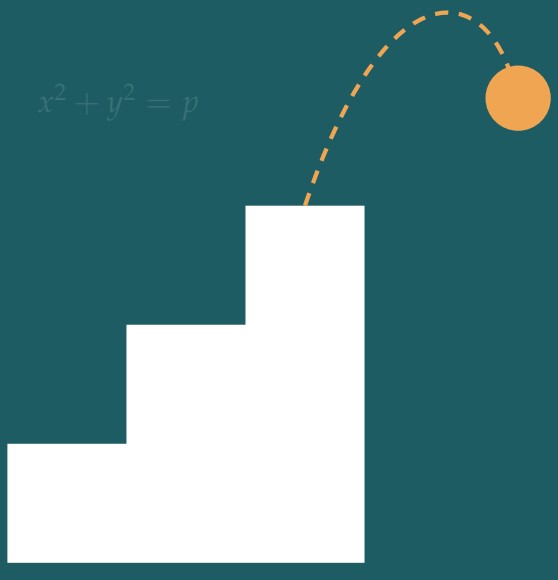
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2007 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

2007 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2007 年详细解答

2007 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 数字排列的计数 纯数学 . . . 3

第 2 题 反正切恒等式 纯数学 . . . 6

第 3 题 三角函数积分 纯数学 . . . 10

第 4 题 二次方程的公共根 纯数学 . . . 13

第 5 题 多面体的几何 纯数学 . . . 16

第 6 题 丢番图方程 纯数学 . . . 20

第 7 题 直线之间的距离 纯数学 . . . 23

第 8 题 三次曲线的相切 纯数学 . . . 27

第 9 题 斜坡上的极限摩擦力 力学 . . . 33

第 10 题 位移–时间图 力学 . . . 37

第 11 题 抛体运动 力学 . . . 41

第 12 题 条件概率 概率与统计 . . . 45

第 13 题 条件概率 概率与统计 . . . 49

第 14 题 道路参数的唯一性 概率与统计 . . . 53

STEP II

第 1 题 二项式近似 纯数学 . . . 58

第 2 题 三次曲线与面积 纯数学 . . . 61

第 3 题 代换积分法 纯数学 . . . 65

第 4 题 三角方程 纯数学 . . . 68

第 5 题 迭代函数 纯数学 . . . 72

第 6 题 一阶微分方程 纯数学 . . . 76

第 7 题 琴生不等式 纯数学 . . . 79

第 8 题 以向量证明塞瓦定理 纯数学 . . . 82

第 9 题 斜碰撞与恢复系数 力学 86
第 10 题 涉及摩擦力的静力学 力学 90
第 11 题 三维空间中的抛体 力学 93
第 12 题 几何分布的最大值 概率与统计 97
第 13 题 生日问题 概率与统计 100
第 14 题 分段概率密度函数 概率与统计 104

STEP III

第 1 题 正切和与韦达定理 纯数学 107
第 2 题 二项式幂级数 纯数学 110
第 3 题 斐波那契恒等式 纯数学 114
第 4 题 参数曲线与曲率 纯数学 117
第 5 题 高阶导数与数学归纳法 纯数学 121
第 6 题 复数与圆 纯数学 125
第 7 题 代换积分法 纯数学 129
第 8 题 里卡蒂方程 纯数学 133
第 9 题 简谐运动 力学 136
第 10 题 斜面上的抛体 力学 140
第 11 题 滑过边缘 力学 144
第 12 题 随机和 概率与统计 148
第 13 题 随机漫步的期望值 概率与统计 151
第 14 题 几何概率 概率与统计 155

详细解答

STEP I 2007，第 1 题

纯数学 · 数字排列的计数

STEP I 2007，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/07-s1-q1

关键思路

不要尝试把平衡数逐一列出。相反，应按两组数字之和的共同值 s 把它们分类。对于每个 s，前面一对数字和后面一对数字可以独立地选取，因此数字和为该值的平衡数的数目，就是两个计数的乘积。于是整个问题归结为计算把 s 写成两个数字之和的方法数目，而这等同于计算正方形网格某条对角线上的格点（lattice point）数目。

解答

把一个四位平衡数写成 $abcd$，其中每个数字都属于 $\{0,1,\ldots,k\}$，$a \geq 1$（首位数字不能是 0），并且

$$a+b=c+d=s$$

其中 s 满足 $0 \leq s \leq 2k$。对于每个 s，设

- $N(s)$ 为满足 $c+d=s$ 的数字有序对（ordered pair）(c,d) 的数目；
- $M(s)$ 为满足 $a \geq 1$ 及 $a+b=s$ 的数字有序对 (a,b) 的数目。

一旦固定了 s，任何一个合乎条件的前数对都可以配上任何一个合乎条件的后数对，所以数字和为 s 的平衡数共有 $M(s)N(s)$ 个。不同的 s 值给出不同的数，因此

$$\text{总数} = \sum_{s=0}^{2k} M(s)\,N(s).$$

计算数对的数目

$N(s)$。若 $0 \le s \le k$，则 c 可以是 $0, 1, \ldots, s$ 中的任何一个，而 $d = s - c$ 自动是一个数字。所以 $N(s) = s + 1$。若 $k < s \le 2k$，则 $d = s - c \le k$ 迫使 $c \ge s - k$，所以 c 由 $s - k$ 取到 k，而 $N(s) = 2k + 1 - s$。

$M(s)$。这个计数与上面相同，只是要减去 $a = 0$ 的那一对（如果它存在的话）。那一对是 $(0, s)$，而它只在 $s \le k$ 时才是容许的。所以

$$M(s) = \begin{cases} s, & 0 \le s \le k, \\ 2k + 1 - s, & k < s \le 2k. \end{cases}$$

特别地，$M(0) = 0$：没有平衡数的数字和是 0，这是对的，因为 $a \ge 1$。

> **想一想** 为甚么我们不能只计算所有满足 $a + b = c + d$ 的四位数字串，然后就此作结?

因为像 0312 这样的数字串并不是四位数。条件 $a \ge 1$ 只影响前数对，而且只影响较小的和，这正正就是 $M(s)$ 与 $N(s)$ 只在 $s \le k$ 时才有分别的原因。

第 (i) 部分

这里 $k = 4$，而和的取值由 0 到 8：

s	0	1	2	3	4	5	6	7	8
$M(s)$	0	1	2	3	4	4	3	2	1
$N(s)$	1	2	3	4	5	4	3	2	1
$M(s)N(s)$	0	2	6	12	20	16	9	4	1

最底一行加起来是 $(2 + 6 + 12 + 20) + (16 + 9 + 4 + 1) = 40 + 30 = 70$。所以用数字 0 至 4 可以组成七十个四位平衡数，正如题目所求。

第 (ii) 部分

现在利用 M 和 N 的公式，在 $s = k$ 处把和式拆开：

$$\text{总数} = \sum_{s=1}^{k} s(s+1) \; + \; \sum_{s=k+1}^{2k} (2k + 1 - s)^2.$$

（$s=0$ 的一项是零。）在第二个和式中设 $r=2k+1-s$。当 s 由 $k+1$ 取到 $2k$ 时，r 由 k 递减至 1，所以第二个和式就是 $\sum_{r=1}^{k} r^2$。因此

$$\begin{aligned}\text{总数} &= \sum_{r=1}^{k}(r^2+r)+\sum_{r=1}^{k} r^2 = 2\sum_{r=1}^{k} r^2+\sum_{r=1}^{k} r\\ &= \tfrac{1}{3}k(k+1)(2k+1)+\tfrac{1}{2}k(k+1)\\ &= \tfrac{1}{6}k(k+1)\big(2(2k+1)+3\big)\\ &= \tfrac{1}{6}k(k+1)(4k+5),\end{aligned}$$

其中用到题目给出的 $\sum r^2$ 恒等式（$r=0$ 的一项没有贡献）以及 $\sum_{r=1}^{k} r=\frac{1}{2}k(k+1)$。这就是所要求的结果。

验算

- 把 $k=4$ 代入公式，得 $\frac{1}{6}\cdot 4\cdot 5\cdot 21=70$，与第 (i) 部分吻合。
- 当 $k=1$ 时，公式给出 3。直接列举，只用 0 和 1 的四位平衡数是 1001、1010 和 1111。
- 对所有满足 $a\geq 1$ 的数字串 *abcd* 作穷举（brute-force）计数，结果在 k 由 1 到 15 的每个情况下都与公式相符（例如 $k=9$，即用上全部十进制数字时，结果是 615）。

常见错误

- 忘记首位数字的条件，结果得出的是 $\sum N(s)^2=\frac{1}{3}(k+1)(2k^2+4k+3)$，当 $k=4$ 时这是 85。
- 在 $s>k$ 时算错数对的数目：它是 $2k+1-s$，而不是 $s+1$，因为两个数字都不能超过 k。
- 把中间的和 $s=k$ 在拆开后的两半中都计算了，或者两半都没有计算。

STEP 之外

计数 $N(s)$ 是 $(1+x+\cdots+x^k)^2$ 中 x^s 的系数，而平衡数的总数则是若干个这类多项式之积的某个系数。这种*生成函数*（generating function）方法同样能处理 $2n$ 位的平衡数，而在那种情况下直接计数很快便会变得难以应付。它是大学本科组合数学（combinatorics）的核心工具。想做更多同类的计数问题，可试试 Siklos 的 *Advanced Problems in Mathematics*（进阶数学问题集）。

详细解答

STEP I 2007，第 2 题

纯数学 · 反正切恒等式

STEP I 2007，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/07-s1-q2

关键思路

只要对等式两边取正切，再运用 $\tan(A+B)$ 的加法公式（addition formula），一条关于反正切（inverse tangent）的等式就会变成一条关于普通分数的等式。代价是 tan 每隔 π 便重复一次，所以 $\tan(A+B)=1$ 只告诉我们 $A+B=\frac{1}{4}\pi+k\pi$；我们还需要交代一下各个角的大小，才能确定 k。在此之后，每一部分都是一小段数论（number theory）：作因式分解（factorise），或者运用整除性（divisibility）论证。

解答

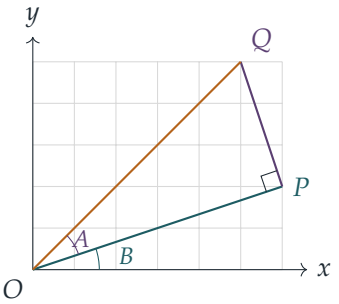
第 (i) 部分

由于 $\tan A=\frac{1}{2}$ 及 $\tan B=\frac{1}{3}$，

$$\tan(A+B)=\frac{\tan A+\tan B}{1-\tan A\tan B}=\frac{\frac{1}{2}+\frac{1}{3}}{1-\frac{1}{6}}=\frac{\frac{5}{6}}{\frac{5}{6}}=1.$$

现在 A 和 B 都是锐角（acute），所以 $0<A+B<\pi$，而在 $(0,\pi)$ 内正切为 1 的角只有 $\frac{1}{4}\pi$。因此 $A+B=\frac{1}{4}\pi$，正如所求。

下图在方格纸上展示同一事实。取 $O=(0,0)$、$P=(6,2)$ 及 $Q=(5,5)$，直线 OP 与 x 轴的夹角为 $\arctan\frac{1}{3}$；$PQ=(-1,3)$ 垂直于 $OP=(6,2)$，而且长度是它的一半，所以角 POQ 是 $\arctan\frac{1}{2}$；而 Q 位于直线 $y=x$ 上，即在 $\frac{1}{4}\pi$ 的方向。



第 (i) 部分：整数 p 和 q

首先排除 $pq = 1$ 的情况,因为这时加法公式会出现除以零。此时 $p = q = 1$ 或 $p = q = -1$，左边分别是 $\frac{1}{2}\pi$ 和 $-\frac{1}{2}\pi$，两者都不是 $\frac{1}{4}\pi$。所以 $pq \neq 1$，对题目所给等式的两边取正切，得

$$\frac{\frac{1}{p} + \frac{1}{q}}{1 - \frac{1}{pq}} = 1.$$

分子和分母同乘以 pq，得 $\dfrac{p+q}{pq-1} = 1$，所以

$$p + q = pq - 1 \quad \Longrightarrow \quad pq - p - q + 1 = 2 \quad \Longrightarrow \quad (p-1)(q-1) = 2,$$

这就是所要求的结果。

由于 $p-1$ 和 $q-1$ 是乘积为 2 的整数，数对 $(p-1, q-1)$ 必为以下其中之一：

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

由此得 $(p,q) = (2,3)$、$(3,2)$、$(0,-1)$ 或 $(-1,0)$。最后两个要舍去，因为 p 和 q 都不是零。所以 $\{p,q\} = \{2,3\}$，而根据第 (i) 部分的前半，两种次序确实都满足该等式。

? **想一想** 我们只证明了解必须满足 $(p-1)(q-1) = 2$。为甚么事后还要检查这些候选答案?

取正切会失去信息：$(p-1)(q-1) = 2$ 等价于 $\tan(\arctan\frac{1}{p} + \arctan\frac{1}{q}) = 1$，而后者也容许和为 $-\frac{3}{4}\pi$ 之类的值。这里的候选数对要么含有 0（于是 $\frac{1}{p}$ 没有意义），要么是 2 和 3，而我们已知这一对是成立的。

第 (ii) 部分

设 $\theta = \arctan\frac{1}{r}$ 及 $\phi = \arctan\dfrac{s}{s+t}$。对题目所给的等式取正切，

$$\frac{\frac{1}{r} + \frac{s}{s+t}}{1 - \frac{s}{r(s+t)}} = 1.$$

分母不为零：由于 $r \geq 1$ 及 $t \geq 1$，我们有 $r(s+t) > s$。分子和分母同乘以 $r(s+t)$，

$$\frac{(s+t)+rs}{r(s+t)-s} = 1 \quad \Longrightarrow \quad s+t+rs = rs+rt-s \quad \Longrightarrow \quad t(r-1) = 2s.$$

想一想 关于 s 和 t 的最大公因数（highest common factor）的条件，在哪里派上用场？

等式 $t(r-1) = 2s$ 表示 t 整除 $2s$。由于 $\text{hcf}(s,t) = 1$，t 与 s 没有公共的质因数（prime factor），所以 t 的每个质因数都必须来自那个 2：即 t 整除 2。由于 t 是正整数，可能的值只有

$$t = 1 \quad \text{或} \quad t = 2.$$

- 若 $t = 1$：$r - 1 = 2s$，所以 $r = 2s + 1$。
- 若 $t = 2$：$2(r-1) = 2s$，所以 $r = s + 1$。（这里 s 必须是奇数，因为 $\text{hcf}(s,2) = 1$。）

两种情况都确实会出现。事实上 $0 < \frac{1}{r} \leq 1$ 及 $0 < \frac{s}{s+t} < 1$，所以 θ 和 ϕ 都落在 $(0, \frac{1}{4}\pi]$ 内，并且 $0 < \theta + \phi < \frac{1}{2}\pi$。把上面的代数运算倒推回去，$t(r-1) = 2s$ 给出 $\tan(\theta+\phi) = 1$，于是 $\theta + \phi = \frac{1}{4}\pi$ 准确成立。例如，$t = 1$, $s = 1$ 给出 $\arctan\frac{1}{3} + \arctan\frac{1}{2}$，而 $t = 2$, $s = 1$ 给出 $\arctan\frac{1}{2} + \arctan\frac{1}{3}$：正是第 (i) 部分的恒等式，两种次序各出现一次。

验算

- 用计算机计算，$\arctan 0.5 + \arctan(1/3) = 0.4636 + 0.3218 = 0.7854 = \frac{1}{4}\pi$。
- 用电脑搜索所有满足 $|p|, |q| \leq 200$ 的非零 p, q，只找到 $(2,3)$ 和 $(3,2)$。
- 在 $r < 400$ 及 $s, t < 150$ 且 $\text{hcf}(s,t) = 1$ 的范围内搜索，只找到 $t = 1$, $r = 2s+1$ 和 $t = 2$, $r = s+1$；例如 $\arctan\frac{1}{8} + \arctan\frac{7}{9} = \frac{1}{4}\pi$（$s = 7$, $t = 2$），在机器精度内成立。

常见错误

- 由 $\tan(A+B) = 1$ 直接推断 $A + B = \frac{1}{4}\pi$，却没有说明为何 $A + B$ 不可能是 $\frac{5}{4}\pi$。
- 忘记了 2 的负因数分解 $(-1)(-2)$ 和 $(-2)(-1)$，或者忘记因为它们会给出零而要把它们舍去。
- 写“t 整除 $2s$，所以 t 整除 2”而没有引用 $\text{hcf}(s,t) = 1$：这正是需要用到该假设的地方。

STEP 之外

tan 的加法公式其实就是复数（complex numbers）的乘法：$\arctan\frac{1}{p}$ 是 $p+\mathrm{i}$ 的辐角（argument），所以第 (i) 部分所说的，就是 $(2+\mathrm{i})(3+\mathrm{i}) = 5+5\mathrm{i}$ 的辐角为 $\frac{1}{4}\pi$。像梅钦（Machin）公式 $4\arctan\frac{1}{5} - \arctan\frac{1}{239} = \frac{1}{4}\pi$ 这类恒等式，曾被用来把 π 计算到数百位，它们正是以完全相同的方式，由高斯整数（Gaussian integers）的因数分解得来。Needham 的 *Visual Complex Analysis* 把复数乘法的几何阐述得十分精彩，而 Davenport 的 *The Higher Arithmetic* 则涵盖了第 (ii) 部分所用的整除性论证。

详细解答

STEP I 2007，第 3 题

纯数学 · 三角函数积分

STEP I 2007，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/07-s1-q3

关键思路

$\cos^4\theta$ 和 $\sin^4\theta$ 单独来看都不易积分，但只要利用 $\cos^2\theta+\sin^2\theta=1$ 把两者的*差*与*和*化为 2θ 的函数，这两个式子就都容易积分（integrate）了。把和与差分别积分，便得到关于所求两个积分的两条简单线性方程。同样的技巧再升一级，就能处理六次方的情况。

解答

恒等式

把平方差（difference of two squares）因式分解，

$$\cos^4\theta-\sin^4\theta=(\cos^2\theta-\sin^2\theta)(\cos^2\theta+\sin^2\theta)=\cos 2\theta\cdot 1=\cos 2\theta.$$

至于两者之和，我们用配方法（completing the square）：

$$\begin{aligned}\cos^4\theta+\sin^4\theta&=(\cos^2\theta+\sin^2\theta)^2-2\sin^2\theta\cos^2\theta\\&=1-\tfrac{1}{2}(2\sin\theta\cos\theta)^2\\&=1-\tfrac{1}{2}\sin^2 2\theta,\end{aligned}$$

其中用到 $\sin 2\theta=2\sin\theta\cos\theta$。两个恒等式（identity）均已得证。

四次方

记 $C_4=\int_0^{\pi/2}\cos^4\theta\,\mathrm{d}\theta$ 及 $S_4=\int_0^{\pi/2}\sin^4\theta\,\mathrm{d}\theta$。把第一个恒等式积分，

$$C_4-S_4=\int_0^{\pi/2}\cos 2\theta\,\mathrm{d}\theta=\left[\tfrac{1}{2}\sin 2\theta\right]_0^{\pi/2}=0.$$

对于第二个恒等式，我们需要 $\int_0^{\pi/2} \sin^2 2\theta \,\mathrm{d}\theta$。利用 $\sin^2 2\theta = \frac{1}{2}(1-\cos 4\theta)$，

$$\int_0^{\pi/2} \sin^2 2\theta \,\mathrm{d}\theta = \left[\tfrac{1}{2}\theta - \tfrac{1}{8}\sin 4\theta\right]_0^{\pi/2} = \frac{\pi}{4},$$

因为 $\sin 2\pi = 0$。因此

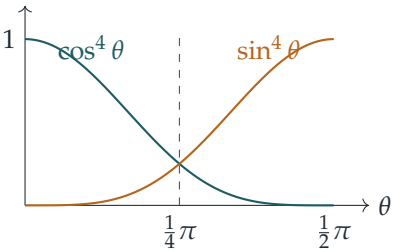
$$C_4 + S_4 = \int_0^{\pi/2} \left(1 - \tfrac{1}{2}\sin^2 2\theta\right)\mathrm{d}\theta = \frac{\pi}{2} - \frac{1}{2}\cdot\frac{\pi}{4} = \frac{3\pi}{8}.$$

把两条方程相加和相减，

$$\int_0^{\pi/2} \cos^4\theta \,\mathrm{d}\theta = \int_0^{\pi/2} \sin^4\theta \,\mathrm{d}\theta = \frac{3\pi}{16}.$$

想一想 我们能否不经任何计算就看出 $C_4 = S_4$?

可以。代换（substitution）$\phi = \frac{1}{2}\pi - \theta$ 把 $\cos\theta$ 变成 $\sin\phi$，并把上下限 0 和 $\frac{1}{2}\pi$ 对调，所以对任何 n 都有 $\int_0^{\pi/2} \cos^n\theta \,\mathrm{d}\theta = \int_0^{\pi/2} \sin^n\phi \,\mathrm{d}\phi$。$\cos^4\theta$ 与 $\sin^4\theta$ 的图像关于直线 $\theta = \frac{1}{4}\pi$ 互为镜像，因此两者所围的面积相等：



所以第一个恒等式其实只是为了求和才用得上；差的积分等于 0，不过是这种对称性（symmetry）的另一种面貌。

六次方

我们重复同一方法。设 $c = \cos\theta$、$s = \sin\theta$，并有 $c^2 + s^2 = 1$：

$$\begin{aligned} c^6 - s^6 &= (c^2 - s^2)(c^4 + c^2 s^2 + s^4) = \cos 2\theta\left(1 - \tfrac{1}{2}\sin^2 2\theta + \tfrac{1}{4}\sin^2 2\theta\right) \\ &= \cos 2\theta\left(1 - \tfrac{1}{4}\sin^2 2\theta\right), \end{aligned}$$

其中用到已证明的恒等式以及 $c^2 s^2 = \frac{1}{4}\sin^2 2\theta$。至于两者之和，我们利用 $c^2 + s^2$ 的立方：

$$c^6 + s^6 = (c^2 + s^2)^3 - 3c^2 s^2 (c^2 + s^2) = 1 - \tfrac{3}{4}\sin^2 2\theta.$$

以 C_6 和 S_6 表示这两个积分。由于 $\frac{\mathrm{d}}{\mathrm{d}\theta}\sin^3 2\theta = 6\sin^2 2\theta\cos 2\theta$，

$$C_6 - S_6 = \left[\tfrac{1}{2}\sin 2\theta - \tfrac{1}{24}\sin^3 2\theta\right]_0^{\pi/2} = 0,$$

这与对称性论证所预料的结果一致。再次利用 $\int_0^{\pi/2} \sin^2 2\theta \,\mathrm{d}\theta = \frac{1}{4}\pi$，

$$C_6 + S_6 = \frac{\pi}{2} - \frac{3}{4}\cdot\frac{\pi}{4} = \frac{5\pi}{16}.$$

因此

$$\int_0^{\pi/2} \cos^6\theta \,\mathrm{d}\theta = \int_0^{\pi/2} \sin^6\theta \,\mathrm{d}\theta = \frac{5\pi}{32}.$$

验算

- 幂次愈高，答案愈小：$n = 0$ 时为 $\frac{1}{2}\pi$，$n = 2$ 时为 $\frac{1}{4}\pi$，然后是 $\frac{3}{16}\pi$ 和 $\frac{5}{32}\pi$。这是必然的，因为在该区间上 $0 \le \cos\theta \le 1$。
- 数值上，用辛普森法则（Simpson's rule）可得 $\int_0^{\pi/2} \cos^4\theta \,\mathrm{d}\theta \approx 0.589049$ 及 $\int_0^{\pi/2} \cos^6\theta \,\mathrm{d}\theta \approx 0.490874$，与 $3\pi/16$ 和 $5\pi/32$ 吻合；正弦的积分在同样的准确度下亦与这些数值一致。
- 把 $\theta = \frac{1}{4}\pi$ 代入两个关于和的恒等式，得 $\frac{1}{4} + \frac{1}{4} = 1 - \frac{1}{2}$ 及 $\frac{1}{8} + \frac{1}{8} = 1 - \frac{3}{4}$，两者皆成立。

常见错误

- 仿照 $\cos^2\theta + \sin^2\theta = 1$ 而写出 $\cos^4\theta + \sin^4\theta = 1$：这样就漏掉了交叉项 $2\sin^2\theta\cos^2\theta$。
- 对 $\cos 2\theta$ 或 $\cos 4\theta$ 积分时，忘记由链式法则（chain rule）而来的因子 $\frac{1}{2}$；在本题中，这对 $\cos 2\theta$ 没有影响（反正积分是 0），但对 $\sin^2 2\theta$ 就有影响。

STEP 之外

这些都是沃利斯（Wallis）积分 $I_n = \int_0^{\pi/2} \sin^n\theta \,\mathrm{d}\theta$ 的特例，它们满足递推公式（reduction formula）$I_n = \frac{n-1}{n} I_{n-2}$（用分部积分法（integration by parts）即可得到）。你可以验证 $I_4 = \frac{3}{4}\cdot\frac{1}{2}\cdot\frac{\pi}{2}$ 和 $I_6 = \frac{5}{6} I_4$ 正好给出上面的答案。比较 I_{2n} 与 I_{2n+1}，便可推导出沃利斯著名的 π 的无穷乘积（infinite product）。Siklos 的 *Advanced Problems in Mathematics* 有更多 STEP 风格的递推公式练习，而 Bartle 与 Sherbert 的 *Introduction to Real Analysis* 则把当中的极限论证严格化。

详细解答

STEP I 2007，第 4 题

纯数学 · 二次方程的公共根

STEP I 2007，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/07-s1-q4

关键思路

第一部分是 $x^3+y^3+z^3-3xyz$ 的经典因式分解（factorisation）：一旦知道 $x+b+c$ 是因式，比较系数便可得出那个二次式。第二部分有两个二次方程共用一个根。标准的做法是*消去*（eliminate）：把两条方程组合起来，使 k 的其中一个幂消去。用两种不同的方法这样做，便可把 k 和 k^2 写成比的形式，然后由 $k^2=(k)^2$ 得出一个只涉及系数的关系式。这个关系式原来就是第一部分取 $x=a$ 的情况。

解答

因式分解

把 $x=-(b+c)$ 代入 $P(x)=x^3-3xbc+b^3+c^3$：

$$\begin{aligned}P(-b-c)&=-(b+c)^3+3bc(b+c)+b^3+c^3\\&=-(b^3+3b^2c+3bc^2+c^3)+3b^2c+3bc^2+b^3+c^3=0,\end{aligned}$$

所以根据因式定理（factor theorem），$x+b+c$ 是因式。比较系数（P 的 x^2 项为零，常数项为 b^3+c^3）可得

$$x^3-3xbc+b^3+c^3=(x+b+c)\,\mathrm{Q}(x),\qquad \mathrm{Q}(x)=x^2+b^2+c^2-xb-bc-cx.$$

把右边展开即可确认：x^2 项是 $x^2(b+c)-x^2(b+c)=0$，x 项合并后是 $-3bcx$，而常数项是 b^3+c^3。

至于平方和，x^2、b^2、c^2 各在 $2\mathrm{Q}(x)$ 中出现两次，而每个交叉项则以 $-2xb$ 等形式出现，所以

$$2\mathrm{Q}(x)=(x-b)^2+(b-c)^2+(c-x)^2.$$

公共根

由于 k 是两条方程的根，

$$ak^2 + bk + c = 0, \tag{1}$$
$$bk^2 + ck + a = 0. \tag{2}$$

要消去 k^2，取 $b \times (1) - a \times (2)$：

$$(b^2 - ac)k + bc - a^2 = 0, \qquad \text{所以} \qquad (ac - b^2)\,k = bc - a^2.$$

若改为消去 k 项，则取 $c \times (1) - b \times (2)$：

$$(ac - b^2)k^2 + c^2 - ab = 0, \qquad \text{所以} \qquad (ac - b^2)\,k^2 = ab - c^2.$$

这就是涉及 k^2 的类似式子。

现在利用 $k^2 = k \cdot k$。由于 $ac \neq b^2$，我们可以除以 $ac - b^2$，得 $k = \dfrac{bc - a^2}{ac - b^2}$，于是

$$(ac - b^2)\,\frac{(bc - a^2)^2}{(ac - b^2)^2} = ab - c^2 \quad \Longrightarrow \quad (ac - b^2)(ab - c^2) = (bc - a^2)^2,$$

其中最后一步是把全式乘以 $ac - b^2$。这就是所求的结果。

把两边展开，

$$a^2bc - ac^3 - ab^3 + b^2c^2 = b^2c^2 - 2a^2bc + a^4,$$
$$\text{所以} \quad 0 = a^4 + ab^3 + ac^3 - 3a^2bc = a\left(a^3 - 3abc + b^3 + c^3\right).$$

由于 $a \neq 0$，我们得到 $a^3 - 3abc + b^3 + c^3 = 0$，正如所求。

想一想 我们在哪里见过 $a^3 - 3abc + b^3 + c^3$?

它就是第一部分取 $x = a$ 的情况。所以

$$(a + b + c)\,\mathrm{Q}(a) = 0, \qquad 2\mathrm{Q}(a) = (a - b)^2 + (b - c)^2 + (c - a)^2.$$

因此，要么 $a + b + c = 0$，要么 $\mathrm{Q}(a) = 0$。

情况 $\mathrm{Q}(a) = 0$。因为 a、b、c 是实数，三个平方都是非负的，而它们的和为零，只可能是每一个都为零。所以 $a = b = c$，两条方程都是 $a(y^2 + y + 1) = 0$：它们完全相同。

情况 $a + b + c = 0$。把 $c = -a - b$ 代入 k 的公式：

$$bc - a^2 = -ab - b^2 - a^2, \qquad ac - b^2 = -a^2 - ab - b^2,$$

所以 $bc - a^2 = ac - b^2$，而由于这个值不为零，$k = 1$。（事实上 $y = 1$ 满足两条方程，因为这时两条方程都变成 $a + b + c = 0$。）

所以要么 $k=1$，要么两条方程完全相同，全题至此完成。

留意在方程相同的情况下，$a=b=c$ 会得出 $ac=b^2$，而这已被题目的假设排除。所以在题设条件下，公共根其实必定是 $k=1$；“相同”这个可能性，是把条件 $ac\neq b^2$ 拿走之后才会出现的，这时公共根是复数的单位立方根（cube roots of unity）。

验算

取 $a=1$、$b=2$、$c=-3$，则 $a+b+c=0$。两条方程是 $y^2+2y-3=(y-1)(y+3)=0$ 和 $2y^2-3y+1=(2y-1)(y-1)=0$，它们共用 $k=1$。这里 $ac-b^2=-7$、$bc-a^2=-7$ 及 $ab-c^2=-7$，所以 $(ac-b^2)k=bc-a^2$、$(ac-b^2)k^2=ab-c^2$ 和 $(-7)(-7)=(-7)^2$ 全部成立，而且 $a^3-3abc+b^3+c^3=1+18+8-27=0$。我们亦以精确算术，用一千组随机的三元数组检验过第一部分的两个恒等式，以及展开式 $(ac-b^2)(ab-c^2)-(bc-a^2)^2=-a(a^3-3abc+b^3+c^3)$。

常见错误

- 消去常数项（$a\times(1)-c\times(2)$）会得出 $k\big((a^2-bc)k+ab-c^2\big)=0$，还需要 $k\neq 0$ 才用得上。$c\times(1)-b\times(2)$ 这个组合则可避免这个问题。
- 除以 $ac-b^2$ 时，没有说明题目已保证它不为零。
- 由 $\mathrm{Q}(a)=0$ 推出 $a=b=c$ 时，没有用到 a、b、c 是实数：平方和的论证对复数并不成立。

STEP 之外

两个多项式有公共根的条件，一般由结式（resultant）给出，那是一个由它们的系数构成的行列式（determinant）；关系式 $(ac-b^2)(ab-c^2)=(bc-a^2)^2$ 正正就是令这两个二次式的结式等于零。$a^3+b^3+c^3-3abc$ 这个式子也是一个循环矩阵（circulant matrix）的行列式，而且它可以完全分解为 $(a+b+c)(a+\omega b+\omega^2 c)(a+\omega^2 b+\omega c)$，其中 ω 是一个复数的单位立方根。想进一步了解多项式及其根，可参阅 Allenby 的 *Rings, Fields and Groups*（环、域与群）。

详细解答

STEP I 2007，第 5 题

纯数学 • 多面体的几何

STEP I 2007，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/07-s1-q5

关键思路

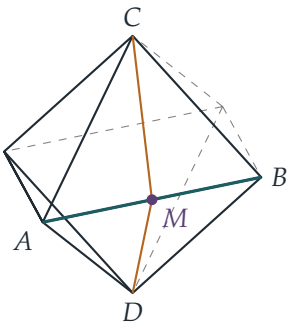
正八面体（regular octahedron）可以非常整齐地放进坐标系：它的六个顶点可取为各坐标轴正、负方向上距离原点一个单位的点。用了这个模型，两部分都只是向量（vector）的运算。两个面之间的夹角，是在垂直于它们公共棱的平面上量度的，所以我们取两条直线：它们都由公共棱的中点出发，分别横越两个面，到达对面的顶点。至于各面的中心，只是三个顶点的平均。

解答

任何两个正八面体都只相差一个缩放比例，而角度和体积之比在缩放之下都不会改变。因此我们不妨考虑顶点为

$$(\pm 1, 0, 0), \qquad (0, \pm 1, 0), \qquad (0, 0, \pm 1).$$

的正八面体。这些点之中，任何两个不是相对的点都相距 $\sqrt{2}$，由此得出 12 条棱；而各个面就是每条坐标轴上各取一个顶点所成的 8 个三角形，每一种正负号的选择对应一个面。每个面都是边长为 $\sqrt{2}$ 的等边三角形，所以这确实是一个正八面体。



第 (i) 部分

取两个有公共棱的面；根据对称性，每条棱的情况都一样，所以只考虑其中一条便足够。设 $A=(1,0,0)$、$B=(0,1,0)$、$C=(0,0,1)$ 及 $D=(0,0,-1)$，则面 ABC 与面 ABD 相交于 AB。设 $M=(\frac{1}{2},\frac{1}{2},0)$ 为 AB 的中点。

? **想一想** 两个平面相交于一条直线，它们之间的夹角应怎样量度?

