\right)\left(\sqrt{c^2 - b^3} - c\right) = (c^2 - b^3) - c^2 = -b^3,$$

所以，取实立方根，得 $u' = -b/u$ 及 $-(u' + b/u') = u + b/u$。因此无论 c 是正是负，公式都成立。（留意在两种情况下均有 $u \neq 0$，因为当 $b > 0$ 时 $\sqrt{c^2 - b^3} < |c|$。）

另外两个根

记 $r = u + \frac{b}{u}$。由于 r 是一个根，$x - r$ 是因式，在

$$x^3 - 3bx - 2c = (x - r)(x^2 + rx + q) = x^3 + (q - r^2)x - rq$$

中比较系数（comparing coefficients）（x^2 项自动抵消），必有 $q = r^2 - 3b$。展开得

$$q = u^2 + 2b + \frac{b^2}{u^2} - 3b = u^2 + \frac{b^2}{u^2} - b,$$

所以另外两个根满足

$$x^2 + \left(u + \frac{b}{u}\right)x + u^2 + \frac{b^2}{u^2} - b = 0,$$

这正是所要证明的。（作为一致性的检查，常数项亦吻合：$rq = u^3 + b^3/u^3$，而有理化后 $b^3/u^3 = b^3/(c + \sqrt{c^2 - b^3}) = c - \sqrt{c^2 - b^3}$，所以 $rq = 2c$。）

要解这个二次方程，记 $v = b/u$，于是 $uv = b$，而二次方程成为

$$x^2 + (u+v)x + (u^2 - uv + v^2) = 0.$$

现在 $\omega^3 = 1$ 且 $1 + \omega + \omega^2 = 0$。试取 $x_1 = \omega u + \omega^2 v$ 和 $x_2 = \omega^2 u + \omega v$：

$$x_1 + x_2 = (\omega + \omega^2)(u+v) = -(u+v),$$
$$x_1 x_2 = \omega^3 u^2 + \omega^2 uv + \omega^4 uv + \omega^3 v^2 = u^2 + (\omega^2 + \omega)uv + v^2 = u^2 - uv + v^2.$$

这正是该二次方程的两根之和与两根之积，所以它的根是

$$x = \omega u + \omega^2 \frac{b}{u} \qquad 及 \qquad x = \omega^2 u + \omega \frac{b}{u}.$$

由于 $\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}\mathrm{i}$ 及 $\omega^2 = -\frac{1}{2} - \frac{\sqrt{3}}{2}\mathrm{i}$，这两个根亦可写成 $-\frac{1}{2}\left(u + \frac{b}{u}\right) \pm \frac{\sqrt{3}}{2}\left(u - \frac{b}{u}\right)\mathrm{i}$。

解 $x^3 - 6x = 6$

这是 $x^3 - 3bx = 2c$，其中 $b = 2$，$c = 3$，而 $c^2 = 9 \geq 8 = b^3 > 0$，所以上述结果适用。这里

$$u = \left(3 + \sqrt{9-8}\right)^{1/3} = 4^{1/3} = 2^{2/3}, \qquad \frac{b}{u} = \frac{2}{2^{2/3}} = 2^{1/3}.$$

所以各根为

$$x = 2^{2/3} + 2^{1/3}, \qquad x = -\tfrac{1}{2}\left(2^{2/3} + 2^{1/3}\right) \pm \tfrac{\sqrt{3}}{2}\left(2^{2/3} - 2^{1/3}\right)\mathrm{i},$$

其中那对复数根（complex roots）就是 $\omega\, 2^{2/3} + \omega^2\, 2^{1/3}$ 及其共轭（conjugate）。

验算

- 实根是 $2^{2/3} + 2^{1/3} \approx 2.847322$，而 $2.847322^3 - 6(2.847322) \approx 23.083930 - 17.083932 \approx 6.0000$。
- 三个根之和应为 0（没有 x^2 项）：$(u+v) - (u+v) = 0$。它们的积应为 6：$r(u^2 - uv + v^2) = u^3 + v^3 = 4 + 2 = 6$。
- 若 $c^2 = b^3$，则 $u = \sqrt{b} = b/u$，两个复数根都变成 $-\sqrt{b}$，即一个重根（repeated root），这可由 $x^3 - 3bx - 2b^{3/2} = (x - 2\sqrt{b})(x + \sqrt{b})^2$ 印证。

常见错误

- 忘记了 $\cosh T \geq 1$，所以双曲函数的论证只能直接用于 $c > 0$；$c < 0$ 的情况需要代换 $x \to -x$。

- 把复数根写成 $\omega u + \omega \frac{b}{u}$：两个立方根的积必须保持等于 b，所以 ω 的幂相乘必须等于 1。

STEP 之外

这就是三次方程的卡尔达诺公式（Cardano's formula），而 $u + b/u$ 这个形式（其中 u^3 是某个二次方程的根），正是十六世纪最初发现它时的模样。当 $c^2 < b^3$ 时，虽然三个根全是实数，平方根却是虚数（即所谓“不可约情况”（irreducible case）），这时便由三角形式 $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$ 接手。至于一般的五次方程为何没有这类公式，那就是伽罗瓦理论（Galois theory）的故事了；Allenby 的 *Rings, Fields and Groups*（环、域与群）是很好的入门读物。

详细解答

STEP III 2005，第 7 题

纯数学 · 代换积分法

STEP III 2005，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/05-s3-q7

关键思路

题目开首的结果告诉我们：只要被积函数（integrand）外面另有一个单独的因子 $\frac{1}{x}$，藏在被积函数里的 x^m 就可以当作一个新变量 $u = x^m$ 来处理。因此，对每一个积分，我们先从分母抽出一个因子 x，看看剩下的是哪一个幂 x^m、函数 f 又是甚么，然后改为计算容易得多的 u 积分。

解答

一般结果

已知 $\mathrm{F}'(u) = \dfrac{1}{u\,\mathrm{f}(u)}$。用链式法则（chain rule）对 $\mathrm{F}(x^m)$ 求导（取 $x > 0$，使 x^m 对任何实数 m 都有意义），

$$\frac{\mathrm{d}}{\mathrm{d}x}\,\mathrm{F}(x^m) = \mathrm{F}'(x^m)\cdot mx^{m-1} = \frac{mx^{m-1}}{x^m\,\mathrm{f}(x^m)} = \frac{m}{x\,\mathrm{f}(x^m)}.$$

所以 $\mathrm{F}(x^m)$ 是 $\dfrac{m}{x\,\mathrm{f}(x^m)}$ 的一个原函数（antiderivative），这正是所要证明的结果：

$$\int \frac{m}{x\,\mathrm{f}(x^m)}\,\mathrm{d}x = \mathrm{F}(x^m) + c.$$

换句话说，这就是把代换（substitution）$u = x^m$、$\mathrm{d}u = mx^{m-1}\,\mathrm{d}x$ 倒过来看。条件 $m \neq 0$ 在下文十分重要：我们将会除以 m。

实际运用时，我们把它写成

$$\int \frac{1}{x\,\mathrm{f}(x^m)}\,\mathrm{d}x = \frac{1}{m}\,\mathrm{F}(x^m) + c. \qquad (*)$$

第 (i) 部分

抽出一个因子 x，

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

这符合 $(*)$ 的形式，其中 $m = n - 1$，$\mathrm{f}(u) = u - 1$。我们需要 $n \neq 1$；当 $n = 1$ 时，分母 $x^n - x$ 恒等于零，根本没有积分可求。

接着要求 F，即 $\dfrac{1}{u(u-1)}$ 的一个原函数。由部分分式（partial fractions）得

$$\frac{1}{u(u-1)} = \frac{1}{u-1} - \frac{1}{u}, \qquad \text{所以} \qquad \mathrm{F}(u) = \ln|u-1| - \ln|u| = \ln\left|\frac{u-1}{u}\right|.$$

把 $u = x^{n-1}$ 代入 $(*)$：

$$\boxed{\int \frac{1}{x^n - x}\,\mathrm{d}x = \frac{1}{n-1}\ln\left|\frac{x^{n-1}-1}{x^{n-1}}\right| + c, \qquad n \neq 1.}$$

有了绝对值（modulus）符号，这条公式在被积函数有定义的任何区间上都成立（例如在 $0 < x < 1$ 上和在 $x > 1$ 上都成立；当 n 是整数时，在 $x < 0$ 上也成立）。

第 (ii) 部分

这里把 x 从平方根中抽出来时要小心一点。我们只考虑 $x > 0$，这时 x^n 对每个 n 都有定义，而且 $\sqrt{x^2} = x$。于是

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

所以被积函数是

$$\frac{1}{x\sqrt{1 + x^{n-2}}},$$

这符合 $(*)$ 的形式，其中 $m = n - 2$，$\mathrm{f}(u) = \sqrt{1+u}$。

? 想一想 当 $n = 2$ 时会怎样?