我们在公共直线上取一点，然后在每个平面上，作通过该点并垂直于公共直线的直线；这两条直线之间的夹角，就是两个面之间的夹角。由于每个面都是等边三角形，三角形 ABC 的中线（median）MC 垂直于 AB，三角形 ABD 的中线 MD 亦然。所以我们要求的角就是 $\angle CMD$。

现在

$$\overrightarrow{MC}=\left(-\tfrac{1}{2},-\tfrac{1}{2},1\right),\qquad \overrightarrow{MD}=\left(-\tfrac{1}{2},-\tfrac{1}{2},-1\right),$$

而两者都垂直于 $\overrightarrow{AB}=(-1,1,0)$，正如所料。两个向量长度的平方都是 $\frac{1}{4}+\frac{1}{4}+1=\frac{3}{2}$，而

$$\overrightarrow{MC}\cdot\overrightarrow{MD}=\tfrac{1}{4}+\tfrac{1}{4}-1=-\tfrac{1}{2}.$$

因此

$$\cos\angle CMD=\frac{-\frac{1}{2}}{\frac{3}{2}}=-\frac{1}{3},$$

所以两个面之间的夹角是 $\arccos\left(-\frac{1}{3}\right)$，约为 $109.5°$，正是所求。余弦值是负数，所以这个角是钝角：正八面体沿棱的地方不及正方体那么尖锐。

若不用坐标，同一个计算就是在三角形 CMD 中运用余弦定理（cosine rule）。设棱长为 a，则中线 $MC=MD=\frac{\sqrt{3}}{2}a$，而 C、D 是相对的顶点：A、C、$(-1,0,0)$、D 四点构成一个边长为 a 的正方形，所以 $CD=\sqrt{2}\,a$。于是

$$2a^2=\tfrac{3}{4}a^2+\tfrac{3}{4}a^2-2\cdot\tfrac{3}{4}a^2\cos\angle CMD\quad\Longrightarrow\quad\cos\angle CMD=-\tfrac{1}{3}.$$

第 (ii) 部分

一个面的中心是它三个顶点的平均。每个面都有一个顶点是 $(\pm1,0,0)$，一个是 $(0,\pm1,0)$，一个是 $(0,0,\pm1)$，所以它的中心是

$$\left(\pm\tfrac{1}{3},\pm\tfrac{1}{3},\pm\tfrac{1}{3}\right),$$

八种正负号的选择给出八个中心。这些点正好是一个边长为 $\frac{2}{3}$ 的正方体的顶点，所以

$$V_{\text{cube}} = \left(\tfrac{2}{3}\right)^3 = \tfrac{8}{27}.$$

正八面体是由两个以正方形为底的棱锥（pyramid），沿着顶点为 $(\pm1,0,0),(0,\pm1,0)$ 的正方形黏合而成。这个正方形的对角线长度为 2，所以它的面积是 $\frac{1}{2}\cdot 2\cdot 2 = 2$，而每个棱锥的高是 1。因此

$$V_{\text{oct}} = 2\times\tfrac{1}{3}\times 2\times 1 = \tfrac{4}{3}.$$

两者之比为

$$\frac{V_{\text{oct}}}{V_{\text{cube}}} = \frac{4/3}{8/27} = \frac{9}{2},$$

所以正八面体的体积与正方体的体积之比是 $9:2$。

验算

- $\arccos(-\frac{1}{3}) \approx 109.47^\circ$。对坐标模型的全部 12 条棱，以数值方式重复上述中点计算，每次都得出 $\cos = -\frac{1}{3}$。
- 正八面体就是 $|x|+|y|+|z|\le 1$。在 $[-1,1]^3$ 内随机取 10^6 个点，以蒙地卡罗（Monte Carlo）方法估算它的体积，结果是 1.330，而准确值是 $\frac{4}{3}$。
- 正方体位于正八面体之内（每个中心都满足 $|x|+|y|+|z|=1$），所以比值必定大于 1。而 $\frac{9}{2}$ 确实如此。

常见错误

- 取了向外法向量（outward normal）$(1,1,1)$ 与 $(1,1,-1)$ 之间的夹角，得出 $\arccos\frac{1}{3}$。两个面本身之间的夹角是它的补角。
- 沿着不垂直于公共棱的直线量度夹角，例如沿两条棱 CA 和 DA。在这里那个角是 90°，但它并不是两个面之间的夹角。
- 以为正方体的边长就是两个相对的面的中心之间的距离。那个距离 $\frac{2}{\sqrt{3}}$ 其实是正方体的一条空间对角线。

STEP 之外

正方体与正八面体是互为对偶（dual）的多面体：把其中一个的各面中心连接起来，就得到另一个，面与顶点的角色互相对调。对偶性是第一个迹象，显示这两个立体的对称群（symmetry group）是同一个群，其阶为 48；你会在群论的入门课程中遇到它。想了解柏拉图立体（Platonic solids）的对称群，可参阅 Armstrong 的 *Groups and Symmetry*（一本关于群与对称的入门书）。

详细解答

STEP I 2007，第 6 题

纯数学 · 丢番图方程

> **STEP I 2007，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/07-s1-q6

关键思路

两部分都由幂的差出发，而 $x-y$ 是每一个这类式子的因式。约去这个因式（因为 $d \neq 0$，所以可以这样做）之后，余下的是一个关于 x 和 y 的对称式（symmetric expression）。这个式子连同 $x-y=d$，就可以确定 x 和 y。要得到整数解，我们再选取适当的 d，使公式算出来的都是整数。在第 (ii) 部分，这就是要令一个平方根成为有理数（rational number）。

解答

第 (i) 部分

把左方因式分解（factorise），

$$(x-y)(x+y) = (x-y)^3.$$

由于 $x-y=d\neq 0$，我们可以除以 $x-y$，得 $x+y=(x-y)^2=d^2$。把这式与 $x-y=d$ 相加和相减，得

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

要找整数对，设 $x=\sqrt{m}$ 及 $y=\sqrt{n}$。那么 $x^2-y^2=m-n$，而所要求的条件正好是 $x^2-y^2=(x-y)^3$。取 $d=\sqrt{m}-\sqrt{n}$，由于 $m>n$，它不等于零，于是

$$m=\left(\frac{d(d+1)}{2}\right)^2, \qquad n=\left(\frac{d(d-1)}{2}\right)^2.$$

若 d 是正整数，则 $d(d\pm 1)$ 是两个连续整数之积，所以是偶数。于是 $\sqrt{m}$ 和 $\sqrt{n}$ 都是整数，m 和 n 也是整数。此外，因为 $d>0$，所以 $m>n$。我们需要 $n>100$，即 $\frac{1}{2}d(d-1)>10$。

取 $d=5$ 得 $\frac{1}{2}d(d-1)=10$，并不足够，所以我们取 $d=6$：

$$\sqrt{m}=21,\quad \sqrt{n}=15,\qquad m=441,\quad n=225.$$

验证：$m-n=216=6^3=(21-15)^3$。（其他答案同样成立，例如 $d=7$ 时 $m=784$，$n=441$。）

第 (ii) 部分

再次因式分解：$x^3-y^3=(x-y)(x^2+xy+y^2)$。除以 $x-y=d\neq 0$，得

$$x^2+xy+y^2=d^3.$$

我们还有 $(x-y)^2=d^2$，即 $x^2-2xy+y^2=d^2$。从上一条方程减去这一条，得 $3xy=d^3-d^2$，正是所要证明的结果。

现在代入 $y=x-d$：

$$3x(x-d)=d^3-d^2\quad\Longrightarrow\quad 3x^2-3dx-(d^3-d^2)=0.$$

由二次公式（quadratic formula），

$$x=\frac{3d\pm\sqrt{9d^2+12(d^3-d^2)}}{6}=\frac{3d\pm\sqrt{3d^2(4d-1)}}{6}.$$

现在 $\sqrt{3d^2(4d-1)}=|d|\sqrt{3(4d-1)}$，而 $\pm$ 已经涵盖正负两种情况，所以可以把 $|d|$ 换成 d。乘以 2，

$$2x=d\pm\frac{d\sqrt{3(4d-1)}}{3}=d\pm d\sqrt{\frac{4d-1}{3}},$$

这就是所要求的结果。

想一想 怎样选取 d，才能使 x 和 y 都是整数?

若 $m=x$ 和 $n=y$ 是整数，则 $d=m-n$ 是非零整数。要使 x 是有理数，$\sqrt{(4d-1)/3}$ 必须是有理数。把它记作 k。那么 $k^2=(4d-1)/3$ 是整数，而整数的平方根只有在该整数是完全平方数（perfect square）时才是有理数，所以 k 是整数。因此我们需要

$$4d-1=3k^2\quad \text{对某个整数}k\geq 0.$$

左方是奇数，所以 k 是奇数。于是 $2x=d(1\pm k)$ 是偶数，所以 x 是整数，$y=x-d$ 也是整数。

- $k=1$ 给出 $d=1$，于是 $x=1$，$y=0$ 或 $x=0$，$y=-1$。两组都不是由两个正整数组成。
- $k=3$ 给出 $d=7$，于是 $2x=7\pm 21$。取正号得 $x=14$，$y=7$。

所以 $m = 14$，$n = 7$ 符合要求：

$$m^3 - n^3 = 2744 - 343 = 2401 = 7^4 = (m-n)^4.$$

次序是重要的：若取 $m = 7$，$n = 14$，左方是 -2401，右方却是 $+2401$。

验算

- 第 (i) 部分：$21^2 - 15^2 = 441 - 225 = 216$，而 $(21-15)^3 = 216$。
- 第 (ii) 部分：取 $d = 7$，$3xy = 3 \cdot 14 \cdot 7 = 294$，而 $d^3 - d^2 = 343 - 49 = 294$。
- 用电脑在 $1 \le m, n < 200$ 的范围内搜索，恰好找到 $(14,7)$、$(57,38)$ 和 $(148,111)$。它们来自 $k = 3, 5, 7$（即 $d = 7, 19, 37$），与公式吻合。搜索结果亦确认，在第 (i) 部分满足 $n > 100$ 的解之中，$(441,225)$ 是最小的。

常见错误

- 忘记约去因式时需要 $x - y \neq 0$。这正是题目假设 $d \neq 0$ 的原因。
- 在第 (i) 部分取 $d = 5$，得 $n = 100$，不符合严格不等式（strict inequality）$n > 100$。
- 不加说明就写 $\sqrt{d^2} = d$。它其实等于 $|d|$，是因为有 $\pm$，我们才可以去掉绝对值符号。

STEP 之外

第 (ii) 部分把丢番图方程（Diophantine equation）$m^3 - n^3 = (m-n)^4$ 完全解出：对奇数 $k \ge 3$ 取 $d = (3k^2+1)/4$，解就是 $m = \frac{1}{2}d(k+1)$ 和 $n = \frac{1}{2}d(k-1)$。要求一个平方根是有理数，是把这类方程化为简单得多的条件 $4d - 1 = 3k^2$ 的标准方法。这类方程，以及它们背后二次型（quadratic form）的算术，可参阅 Davenport 的 *The Higher Arithmetic*（一本数论入门经典）。

详细解答

STEP I 2007，第 7 题

纯数学 · 直线之间的距离

STEP I 2007，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/07-s1-q7

关键思路

在两条直线上各写出一个一般点，相减，再把长度平方：这样 D^2 就是两个参数（parameter）的二次式（quadratic）。窍门在于把这个二次式写成平方和（sum of squares），因为当每个平方都取到最小时，平方和也就最小。在第 (ii) 部分，同样的方法仍然适用，而对于不同的 k 值，这些平方如何拆开，正好告诉我们背后的几何。

解答

第 (i) 部分

L_1 上的一般点是 $P = (1+2\lambda,\ 2\lambda,\ 2-3\lambda)$，$L_2$ 上的一般点是 $Q = (4+\mu,\ -2+2\mu,\ 9-2\mu)$。所以

$$\overrightarrow{PQ} = \begin{pmatrix} 3+\mu-2\lambda \\ -2+2\mu-2\lambda \\ 7-2\mu+3\lambda \end{pmatrix}$$

而

$$D^2 = (3+\mu-2\lambda)^2 + (-2+2\mu-2\lambda)^2 + (7-2\mu+3\lambda)^2.$$

逐个系数合并同类项：

$$\mu^2:\ \ 1+4+4=9, \qquad \lambda^2:\ \ 4+4+9=17, \qquad \lambda\mu:\ \ -4-8-12=-24,$$
$$\mu:\ \ 6-8-28=-30, \qquad \lambda:\ \ -12+8+42=38, \qquad \text{常数项}:\ \ 9+4+49=62,$$

所以 $D^2 = 9\mu^2 + 17\lambda^2 - 24\lambda\mu - 30\mu + 38\lambda + 62$。

现在把题目给出的形式展开：

$$\begin{aligned}(3\mu-4\lambda-5)^2+(\lambda-1)^2+36 &= \left(9\mu^2+16\lambda^2+25-24\lambda\mu-30\mu+40\lambda\right)\\ &\quad+\left(\lambda^2-2\lambda+1\right)+36\\ &= 9\mu^2+17\lambda^2-24\lambda\mu-30\mu+38\lambda+62.\end{aligned}$$

两者相同，所以

$$D^2=(3\mu-4\lambda-5)^2+(\lambda-1)^2+36,$$

正如所求。

想一想 为甚么我们可以令两个平方同时为零?

每个平方都不小于 0，所以无论 λ 和 μ 怎样选，都有 $D^2\geq 36$。等号成立需要 $\lambda=1$ 及 $3\mu=4\lambda+5=9$，即 $\mu=3$。这两个条件互不冲突（第一个条件确定 λ，然后第二个条件确定 μ），所以 $D^2=36$ 确实可以取到。因此两条直线之间的最短距离是

$$D_{\min}=6,$$

并在以下两点取到：

$$\lambda=1:\ P=(3,\,2,\,-1)\quad \text{位于}L_1,\qquad \mu=3:\ Q=(7,\,4,\,3)\quad \text{位于}L_2.$$

作为验算，$\overrightarrow{PQ}=(4,2,4)$ 的长度是 $\sqrt{16+4+16}=6$，而且它与两个方向向量（direction vector）都垂直（perpendicular）：$(4,2,4)\cdot(2,2,-3)=8+4-12=0$ 及 $(4,2,4)\cdot(1,2,-2)=4+4-8=0$。连接两条直线的最短线段，正应如此。

第 (ii) 部分

L_3 上的一般点是 $P=(2,\,3+\alpha,\,5)$，L_4 上的一般点是 $Q=\left(3+4k\beta,\,3+(1-k)\beta,\,-2-3k\beta\right)$。所以

$$\overrightarrow{PQ}=\begin{pmatrix}1+4k\beta\\(1-k)\beta-\alpha\\-7-3k\beta\end{pmatrix}$$

而

$$\begin{aligned}D^2&=(1+4k\beta)^2+\left((1-k)\beta-\alpha\right)^2+(7+3k\beta)^2\\&=\left((1-k)\beta-\alpha\right)^2+25k^2\beta^2+50k\beta+50\\&=\left((1-k)\beta-\alpha\right)^2+25(k\beta+1)^2+25,\end{aligned}$$

这是因为 $(1+4k\beta)^2+(7+3k\beta)^2=1+8k\beta+16k^2\beta^2+49+42k\beta+9k^2\beta^2$。

无论 β 是甚么，我们都可以选 $\alpha=(1-k)\beta$ 令第一个平方为零，因为 α 没有在其他地方出现。所以问题变成 $25(k\beta+1)^2$ 可以小到甚么程度，而两种情况正是在这里分开。

想一想 当 $k=0$ 时，有甚么不同?

若 $k\neq 0$，我们可以取 $\beta=-1/k$，令 $k\beta+1=0$。若 $k=0$，无论 β 是甚么，$25(k\beta+1)^2$ 这一项都固定在 25。

情况 $k\neq 0$。每个平方都是非负的，所以 $D^2\geq 25$，等号在 $\beta=-1/k$、$\alpha=(1-k)\beta=(k-1)/k$ 时成立。最短距离是

$$D_{\min}=5,$$

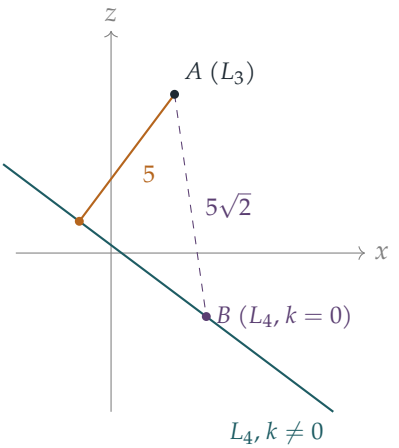
与 k 的（非零）值无关。最接近的两点是 $P=\left(2,\,3+\frac{k-1}{k},\,5\right)$ 和 $Q=\left(-1,\,3+\frac{k-1}{k},\,1\right)$，其中 $\overrightarrow{PQ}=(-3,0,-4)$，长度为 5。由于 $D_{\min}>0$，两条直线永不相交；而当 $k\neq 0$ 时，它们的方向 $(0,1,0)$ 与 $(4k,1-k,-3k)$ 并不平行（后者的第一和第三个分量不同时为零）。所以这两条直线是*异面*（*skew*）直线，而 5 就是它们的公垂线（common perpendicular）的长度，这条公垂线是唯一的。

情况 $k=0$。这时 L_4 的方向是 $(0,1,0)$，与 L_3 相同：两条直线*平行*（*parallel*）。此时 $D^2=(\beta-\alpha)^2+50$，只要 $\alpha=\beta$ 就取到最小值，所以

$$D_{\min}=\sqrt{50}=5\sqrt{2},$$

而且有无穷多对点可以取到这个值：其中一条直线上的任意一点，配上另一条直线上正对着它的那一点（y 坐标相同）。这其实就是两条平行线之间的固定距离。

几何意义。要看清这两条直线，最容易的方法是沿着 y 轴直望过去，即投影（projection）到 xz 平面上。由于 L_3 与 y 轴平行，它投影成单独一点 $A=(2,5)$，而与 L_3 垂直的距离在投影下保持不变。当 $k\neq 0$ 时，L_4 投影成通过 $(3,-2)$、方向为 $(4,-3)$ 的直线（$1-k$ 那个分量消失了），而 A 到这条直线的距离是 $\left|1\cdot(-3)-(-7)\cdot 4\right|/5=5$，与 k 无关。当 $k=0$ 时，L_4 投影成单独一点 $B=(3,-2)$，距离是 $AB=\sqrt{1+49}=5\sqrt{2}$。



所以，当 $k \neq 0$ 时答案是 5（异面直线，只有一条最短线段，它与两条直线都垂直）；当 $k = 0$ 时答案是 $5\sqrt{2}$（平行直线，距离处处相同）。

验算

在第 (i) 部分，代入 $\lambda = 1$、$\mu = 3$ 所得的两点相距 6，而且 $\overrightarrow{PQ}$ 与两条直线都垂直。在 (λ, μ) 的网格上作数值搜索，找不到低于 $D^2 = 36$ 的值。在第 (ii) 部分，对 $k = 0.001,\ 0.3,\ 1,\ -2,\ 5$ 以数值方法求 D^2 的最小值，每次都得到 25，而 $k = 0$ 时则得到 50。留意 $D_{\min}$ 在 $k = 0$ 处由 5 跳到 $5\sqrt{2}$。当 k 很小时，最接近的两点位于很远的地方，在 $\beta = -1/k$ 附近，所以极限不一定连续。

常见错误

- 忘记检查两个平方能否同时为零。这里是可以的，但必须明确说明，不能想当然。
- 在 $k = 0$ 时除以 k：$\beta = -1/k$ 这个选择只在 $k \neq 0$ 时可用，这正是要分两种情况的原因。
- 把 $k \neq 0$ 时的两条直线说成“相交”。它们不平行，但 $D_{\min} = 5 > 0$，所以它们是异面直线。

STEP 之外

多变量的配方法（completing the square）是二次型（quadratic forms）理论的起点：把 D^2 写成平方和，就是把一个对称矩阵（symmetric matrix）对角化（diagonalising），而求最小值则是一个最小平方法（least squares）问题。我们最后验证的垂直关系，就是最小平方法中的“正规方程”（normal equations），而异面直线距离的一般公式 $|(\mathbf{b}-\mathbf{a})\cdot(\mathbf{u}\times\mathbf{v})|/|\mathbf{u}\times\mathbf{v}|$ 也是来自同一套几何。Strang 的 *Introduction to Linear Algebra*（线性代数入门）对投影和最小平方法有十分精彩的讲解。

详细解答

STEP I 2007，第 8 题

纯数学 · 三次曲线的相切

STEP I 2007，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/07-s1-q8

关键思路

两条曲线在某一点相切（touch），当且仅当求交点的方程在该处有重根（repeated root）。因此应考虑两个三次式之差，并预期它含有因式 $(x-2)^2$。含 a 各项的系数 $1, -6, 12, -8$ 正是 $(x-2)^3$ 的二项式系数（binomial coefficients），看出这一点后，因式分解（factorisation）和绘画略图都会容易得多。

解答

在 $(2,8)$ 相切及另一个交点

把含 a 的各项归在一起，

$$ax^3 - 6ax^2 + 12ax - 8a = a(x-2)^3, \qquad \text{所以} \qquad y = a(x-2)^3 + 12x - 16. \qquad (*)$$

它与 $y = x^3$ 的交点，就是下式的根：

$$a(x-2)^3 + 12x - 16 - x^3 = 0.$$

现在 $x^3 - 12x + 16$ 在 $x = 2$ 时为零，除去相应因式后得 $x^3 - 12x + 16 = (x-2)^2(x+4)$。因此

$$\begin{aligned} a(x-2)^3 + 12x - 16 - x^3 &= a(x-2)^3 - (x-2)^2(x+4) \\ &= (x-2)^2\big[(a-1)x - 2a - 4\big]. \end{aligned}$$

在 $x = 2$ 处，两条曲线都通过 $(2,8)$，而因式 $(x-2)^2$ 表明 $x = 2$ 是交点方程的重根。等价地说，两者的斜率（gradient）相同：由 $(*)$ 得 $y' = 3a(x-2)^2 + 12$，它在 $x = 2$ 时等

于 12，而 $y = x^3$ 在该处的斜率是 $3 \cdot 2^2 = 12$。所以对 a 的每一个值，两条曲线都在 $(2, 8)$ 相切。

其余的交点都来自 $(a-1)x = 2a + 4$。

- 若 $a \neq 1$，则 $x = \dfrac{2a+4}{a-1}$，而它永远不会等于 2（否则需要 $2a + 4 = 2a - 2$）。所以恰好有另外一个交点，
$$\left(\frac{2a+4}{a-1}, \left(\frac{2a+4}{a-1}\right)^3\right).$$
- 若 $a = 1$，方括号内是常数 -6，所以没有其他交点。

想一想 哪一条曲线在上方？在甚么范围内？

因式分解已顺带回答了这个问题。由于 $(x-2)^2 \geq 0$，$(*)$ 减去 x^3 的正负号就是线性因式 $(a-1)x - 2a - 4$ 的正负号。由此可知曲线 $(*)$ 在另一个交点的哪一侧位于 $y = x^3$ 之上，而绘画略图所需的资料，大部分就在这里。

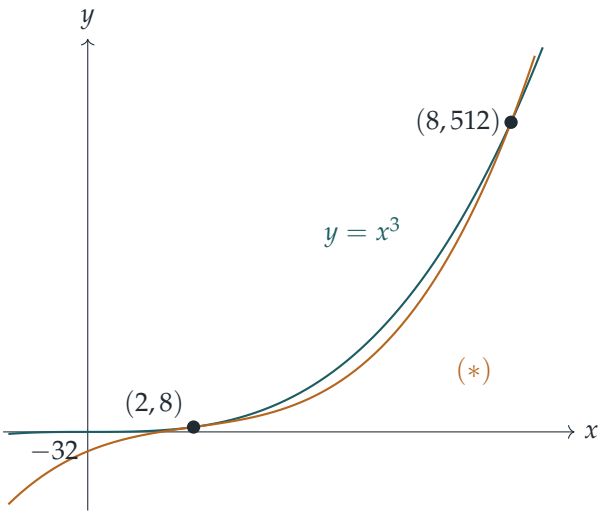
在全部三幅略图中，$(*)$ 在 $(2, 8)$ 都有一个拐点（point of inflection），该处斜率为 12，因为当 $a \neq 0$ 时 $y'' = 6a(x-2)$ 在该处变号。所以 $(*)$ 与 $y = x^3$ 在 $(2, 8)$ 有共同的切线（tangent）$y = 12x - 16$；$y = x^3$ 在该处向上凹（concave up），而 $(*)$ 则不然。

第 (i) 部分： $a = 2$

这时 $y = 2(x-2)^3 + 12x - 16$，而

$$y - x^3 = (x-2)^2(x-8).$$

另一个交点是 $(8, 512)$。曲线 $(*)$ 在 $x < 8$ 时位于 $y = x^3$ 之下（在 $x = 2$ 处相切），在 $x > 8$ 时位于其上。它的斜率 $6(x-2)^2 + 12$ 恒为正数，所以它是递增的，没有转向点（turning points）。它与 y 轴相交于 $(0, -32)$，并与 x 轴相交一次，位置在 $x \approx 1.37$ 附近。



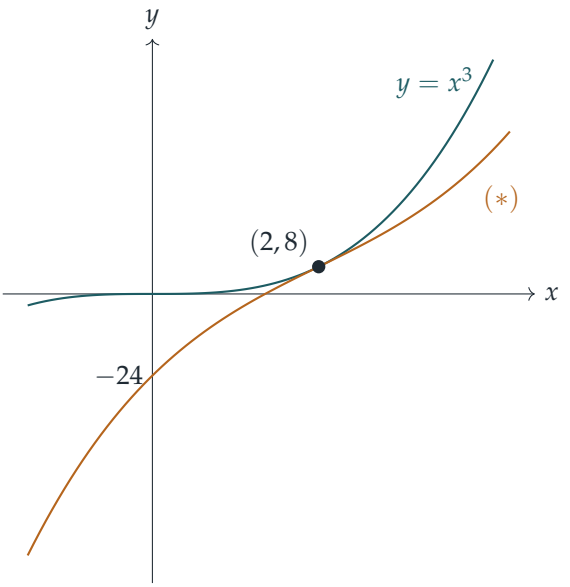
（纵向比例被大幅压缩，所以在 $(2,8)$ 的相切看起来是平的；两条曲线在该处有共同的切线。）

第 (ii) 部分： $a = 1$

这时 $y = (x-2)^3 + 12x - 16 = x^3 - 6x^2 + 24x - 24$，而

$$y - x^3 = -6(x-2)^2 \leq 0.$$

所以除了在 $(2,8)$ 相切之外，$(*)$ 处处位于 $y = x^3$ 之下；两者没有其他交点。无论向哪一个方向离开 $x = 2$，两者的差距 $6(x-2)^2$ 都会愈来愈大。斜率 $3(x-2)^2 + 12$ 恒为正数，所以 $(*)$ 是递增的。它与 y 轴相交于 $(0,-24)$，并与 x 轴相交一次，位置在 $x \approx 1.36$ 附近。



第 (iii) 部分： $a = -2$

现在另一个交点满足 $x = \dfrac{-4+4}{-3} = 0$，所以它就是原点 $(0,0)$，而

$$y - x^3 = (x-2)^2(-3x) = -3x(x-2)^2.$$

所以 $(*)$ 在 $x < 0$ 时位于 $y = x^3$ 之上，在 $x > 0$ 时位于其下，并在 $(2,8)$ 相切。

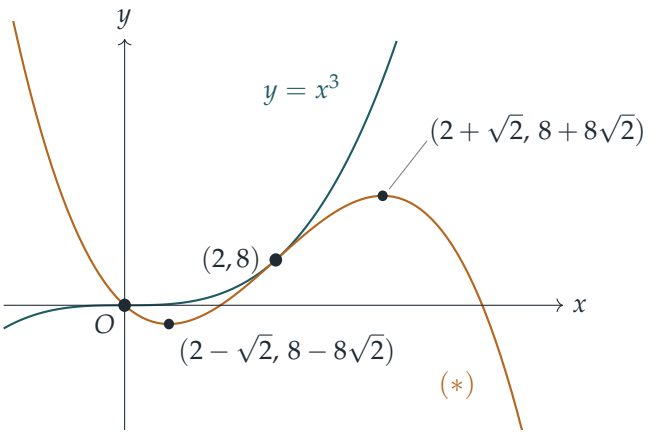
展开后，$y = -2(x-2)^3 + 12x - 16 = -2x^3 + 12x^2 - 12x = -2x(x^2 - 6x + 6)$，它与 x 轴相交于 $x = 0$ 和 $x = 3 \pm \sqrt{3}$（约为 1.27 和 4.73）。斜率是

$$y' = -6(x-2)^2 + 12,$$

它在 $(x-2)^2 = 2$ 时为零，即 $x = 2 \pm \sqrt{2}$。代入后，

$$x = 2 - \sqrt{2}: \quad y = 8 - 8\sqrt{2} \approx -3.31 \quad \text{(极小值)},$$
$$x = 2 + \sqrt{2}: \quad y = 8 + 8\sqrt{2} \approx 19.3 \quad \text{(极大值)}.$$

由于首项系数是负数，当 $x \to -\infty$ 时 $y \to +\infty$，当 $x \to +\infty$ 时 $y \to -\infty$。在原点，$(*)$ 的斜率是 -12，而 $y = x^3$ 的斜率是 0，所以两条曲线在该处确实互相穿越。



验算

- 在 $x = 2$ 处，对每一个 a，$(*)$ 都给出 $a \cdot 0 + 24 - 16 = 8$，且 $y' = 12$，与 $y = x^3$ 吻合。
- 取 $a = 2$ 和 $x = 8$，由简洁的形式得 $2 \cdot 6^3 + 96 - 16 = 512 = 8^3$，确认了点 $(8, 512)$。
- 取 $a = 0$，$(*)$ 就是切线 $y = 12x - 16$，而公式给出另一个交点在 $x = -4$；事实上 $x^3 - 12x + 16 = (x-2)^2(x+4)$。
- 我们用数千组随机有理数 a 和 x 精确地验证了这个因式分解，并以数值方法核对了第 (iii) 部分的转向值 $8 \pm 8\sqrt{2} \approx 19.31, -3.31$。

常见错误

- 遗漏 $a = 1$ 的情况：这时另一个交点的公式失效，因为根本没有另一个交点。
- 只证明两条曲线都通过 $(2, 8)$。相切还需要斜率相等，或等价地，需要有重根。
- 在第 (iii) 部分，把两条曲线画成在原点相切。它们在该处是互相穿越的，斜率并不相同。

STEP 之外

$f(x) - g(x)$ 的重根是相切在代数上的投影：两条曲线在该点一阶吻合。再推进一步，切触阶（order of contact）数算的是有多少阶导数相同，而这正是泰勒多项式（Taylor polynomials）背后的想法：泰勒多项式就是在某一点与函数有最高可能切触阶的多项式。Bartle 与 Sherbert 的 *Introduction to Real Analysis*（实分析入门）严谨地建立了泰勒定理。Siklos 的 *Advanced Problems in Mathematics*（高等数学难题集）有更多 STEP 风格的曲线描绘题

目。

详细解答

STEP I 2007，第 9 题

力学 · 斜坡上的极限摩擦力

STEP I 2007，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：

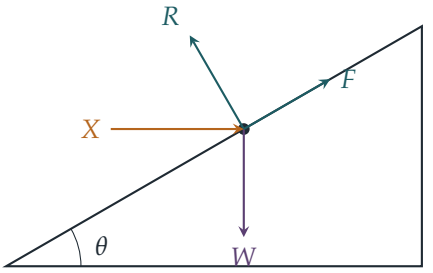
step.maths.org/questions/07-s1-q9

关键思路

这题是极限摩擦（limiting friction）的两种情况，两者的分别只在于摩擦力的方向。在每种情况下，我们沿斜面方向及垂直于斜面方向分解力，再消去法向反作用力（normal reaction），便可把每个水平力写成 W 乘以一个三角函数式的比。将两式相除可消去 W，剩下一条只涉及 k、μ 和 θ 的关系式。至于不等式，则来自 $\sin\theta\cos\theta = \frac{1}{2}\sin 2\theta$ 永远不会超过 $\frac{1}{2}$ 这个事实。

解答

我们把物体模型化为质点（particle），并假设两种情况下摩擦力均达到极限，即 $F = \mu R$。为使代数运算简洁，记 $s = \sin\theta$ 及 $c = \cos\theta$。



水平力把质点压向斜面，所以它有一个沿斜面向上的分量 Xc，以及一个垂直压向斜面的分量 Xs。重量则有沿斜面向下的分量 Ws 和垂直压向斜面的分量 Wc。

即将下滑：力为 X

质点正处于即将向下滑动的状态，所以摩擦力 μR 沿斜面向上（如图所示）。先沿垂直于斜面的方向分解，再沿斜面方向分解，得

$$R = Wc + Xs, \qquad Xc + \mu R = Ws.$$

代入 R，

$$Xc + \mu Wc + \mu Xs = Ws \quad \Longrightarrow \quad X(c + \mu s) = W(s - \mu c). \tag{1}$$

即将上滑：力为 kX

现在质点正处于即将向上滑动的状态，所以摩擦力沿斜面向下。同理，

$$R' = Wc + kXs, \qquad kXc = Ws + \mu R',$$

所以

$$kX(c - \mu s) = W(s + \mu c). \tag{2}$$

想一想 这里眼前的每一项，我们都可以随意拿来做除数吗?

可以，这是由题目的物理情境保证的。题目说需要一个正的力 X 才能阻止质点下滑，即是说没有这个力质点便会滑下，这表示 $\tan\theta > \mu$，即 $s - \mu c > 0$。于是由 (1) 得 $X > 0$，而 $c + \mu s > 0$ 本来就成立。水平推力若要能够把质点推上斜面，(2) 便需要 $c - \mu s > 0$，即 $\tan\theta < 1/\mu$。两者合起来必有 $\mu < \tan\theta < 1/\mu$，特别是 $\mu < 1$；这一点到最后会用得着。

关系式

将 (2) 除以 (1)，X 和 W 同时消去：

$$k\,\frac{c - \mu s}{c + \mu s} = \frac{s + \mu c}{s - \mu c} \quad \Longrightarrow \quad k(c - \mu s)(s - \mu c) = (s + \mu c)(c + \mu s).$$

把两边展开，并利用 $s^2 + c^2 = 1$：

$$(c - \mu s)(s - \mu c) = sc - \mu c^2 - \mu s^2 + \mu^2 sc = (1 + \mu^2)sc - \mu,$$
$$(s + \mu c)(c + \mu s) = sc + \mu s^2 + \mu c^2 + \mu^2 sc = (1 + \mu^2)sc + \mu.$$

所以 $k(1 + \mu^2)sc - k\mu = (1 + \mu^2)sc + \mu$，整理后得

$$(k - 1)(1 + \mu^2)\sin\theta\cos\theta = \mu(k + 1),$$

这正是所需的结果。

不等式

右边是正数（因为 $\mu > 0$ 及 $k > 0$），而当 θ 为锐角时 $(1+\mu^2)\sin\theta\cos\theta$ 亦是正数；因此 $k > 1$。（这并不令人意外：把质点推上去所需的力，自然比仅仅把它顶住的大。）现在

$$\sin\theta\cos\theta = \tfrac{1}{2}\sin 2\theta \le \tfrac{1}{2},$$

而由于 $(k-1)(1+\mu^2) > 0$，我们可以把这个不等式两边乘以它：

$$\mu(k+1) = (k-1)(1+\mu^2)\sin\theta\cos\theta \le \tfrac{1}{2}(k-1)(1+\mu^2).$$

两边乘以 2，再把含 k 的项集合起来：

$$2\mu k + 2\mu \le k(1+\mu^2) - (1+\mu^2) \quad\Longrightarrow\quad k(1-2\mu+\mu^2) \ge 1 + 2\mu + \mu^2,$$

即 $k(1-\mu)^2 \ge (1+\mu)^2$。上文已证明 $\mu < 1$，所以 $(1-\mu)^2 > 0$，我们可以除以它而不改变不等号的方向：

$$k \ge \frac{(1+\mu)^2}{(1-\mu)^2}.$$

等号成立当且仅当 $\sin 2\theta = 1$，即 $\theta = 45°$。

> ? **想一想** 为甚么在最后一步之前，必须先知道 $\mu < 1$?

若 $\mu = 1$，我们便是除以零（而且上一行会变成 $0 \ge 4$，这是错的）；若 $\mu > 1$，相除虽然仍然合法，但 θ 须满足的条件早已不成立。从物理上看，当 $\mu \ge 1$ 时，$\tan\theta > \mu \ge 1$ 和 $\tan\theta < 1/\mu \le 1$ 这两个要求互相矛盾：要么质点根本不需要外力顶住，要么任何水平推力都无法令它开始向上滑。所以题目其实隐含了 $\mu < 1$。

验算

- 取 $\mu = \frac{1}{2}$ 及 $\theta = 45°$，则 $s = c = 1/\sqrt{2}$。由 (1) 和 (2)，$X = W\cdot\frac{1/2}{3/2} = \frac{1}{3}W$ 及 $kX = W\cdot\frac{3/2}{1/2} = 3W$，所以 $k = 9 = (1.5/0.5)^2$：这正是等号成立的情况。若改取 $\theta = 60°$，同样的公式给出 $k \approx 25.2 > 9$。
- 若 $\mu \to 0$，则 $X = kX = W\tan\theta$ 且 $k \to 1$，这与恒等式（两边皆为 0）及下界 1 都吻合。
- 以数值方法检验恒等式和不等式：随机选取 10^5 组 $\mu \in (0,1)$ 和符合条件的 θ，没有发现任何反例。

常见错误

- 取 $R = W\cos\theta$：水平力同样会把质点压向斜面，令反作用力增加 $X\sin\theta$。

- 两种情况下都让摩擦力指向同一方向。摩擦力总是与即将发生的运动方向相反。
- 除以 $(1-\mu)^2$ 而没有说明它为何是正数（且不等于零）。

STEP 之外

要看清这个结果，有一个巧妙的方法，就是引入摩擦角（angle of friction）λ，其中 $\tan\lambda = \mu$：法向反作用力与极限摩擦力可合成为单一个力，它与法线成 λ 角，而 (1) 和 (2) 则变成 $X = W\tan(\theta-\lambda)$ 及 $kX = W\tan(\theta+\lambda)$。所以 $k = \tan(\theta+\lambda)/\tan(\theta-\lambda)$，而该下界所说的，就是这个比在 $\theta = 45^\circ$ 时最小。这类运用摩擦角的论证，以及“摩擦锥”（friction cone）的概念，在大学的静力学（statics）和接触理论中会再次出现。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）对摩擦和受力分析的处理十分严谨，正合大学一年级课程的要求。

详细解答

STEP I 2007，第 10 题

力学 · 位移–时间图

> **STEP I 2007，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/07-s1-q10

关键思路

在位移—时间图（displacement–time graph）上，每名骑兵和每支军队都是一条直线（或两段直线），而“彼此相遇”不过是两条线相交。因此我们以撒克逊军队为起点量度位移，写出每条线的方程，再求它们的交点。题目中的不等式，正正就是各个交点按我们预期的次序出现的条件；而最后一部分所问的，是其中一个条件不成立时会发生甚么事。

解答

建立图像

我们把军队和骑兵看成沿一条直线运动的点。设 s 为由（静止的）撒克逊军队起计、朝诺曼军队方向的位移，并在两军相距 d 时开始计时。于是：

- 撒克逊军队：$s = 0$。
- 诺曼军队：$s = d - ut$。
- 撒克逊骑兵，去程：$s = xt$。他到达诺曼军队时 $xt = d - ut$，即在 $t_S = \dfrac{d}{x+u}$，位置为 $s = \dfrac{xd}{x+u}$。他以速率 x 折返，

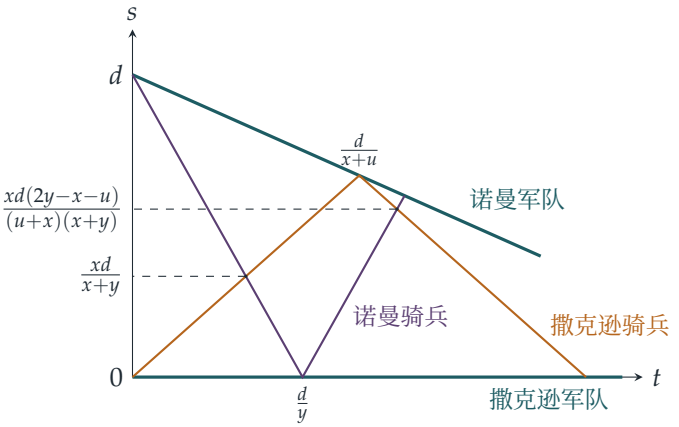
  $$s = \frac{xd}{x+u} - x\left(t - \frac{d}{x+u}\right) = \frac{2xd}{x+u} - xt,$$

  并在 $t = \dfrac{2d}{x+u}$ 回到本阵。
- 诺曼骑兵，去程：$s = d - yt$。他在 $t_N = \dfrac{d}{y}$ 到达撒克逊军队，然后以速率 y 折返，其

直线为

$$s = y\left(t - \frac{d}{y}\right) = yt - d.$$

由于 $y > u$，他确实能追上自己的军队，此时 $yt - d = d - ut$，即在 $t = \dfrac{2d}{y+u}$。



（图中取 $u = 1$、$x = 2$、$y = 4$、$d = 1$。）

第一次相遇

两名骑兵都在去程，所以我们求 $s = xt$ 与 $s = d - yt$ 的交点：$(x+y)t = d$，得

$$t = \frac{d}{x+y}, \qquad s = \frac{xd}{x+y}.$$

这确实发生在两人掉头之前，因为 $\frac{d}{x+y}$ 比 $\frac{d}{y}$ 和 $\frac{d}{x+u}$ 都小（后者用到 $y > u$）。

第二次相遇

现在两名骑兵都在回程：我们求 $s = \dfrac{2xd}{x+u} - xt$ 与 $s = yt - d$ 的交点。两边同时加上 $xt + d$，

$$(x+y)t = d + \frac{2xd}{x+u} = \frac{d(3x+u)}{x+u}, \qquad \text{所以} \qquad t_2 = \frac{d(3x+u)}{(x+u)(x+y)}.$$

于是

$$s = \frac{2xd}{x+u} - \frac{xd(3x+u)}{(x+u)(x+y)} = \frac{xd\big(2(x+y) - (3x+u)\big)}{(x+u)(x+y)} = \frac{xd(2y-x-u)}{(u+x)(x+y)},$$

这就是题目要求的第二个距离。

想一想 我们怎样知道在时刻 t_2 两名骑兵真的都在回程?

这正是那些不等式派上用场的地方。诺曼骑兵已经掉头，若 $t_2 \geq \frac{d}{y}$:

$$y(3x+u) \geq (x+u)(x+y) \iff 3xy \geq x^2+xy+ux \iff u \leq 2y-x.$$

撒克逊骑兵已经掉头，若 $t_2 \geq \frac{d}{x+u}$:

$$3x+u \geq x+y \iff u \geq y-2x.$$

所以条件 $y-2x < u < 2y-x$ 恰好保证 t_2 在两个掉头时刻之后。同一个条件 $u < 2y-x$ 亦使 $s > 0$, 即撒克逊骑兵尚未回到本阵。(还有一个题目没有明言的要求：诺曼骑兵不可先回到自己的军队。把 $t_2 \leq \frac{2d}{y+u}$ 化简, 可得 $(x-u)(2x-y+u) \geq 0$; 由于 $2x-y+u > 0$, 这表示 $x \geq u$，即撒克逊骑兵至少与诺曼军队一样快。这肯定是题目的原意：否则他一掉头，诺曼军队便会越过他。)

情况 (i)： $u > 2y-x$

此时 $t_2 < \frac{d}{y}$：撒克逊骑兵回程的直线与诺曼骑兵去程的直线先相交。由于 $y > u$，这个情况需要 $x > y$：撒克逊骑兵是较快的一方。他到达诺曼军队后掉头，在两人都朝撒克逊军队前进时从后追上诺曼骑兵。(求 $s = \frac{2xd}{x+u} - xt$ 与 $s = d - yt$ 的交点，可知这发生在距撒克逊军队 $\frac{xd(x+u-2y)}{(x+u)(x-y)}$ 处。) 他在 $\frac{2d}{x+u}$ 回到本阵，早于诺曼骑兵到达的时刻 $\frac{d}{y}$，所以两人不会第二次迎面相遇：诺曼骑兵只是到达撒克逊军队，而他的对手早已回来了。

情况 (ii)： $u < y-2x$

此时 $t_2 < \frac{d}{x+u}$：诺曼骑兵快得多。他到达撒克逊军队后掉头，越过仍在朝诺曼军队前进的撒克逊骑兵。求 $s = yt - d$ 与 $s = xt$ 的交点，可知这发生在 $t = \frac{d}{y-x}$，距撒克逊军队 $\frac{xd}{y-x}$ 处。其后诺曼骑兵在 $\frac{2d}{y+u}$ 回到本阵，早于撒克逊骑兵到达诺曼军队的时刻 $\frac{d}{x+u}$ (因为 $2(x+u) < y+u$)，此后两人在战场上再无相遇。

验算

- 取 $d=1$、$u=1$、$x=2$、$y=4$ (如图所示)，公式给出相遇位置为 $\frac{1}{3}$ 和 $\frac{5}{9}$。以 10^{-5} 为时间步长用数值方法逐步模拟这个运动，得到的交点在 0.3333 和 0.5556。
- 若 $y = x+u$，两个距离相等，都是 $\frac{x}{x+y}d$。这是一个很适合用来检验的特殊值。
- 当 $u \to 2y-x$ 时，第二个距离趋向 0：第二次相遇正好发生在撒克逊军队处，这就是与情况 (i) 的分界。

常见错误

- 把骑兵的回程写成 $s = -xt$ 或 $s = yt$，忘记把时间原点移到掉头的一刻。
- 求撒克逊骑兵的掉头位置时，用了诺曼军队在 $t = 0$ 的位置，而不是他到达那一刻的位置。
- 在最后一部分只说“他们不会相遇”。在两个情况中，他们其实都有第二次相遇，只是两人同向而行：一人从后越过另一人。

STEP 之外

这类图像称为世界线（world lines），是狭义相对论（special relativity）的基本工具。在那里，运动中的观察者之间来回发送的光信号，正是用这种方法分析的（雷达法），而直线的斜率则记录了相对速度。这里的代数运算，即把 u 加到一个速率上、又从另一个速率中减去，就是伽利略（Galileo）式的相对速度。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对伽利略图景和相对论图景都有论述。

详细解答

STEP I 2007，第 11 题

力学 • 抛体运动

STEP I 2007，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：

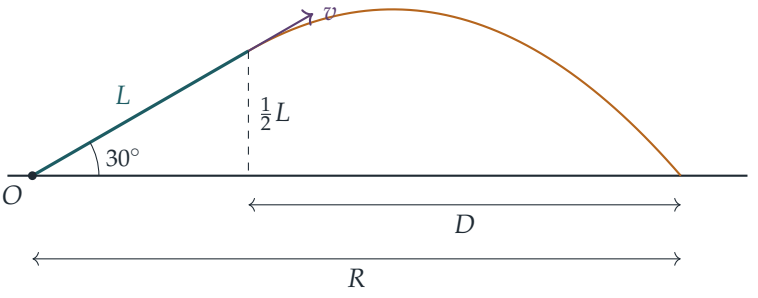
step.maths.org/questions/07-s1-q11

关键思路

把运动分成两个阶段。在光滑（smooth）管内，沿管方向唯一的力就是重力的分量，所以用能量（energy）（或等加速度公式（constant-acceleration formulae））便可求出离管时的速率。之后它就是一个普通的抛体（projectile），从高于终点水平 $\frac{1}{2}L$ 的一点以 30° 射出。消去飞行时间便得到关于 D 的二次方程；对它作隐函数微分，就可求得 $\mathrm{d}D/\mathrm{d}L$，完全不必解出 D。

解答

我们把粒子（particle）看成质点（point mass），把管看成光滑的（因此法向反作用力（normal reaction）不作功，沿管方向亦没有分量），并忽略空气阻力（air resistance）。



在管内

沿管方向分解力，唯一的分量是沿斜面向下的 $mg\sin 30^\circ = \frac{1}{2}mg$，所以减速度（deceleration）为 $\frac{1}{2}g$。经过距离 L 后，

$$v^2 = u^2 - 2\cdot\tfrac{1}{2}g\cdot L = u^2 - gL,$$

其中 v 是离管时的速率。（粒子要能离开管，必须有 $u^2 > gL$。）管的顶端比下端高 $L\sin 30^\circ = \frac{1}{2}L$，与下端的水平距离为 $L\cos 30^\circ = \frac{\sqrt{3}}{2}L$。

抛体阶段

粒子沿管的方向离开，所以它的速度的水平分量为 $\frac{\sqrt{3}}{2}v$，垂直分量为向上的 $\frac{1}{2}v$。以离管点为原点量度，经过时间 t 后

$$x = \tfrac{\sqrt{3}}{2}vt, \qquad y = \tfrac{1}{2}vt - \tfrac{1}{2}gt^2.$$

当 $y = -\frac{1}{2}L$ 时，粒子回到下端的水平，此刻 $x = D$。由第一式得 $t = \dfrac{2D}{\sqrt{3}\,v}$，代入第二式，

$$\frac{D}{\sqrt{3}} - \frac{2gD^2}{3v^2} = -\frac{L}{2}.$$

全式乘以 $-6v^2$，

$$4gD^2 - 2\sqrt{3}\,v^2D - 3Lv^2 = 0,$$

再代入 $v^2 = u^2 - gL$，

$$4gD^2 - 2\sqrt{3}\,(u^2 - gL)D - 3L(u^2 - gL) = 0,$$

这就是所求的方程。（它的两个根之中，D 是正的那一个：两根之积为 $-3Lv^2/(4g) < 0$。）

导数

当 u 固定时，D 取决于 L，所以我们把方程对 L 求导，凡同时含有两者的项都用积法则（product rule）：

$$8gD\frac{\mathrm{d}D}{\mathrm{d}L} - 2\sqrt{3}(u^2 - gL)\frac{\mathrm{d}D}{\mathrm{d}L} + 2\sqrt{3}\,gD - 3(u^2 - gL) + 3gL = 0.$$

最后两项合并为 $-3(u^2 - 2gL)$。把含 $\mathrm{d}D/\mathrm{d}L$ 的项集中起来，

$$\left(8gD - 2\sqrt{3}(u^2 - gL)\right)\frac{\mathrm{d}D}{\mathrm{d}L} = -\left(2\sqrt{3}\,gD - 3(u^2 - 2gL)\right),$$

所以

$$\frac{\mathrm{d}D}{\mathrm{d}L} = -\frac{2\sqrt{3}\,gD - 3(u^2 - 2gL)}{8gD - 2\sqrt{3}(u^2 - gL)},$$

正如所求。

想一想 为甚么 $-3L(u^2-gL)$ 这一项会产生两部分?

因为 L 在其中出现了两次: 把 $-3Lu^2+3gL^2$ 求导得 $-3u^2+6gL$, 即 $-3(u^2-gL)+3gL$。忘记括号内的 L, 是漏掉答案中 $2gL$ 的最常见原因。

R 的驻值 (stationary value)

粒子最终相对于下端的水平位移, 等于管的水平跨度加上 D:

$$R=\tfrac{\sqrt3}{2}L+D,\qquad \text{所以}\qquad \frac{\mathrm{d}R}{\mathrm{d}L}=\frac{\sqrt3}{2}+\frac{\mathrm{d}D}{\mathrm{d}L}.$$

因此 $\mathrm{d}R/\mathrm{d}L=0$ 当且仅当 $\mathrm{d}D/\mathrm{d}L=-\frac{\sqrt3}{2}$, 即 (两边乘以分母, 下文会验证分母不为零)

$$2\sqrt3\,gD-3(u^2-2gL)=\tfrac{\sqrt3}{2}\big(8gD-2\sqrt3(u^2-gL)\big)=4\sqrt3\,gD-3(u^2-gL).$$

两边的 $3u^2$ 项互相抵消, 余下 $6gL-2\sqrt3\,gD=3gL$, 所以 $2\sqrt3\,gD=3gL$, 即

$$2D=L\sqrt3.$$

每一步都是可逆的, 所以当 $2D=L\sqrt3$ 时 $\mathrm{d}R/\mathrm{d}L=0$, 正如所求。

要求出 R, 把 $D=\frac{\sqrt3}{2}L$ 代入二次方程:

$$4g\cdot\tfrac34L^2-2\sqrt3(u^2-gL)\cdot\tfrac{\sqrt3}{2}L-3L(u^2-gL)=3gL^2-6L(u^2-gL)=0.$$

由于 $L>0$, 由此得 $gL=2(u^2-gL)$, 所以 $L=\dfrac{2u^2}{3g}$。(这时 $v^2=u^2-gL=\frac13u^2>0$, 所以粒子确实会离开管, 而分母 $8gD-2\sqrt3v^2=\frac{8\sqrt3}{3}u^2-\frac{2\sqrt3}{3}u^2=2\sqrt3\,u^2$ 不为零。) 最后

$$R=\tfrac{\sqrt3}{2}L+D=\sqrt3\,L=\frac{2\sqrt3\,u^2}{3g}=\frac{2u^2}{\sqrt3\,g}.$$

验算

- 当 $L=0$ 时, 二次方程给出 $D=\sqrt3u^2/(2g)$, 正是以 30° 发射时大家熟悉的射程 (range) $u^2\sin60^\circ/g$。
- 当 $L=u^2/g$ 时, 粒子刚好到达顶端而 $v=0$, 所以 $D=0$, 同样有 $R=\frac{\sqrt3}{2}u^2/g$。驻值 $2u^2/(\sqrt3g)\approx1.155\,u^2/g$ 出现在这两个情况之间, 而且比两者都大, 所以它是极大值。
- 数值上, 取 $u=g=1$, 让 L 扫过 $[0,1]$, 所得的最大 R 为 1.1547, 出现于 $L\approx0.6667$, 与 $2/\sqrt3$ 及 $L=\frac23$ 吻合; 直接逐步模拟运动, 亦能重现关于 D 的二次方程。

常见错误

- 把管内的减速度当作 g，而不是 $g\sin 30^\circ = \frac{1}{2}g$。
- 令 $y = 0$ 而不是 $y = -\frac{1}{2}L$：粒子的落点低于它的发射点。
- 写成 $R = D$，忘记了管本身的水平长度 $\frac{\sqrt{3}}{2}L$。

STEP 之外

不解方程 $F(D, L) = 0$ 而直接对它求导，就是隐函数微分法（implicit differentiation）；到了大学一年级的分析课，它会发展成隐函数定理（implicit function theorem）：$\mathrm{d}D/\mathrm{d}L = -F_L/F_D$，凡 $F_D \neq 0$ 之处皆成立，而这正是我们验证过的那个分母。力学方面，Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）是很好的下一步。

详细解答

STEP I 2007，第 12 题

概率与统计 · 条件概率

STEP I 2007，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/07-s1-q12

关键思路

抽第二粒糖果时，我们并不知道袋中的情况，因为那取决于之前发生了甚么事。所以我们要考虑条件概率（conditional probability）：按第二次抽取之前发生的所有情况分类，求出每种情况下抽到红色的机会，再用全概率定理（law of total probability）加起来。第 (ii) 部分有一条好用的捷径：一个袋刚收到一粒随机糖果后，袋中红色糖果的期望数目（expected number）很容易写出来，所以我们不必就转移的那粒糖果的颜色重复分类。

解答

第 (i) 部分

设 R_1 为第一粒糖果是红色的事件，R_2 为第二粒糖果是红色的事件。每粒糖果被首先抽中的可能性均等，所以

$$\mathrm{P}(R_1) = \frac{a}{N}.$$

至于 R_2，按第一粒糖果的颜色分类。若第一粒是红色，则余下 $N-1$ 粒糖果，其中 $a-1$ 粒是红色；若不是红色，则余下 $N-1$ 粒，其中 a 粒是红色。（这里 $N \geq 2$ 保证了第二粒糖果存在。）由全概率定理，

$$\begin{aligned}\mathrm{P}(R_2) &= \mathrm{P}(R_1)\,\mathrm{P}(R_2 \mid R_1) + \mathrm{P}(R_1')\,\mathrm{P}(R_2 \mid R_1') \\ &= \frac{a}{N}\cdot\frac{a-1}{N-1} + \frac{N-a}{N}\cdot\frac{a}{N-1} \\ &= \frac{a\big[(a-1)+(N-a)\big]}{N(N-1)} = \frac{a(N-1)}{N(N-1)} = \frac{a}{N}.\end{aligned}$$

所以 $\mathrm{P}(R_2) = \mathrm{P}(R_1)$，正是所求。（用对称性（symmetry）也可得出同一结论：由两粒不同糖果组成的每个有序对出现的可能性均等，所以第二粒糖果是某一粒指定糖果的机会，与第一粒是该糖果的机会相同。）

第 (ii) 部分

这里的“第一粒糖果”是第一次掷硬币后转移的那一粒，“第二粒糖果”是第二次掷硬币后选出的那一粒。设 R_1、R_2 分别为这两粒糖果是红色的事件。两次掷硬币互相独立（independent），而每次掷硬币的结果亦与选中哪粒糖果无关。

第一粒糖果。 按第一次掷硬币的结果分类，

$$\mathrm{P}(R_1) = p \cdot \frac{a}{N} + q \cdot \frac{b}{N} = \frac{pa + qb}{N}.$$

第二粒糖果。 两次掷硬币共有四种情况。

正面、正面。先从第 1 袋取出一粒糖果，然后再从第 1 袋取出另一粒。这正是把第 (i) 部分的情况套用于第 1 袋（有 N 粒糖果，其中 a 粒红色），所以第二粒糖果是红色的概率为 a/N。

正面、反面。一粒糖果由第 1 袋移到第 2 袋，第 2 袋现在有 $N+1$ 粒糖果；然后我们从第 2 袋抽取。第 2 袋有 $b+1$ 粒红色糖果的概率为 a/N，有 b 粒红色糖果的概率为 $1 - a/N$，所以

$$\mathrm{P}(R_2 \mid \mathrm{HT}) = \frac{a}{N} \cdot \frac{b+1}{N+1} + \left(1 - \frac{a}{N}\right) \frac{b}{N+1} = \frac{b + \frac{a}{N}}{N+1} = \frac{Nb + a}{N(N+1)}.$$

想一想 为甚么第二行可以化简得如此干净俐落?

分子就是转移之后第 2 袋中红色糖果的期望数目：原有的 b 粒，加上新来的糖果是红色的机会 a/N。由于在这种情况下第 2 袋的大小固定为 $N+1$，抽到红色的机会就是这个期望数目除以 $N+1$。

反面、正面。把两个袋对调，用同样的论证可知，第 1 袋现在有 $N+1$ 粒糖果，其中红色糖果的期望数目为 $a + b/N$，所以

$$\mathrm{P}(R_2 \mid \mathrm{TH}) = \frac{Na + b}{N(N+1)}.$$

反面、反面。把第 (i) 部分套用于第 2 袋，得概率为 b/N。

现在把四种情况合并，它们的概率分别为 p^2、pq、qp、q^2：

$$\begin{aligned}\mathrm{P}(R_2) &= p^2\frac{a}{N} + pq\,\frac{(Nb+a)+(Na+b)}{N(N+1)} + q^2\frac{b}{N}\\ &= p^2\frac{a}{N} + pq\,\frac{(N+1)(a+b)}{N(N+1)} + q^2\frac{b}{N}\\ &= \frac{p^2a + pqa + pqb + q^2b}{N}\\ &= \frac{pa(p+q)+qb(p+q)}{N} = \frac{pa+qb}{N},\end{aligned}$$

最后一步用了 $p+q=1$。这等于 $\mathrm{P}(R_1)$，正是我们要证明的结果。

验算

- 若 $p=1$，一切都在第 1 袋中发生，两粒糖果的答案 a/N 正好就是第 (i) 部分。若 $p=0$，则得到第 2 袋的第 (i) 部分。
- 若 $a=b=0$，则没有红色糖果，两个概率都是 0；若 $a=b=N$，则每粒糖果都是红色，两个概率都是 1。
- 取 $N=5$、$a=2$、$b=4$、$p=0.3$，公式给出 $(0.6+2.8)/5=0.68$。把整个过程模拟 200 000 次，第一粒糖果的结果为 0.679，第二粒为 0.681；而对所有较小的 N、a、b 逐一穷举计算，结果亦吻合。

常见错误

- 忘记了转移之后，收到糖果的袋有 $N+1$ 粒糖果而不是 N 粒，而送出糖果的袋则有 $N-1$ 粒。
- 在 HH 和 TT 两种情况中，把第二次抽取当作与第一次独立。其实两者并不独立，但第 (i) 部分告诉我们答案始终不变，所以应直接引用，而不必重新计算。

STEP 之外

这两部分都是同一个深刻想法的例子：如果你从瓮中抽取一件东西而不看抽到的是甚么，你对瓮的认识并没有改变，所以下一次抽取的分布（distribution）依然相同。严格来说，这一连串抽取是可交换的（exchangeable），而在第 (ii) 部分中，$p\times$（第 1 袋中红色所占比例）$+$ $q\times$（第 2 袋中红色所占比例）这个量，即下一粒抽出的糖果是红色的机会，是一个鞅（martingale）：它未来的期望值等于它现在的值。鞅是现代概率论的核心，由赌博到金融都有它的身影。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）是

认识条件期望（conditional expectation）和这些想法的好起点。

详细解答

STEP I 2007，第 13 题

概率与统计 • 条件概率

STEP I 2007，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/07-s1-q13

本题在 *Beyond STEP* 第 12.1 节精确地取条件中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/conditioning.html

关键思路

虽然圆片是逐个取出的，但每一组四个圆片出现的可能性都相同，所以大部分工夫都花在数组合（combination）上。真正值得留意的是各个条件。知道第一个取出的是哪一个圆片，或者只知道某个指定的圆片确实被取出，余下三个圆片都仍然是从剩下十个之中随机选出的。只知道有某个有号码的圆片被取出，则是性质完全不同的信息，而最后两部分正是为了显示这个分别而设。

解答

设 N 为取出的四个圆片之中有号码的圆片数目。由于袋子经过摇匀，每一个由四个不同圆片组成的有序序列出现的可能性都相同，因此每一个由四个圆片组成的无序集合出现的可能性也相同。这样的集合共有

$$\binom{11}{4} = 330$$

个。要组成一个恰好有 k 个有号码圆片的集合，须从 5 个有号码的圆片中选 k 个，再从 6 个空白圆片中选 $4-k$ 个，所以

$$\mathrm{P}(N=k) = \frac{\binom{5}{k}\binom{6}{4-k}}{330}.$$

第 (i) 部分

四个全部有号码，即 $k = 4$：

$$\mathrm{P}(N = 4) = \frac{\binom{5}{4}\binom{6}{0}}{330} = \frac{5}{330} = \frac{1}{66}.$$

换一个算法，按次序逐个抽取，$\frac{5}{11} \cdot \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} = \frac{1}{66}$。

第 (ii) 部分

设 T_1 为“3”号圆片最先被取出这个事件（event）。已知 T_1 发生，摇匀的袋子里现在有 10 个圆片，其中 4 个有号码、6 个空白，而接着取出的三个圆片是从这些圆片中随机选出的。四个圆片全部有号码，当且仅当这三个全部有号码，所以

$$\mathrm{P}(N = 4 \mid T_1) = \frac{\binom{4}{3}}{\binom{10}{3}} = \frac{4}{120} = \frac{1}{30}.$$

第 (iii) 部分

仍然已知 T_1 发生，我们需要从余下十个圆片抽出的三个之中，恰好再有一个有号码的圆片：

$$\mathrm{P}(N = 2 \mid T_1) = \frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} = \frac{4 \times 15}{120} = \frac{1}{2}.$$

第 (iv) 部分

现在设 T 为“3”号圆片在某一次被取出这个事件。

> ? **想一想** “3”号圆片是最先取出，还是在其他某一次取出，有没有分别?

没有分别。已知 T 发生，取出的四个圆片就是“3”号圆片连同另外三个，而根据对称性（symmetry），从其余十个圆片中选出三个的每一个集合，出现的可能性都相同。这正是第 (iii) 部分的情况，当中抽取的次序从来都无关重要。所以答案同样是 $\frac{1}{2}$。

直接验证如下：$\mathrm{P}(T) = \binom{10}{3}/\binom{11}{4} = \frac{120}{330} = \frac{4}{11}$，而包含“3”号圆片、另外一个有号码圆片和两个空白圆片的集合共有 $\binom{4}{1}\binom{6}{2} = 60$ 个，所以

$$\mathrm{P}(N = 2 \mid T) = \frac{60/330}{120/330} = \frac{1}{2}.$$

第 (v) 部分

设 F 为第一个圆片有号码这个事件。按第一个圆片上的号码把 F 分拆：F 是“j 号圆片最先被取出”这五个事件（$j = 1, \ldots, 5$）的不相交并集（disjoint union），每个事件的概率

都是 $\frac{1}{11}$。第 (iii) 部分的论证并没有用到号码是 3 这一点，所以对每个 j 都有

$$\mathrm{P}(N=2 \mid j \text{ 号圆片最先}) = \tfrac{1}{2}.$$

因此，根据全概率公式（law of total probability），

$$\mathrm{P}(N=2 \mid F) = \sum_{j=1}^{5} \frac{1/11}{5/11} \cdot \frac{1}{2} = \frac{1}{2}.$$

（直接的做法：已知 F 发生，其余三个圆片是从 4 个有号码和 6 个空白的圆片中随机抽出的三个，这又回到第 (iii) 部分。）

第 (vi) 部分

设 A 为最少有一个有号码的圆片被取出这个事件，即 A 就是事件 $N \geq 1$。这与之前的信息并不属于同一类：它没有指明某一个圆片，再让其余的保持随机。我们改为使用条件概率（conditional probability）的定义。由于 $N=2$ 蕴含 A，

$$\mathrm{P}(N=2 \mid A) = \frac{\mathrm{P}(N=2)}{\mathrm{P}(N \geq 1)}.$$

> **想一想** 为甚么不可以直接说“已知有一个有号码的圆片，所以其余三个是从剩下十个之中随机抽出的”？

因为我们不知道已知的是哪一个圆片。一个有两个有号码圆片的集合，与一个有四个的集合，同样满足 A；并没有哪一个圆片可以让我们单独抽出来放在一旁。所以这个条件是按 N 本身的值为各个集合重新加权，我们必须老老实实地数。

我们有

$$\mathrm{P}(N=2) = \frac{\binom{5}{2}\binom{6}{2}}{330} = \frac{10 \times 15}{330} = \frac{150}{330}, \qquad \mathrm{P}(N=0) = \frac{\binom{6}{4}}{330} = \frac{15}{330},$$

所以 $\mathrm{P}(N \geq 1) = \frac{315}{330}$，而

$$\mathrm{P}(N=2 \mid A) = \frac{150}{315} = \frac{10}{21}.$$

验算

N 的分布（distribution）加起来应该正确：当 $k = 0, \ldots, 4$ 时，$\binom{5}{k}\binom{6}{4-k}$ 的数目分别是 $15, 100, 150, 60, 5$，总和为 330。在第 (ii) 和第 (iii) 部分，其余有号码圆片数目的条件分布，在 120 个之中的数目分别是 $20, 60, 36, 4$，总和同样正确。我们亦以精确分数列出全部 $11 \times 10 \times 9 \times 8 = 7920$ 个有序抽取结果，核对了全部六个答案（$\frac{1}{66}$、$\frac{1}{30}$、$\frac{1}{2}$、$\frac{1}{2}$、$\frac{1}{2}$、$\frac{10}{21}$）。留意 $\frac{10}{21} \approx 0.476$ 略小于 $\frac{1}{2}$，这是合理的：以 $N \geq 1$ 为条件只剔除了 $N=0$ 的情况，所以与知道某个指定圆片被取出相比，它在 $N=1$ 上留下了较多的权重。

常见错误

- 把第 (iv) 和第 (vi) 部分当作同一类条件。“3 号圆片被取出”指明了一个圆片；“有一个有号码的圆片被取出”则没有，两者的答案并不相同。
- 在第 (vi) 部分，忘记了条件是 $N \geq 1$，反而除以 $\mathrm{P}(N = 1)$ 或 $\frac{5}{11}$。
- 分子数的是有序抽取，分母数的却是无序集合。两者选定其一，然后贯彻始终。

STEP 之外

第 (v) 和第 (vi) 部分之间的对比，与著名的“男孩还是女孩”悖论（“boy or girl” paradox）是同一个陷阱：得知某个指定的孩子是男孩，与得知最少有一个是男孩，会得出不同的答案。在大学里，这一点会有严格的表述：以一个事件为条件，就是把概率测度（probability measure）限制在该事件上再重新归一化（renormalise），而这里的数目正是超几何分布（hypergeometric distribution）。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门），以及 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论），都仔细讨论了这些悖论。

详细解答

STEP I 2007，第 14 题

概率与统计 · 泊松参数的唯一性

> **STEP I 2007，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/07-s1-q14

关键思路

每一项资料都给出一条形如 $h(\lambda) = $ 常数 的方程，其中 h 是平均值（mean）的某个函数。这条方程能否唯一地确定 λ，完全取决于 h 在 $\lambda > 0$ 上的形状：严格单调（strictly monotonic）的 h 取每个值最多一次，而先升后降的 h 则会取大部分的值两次，只有最大值才取一次。因此，整道题的重点就是找出每个函数在哪里单调、在哪里达到顶点。

解答

全题中，$\lambda > 0$（泊松分布（Poisson distribution）的平均值必为正数），而且 $\mathrm{P}(X = k) = \mathrm{e}^{-\lambda}\lambda^k/k!$，其中 $k = 0, 1, 2, \ldots$

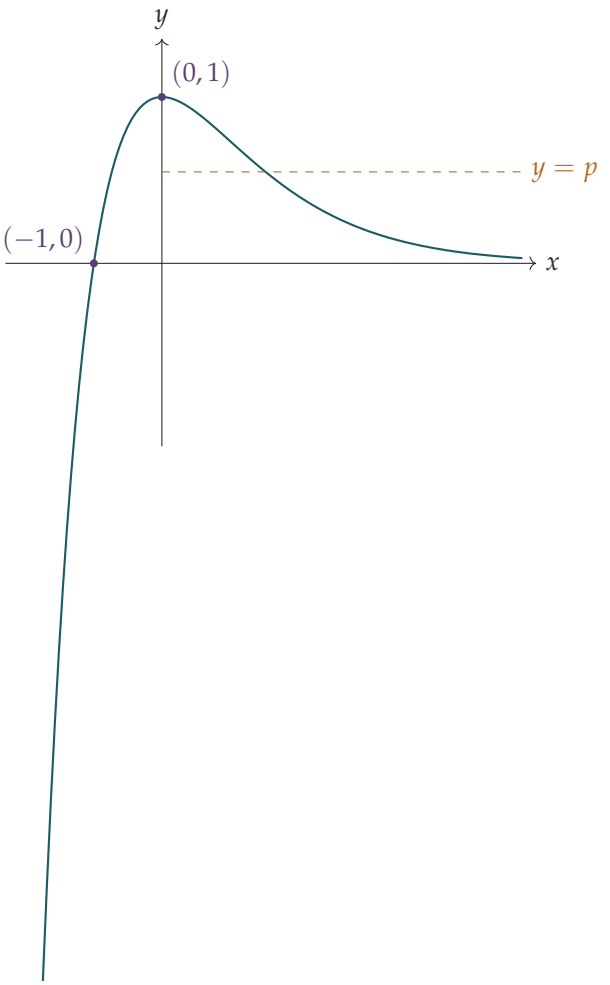
第 (i) 部分

设 $f(x) = (x+1)\mathrm{e}^{-x}$。由乘积法则（product rule），

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

由于 $\mathrm{e}^{-x} > 0$，$f'(x)$ 的正负号与 x 相反：f 在 $x < 0$ 时递增，在 $x > 0$ 时递减。所以唯一的转向点（turning point）是位于 $(0, 1)$ 的极大点，这也正是曲线与 y 轴相交之处。曲线与 x 轴相交于 $x + 1 = 0$ 之处，即 $(-1, 0)$。

至于两端的情况：当 $x \to \infty$ 时，e^{-x} 压倒线性因子，$f(x) \to 0$，并且是从上方趋近（x 轴是一条渐近线（asymptote））；当 $x \to -\infty$ 时，两个因子的大小都不断增长，$f(x) \to -\infty$。



现在考虑概率。我们有

$$\mathrm{P}(X \geq 2) = 1 - \mathrm{P}(X = 0) - \mathrm{P}(X = 1) = 1 - \mathrm{e}^{-\lambda} - \lambda \mathrm{e}^{-\lambda} = 1 - f(\lambda),$$

所以题目所给的资料恰好等于说

$$f(\lambda) = (1 + \lambda)\mathrm{e}^{-\lambda} = p.$$

在 $\lambda > 0$ 上，函数 f 连续（continuous）而且严格递减（因为 $f'(\lambda) = -\lambda \mathrm{e}^{-\lambda} < 0$），由 $f(0) = 1$ 开始下降，当 $\lambda \to \infty$ 时趋向 0。所以它取 $(0, 1)$ 内的每个值恰好一次：由中间值定理（intermediate value theorem）可知最少有一个解，而严格单调性则排除了第二个解。在略图中，当 $0 < p < 1$ 时，水平线 $y = p$ 与右边的分支恰好相交一次。因此 λ 被唯一确定。

第 (ii) 部分

这里 $\mathrm{P}(X=1)=\lambda \mathrm{e}^{-\lambda}=q$。记 $g(\lambda)=\lambda \mathrm{e}^{-\lambda}$，则

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

所以 g 在 $(0,1]$ 上严格递增，在 $[1,\infty)$ 上严格递减，最大值为 $g(1)=\mathrm{e}^{-1}$。此外，当 $\lambda\to 0^+$ 及当 $\lambda\to\infty$ 时，都有 $g(\lambda)\to 0$。（事实上 $g(\lambda)=\mathrm{e}^{-1}f(\lambda-1)$，所以 g 的图像就是第 (i) 部分的曲线向右平移一个单位，再沿垂直方向按 e^{-1} 的比例缩放。）

? **想一想** 一条水平线与一条先升至顶点、然后下降的曲线，可以相交多少次?

每一段单调的部分最多被击中一次，所以：

- 若 $0<q<\mathrm{e}^{-1}$，则在 $(0,1)$ 内有一个解，在 $(1,\infty)$ 内有另一个解，所以 λ 有两个可能的值；
- 若 $q>\mathrm{e}^{-1}$，则没有解；
- 若 $q=\mathrm{e}^{-1}$，则唯一的解就是顶点本身，即 $\lambda=1$。

所以这项资料只对恰好一个 q 值确定唯一的 λ，即

$$q=\mathrm{e}^{-1},\qquad \lambda=1.$$

第 (iii) 部分

根据条件概率（conditional probability）的定义，并且由于事件 $\{X=1\}$ 包含于 $\{X\le 2\}$ 之内，

$$\mathrm{P}(X=1\mid X\le 2)=\frac{\mathrm{P}(X=1)}{\mathrm{P}(X\le 2)}=\frac{\lambda \mathrm{e}^{-\lambda}}{\mathrm{e}^{-\lambda}\left(1+\lambda+\frac{1}{2}\lambda^2\right)}=\frac{2\lambda}{2+2\lambda+\lambda^2}.$$

把它记作 $k(\lambda)$。将分子和分母同时除以 λ（由于 $\lambda>0$，这是容许的），

$$k(\lambda)=\frac{2}{\lambda+\dfrac{2}{\lambda}+2}.$$

? **想一想** 当 $\lambda>0$ 时，$\lambda+2/\lambda$ 的最小值是多少?