这时 $m = 0$，一般结果并不适用。不过这个积分可以直接计算，十分容易：被积函数是 $\dfrac{1}{\sqrt{2x^2}} = \dfrac{1}{\sqrt{2}\,x}$，所以

$$\int \frac{1}{\sqrt{x^2 + x^2}}\,\mathrm{d}x = \frac{1}{\sqrt{2}}\ln x + c \qquad (n = 2).$$

当 $n \neq 2$ 时，我们需要 $\mathrm{F}(u) = \displaystyle\int \frac{1}{u\sqrt{1+u}}\,\mathrm{d}u$，其中 $u = x^{n-2} > 0$。代入 $t = \sqrt{1+u}$，则 $u = t^2 - 1$，$\mathrm{d}u = 2t\,\mathrm{d}t$，且 $t > 1$：

$$\int \frac{1}{u\sqrt{1+u}}\,\mathrm{d}u = \int \frac{2t}{(t^2-1)\,t}\,\mathrm{d}t = \int\left(\frac{1}{t-1} - \frac{1}{t+1}\right)\mathrm{d}t = \ln\left(\frac{t-1}{t+1}\right),$$

由于 $t>1$，这里不需要绝对值符号。这个结果可以化简得很整齐：把分子和分母同乘以 $t-1$，并利用 $t^2-1=u$，

$$\frac{t-1}{t+1}=\frac{(t-1)^2}{t^2-1}=\frac{\left(\sqrt{1+u}-1\right)^2}{u},\qquad \text{所以}\qquad \mathrm{F}(u)=2\ln\left(\sqrt{1+u}-1\right)-\ln u.$$

现在把 $u=x^{n-2}$ 代入 $(*)$，其中 $m=n-2$。$-\ln u$ 一项变成 $-(n-2)\ln x$，再除以 $m=n-2$ 便成为 $-\ln x$：

$$\boxed{\int\frac{1}{\sqrt{x^n+x^2}}\,\mathrm{d}x=\frac{2}{n-2}\ln\left(\sqrt{1+x^{n-2}}-1\right)-\ln x+c,\qquad n\neq 2,\ x>0.}$$

答案也可以等价地写成 $\dfrac{1}{n-2}\ln\left(\dfrac{\sqrt{1+x^{n-2}}-1}{\sqrt{1+x^{n-2}}+1}\right)+c$；两种形式都可以。

验算

- 把答案求导，看看能否还原。对于 (i)，取 $n=2$，公式给出 $\ln|x-1|-\ln|x|$，其导数是 $\frac{1}{x-1}-\frac{1}{x}=\frac{1}{x^2-x}$，正如所料。
- 对于 (i),取 $n=3$,公式给出 $\displaystyle\int_2^3\frac{\mathrm{d}x}{x^3-x}=\frac{1}{2}\ln\frac{32}{27}\approx 0.084950$,与辛普森法则(Simpson's rule）的结果吻合。
- 对于 (ii)，取 $n=3$，公式给出 $\displaystyle\int_1^3\frac{\mathrm{d}x}{\sqrt{x^3+x^2}}=\ln\frac{3+2\sqrt{2}}{3}\approx 0.664135$，同样与数值积分（numerical integration）的结果吻合。我们还用许多不同的 n 值检验过这两条公式，包括负数和非整数的值。

常见错误

- 漏掉因子 $\frac{1}{m}$：一般结果的分子有 m，而我们的被积函数没有，所以必须除以 m。
- 不加说明就写 $\sqrt{x^2}=x$。这只在 $x\geq 0$ 时成立；当 $x<0$ 时，它等于 $-x$（或写成 $|x|$）。
- 遗漏例外的值：(i) 中的 $n=1$ 和 (ii) 中的 $n=2$，这时 $m=0$。

STEP 之外

题目开首的结果，其实是从另一面看代换法则：代换不过是把链式法则倒过来用，而在大学第一门分析（analysis）课程中，代换法则正是这样证明的，当中会对所涉及的函数仔细列明假设。像 (ii) 这类含有多项式平方根的积分，会引出一个问题：到底哪些积分可以用初等函数（elementary functions）表示? 例如 $\int\mathrm{d}x/\sqrt{1+x^4}$ 就不可以。如想看微积

分基本定理（fundamental theorem of calculus）和代换法则的严谨论述，可参阅 Abbott 的 *Understanding Analysis*。

详细解答

STEP III 2005，第 8 题

纯数学 · 复数与圆

STEP III 2005，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/05-s3-q8

关键思路

阿根图（Argand diagram）上的每一个长度，都可以用 $|z|^2 = zz^*$ 化为代数式。一旦得到第一条恒等式，直角就变成毕氏定理（Pythagoras′ theorem），而切线（tangent）就变成切线与半径之间的直角。于是“OA 与圆相切”就化为单独一条方程，由此定出半径的平方为 $cc^* - aa^*$。此后，“P 在圆上”与“P' 在圆上”都归结为同一条方程。

解答

恒等式与直角

由于 $|z|^2 = zz^*$，而且取共轭（conjugate）与减法可以交换次序，

$$\begin{aligned} |a-c|^2 &= (a-c)(a-c)^* = (a-c)(a^*-c^*) \\ &= aa^* + cc^* - ac^* - ca^*, \end{aligned}$$

这正是所要证明的。

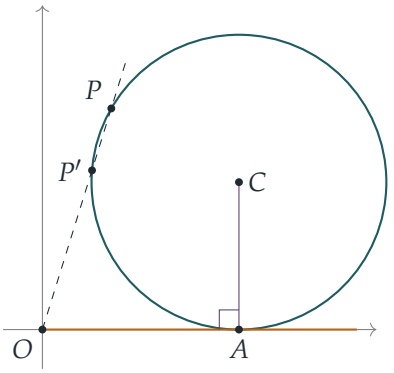
三角形 OAC 的边长为 $OA = |a|$、$OC = |c|$ 和 $AC = |a-c|$，由于 a, c 非零且互不相同，三者皆不为零。根据毕氏定理及其逆定理（converse），在 A 的角是直角，当且仅当 $OC^2 = OA^2 + AC^2$，即

$$cc^* = aa^* + \left(aa^* + cc^* - ac^* - ca^*\right) \iff 2aa^* = ac^* + ca^*.$$

这里每一步都可逆，所以“当且仅当”成立。

（毕氏定理的逆定理同时排除了退化的“三角形”。假如 O、A、C 共线（collinear），则三个长度之中有一个等于另外两个之和；当三个长度皆为正数时，这不可能与 $OC^2 = OA^2 + AC^2$ 同时成立。）

相切的圆



切线在切点处与半径垂直。因此 OA 在 A 与圆相切，即表示角 OAC 是直角，而根据第一部分，

$$2aa^* = ac^* + ca^*. \qquad (*)$$

半径 r 满足 $r^2 = |a-c|^2 = aa^* + cc^* - (ac^* + ca^*) = cc^* - aa^*$，当中用到 $(*)$。（这其实又是毕氏定理：$r^2 = OC^2 - OA^2$。）

设 $k = bb^* > 0$ 及 $s = abc^* + a^*b^*c$。则

$$|ab-c|^2 = (ab-c)(a^*b^* - c^*) = aa^*k - s + cc^*,$$
$$\left|\frac{a}{b^*} - c\right|^2 = \left(\frac{a}{b^*} - c\right)\left(\frac{a^*}{b} - c^*\right) = \frac{aa^*}{k} - \frac{ac^*}{b^*} - \frac{a^*c}{b} + cc^*.$$

把第二式乘以 $k = bb^*$，得

$$k\left|\frac{a}{b^*} - c\right|^2 = aa^* - s + k\,cc^*. \qquad (**)$$

P **在圆上**。$|ab-c|^2 = cc^* - aa^*$ 成立，当且仅当

$$aa^*k - s + cc^* = cc^* - aa^* \iff aa^*(k+1) = s.$$

P' **在圆上**。$|a/b^* - c|^2 = cc^* - aa^*$ 成立，当且仅当（乘以 $k \neq 0$ 并利用 $(**)$）

$$aa^* - s + k\,cc^* = k\,cc^* - k\,aa^* \iff aa^*(k+1) = s.$$

两个条件完全相同，所以 P 在圆上当且仅当 P' 在圆上。

逆命题

现在不再假设相切。设圆的圆心为 C 并通过 A，于是 $r^2 = |a-c|^2 = aa^* + cc^* - ac^* - ca^*$。再设 ab 和 a/b^* 都在圆上，其中 $k = bb^* \neq 1$。利用同样的展开式，

$$aa^*k - s + cc^* = r^2,$$
$$aa^* - s + k\,cc^* = kr^2 \qquad \text{（由 } (**)\text{）}.$$

从第一条方程减去第二条，便可消去 s：

$$(k-1)\,aa^* - (k-1)\,cc^* = (1-k)\,r^2.$$

> **想一想** 为甚么题目坚持要 $bb^* \neq 1$?

若 $bb^* = 1$，则 $b^* = 1/b$，于是 $a/b^* = ab$，两点重合。“两点都在圆上”这时只是单独一个条件，当然不足以推出相切。在我们的代数运算中，这正正就是除以 $k-1 \neq 0$ 的那一步：

$$r^2 = cc^* - aa^*.$$

与 $r^2 = aa^* + cc^* - ac^* - ca^*$ 比较，得 $2aa^* = ac^* + ca^*$。根据第一部分，角 OAC 是直角。因此 OA 在圆上的点 A 处与半径 CA 垂直，即 OA 是圆的切线。

验算

- 这个结果背后有一个简洁的几何原因。由于 $\dfrac{a}{b^*} = \dfrac{ab}{bb^*}$，点 P' 就是把 P 按正实数 $1/|b|^2$ 缩放而得。所以 O、P、P' 共线，且 $OP \cdot OP' = |a|^2 = OA^2$，这正是切割线定理（tangent–secant theorem）（见图）。
- 数值上，取 $a = 2$、$c = 2 + 1.5\mathrm{i}$（于是 $r^2 = 2.25 = |c|^2 - |a|^2$），以及圆上的 $P = 0.701 + 2.25\mathrm{i}$。则 $b = P/2$ 及 $P' = a/b^* \approx 0.505 + 1.620\mathrm{i}$，而 $|P' - c|^2 \approx 2.250$。
- 我们用电脑以随机的 a, b, c 验证了恒等式和相切的情况。另外又取了 1000 个随机的非相切圆，并设 $|b| \neq 1$：每一次，只要 P 在圆上，P' 就必定不在圆上。

常见错误

- 正确写出 $(a-c)^* = a^* - c^*$，其后却把 $|a-c|^2$ 展开成 $(a-c)^2$。模（modulus）的平方必须用到共轭。
- 在逆命题中，一开始就假设相切（从而假设 $r^2 = cc^* - aa^*$）。这正是你要证明的结论。只可从 $r = |a-c|$ 出发。
- 除以 $1 - bb^*$ 时没有说明它为何不等于零。

STEP 之外

由于 $|a|^2/(ab)^* = aa^*/(a^*b^*) = a/b^*$，点 P' 是 P 在 $z \mapsto |a|^2/z^*$ 之下的像。这个映射就是关于圆 $|z| = |a|$ 的反演（inversion）：它把每一点送到由 O 出发的同一条射线上、距离为 $|a|^2/OP$ 的那一点。本题说明，与反演圆正交（orthogonal）的圆会被映射到自身。反演以及由它生成的莫比乌斯变换（Möbius transformations），是大学一年级复分析（complex analysis）和几何的核心内容。Needham 的 *Visual Complex Analysis* 用了整整一章的篇幅作出精彩的论述。

详细解答

STEP III 2005，第 9 题

力学 · 碰撞与恢复系数

STEP III 2005，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：

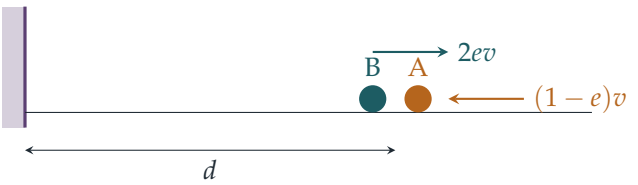
step.maths.org/questions/05-s3-q9

关键思路

这两个质量看似古怪，其实是特意选定的，好让第一次碰撞得出非常整齐的答案。求出第二次碰撞之后，把它发生前一刻的速度与第一次碰撞前一刻的速度比较：两者相同，只是各自乘以 e。碰撞定律对速度而言是线性（linear）的，所以整个过程会以较小的尺度重演，其后的每一次碰撞也就一并得出。

解答

我们把 A 和 B 视为光滑水平直线上的质点（particle），每次碰撞均应用牛顿实验定律（Newton's experimental law）。取朝向墙壁的方向为正，距离则由墙壁量起。



（图中所示为第一次碰撞前的一刻，两个质点画得稍为分开。）

第一次碰撞

设碰撞后 A 和 B 的速度（朝向墙壁）分别为 v_A 和 v_B。碰撞前，A 的速度为 $(1-e)v$，B 的速度为 $-2ev$。

由动量守恒（conservation of momentum）得

$$\begin{aligned}4em\,v_A+(1-e)^2m\,v_B &= 4em(1-e)v-(1-e)^2m\cdot 2ev\\ &= 2e(1-e)v\big[2-(1-e)\big]m = 2e(1-e)(1+e)v\,m.\end{aligned}$$

接近速率（speed of approach）为 $(1-e)v+2ev=(1+e)v$，故由牛顿定律得

$$v_B-v_A=e(1+e)v.$$

把 $v_B=v_A+e(1+e)v$ 代入动量方程，再除以 m：

$$\big[4e+(1-e)^2\big]v_A = 2e(1-e)(1+e)v-(1-e)^2e(1+e)v = e(1-e)(1+e)v\big[2-(1-e)\big].$$

由于 $4e+(1-e)^2=(1+e)^2$，右方等于 $e(1-e)(1+e)^2v$，因此

$$v_A=e(1-e)v,\qquad v_B=e(1-e)v+e(1+e)v=2ev.$$

两者皆为正数：碰撞后，**A 以速率** $e(1-e)v$ **朝向墙壁运动，B 以速率** $2ev$ **朝向墙壁运动**。

第二次碰撞

B 较快，所以会抛离 A，并在时间 $t_1=d/(2ev)$ 之后到达墙壁。在这段时间内，A 移动了 $e(1-e)v\cdot\frac{d}{2ev}=\frac{1}{2}(1-e)d$，所以 A 与墙壁的距离为

$$d-\tfrac{1}{2}(1-e)d=\tfrac{1}{2}(1+e)d$$

。B 以速率 $e\cdot 2ev=2e^2v$ 反弹，此时正远离墙壁，而 A 仍以速率 $e(1-e)v$ 朝向墙壁运动。两者互相接近，接近速率为 $e(1-e)v+2e^2v=e(1+e)v>0$，所以它们必定再次相遇（而且是在 A 到达墙壁之前，因为 B 位于 A 与墙壁之间）。所需时间为

$$t_2=\frac{\frac{1}{2}(1+e)d}{e(1+e)v}=\frac{d}{2ev},$$

其间 B 由墙壁起计移动了 $2e^2v\cdot\frac{d}{2ev}=ed$。所以**第二次碰撞确会发生，位置距离墙壁** de，正如题目所求。

其后的碰撞

> ? **想一想** 怎样才能不必把所有功夫重做一遍，便推断出还有更多碰撞?

考察第二次碰撞前一刻的状态: A 以速率 $e(1-e)v$ 朝向墙壁运动, B 以速率 $2e^2v=2e(ev)$ 远离墙壁。这正好是把 v 换成 ev、把 d 换成 de 之后的初始状态（质量和 e 均不变）。上文所做的一切对任何 $v>0$ 和 $d>0$ 都成立，所以把 (v,d) 换成 (ev,de) 后，同一论证依

然适用：第三次碰撞会发生，位置距离墙壁 $(de)e = de^2$，而它发生前一刻的状态就是取 (e^2v, de^2) 的初始状态。由归纳法（induction）可知，对每个 n 都有第 $(n+1)$ 次碰撞，所以碰撞会永远持续下去。

第 n 次碰撞前一刻的状态，就是把 v 和 d 分别换成 $e^{n-1}v$ 和 de^{n-1} 的初始状态，所以

$$\text{第 } n \text{ 次碰撞发生于距离墙壁} de^{n-1} \text{ 之处。}$$

由第二部分可知，由一次碰撞到下一次碰撞所需的时间为 $t_1 + t_2 = d/(ev)$。若由第 n 次碰撞起计，这个时间变成

$$\frac{de^{n-1}}{e \cdot e^{n-1}v} = \frac{d}{ev},$$

与 n 无关。所以相继两次碰撞之间的时间全部相等，均为 $d/(ev)$。每一个循环中，距离和速率都按因子 e 缩小，这两个效应在时间上互相抵销。

验算

- 第一次碰撞的动量：取 $e = \frac{1}{2}$、$m = v = 1$，两个质量分别为 2 和 $\frac{1}{4}$；碰撞前，$2 \cdot \frac{1}{2} - \frac{1}{4} \cdot 1 = \frac{3}{4}$；碰撞后，$2 \cdot \frac{1}{4} + \frac{1}{4} \cdot 1 = \frac{3}{4}$。相对速率由 $\frac{3}{2}$ 变为 $\frac{3}{4}$，即乘以因子 e。
- 我们以分数作精确运算，就 200 组随机选取的 e、v、d、m 模拟了这个运动，每组计算八次碰撞：每次碰撞的位置都是 de^{n-1}，而相隔的时间全部是 $d/(ev)$。当 $e = \frac{1}{2}$、$v = d = 1$ 时，碰撞发生于时间 $0, 2, 4, 6$，距离为 $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$。