由算术平均—几何平均不等式（AM–GM inequality），$\lambda+\dfrac{2}{\lambda}\ge 2\sqrt{\lambda\cdot\dfrac{2}{\lambda}}=2\sqrt{2}$，等号成立当且仅当 $\lambda=2/\lambda$，即 $\lambda=\sqrt{2}$。再者，$\lambda+2/\lambda$ 在 $(0,\sqrt{2}]$ 上严格递减，在 $[\sqrt{2},\infty)$ 上严格递增（它的导数 $1-2/\lambda^2$ 只在 $\sqrt{2}$ 处变号），而且在两端都趋向 ∞。所以 k 与第

(ii) 部分的 g 一样，具有先升后降的形状，其最大值为

$$k(\sqrt{2}) = \frac{2}{2\sqrt{2}+2} = \frac{1}{\sqrt{2}+1} = \sqrt{2}-1.$$

与之前完全一样：当 $0 < r < \sqrt{2}-1$ 时，λ 有两个值；当 $r > \sqrt{2}-1$ 时，没有任何值；当 $r = \sqrt{2}-1$ 时，恰好有一个值。所以 λ 只在以下情况被唯一确定：

$$r = \sqrt{2}-1, \qquad \lambda = \sqrt{2}.$$

另一方法：在 $k(\lambda) = r$ 中消去分母，得到二次方程 $r\lambda^2 - 2(1-r)\lambda + 2r = 0$。它的两根之积为 $2 > 0$，两根之和为 $2(1-r)/r > 0$，所以若有实根，两根必定都是正数；方程恰好有一个根，当且仅当判别式（discriminant）$4\left[(1-r)^2 - 2r^2\right] = 4(1-2r-r^2)$ 等于零，由此得 $r = \sqrt{2}-1$（取正根），以及重根 $\lambda = (1-r)/r = \sqrt{2}$。

验算

- 第 (ii) 部分：$\lambda = 1$ 给出 $\mathrm{P}(X=1) = \mathrm{e}^{-1} \approx 0.3679$；以数值方法在 $0 < \lambda < 40$ 的范围内扫描 $\lambda \mathrm{e}^{-\lambda}$，其值从不超过这个数，而 $q = 0.3$ 则给出两个交点。
- 第 (iii) 部分：取 $\lambda = \sqrt{2}$，则 $\dfrac{2\sqrt{2}}{2+2\sqrt{2}+2} = \dfrac{\sqrt{2}}{2+\sqrt{2}} = \sqrt{2}-1 \approx 0.4142$，数值扫描亦确认这就是最大值；$r = 0.35$ 给出 λ 的两个值，$r = 0.45$ 则没有任何值。
- 合理性检查：当 $\lambda \to 0$ 时，几乎全部概率质量都集中在 0；当 $\lambda \to \infty$ 时，在已知 $X \le 2$ 的条件下，概率质量堆积在 2。所以在这两个极限情况下都有 $\mathrm{P}(X=1 \mid X \le 2) \to 0$，与公式所示一致。

常见错误

- 忘记 $\lambda > 0$：在第 (i) 部分，曲线在 $x < 0$ 时同样存在，但只有 $x > 0$ 的分支才与概率有关。
- 在第 (i) 部分声称解是唯一的，却没有说明原因：你既要说明解存在（连续性，以及函数值由 1 一直下降至 0），也要说明没有第二个解（严格单调性）。
- 在第 (iii) 部分使用 $\mathrm{P}(X=1 \text{ 及 } X \le 2) = \mathrm{P}(X=1)\,\mathrm{P}(X \le 2)$；这两个事件并非独立（independent），而是 $\{X=1\}$ 包含于 $\{X \le 2\}$ 之内。

STEP 之外

这是初尝*可识别性*（identifiability）的滋味：在统计学中，我们会问一个观察得到的量能否确定模型的参数，而答案是肯定的，当且仅当由参数到可观察量的映射是一对一（one-

to-one）的。在这里，$\mathrm{P}(X \geq 2)$ 能识别 λ，但 $\mathrm{P}(X = 1)$ 和 $\mathrm{P}(X = 1 \mid X \leq 2)$ 一般都不能。当你要求似然（likelihood）的最大值时，同样的先升后降形状会再次出现：$\lambda \mathrm{e}^{-\lambda}$ 就是在看到单一观察值 $X = 1$ 之后 λ 的似然，而它在 $\lambda = 1$ 处的顶点就是最大似然估计（maximum likelihood estimate）。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门）写得生动活泼，是进一步了解泊松分布和条件概率的好去处。

详细解答

STEP II 2007，第 1 题

纯数学 · 二项式近似

STEP II 2007，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/07-s2-q1

关键思路

二项式级数（binomial series）只有在那个“小量”真的很小时才会快速收敛（converge），所以整道题的关键，就是把根式（surd）改写成 $\sqrt{1+\text{某个极小的数}}$ 的简单有理数倍。为此，我们要找一个完全平方数（或完全立方数），使它乘以 3 或 6 之后刚好落在 10 的某个幂旁边。这样 k 相对于 100（或 1000）就很小，只需寥寥数项便能达到所需的准确度。

解答

第 (i) 部分

设 $x=\frac{k}{100}$，$(1+x)^{1/2}$ 的二项式级数的系数为

$$\binom{\frac12}{1}=\frac12,\qquad \binom{\frac12}{2}=\frac{\frac12\cdot(-\frac12)}{2!}=-\frac18,\qquad \binom{\frac12}{3}=\frac{\frac12\cdot(-\frac12)\cdot(-\frac32)}{3!}=\frac{1}{16}.$$

代入 $x=k/100$，得

$$\left(1+\frac{k}{100}\right)^{1/2}\approx 1+\frac{k}{200}-\frac{k^2}{80\,000}+\frac{k^3}{16\,000\,000}.$$

(a) 取 $k=8$，左方就是 $\sqrt{1.08}$，而关键在于 $108=3\times 36$：

$$\sqrt{1.08}=\frac{\sqrt{108}}{10}=\frac{6\sqrt3}{10}=\frac{3\sqrt3}{5},\qquad \text{所以}\qquad \sqrt3=\frac53\sqrt{1.08}.$$

由级数可得

$$\sqrt{1.08}\approx 1+0.04-0.0008+0.000032=1.039232,$$

因此

$$\sqrt{3} \approx \frac{5}{3} \times 1.039232 = 1.7320533\ldots \approx 1.73205\text{(准确至小数点后五位)。}$$

(b)

? **想一想** k 取甚么值，才会令 $1 + \frac{k}{100}$ 等于某个简单平方数的 6 倍?

我们希望 $100 + k = 6m^2$，其中 m 为整数，而且 $6m^2$ 接近 100，这样 k 才会小。由于 $6 \times 16 = 96$，取 $m = 4$ 及 $k = -4$:

$$\sqrt{0.96} = \frac{\sqrt{96}}{10} = \frac{4\sqrt{6}}{10} = \frac{2\sqrt{6}}{5}, \qquad \text{所以} \qquad \sqrt{6} = \frac{5}{2}\sqrt{0.96}.$$

当 $k = -4$ 时，首项之后的每一项都是负数:

$$\sqrt{0.96} \approx 1 - 0.02 - 0.0002 - 0.000004 = 0.979796,$$

于是

$$\sqrt{6} \approx \frac{5}{2} \times 0.979796 = 2.44949\text{(准确至小数点后五位)。}$$

这个 k 的大小只有 (a) 中的一半，所以这里被略去的各项更加微小。(例如由 $1.5 = 6/4$ 得出的 $k = 50$，同样可以准确地表示 $\sqrt{6}$，但当 $x = \frac{1}{2}$ 时，级数收敛得太慢，远远不足以准确至小数点后五位。)

第 (ii) 部分

取首两项，得

$$\left(1 + \frac{k}{1000}\right)^{1/3} \approx 1 + \frac{k}{3000}.$$

? **想一想** 哪一个立方数乘以 3 之后，刚好略大于 1000?

我们需要 $1000 + k = 3m^3$，而且 k 要小。在 $1000/3 \approx 333$ 附近的立方数有 $6^3 = 216$ 和 $7^3 = 343$，而 $3 \times 343 = 1029$ 非常接近 1000。所以取 $k = 29$:

$$\sqrt[3]{1.029} = \frac{\sqrt[3]{3 \times 7^3}}{10} = \frac{7\sqrt[3]{3}}{10}, \qquad \text{所以} \qquad \sqrt[3]{3} = \frac{10}{7}\sqrt[3]{1.029}.$$

利用两项的近似式，

$$\sqrt[3]{3} \approx \frac{10}{7}\left(1 + \frac{29}{3000}\right) = \frac{10}{7} \cdot \frac{3029}{3000} = \frac{3029}{2100},$$

这就是所求的近似值。

验算

用计算机可得 $\sqrt{3} = 1.7320508\ldots$ 及 $\sqrt{6} = 2.4494897\ldots$，两者都与上面的答案在小数点后五位之内吻合。至于第 (ii) 部分，$3029/2100 = 1.44238\ldots$，而 $\sqrt[3]{3} = 1.44225\ldots$。误差约为 1.3×10^{-4}，与第一个被略去的项的大小 $\frac{10}{7} \cdot \frac{1}{9}\left(\frac{29}{1000}\right)^2 \approx 1.3 \times 10^{-4}$ 相符，所以这个近似值略为高估了真实值。

常见错误

- 算错第三个系数：$\frac{1}{2} \cdot \left(-\frac{1}{2}\right) \cdot \left(-\frac{3}{2}\right)/6 = +\frac{1}{16}$，而不是 $-\frac{1}{16}$。
- 计算级数之后，忘记乘回 $\frac{5}{3}$、$\frac{5}{2}$ 或 $\frac{10}{7}$。
- 选了一个很大的 k（例如 (b) 中的 $k = 50$），这时只取三四项根本远远不够。

STEP 之外

究竟需要多少项，归根究柢是要为泰勒级数（Taylor series）的余项（remainder）定出上界。有了拉格朗日（Lagrange）余项公式，你就可以证明（而不只是相信）这些近似值确实准确至小数点后五位。把问题重新缩放到级数快速收敛的范围内，这个技巧正是计算机和电脑程序库实际计算方根和对数的方法。关于泰勒级数及其余项的理论，可参阅 Abbott 的 *Understanding Analysis*（一本实分析入门教材）。

详细解答

STEP II 2007，第 2 题

纯数学 · 三次曲线与面积

STEP II 2007，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/07-s2-q2

关键思路

三次曲线（cubic）转向点（turning point）的一切资料，都藏在它的导数（derivative）里。由于转向点位于 $x = p$ 和 $x = q$，导数必定是 $(x-p)(x-q)$ 的倍数，比较系数（comparing coefficients）便立即得出 b 和 c。之后，求 $m - n$ 最简洁的方法，是把导数在两个转向点之间积分（integrate）；而只要留意到 $y - n$ 在 $x = q$ 处有一个二重根（double root），面积也就容易计算了。

解答

第 (i) 部分

求导，得

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 6x^2 - 2bx + c.$$

这个二次式在转向点 $x = p$ 和 $x = q$ 处等于零，而它的首项系数是 6，所以

$$6x^2 - 2bx + c = 6(x-p)(x-q) = 6x^2 - 6(p+q)x + 6pq.$$

比较系数，得

$$b = 3(p+q), \qquad c = 6pq.$$

以下两项事实稍后会用到。第一，由于 x^3 的系数为正，$\mathrm{d}y/\mathrm{d}x = 6(x-p)(x-q)$ 在两根之外为正，在两根之间为负，所以曲线先升、后降、再升：极大点（maximum）在先，故 $p < q$。

第二，条件 $n > 0$ 对 q 有所限制。代入 $x = q$，

$$n = 2q^3 - 3(p+q)q^2 + 6pq^2 = q^2(3p - q),$$

所以 $n > 0$ 即 $q < 3p$。综合起来，$0 < p < q < 3p$。

第 (ii) 部分

曲线通过原点，而且

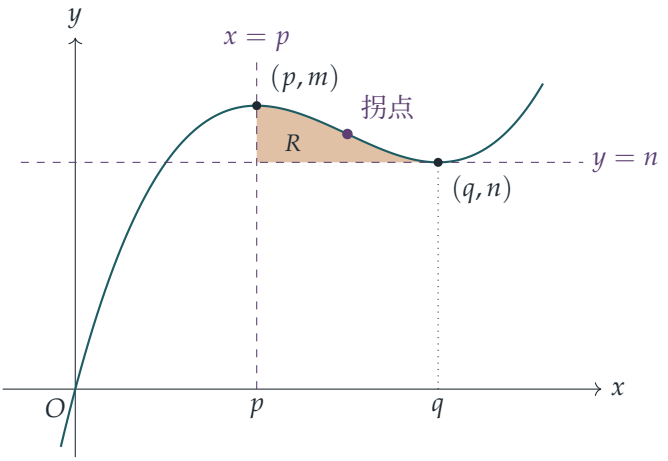
$$y = x\big(2x^2 - bx + c\big),$$

其中二次因式的判别式（discriminant）为

$$b^2 - 8c = 9(p+q)^2 - 48pq = 3(3p-q)(p-3q).$$

由于 $q < 3p$ 及 $p < q < 3q$，第一个括号为正，第二个括号为负，所以判别式为负，原点是曲线与 x 轴唯一的交点。因此曲线由 $-\infty$ 向上而来，在 O 穿过横轴，升至极大点 (p, m)，再降至极小点（minimum）(q, n)（此时仍在横轴上方），然后一直上升。

拐点（point of inflection）位于 $\mathrm{d}^2y/\mathrm{d}x^2 = 12x - 2b = 0$ 之处，即 $x = \frac{b}{6} = \frac{1}{2}(p+q)$，正好在两个转向点的正中间。



区域 R（阴影部分）位于 $x = p$ 与 $x = q$ 之间，在曲线之下、直线 $y = n$ 之上；曲线在极小点处与 $y = n$ 相切。

对称性。设 $s = \frac{1}{2}(p+q)$, $h = q - p$，并令 $x = s + t$，即从拐点起量度。于是 $x - p = t + \frac{1}{2}h$ 及 $x - q = t - \frac{1}{2}h$，所以

$$\frac{\mathrm{d}y}{\mathrm{d}t} = 6\big(t^2 - \tfrac{1}{4}h^2\big),$$

这是 t 的偶函数（even function）。因此它由 0 起的积分，即 $y - y(s) = 2t^3 - \frac{3}{2}h^2t$，是 t 的奇函数（odd function）。所以 $y(s+t) - y(s) = -\big(y(s-t) - y(s)\big)$：曲线具有绕其拐点的 2 阶旋转对称（*rotational symmetry*）（绕该点旋转半周，曲线与自身重合）。特别地，这个半周旋转把两个转向点互换，所以拐点是 (p, m) 与 (q, n) 的中点，即 $\big(\frac{1}{2}(p+q), \frac{1}{2}(m+n)\big)$。

第 (iii) 部分

由极大点到极小点的落差，就是斜率的积分：

$$m-n=y(p)-y(q)=-\int_p^q 6(x-p)(x-q)\,\mathrm{d}x.$$

> **想一想** 有没有不必展开就能快速计算 $\int_p^q (x-p)(x-q)\,\mathrm{d}x$ 的方法？

代入 $x=p+u$，于是 $x-q=u-h$，其中 $h=q-p$，而 u 由 0 变到 h：

$$\int_0^h u(u-h)\,\mathrm{d}u=\frac{h^3}{3}-\frac{h^3}{2}=-\frac{h^3}{6}.$$

因此

$$m-n=-6\Big(-\frac{(q-p)^3}{6}\Big)=(q-p)^3,$$

正如所求。

第 (iv) 部分

在 $p\le x\le q$ 上，曲线由 m 递减至 n，所以它位于 $y=n$ 之上或与之相接，而

$$\text{区域面积}R=\int_p^q (y-n)\,\mathrm{d}x.$$

三次式 $y-n=2x^3-bx^2+cx-n$ 在 $x=q$ 处等于零，其导数在该处亦等于零，所以 $x=q$ 是二重根，并且

$$y-n=2(x-q)^2(x-r)$$

对某个 r 成立。比较 x^2 的系数，$-2(2q+r)=-b=-3(p+q)$，所以 $r=\frac{1}{2}(3p-q)$。同样设 $h=q-p$，并令 $u=x-q$（由 $-h$ 变到 0），

$$x-r=u+q-\tfrac{1}{2}(3p-q)=u+\tfrac{3}{2}h.$$

因此

$$\begin{aligned}\text{区域面积}R&=\int_{-h}^0 2u^2\big(u+\tfrac{3}{2}h\big)\,\mathrm{d}u=\Big[\tfrac{1}{2}u^4+hu^3\Big]_{-h}^0\\&=-\big(\tfrac{1}{2}h^4-h^4\big)=\tfrac{1}{2}h^4=\tfrac{1}{2}(q-p)^4,\end{aligned}$$

这就是所求的结果。

验算

取 $p=1$，$q=2$（符合条件，因为 $q<3p$）。于是 $b=9$，$c=12$，而 $y=2x^3-9x^2+12x$ 满足 $y'=6(x-1)(x-2)$，且 $m=y(1)=5$ 及 $n=y(2)=4>0$。所以 $m-n=1=$

$(q-p)^3$。拐点是 (1.5, 4.5)，即两个转向点的中点；把 $y-4$ 在 $[1,2]$ 上作数值积分，得 $0.5000 = \frac{1}{2}(q-p)^4$。我们亦以 2000 对满足 $0 < p < q < 3p$ 的随机数值验证了全部四个结果。

常见错误

- 取 $q < p$：三次项系数为正时，极大点必定在极小点的左边。
- 积分时用了 y 而不是 $y-n$，这样会多算了直线 $y=n$ 之下、高度为 n 的长方形。
- 把对称性说成是对直线 $x = \frac{1}{2}(p+q)$ 的反射；三次曲线具有的是绕一点的半周旋转对称，而不是镜像对称。

STEP 之外

每一条三次曲线都绕其拐点成半周旋转对称，因为平移 x 以消去 x^2 项（即卡尔达诺（Cardano）公式中所用的“缺项三次式”（depressed cubic）代换）之后，余下的是一个奇函数加一个常数。第 (iv) 部分的二重根技巧，是相切与根的重数（multiplicity）理论的起点，这套理论在代数几何（algebraic geometry）中占有核心地位。Siklos 的 *Advanced Problems in Mathematics* 载有更多同类风格的题目。

详细解答

STEP II 2007，第 3 题

纯数学 · 代换积分法

STEP II 2007，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/07-s2-q3

关键思路

一切都建基于 $1/(a^2+x^2)$ 的标准积分（standard integral），所以整道题的重点，就是在代换（substitution）之后认出这个形状。在 (i)(a) 中，分子（numerator）$\cos x$ 正是 $\sin x$ 的导数（derivative），这是个很强的提示。在 (i)(b) 中，半角代换（half-angle substitution）把 $1+\sin^2 x$ 变成 t 的四次式（quartic），而关键在于看出同样的做法可以倒过来用：在 (ii) 中，我们要找出哪一个三角积分会产生新的四次式。

解答

标准积分

设 $a \neq 0$，并取 $x = a\tan\theta$，其中 $-\frac{1}{2}\pi < \theta < \frac{1}{2}\pi$。当 θ 走遍这个区间时，x 恰好取遍每个实数一次（不论 a 的正负），所以这个代换是有效的，而且由 $\tan\theta = x/a$ 可得 $\theta = \arctan(x/a)$。现在 $\mathrm{d}x = a\sec^2\theta\,\mathrm{d}\theta$，而

$$a^2 + x^2 = a^2(1+\tan^2\theta) = a^2\sec^2\theta,$$

所以

$$\int \frac{1}{a^2+x^2}\,\mathrm{d}x = \int \frac{a\sec^2\theta}{a^2\sec^2\theta}\,\mathrm{d}\theta = \int \frac{1}{a}\,\mathrm{d}\theta = \frac{\theta}{a} + \text{常数} = \frac{1}{a}\arctan\frac{x}{a} + \text{常数},$$

这正是所要证明的结果。

第 (i)(a) 部分

代入 $u = \sin x$，则 $\mathrm{d}u = \cos x \,\mathrm{d}x$，而当 x 由 0 变到 $\frac{1}{2}\pi$ 时，u 由 0 变到 1。利用标准积分并取 $a = 1$，

$$I = \int_0^1 \frac{1}{1+u^2}\,\mathrm{d}u = \Big[\arctan u\Big]_0^1 = \frac{\pi}{4}.$$

第 (i)(b) 部分

取 $t = \tan\frac{1}{2}x$，便有以下标准的半角结果

$$\sin x = \frac{2t}{1+t^2}, \qquad \cos x = \frac{1-t^2}{1+t^2}, \qquad \mathrm{d}x = \frac{2}{1+t^2}\,\mathrm{d}t,$$

最后一式是因为 $\mathrm{d}t/\mathrm{d}x = \frac{1}{2}\sec^2\frac{1}{2}x = \frac{1}{2}(1+t^2)$。积分上下限（limits）$x = 0$ 和 $x = \frac{1}{2}\pi$ 分别变成 $t = 0$ 和 $t = \tan\frac{1}{4}\pi = 1$。接着，

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

把各部分合起来，$1+t^2$ 的因子完全约去：

$$I = \int_0^1 \frac{1-t^2}{1+t^2}\cdot\frac{(1+t^2)^2}{1+6t^2+t^4}\cdot\frac{2}{1+t^2}\,\mathrm{d}t = 2\int_0^1 \frac{1-t^2}{1+6t^2+t^4}\,\mathrm{d}t.$$

两边除以 2，便得 $\displaystyle\int_0^1 \frac{1-t^2}{1+6t^2+t^4}\,\mathrm{d}t = \frac{1}{2}I$，这正是所要证明的结果。（由 (i)(a)，它的值是 $\pi/8$。）

第 (ii) 部分

> **想一想** 四次式 $1+14t^2+t^4$ 可能是从哪里来的?

在 (i)(b) 中，$6t^2$ 的来源是：把 $1+t^2$ 平方得到的 $2t^2$，再加上由 $\sin^2 x$ 得到的 $4t^2 = (2t)^2$。要得到 $14t^2$，我们需要额外的 $12t^2 = 3(2t)^2$，也就是以 $3\sin^2 x$ 取代 $\sin^2 x$。事实上，取 $t = \tan\frac{1}{2}x$，

$$1 + 3\sin^2 x = \frac{(1+t^2)^2 + 12t^2}{(1+t^2)^2} = \frac{1+14t^2+t^4}{(1+t^2)^2}.$$

因此，把 (i)(b) 的计算原封不动地重做一次，只是以 $1+3\sin^2 x$ 取代 $1+\sin^2 x$，便得

$$\int_0^{\frac{1}{2}\pi} \frac{\cos x}{1+3\sin^2 x}\,\mathrm{d}x = 2\int_0^1 \frac{1-t^2}{1+14t^2+t^4}\,\mathrm{d}t.$$

仿照 (i)(a)，用 $u = \sin x$ 计算左边：

$$\int_0^1 \frac{1}{1+3u^2}\,\mathrm{d}u = \frac{1}{3}\int_0^1 \frac{1}{\frac{1}{3}+u^2}\,\mathrm{d}u = \frac{1}{3}\cdot\sqrt{3}\Big[\arctan\left(\sqrt{3}\,u\right)\Big]_0^1 = \frac{1}{\sqrt{3}}\cdot\frac{\pi}{3},$$

这里用了标准积分，并取 $a = 1/\sqrt{3}$。取其一半，

$$\int_0^1 \frac{1-t^2}{1+14t^2+t^4}\,\mathrm{d}t = \frac{\pi}{6\sqrt{3}} = \frac{\sqrt{3}\,\pi}{18}.$$

验算

用辛普森法则（Simpson's rule）可得 $I \approx 0.785398$（即 $\pi/4$），(i)(b) 的积分 ≈ 0.392699（即 $\pi/8$），而 (ii) 的积分 ≈ 0.302300，与 $\sqrt{3}\,\pi/18 \approx 0.302300$ 吻合。再做一个简单的合理性检查：两个四次式的被积函数（integrand）在 $t = 0$ 时都等于 1，在 $t = 1$ 时都等于 0，而分母含 $14t^2$ 的那个在整个区间上都较小，所以它的积分应该小于 $\pi/8 \approx 0.393$，事实也确是如此。

常见错误

- 忘记了来自 $\mathrm{d}x$ 的因子 $\frac{2}{1+t^2}$，以致漏掉 (i)(b) 中的因子 2。
- 写成 $\int \frac{\mathrm{d}u}{1+3u^2} = \arctan(\sqrt{3}u)$ 而漏掉因子 $\frac{1}{\sqrt{3}}$；应先把 3 提出来，或者代入 $v = \sqrt{3}u$。
- 代换之后仍然沿用 x 的上下限：$x = \frac{1}{2}\pi$ 对应的是 $t = 1$。

STEP 之外

还有一条直接的路径：把分子和分母同时除以 t^2，再取 $v = t + 1/t$，便可把 $\frac{1-t^2}{1+kt^2+t^4}$ 变成 $-\frac{1}{v^2+k-2}$，这又是一个 arctan 积分。这里带出的一般道理是：$\sin x$ 和 $\cos x$ 的有理函数（rational function），在代换 $t = \tan\frac{1}{2}x$ 之下总会变成 t 的有理函数，因而原则上总是可以积分的；这正是以初等函数表示积分的系统理论的起点。想做更多同类风格的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

STEP II 2007，第 4 题

纯数学 · 三角方程

STEP II 2007，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/07-s2-q4

关键思路

用复角公式（compound-angle formulae）把所有项展开，再除以 $\cos A \cos B$：方程便化成关于 $t = \tan A$ 和 $u = \tan B$ 的关系式，其中有一个 tu 项、一个 t 项、一个 u 项和一个常数项。这样的式子能因式分解（factorise）成两个一次括号之积，当且仅当“配成矩形”之后没有余项，而这个余项原来是 $\beta^2 + \gamma^2 - \alpha^2$ 的倍数。其后三条方程都有 $\alpha = 2$, $\beta = \sqrt{3}$, $\gamma = 1$，它们满足这个条件；唯一要做的工夫就是正确地选取 A 和 B。

解答

因式分解

由复角公式，

$$\begin{aligned}&\alpha \sin(A - B) + \beta \cos(A + B) - \gamma \sin(A + B)\\ &\quad = \alpha(\sin A \cos B - \cos A \sin B) + \beta(\cos A \cos B - \sin A \sin B)\\ &\qquad - \gamma(\sin A \cos B + \cos A \sin B).\end{aligned}$$

由于 $\cos A \cos B \neq 0$，我们可以除以它。记 $t = \tan A$, $u = \tan B$，则方程等价于 $P(t, u) = 0$，其中

$$P(t, u) = \alpha(t - u) + \beta(1 - tu) - \gamma(t + u) = -\beta tu + (\alpha - \gamma)t - (\alpha + \gamma)u + \beta.$$

由于 $\beta \neq 0$，设

$$m = -\frac{\alpha + \gamma}{\beta}, \qquad n = \frac{\alpha - \gamma}{\beta}.$$

于是 $-\beta(t-m)(u-n) = -\beta tu + \beta n\,t + \beta m\,u - \beta mn = -\beta tu + (\alpha-\gamma)t - (\alpha+\gamma)u + \dfrac{\alpha^2-\gamma^2}{\beta}$，除常数项外与 P 完全吻合。因此，对所有 t 和 u，

$$P(t,u) = -\beta(t-m)(u-n) + \frac{\beta^2+\gamma^2-\alpha^2}{\beta}. \qquad (*)$$

充分性。当 $\alpha^2 = \beta^2+\gamma^2$ 时，$(*)$ 中的常数项为零，所以（因为 $\beta \neq 0$）方程就是 $(\tan A - m)(\tan B - n) = 0$，其中 m 和 n 只取决于 α, β, γ。

必要性。假设方程可化为 $(t-m')(u-n') = 0$，其中 m', n' 为常数，即 $P(t,u) = 0$ 成立当且仅当 $t = m'$ 或 $u = n'$。记 $c = (\beta^2+\gamma^2-\alpha^2)/\beta$，并假设 $c \neq 0$。这时每一对 (m', u) 都是解，所以 $P(m', u) = -\beta(m'-m)(u-n) + c$ 对每个 u 都等于零。若 $m' \neq m$，这是 u 的一个非常数线性函数（linear function），不会恒等于零；若 $m' = m$，它等于 $c \neq 0$，永不为零。无论哪种情况都得出矛盾，所以 $c = 0$，即 $\alpha^2 = \beta^2+\gamma^2$。（若把“可化为”理解成“$P$ 是 $(t-m)(u-n)$ 的常数倍”，比较 tu 的系数便知这个倍数必为 $-\beta$，然后 $(*)$ 同样迫使 $c = 0$。）

> **想一想** 在下面各部分中，如果 $\cos x = 0$，即我们不能作除法的时候，会发生甚么事?

把 $(*)$ 两边乘回 $\cos A \cos B$。当 $\alpha^2 = \beta^2+\gamma^2$ 时，我们得到

$$\alpha\sin(A-B) + \beta\cos(A+B) - \gamma\sin(A+B) = -\beta(\sin A - m\cos A)(\sin B - n\cos B), \qquad (**)$$

这时两边都只是正弦和余弦的组合，所以这个恒等式（identity）对所有 A 和 B 都成立（两边都是连续的，而且只要 $\cos A\cos B \neq 0$ 便相等）。使用 $(**)$，我们就毋须另外处理 $\cos A = 0$ 的情况。

在全部三个部分中，$\alpha = 2, \beta = \sqrt{3}, \gamma = 1$，而 $\alpha^2 = 4 = \beta^2+\gamma^2$。所以

$$m = -\frac{3}{\sqrt{3}} = -\sqrt{3}, \qquad n = \frac{1}{\sqrt{3}},$$

方程就是 $(\sin A + \sqrt{3}\cos A)(\sin B - \frac{1}{\sqrt{3}}\cos B) = 0$。留意 $\sin A + \sqrt{3}\cos A = 0$ 必然令 $\cos A \neq 0$（否则 $\sin A = 0$ 亦成立），所以它等同于 $\tan A = -\sqrt{3}$；同样地，第二个括号为零当且仅当 $\tan B = \frac{1}{\sqrt{3}}$。

第 (i) 部分

取 $A = x$ 及 $B = \frac{1}{4}\pi$，使得 $A - B = x - \frac{1}{4}\pi$ 及 $A + B = x + \frac{1}{4}\pi$。由于 $\tan\frac{1}{4}\pi = 1 \neq \frac{1}{\sqrt{3}}$，第二个括号不为零，所以我们需要 $\tan x = -\sqrt{3}$:

$$x = \tfrac{2}{3}\pi \quad 或 \quad x = \tfrac{5}{3}\pi.$$

第 (ii) 部分

现在 $A = x$ 及 $B = \frac{1}{6}\pi$。但 $\tan\frac{1}{6}\pi = \frac{1}{\sqrt{3}} = n$，所以第二个括号是 $\sin\frac{1}{6}\pi - \frac{1}{\sqrt{3}}\cos\frac{1}{6}\pi = \frac{1}{2} - \frac{1}{2} = 0$。由 $(**)$，左方减右方恒等于零：这条方程是一个恒等式，每一个满足 $0 \le x < 2\pi$ 的 x 都是解。

第 (iii) 部分

我们需要 $A - B = x + \frac{1}{3}\pi$ 及 $A + B = 3x$。两式相加和相减，得

$$A = 2x + \tfrac{1}{6}\pi, \qquad B = x - \tfrac{1}{6}\pi.$$

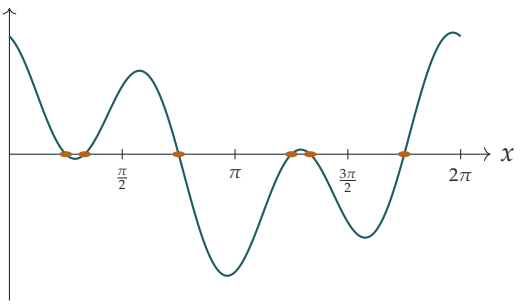
所以解由 $\tan(2x + \frac{1}{6}\pi) = -\sqrt{3}$ 或 $\tan(x - \frac{1}{6}\pi) = \frac{1}{\sqrt{3}}$ 给出。

对于前者，$2x + \frac{1}{6}\pi = \frac{2}{3}\pi + k\pi$，所以 $x = \frac{1}{4}\pi + \frac{1}{2}k\pi$，其中 k 为整数。在 $[0, 2\pi)$ 内，这给出 $x = \frac{1}{4}\pi,\ \frac{3}{4}\pi,\ \frac{5}{4}\pi,\ \frac{7}{4}\pi$。

对于后者，$x - \frac{1}{6}\pi = \frac{1}{6}\pi + k\pi$，所以 $x = \frac{1}{3}\pi + k\pi$，给出 $x = \frac{1}{3}\pi,\ \frac{4}{3}\pi$。

这六个值互不相同，所以解为

$$x = \tfrac{1}{4}\pi,\ \tfrac{1}{3}\pi,\ \tfrac{3}{4}\pi,\ \tfrac{5}{4}\pi,\ \tfrac{4}{3}\pi,\ \tfrac{7}{4}\pi.$$



(iii) 的左方减右方的图像，恰好在标示的六个点穿过零。

验算

- 在 (i) 中，当 $x = \frac{2}{3}\pi$ 时：$2\sin\frac{5}{12}\pi \approx 1.932$，$\sqrt{3}\cos\frac{11}{12}\pi \approx -1.673$，而 $\sin\frac{11}{12}\pi \approx 0.259$；的确 $1.932 - 1.673 = 0.259$。
- 在 (ii) 中，试 $x = 0$：$2\sin(-\frac{1}{6}\pi) + \sqrt{3}\cos\frac{1}{6}\pi = -1 + \frac{3}{2} = \frac{1}{2} = \sin\frac{1}{6}\pi$。其他任何 x 也同样成立。
- 在 (iii) 中，当 $x = \frac{1}{4}\pi$ 时：$2\sin\frac{7}{12}\pi \approx 1.932$，而 $\sin\frac{3}{4}\pi - \sqrt{3}\cos\frac{3}{4}\pi \approx 0.707 + 1.225 = 1.932$。在 $[0, 2\pi)$ 上以数值方法搜索根，找到的正好是上述六个根。

常见错误

- 在 (iii) 中取 $A = x + \frac{1}{3}\pi$，并由 $3x$ 得出 B：你必须联立求解 $A - B = x + \frac{1}{3}\pi$, $A + B = 3x$。
- 在 (iii) 中只取到 $k = 0, 1$ 便停下而漏掉解：$\tan(2x + \frac{1}{6}\pi)$ 的周期是 $\frac{1}{2}\pi$，所以该括号给出四个解。
- 没有察觉 (ii) 是恒等式，反而去找个别的解。

STEP 之外

式子 $-\beta tu + (\alpha - \gamma)t - (\alpha + \gamma)u + \beta$ 是圆锥曲线（conic）的一种退化情况：它的零点集是一条等轴双曲线（rectangular hyperbola），除非某个类似行列式（determinant）的量为零，这时它便分裂成两条直线。判断一个二次型（quadratic form）何时可以因式分解，是用矩阵和特征值（eigenvalues）为圆锥曲线分类的第一步，这是大学一年级线性代数（linear algebra）的课题；见 Strang, *Introduction to Linear Algebra*（线性代数入门教科书）。Siklos, *Advanced Problems in Mathematics*（进阶数学难题集）中有更多同类风格的三角方程。

详细解答

STEP II 2007，第 5 题

纯数学 · 迭代函数

STEP II 2007，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/07-s2-q5

关键思路

$\dfrac{x+\sqrt{3}}{1-\sqrt{3}\,x}$ 这个形状，正是 $\tan(A+B)$ 的复角公式（compound-angle formula），其中 $\tan B=\sqrt{3}$，即 $B=\frac{\pi}{3}$。所以 f 的作用不过是把角度加上 $\frac{\pi}{3}$；而 tan 的周期（period）是 π，因此走三步便回到起点。第 (iii) 部分可以用同一招，只是换成 sin 和 $\frac{\pi}{6}$，但当中有个陷阱：$\sqrt{1-t^2}$ 只有在角度位于 $[-\frac{\pi}{2},\frac{\pi}{2}]$ 之内时才是余弦，而这一点会改变答案。

解答

第 (i) 部分

设 $y=f(x)=\dfrac{x+\sqrt{3}}{1-\sqrt{3}\,x}$。那么 $f^2(x)=\dfrac{y+\sqrt{3}}{1-\sqrt{3}\,y}$，把分子和分母同乘以 $1-\sqrt{3}\,x$，得

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

现在再应用 f 一次，这次取 $y=f^2(x)$，并把分子和分母同乘以 $1+\sqrt{3}\,x$：

$$f^3(x)=\frac{(x-\sqrt{3})+\sqrt{3}(1+\sqrt{3}\,x)}{(1+\sqrt{3}\,x)-\sqrt{3}(x-\sqrt{3})}=\frac{4x}{4}=x.$$

所以 f^3 是恒等映射（identity）。（题目排除的那两个点，正好保证每一步都有定义：$f(x)$ 需要 $x\neq 1/\sqrt{3}$，$f^2(x)$ 需要 $x\neq -1/\sqrt{3}$，而 f 把 $\mathbb{R}\setminus\{\pm 1/\sqrt{3}\}$ 映满到它自身，所以我们可以一直做下去。）

由于 $2007=3\times 669$，

$$f^{2007}(x)=\left(f^3\right)^{669}(x)=x.$$

第 (ii) 部分

设 $x = \tan\theta$。对任何使 $\tan\alpha$ 有定义且 $1 - \sqrt{3}\tan\alpha \neq 0$ 的角 α，利用 $\tan\frac{\pi}{3} = \sqrt{3}$ 和复角公式可得

$$f(\tan\alpha) = \frac{\tan\alpha + \tan\frac{\pi}{3}}{1 - \tan\alpha\tan\frac{\pi}{3}} = \tan\big(\alpha + \tfrac{\pi}{3}\big).$$

取 $\alpha = \theta$，这就是 $n = 1$ 的情况。若 $f^k(x) = \tan(\theta + \frac{1}{3}k\pi)$，则以 $\alpha = \theta + \frac{1}{3}k\pi$ 代入上述公式，得

$$f^{k+1}(x) = f\big(\tan(\theta + \tfrac{1}{3}k\pi)\big) = \tan\big(\theta + \tfrac{1}{3}(k+1)\pi\big),$$

所以根据归纳法（induction），对每个正整数 n 都有 $f^n(x) = \tan(\theta + \frac{1}{3}n\pi)$。

整个过程不会出问题：公式只会在某个角 $\theta + \frac{1}{3}k\pi$ 与 $\frac{\pi}{2}$ 或 $\frac{\pi}{6}$ 模（modulo）π 同余时失效，而这会迫使 $\theta = \pm\frac{\pi}{6}$ 或 $\pm\frac{\pi}{2}$，即 $x = \pm 1/\sqrt{3}$ 或 x 为无穷大，这些情况都已被排除。这也解释了第 (i) 部分：加三次 $\frac{\pi}{3}$ 就是加上 π，即 tan 的周期。

第 (iii) 部分

由于 $|t| \leq 1$，我们可以写 $t = \sin\varphi$，其中 $\varphi = \arcsin t \in [-\frac{\pi}{2}, \frac{\pi}{2}]$。

想一想 $\sqrt{1 - \sin^2\alpha}$ 是否总是等于 $\cos\alpha$?