常见错误

- 混淆正负号：B 起初是远离墙壁运动的，所以按我们的约定，它的初速度是负数。
- 忘记墙壁的恢复系数（coefficient of restitution）同样是 e，所以 B 离开墙壁时的速率是 $2e^2v$，而不是 $2ev$。
- 用蛮力把第二次碰撞重新计算一遍，而没有留意到当中的缩放关系；这样做也行得通，但冗长得多，而且容易看漏当中的规律。

STEP 之外

这里的推论用到线性和自相似性（self-similarity）：若把每个速度都乘以同一个常数，碰撞的方程保持不变，所以一个除尺度外能重现自身的状态会永远重复下去。同一想法也是分析弹跳无限多次的球的基础，更广泛而言，亦是量纲分析（dimensional analysis）的基础。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门）正是以这种严谨的方式处理碰撞和动量。

详细解答

STEP III 2005，第 10 题

力学 · 弹性带与摩擦力

STEP III 2005，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/05-s3-q10

关键思路

一切都取决于一个变量：两圆盘圆心之间的距离 d。橡皮圈的长度由 d 决定，因此它的张力（tension）和所储存的能量都是 d 的简单函数。由此出发，能量的考虑（弹性势能（elastic energy）对比克服摩擦力（friction）所作的功）告诉我们圆盘在哪里停下，而比较橡皮圈的拉力与极限摩擦力（limiting friction），则告诉我们圆盘停下后会否保持静止。

解答

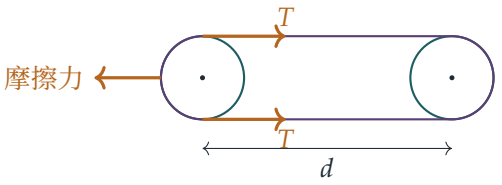
我们把两个圆盘视为完全相同，橡皮圈视为轻而有弹性，且与圆盘的接触是光滑（smooth）的，并假设圆盘一旦滑动，摩擦力的大小就是 μmg。由对称性，两个圆盘总是沿相反方向移动相等的距离，所以橡皮圈的两段直线部分始终保持平行。

把橡皮圈表示为 d 的函数。 当两圆心相距 d 时，橡皮圈绕过每个圆盘半周，再沿着两段长度为 d 的直线部分伸展。所以它的长度是 $2\pi r + 2d$，伸长量（extension）是 $x = 2d$。根据虎克定律（Hooke's law），取弹性模量（modulus of elasticity）$\lambda = \frac{\pi mg}{12}$ 及自然长度（natural length）$2\pi r$，

$$T = \frac{\lambda x}{2\pi r} = \frac{\pi mg}{12} \cdot \frac{2d}{2\pi r} = \frac{mg\,d}{12r}, \qquad E = \frac{\lambda x^2}{2 \cdot 2\pi r} = \frac{mg\,d^2}{12r}.$$

由于接触是光滑的，整条橡皮圈上的张力处处相同。每个圆盘都受两段直线部分拉扯，两者都平行于圆心连线，并指向另一个圆盘。所以每个圆盘所受的合拉力是

$$P = 2T = \frac{mg\,d}{6r}.$$



上图为俯视图，显示左边圆盘向另一个圆盘滑动时所受的水平力。它的重量与法向反作用力（normal reaction）都是铅垂的，并且互相抵消。

释放时 $d = 6r$，所以 $E_0 = 3mgr$ 而 $P = mg$。接触时 $d = 2r$，所以 $E = \frac{1}{3}mgr$。

圆盘互相靠近时的能量。 当圆盘从 $d = 6r$ 向内移动时，每个圆盘都要克服摩擦力 μmg 移动 $\frac{1}{2}(6r - d)$ 的距离。所以克服摩擦力所作的总功是 $\mu mg(6r - d)$。因此总动能（kinetic energy）为

$$\begin{aligned} K(d) &= 3mgr - \frac{mg\,d^2}{12r} - \mu mg(6r - d) \\ &= mg\,(6r - d)\left(\frac{6r + d}{12r} - \mu\right), \end{aligned}$$

其中用到 $3r - \frac{d^2}{12r} = \frac{(6r-d)(6r+d)}{12r}$。

第 (i) 部分

起动。释放时每个圆盘所受的拉力是 $P = mg$，而摩擦力最多只能提供 μmg。所以圆盘开始滑动，当且仅当 $mg > \mu mg$，即 $\mu < 1$。

停下。假设 $\mu < 1$。对于 $d < 6r$，$K(d)$ 只会在 $\frac{6r+d}{12r} = \mu$ 之处为零，即在

$$d^* = 12\mu r - 6r.$$

当 d 由 $6r$ 开始减少时，第二个因子 $\frac{6r+d}{12r} - \mu$ 由 $1 - \mu > 0$ 开始递减，所以在 d 到达 d^* 之前 $K > 0$。圆盘在相撞之前停下，当且仅当在接触之前已到达 d^*，即 $2r < d^* < 6r$：

$$8r < 12\mu r < 12r \iff \tfrac{2}{3} < \mu < 1.$$

若 $\mu \le \frac{2}{3}$，则 $K(2r) \ge 0$，圆盘会先碰上对方。（当 $\mu = \frac{2}{3}$ 时，它们刚好以零速率互相接触，这不算是“之前”。）

> **想一想** 圆盘在 d^* 停下之后，橡皮圈会否令它们再次移动？

在 d^* 处，拉力是 $P = \frac{mg\,d^*}{6r} = (2\mu - 1)mg$。由于 $\mu < 1$，这小于极限摩擦力 μmg，所以静摩擦力（static friction）能把圆盘固定，它们保持静止。结合两个条件，圆盘开始滑动但在相撞之前停下，当且仅当 $\frac{2}{3} < \mu < 1$。 ■

第 (ii) 部分

若圆盘至少相撞一次，它们首次到达 $d = 2r$ 时必定是向内移动，而靠近阶段的公式仍然有效。所以紧接首次碰撞之前的总动能是

$$K(2r) = mg\,(4r)\left(\frac{8r}{12r} - \mu\right) = 4mgr\left(\tfrac{2}{3} - \mu\right) = \tfrac{4}{3}mgr(2 - 3\mu),$$

正如所求。（此值为正，恰好当 $\mu < \frac{2}{3}$，与第 (i) 部分一致。）

第 (iii) 部分

首先，由 $\mu^2 < \frac{4}{9}$ 得 $\mu < \frac{2}{3}$，所以根据第 (i) 和第 (ii) 部分，圆盘确实会相撞。碰撞中损失一半动能，所以它们从 $d = 2r$ 反弹时的总动能是

$$K_1 = \tfrac{2}{3}mgr(2 - 3\mu).$$

现在圆盘互相远离。橡皮圈的拉力和摩擦力都指向另一个圆盘，所以每个圆盘都在减速。当间距由 $2r$ 增至 d 时，弹性势能增加了 $\frac{mg(d^2-4r^2)}{12r}$，而克服摩擦力所作的功是 $\mu mg(d - 2r)$，所以

$$K(d) = K_1 - \frac{mg(d^2 - 4r^2)}{12r} - \mu mg(d - 2r).$$

这是 d 的严格递减函数，所以它恰好有一个零点，就是圆盘停下时的间距 d_1。设 $d_1 = 2r + u$，并乘以 $\frac{12r}{mg}$：

$$u^2 + 4r(1 + 3\mu)\,u - 8r^2(2 - 3\mu) = 0.$$

取正根，并化简判别式（discriminant），$(1 + 3\mu)^2 + 2(2 - 3\mu) = 9\mu^2 + 5$，得

$$u = 2r\left(\sqrt{9\mu^2 + 5} - (1 + 3\mu)\right), \qquad d_1 = 2r\left(\sqrt{9\mu^2 + 5} - 3\mu\right).$$

（可以验证 $2r < d_1 \le 6r$，所以橡皮圈保持绷紧，其能量公式仍然适用。）

> ? **想一想** 圆盘在 d_1 处瞬间静止。甚么因素决定它们会否留在那里？

若橡皮圈的拉力不超过极限摩擦力，它们便保持静止：

$$\frac{mg\,d_1}{6r} \le \mu mg \iff \sqrt{9\mu^2 + 5} - 3\mu \le 3\mu \iff 9\mu^2 + 5 \le 36\mu^2 \iff \mu^2 \ge \tfrac{5}{27},$$