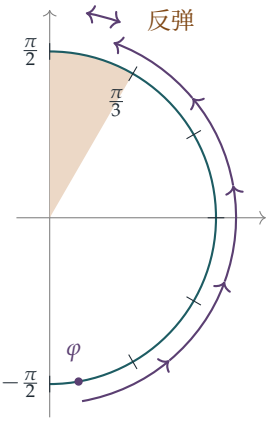
只有当 $\cos\alpha \geq 0$ 时才是。对于角 $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$，这没有问题，所以

$$g(\sin\alpha) = \cos\tfrac{\pi}{6}\sin\alpha + \sin\tfrac{\pi}{6}\cos\alpha = \sin\big(\alpha + \tfrac{\pi}{6}\big). \qquad (*)$$

要再次应用 $(*)$，我们必须把结果改写成 $[-\frac{\pi}{2}, \frac{\pi}{2}]$ 内某个角的 sin。这里分两种情况。

- 若 $\alpha \leq \frac{\pi}{3}$，则 $\alpha + \frac{\pi}{6} \leq \frac{\pi}{2}$ 已经在范围之内。
- 若 $\frac{\pi}{3} \leq \alpha \leq \frac{\pi}{2}$，则 $\alpha + \frac{\pi}{6}$ 可能会超过 $\frac{\pi}{2}$，但 $\sin(\alpha + \frac{\pi}{6}) = \sin\big(\pi - \alpha - \frac{\pi}{6}\big) = \sin\big(\frac{5\pi}{6} - \alpha\big)$，而且 $\frac{5\pi}{6} - \alpha \in [\frac{\pi}{3}, \frac{\pi}{2}]$。

（当 $\alpha = \frac{\pi}{3}$ 时，两者都给出 $\frac{\pi}{2}$。）所以这个范围内的角每次上升 $\frac{\pi}{6}$，直至进入 $[\frac{\pi}{3}, \frac{\pi}{2}]$；此后它便按 $\alpha \mapsto \frac{5\pi}{6} - \alpha$ 被反射，而再反射一次就会还原：这个角在两个值之间来回反弹。



每应用一次 g，角就沿单位圆的右半边移动 $\frac{\pi}{6}$；一旦进入阴影区域 $[\frac{\pi}{3}, \frac{\pi}{2}]$，它便来回翻转。

现在设 m 为所需的上升次数，即满足 $\varphi + \frac{m\pi}{6} \geq \frac{\pi}{3}$ 的最小整数 $m \geq 0$。由于 $\varphi \geq -\frac{\pi}{2}$，故 $m \leq 5$；以 t 表示则为：

t	$[\frac{\sqrt{3}}{2}, 1]$	$[\frac{1}{2}, \frac{\sqrt{3}}{2})$	$[0, \frac{1}{2})$	$[-\frac{1}{2}, 0)$	$[-\frac{\sqrt{3}}{2}, -\frac{1}{2})$	$[-1, -\frac{\sqrt{3}}{2})$
m	0	1	2	3	4	5

写 $g^n(t) = \sin\varphi_n$，其中 $\varphi_n \in [-\frac{\pi}{2}, \frac{\pi}{2}]$，且 $\varphi_0 = \varphi$。当 $n \leq m$ 时，之前的每个角都小于 $\frac{\pi}{3}$，所以 $\varphi_n = \varphi + \frac{n\pi}{6}$。于是 $\varphi_m \in [\frac{\pi}{3}, \frac{\pi}{2}]$，所以 $\varphi_{m+1} = \frac{5\pi}{6} - \varphi_m$，$\varphi_{m+2} = \varphi_m$，如此类推，交替出现。由于 $\sin\varphi_{m+1} = \sin(\varphi_m + \frac{\pi}{6}) = \sin(\varphi + \frac{(m+1)\pi}{6})$，我们得到（其中 $\varphi = \arcsin t$，m 取自上表）

$$g^n(t) = \begin{cases} \sin(\varphi + \frac{n\pi}{6}), & n \leq m+1, \\ \sin(\varphi + \frac{m\pi}{6}), & n > m+1,\ n-m \text{ 为偶数}, \\ \sin(\varphi + \frac{(m+1)\pi}{6}), & n > m+1,\ n-m \text{ 为奇数}. \end{cases}$$

特别地，对每个 t 和每个 $n \geq 5$，都有 $g^{n+2}(t) = g^n(t)$。

例如，$t = 0$ 时 $\varphi = 0$，$m = 2$，其轨道（orbit）为 $0 \mapsto \frac{1}{2} \mapsto \frac{\sqrt{3}}{2} \mapsto 1 \mapsto \frac{\sqrt{3}}{2} \mapsto 1 \mapsto \cdots$。

验算

- 取 $x = 1$：$f(1) = \frac{1+\sqrt{3}}{1-\sqrt{3}} = -2-\sqrt{3}$，$f^2(1) = \frac{1-\sqrt{3}}{1+\sqrt{3}} = \sqrt{3}-2$，且 $f(\sqrt{3}-2) = 1$，与 $\tan\frac{\pi}{4}$、$\tan\frac{7\pi}{12}$、$\tan\frac{11\pi}{12}$ 相符。
- 直接计算，$g(\frac{\sqrt{3}}{2}) = \frac{3}{4} + \frac{1}{4} = 1$ 及 $g(1) = \frac{\sqrt{3}}{2}$：反弹确实存在。想当然地猜 $g^5(0)$ 等于 $\sin\frac{5\pi}{6} = \frac{1}{2}$ 是错的；逐次迭代得到的是 1。

- 数值上，从 5000 个随机的 $t \in [-1,1]$ 出发，把 g 迭代十九次，结果与方框内的公式相差不超过 2×10^{-12}。

常见错误

- 在 (iii) 中对所有 n 都答 $\sin(\varphi + \frac{n\pi}{6})$。这忽略了 $\sqrt{1-t^2} \geq 0$，所以角永远不能越过 $\frac{\pi}{2}$。
- 计算 f^2 时的运算失误：应先把分子和分母同乘以 $1 - \sqrt{3}\,x$，然后才化简。

STEP 之外

形如 $x \mapsto \frac{ax+b}{cx+d}$ 的映射称为莫比乌斯变换（Möbius transformation），而把两个这样的变换复合，就相当于把它们的矩阵 $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ 相乘。这里 f 的矩阵是 $2\left(\begin{smallmatrix} \cos\frac{\pi}{3} & \sin\frac{\pi}{3} \\ -\sin\frac{\pi}{3} & \cos\frac{\pi}{3} \end{smallmatrix}\right)$，即一个经缩放、转角为 $\frac{\pi}{3}$ 的旋转，它的三次方是单位矩阵的倍数，这就是 $f^3(x) = x$ 的原因。第 (iii) 部分让我们浅尝动力系统（dynamical systems）的滋味：那些点是最终周期（eventually periodic）的，而不是周期的。Needham 的 *Visual Complex Analysis*（以几何直观讲解复分析的著作）对莫比乌斯变换有十分优美的几何处理。

详细解答

STEP II 2007，第 6 题

纯数学 · 一阶微分方程

STEP II 2007，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/07-s2-q6

关键思路

第 (i) 部分给出两个函数，它们的导数（derivative）的分母都是 $\sqrt{3+x^2}$；把两者适当地组合起来，所得的导数正好是 $\sqrt{3+x^2}$ 本身的倍数，于是所求的积分（integral）便唾手可得。第 (ii) 部分的微分方程（differential equation）是 $\frac{\mathrm{d}y}{\mathrm{d}x}$ 的二次方程。解这个二次方程，便得到两条普通的一阶方程，而每一条所需的正是第 (i) 部分的积分。

解答

第 (i) 部分

设 $s=\sqrt{3+x^2}$，则对任何实数 x 都有 $s>0$，并且 $\dfrac{\mathrm{d}s}{\mathrm{d}x}=\dfrac{x}{s}$。另外，由于 $s>|x|$，所以 $x+s>0$，因此该对数对所有 x 都有定义。

根据链式法则（chain rule），

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln(x+s)=\frac{1+\dfrac{x}{s}}{x+s}=\frac{\dfrac{s+x}{s}}{x+s}=\frac{1}{\sqrt{3+x^2}}.$$

根据积法则（product rule），

$$\frac{\mathrm{d}}{\mathrm{d}x}(xs)=s+x\cdot\frac{x}{s}=\frac{s^2+x^2}{s}=\frac{3+2x^2}{\sqrt{3+x^2}}.$$

? **想一想** 这两个导数怎样给出 $\int\sqrt{3+x^2}\,\mathrm{d}x$?

我们希望分母为 s 而分子为 $3+x^2$，因为 $\dfrac{3+x^2}{s}=s$。第二个导数的分子是 $3+2x^2$；加

上第一个导数的 3 倍，便得

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big(xs+3\ln(x+s)\Big)=\frac{3+2x^2+3}{s}=\frac{2(3+x^2)}{s}=2\sqrt{3+x^2}.$$

因此

$$\int\sqrt{3+x^2}\,\mathrm{d}x=\tfrac{1}{2}x\sqrt{3+x^2}+\tfrac{3}{2}\ln\big(x+\sqrt{3+x^2}\,\big)+C.$$

第 (ii) 部分

设 $p=\dfrac{\mathrm{d}y}{\mathrm{d}x}$。原方程就是 p 的二次方程 $3p^2+2xp-1=0$，所以

$$p=\frac{-2x\pm\sqrt{4x^2+12}}{6}=\frac{-x\pm\sqrt{3+x^2}}{3}.$$

判别式（discriminant）$4x^2+12$ 恒为正数，所以在每一个 x 处都有两个相异实根。两根之差 $\frac{2}{3}\sqrt{3+x^2}$ 永不为零，所以一个解不可能中途由一个根转到另一个根。（导数具有介值性质（intermediate value property），所以它不可能在两个始终相差最少 $\frac{2}{3}\sqrt{3}$ 的值之间跳跃。）因此，每一个解都对所有 x 满足

$$\frac{\mathrm{d}y}{\mathrm{d}x}=-\frac{x}{3}+\frac{1}{3}\sqrt{3+x^2}\qquad\text{或}\qquad\frac{\mathrm{d}y}{\mathrm{d}x}=-\frac{x}{3}-\frac{1}{3}\sqrt{3+x^2}$$

其中之一。利用第 (i) 部分的结果积分，

$$y=-\frac{x^2}{6}\pm\Big(\tfrac{1}{6}x\sqrt{3+x^2}+\tfrac{1}{2}\ln\big(x+\sqrt{3+x^2}\,\big)\Big)+C.$$

当 $x=1$ 时，$\sqrt{3+x^2}=2$，而 $\ln(x+\sqrt{3+x^2})=\ln 3$。

取 $+$ 号：由 $y(1)=0$ 得 $0=-\frac{1}{6}+\frac{1}{3}+\frac{1}{2}\ln 3+C$，所以 $C=-\frac{1}{6}-\frac{1}{2}\ln 3$。把对数项合并，

$$y=\frac{x\sqrt{3+x^2}-x^2-1}{6}+\frac{1}{2}\ln\left(\frac{x+\sqrt{3+x^2}}{3}\right).$$

取 $-$ 号：由 $y(1)=0$ 得 $0=-\frac{1}{6}-\frac{1}{3}-\frac{1}{2}\ln 3+C$，所以 $C=\frac{1}{2}+\frac{1}{2}\ln 3$，于是

$$y=\frac{3-x^2-x\sqrt{3+x^2}}{6}-\frac{1}{2}\ln\left(\frac{x+\sqrt{3+x^2}}{3}\right).$$

这就是通过 $(1,0)$ 的两个解，两者对所有实数 x 都成立。

验算

- 把 $x=1$ 代入每个答案：分式部分是 $\frac{2-1-1}{6}=0$（或 $\frac{3-1-2}{6}=0$），而 $\ln\frac{1+2}{3}=0$。所以两个答案都有 $y(1)=0$。

- p 的两个根之和是 $-\frac{2x}{3}$，所以两个解之和应为 $\frac{1-x^2}{3}$。平方根项和对数项确实互相抵消，剩下的正好是 $\frac{2-2x^2}{6} = \frac{1-x^2}{3}$。
- 数值上，用辛普森法则（Simpson's rule）可得 $\int_{-1}^{2} \sqrt{3+x^2}\,\mathrm{d}x \approx 5.949681$，与第 (i) 部分的公式吻合。用龙格－库塔法（Runge–Kutta）把方程的每一个分支由 $(1,0)$ 逐步计算至 $x = 4$，得 $y \approx 0.584957$ 和 $y \approx -5.584957$，与两条公式吻合。（两者之和是 $-5 = \frac{1-16}{3}$。）

常见错误

- 把 $3p^2 + 2xp = 1$ 当作线性方程处理，或尝试分离变量（separate variables），而不是把它当作 p 的二次方程来解。
- 两个分支共用同一个常数 C。每个符号都需要各自的常数，并由 $y(1) = 0$ 确定。
- 漏掉 $\sqrt{3+x^2}$ 前面的因子 $\frac{1}{3}$，以致使用第 (i) 部分的积分时倍数出错。

STEP 之外

形如 $F\left(x, y, \frac{\mathrm{d}y}{\mathrm{d}x}\right) = 0$ 而对导数并非线性的方程，称为隐式（implicit）方程。通过同一点可以有多个解，方程中 p 的每一个根各对应一个。在两个根重合之处，可能出现奇解（singular solutions），即一族曲线的包络线（envelope），克莱罗方程（Clairaut's equation）便是一例。本题的两个根永不重合，所以恰好有两个解。第 (i) 部分的 $\ln\left(x + \sqrt{3+x^2}\right)$ 等于 $\operatorname{arsinh}(x/\sqrt{3})$ 加上一个常数，而该积分其实就是标准的代换（substitution）$x = \sqrt{3}\sinh u$，只是换了个面貌。Boyce 与 DiPrima 合著的 *Elementary Differential Equations*（初等微分方程教科书）涵盖一阶方程，以及其解的存在性与唯一性（existence and uniqueness）。

详细解答

STEP II 2007，第 7 题

纯数学 · 琴生不等式

STEP II 2007，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/07-s2-q7

本题在 *Beyond STEP* 第 3.3 节凸性的威力中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch03/convexity.html

关键思路

延森不等式（Jensen's inequality）指出：对于凹函数（concave function），函数值的平均不超过平均值处的函数值。每一部分的关键，都在于看出该用哪一个凹函数，以及该对哪些数取平均。就三角形而言，各角之和固定，所以它们的平均值是已知的。至于算术—几何平均不等式（AM–GM inequality），对数（logarithm）能把积化为和。(a) 和 (b) 两部分则需要选对代入 AM–GM 的一组数，当中包括一些常数。

解答

凹性

若 $f(x) = \sin x$，则 $f''(x) = -\sin x$。当 $0 < x < \pi$ 时，$\sin x > 0$，所以 $f''(x) < 0$，故 $\sin x$ 在该区间上是凹的。

若 $f(x) = \ln x$，则 $f'(x) = 1/x$ 及 $f''(x) = -1/x^2$。对所有 $x > 0$，这都是负数，所以 $\ln x$ 在 $x > 0$ 时是凹的。

第 (i) 部分

三角形的各角满足 $A, B, C > 0$ 及 $A + B + C = \pi$，所以每个角都在 $(0, \pi)$ 之内，而 sin 在此区间上是凹的。取 $n = 3$，由延森不等式得

$$\frac{\sin A + \sin B + \sin C}{3} \leq \sin\left(\frac{A + B + C}{3}\right) = \sin\frac{\pi}{3} = \frac{\sqrt{3}}{2}.$$

两边乘以 3，

$$\sin A + \sin B + \sin C \leq \frac{3\sqrt{3}}{2},$$

正是所求。等号成立当且仅当 $A = B = C = \pi/3$，即等边三角形。

第 (ii) 部分

取 $f(x) = \ln x$，它在 $x > 0$ 上是凹的，并设 $x_k = t_k > 0$。由延森不等式得

$$\frac{1}{n}\sum_{k=1}^{n} \ln t_k \leq \ln\left(\frac{t_1 + t_2 + \cdots + t_n}{n}\right).$$

左边等于 $\frac{1}{n}\ln(t_1 t_2 \cdots t_n) = \ln \sqrt[n]{t_1 t_2 \cdots t_n}$。由于 e^x 是递增函数，我们可以对两边取指数而不改变不等号的方向：

$$\sqrt[n]{t_1 t_2 \cdots t_n} \leq \frac{t_1 + t_2 + \cdots + t_n}{n},$$

等号成立当且仅当 $t_1 = t_2 = \cdots = t_n$。（当 $n = 1$ 时，这显然是等式，与上述结论一致。）

(a)

想一想 我们应该对哪四个数取平均?

左边有四项，所以取 $n = 4$。我们需要几何平均（geometric mean）得出 xyz，而 x^4, y^4, z^4 给出 $\sqrt[4]{x^4 y^4 z^4} = xyz$。常数 16 就是第四项，而 $\sqrt[4]{16} = 2$。因此取 $t_1 = x^4$，$t_2 = y^4$，$t_3 = z^4$，$t_4 = 16$，全部为正数：

$$\frac{x^4 + y^4 + z^4 + 16}{4} \geq \sqrt[4]{16\, x^4 y^4 z^4} = 2xyz,$$

其中取四次方根时用到 $x, y, z > 0$。两边乘以 4 得

$$x^4 + y^4 + z^4 + 16 \geq 8xyz,$$

正是所求。等号成立需要 $x^4 = y^4 = z^4 = 16$，即 $x = y = z = 2$。

(b) 现在出现的是五次方，所以我们取 $n = 5$，并让 x^5, y^5, z^5 成为其中三个数。要令五次方根恰好得出 xyz，其余两个数的积必须是 1。最自然的选择是 $t_4 = t_5 = 1$：

$$\frac{x^5 + y^5 + z^5 + 1 + 1}{5} \geq \sqrt[5]{x^5 y^5 z^5 \cdot 1 \cdot 1} = xyz.$$

移项整理后，

$$x^5 + y^5 + z^5 - 5xyz \geq -2$$

对所有正数 x, y, z 成立。

想一想 这是否已证明最小值是 -2?

单凭这一点并不足够：下界（lower bound）必须能够取到，才是最小值。AM–GM 的等号在五个数全部相等时成立，即 $x^5 = y^5 = z^5 = 1$，所以 $x = y = z = 1$。此时该式的值为 $1 + 1 + 1 - 5 = -2$。因此最小值是

$$\boxed{-2}, \quad \text{取得于} x = y = z = 1.$$

验算

- 在 (i) 中，等腰直角三角形给出 $1 + 2 \cdot \frac{\sqrt{2}}{2} \approx 2.414$，小于 $3\sqrt{3}/2 \approx 2.598$。随机抽取 200 000 个三角形，结果从未超过 2.598。
- 在 (a) 中，当 $x = y = z = 2$ 时，两边都等于 64。
- 在 (b) 中，从随机起点出发进行数值最小化，结果总是收敛到 $(1, 1, 1)$，其值为 -2。此外，$x = y = z = 1.1$ 给出 $3(1.61051) - 5(1.331) \approx -1.82 > -2$。

常见错误

- 忘记了取指数之所以保持不等号方向，全因 e^x 是递增函数；这一点值得写上一句说明。
- 在 (b) 中，写到“≥ -2”便停笔，没有举出令该式达到 -2 的 x、y、z 的值。
- 对 x^5, y^5, z^5 使用 $n = 3$ 的 AM–GM，这样只会得到 $x^5 + y^5 + z^5 \geq 3(xyz)^{5/3}$，无法与 $5xyz$ 直接比较。

STEP 之外

延森不等式是关于凸（convex）函数或凹函数的命题，它真正的归宿是分析学和概率论：用期望值（expectation）的语言来说，它就是对凹函数 f 有 $\mathrm{E}[f(X)] \leq f(\mathrm{E}[X])$。(a) 和 (b) 中的技巧，即补上常数令几何平均恰好符合所需，正是加权 AM–GM 不等式和赫尔德（Hölder）不等式理论的起点。想进一步了解不等式和凸性，可参阅 Bartle 与 Sherbert 的 *Introduction to Real Analysis*（实分析入门）；至于概率版本，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）。

详细解答

STEP II 2007，第 8 题

纯数学 · 以向量证明塞瓦定理

STEP II 2007，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/07-s2-q8

关键思路

平面上每一点都可以写成 $\beta\mathbf{b}+\gamma\mathbf{c}$，而第 (i) 部分正好告诉我们这样的点何时位于 BC 上：两个系数之和为 1。因此，要找出一条直线与 BC（或 CA、AB）的交点，我们只需把该直线上的一般点用 $\mathbf{b}$ 和 $\mathbf{c}$ 表示，再对系数加上适当的条件。这样每个比都可以用 β 和 γ 表示，而三者的乘积会全部约简。

解答

在整个解答中，像 $BD:DC$ 这样的比是指有向比（signed ratio）：若 $\overrightarrow{BD}=k\overrightarrow{DC}$，则 $BD/DC=k$；当 D 位于线段 BC 之外时，这个比是负数。由于 A、B、C 不共线（collinear），$\mathbf{b}$ 和 $\mathbf{c}$ 互不平行且均非零，所以我们可以比较系数：由 $p\mathbf{b}+q\mathbf{c}=p'\mathbf{b}+q'\mathbf{c}$ 必可得出 $p=p'$ 及 $q=q'$。

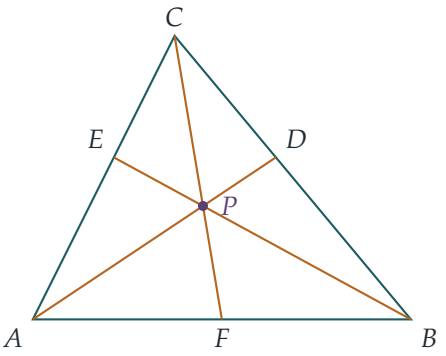
第 (i) 部分

若 $s+t=1$，则

$$s\mathbf{b}+t\mathbf{c}=(1-t)\mathbf{b}+t\mathbf{c}=\mathbf{b}+t(\mathbf{c}-\mathbf{b}),$$

这是通过 B 且方向为 $\overrightarrow{BC}$ 的直线的向量方程（vector equation），其中 t 为参数。当 t 取遍所有实数时，X 走遍整条直线。因此 X 的轨迹（locus）是通过 B 和 C 的（无限延伸的）直线；线段 BC 则是其中 $0\le t\le 1$ 的部分。

反过来，比较系数可知，点 $\beta\mathbf{b}+\gamma\mathbf{c}$ 位于直线 BC 上当且仅当 $\beta+\gamma=1$。这就是我们从现在起使用的判别方法。特别地，$\beta+\gamma\neq 1$ 表示 P 不在 BC 上，而 $\beta,\gamma\neq 0$ 表示 P 不在 CA 或 AB 上。



第 (ii) 部分

直线 AP 上的一般点是 $\lambda(\beta\mathbf{b}+\gamma\mathbf{c})=\lambda\beta\,\mathbf{b}+\lambda\gamma\,\mathbf{c}$。根据第 (i) 部分，它位于 BC 上正好是当 $\lambda\beta+\lambda\gamma=1$，即

$$\lambda=\frac{1}{\beta+\gamma}.$$

（这需要 $\beta+\gamma\neq 0$。若 $\beta+\gamma=0$，则 $\overrightarrow{AP}=\beta(\mathbf{b}-\mathbf{c})$ 与 BC 平行，D 并不存在，所以题目其实隐含了 $\beta+\gamma\neq 0$ 的假设。）因此

$$\mathbf{d}=\frac{\beta\mathbf{b}+\gamma\mathbf{c}}{\beta+\gamma}.$$

现在

$$\overrightarrow{BD}=\mathbf{d}-\mathbf{b}=\frac{\beta\mathbf{b}+\gamma\mathbf{c}-(\beta+\gamma)\mathbf{b}}{\beta+\gamma}=\frac{\gamma}{\beta+\gamma}(\mathbf{c}-\mathbf{b}),$$
$$\overrightarrow{DC}=\mathbf{c}-\mathbf{d}=\frac{(\beta+\gamma)\mathbf{c}-\beta\mathbf{b}-\gamma\mathbf{c}}{\beta+\gamma}=\frac{\beta}{\beta+\gamma}(\mathbf{c}-\mathbf{b}).$$

两者都是同一个向量 $\mathbf{c}-\mathbf{b}$ 的倍数，所以 $\beta\overrightarrow{BD}=\gamma\overrightarrow{DC}$，又因为 $\beta\neq 0$，

$$BD:DC=\gamma:\beta,$$

这正是所需的结果。

第 (iii) 部分

我们用同样的方法求 E 和 F，不过这次的目标直线是 CA（$\mathbf{c}$ 的倍数）和 AB（$\mathbf{b}$ 的倍数）。

> ? **想一想** 怎样判断一点 $p\mathbf{b}+q\mathbf{c}$ 是否在直线 CA 上?

直线 CA 通过原点 A，方向为 $\mathbf{c}$，所以线上的点正好就是 $\mathbf{c}$ 的所有倍数：$\mathbf{b}$ 的系数必须为零。同样地，一点在 AB 上，正好是当它的 $\mathbf{c}$ 系数为零。

求 E。直线 BP 上的一般点是

$$\mathbf{b}+\mu(\beta\mathbf{b}+\gamma\mathbf{c}-\mathbf{b})=\big(1-\mu(1-\beta)\big)\mathbf{b}+\mu\gamma\,\mathbf{c}.$$

令 $\mathbf{b}$ 的系数为零，得 $\mu=1/(1-\beta)$（我们需要 $\beta\neq1$；若 $\beta=1$，则 $\overrightarrow{BP}=\gamma\mathbf{c}$ 与 CA 平行，E 并不存在）。所以

$$\mathbf{e}=e\,\mathbf{c},\qquad e=\frac{\gamma}{1-\beta}.$$

于是 $\overrightarrow{CE}=(e-1)\mathbf{c}$ 及 $\overrightarrow{EA}=-e\,\mathbf{c}$，而且因为 $\gamma\neq0$，所以 $e\neq0$，因此

$$\frac{CE}{EA}=\frac{e-1}{-e}=\frac{1-e}{e}=\frac{(1-\beta-\gamma)/(1-\beta)}{\gamma/(1-\beta)}=\frac{1-\beta-\gamma}{\gamma}.$$

求 F。把 B 和 C 的角色互换（即 $\mathbf{b}\leftrightarrow\mathbf{c}$ 及 $\beta\leftrightarrow\gamma$），在假设 $\gamma\neq1$ 之下，直线 CP 与 AB 相交于 $\mathbf{f}=f\,\mathbf{b}$，其中 $f=\beta/(1-\gamma)$。于是 $\overrightarrow{AF}=f\,\mathbf{b}$ 及 $\overrightarrow{FB}=(1-f)\mathbf{b}$，其中 $1-f=(1-\beta-\gamma)/(1-\gamma)\neq0$，因为 $\beta+\gamma\neq1$。所以

$$\frac{AF}{FB}=\frac{f}{1-f}=\frac{\beta}{1-\beta-\gamma}.$$

乘积。中间的因子用第 (ii) 部分的结果，得

$$\frac{AF}{FB}\times\frac{BD}{DC}\times\frac{CE}{EA}=\frac{\beta}{1-\beta-\gamma}\times\frac{\gamma}{\beta}\times\frac{1-\beta-\gamma}{\gamma}=1,$$

其中每个被约去的因子都不是零，因为 $\beta\neq0$、$\gamma\neq0$ 及 $\beta+\gamma\neq1$。这就是所需的结果。（它对有向比成立，因此对长度本身也成立。）

验算

- 形心（centroid）：取 $\beta=\gamma=\frac{1}{3}$，P 便是形心。这时 $BD:DC=1:1$、$CE/EA=\frac{1/3}{1/3}=1$ 及 $AF/FB=1$，所以三条塞瓦线（cevian）都是中线（median），正如所料。
- 取 $A=(0,0)$、$B=(4,0)$、$C=(1.5,3)$、$\beta=0.3$、$\gamma=0.4$（即图中的情况），我们得到 $D=(2.571,1.714)$、$E=\frac{4}{7}\mathbf{c}$ 及 $F=(2,0)$，从而 $\frac{AF}{FB}=1$、$\frac{BD}{DC}=\frac{4}{3}$、$\frac{CE}{EA}=\frac{3}{4}$，乘积为 1。
- 我们以超过 1700 个随机三角形和随机的 β,γ（包括 P 在三角形之外的情况），用分数精确地核对了 D、E、F 的公式、全部三个有向比以及乘积。

常见错误

- 把 (i) 中的轨迹说成线段 BC：s 和 t 并没有限定为正数，所以轨迹是整条直线。

- 把比倒转了：$\overrightarrow{BD}$ 与 $\mathbf{c}$ 的系数 γ 成正比，所以 γ 大的时候，D 会被拉向 C。
- 忘记了沿 BP 和 CP 的参数是从 B 和 C 起计，而不是从原点起计。

STEP 之外

第 (iii) 部分就是塞瓦定理（Ceva's theorem），而 $(1-\beta-\gamma, \beta, \gamma)$ 是 P 相对于三角形 ABC 的重心坐标（barycentric coordinates）：它们与三角形 PBC、PCA、PAB 的面积成正比，而那三个比不过是这些面积之间的商。重心坐标是仿射几何（affine geometry）最自然的框架：在仿射几何中，直线以及沿直线的比都有意义，但长度和角度则没有；重心坐标亦会在电脑图像（computer graphics）和有限元法（finite-element method）中再次出现。Strang 的 *Introduction to Linear Algebra*（线性代数入门）有系统地处理相对于一组基（basis）的坐标，是很好的入门读物。

详细解答

STEP II 2007，第 9 题

力学 · 斜碰撞与恢复系数

> **STEP II 2007，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/07-s2-q9

关键思路

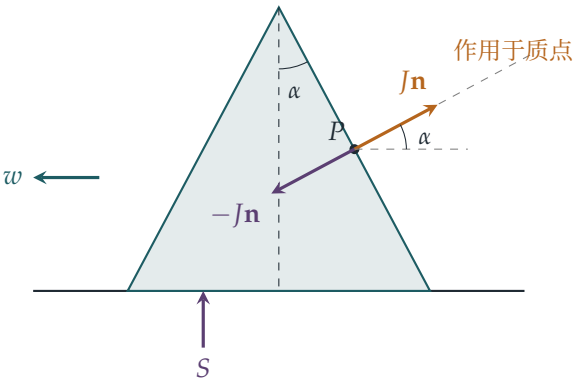
所有接触面都是光滑的（smooth），所以质点与圆锥之间唯一的冲量（impulse）沿碰撞点的公法线（common normal）方向作用，而桌面只能沿铅垂方向施力。由此我们同时得知两件事：质点沿锥面方向的速度不变，而质点与圆锥合起来的水平动量（momentum）守恒。把这些与沿法线方向的牛顿恢复定律（Newton's law of restitution）结合起来，便可求出圆锥的速率。第 (ii) 部分其实就是把第 (i) 部分应用于下落质点速度的法向分量。

解答

把质点视为一点，把圆锥视为刚体（rigid body）；按题目所述，圆锥既不转动，亦不作铅垂方向的运动，所以刚碰撞后它以某速率 w 沿水平方向滑动。我们在通过圆锥的轴和碰撞点 P 的铅垂平面内考虑。由于通过 P 的母线（generator）与铅垂的轴成 α 角，在 P 的向外法线与水平方向成 α 角。取单位向量

$$\mathbf{n} = (\cos\alpha,\ \sin\alpha)\ \text{（向外法线）}, \qquad \mathbf{t} = (\sin\alpha,\ -\cos\alpha)\ \text{（沿斜面向下）},$$

其中第一个坐标为水平方向（远离轴），第二个坐标为铅垂向上。



设作用于质点的冲量为 $J\mathbf{n}$（$J > 0$），则圆锥所受的冲量为 $-J\mathbf{n}$，其水平部分的大小为 $J\cos\alpha$，方向远离 P。桌面提供向上的铅垂冲量 $S = J\sin\alpha$，刚好抵消 $-J\mathbf{n}$ 的向下部分；这就是圆锥没有铅垂速度的原因。所以

$$Mw = J\cos\alpha, \tag{1}$$

而圆锥沿水平方向远离 P 运动。

第 (i) 部分

质点到达时的速度为 $-u\mathbf{n}$，它没有沿 $\mathbf{t}$ 方向的分量。由于接触面光滑，作用于质点的冲量完全沿 $\mathbf{n}$ 方向，不能改变质点沿 $\mathbf{t}$ 方向的分量。因此碰撞后该分量仍然是零，质点的速度是 $\mathbf{n}$ 的倍数：它沿 P 处的法线运动（向外运动；又或者，若圆锥退得够快，则仍然向内运动，只是较慢）。

碰撞后质点的速度为 $(-u + J/m)\,\mathbf{n}$。圆锥以速率 w 沿水平方向远离 P 运动（即图中向左），所以它的速度是 $(-w, 0)$，其沿 $\mathbf{n}$ 方向的分量为 $-w\cos\alpha$。

> **想一想** 牛顿恢复定律比较的是沿哪个方向的相对速度？圆锥在当中有甚么贡献？

恢复定律沿公法线 $\mathbf{n}$ 方向起作用。碰撞前，沿 $\mathbf{n}$ 方向的接近速率为 u（圆锥静止）。碰撞后的分离速率，等于质点的 $\mathbf{n}$ 分量减去圆锥的相应分量：

$$\left(-u + \frac{J}{m}\right) - (-w\cos\alpha) = eu \quad\Longrightarrow\quad J = m\big((1+e)u - w\cos\alpha\big).$$

代入 (1)，得

$$Mw = m\cos\alpha\,\big((1+e)u - w\cos\alpha\big) \quad\Longrightarrow\quad \boxed{w = \frac{m(1+e)u\cos\alpha}{M + m\cos^2\alpha}}.$$

换一个说法，水平动量守恒（桌面的冲量是铅垂的）：$m(-u\cos\alpha) = m\big(-u + \frac{J}{m}\big)\cos\alpha - Mw$，这又是 (1)。

第 (ii) 部分

现在质点到达时的速度为 $(0,-u)$。把它分解：

$$(0,-u)\cdot\mathbf{n} = -u\sin\alpha, \qquad (0,-u)\cdot\mathbf{t} = u\cos\alpha.$$

切向部分 $u\cos\alpha$ 不受碰撞影响（同样因为接触面光滑），之后亦再无作用：冲量 $J\mathbf{n}$、方程 (1) 和恢复定律的方程都只涉及法向分量。所以第 (i) 部分的推导完全适用，只需把法向接近速率 u 换成 $u\sin\alpha$：

$$w = \frac{m(1+e)(u\sin\alpha)\cos\alpha}{M+m\cos^2\alpha} = \frac{mu(1+e)\sin\alpha\cos\alpha}{M+m\cos^2\alpha},$$

正是所求。

对 α 求最大值。固定 M, m, e, u，我们需要在 $0<\alpha<\frac{\pi}{2}$ 的范围内求 $f(\alpha)=\dfrac{\sin\alpha\cos\alpha}{M+m\cos^2\alpha}$ 的最大值。

> ? **想一想** 直接对这个商求导相当繁复。有没有一个代换可以令它变得简洁？

把分子和分母同除以 $\cos^2\alpha$，并设 $T=\tan\alpha>0$，利用 $\sec^2\alpha = 1+T^2$：

$$f = \frac{\tan\alpha}{M\sec^2\alpha+m} = \frac{T}{MT^2+(M+m)} = \frac{1}{MT+\dfrac{M+m}{T}}.$$

根据算术—几何平均不等式（AM–GM inequality）（两项皆为正数），

$$MT+\frac{M+m}{T} \geq 2\sqrt{M(M+m)},$$

等号成立当且仅当 $MT=(M+m)/T$，即 $T^2=(M+m)/M$。所以 f 在该处取得最大值，w 亦然。由于 $T\mapsto\alpha$ 是由 $(0,\infty)$ 到 $(0,\frac{\pi}{2})$ 的双射（bijection），这是唯一的最大值，而且

$$\cos^2\alpha = \frac{1}{1+\tan^2\alpha} = \frac{1}{1+\frac{M+m}{M}} = \frac{M}{2M+m}, \qquad \text{所以} \qquad \cos\alpha = \sqrt{\frac{M}{2M+m}}$$

（因为 α 是锐角，故取正根）。最大速率为 $w_{\max} = \dfrac{mu(1+e)}{2\sqrt{M(M+m)}}$。

验算

- 若 $\alpha\to\frac{\pi}{2}$，圆锥变成一块平的圆盘：铅垂下落不会产生水平推力，故 $w\to 0$。在第 (i) 部分中，当 $\alpha\to\frac{\pi}{2}$ 时法线是铅垂的，同样有 $w\to 0$。吻合。
- 在第 (i) 部分中，当 $\alpha\to 0$（一面铅垂的墙）时，我们得到 $w=\dfrac{m(1+e)u}{M+m}$，这正是与一个质量为 M 的自由方块正碰时大家熟悉的答案。

- 若 $m \ll M$，最佳角度满足 $\cos^2\alpha \to \frac{1}{2}$，即 $\alpha = 45°$，这正是 $\sin\alpha\cos\alpha$ 取最大值之处。
- 数值上，取 $M = 3, m = 1, e = \frac{1}{2}, u = 4, \alpha = \frac{\pi}{6}$，第 (ii) 部分给出 $w = \frac{2\sqrt{3}}{5} \approx 0.693$。以随机选取的参数直接计算冲量（由恢复定律解出 J），结果与两条公式都一致；对 α 作网格搜索，亦确认了取得最大值的角度。

常见错误

- 忘记圆锥会移动，写恢复定律的方程时，把分离速率当成只是质点的速率 eu。圆锥的速度沿法线方向有分量 $w\cos\alpha$。
- 沿法线方向应用动量守恒。桌面的冲量沿 $\mathbf{n}$ 方向有分量，所以只有水平动量守恒。
- 以为法线与铅垂方向成 α 角。与铅垂方向成 α 角的是锥面，所以法线是与水平方向成 α 角。

STEP 之外

这里的桌面起着约束（constraint）的作用：它令圆锥失去一个自由度（degree of freedom），而它的冲量则是为了维持这个约束所需的任何值。到了大学，这会发展成冲击约束（impulsive constraints）的理论，在拉格朗日力学（Lagrangian mechanics）中以广义坐标（generalised coordinates）处理十分简洁；而向法线方向投影的做法，与受约束物体的有效质量（effective mass）背后是同一个想法。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对二维的碰撞和冲量有仔细的论述。

详细解答

STEP II 2007，第 10 题

力学 · 涉及摩擦力的静力学

STEP II 2007，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/07-s2-q10

关键思路

这是一道静力学（statics）问题，唯一棘手的地方是质心（centre of mass）的位置。一旦知道 G 在轴上的位置，对接触点取力矩（taking moments）便可同时消去法向反作用力（normal reaction）和摩擦力（friction），直接求出 P。然后沿水平方向分解力，可知摩擦力的大小等于 P，余下的交给摩擦定律 $|F| \leq \mu R$ 便可。式中出现绝对值（modulus），是因为视乎 α 的大小，P 可能要推，也可能要拉。

解答

模型

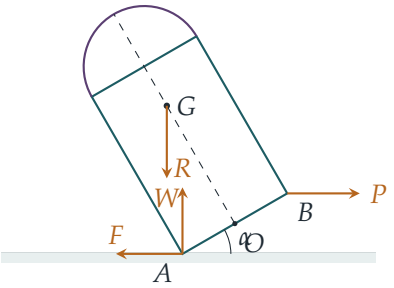
我们把这个立体视为刚体（rigid body），它与平面只在一点 A 接触，这一点是平底面（即圆柱体不与半球相连的一端）边缘上的最低点。设 B 为该底面的最高点，则 AB 是一条长度为 $2r$ 的直径，与水平方向成 α 角。作用力有：作用于质心 G 的重量 W；作用于 A 的法向反作用力 R 和摩擦力 F；以及作用于 B 的水平力 P。

质心

沿轴由底面的圆心 O 起量度距离。圆柱体的体积为 $3\pi r^3$，故质量为 $3\pi r^3\rho$，其质心位于 $\frac{3}{2}r$ 处。半球的体积为 $\frac{2}{3}\pi r^3$，密度为 3ρ，故质量为 $2\pi r^3\rho$；它的质心在其平面之外 $\frac{3}{8}r$ 处，而该平面位于 $3r$ 处，所以质心位于 $\frac{27}{8}r$ 处。总质量为 $5\pi r^3\rho$，而

$$OG = \frac{3\cdot\frac{3}{2}r + 2\cdot\frac{27}{8}r}{5} = \frac{\frac{9}{2}r + \frac{27}{4}r}{5} = \frac{9}{4}r.$$

因此 G 在轴上，距离底面 $\frac{9}{4}r$，并且 $W = 5\pi r^3\rho g$。



平衡

取 A 为原点，x 轴沿水平方向（由 A 指向 B 的垂足），y 轴铅垂向上。设 P 在 x 方向的水平分量为 X，故 X 可正可负；同样地，设摩擦力在 x 方向为 F。

底面由 A 出发，沿方向 $(\cos\alpha, \sin\alpha)$ 伸展，所以 $B = (2r\cos\alpha, 2r\sin\alpha)$ 而 $O = (r\cos\alpha, r\sin\alpha)$。轴垂直于底面，指向立体内部，方向为 $(-\sin\alpha, \cos\alpha)$。因此 G 的水平坐标为

$$x_G = r\cos\alpha - \tfrac{9}{4}r\sin\alpha.$$

沿铅垂方向分解，得 $R = W$。沿水平方向分解，得 $F + X = 0$，故 $|F| = |X|$。

> **想一想** 我们应该对哪一点取力矩?