其中两边皆为正数，所以可以取平方。因此若 $\frac{5}{27} < \mu^2 < \frac{4}{9}$，圆盘会相撞、反弹、单调地减速至在 d_1 处静止，然后被静摩擦力永久固定在那里。它们在首次碰撞之后恰好停下一次。 ■

验算

- 当 $\mu = \frac{2}{3}$ 时，第 (i) 部分的停止点 $d^* = 12\mu r - 6r$ 和第 (ii) 部分的动能都给出“以零速率到达 $d = 2r$”，所以两部分在边界上吻合。
- 取 $\mu = 0.5$（在第 (iii) 部分的范围之内）：$K(2r) = \frac{2}{3}mgr$，$K_1 = \frac{1}{3}mgr$，而 $d_1 = 2r(\sqrt{7.25} - 1.5) \approx 2.385r$。该处的拉力 $\approx 0.397mg < 0.5mg$，所以圆盘留在原处。以逐步数值模拟这个运动，重现了以上所有结果，亦重现了多个 $\mu \in (\frac{2}{3}, 1)$ 所对应的停止点 $d^* = 12\mu r - 6r$。对于 $\mu^2 < \frac{5}{27}$（例如 $\mu = 0.42$），模拟显示圆盘停下后再次移动，与预测相符。

常见错误

- 把伸长量当作 d 而不是 $2d$：两段直线部分都被拉长。
- 每个圆盘只计算一个张力：橡皮圈以两股拉着每个圆盘，合共 $2T$。
- 忘记检查圆盘停下后是否保持静止。“速率为零”与“永久静止”是两个不同的条件。

STEP 之外

摩擦力令这成为一个分段系统。每当速度改变正负号，运动方程就会改变；而当外加的力低于摩擦力的临界值时，便会出现“卡住”的状态。这类系统（黏滑运动，stick–slip motion）在大学一年级的动力学课程中会以相平面（phase-plane）方法研究。当 μ 很小时，带有能量损失的反复碰撞会在有限时间内产生无限多次反弹，这值得细想。关于能量方法和摩擦力的更多内容，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）。

详细解答

STEP III 2005，第 11 题

力学 • 刚体平衡

STEP III 2005，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/05-s3-q11

关键思路

转轴与各杆组成一个绕固定水平轴转动的刚体（rigid body），所以只需一个角 θ 便能描述整个系统，而能量自然是最合适的工具。势能（potential energy）是三个余弦之和，它们的角度相差 $\frac{2}{3}\pi$；把它们合并，便得到单一的表达式 $A\cos\theta + B\sin\theta$，这其实是一个乔装了的余弦波。一旦写成这个形式，平衡位置、它们的稳定性，以及越过最高点所需的能量，全都可以直接读出。

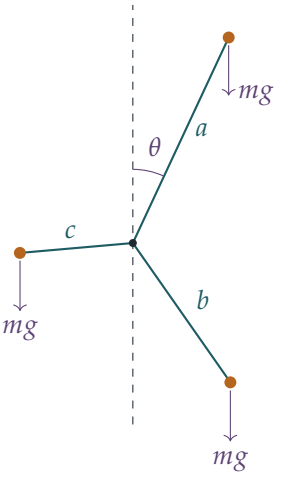
解答

能量

我们把各粒子视为轻质刚杆上的质点，并假设轴承光滑，因此没有能量损失。如图所示，所有角度均由向上的铅垂线起按顺时针方向量度。于是长度为 a、b、c 的杆与向上的铅垂线所成的角分别是 θ、$\theta + \frac{2}{3}\pi$ 和 $\theta + \frac{4}{3}\pi$，所以各粒子在转轴之上的高度为

$$a\cos\theta, \qquad b\cos\left(\theta + \tfrac{2}{3}\pi\right), \qquad c\cos\left(\theta + \tfrac{4}{3}\pi\right)$$

。



利用 $\cos\frac{2}{3}\pi = \cos\frac{4}{3}\pi = -\frac{1}{2}$、$\sin\frac{2}{3}\pi = \frac{\sqrt{3}}{2}$ 和 $\sin\frac{4}{3}\pi = -\frac{\sqrt{3}}{2}$，可得

$$\cos\left(\theta + \tfrac{2}{3}\pi\right) = -\tfrac{1}{2}\cos\theta - \tfrac{\sqrt{3}}{2}\sin\theta, \qquad \cos\left(\theta + \tfrac{4}{3}\pi\right) = -\tfrac{1}{2}\cos\theta + \tfrac{\sqrt{3}}{2}\sin\theta.$$

取转轴为势能的零点，

$$\begin{aligned} V &= mg\left[a\cos\theta + b\cos\left(\theta + \tfrac{2}{3}\pi\right) + c\cos\left(\theta + \tfrac{4}{3}\pi\right)\right] \\ &= mg\left[\tfrac{1}{2}(2a - b - c)\cos\theta - \tfrac{\sqrt{3}}{2}(b - c)\sin\theta\right]. \end{aligned}$$

每个粒子都以角速度（angular velocity）$\dot\theta$ 绕转轴作圆周运动，所以它们的速率分别是 $a\dot\theta$、$b\dot\theta$、$c\dot\theta$。各杆为轻杆，因此总能量为

$$E = \tfrac{1}{2}m\left(a^2 + b^2 + c^2\right)\dot\theta^2 + \tfrac{1}{2}mg\left[(2a - b - c)\cos\theta - \sqrt{3}\,(b - c)\sin\theta\right].$$

平衡

系统在 $\mathrm{d}V/\mathrm{d}\theta = 0$ 之处处于平衡（equilibrium），即

$$-\tfrac{1}{2}mg\left[(2a - b - c)\sin\theta + \sqrt{3}\,(b - c)\cos\theta\right] = 0.$$

由于 $a > b$ 且 $a > c$，我们有 $2a - b - c > 0$。所以在平衡位置 $\cos\theta \neq 0$：若 $\cos\theta = 0$，则 $\sin\theta = \pm 1$，方括号内的值便是 $\pm(2a - b - c) \neq 0$。两边除以 $\cos\theta$，得

$$\tan\theta = -\frac{(b - c)\sqrt{3}}{2a - b - c},$$

这正是所求。这个论证也可以倒过来：任何满足此方程的 θ 都使 $\mathrm{d}V/\mathrm{d}\theta = 0$。所以平衡位置恰好就是此方程的解。

恰好两个平衡位置，以及哪一个稳定

正切函数的周期为 π，并且在每个有定义而长度为 π 的区间上，每个实数值都恰好取一次。所以方程在 $0 \le \theta < 2\pi$ 内恰好有两个解，两者相差 π。这就是两个平衡位置。

要判断稳定性（stability），可把 V 写成单一余弦。设

$$A = \tfrac{1}{2}(2a - b - c), \qquad B = -\tfrac{\sqrt{3}}{2}(b - c),$$

则 $V = mg(A\cos\theta + B\sin\theta)$。于是

$$\begin{aligned} 4(A^2 + B^2) &= (2a - b - c)^2 + 3(b - c)^2 \\ &= 4a^2 + 4b^2 + 4c^2 - 4ab - 4bc - 4ca \\ &= 2\left[(a - b)^2 + (b - c)^2 + (c - a)^2\right] = 4R^2, \end{aligned}$$

所以 $A^2 + B^2 = R^2$，而由于各长度互不相同，$R > 0$。选取 α 使 $\cos\alpha = A/R$ 及 $\sin\alpha = B/R$。由于 $A > 0$ 且 $B < 0$，故 $-\frac{1}{2}\pi < \alpha < 0$。于是

$$V = mgR\cos(\theta - \alpha).$$

两个平衡位置是 $\theta = \alpha$，该处 $V = mgR$ 为极大值；以及 $\theta = \alpha + \pi$，该处 $V = -mgR$ 为极小值。

> **想一想** 为甚么 V 取极小值就代表稳定?