对 A 取力矩：这样 R 和 F 都没有力矩，只剩下 W 和 P。重量沿铅垂方向作用并通过 G，所以它对 A 的力矩为 Wx_G（当 $x_G > 0$ 时为顺时针方向）。作用于 B 的水平力位于高度 $2r\sin\alpha$ 处，所以它的力矩为 $X \cdot 2r\sin\alpha$（当 $X < 0$ 时为逆时针方向，当 $X > 0$ 时为顺时针方向）。取顺时针方向为正，

$$Wx_G + 2r\sin\alpha\, X = 0 \quad\Longrightarrow\quad X = -\frac{W\left(r\cos\alpha - \frac{9}{4}r\sin\alpha\right)}{2r\sin\alpha} = W\left(\tfrac{9}{8} - \tfrac{1}{2}\cot\alpha\right),$$

其中我们除以了 $\sin\alpha$，由于 $0 < \alpha < \frac{1}{2}\pi$，它是正数。

摩擦条件

由于平面是粗糙的，而立体在 A 处不滑动，我们需要 $|F| \le \mu R$。利用 $|F| = |X|$ 和 $R = W > 0$，

$$W\left|\tfrac{9}{8} - \tfrac{1}{2}\cot\alpha\right| \le \mu W \quad\Longrightarrow\quad \mu \ge \left|\tfrac{9}{8} - \tfrac{1}{2}\cot\alpha\right|,$$

这正是所要求的结果。

X 的正负号告诉我们 P 的作用方向。若 $\tan\alpha > \frac{4}{9}$，则 G 越过了 A，位于远离 B 的一侧，立体本身会向后翻倒，因此 P 必须把 B 向外拉离立体；若 $\tan\alpha < \frac{4}{9}$，则 P 必须向内推。无论哪种情况，摩擦力都与 P 方向相反，而绝对值正好涵盖这两种情况。

验算

- 若 $\tan\alpha = \frac{4}{9}$，则 $x_G = 0$：G 在 A 的正上方，立体自行平衡，$P = 0$，无需摩擦力。所得的界给出 $\mu \geq |\frac{9}{8} - \frac{9}{8}| = 0$，与预期相符。
- 当 $\alpha = 45°$ 时，所得的界为 $\mu \geq \frac{5}{8}$。
- 以数值方法验证：把立体切成薄圆片，得总质量为 $5\pi\rho r^3$ 及 $OG = 2.2500r$；再对 10 000 个随机的 α 值直接解三条平衡方程，每次所得的 $|F|/R$ 都等于 $|\frac{9}{8} - \frac{1}{2}\cot\alpha|$。

常见错误

- 忘记半球的密度是三倍，因而把它的质量当作 $\frac{2}{3}\pi r^3\rho$；这会令 G 的位置改变，常数 $\frac{9}{8}$ 也随之改变。
- 把半球质心的 $\frac{3}{8}r$ 由圆柱体的底面量起，而不是由半球本身位于高度 $3r$ 的平面量起。
- 预先假定 P（或摩擦力）的方向，因而漏掉绝对值：无论 P 向哪一方作用，结果都必须成立。

STEP 之外

求组合体的质心，其实是体积分（volume integral）的离散版本：$\bar{\mathbf{r}} = \int \rho\,\mathbf{r}\,\mathrm{d}V \big/ \int \rho\,\mathrm{d}V$，而半球的 $\frac{3}{8}r$ 本来就是这样推导出来的。在大学一年级的力学课程中，这会发展成刚体理论，其中把同样的积分里的 $\mathbf{r}$ 换成 r^2，便得到转动惯量（moments of inertia）。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学导论）对这两个概念都有细致的阐述。

详细解答

STEP II 2007，第 11 题

力学 • 三维空间中的抛体

STEP II 2007，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/07-s2-q11

关键思路

把整个运动写成一个位置向量（position vector）：水平速度保持不变，只有垂直分量受重力影响。一旦把 $\mathbf{r}(t)$ 写在纸上，与 O 的距离就是 $|\mathbf{r}|$，而最短距离只需求一个二次式的最小值便可得到。最后一部分之所以行得通，是因为子弹极快，在它飞行期间质点几乎没有移动。

解答

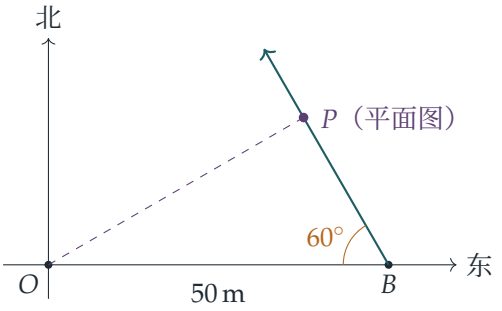
我们把质点（particle）视为只在重力作用下运动的一点（$g = 10\,\mathrm{m\,s^{-2}}$，不计空气阻力），并取 O 为原点，于是 B 的位置向量为 $50\mathbf{i}$。

第 (i) 部分

初速度的分量。若 $\tan\alpha = \frac{1}{2}$，则 $\cos\alpha = \frac{2}{\sqrt{5}}$ 及 $\sin\alpha = \frac{1}{\sqrt{5}}$（边长为 1、2、$\sqrt{5}$ 的三角形）。所以初速度的

$$\text{水平速率为} 25\cdot\frac{2}{\sqrt{5}} = 10\sqrt{5}, \qquad \text{垂直速率为} 25\cdot\frac{1}{\sqrt{5}} = 5\sqrt{5}.$$

水平方向是哪一边？该方向与直线 OB 成 60° 角。假如质点大致朝东、远离 O 而去，它根本不可能经过 O 附近；要从 O 的北面经过，它必须折返向西并偏向北方。所以水平方向与 $\overrightarrow{BO}$（向西）成 60° 角，并在其北侧，由此得单位向量（unit vector）$-\frac{1}{2}\mathbf{i} + \frac{\sqrt{3}}{2}\mathbf{j}$。



因此水平速度为

$$10\sqrt{5}\left(-\tfrac{1}{2}\mathbf{i}+\tfrac{\sqrt{3}}{2}\mathbf{j}\right)=-5\sqrt{5}\,\mathbf{i}+5\sqrt{15}\,\mathbf{j},$$

而且它保持不变，因为重力只沿垂直方向作用。垂直方向上，$z=5\sqrt{5}\,t-\frac{1}{2}(10)t^2$。加上起点 $50\mathbf{i}$，得

$$\mathbf{r}(t)=\left(50-5\sqrt{5}\,t\right)\mathbf{i}+5\sqrt{15}\,t\,\mathbf{j}+\left(5\sqrt{5}\,t-5t^2\right)\mathbf{k}.$$

与 O 的距离。把各分量平方后相加，

$$\begin{aligned}|\mathbf{r}|^2&=\left(2500-500\sqrt{5}\,t+125t^2\right)+375t^2+\left(125t^2-50\sqrt{5}\,t^3+25t^4\right)\\&=25t^4-50\sqrt{5}\,t^3+625t^2-500\sqrt{5}\,t+2500.\end{aligned}$$

把它与下式比较：

$$\begin{aligned}25\left(t^2-\sqrt{5}\,t+10\right)^2&=25\left(t^4+5t^2+100-2\sqrt{5}\,t^3+20t^2-20\sqrt{5}\,t\right)\\&=25t^4-50\sqrt{5}\,t^3+625t^2-500\sqrt{5}\,t+2500,\end{aligned}$$

两者相同。所以 $|\mathbf{r}|=5\left|t^2-\sqrt{5}\,t+10\right|$。二次式 $t^2-\sqrt{5}\,t+10$ 的判别式（discriminant）为 $5-40<0$，故它恒为正数，绝对值符号可以除去：

$$|\mathbf{r}|=5\left(t^2-\sqrt{5}\,t+10\right)\text{ m},$$

正是所求。

点 P。配方（completing the square）得

$$t^2-\sqrt{5}\,t+10=\left(t-\tfrac{\sqrt{5}}{2}\right)^2+\tfrac{35}{4},$$

所以当 $t=\frac{\sqrt{5}}{2}$ 时距离最小，而最短距离为 $5\cdot\frac{35}{4}=\frac{175}{4}=43.75\,\text{m}$。代入 $t=\frac{\sqrt{5}}{2}$，此时 $\sqrt{5}\,t=\frac{5}{2}$ 及 $t^2=\frac{5}{4}$：

$$\overrightarrow{OP}=\tfrac{75}{2}\,\mathbf{i}+\tfrac{25\sqrt{3}}{2}\,\mathbf{j}+\tfrac{25}{4}\,\mathbf{k}\quad\text{（单位为米）}.$$

$\overrightarrow{OP}$ 的水平部分，向东分量为 $\frac{75}{2}$，向北分量为 $\frac{25\sqrt{3}}{2}$。方位角（bearing）θ 是由正北起顺时针量度的，所以

$$\tan\theta=\frac{\text{东}}{\text{北}}=\frac{75/2}{25\sqrt{3}/2}=\sqrt{3},$$

又因为两个分量均为正数，故 $\theta=60^\circ$：由 O 测得 P 的方位角为 060°。

第 (ii) 部分

速度的垂直分量为

$$\dot{z} = 5\sqrt{5} - 10t,$$

它恰好在 $t = \frac{\sqrt{5}}{2}$ 时为零。这正是质点位于 P 的时刻，所以质点在 P 处达到最高点（高度为 $\frac{25}{4}$ m）。

第 (iii) 部分

子弹以 $350\,\mathrm{m\,s^{-1}}$ 沿直线由 O 飞往 P（我们忽略重力对它的影响），所走的距离为 $OP = \frac{175}{4}$ m。所需时间为

$$T = \frac{175/4}{350} = \frac{1}{8}\,\mathrm{s}.$$

? **想一想** 多过了这 $\frac{1}{8}$ 秒之后，质点在哪里?

由第 (ii) 部分，质点在 P 处的垂直速度为零，所以它在该处的速度纯粹是水平的，大小为 $10\sqrt{5}\,\mathrm{m\,s^{-1}}$。在其后的 T 秒内，它水平移动了 $10\sqrt{5}\,T$，并下坠了 $\frac{1}{2}gT^2$。取 $T = \frac{1}{8}$，

$$\text{水平位移} = \frac{10\sqrt{5}}{8} \approx 2.80\,\mathrm{m}, \qquad \text{下坠距离} = 5 \cdot \frac{1}{64} \approx 0.08\,\mathrm{m}.$$

这两段位移互相垂直，所以质点与 P 的距离为

$$\sqrt{\frac{500}{64} + \frac{25}{4096}} = \frac{5\sqrt{1281}}{64} \approx 2.80\,\mathrm{m},$$

即约为 3 m，正是所求。（下坠的影响微乎其微：答案基本上就是 $10\sqrt{5}/8$。）神枪手射失了。

验算

- 当 $t = 0$ 时，公式给出 $5 \cdot 10 = 50\,\mathrm{m}$，即距离 OB。吻合。
- $|\overrightarrow{OP}|^2 = 37.5^2 + 21.65^2 + 6.25^2 = 1406.25 + 468.75 + 39.06 = 1914.06 = 43.75^2$，所以 P 的位置向量与最短距离一致。
- 以数值方法扫描 $|\mathbf{r}(t)|$，得最小值 43.75 m，出现于 $t \approx 1.118 = \frac{\sqrt{5}}{2}$；再让质点由 P 继续运动 0.125 s，得相距 2.796 m。

常见错误

- 把水平方向取为 $\frac{1}{2}\mathbf{i} + \frac{\sqrt{3}}{2}\mathbf{j}$（远离 O）。这样质点永远不会从 O 的北面经过，距离也算不出题目所给的式子。

► 把方位角写成 30°：那是由正东量起的角，但方位角是由正北起顺时针量度的。

STEP 之外

第 (iii) 部分是拦截问题（interception problem）的入门版本：要击中移动的目标，必须瞄准它将会到达的位置，这就要解 $|\mathbf{r}_{\text{目标}}(t) - \mathbf{r}_{\text{子弹}}(t)| = 0$，同时求出发射方向和时间。求 $|\mathbf{r}(t)|$ 的最小值，也正是轨道问题和碰撞问题中“最接近距离”（distance of closest approach）背后的想法；在这类问题中，把距离平方以得到多项式，是标准的第一步。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）对这类向量运动学（vector kinematics）有严谨而完整的阐述。

详细解答

STEP II 2007，第 12 题

概率与统计 · 几何分布的最大值

STEP II 2007，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/07-s2-q12

关键思路

先撇开题目对程序那段颇为繁琐的描述，逐颗骰子单独来看。每颗骰子都一直掷到首次出现六点为止，而整个实验在两颗骰子中较慢的那一颗掷出六点时结束。因此，掷骰次数就是两个独立（independent）几何随机变量（geometric random variable）的最大值，而处理最大值最方便的途径，是考虑累积概率（cumulative probability）$\mathrm{P}(\max \le r)$。

解答

第 (i) 部分

设 X 为第一颗骰子首次出现六点时的掷骰次序，并以同样方式为第二颗骰子定义 Y。骰子的每一次投掷都与其他各次投掷独立，所以 X 与 Y 独立，而且两者都服从几何分布：

$$\mathrm{P}(X = k) = q^{k-1}p, \qquad \mathrm{P}(X \le r) = 1 - \mathrm{P}(\text{首 } r \text{ 次投掷全非六点}) = 1 - q^r.$$

题目的程序怎样与此吻合？当两颗骰子都未出现过六点时，我们两颗一起掷；一旦其中一颗出现了六点，便只继续掷另一颗。无论哪种情况，每颗骰子都恰好掷到它首次出现六点为止（之后不再掷），而当两颗都出现过六点，我们便立即停止。如果两颗骰子在同一次投掷中首次出现六点，那就是两颗同时掷出六点的情况。因此掷骰次数为

$$N = \max(X, Y).$$

? **想一想** 为甚么要处理 $\mathrm{P}(N \le r)$，而不直接处理 $\mathrm{P}(N = r)$？

事件 $N \le r$ 就是“$X \le r$ 而且 $Y \le r$”，而独立性可把它化为乘积。事件 $N = r$ 则要分成几种麻烦的情况（一颗骰子在第 r 次完成而另一颗较早完成，或两颗都在第 r 次完成）。

所以

$$\mathrm{P}(N \le r) = \mathrm{P}(X \le r)\,\mathrm{P}(Y \le r) = (1-q^r)^2,$$

而对于 $r \ge 1$，

$$\begin{aligned}
\mathrm{P}(N = r) &= \mathrm{P}(N \le r) - \mathrm{P}(N \le r-1) \\
&= (1-q^r)^2 - (1-q^{r-1})^2 \\
&= \left[(1-q^r) - (1-q^{r-1})\right]\left[(1-q^r) + (1-q^{r-1})\right] \\
&= (q^{r-1} - q^r)(2 - q^{r-1} - q^r) \\
&= pq^{r-1}\,(2 - q^{r-1} - q^r),
\end{aligned}$$

当中先用了平方差（difference of two squares），再用了 $q^{r-1} - q^r = q^{r-1}(1-q) = pq^{r-1}$。这就是所要求的结果。

至于期望值（expectation），把这个概率拆成两部分：

$$\mathrm{P}(N = r) = 2pq^{r-1} - pq^{2r-2}(1+q).$$

把题目给出的结果求导（或把它除以 x），可得对于 $|x| < 1$ 有 $\sum_{r\ge1} rx^{r-1} = (1-x)^{-2}$。分别取 $x = q$ 和 $x = q^2$ 代入（当 $0 < p \le 1$ 时两者都小于 1）：

$$\begin{aligned}
\mathrm{E}(N) &= 2p\sum_{r=1}^{\infty} rq^{r-1} - p(1+q)\sum_{r=1}^{\infty} r(q^2)^{r-1} \\
&= \frac{2p}{(1-q)^2} - \frac{p(1+q)}{(1-q^2)^2} \\
&= \frac{2}{p} - \frac{p(1+q)}{p^2(1+q)^2} \\
&= \frac{2}{p} - \frac{1}{p(1+q)} \\
&= \frac{1+2q}{p(1+q)},
\end{aligned}$$

其中用了 $1-q = p$ 和 $1-q^2 = (1-q)(1+q) = p(1+q)$。若只以 p 表示（代入 $q = 1-p$），

$$\mathrm{E}(N) = \frac{3-2p}{p(2-p)}.$$

第 (ii) 部分

进行大量实验后，掷骰次数的平均值应接近 $\mathrm{E}(N)$，所以我们解 $\mathrm{E}(N) = m$ 来估计 p：

$$m\,p(2-p) = 3-2p \quad\Longleftrightarrow\quad mp^2 - 2(m+1)p + 3 = 0,$$

所以

$$p = \frac{(m+1) \pm \sqrt{(m+1)^2 - 3m}}{m} = \frac{m+1 \pm \sqrt{m^2 - m + 1}}{m}.$$

想一想 我们要取哪一个根?

每次实验最少要掷一次,所以 $m \geq 1$。于是 $\sqrt{m^2-m+1} \geq 1$(因为 $m^2-m=m(m-1) \geq 0$),所以取正号的根至少是 $(m+2)/m > 1$,这对概率来说是不可能的。根据同一个不等式,取负号的根至多是 $m/m = 1$,而且它是正数,因为 $(m+1)^2 > m^2-m+1$。所以估计值为

$$\hat{p} = \frac{m+1-\sqrt{m^2-m+1}}{m}.$$

验算

- 若 $p=1$,每次实验都只需掷一次:$\mathrm{E}(N) = \frac{1}{1\cdot 1} = 1$,而 $m=1$ 给出 $\hat{p} = (2-1)/1 = 1$。
- 各概率之和为 1:$\sum_r \mathrm{P}(N=r)$ 逐项相消(telescope)后得 $\lim_{r\to\infty}(1-q^r)^2 = 1$。
- 对于公平的骰子,$p=\frac{1}{6}$ 给出 $\mathrm{E}(N) = \frac{8/3}{11/36} = \frac{96}{11} \approx 8.73$,比一颗骰子平均所需的 6 次稍多,这正合预期。一次包含 200,000 次实验的模拟得出平均值 8.71,而把 $m = 96/11$ 代入 $\hat{p}$ 的公式,则得回 $\frac{1}{6}$。

常见错误

- 把每颗骰子的投掷分开计算:两颗骰子同时掷一次只算作一次投掷,这正是 $N = \max(X,Y)$ 的原因。
- 忘记 $pq^{2r-2}(1+q)$ 这一项所需的求和是取 $x=q^2$,而不是 $x=q$。
- 在第 (ii) 部分保留两个根,而没有检查哪一个才可以是概率。

STEP 之外

$\mathrm{P}(\max \leq r) = \prod \mathrm{P}(X_i \leq r)$ 这个技巧,是处理独立变量的最大值(把符号换成 $\geq$ 则可处理最小值)的标准方法,也是*顺序统计量*(order statistics)的起点。第 (ii) 部分用的是*矩估计法*(method of moments):令样本平均值等于理论平均值,再解出参数。大学一年级的统计学会把它与最大似然法(maximum likelihood)比较,并探讨这类估计量(estimator)有多准确。至于概率方面,Grimmett 与 Welsh 的 *Probability: An Introduction*(概率论入门教材)是很好的下一步。

详细解答

STEP II 2007，第 13 题

概率与统计 · 生日问题

> **STEP II 2007，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/07-s2-q13

本题在 *Beyond STEP* 第 9.9 节生日问题中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/birthday.html

关键思路

没有任何两位宾客生日相同的概率（probability），是若干个形如 $\frac{365-r}{365}$ 的因子的乘积，而乘积并不好处理。第一部分把每个因子化成指数函数（exponential），而指数函数的乘积等于以和为指数的指数函数，这个和可以准确地求出。至于主人的问题，各位宾客的事件互相独立（independent），所以互补事件（complement）的概率只是单独一个幂，同一个近似方法照样适用。

解答

近似式

对任意实数 x，

$$\mathrm{e}^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots .$$

设 $x = r/n$。由于 r 远小于 n，x 很小，x^2 及其后各项与 x 相比可以忽略不计。所以

$$\mathrm{e}^{-r/n} \approx 1 - \frac{r}{n} = \frac{n-r}{n},$$

这正是所求。

最少两位宾客生日相同

设 A 为 k 位宾客中最少有两位生日相同的事件。求 $\mathrm{P}(A')$，即 k 个生日全部不同的概率，会比较容易。逐一考虑各位宾客。第一位的生日可以是任何一天。已知首 r 位宾客有 r 个不同的生日，第 $(r+1)$ 位宾客必须避开这 r 天，其概率为 $\frac{365-r}{365}$，因为各人的生日互相独立，并且在 365 天上均匀分布（uniform）。把这些概率相乘，

$$\mathrm{P}(A') = \prod_{r=0}^{k-1} \frac{365-r}{365}.$$

若派对的人数不多，这里每个 r 都远小于 365，所以由第一部分，取 $n = 365$，

$$\begin{aligned}\mathrm{P}(A') &\approx \prod_{r=0}^{k-1} \mathrm{e}^{-r/365} = \exp\Big(-\frac{1}{365}\sum_{r=0}^{k-1} r\Big) \\ &= \exp\Big(-\frac{1}{365}\cdot\frac{k(k-1)}{2}\Big) = \mathrm{e}^{-k(k-1)/730}.\end{aligned}$$

（$r = 0$ 的因子恰好是 $1 = \mathrm{e}^0$，所以把它包括在内并无影响。）因此

$$\mathrm{P}(A) = 1 - \mathrm{P}(A') \approx 1 - \mathrm{e}^{-k(k-1)/730},$$

这就是所求的结果。

> ? **想一想** 一次过对这么多个因子取近似值，难道不令人担心吗？误差可能会累积起来。

每个因子的误差属 $(r/365)^2$ 的数量级，而指数本身则属 $k^2/730$ 的数量级。对于我们所需的 k 值（20 左右），最大的 $r/365$ 约为 0.06，指数的总相对误差只有百分之几。当 $k = 23$ 时，近似式给出 0.500，而真确值为 0.507。

机会过半所需的最少人数

我们需要

$$1 - \mathrm{e}^{-k(k-1)/730} \geq \tfrac{1}{2} \iff \mathrm{e}^{-k(k-1)/730} \leq \tfrac{1}{2} \iff \frac{k(k-1)}{730} \geq \ln 2,$$

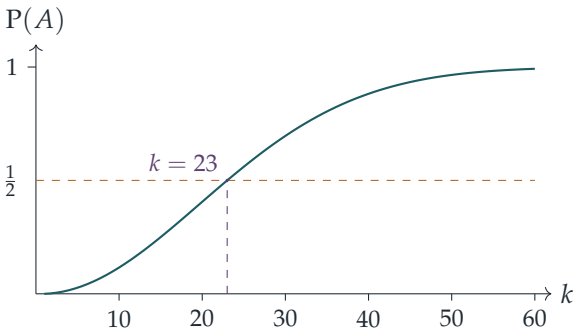
其中先取对数（对数函数是递增的，所以不等号方向不变），再乘以 -1（不等号方向因而倒转）。利用 $\ln 2 \approx \frac{253}{365}$，条件变成

$$k(k-1) \geq 730 \times \frac{253}{365} = 2 \times 253 = 506.$$

由于当 $k \geq 1$ 时 $k(k-1)$ 随 k 递增，而且

$$22 \times 21 = 462 < 506, \qquad 23 \times 22 = 506,$$

所以最小的值是 $k = 23$。



与主人生日相同

现在设 m 为宾客人数（主人不计算在内）。设 B 为最少有一位宾客与主人生日相同的事件。无论主人的生日是哪一天，每位宾客的生日不是这一天的概率都是 $\frac{364}{365}$，而且各人互相独立，所以

$$\mathrm{P}(B') = \left(\frac{364}{365}\right)^m \approx \left(\mathrm{e}^{-1/365}\right)^m = \mathrm{e}^{-m/365},$$

这里用了第一部分的结果，取 $n = 365$ 和 $r = 1$。这一次，无论 m 有多大，$r/n = 1/365$ 都确实很小。

我们需要 $\mathrm{P}(B) \geq \frac{1}{2}$，即 $\mathrm{e}^{-m/365} \leq \frac{1}{2}$。与之前完全一样，这等价于

$$\frac{m}{365} \geq \ln 2 \approx \frac{253}{365} \iff m \geq 253.$$

所以最少要有 253 位宾客。

> **想一想** 为甚么答案比 23 大这么多?

在第一个问题中，任何一对宾客都可以生日相同，而 23 位宾客可组成 $\binom{23}{2} = 253$ 对。在第二个问题中，只有 m 对“宾客与主人”的配对才算数。要有同样多的配对机会，便需要 $m = 253$，这正是 253 在两个答案中都出现的原因。

验算

- 准确计算乘积，可以确认两个答案。有人生日相同的真确概率，22 位宾客时为 0.4757，23 位时为 0.5073。$1 - (364/365)^m$ 的真确值在 $m = 252$ 时小于 $\frac{1}{2}$，在 $m = 253$ 时为 0.5005。
- $\frac{253}{365} = 0.69315\ldots$，而 $\ln 2 = 0.69315\ldots$，两者准确至小数点后五位都相同，所以题目给出的近似值不影响任何一个答案。
- 模拟 200 000 场各有 23 人的派对，结果有 50.7% 的派对出现生日相同的情况。

常见错误

- 计算了两位指定宾客生日相同的概率，或者为每一对宾客各加一次 $\frac{1}{365}$。互相重叠的事件，其概率不能直接相加；改为考虑互补事件便可避开这个问题。
- 在第一个乘积中每个因子都用 $\frac{364}{365}$。第 $(r+1)$ 位宾客必须避开 r 天，而不只是一天。
- 在最后一部分把主人也算进宾客人数内。主人是另外一个人，所以需要 253 位宾客。

STEP 之外

这里的经验法则（heuristic）是：把 k 件物件放进 N 个盒子，一旦 k 达到大约 $\sqrt{2N\ln 2}\approx 1.18\sqrt{N}$，便很可能出现重复。密码学（cryptography）中的“生日攻击”（birthday attack）正是以此为基础：要在一个 N 位元的杂凑（hash）中找出碰撞（collision），只需大约 $2^{N/2}$ 次尝试，而不是 2^N 次。这个近似式的背后是泊松极限（Poisson limit）：生日相同的配对数目近似服从平均值为 $\binom{k}{2}/365$ 的泊松分布，而 $\mathrm{P}(\text{没有配对相同})\approx \mathrm{e}^{-k(k-1)/730}$。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门教材）详细讨论了生日问题及其泊松近似。

详细解答

STEP II 2007，第 14 题

概率与统计 · 分段概率密度函数

STEP II 2007，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/07-s2-q14

关键思路

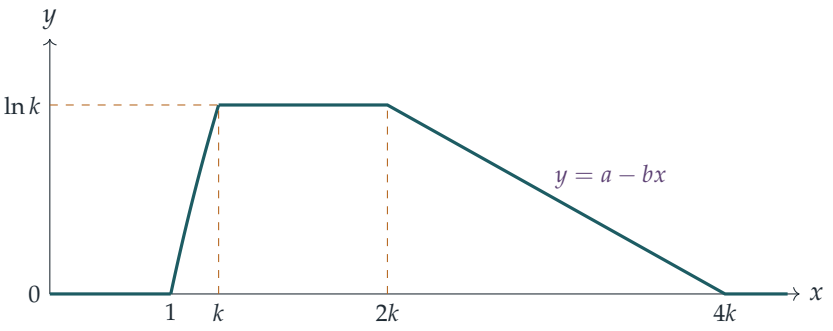
“连续”这两个字举足轻重。分段定义的密度函数（density function）在每个分界点都必须衔接起来，由此我们马上得到关于 a 和 b 的两条方程。余下的一个条件，即图像下的总面积为 1，便可确定 k。至于中位数（median），我们逐段把面积累加，直至超过二分之一为止。

解答

全题之中 $k > 1$：否则区间 $1 \le x \le k$ 便是空的，而且无论如何我们都需要 $\ln k > 0$，密度函数才会是非负，并且在 $x = 1$ 之后不恒等于零。

第 (i) 部分

在 $[1, k]$ 上，图像是常见的对数曲线，在 $x = 1$ 时由 0 开始（因此与取值为零的一段连续地衔接），递增而且是凹的（concave）。它在 $x = k$ 时到达高度 $\ln k$，保持在 $\ln k$ 的水平直至 $x = 2k$，然后沿直线下降，在 $x = 4k$ 时降至 0（下文会证明连续性恰好迫使如此），其后一直为零。草图采用了我们在第 (ii) 部分求得的数值：$k = \mathrm{e}^{1/3} \approx 1.40$，因此平台的高度为 $\frac{1}{3}$。



第 (ii) 部分

连续性（*continuity*）。在 $x = k$ 处，两条公式都给出 $\ln k$，所以这里没有新的资料。在 $x = 2k$ 和 $x = 4k$ 处，连续性要求

$$a - 2bk = \ln k, \qquad a - 4bk = 0.$$

两式相减，得 $2bk = \ln k$，所以

$$b = \frac{\ln k}{2k}, \qquad a = 4bk = 2\ln k.$$

总概率。我们把面积分成三部分。用分部积分法（integration by parts）（取 $u = \ln x$, $\mathrm{d}v = \mathrm{d}x$），

$$\int_1^k \ln x \,\mathrm{d}x = \Big[x\ln x - x\Big]_1^k = k\ln k - k + 1.$$

平台是一个阔 k、高 $\ln k$ 的长方形，面积为 $k\ln k$。最后一部分是一个三角形，底为 $4k - 2k = 2k$，高为 $\ln k$，面积为 $\frac{1}{2}\cdot 2k\cdot \ln k = k\ln k$。所以

$$1 = (k\ln k - k + 1) + k\ln k + k\ln k = 3k\ln k - k + 1,$$

化简后得

$$k(3\ln k - 1) = 0.$$

由于 $k > 1$，不可能有 $k = 0$，所以 $\ln k = \frac{1}{3}$，而

$$k = \mathrm{e}^{1/3} \approx 1.396, \qquad a = 2\ln k = \tfrac{2}{3}, \qquad b = \frac{\ln k}{2k} = \frac{1}{6\,\mathrm{e}^{1/3}} \approx 0.119.$$

留意 $b > 0$，所以最后一段确实如图所示向下倾斜。

第 (iii) 部分

我们把概率逐段累加，找出中位数 m 落在哪一段。利用 $\ln k = \frac{1}{3}$，

$$\mathrm{P}(X \le k) = k\ln k - k + 1 = 1 - \tfrac{2k}{3} \approx 0.070,$$

而平台再加上 $k\ln k = \frac{k}{3}$，所以

$$\mathrm{P}(X \le 2k) = 1 - \tfrac{2k}{3} + \tfrac{k}{3} = 1 - \tfrac{k}{3} \approx 0.535.$$

想一想 中位数落在图像的哪一段上?

由于 $0.070 < \frac{1}{2} < 0.535$，中位数落在平台上，即 $k \le m \le 2k$。在这一段，密度函数恒等于 $\frac{1}{3}$，所以

$$\mathrm{P}(X \le m) = 1 - \tfrac{2k}{3} + \tfrac{1}{3}(m-k) = \tfrac{1}{2}.$$

乘以 3：$3 - 2k + m - k = \frac{3}{2}$，所以

$$m = 3k - \tfrac{3}{2} = 3\mathrm{e}^{1/3} - \tfrac{3}{2} \approx 2.69.$$

这个值确实介乎 $k \approx 1.40$ 与 $2k \approx 2.79$ 之间，正如所应有的那样。

验算

- 取 $k = \mathrm{e}^{1/3}$、$a = \frac{2}{3}$、$b = 1/(6\mathrm{e}^{1/3})$，对密度函数作数值积分（中点法则），得总面积为 1.0000000，而 $\int_1^{2.68684} f(x)\,\mathrm{d}x = 0.5000000$。
- 中位数的合理性检查：略多于一半的概率落在 $2k$ 的左边（约 0.535），所以中位数应该在 $2k \approx 2.79$ 稍左的位置；果然 $m \approx 2.69$。

常见错误

- 忘记了 $\int_1^k \ln x\,\mathrm{d}x$ 的下限：$-[x\ln x - x]$ 在 $x = 1$ 处给出的 $+1$，正是令关于 k 的方程可以如此俐落地因式分解的原因。
- 因为中位数在支集（support）$[1, 4k]$ 的“中间”，便假设它落在倾斜的一段上；务必先把面积累加。

STEP 之外

本题的密度函数由几段拼合而成，而各常数则由连续性和归一化（normalisation）确定。到了大学，你会遇到*累积分布函数*（cumulative distribution function）$F(x) = \mathrm{P}(X \le x)$，并以它作为最基本的对象：每一个随机变量（random variable），不论是离散的还是连续的，都有累积分布函数，而中位数这类量不过是 $F(m) = \frac{1}{2}$ 的解。Grimmett and Welsh, *Probability: An Introduction*（概率论入门教材）清楚地阐述了这个观点。

详细解答

STEP III 2007，第 1 题

纯数学 · 正切和与韦达定理

STEP III 2007，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/07-s3-q1

关键思路

多个角之和的正切，是各个角的正切的对称函数（symmetric function）之比；而多项式的根的对称函数，正正就是韦达定理（Vieta's formulae）以系数表示出来的东西。至于最后一部分，窍门是把那条三角方程化为 $t = \tan\theta$ 的四次方程（quartic）；其系数的对称性便会迫使角之和的正切等于零。

解答

四个角之和的正切

把和拆成 $A + B$，其中 $A = \theta_1 + \theta_2$，$B = \theta_3 + \theta_4$，于是

$$\tan A = \frac{t_1 + t_2}{1 - t_1 t_2}, \qquad \tan B = \frac{t_3 + t_4}{1 - t_3 t_4}.$$

然后，把 $\tan(A + B) = \dfrac{\tan A + \tan B}{1 - \tan A \tan B}$ 的分子和分母同乘以 $(1 - t_1 t_2)(1 - t_3 t_4)$，得

$$\tan(A + B) = \frac{(t_1 + t_2)(1 - t_3 t_4) + (t_3 + t_4)(1 - t_1 t_2)}{(1 - t_1 t_2)(1 - t_3 t_4) - (t_1 + t_2)(t_3 + t_4)}.$$

分子是 $t_1 + t_2 + t_3 + t_4 - (t_1 t_3 t_4 + t_2 t_3 t_4 + t_1 t_2 t_3 + t_1 t_2 t_4)$，而分母是 $1 - t_1 t_2 - t_3 t_4 + t_1 t_2 t_3 t_4 - (t_1 t_3 + t_1 t_4 + t_2 t_3 + t_2 t_4)$。以

$$\sigma_1 = \sum t_i, \quad \sigma_2 = \sum_{i<j} t_i t_j, \quad \sigma_3 = \sum_{i<j<k} t_i t_j t_k, \quad \sigma_4 = t_1 t_2 t_3 t_4$$

表示各个初等对称函数（elementary symmetric functions），上式便是

$$\boxed{\tan(\theta_1 + \theta_2 + \theta_3 + \theta_4) = \frac{\sigma_1 - \sigma_3}{1 - \sigma_2 + \sigma_4}.}$$

（规律与两个角或三个角的情况相同：奇数次的对称函数正负号交替地放在分子，偶数次的则放在分母。）

以系数表示

若 $t_1,\dots,t_4$ 是 $at^4+bt^3+ct^2+dt+e=0$ 的根，则 $at^4+bt^3+ct^2+dt+e=a(t-t_1)(t-t_2)(t-t_3)(t-t_4)$，比较系数（韦达定理）得

$$\sigma_1=-\frac{b}{a},\qquad \sigma_2=\frac{c}{a},\qquad \sigma_3=-\frac{d}{a},\qquad \sigma_4=\frac{e}{a}.$$

代入后，分子和分母同乘以 $a\neq 0$，得

$$\boxed{\tan(\theta_1+\theta_2+\theta_3+\theta_4)=\frac{d-b}{a-c+e}.}$$

最后一部分

由于 $1+\cos 2\theta=2\cos^2\theta$，方程即为

$$2p\cos^2\theta+\cos\theta\cos\alpha+\sin\theta\sin\alpha=0.$$

除以 $\cos^2\theta$（我们不理会 $\cos\theta=0$ 这个特殊情况），得

$$2p=-(\cos\alpha+t\sin\alpha)\sec\theta,\qquad t=\tan\theta.$$

$\sec\theta$ 并不是 t 的有理函数(rational function),所以我们把两边平方,并利用 $\sec^2\theta=1+t^2$:

$$4p^2=(\cos\alpha+t\sin\alpha)^2(1+t^2).$$

展开后，

$$\sin^2\alpha\,t^4+2\sin\alpha\cos\alpha\,t^3+t^2+2\sin\alpha\cos\alpha\,t+(\cos^2\alpha-4p^2)=0,\qquad (*)$$

其中 t^2 的系数用了 $\sin^2\alpha+\cos^2\alpha=1$。

? **想一想** 平方可能会引入额外的解。各个 $\tan\theta_i$ 真的就是 $(*)$ 的四个根吗?

平方之后，$p\cos 2\theta-\cos(\theta-\alpha)+p=0$ 的解也同样满足 $(*)$。不过，把 θ 换成 $\theta+\pi$，$t=\tan\theta$ 和 $\cos^2\theta$ 都保持不变，而 $\cos(\theta-\alpha)$ 则会变号。因此，对于 $(*)$ 的每一个实根 t，在 $[0,2\pi)$ 内正切等于 t 的两个角之中，恰好有一个满足原来的方程（不会两个都满足，因为那样便需要 $\cos(\theta-\alpha)=0$，从而 $p\cos^2\theta=0$，属于特殊情况）。所以，$[0,2\pi)$ 内的解与 $(*)$ 的实根一一对应（one-to-one correspondence），而题目所给的四个角的正切，正是 $(*)$ 的四个根。

现在应用前面的结果，取 $a=\sin^2\alpha$，$b=d=2\sin\alpha\cos\alpha$，$c=1$，$e=\cos^2\alpha-4p^2$（我们需要 $a\neq 0$，即 $\sin\alpha\neq 0$；否则 $(*)$ 便不是四次方程，这又是一个特殊情况）。分子是 $d-b=0$，而分母是

$$a-c+e=\sin^2\alpha-1+\cos^2\alpha-4p^2=-4p^2,$$

它不等于零（若 $p=0$，方程只是 $\cos(\theta-\alpha)=0$，只有两个解）。因此

$$\tan(\theta_1+\theta_2+\theta_3+\theta_4)=0,$$

所以对某个整数 n 有 $\theta_1+\theta_2+\theta_3+\theta_4=n\pi$，这正是所要证明的。

验算

- 在四个角的公式中取 $t_3=t_4=0$：公式便简化为 $(t_1+t_2)/(1-t_1t_2)$，即我们熟悉的两角公式。
- 取 $p=\frac{1}{2}$ 和 $\alpha=3$，在 $[0,2\pi)$ 上以数值方法解方程，得到四个根，它们的和除以 π 后是一个整数，误差在 10^{-9} 以内。（并非每一组 p 和 α 都给出四个解：$p=\frac{1}{2}$、$\alpha=1$ 便只有两个解。）我们随机试过的每一组 (p,α)，只要给出四个解，结果都一样；我们亦用数以千计的随机角和随机四次方程验证过两条公式。

常见错误

- 韦达定理中的正负号出错：相间的系数 b 和 d 是带负号的。
- 试图不经平方就直接以 $\tan\theta$ 表示 $\cos(\theta-\alpha)$，又或者平方之后没有检查是否混入了不想要的角。

STEP 之外

以初等对称函数表示角之和的 tan，是一条重要定理的其中一个例子：根的每一个对称多项式（symmetric polynomial），都是系数的多项式。这个结果也可以用复数（complex numbers）漂亮地解释，因为 $\prod(1+\mathrm{i}t_k)=\prod\sec\theta_k\,\mathrm{e}^{\mathrm{i}\sum\theta_k}$，而它的实部和虚部分别是 $1-\sigma_2+\sigma_4$ 和 $\sigma_1-\sigma_3$。Needham 的 *Visual Complex Analysis* 深入发展了这种几何的思考方式，而 Siklos 的 *Advanced Problems in Mathematics* 则收录了更多同类风格的 STEP 题目。

详细解答

STEP III 2007，第 2 题

纯数学 · 二项式幂级数

STEP III 2007，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/07-s3-q2

关键思路

一切都源自同一个二项级数 (binomial series)。只要知道 $(1-4x)^{-1/2}$ 的系数，对级数求导便会带出一个因子 n（把 $(n!)^2$ 变成 $n!\,(n-1)!$），而对它积分则会除以 $n+1$（把 $(n!)^2$ 变成 $n!\,(n+1)!$）。因此，后面每一个和都是第一个级数的导数或积分，在 $|x|<\frac{1}{4}$ 内一个精心挑选的点上取值。

解答

第 (i) 部分

把奇数的乘积同时乘以和除以所欠缺的偶数 $2\cdot4\cdot6\cdots(2n)$：

$$1\cdot3\cdot5\cdots(2n-1)=\frac{1\cdot2\cdot3\cdot4\cdots(2n-1)(2n)}{2\cdot4\cdots(2n)}=\frac{(2n)!}{2^n\,(1\cdot2\cdots n)}=\frac{(2n)!}{2^n\,n!},$$

因为 n 个偶数 $2k$ 各自提供一个因子 2 和一个因子 k。

当 $|4x|<1$，即 $|x|<\frac{1}{4}$ 时，由二项级数得

$$(1-4x)^{-1/2}=1+\sum_{n=1}^{\infty}\frac{\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(-\frac{2n-1}{2}\right)}{n!}(-4x)^n.$$

分子有 n 个因子，每个都是负数，所以它等于 $(-1)^n\,\dfrac{1\cdot3\cdots(2n-1)}{2^n}$。这个 $(-1)^n$ 与 $(-4)^n$ 中的符号互相抵消，而 $4^n/2^n=2^n$，所以 x^n 的系数是

$$\frac{2^n\cdot1\cdot3\cdots(2n-1)}{n!}=\frac{2^n}{n!}\cdot\frac{(2n)!}{2^n\,n!}=\frac{(2n)!}{(n!)^2},$$

这里用了第一个结果。因此，当 $|x| < \frac{1}{4}$ 时，

$$\frac{1}{\sqrt{1-4x}} = 1 + \sum_{n=1}^{\infty} \frac{(2n)!}{(n!)^2}\, x^n,$$

正是所求。

第 (ii) 部分

幂级数（power series）在其收敛区间（interval of convergence）之内可以逐项求导（differentiate term by term），这里的收敛区间是 $|x| < \frac{1}{4}$。对两边求导，

$$\frac{\mathrm{d}}{\mathrm{d}x}(1-4x)^{-1/2} = -\tfrac{1}{2}(1-4x)^{-3/2}\cdot(-4) = 2(1-4x)^{-3/2},$$

而右边的常数项消失，x^n 变成 nx^{n-1}。由于 $n/(n!)^2 = 1/\big(n!\,(n-1)!\big)$，

$$2(1-4x)^{-3/2} = \sum_{n=1}^{\infty} \frac{(2n)!}{n!\,(n-1)!}\, x^{n-1}.$$

乘以 x 便得到我们想要的幂 x^n：

$$\sum_{n=1}^{\infty} \frac{(2n)!}{n!\,(n-1)!}\, x^n = \frac{2x}{(1-4x)^{3/2}} \qquad (|x| < \tfrac{1}{4}).$$

现在代入 $x = \frac{6}{25}$,这是容许的,因为 $\frac{6}{25} < \frac{6}{24} = \frac{1}{4}$。这时 $1-4x = \frac{1}{25}$,所以 $(1-4x)^{3/2} = \frac{1}{125}$，于是

$$\sum_{n=1}^{\infty} \frac{(2n)!}{n!\,(n-1)!} \left(\frac{6}{25}\right)^n = 2\cdot\frac{6}{25}\cdot 125 = 60.$$

第 (iii) 部分

> **想一想** 现在阶乘是 $(n+1)!\,n!$ 而不是 $(n!)^2$。对级数做甚么运算才会得出这个结果?

我们需要把第 n 个系数除以 $n+1$，因为 $(n+1)! = (n+1)\,n!$。对 x^n 积分正好做到这一点：$\int_0^x t^n\,\mathrm{d}t = \dfrac{x^{n+1}}{n+1}$。幂级数在其收敛区间之内同样可以逐项积分（term-by-term integration），所以当 $|x| < \frac{1}{4}$ 时，我们把第 (i) 部分的级数由 0 积分至 x：

$$\int_0^x (1-4t)^{-1/2}\,\mathrm{d}t = \Big[-\tfrac{1}{2}(1-4t)^{1/2}\Big]_0^x = \frac{1-\sqrt{1-4x}}{2},$$

于是

$$\frac{1-\sqrt{1-4x}}{2} = x + \sum_{n=1}^{\infty} \frac{(2n)!}{n!\,(n+1)!}\, x^{n+1}.$$

接着我们要对上目标和式中的幂。由于 $3^{2n} = 9^n = 9^{n+1}/9$,

$$\frac{2^{n+1}}{3^{2n}} = 9\left(\frac{2}{9}\right)^{n+1},$$

所以目标和式等于我们的级数在 $x = \frac{2}{9}$ 处的值的 9 倍。这个值是合理的, 因为 $\frac{2}{9} < \frac{2}{8} = \frac{1}{4}$。在该处，$1 - 4x = \frac{1}{9}$ 而 $\sqrt{1-4x} = \frac{1}{3}$，由此得

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{2^{n+1}(2n)!}{3^{2n}(n+1)!\,n!} &= 9\sum_{n=1}^{\infty} \frac{(2n)!}{n!\,(n+1)!}\left(\frac{2}{9}\right)^{n+1} \\ &= 9\left(\frac{1-\frac{1}{3}}{2} - \frac{2}{9}\right) = 9\left(\frac{1}{3} - \frac{2}{9}\right) = 1,\end{aligned}$$

正是所求。留意首项 x（即常数 1 的积分）必须减去，因为和式由 $n = 1$ 开始。

验算

- 在 (i) 中，$n = 2$ 给出 $1 \cdot 3 = 3$ 和 $\frac{4!}{2^2 2!} = 3$；级数的开首是 $1 + 2x + 6x^2 + 20x^3 + \cdots$，与平方后的结果吻合：$(1 + 2x + 6x^2)^2(1 - 4x) = 1 + O(x^3)$。
- 以精确分数把 (ii) 的首 3000 项相加, 得 60.000…; 而 (iii) 的首 400 项之和为 1.000…，准确至机器精度。
- 在 (iii) 中，首项（$n = 1$）是 $\frac{4\cdot 2}{9\cdot 2} = \frac{4}{9}$，已占总和 1 的相当大部分；其后各项大致按 $(8/9)^n$ 的速度递减。

常见错误

- 弄错二项式系数（binomial coefficient）的正负号：分子有 n 个负因子，$(-4)^n$ 中也有 n 个，所以它们刚好互相抵消。
- 忘记检查 $\frac{6}{25}$ 和 $\frac{2}{9}$ 是否位于 $|x| < \frac{1}{4}$ 之内；在这个范围之外，级数会发散（diverge）。
- 在 (iii) 中，忘记了对常数 1 积分所得的 x 项，这样会得出答案 3 而不是 1。

STEP 之外

第 (iii) 部分出现的数 $\frac{(2n)!}{n!\,(n+1)!} = \frac{1}{n+1}\binom{2n}{n}$ 就是卡塔兰数(Catalan numbers)$1, 1, 2, 5, 14, \ldots$，它们可用来计数加括号的方式、二元树（binary trees）以及许多其他结构的数目，而 $\frac{1-\sqrt{1-4x}}{2x}$ 是它们的母函数（generating function）。我们未经证明便使用的事实，即幂级数在严格位于其收敛半径（radius of convergence）之内的范围可以逐项求导和逐项积

分，是大学一年级分析学（analysis）的标准定理；参见 Abbott 的 *Understanding Analysis*（分析学入门教材）。

详细解答

STEP III 2007，第 3 题

纯数学 · 斐波那契恒等式

STEP III 2007，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/07-s3-q3

关键思路

这里每一条恒等式的证法都一样：取下标最大的那一项，用递推关系（recurrence）$F_m = F_{m-1} + F_{m-2}$ 把它拆开。这样做一次，(ii) 中的式子就变成所有下标都减一的式子；做两次，便回到同一形状而下标全部减二，这正是归纳法（induction）所需要的。至于 (iv) 的整除性（divisibility），留意 $F_{2n-1} + F_{2n} = F_{2n+1}$，所以 F_{2n-1}^2 与 F_{2n}^2 除以 F_{2n+1} 时余数相同。

解答

第 (i) 部分

由 $F_1 = F_2 = 1$ 开始，把相邻两项相加：

$$F_3 = 2, \quad F_4 = 3, \quad F_5 = 5, \quad F_6 = 8, \quad F_7 = 13, \quad F_8 = 21.$$

第 (ii) 部分

设 $k \geq 1$，使每一项都有定义。把下标最大的项 F_{2k+3} 换成 $F_{2k+2} + F_{2k+1}$：

$$\begin{aligned} F_{2k+3}F_{2k+1} - F_{2k+2}^2 &= (F_{2k+2} + F_{2k+1})F_{2k+1} - F_{2k+2}^2 \\ &= F_{2k+1}^2 + F_{2k+2}\left(F_{2k+1} - F_{2k+2}\right) \\ &= F_{2k+1}^2 - F_{2k+2}F_{2k}, \end{aligned}$$

因为由递推关系可知 $F_{2k+2} - F_{2k+1} = F_{2k}$。这就是 $-F_{2k+2}F_{2k} + F_{2k+1}^2$，正如所求。

第 (iii) 部分

以 $P(n)$ 表示命题 $F_{2n+1}F_{2n-1} - F_{2n}^2 = 1$。

基础情况（*base case*）。当 $n = 1$ 时：$F_3F_1 - F_2^2 = 2 \cdot 1 - 1^2 = 1$，所以 $P(1)$ 成立。

归纳步骤（*inductive step*）。假设 $P(k)$ 对某个 $k \geq 1$ 成立。由第 (ii) 部分，

$$F_{2k+3}F_{2k+1} - F_{2k+2}^2 = F_{2k+1}^2 - F_{2k+2}F_{2k}.$$

> **想一想** 第 (ii) 部分只把下标减了一。怎样才能一路退回到 $P(k)$?

再用一次同样的技巧，这次拆开下标最大的项 $F_{2k+2} = F_{2k+1} + F_{2k}$：

$$\begin{aligned} F_{2k+1}^2 - F_{2k+2}F_{2k} &= F_{2k+1}^2 - (F_{2k+1} + F_{2k})F_{2k} \\ &= F_{2k+1}\left(F_{2k+1} - F_{2k}\right) - F_{2k}^2 \\ &= F_{2k+1}F_{2k-1} - F_{2k}^2, \end{aligned}$$

当中用到 $F_{2k+1} - F_{2k} = F_{2k-1}$（因为 $2k+1 \geq 3$，所以成立）。根据归纳假设（hypothesis）$P(k)$，这个式子等于 1。把以上两组等式连起来，

$$F_{2k+3}F_{2k+1} - F_{2k+2}^2 = 1,$$

这就是 $P(k+1)$。由归纳法，$P(n)$ 对所有 $n \geq 1$ 成立。

推论。把 $P(n)$ 移项，

$$F_{2n}^2 + 1 = F_{2n+1}F_{2n-1},$$

而 F_{2n-1} 是整数（由递推关系可知每个 F_m 都是整数），所以 $F_{2n}^2 + 1$ 可被 F_{2n+1} 整除。

第 (iv) 部分

由递推关系，$F_{2n-1} = F_{2n+1} - F_{2n}$。两边平方，

$$F_{2n-1}^2 = F_{2n+1}^2 - 2F_{2n+1}F_{2n} + F_{2n}^2,$$

所以

$$F_{2n-1}^2 + 1 = F_{2n+1}\left(F_{2n+1} - 2F_{2n}\right) + \left(F_{2n}^2 + 1\right).$$

第一个括号所在的一项是 F_{2n+1} 的倍数（它的余因子（cofactor）是整数），而由第 (iii) 部分可知 $F_{2n}^2 + 1$ 也是 F_{2n+1} 的倍数。因此 $F_{2n-1}^2 + 1$ 可被 F_{2n+1} 整除。

事实上我们还可以求出商（quotient）。利用 $F_{2n}^2 + 1 = F_{2n+1}F_{2n-1}$，

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

而当 $n \geq 2$ 时，括号内的式子是 $F_{2n-1} - F_{2n-2} = F_{2n-3}$。所以 $F_{2n-1}^2 + 1 = F_{2n+1}F_{2n-3}$，与 (iii) 中的恒等式（identity）十分相似。

验算

取 $n = 3$：$F_6^2 + 1 = 65 = 13 \times 5 = F_7F_5$，而 $F_5^2 + 1 = 26 = 13 \times 2 = F_7F_3$。取 $n = 4$：$F_7^2 + 1 = 170 = 34 \times 5 = F_9F_5$。我们亦用电脑验证过，对于直至 37 的每一个 n，(ii) 和 (iii) 中的恒等式、两个整除性命题以及商 F_{2n-3} 全部成立。

常见错误

- 只用 (ii) 来做归纳：它只把下标减一，所得的式子 $F_{2k+1}^2 - F_{2k+2}F_{2k}$ 还不是归纳假设，必须再用一次递推关系。
- 忘记基础情况，或验证时用错下标（$n = 1$ 涉及 F_1、F_2、F_3）。

STEP 之外

第 (iii) 部分是卡西尼恒等式（Cassini's identity）$F_{m+1}F_{m-1} - F_m^2 = (-1)^m$ 的一半，而这条恒等式用行列式（determinant）即可立刻得出：矩阵 $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^m$ 的元素是 F_{m+1}、F_m、F_m、F_{m-1}，而它的行列式是 $(-1)^m$。像 (iv) 这类整除性的结果，是线性递推数列（linear recurrence）算术的起点，当中大部分工作都靠同余（congruence）完成。想进一步学习数论（number theory），Davenport 的 *The Higher Arithmetic*（高等算术）是很好的下一步；想了解矩阵的观点，则可读 Strang 的 *Introduction to Linear Algebra*（线性代数导论）。

详细解答

STEP III 2007，第 4 题

纯数学 · 参数曲线与曲率

STEP III 2007，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/07-s3-q4

关键思路

一切都建基于 $\dot{x}$ 的一条简洁公式。看似棘手的一项 $\ln\tan\frac{1}{2}t$ 求导后得 $1/\sin t$，于是 $\dot{x}$ 便化简为 $a\cos^2 t/\sin t$。一旦求得 $\dot{x}$ 和 $\dot{y}$，斜率（gradient）、切线（tangent）、曲率（curvature）和法线（normal）都可以用直接的坐标几何逐一得出。把单位切向量（unit tangent）写成 $(\cos t, \sin t)$，可令最后一部分十分简短。

解答

斜率

由于 $\dfrac{\mathrm{d}}{\mathrm{d}t}\ln\tan\frac{1}{2}t = \dfrac{\frac{1}{2}\sec^2\frac{1}{2}t}{\tan\frac{1}{2}t} = \dfrac{1}{2\sin\frac{1}{2}t\cos\frac{1}{2}t} = \dfrac{1}{\sin t}$，我们得到

$$\dot{x} = a\left(-\sin t + \frac{1}{\sin t}\right) = a\,\frac{1-\sin^2 t}{\sin t} = \frac{a\cos^2 t}{\sin t},$$
$$\dot{y} = a\cos t.$$

当 $0 < t < \frac{1}{2}\pi$ 时两者皆为正数，所以

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\dot{y}}{\dot{x}} = a\cos t\cdot\frac{\sin t}{a\cos^2 t} = \tan t,$$

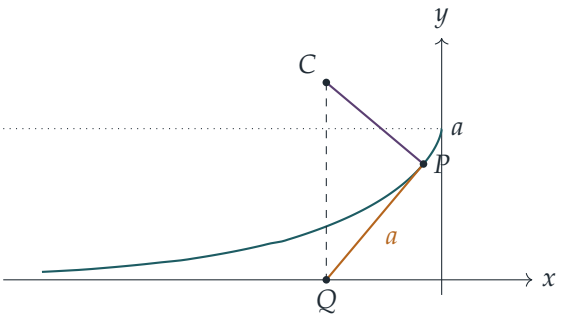
正如题目所求。

草图

▶ 当 $t \to 0^+$ 时，$\tan\frac{1}{2}t \to 0^+$，所以 $x \to -\infty$ 而 $y \to 0$。x 轴是一条渐近线（asymptote）。

- 当 $t \to \frac{1}{2}\pi^-$ 时，$x \to a(0+\ln 1) = 0$ 而 $y \to a$。曲线终止于 $(0,a)$，该处的斜率 $\tan t \to \infty$，所以该处的切线是垂直的。
- 由于 $\dot{x} > 0$，x 随 t 递增，斜率 $\tan t$ 亦然。所以曲线稳步上升，并且向上凹（concave upwards），而 y 介乎 0 与 a 之间。

这条曲线就是曳物线（tractrix）。图中亦标示了后面各部分所用的点 P、Q 和 C。



切线：$PQ = a$

在 $P = (x,y)$ 的切线斜率为 $\tan t$。由 P 沿水平方向往回走 $y/\tan t$ 的距离，切线便与 $y = 0$ 相交，所以

$$x_Q = x - \frac{y}{\tan t} = x - \frac{a\sin t\cos t}{\sin t} = x - a\cos t.$$

因此 $\overrightarrow{QP} = (a\cos t,\ a\sin t)$，并且

$$PQ = \sqrt{a^2\cos^2 t + a^2\sin^2 t} = a.$$

所以由曲线至 x 轴的切线长度恒为 a。（留待后用：$x_Q = a\ln\tan\frac{1}{2}t$。）

曲率半径

把 $\dot{x}$ 写成 $\dot{x} = a\Big(\dfrac{1}{\sin t} - \sin t\Big)$，求导便容易得多：

$$\ddot{x} = a\Big(-\frac{\cos t}{\sin^2 t} - \cos t\Big), \qquad \ddot{y} = -a\sin t.$$

于是

$$\begin{aligned}\dot{x}\ddot{y} - \dot{y}\ddot{x} &= \frac{a\cos^2 t}{\sin t}(-a\sin t) + a\cos t\cdot a\Big(\frac{\cos t}{\sin^2 t} + \cos t\Big)\\ &= a^2\Big(-\cos^2 t + \frac{\cos^2 t}{\sin^2 t} + \cos^2 t\Big) = a^2\cot^2 t,\end{aligned}$$

以及

$$\dot{x}^2+\dot{y}^2 = a^2\cos^2 t\left(\frac{\cos^2 t}{\sin^2 t}+1\right) = \frac{a^2\cos^2 t}{\sin^2 t} = a^2\cot^2 t.$$

由于在我们的范围内 $\cot t > 0$，故 $(\dot{x}^2+\dot{y}^2)^{3/2} = a^3\cot^3 t$，因此

$$\rho = \frac{a^3\cot^3 t}{a^2\cot^2 t} = a\cot t,$$

正如题目所求。

曲率中心：CQ 是垂直的

在 P 的切线沿单位向量 $(\cos t, \sin t)$ 的方向，所以法线沿 $(-\sin t, \cos t)$ 或其反方向。

> ? **想一想** 两个法线方向之中，哪一个是“在曲线上方”？

由于 $\cos t > 0$，向量 $(-\sin t, \cos t)$ 的 y 分量为正数。由 P 沿这个方向走一小段，便会到达曲线上方，因为曲线向上凹（而且无论如何，另一个方向只会把我们带往下方）。所以

$$\overrightarrow{OC} = \overrightarrow{OP} + \rho(-\sin t, \cos t) = \big(x - a\cot t\sin t,\ a\sin t + a\cot t\cos t\big).$$

根据切线部分的结果，其 x 坐标为

$$x_C = x - a\cos t = x_Q$$

。所以 C 和 Q 的 x 坐标相同，CQ 平行于 y 轴，正如题目所求。（y 坐标可化简为 $a\sin t + \dfrac{a\cos^2 t}{\sin t} = \dfrac{a}{\sin t}$，所以 $C = \big(a\ln\tan\tfrac{1}{2}t,\ a\operatorname{cosec} t\big)$。）

验算

- 取 $a = 1$ 和 $t = \frac{5\pi}{18}$（$50°$），得 $P \approx (-0.1202, 0.7660)$ 和 $Q \approx (-0.7630, 0)$，由此 $PQ = 1.0000$。公式给出 $C \approx (-0.7630, 1.3054)$，正好在 Q 的正上方。
- 以 2000 组随机的 a 和 t 值作数值检验，其中 $\dot{x}$、$\ddot{x}$ 以有限差分（finite differences）求得，结果与 $\mathrm{d}y/\mathrm{d}x = \tan t$、$PQ = a$、$\rho = a\cot t$ 和 $x_C = x_Q$ 相符，误差只在舍入误差范围之内。
- 当 $t \to \frac{1}{2}\pi$ 时，$\rho \to 0$：曲线趋近 $(0, a)$ 时弯得愈来愈急（它对 y 轴的镜像与它在该处相接，形成一个尖点（cusp））。当 $t \to 0$ 时，$\rho \to \infty$，与曲线逐渐贴平于渐近线的情况吻合。

常见错误

- 对 $\ln\tan\frac{1}{2}t$ 求导时，漏掉由链式法则（chain rule）而来的因子 $\frac{1}{2}$。答案必须是 $1/\sin t$。

- 把 $\dot{x}$ 留作 $a(-\sin t + \operatorname{cosec} t)$ 而不加化简。正是 $a\cos^2 t / \sin t$ 这个形式，才令其后每一部分都整齐利落。
- 采用法线方向 $(\sin t, -\cos t)$，这个方向指向曲线下方，会把 C 放在错误的位置。

STEP 之外

设有人沿 x 轴行走，用一条长度为 a 的绳拖着一件重物，重物所走的路径就是曳物线，而 $PQ = a$ 正是这个性质。点 C 描出曲线的*渐屈线*（evolute），即其曲率中心的轨迹（locus）。这里 $C = (a\ln\tan\frac{1}{2}t,\ a\operatorname{cosec} t)$ 原来位于一条悬链线（catenary）上。把曳物线绕其渐近线旋转，便得到伪球面（pseudosphere），这是一个具有恒定负曲率的曲面，也是双曲几何（hyperbolic geometry）的一个模型。Needham 的 *Visual Complex Analysis* 对伪球面有一段优美的几何论述。

详细解答

STEP III 2007，第 5 题

纯数学 · 高阶导数与数学归纳法

STEP III 2007，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/07-s3-q5

关键思路

只要把 r 和 $\mathrm{d}\theta/\mathrm{d}x$ 都单用 θ 表示，一切便变得简单。由 $\coth\theta = x$ 可得 $r = 1/\sinh\theta$，继而有 $\mathrm{d}r/\mathrm{d}x = \cosh\theta$ 及 $\mathrm{d}\theta/\mathrm{d}x = -\sinh^2\theta$。对 $\cosh n\theta / r^n$ 求导，所得的恰好就是 $\cosh(n+1)\theta$ 的复角公式（compound-angle formula），于是便可用归纳法（induction）证明。

解答

准备工作

由于 $x > 1$，方程 $\coth\theta = x$ 有唯一满足 $\theta > 0$ 的解，因此 $\sinh\theta > 0$。利用 $\coth^2\theta - 1 = \operatorname{cosech}^2\theta$，

$$r^2 = x^2 - 1 = \coth^2\theta - 1 = \frac{1}{\sinh^2\theta}, \qquad \text{所以} \qquad r = \frac{1}{\sinh\theta},$$

这里取正根，因为 r 和 $\sinh\theta$ 都是正数。

整个解答只靠两个导数（derivative）。把 $\coth\theta = x$ 对 x 求导，

$$-\operatorname{cosech}^2\theta\,\frac{\mathrm{d}\theta}{\mathrm{d}x} = 1 \quad \Longrightarrow \quad \frac{\mathrm{d}\theta}{\mathrm{d}x} = -\sinh^2\theta = -\frac{1}{r^2}.$$

把 $r^2 = x^2 - 1$ 求导，得 $2r\,\dfrac{\mathrm{d}r}{\mathrm{d}x} = 2x$，所以

$$\frac{\mathrm{d}r}{\mathrm{d}x} = \frac{x}{r} = \coth\theta\,\sinh\theta = \cosh\theta.$$

首两个导数

由于 $y = \ln r^2 = 2\ln r$,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{2}{r}\frac{\mathrm{d}r}{\mathrm{d}x} = \frac{2\cosh\theta}{r},$$

正是所求。至于二阶导数，最简洁的做法是写成 $1/r = \sinh\theta$，于是 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = 2\cosh\theta\sinh\theta = \sinh 2\theta$。由链式法则（chain rule），

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = 2\cosh 2\theta\,\frac{\mathrm{d}\theta}{\mathrm{d}x} = -\frac{2\cosh 2\theta}{r^2},$$

正是所求。

三阶导数

现在把二阶导数写成 $\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2} = -2\cosh 2\theta\sinh^2\theta$，并运用积法则（product rule），其中 $\mathrm{d}\theta/\mathrm{d}x = -\sinh^2\theta$:

$$\begin{aligned}\frac{\mathrm{d}^3y}{\mathrm{d}x^3} &= -2\big(2\sinh 2\theta\sinh^2\theta + 2\cosh 2\theta\sinh\theta\cosh\theta\big)\frac{\mathrm{d}\theta}{\mathrm{d}x}\\ &= 4\sinh^3\theta\big(\cosh 2\theta\cosh\theta + \sinh 2\theta\sinh\theta\big)\\ &= 4\sinh^3\theta\cosh 3\theta,\end{aligned}$$

最后一步用了恒等式 $\cosh(A+B) = \cosh A\cosh B + \sinh A\sinh B$。所以

$$\frac{\mathrm{d}^3y}{\mathrm{d}x^3} = \frac{4\cosh 3\theta}{r^3}.$$

一般公式

从 $2\cosh\theta/r$、$-2\cosh 2\theta/r^2$、$4\cosh 3\theta/r^3$ 这个规律可见，系数应为 $2\cdot 0!$、$-2\cdot 1!$、$2\cdot 2!$，因此我们猜想

$$\frac{\mathrm{d}^ny}{\mathrm{d}x^n} = (-1)^{n-1}\,\frac{2\,(n-1)!\,\cosh n\theta}{r^n} \qquad (n \geq 1).$$

? **想一想** 刚才求三阶导数的那个计算，推广到一般情况会是怎样的?

那就是求 $\cosh n\theta/r^n = \cosh n\theta\,\sinh^n\theta$ 的导数。由积法则和链式法则，并利用 $\mathrm{d}\theta/\mathrm{d}x = -\sinh^2\theta$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\big(\cosh n\theta\,\sinh^n\theta\big) &= \big(n\sinh n\theta\,\sinh^n\theta + n\cosh n\theta\,\sinh^{n-1}\theta\cosh\theta\big)\big(-\sinh^2\theta\big)\\ &= -n\sinh^{n+1}\theta\,\big(\cosh n\theta\cosh\theta + \sinh n\theta\sinh\theta\big)\\ &= -n\sinh^{n+1}\theta\,\cosh(n+1)\theta,\end{aligned}$$

即

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\cosh n\theta}{r^n}\right) = -\frac{n\cosh(n+1)\theta}{r^{n+1}}.$$

现在进行归纳。公式对 $n = 1$ 成立（它就是上面求得的一阶导数）。假设公式对某个 $n = k \geq 1$ 成立。那么，把上面的计算用于 $n = k$，

$$\frac{\mathrm{d}^{k+1}y}{\mathrm{d}x^{k+1}} = (-1)^{k-1}\,2\,(k-1)!\cdot\left(-\frac{k\cosh(k+1)\theta}{r^{k+1}}\right) = (-1)^k\,\frac{2\,k!\,\cosh(k+1)\theta}{r^{k+1}},$$

这正是 $n = k+1$ 时的公式。由归纳法，

$$\boxed{\frac{\mathrm{d}^n y}{\mathrm{d}x^n} = (-1)^{n-1}\,\frac{2\,(n-1)!\,\cosh n\theta}{r^n} \quad \text{对所有} n \geq 1.}$$

（$n = 2$ 和 $n = 3$ 的情况与我们直接求得的结果一致。）

另一途径

部分分式（partial fractions）提供了一个独立的推导。由于 $y = \ln(x-1) + \ln(x+1)$，

$$\frac{\mathrm{d}^n y}{\mathrm{d}x^n} = (-1)^{n-1}(n-1)!\left[(x-1)^{-n} + (x+1)^{-n}\right].$$

但 $x \pm 1 = \dfrac{\cosh\theta \pm \sinh\theta}{\sinh\theta} = r\,\mathrm{e}^{\pm\theta}$，所以方括号内的式子等于 $r^{-n}\left(\mathrm{e}^{-n\theta} + \mathrm{e}^{n\theta}\right) = 2\cosh n\theta / r^n$，得出相同的公式。这也解释了这个代换的来历：$\theta$ 就是 $\ln\dfrac{x+1}{x-1}$ 的一半。

验算

- 当 $x = 2$ 时：由部分分式得 $\mathrm{d}^3y/\mathrm{d}x^3 = 2\left(1 + \frac{1}{27}\right) = \frac{56}{27} \approx 2.0741$，而以 $\theta = \tanh^{-1}\frac{1}{2}$、$r = \sqrt{3}$ 代入 $4\cosh 3\theta / r^3$ 亦得相同的值。
- 我们在 200 个满足 $x > 1$ 的随机点上，把首三个导数与 $\ln(x^2-1)$ 的有限差分（finite differences）作了比较，又就 $n \leq 14$ 把一般公式与部分分式的形式作了比较（两者相符至约 10^{-14}）。
- 正负号：当 $x > 1$ 时，$\ln(x^2-1)$ 递增而且是凹的（concave），与 $y' > 0$ 及 $y'' < 0$ 吻合。

常见错误

- 写成 $\mathrm{d}\theta/\mathrm{d}x = +\sinh^2\theta$：$\coth\theta$ 的导数是 $-\operatorname{cosech}^2\theta$，而正是这个负号令各阶导数的正负号交替出现。

- 阶乘（factorial）差了一位：第 n 阶导数所带的是 $(n-1)!$，因为一阶导数的系数是 $2 = 2 \cdot 0!$。
- 只验证几个情况便停下来：题目要求“with proof”（附证明），就必须写出一般的归纳步骤。

STEP 之外

每一阶导数都有如此简洁的闭合形式（closed form），让我们一窥复分析（complex analysis）的面貌：$\ln(x^2-1)$ 在 $x = \pm 1$ 处有奇点（singularities），而它的高阶导数正是由到这些奇点的距离所支配，在这里就是 $x \mp 1 = re^{\mp\theta}$。在大学一年级的分析课程中，要估计泰勒级数（Taylor series）余项的界，并求出其收敛半径（radius of convergence），所需要的正是这类公式。Needham 的 *Visual Complex Analysis* 把奇点为何支配幂级数（power series）解释得十分精彩。

详细解答

STEP III 2007，第 6 题

纯数学 · 复数与圆

STEP III 2007，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/07-s3-q6

关键思路

在以原点为圆心、半径为 a 的圆上，取共轭（conjugation）与反演（inversion）是同一回事：$pp^* = a^2$，所以 $p^* = a^2/p$。单凭这一点，题目中每一个共轭都可化为普通的分式，恒等式（identity）随即得出。至于垂直，判别方法是两条弦所对应的向量之比为纯虚数（purely imaginary）。利用恒等式，这个判别法便化为一个简单的乘积条件。对于多边形，这个乘积条件就是一个递推关系（recurrence）：一旦选定 B_1，其余每一个 B_k 都随之确定，唯一的问题是这条链绕回 B_1 时能否闭合。

解答

恒等式

由于 P 在圆上，$|p| = a$，所以 $pp^* = a^2$，即 $p^* = a^2/p$；同理 $q^* = a^2/q$。因此

$$p^* - q^* = \frac{a^2}{p} - \frac{a^2}{q} = \frac{a^2(q-p)}{pq} = -\frac{a^2(p-q)}{pq}.$$

因为 $p \neq q$（而且由 $a > 0$ 知 $p, q \neq 0$），上式不等于零，所以可以移项得

$$pq = -a^2 \frac{p-q}{p^* - q^*},$$

这正是所求。

互相垂直的弦

弦（chord）PQ 的方向由复数 $p-q$ 给出，而 RS 的方向由 $r-s$ 给出。两个非零复数互相垂直，当且仅当它们的比

$$w=\frac{p-q}{r-s}$$

是纯虚数，因为这个比的辐角（argument）就是两者之间的夹角。

> **想一想** 怎样把“w 是纯虚数”化为一条可以运用的方程?