这里 $V''(\theta) = -mgR\cos(\theta - \alpha)$。在 $\theta = \alpha + \pi$ 时，它等于 $mgR > 0$，所以微小的位移会令势能上升。总能量既然固定，系统便不能走得太远，只会在这个位置附近振动。在 $\theta = \alpha$ 时，$V'' = -mgR < 0$：微小的位移令 V 下降，动能（kinetic energy）增加，系统便会远离该位置。所以稳定平衡位置是 $\theta = \alpha + \pi$，即满足 $\frac{1}{2}\pi < \theta < \pi$ 的那一个，该处 $\cos\theta < 0$ 而 $\sin\theta > 0$。在这个位置，最长的杆 a 指向水平线以下，与向下的铅垂线成角 $|\alpha|$。另一个平衡位置满足 $-\frac{1}{2}\pi < \theta < 0$，杆 a 在水平线以上，是不稳定的。

完整转圈

由于轴承光滑，而重力是保守力（conservative force），所以能量守恒。记 $I = m(a^2 + b^2 + c^2)$，则

$$\tfrac{1}{2}I\dot{\theta}^2 + mgR\cos(\theta - \alpha) = E.$$

要完整地转圈，系统必须越过每一个角度，尤其要越过 $\theta = \alpha$，该处 V 取其最大值 mgR。由于处处都有 $\frac{1}{2}I\dot{\theta}^2 \ge 0$，这就需要 $E \ge mgR$。设 ω 为系统经过稳定位置 $\theta = \alpha + \pi$ 时的角速度，该处 $V = -mgR$。于是 $E = \frac{1}{2}I\omega^2 - mgR$，条件变成

$$\tfrac{1}{2}I\omega^2 - mgR \ge mgR \quad \Longleftrightarrow \quad \omega^2 \ge \frac{4mgR}{I} = \frac{4gR}{a^2 + b^2 + c^2}.$$

所以系统经过稳定平衡位置时，角速度必须至少为 $\sqrt{4gR/(a^2+b^2+c^2)}$，这正是所求。

（在等号成立的临界情况下，系统会缓缓趋近不稳定平衡位置，要经过无限长的时间才到达。严格来说，要完整转一圈，ω 必须严格大于这个值；题目只要求必要的下界。）

验算

- 若 $a=b=c$，则 $R=0$ 且 $V\equiv 0$：三个相等的质量在任何位置都平衡（它们的质心在轴上），而且无论转得多慢都能转过一圈。这与下界趋向零吻合。
- 若 $b=c$，正切条件给出 $\tan\theta=0$。根据对称性，长杆指向正上方（不稳定）或正下方（稳定）。
- 以随机长度作数值检验，可确认 V 的形式、恒等式 $A^2+B^2=R^2$，以及恰好有两个平衡位置，其稳定性亦如上所述。取 $a=3$、$b=2$、$c=0.5$ 及 $g=9.81$，临界角速度为 $2.5406\ \mathrm{rad\,s^{-1}}$。对运动方程逐步作数值计算可见，比这个值低 2% 时系统会摆回来，而高 2% 时则会转过一整圈。

常见错误

- 把位于 $\theta-\frac{2}{3}\pi$ 和 $\theta+\frac{2}{3}\pi$ 的两根杆与图中的位置对调，会令 $\sin\theta$ 项的符号反转，正切条件中的符号亦随之反转。
- 断言“有两个解”，却没有指出 $\cos\theta=0$ 不可能是平衡位置，因而除以 $\cos\theta$ 不会漏掉任何解。
- 把动能写成 $\frac{1}{2}m(a+b+c)^2\dot\theta^2$ 或类似的式子。每个粒子各有自己的速率 $r\dot\theta$，所以转动惯量（moment of inertia）是 $m(a^2+b^2+c^2)$。

STEP 之外

这个系统是一个复摆(compound pendulum),即绕固定轴转动的刚体。它的能量只取决于转动惯量和质心(centre of mass)的位置,而这里质心与轴的距离为 $R/3$。把 $\frac{1}{2}I\dot\theta^2+V(\theta)=E$ 看成 $(\theta,\dot\theta)$ 相平面（phase plane）上的一条曲线，振动在分界线（separatrix）之内，转动在分界线之外，这个图像是大学一年级动力学的核心概念。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学导论）涵盖了刚体转动，以及处理这类问题的能量方法。

详细解答

STEP III 2005，第 12 题

概率与统计 · 无偏估计量的组合

STEP III 2005，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/05-s3-q12

关键思路

整道题都建基于独立（independent）随机变量（random variable）的线性组合（linear combination）的两个事实：期望值（expectation）可以相加，而方差（variance）则要乘上系数的平方后相加。要求某个期望值无论真实的单圈时间是多少都符合要求，就等于要比较系数，于是每个条件都化成一条关于常数的方程。这样，X 就是比赛时间的一个合理估计，而 Y 则利用两种计时方法之间的差异来估计未知的 σ。

解答

设真实的单圈时间为 $\mu_1, \ldots, \mu_4$，真实的比赛时间为 $\mu = \mu_1 + \mu_2 + \mu_3 + \mu_4$。记

$$S = T_1 + T_2 + T_3 + T_4.$$

那么 $\mathrm{E}(S) = \mu_1 + \cdots + \mu_4 = \mu$，而且由于 $T_1, \ldots, T_4$ 互相独立（各计时员互相独立），$\mathrm{Var}(S) = 4\sigma^2$。另外 $\mathrm{E}(T) = \mu$，$\mathrm{Var}(T) = \sigma^2$，而 T 与 S 独立，因为 S 只取决于另外四位计时员。

求 X

对于 $X = aT + bS$，

$$\mathrm{E}(X) = a\mu + b\mu = (a+b)\mu.$$

无论真实的单圈时间是多少，这个值都必须等于 μ，特别是对某个 $\mu \neq 0$ 也要成立；因此

$$a + b = 1.$$

由 T 与 S 的独立性，

$$\operatorname{Var}(X) = a^2 \operatorname{Var}(T) + b^2 \operatorname{Var}(S) = \sigma^2\left(a^2 + 4b^2\right) = \sigma^2\left(a^2 + 4(1-a)^2\right).$$

二次式 $a^2 + 4(1-a)^2 = 5a^2 - 8a + 4$ 的首项系数为正，所以它在导数 $10a - 8$ 等于零之处取得最小值，即 $a = \frac{4}{5}$。（用配方法（completing the square）：$5a^2 - 8a + 4 = 5\left(a - \frac{4}{5}\right)^2 + \frac{4}{5}$。）因此 $a = \frac{4}{5}$，$b = \frac{1}{5}$，而

$$X = \tfrac{4}{5}T + \tfrac{1}{5}(T_1 + T_2 + T_3 + T_4), \qquad \operatorname{Var}(X) = \tfrac{4}{5}\sigma^2.$$

这个结果合情合理：T 的可靠程度是 S 的四倍（方差是 σ^2 对 $4\sigma^2$），所以它得到四倍的权重。

求 Y

对于 $Y = cT + dS$，我们利用 $\mathrm{E}(Y^2) = \operatorname{Var}(Y) + \big(\mathrm{E}(Y)\big)^2$：

$$\mathrm{E}(Y^2) = \sigma^2(c^2 + 4d^2) + (c+d)^2\mu^2.$$

想一想 无论真实的单圈时间是多少，这个式子怎样才能都等于 σ^2？

常数 c、d 不可以依赖 μ，但 μ 可以是任何正数。这个式子是 $\sigma^2(c^2 + 4d^2)$ 加上 μ^2 的某个倍数，所以要它在 μ 变动时仍然等于 σ^2，μ^2 的系数就必须为零：$(c+d)^2 = 0$，所以 $d = -c$。接着我们需要 $\sigma^2(c^2 + 4c^2) = \sigma^2$，即 $c^2 = \frac{1}{5}$。取 $c = 1/\sqrt{5}$，

$$Y = \frac{T - (T_1 + T_2 + T_3 + T_4)}{\sqrt{5}}$$

（取另一个符号会得到 $-Y$，同样可行，因为要紧的只是 Y^2）。所以 Y 量度的是两种计时结果之间的差异；它的期望值是 0，而它的平方是 σ^2 的无偏估计（unbiased estimate）。

这一场比赛

当 $T = 220$ 及 $S = 220.5$ 时，

$$X = \tfrac{4}{5}(220) + \tfrac{1}{5}(220.5) = 176 + 44.1 = 220.1,$$
$$Y = \frac{220 - 220.5}{\sqrt{5}} = -\frac{0.5}{\sqrt{5}}, \qquad Y^2 = \frac{0.25}{5} = 0.05.$$

所以我们估计 $\sigma^2 \approx 0.05$，因而

$\operatorname{Var}(X) = \frac{4}{5}\sigma^2 \approx \frac{4}{5}(0.05) = 0.04$，　　就标准差（standard deviation）而言，$X \approx 0.2$ 秒.

因此我们对真实比赛时间的估计是 220.1 ± 0.2 秒：真实时间大致落在区间

$$[219.9,\ 220.3] \text{ 秒},$$

之内，这里在两边各取一个标准差。如果想要更大的把握，可以在两边各取两个标准差，得到 $[219.7,\ 220.5]$ 秒；若误差服从正态分布（normal distribution），这大约是一个 95% 的区间。要留意 σ^2 只是由 Y^2 的单一个值估计出来的，所以区间的宽度本身也只是一个粗略的估计。

验算

- 当 $a = \frac{4}{5}$ 时，$\operatorname{Var}(X) = \frac{4}{5}\sigma^2$ 比 $\operatorname{Var}(T) = \sigma^2$ 和 $\operatorname{Var}(S) = 4\sigma^2$ 都要小，这正是结合两个独立量度所应有的效果。
- $X = 220.1$ 介乎两个量度值之间，而且远较接近 $T = 220$，即较可靠的那一个。
- 以 200 000 场比赛（$\sigma = 0.7$）进行模拟，所得 X 的平均值与真实时间相差不超过 0.001，$\operatorname{Var}(X) \approx 0.393$，对比 $\frac{4}{5}(0.49) = 0.392$；而 Y^2 的平均值为 0.491，对比 $\sigma^2 = 0.49$。

常见错误

- 把方差取作 $\operatorname{Var}(S) = \sigma^2$ 或 $16\sigma^2$：正确的是 $4\sigma^2$，即四个独立方差之和。
- 忘记了 $\mathrm{E}(Y^2)$ 中的 $(\mathrm{E}\,Y)^2$ 一项，而正是这一项迫使 $d = -c$。
- 把 $Y^2 = 0.05$ 当作 X 的方差，而没有乘上 $\frac{4}{5}$。

STEP 之外

X 是两个估计的最小方差无偏线性组合：一般的法则是以每个独立估计的方差的倒数作为其权重，这是荟萃分析（meta-analysis）中逆方差加权（inverse-variance weighting）的基础，也是最小二乘法（least squares）的高斯–马可夫定理（Gauss–Markov theorem）的基础。Y^2 是 σ^2 的无偏估计量（estimator），就像以 $n-1$ 作除数的样本方差（sample variance）一样。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）仔细讲解了和的期望值与方差，而大学的第一门统计课程则会正式探讨估计量与置信区间（confidence interval）。

详细解答

STEP III 2005，第 13 题

概率与统计 • 几何分布

STEP III 2005，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/05-s3-q13

关键思路

由 $w+1$ 至 n 的牌完全不起作用：它们既不派钱，也不会令游戏结束。因此在第 (i) 部分，我们大可不理会它们，只看抽出 $0, 1, \ldots, w$ 的那些抽牌；每一次这样的抽牌，都是从 $w+1$ 个等可能的数字中公平地抽出一个，于是总奖金就是一个几何随机变量（geometric random variable）。在第 (ii) 部分，各次抽牌不再独立（independent），我们把总奖金拆成若干个指示变量（indicator）之和，每张有奖金的牌对应一个，然后利用期望值的线性性质（linearity of expectation）。

解答

若某次抽牌抽出 $0, 1, \ldots, w$ 其中之一，称它为*有关*的抽牌；若抽出 $w+1, \ldots, n$ 其中之一，则称它为*无关*的抽牌。

第 (i) 部分

为何两个事件的概率相同。无关的抽牌不会改变任何事情：玩家甚么也得不到，游戏继续进行。因此玩家的总奖金只取决于有关抽牌的序列：每次非零的有关抽牌可得 £1，而第一次抽出 0 的有关抽牌令游戏结束。所以，玩家恰好赢得 £3，当且仅当在有关的抽牌之中，首三次抽出非零的数字，而第四次抽出 0。这两者是同一个事件，故概率相同。

（每次抽牌都独立地以概率 $\frac{w+1}{n+1} > 0$ 为有关的抽牌，所以有关的抽牌以概率 1 有无限多次，当中自然有第一、第二、第三和第四次；游戏以概率 1 结束。）

想一想 为甚么不论之前发生过甚么，每次有关的抽牌抽出 $0,1,\ldots,w$ 中任何一个数的可能性都相同?

由于每次抽出的牌都会放回，各次抽牌互相独立，而且每次都在 $\{0,1,\ldots,n\}$ 上均匀分布（uniform）。已知某一次抽牌是有关的抽牌，它就在 $\{0,1,\ldots,w\}$ 上均匀分布，因为这 $w+1$ 个结果本来就是等可能的。各次抽牌互相独立，意味着先后各次有关抽牌的数值也互相独立。所以每次有关的抽牌以概率 $\frac{1}{w+1}$ 抽出 0，以概率 $\frac{w}{w+1}$ 抽出非零的数，并且与其他各次独立。

因此

$$P(\text{恰好赢得£3}) = \left(\frac{w}{w+1}\right)^3 \cdot \frac{1}{w+1} = \frac{w^3}{(w+1)^4},$$

正如所求。

一般的 r。把 3 换成 r，用完全相同的论证（首 r 次有关的抽牌非零，第 $(r+1)$ 次为零），可得

$$P(\text{恰好赢得£}r) = \frac{w^r}{(w+1)^{r+1}}, \qquad r = 0,1,2,\ldots$$

设 $q = \frac{w}{w+1}$，上式就是 $(1-q)q^r$，即几何分布（geometric distribution）；这些概率之和为 1，理应如此。

总奖金的期望值。设 X 为以英镑计的总奖金。当 $|q|<1$ 时，对 $\sum_{r\geq 0} q^r = \frac{1}{1-q}$ 求导，再乘以 q，得 $\sum_{r\geq 0} rq^r = \frac{q}{(1-q)^2}$。所以

$$E(X) = \sum_{r=0}^{\infty} r(1-q)q^r = (1-q)\cdot\frac{q}{(1-q)^2} = \frac{q}{1-q} = \frac{w/(w+1)}{1/(w+1)} = w.$$

总奖金的期望值是 £w。

第 (ii) 部分

现在抽出的牌不再放回，所以这个游戏等同于把整叠牌洗成一个均匀随机的次序，然后逐张翻开，直至 0 出现为止。在这个次序中，牌 $1,\ldots,w$ 之中每有一张排在牌 0 之前，玩家就得到 £1（排在 0 之前的每张牌都会被抽出，排在它之后的则一张也不会）。

对 $k = 1,\ldots,w$，若牌 k 排在牌 0 之前，设 $I_k = 1$，否则设 $I_k = 0$。总奖金为 $Y = I_1 + I_2 + \cdots + I_w$。

想一想 各个 I_k 并不独立。这有没有影响?

没有：不论各变量之间如何相依，和的期望值总是等于期望值之和。所以我们只需求出 $P(I_k = 1)$。在均匀随机的次序中，把牌 k 与牌 0 的位置对调，就是“k 在 0 之前”的次序与“0 在 k 之前”的次序之间的一个双射（bijection），所以两者等可能，$P(I_k = 1) = \frac{1}{2}$。

因此

$$E(Y) = \sum_{k=1}^{w} E(I_k) = \sum_{k=1}^{w} \tfrac{1}{2} = \frac{w}{2},$$

正是第 (i) 部分奖金期望值 w 的一半，正如所求。

验算

- 取 $w = n$（每张非零的牌都有奖金）而且不放回，牌 0 的位置在 $1, \ldots, n+1$ 上均匀分布，所以排在它之前的牌数的平均值是 $\frac{n}{2}$，与 $\frac{w}{2}$ 吻合。
- 放回而 $w = 1$ 时，每次有关的抽牌相当于在 0 与 1 之间掷一枚公平硬币，所以 $P(X = r) = 2^{-(r+1)}$ 而 $E(X) = 1$，与 $w^r/(w+1)^{r+1}$ 及 $E(X) = w$ 吻合。
- 取 $n = 5$、$w = 2$，模拟 2×10^5 局游戏，得 $P(X = 3) \approx 0.0995$，对比 $\frac{8}{81} \approx 0.0988$；平均奖金为 1.995，对比 2。对细小的牌叠（例如 $n = 5$、$w = 3$）穷举所有次序，得出不放回时的平均值为 $\frac{3}{2} = \frac{w}{2}$。

常见错误

- 以 $\frac{1}{n+1}$ 作为某次抽牌抽出 0 的机会。这是每次抽牌的机会，但其间可以出现任意多次无关的抽牌；以有关的抽牌为条件，才得出简洁的 $\frac{1}{w+1}$。
- 遗漏 $r = 0$：第一次有关的抽牌可能就是 0，所以有 $\frac{1}{w+1}$ 的概率一无所获。
- 在第 (ii) 部分试图求出奖金的完整分布。这并无必要；用指示变量和线性性质，两行就可完成。

STEP 之外

不理会无关的抽牌，是*稀疏化*（thinning）的一个例子，即只在有趣的事情发生的时刻观察一个过程；这个想法会在马可夫链（Markov chain）（所谓“跳跃链”（jump chain））和泊松过程（Poisson process）中再次出现。把一个计数写成指示变量之和，再利用期望值的线性性质（即使各指示变量并不独立），是大学一年级概率论与组合数学中最有用的技巧之一。这两个想法在 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门教材）以及 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论导论）中都有详细阐述。

详细解答

STEP III 2005，第 14 题

概率与统计 · 连续随机变量

STEP III 2005，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/05-s3-q14

关键思路

这一题由始至终只靠题目容许我们直接引用的那一个积分，所以功夫全在于看出每个积分需要取哪个 m 和 n。代换（substitution）$x = k^2/u$ 把区间 $[0,k]$ 与 $[k,\infty)$ 对调，而被积函数（integrand）的形状保持不变，这种对称性马上确定了中位数（median）。至于时间 $T = s/V$，同样的倒数技巧说明 T 的概率密度（density）与 V 的属于完全相同的族，只是 k 换成了 s/k。之后，中位数和平均值（mean）都可以从 V 的结果直接读出。

解答

常数 C

在已知结果中取 $m = a$ 和 $n = 2a$。这是容许的，因为 a 是正整数，而且 $2a \geq a$。这时 $n+2 = 2a+2$，$n-m+1 = a+1$，所以

$$\int_0^\infty \frac{x^a}{(x+k)^{2a+2}}\,\mathrm{d}x = \frac{a!\,a!}{(2a+1)!\,k^{a+1}}.$$

密度函数在其取值范围上的积分为 1，所以

$$1 = Ck^{a+1}\cdot\frac{a!\,a!}{(2a+1)!\,k^{a+1}} \qquad\Longrightarrow\qquad C = \frac{(2a+1)!}{a!\,a!},$$

正如所求。

代换与中位数

在左边的积分中，设 $x = k^2/u$，于是 $\mathrm{d}x = -\dfrac{k^2}{u^2}\,\mathrm{d}u$。当 x 由 0 增至 v 时，u 由 ∞ 减至 k^2/v。此外

$$x + k = \frac{k^2}{u} + k = \frac{k(u+k)}{u},$$

所以

$$\frac{x^a}{(x+k)^{2a+2}} = \frac{k^{2a}/u^a}{k^{2a+2}(u+k)^{2a+2}/u^{2a+2}} = \frac{u^{a+2}}{k^2\,(u+k)^{2a+2}}.$$

乘以 $\mathrm{d}x$ 后，k^2/u^2 刚好约去，而负号则把积分上下限对调：

$$\int_0^v \frac{x^a}{(x+k)^{2a+2}}\,\mathrm{d}x = \int_\infty^{k^2/v} \frac{u^a}{(u+k)^{2a+2}}\,(-\mathrm{d}u) = \int_{k^2/v}^\infty \frac{u^a}{(u+k)^{2a+2}}\,\mathrm{d}u,$$

这正是所要证明的恒等式。

想一想 v 取哪个值时，这条恒等式会告诉我们关于中位数的事情?

取 $v = k$，于是 $k^2/v = k$ 也成立。把两边同乘 Ck^{a+1} 后，恒等式变成

$$\begin{aligned}\mathrm{P}(V < k) &= \int_0^k \mathrm{f}(x)\,\mathrm{d}x \\ &= \int_k^\infty \mathrm{f}(u)\,\mathrm{d}u = \mathrm{P}(V > k).\end{aligned}$$

V 是连续（continuous）随机变量，所以 $\mathrm{P}(V = k) = 0$，这两个概率之和为 1。因此两者各为 $\frac{1}{2}$，V 的中位数就是 k。

V 的期望值

期望值（expected value）为

$$\mathrm{E}(V) = Ck^{a+1}\int_0^\infty \frac{x^{a+1}}{(x+k)^{2a+2}}\,\mathrm{d}x.$$

现在取 $m = a+1$ 和 $n = 2a$。这需要 $2a \geq a+1$，而由于 $a \geq 1$，这是成立的。这时 $n - m + 1 = a$，并且

$$\mathrm{E}(V) = \frac{(2a+1)!}{a!\,a!}\,k^{a+1}\cdot\frac{(a+1)!\,(a-1)!}{(2a+1)!\,k^a} = k\cdot\frac{(a+1)!}{a!}\cdot\frac{(a-1)!}{a!} = \frac{(a+1)k}{a}.$$

T 的分布

速率是正数，所以对于 $t > 0$，我们可以把不等式两边同乘 $V/t > 0$ 而不必改变不等号的方向：

$$T < t \iff \frac{s}{V} < t \iff V > \frac{s}{t}.$$

因此

$$\mathrm{P}(T<t)=\mathrm{P}\Big(V>\frac{s}{t}\Big)=\int_{s/t}^{\infty}\frac{Ck^{a+1}x^{a}}{(x+k)^{2a+2}}\,\mathrm{d}x, \qquad (*)$$

正如所求。

现在在 $(*)$ 中作代换 $u=s/x$，于是 $x=s/u$，$\mathrm{d}x=-\dfrac{s}{u^2}\,\mathrm{d}u$。当 x 由 s/t 增至 ∞ 时，u 由 t 减至 0。由于 $x+k=(s+ku)/u$，

$$\mathrm{P}(T<t)=\int_0^t Ck^{a+1}\,\frac{s^a/u^a}{(s+ku)^{2a+2}/u^{2a+2}}\cdot\frac{s}{u^2}\,\mathrm{d}u=\int_0^t\frac{Ck^{a+1}s^{a+1}\,u^a}{(s+ku)^{2a+2}}\,\mathrm{d}u.$$

这对每个 $t>0$ 都成立，而当 $t\le 0$ 时 $\mathrm{P}(T<t)=0$。所以对 t 求导，便得到 T 的密度函数：

$$\mathrm{g}(t)=\frac{Ck^{a+1}s^{a+1}\,t^a}{(s+kt)^{2a+2}}\qquad(t>0).$$

为了与 f 比较，从分母抽出 k^{2a+2}，并记 $K=s/k$：

$$\mathrm{g}(t)=\frac{C\,(s/k)^{a+1}\,t^a}{(t+s/k)^{2a+2}}=\frac{CK^{a+1}t^a}{(t+K)^{2a+2}}.$$

这正正是把 k 换成 $K=s/k$ 之后的密度函数 f。所以 T 与 V 属于同一个分布（distribution）族。

中位数与平均值

由于 T 的分布形式相同，只是参数（parameter）为 $K=s/k$，之前的结果可以直接套用：

$$\text{就中位数而言，}\ T=\frac{s}{k},\qquad \mathrm{E}(T)=\frac{(a+1)}{a}\cdot\frac{s}{k}.$$

因此，时间的中位数与速率的中位数之积为

$$\frac{s}{k}\cdot k=s,$$

这就是距离，与题目所述相符。时间的期望值与速率的期望值之积为

$$\mathrm{E}(T)\,\mathrm{E}(V)=\frac{(a+1)s}{ak}\cdot\frac{(a+1)k}{a}=s\Big(1+\frac{1}{a}\Big)^2>s,$$

因为 a 是正整数，所以 $1+1/a>1$。至此全题完成。

验算

- 当 $a=1$，$k=1$ 时，我们得到 $C=3!/1=6$，而对 $6x/(x+1)^4$ 在 $[0,\infty)$ 上作数值积分（numerical integration）得 1.000000。在 $[0,1]$ 上则得 0.500000，印证了中位数。

- 对若干组数值作数值求积，例如 $a = 2$，$k = 0.7$，得 $\mathrm{E}(V) = 1.05000 = \frac{3}{2}(0.7)$。取 $s = 3.2$ 时，得 $\mathrm{E}(T)\mathrm{E}(V) = 7.20000 = 3.2 \times \frac{9}{4}$。
- 以 $a = 2$，$k = s = 1$ 模拟（simulation）2×10^5 个速率，结果是速率中位数 $\times$ 时间中位数 ≈ 1.000001，但平均速率 ≈ 1.504（理论值：1.5）。
- 中位数之积等于 s 这一点，不用任何公式也看得出来。$T = s/V$ 是 V 的递减函数，所以 T 的中位数就是 s 除以 V 的中位数。

常见错误

- 引用已知积分时忘记检查 $n \geq m$。对 $\mathrm{E}(V)$ 而言，这需要 $a \geq 1$，这正是 a 必须为正整数的原因。对于更高阶的矩（moment），公式最终会失效，事实上 $\mathrm{E}(V^{a+1})$ 是无限大。
- 对 $(*)$ 求导以求 T 的密度函数时，漏掉因子 s/t^2（链式法则，chain rule）。
- 以为 $\mathrm{E}(s/V) = s/\mathrm{E}(V)$。最后一部分的重点，正是要指出这是错的。

STEP 之外

最后的不等式是琴生不等式（Jensen's inequality）的一个例子：当 $v > 0$ 时 $1/v$ 是凸（convex）函数，所以 $\mathrm{E}(1/V) \geq 1/\mathrm{E}(V)$，等号仅在 V 为常数时成立。相比之下，中位数却可以直接穿过任何单调（monotonic）函数。代换 $V/(V+k)$ 把 V 变成一个 $\mathrm{Beta}(a+1, a+1)$ 变量，它关于 $\frac{1}{2}$ 对称，这是理解中位数为何是 k 的另一个途径。Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门）对琴生不等式、变量代换和 Beta 分布族都有十分易读的讲解。

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

2005 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。