非零的 w 是纯虚数，当且仅当 $w^*=-w$，即 $w/w^*=-1$。对两条弦分别使用恒等式（第二条以 r,s 代替 p,q），

$$\frac{w}{w^*}=\frac{p-q}{p^*-q^*}\cdot\frac{r^*-s^*}{r-s}=\frac{-pq/a^2}{-rs/a^2}=\frac{pq}{rs}.$$

所以 $PQ\perp RS$ 当且仅当 $pq/rs=-1$。由于 $rs\neq 0$，这就是说

$$pq+rs=0.$$

这里每一步都可逆，所以我们证明的比题目要求的稍多：对于这个圆的弦，$PQ\perp RS\iff pq+rs=0$。下文需要用到逆命题（converse）这个方向。

多边形问题

设点 A_k 和 B_k 分别由 α_k 和 β_k 表示。根据上一部分（及其逆命题），那 n 个垂直条件恰好就是

$$\beta_k\beta_{k+1}=-\alpha_k\alpha_{k+1}\quad(k=1,\dots,n-1),\qquad \beta_n\beta_1=-\alpha_n\alpha_1,$$

前提是相邻的 B 互不相同，使每条 B_kB_{k+1} 确实是一条弦。

选定 β_1。由于 $\beta_{k+1}=-\alpha_k\alpha_{k+1}/\beta_k$，首 $n-1$ 个条件便依次确定 $\beta_2,\beta_3,\dots,\beta_n$，而且每个 β_{k+1} 的模（modulus）自动等于 a。算出最初几个：

$$\beta_2=-\frac{\alpha_1\alpha_2}{\beta_1},\qquad \beta_3=-\frac{\alpha_2\alpha_3}{\beta_2}=\frac{\alpha_3\beta_1}{\alpha_1},\qquad \beta_4=-\frac{\alpha_3\alpha_4}{\beta_3}=-\frac{\alpha_1\alpha_4}{\beta_1},$$

再用简单的归纳法（induction）可得

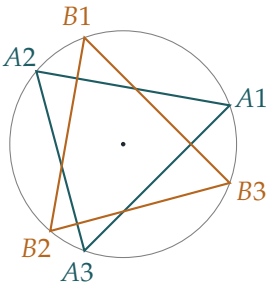
$$\beta_k=\frac{\alpha_k\beta_1}{\alpha_1}\ (k\text{ 为奇数}),\qquad \beta_k=-\frac{\alpha_1\alpha_k}{\beta_1}\ (k\text{ 为偶数}).$$

（若公式对 k 成立，则由 $\beta_{k+1}=-\alpha_k\alpha_{k+1}/\beta_k$ 可得 $k+1$ 时的另一种形式。）现在一切取决于最后一个条件 $\beta_n\beta_1=-\alpha_n\alpha_1$ 是否成立。

$n=3$。这时 $\beta_3=\alpha_3\beta_1/\alpha_1$，所以闭合条件是

$$\frac{\alpha_3\beta_1^2}{\alpha_1}=-\alpha_3\alpha_1\iff\beta_1^2=-\alpha_1^2\iff\beta_1=\pm\mathrm{i}\,\alpha_1.$$

所以 B_1 只有两个选择。取 $\beta_1 = \mathrm{i}\alpha_1$ 时，由公式得 $\beta_2 = \mathrm{i}\alpha_2$ 和 $\beta_3 = \mathrm{i}\alpha_3$；取 $\beta_1 = -\mathrm{i}\alpha_1$ 时，每个 $\beta_k = -\mathrm{i}\alpha_k$。两种情况下，三角形 $B_1B_2B_3$ 都是三角形 $A_1A_2A_3$ 绕圆心旋转一个直角而成。它的顶点互不相同，所以两个选择都确实可行；而把一条弦旋转 $90°$，所得的弦当然与原来的弦垂直。



图中所示的是 $\beta_1 = \mathrm{i}\alpha_1$ 这个选择：琥珀色三角形的每一边，都是蓝绿色三角形的对应边转过 $90°$ 而成。

$n = 4$。这时 $\beta_4 = -\alpha_1\alpha_4/\beta_1$，所以不论 β_1 是甚么，

$$\beta_4\beta_1 = -\alpha_1\alpha_4$$

都自动成立。因此 B_1 可以选在圆上任何位置，有无穷多个选择。唯一的例外，是会令两个相邻的 B 重合的有限多个位置。例如，$\beta_1 = \beta_2$ 需要 $\beta_1^2 = -\alpha_1\alpha_2$，这排除了两个点。四种重合情况各自最多排除两个点，所以最多有八个位置被排除。

$n > 4$。同样的计算对每个 n 都适用，而且只有 n 的奇偶性（parity）起作用：

- 若 n 是奇数，闭合条件是 $\beta_1^2 = -\alpha_1^2$。恰好有两个选择 $\beta_1 = \pm\mathrm{i}\alpha_1$，而 $B_1B_2\cdots B_n$ 就是多边形 $A_1A_2\cdots A_n$ 绕圆心旋转 $\pm 90°$ 而成；
- 若 n 是偶数，闭合条件自动成立。除了最多 $2n$ 个会令相邻的 B 重合的位置外，B_1 可以是圆上任何一点，所以有无穷多个选择。

验算

- 试取 $p = a$，$q = \mathrm{i}a$：这时 $pq = \mathrm{i}a^2$，而 $-a^2(a - \mathrm{i}a)/(a + \mathrm{i}a) = -a^2 \cdot (-\mathrm{i}) = \mathrm{i}a^2$，与预期相符。
- 试取直径 $p = a$，$q = -a$ 和 $r = \mathrm{i}a$，$s = -\mathrm{i}a$，两者互相垂直：$pq + rs = -a^2 + a^2 = 0$。
- 数值方面，我们以 1000 个随机构形测试了恒等式和判别准则 $PQ \perp RS \iff pq + rs = 0$。对 $n = 3, \ldots, 9$，我们由随机的起点出发执行递推。当 n 是奇数时，让 B_1 以 $0.1°$ 为步长绕圆扫描一周，这条链只在 $\beta_1 = \pm\mathrm{i}\alpha_1$ 附近闭合。当 n 是偶数时，它每次都闭合，而且所有弦都互相垂直，误差仅为舍入误差。

常见错误

- 只证明“垂直 $\Rightarrow pq + rs = 0$”。处理多边形时还需要逆命题，才能肯定我们构造出来的 β 确实给出互相垂直的弦。
- 把“垂直即 w 是虚数”写成 $w + w^* = 0$，却没有指出 $w \neq 0$。这里由于各点互不相同，$w \neq 0$ 是有保证的。
- 当 n 是偶数时，说“每个 B_1 都可行”，却没有留意到有少数位置会令两个相邻的 B 重合，以致那条“弦”没有方向。

STEP 之外

映射 $z \mapsto a^2/z^*$ 是关于圆 $|z| = a$ 的*反演*，而圆上的点在这个映射下保持不动（$p = a^2/p^*$），正是整道题目的动力所在。反演是莫比乌斯变换（Möbius transformation）的一种，而莫比乌斯变换在复分析与几何的入门课程中占有核心地位。多边形的论证带有*单值性*（monodromy）的味道：沿着一连串别无选择的步骤绕一圈，然后问是否回到起点，而答案取决于 n 的奇偶性。Needham 的 *Visual Complex Analysis*（《可视化复分析》）把反演和这些几何思想阐述得十分精彩。

详细解答

STEP III 2007，第 7 题

纯数学 · 代换积分法

STEP III 2007，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/07-s3-q7

关键思路

题目不容许我们提及“arctan”或“π”，所以每一项事实都必须靠积分内的变量代换（substitution）得出。每个代换都经过挑选，令被积函数（integrand）$1/(1+u^2)$（或 $1/\sqrt{1-u^2}$）变回自己，或变成另一个，只有积分上下限（limits）改变。比较上下限，便得到 t 与 s 各个值之间的等式；而第 (i) 部分的对称性（symmetry）$u \mapsto 1/u$，正是把这些等式化为 p 的倍数的关键。

解答

第 (i) 部分

取 $x > 0$，并设 $u = v^{-1}$，于是 $\mathrm{d}u = -v^{-2}\,\mathrm{d}v$。当 u 由 0 增至 x 时，v 由 ∞ 减至 $1/x$。另外

$$\frac{1}{1+u^2} = \frac{1}{1+v^{-2}} = \frac{v^2}{v^2+1}.$$

因此

$$t(x) = \int_{\infty}^{1/x} \frac{v^2}{1+v^2}\left(-\frac{1}{v^2}\right)\mathrm{d}v = \int_{1/x}^{\infty} \frac{1}{1+v^2}\,\mathrm{d}v,$$

其中负号用来对调上下限。这就是所求的结果。（这个广义积分（improper integral）收敛（converges），因为被积函数不超过 $1/v^2$。）

以 $1/x$ 代替 x 套用上式，得 $t(1/x) = \int_x^{\infty} \frac{1}{1+v^2}\,\mathrm{d}v$。加上 $t(x) = \int_0^x \frac{1}{1+v^2}\,\mathrm{d}v$，两个积分范围便合并成 $[0,\infty)$：

$$t(x) + t(1/x) = \int_0^{\infty} \frac{1}{1+v^2}\,\mathrm{d}v = \tfrac{1}{2}p \qquad (x > 0).$$

代入 $x=1$，得 $2t(1)=\frac{1}{2}p$。

第 (ii) 部分

当 $u\geq 0$ 时，有 $0\leq y<1$。两边平方，$y^2(1+u^2)=u^2$，所以 $u^2(1-y^2)=y^2$，而

$$u=\frac{y}{\sqrt{1-y^2}},$$

这里取正根，因为 u 与 y 同号。由商法则（quotient rule），

$$\frac{\mathrm{d}u}{\mathrm{d}y}=\frac{\sqrt{1-y^2}+\dfrac{y^2}{\sqrt{1-y^2}}}{1-y^2}=\frac{(1-y^2)+y^2}{(1-y^2)^{3/2}}=\frac{1}{\sqrt{(1-y^2)^3}},$$

正如所求。

现在代入 $t(x)$。由 $u^2=y^2/(1-y^2)$ 得

$$1+u^2=\frac{1}{1-y^2},$$

所以

$$\frac{1}{1+u^2}\frac{\mathrm{d}u}{\mathrm{d}y}=(1-y^2)\cdot\frac{1}{(1-y^2)^{3/2}}=\frac{1}{\sqrt{1-y^2}}.$$

映射 $u\mapsto y$ 是递增（increasing）的（它的反函数的导数为正），而且 $u=0\mapsto y=0$，$u=x\mapsto y=x/\sqrt{1+x^2}$，后者小于 1。所以

$$t(x)=\int_0^{x/\sqrt{1+x^2}}\frac{1}{\sqrt{1-y^2}}\,\mathrm{d}y=s\left(\frac{x}{\sqrt{1+x^2}}\right).$$

取 $x=1$，并利用第 (i) 部分，上式即 $s\left(\frac{1}{\sqrt{2}}\right)=t(1)=\frac{1}{4}p$。

第 (iii) 部分

记 $a=\frac{1}{\sqrt{3}}$，于是 $z=\dfrac{u+a}{1-au}$。在 $0\leq u\leq a$ 上有 $au\leq\frac{1}{3}$，所以分母保持为正，z 在该范围内是 u 的光滑函数（smooth function）。求导得

$$\frac{\mathrm{d}z}{\mathrm{d}u}=\frac{(1-au)+a(u+a)}{(1-au)^2}=\frac{1+a^2}{(1-au)^2}.$$

? **想一想** 为甚么这个代换会令 $1/(1+u^2)$ 保持不变?

通分计算 $1+z^2$。交叉项 $\mp 2au$ 互相抵消：

$$1+z^2=\frac{(1-au)^2+(u+a)^2}{(1-au)^2}=\frac{1+a^2u^2+u^2+a^2}{(1-au)^2}=\frac{(1+a^2)(1+u^2)}{(1-au)^2}.$$

两式相除，

$$\frac{1}{1+z^2}\frac{\mathrm{d}z}{\mathrm{d}u}=\frac{1}{1+u^2},$$

所以被积函数不变，只有上下限移动。（这背后其实是正切的复角公式（tangent addition formula），但我们用不着它。）当 $u=0$ 时，$z=a=\frac{1}{\sqrt{3}}$。当 $u=a$ 时，

$$z=\frac{2/\sqrt{3}}{1-\frac{1}{3}}=\frac{2}{\sqrt{3}}\cdot\frac{3}{2}=\sqrt{3}.$$

因此

$$t\left(\tfrac{1}{\sqrt{3}}\right)=\int_0^{1/\sqrt{3}}\frac{\mathrm{d}u}{1+u^2}=\int_{1/\sqrt{3}}^{\sqrt{3}}\frac{\mathrm{d}z}{1+z^2}.$$

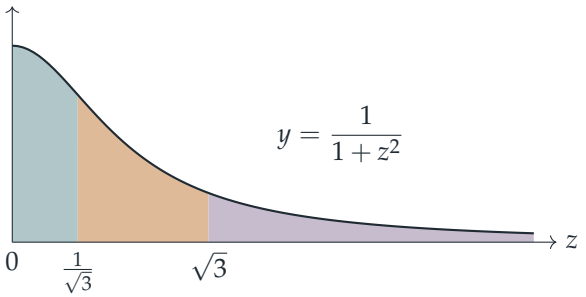
最后一步，把范围 $[0,\infty)$ 在 $\frac{1}{\sqrt{3}}$ 和 $\sqrt{3}$ 处分开：

$$\tfrac{1}{2}p=\int_0^{1/\sqrt{3}}\frac{\mathrm{d}z}{1+z^2}+\int_{1/\sqrt{3}}^{\sqrt{3}}\frac{\mathrm{d}z}{1+z^2}+\int_{\sqrt{3}}^{\infty}\frac{\mathrm{d}z}{1+z^2}.$$

第一段按定义就是 $t(\frac{1}{\sqrt{3}})$，第二段按刚才证明的结果也是 $t(\frac{1}{\sqrt{3}})$。由第 (i) 部分取 $x=\frac{1}{\sqrt{3}}$，第三段是 $\int_{1/x}^{\infty}\frac{\mathrm{d}z}{1+z^2}=t(x)=t(\frac{1}{\sqrt{3}})$，同样如此。因此

$$3\,t\left(\tfrac{1}{\sqrt{3}}\right)=\tfrac{1}{2}p.$$

下图显示这三段：$1/(1+z^2)$ 之下由 0 至 ∞ 的面积，分成三个面积相等的区域。



验算

以数值方法（辛普森法则（Simpson's rule））计算，$t(\frac{1}{\sqrt{3}})\approx 0.5235988$，$\int_{1/\sqrt{3}}^{\sqrt{3}}\frac{\mathrm{d}z}{1+z^2}\approx 0.5235988$，而 $p\approx 3.1415927$，所以 $3t(\frac{1}{\sqrt{3}})\approx 1.5707963\approx\frac{1}{2}p$。同样，$s(\frac{1}{\sqrt{2}})\approx t(1)\approx 0.7853982\approx\frac{1}{4}p$。当然，$p$ 其实就是 π，而这些结果说的是 $\arctan\frac{1}{\sqrt{3}}=\frac{\pi}{6}$，$\arcsin\frac{1}{\sqrt{2}}=\arctan 1=\frac{\pi}{4}$；你可以用这些来验算，但不可写在答案里。

常见错误

- 忘记 $u = 1/v$ 会把积分方向倒转；du 中的负号正好把上下限的次序还原。
- 在第 (iii) 部分，做完代换便停下来：由 $\frac{1}{\sqrt{3}}$ 至 $\sqrt{3}$ 的积分只是三段相等部分的其中一段，而 $\sqrt{3}$ 之后的那一段需要用到第 (i) 部分。
- 没有检查在积分范围内 $1 - au \neq 0$；这里 $au \leq \frac{1}{3}$，所以代换成立。

STEP 之外

这道题让我们一窥分析学家如何严格地定义三角函数（trigonometric functions）：先把 $\arctan x$ 定义为 $\int_0^x \frac{\mathrm{d}u}{1+u^2}$，再把 π 定义为它在无穷远处的极限（limit）的两倍，然后靠这类代换推导出所有性质，完全不用几何。第 (iii) 部分的代换其实是 tan 的复角公式的化身，用同样的方法可以证明：当 $xy < 1$ 时，$t(x) + t(y) = t\left(\frac{x+y}{1-xy}\right)$。Abbott 的 *Understanding Analysis* 说明了如何利用积分和幂级数（power series）为这类定义奠定稳固的基础。

详细解答

STEP III 2007，第 8 题

纯数学 · 里卡蒂方程

STEP III 2007，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/07-s3-q8

关键思路

非线性方程 $(*)$ 是一条里卡蒂方程（Riccati equation）：y^2 项破坏了线性。把 y 写成 u'/u 的倍数，$y'+y^2$ 便会变成 u''/u，平方项随之消失，剩下一条关于 u 的线性二阶方程。第 (i) 部分的前半已经给了我们这条线性方程的两个解，所以余下的工作只是定出常数。至于第 (ii) 部分，我们需要一条新线性方程的两个解，而线索在于：把 $1+x$ 中的 x 换成 $-x$，就得到 $1-x$。

解答

第 (i) 部分

求 a 和 b。当 $u=x$ 时，$u'=1$，$u''=0$，所以需要

$$\mathrm{a}(x)+x\,\mathrm{b}(x)=0.$$

当 $u=\mathrm{e}^{-x}$ 时，$u'=-\mathrm{e}^{-x}$，$u''=\mathrm{e}^{-x}$；除以 $\mathrm{e}^{-x}\neq 0$ 得

$$1-\mathrm{a}(x)+\mathrm{b}(x)=0.$$

由第一条方程得 $\mathrm{b}=-\mathrm{a}/x$（其中 $x\neq 0$），再由第二条方程得 $\mathrm{a}\left(1+\frac{1}{x}\right)=1$。所以

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

（这两个函数在 $x=0$ 时亦满足上述两个条件。）这条方程是线性而且齐次（homogeneous）的，而 x 与 e^{-x} 并非彼此的倍数，所以通解（general solution）是

$$u=Ax+B\mathrm{e}^{-x},$$

其中 A 和 B 为任意常数。

代换。设 $y = \dfrac{u'}{3u}$，其中 $u \neq 0$。由商法则（quotient rule），

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{u''}{3u} - \frac{(u')^2}{3u^2} = \frac{u''}{3u} - 3\left(\frac{u'}{3u}\right)^2 = \frac{u''}{3u} - 3y^2.$$

所以 $y' + 3y^2 = \dfrac{u''}{3u}$：平方项刚好完全抵销，这正是代换中要放入因子 3 的原因。方程 $(*)$ 变成

$$\frac{u''}{3u} + \frac{x}{1+x}\cdot\frac{u'}{3u} = \frac{1}{3(1+x)},$$

乘以 $3u$ 得

$$\frac{\mathrm{d}^2u}{\mathrm{d}x^2} + \frac{x}{1+x}\frac{\mathrm{d}u}{\mathrm{d}x} - \frac{1}{1+x}u = 0,$$

正如所求。这就是本部分开首的方程，其中的 a 和 b 正是我们刚才求得的。

特解（particular solution）。因此 $u = Ax + B\mathrm{e}^{-x}$，而

$$y = \frac{u'}{3u} = \frac{A - B\mathrm{e}^{-x}}{3(Ax + B\mathrm{e}^{-x})}.$$

> ? **想一想** 常数有两个，条件却只有一个。这会有问题吗？

不会。把 u 乘以一个非零常数并不会改变 u'/u，所以只有比 $A : B$ 才有影响，而一阶方程本来就应该只有一个自由常数。由条件“当 $x = 0$ 时 $y = 0$”得 $A - B = 0$。于是 $B = A \neq 0$（否则 $u \equiv 0$），约去 A 后，

$$y = \frac{1 - \mathrm{e}^{-x}}{3(x + \mathrm{e}^{-x})}.$$

分母永不为零，因为由 $\mathrm{e}^{-x} \geq 1 - x$ 可得 $x + \mathrm{e}^{-x} \geq 1$。第 (i) 部分至此完成。

第 (ii) 部分

现在 y^2 项的系数是 1，所以我们用 $y = \dfrac{u'}{u}$。与之前完全一样，

$$y' = \frac{u''}{u} - \left(\frac{u'}{u}\right)^2 = \frac{u''}{u} - y^2,$$

所以 $y' + y^2 = u''/u$。把方程乘以 u，便得到线性方程

$$u'' + \frac{x}{1-x}u' - \frac{1}{1-x}u = 0.$$

这就是把第 (i) 部分的方程中的 x 换成 $-x$ 所得的方程。事实上，若 $v(x) = w(-x)$，其中 w 是第 (i) 部分方程的解，则 $v'(x) = -w'(-x)$ 及 $v''(x) = w''(-x)$，而第 (i) 部分的方程在点 $-x$ 处就是

$$v''(x) + \frac{-x}{1-x}\big(-v'(x)\big) - \frac{1}{1-x}v(x) = 0,$$

这正是新的方程。所以由 x 和 e^{-x} 可得出解 $-x$（或干脆取 x）和 e^{x}。直接验证：当 $u=x$ 时，$0+\frac{x}{1-x}-\frac{x}{1-x}=0$；当 $u=\mathrm{e}^{x}$ 时，$\mathrm{e}^{x}\left(1+\frac{x}{1-x}-\frac{1}{1-x}\right)=\mathrm{e}^{x}(1-1)=0$。因此

$$u=Ax+B\mathrm{e}^{x},\qquad y=\frac{u'}{u}=\frac{A+B\mathrm{e}^{x}}{Ax+B\mathrm{e}^{x}}.$$

当 $x=0$ 时，上式等于 $(A+B)/B$，而它必须等于 2，所以 $A=B$（且 $B\neq 0$）。因此

$$y=\frac{1+\mathrm{e}^{x}}{x+\mathrm{e}^{x}}.$$

这个解在包含 0、满足 $x+\mathrm{e}^{x}>0$ 而且方程有定义的区间上成立，即 $x_0<x<1$，其中 $x_0\approx -0.567$ 是 $x+\mathrm{e}^{x}=0$ 的根。

验算

- 当 $x=0$ 时，第 (ii) 部分的答案给出 $2/1=2$；对它求导得 $y'(0)=\frac{1\cdot 1-2\cdot 2}{1}=-3$，与方程吻合：$1-4-0=-3$。
- 当 $x\to\infty$ 时，第 (i) 部分的答案趋向 0，其表现有如 $\frac{1}{3x}$。这时 $y'\approx -\frac{1}{3x^2}$ 与 $3y^2\approx\frac{1}{3x^2}$ 互相抵销，而 $\frac{x}{1+x}y\approx\frac{1}{3(1+x)}$，与 $(*)$ 一致。
- 我们以数值方法解了第 (ii) 部分的方程（由 $y(0)=2$ 出发，使用龙格－库塔法（Runge–Kutta）），得到 $y(0.9)=1.0297654\ldots$，与公式的结果在小数点后 10 位之内相符。

常见错误

- 在第 (ii) 部分仍然保留因子 $\frac{1}{3}$：乘数必须与 y^2 的系数配合，所以这里应该用 $y=u'/u$。
- 试图只凭一个条件同时定出 A 和 B。其实只有两者之比才会影响 y。
- 在第 (ii) 部分未经验证便再次取 e^{-x}：$1-x$ 中的正负号改变，令它变成 e^{x}。

STEP 之外

形如 $y'+P(x)y^2+Q(x)y=R(x)$ 的方程称为里卡蒂方程，而代换 $y=u'/(Pu)$ 总能把它化为一条线性二阶方程。反过来说，任何二阶线性方程的比 u'/u 都满足一条里卡蒂方程，这就是为甚么这类非线性方程在物理学中不断出现，由量子力学（quantum mechanics）以至控制理论（control theory）都有它们的踪影。Boyce 与 DiPrima 的 *Elementary Differential Equations*（常微分方程入门教材）涵盖了这个课题，以及线性二阶方程的理论（为甚么两个独立的解就能给出通解）。

详细解答

STEP III 2007，第 9 题

力学 · 简谐运动

STEP III 2007，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/07-s3-q9

关键思路

弹簧对两颗小珠的拉力或推力大小相等、方向相反，而且沿同一直线作用；圆环的反作用力（reaction）则沿径向，所以对圆心的总角动量（angular momentum）一直保持为零：两颗小珠的运动互为镜像。这样整个问题就简化为只有一个坐标，再由能量守恒（conservation of energy）便得出关于 θ 的简谐运动（simple harmonic motion，简称 SHM）。要留神的是，θ 是锐角的半角：如果两颗小珠分开得够远，越过了直径的两端，能量的公式就会改变形式，这时的运动必须由两段 SHM 拼合而成。

解答

建立模型

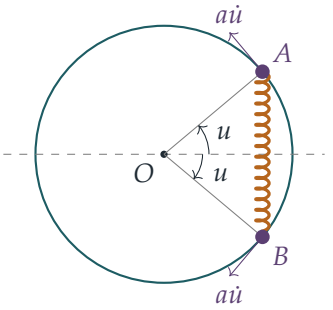
把小珠看成质点（particle），圆环固定而光滑（smooth），弹簧为轻（light）弹簧。作用于小珠的力有：重力和圆环反作用力的垂直分量（两者互相抵消）、反作用力的水平分量（由于圆环光滑，它沿半径方向），以及弹簧的力。弹簧作用于 A 的力和弹簧作用于 B 的力大小相等、方向相反，并且沿同一直线 AB 作用，所以它们对通过 O 的垂直轴的总力矩（moment）为零；径向的力对该轴则完全没有力矩。因此，对该轴的总角动量守恒，而由于小珠从静止开始，总角动量为零。

从一条固定的半径起沿圆环量度角度，设 A 和 B 所在的角度分别为 ϕ_A 和 ϕ_B。角动量为零即 $ma^2(\dot{\phi}_A + \dot{\phi}_B) = 0$，所以 $\phi_A + \phi_B$ 是常数。选取参考半径，使它平分初始的角 AOB；那么在任何时刻都有

$$\phi_A = u, \qquad \phi_B = -u,$$

其中只有一个坐标 u，而开始时 $u = \beta$。每颗小珠的速率都是 $a\dot{u}$，所以动能（kinetic

energy）是 $2\cdot\frac{1}{2}ma^2\dot{u}^2=ma^2\dot{u}^2$。



当 $0<u\le\frac{\pi}{2}$ 时，角 AOB 等于 $2u$；但如果 u 超过 $\frac{\pi}{2}$，小珠便已越过直径的两端，这时角 AOB 等于 $2\pi-2u$。所以

$$\theta=u\quad(0<u\le\tfrac{\pi}{2}),\qquad\theta=\pi-u\quad(\tfrac{\pi}{2}\le u<\pi).$$

小珠在 $u=\beta$ 处由静止释放，由能量守恒得

$$ma^2\dot{u}^2+mk^2a^2(\theta-\alpha)^2=mk^2a^2(\beta-\alpha)^2,\qquad\text{所以}\qquad\dot{u}^2=k^2\left[(\alpha-\beta)^2-(\theta-\alpha)^2\right].$$

“处于压缩状态”（in compression）是指弹簧比它的自然长度（natural length）短。能量在 $\theta=\alpha$ 时最小，所以那就是自然长度的位置；而长度 $2a\sin\theta$ 在 $(0,\frac{\pi}{2})$ 上随 θ 增加而增加，因此 $0<\beta<\alpha$。

$u\le\frac{\pi}{2}$ 时的运动

这时 $\theta=u$，把能量方程对 t 求导，再除以 $2\dot{u}$（只要小珠在运动，这一步就成立，其余时刻则由连续性补足），得

$$\ddot{\theta}=-k^2(\theta-\alpha).$$

这是以 $\theta=\alpha$ 为中心、角频率（angular frequency）为 k 的 SHM。由于小珠在 $\theta=\beta$ 处从静止出发，

$$\theta-\alpha=-(\alpha-\beta)\cos kt,$$

所以两颗小珠互相远离，如果中途没有任何事情打断，便会在 $t=\frac{\pi}{k}$ 时到达 $\theta=2\alpha-\beta$。一切都取决于 $2\alpha-\beta$ 是否小于 $\frac{\pi}{2}$，也就是取决于 β 与 $2\alpha-\frac{\pi}{2}$ 的大小关系。这个临界值落在 $(0,\alpha)$ 之内，正是因为 $\frac{\pi}{4}<\alpha<\frac{\pi}{2}$，所以以下三种情况都有可能出现。

情况 1：$2\alpha-\frac{\pi}{2}<\beta<\alpha$

这时 $2\alpha-\beta<\frac{\pi}{2}$，所以小珠永远不会到达直径的两端，全程都有 $\theta=u$，运动是在 $\theta=\beta$ 与 $\theta=2\alpha-\beta$ 之间的纯粹 SHM。周期（period）为

$$T=\frac{2\pi}{k}.$$

情况 2： $0 < \beta < 2\alpha - \frac{\pi}{2}$

这时 θ 在 SHM 掉头之前就到达 $\frac{\pi}{2}$，所需时间 t_1 满足

$$\frac{\pi}{2} - \alpha = -(\alpha - \beta)\cos kt_1, \qquad \text{所以} \qquad \cos kt_1 = -\lambda, \quad \text{其中}\lambda = \frac{\frac{\pi}{2} - \alpha}{\alpha - \beta}.$$

由于 $\alpha < \frac{\pi}{2}$ 且 $\beta < 2\alpha - \frac{\pi}{2}$，我们有 $0 < \lambda < 1$。因此

$$kt_1 = \pi - \arccos\lambda = \frac{\pi}{2} + \arcsin\lambda.$$

在这一刻，能量方程给出 $\dot{u}^2 = k^2[(\alpha - \beta)^2 - (\frac{\pi}{2} - \alpha)^2] > 0$，所以两颗小珠仍在互相远离，并会继续越过直径的两端。

> **想一想** 小珠越过直径的两端之后，会发生甚么事?

当 $u \geq \frac{\pi}{2}$ 时，我们有 $\theta = \pi - u$，所以能量方程变成

$$\dot{u}^2 = k^2\left[(\alpha - \beta)^2 - (u - (\pi - \alpha))^2\right],$$

这是 u 以 $u = \pi - \alpha$ 为中心的 SHM，角频率同样是 k。它是到达直径之前那段运动关于 $u = \frac{\pi}{2}$ 的镜像。所以小珠再用一段时间 t_1，便在 $u = \pi - \beta$ 处静止（那里再次有 $\theta = \beta$，只是小珠现在位于 O 的另一侧），然后沿原路走完全程回到 $u = \beta$。经过四个这样的阶段后，小珠回到起始位置，所以运动的周期是

$$T = 4t_1 = \frac{4}{k}\left(\frac{\pi}{2} + \arcsin\frac{\frac{\pi}{2} - \alpha}{\alpha - \beta}\right) = \frac{2}{k}\left(\pi + 2\arcsin\frac{\pi - 2\alpha}{2(\alpha - \beta)}\right).$$

（至于角 θ 本身，它并不记录小珠在 O 的哪一侧，所以每隔 $2t_1$ 就重复一次。）

情况 3： $\beta = 2\alpha - \frac{\pi}{2}$

这时 θ 在 $t = \frac{\pi}{k}$ 到达 $2\alpha - \beta = \frac{\pi}{2}$，而 SHM 恰好在这一刻令小珠静止。这时两颗小珠位于直径的两端，所以弹簧沿 AOB 放置，它作用于每颗小珠的力都沿半径方向：这个力没有沿圆环方向的分量，而沿圆环方向也没有其他力作用。小珠处于平衡（equilibrium）而且静止，所以会一直留在那里。因此

$$\beta = 2\alpha - \frac{\pi}{2}$$

就是不会出现振动的那个值。这与两个周期的结果吻合：当 β 从下方趋近这个值时，$\lambda \to 1$，情况 2 的周期趋向 $\frac{4\pi}{k}$，即情况 1 周期的两倍，因为小珠在直径附近逗留的时间愈来愈长。

验算

- 取 $k = 1$，$\alpha = \frac{3\pi}{8}$，$\beta = \frac{\pi}{8}$，由于 $2\alpha - \frac{\pi}{2} = \frac{\pi}{4}$，这属于情况 2。于是 $\lambda = \frac{\pi/8}{\pi/4} = \frac{1}{2}$，$\arcsin\lambda = \frac{\pi}{6}$，而 $T = 2(\pi + \frac{\pi}{3}) = \frac{8\pi}{3} \approx 8.378$。

- 我们用数值方法逐步求解两颗小珠的运动方程（能量写成两颗小珠位置的函数，并不假设任何对称性）。取 $k = 1$，$\alpha = 1.2$，$\beta = 0.5$ 时，小珠在 8.5162 之后回到起点并静止，而公式给出的是 8.51622；取 $\beta = 1.0$（情况 1）时，周期是 $6.2832 \approx 2\pi$。
- 量纲：k 的单位是 s^{-1}（能量等于 ma^2k^2 乘以角度的平方），所以 $\frac{2\pi}{k}$ 是时间。

常见错误

- 把动能写成 $\frac{1}{2}ma^2\dot{\theta}^2$：两颗小珠都在运动，所以动能是 $ma^2\dot{\theta}^2$。漏掉因子 2 会得出错误的角频率 $\sqrt{2}k$。
- 毫无根据地假设小珠对称地运动：这个结论来自对 O 的角动量守恒，而这需要弹簧作用于每颗小珠的力都沿同一直线 AB。
- 忘记 θ 是锐角，于是在所有情况下都答 $\frac{2\pi}{k}$。

STEP 之外

利用守恒量把两个质点的系统简化为一个坐标，正是拉格朗日力学（Lagrangian mechanics）的起点；在那里，对称性（在这题中，是把整个位形绕 O 旋转）会带来守恒定律，而诺特定理（Noether's theorem）把这个想法说得精确。情况 2 的运动是在“双井”（double-well）位势中的振动，它的周期取决于振幅，并且在能量等于中央势垒高度之处出现跳跃。想进一步了解守恒定律和振动，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）。

详细解答

STEP III 2007，第 10 题

力学 · 斜面上的抛体

STEP III 2007，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/07-s3-q10

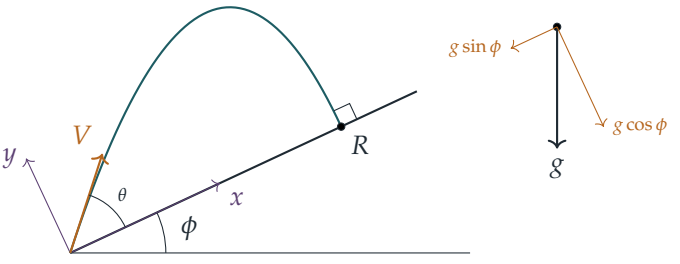
关键思路

取沿斜面及垂直于斜面的坐标轴，重力便化为两个恒定的加速度：沿斜面向下的 $g\sin\phi$ 和指向斜面的 $g\cos\phi$。最特别的条件是质点沿原路折返。光滑平面只能沿垂直于自身的方向施力，所以只有当质点到达时正正垂直撞向平面，平面才能把速度完全反转。两个结果都由这一点得出；余下的只是代数运算，以及一次利用算术平均—几何平均不等式（AM–GM inequality）求最小值。

解答

运动方程

我们把质点（particle）视为只受重力作用的点质量，不计空气阻力。重量 mg 铅垂向下。沿斜面及垂直于斜面分解（resolve）：



加速度在 x 方向为 $-g\sin\phi$，在 y 方向为 $-g\cos\phi$，两者皆为常数。以题目给定的初速度及起点 $(0,0)$ 积分两次，得

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

沿原路折返的条件

质点首次回到平面，是在 $y=0$ 且 $t>0$ 之时，即

$$t=T=\frac{2V\sin\theta}{g\cos\phi}.$$

> ? **想一想** “沿同一路径返回”说明了碰撞的甚么性质?

反弹之后，质点仍然在恒定的重力下运动。若它要沿同一条抛物线（parabola）返回原点，反弹后一刻的速度必须与反弹前一刻的速度刚好相反。（不妨想像把影片倒过来播放。）我们假设平面是光滑（smooth）的，所以它给予质点的冲量（impulse）垂直于平面：沿斜面的分量 $\dot{x}$ 不会因反弹而改变。这个分量要被反转，唯一的可能就是它本身为零。因此质点必须垂直撞击平面，即在 $t=T$ 时 $\dot{x}=0$。（我们还需要反弹把 $\dot{y}$ 反转而不损失速率，所以碰撞必须是完全弹性（perfectly elastic）的。题目把这一点视为理所当然。）

现在 $\dot{x}=V\cos\theta-gt\sin\phi$，所以由 $\dot{x}(T)=0$ 得

$$V\cos\theta=gT\sin\phi=\frac{2V\sin\theta\sin\phi}{\cos\phi}.$$

两边除以 $V\cos\theta$（由于 $0<\theta<\pi/2$，此数为正），

$$2\tan\phi\tan\theta=1,$$

即为所求。

射程

由 $\dot{x}(T)=0$ 可得较简洁的形式 $T=\dfrac{V\cos\theta}{g\sin\phi}$。于是

$$\begin{aligned}R=x(T)&=V\cos\theta\cdot\frac{V\cos\theta}{g\sin\phi}-\frac{1}{2}g\sin\phi\cdot\frac{V^2\cos^2\theta}{g^2\sin^2\phi}\\&=\frac{V^2\cos^2\theta}{g\sin\phi}-\frac{V^2\cos^2\theta}{2g\sin\phi}=\frac{V^2\cos^2\theta}{2g\sin\phi},\end{aligned}$$

这就是所求的结果。（换一个说法：x 在 $\dot{x}=0$ 时最大，所以落点就是这个运动沿斜面向上所能到达的最高点。）

关于 $2V^2/(gR)$ 的恒等式

由射程公式，

$$\frac{2V^2}{gR}=\frac{4\sin\phi}{\cos^2\theta}=4\sin\phi\,(1+\tan^2\theta).$$

利用 $\tan\theta = \dfrac{1}{2\tan\phi}$，即 $\tan^2\theta = \dfrac{\cos^2\phi}{4\sin^2\phi}$，

$$\frac{2V^2}{gR} = 4\sin\phi + \frac{\cos^2\phi}{\sin\phi} = 4\sin\phi + \frac{1-\sin^2\phi}{\sin\phi} = 3\sin\phi + \operatorname{cosec}\phi.$$

最大射程

对给定的 V，R 在 $3\sin\phi + \operatorname{cosec}\phi$ 最小时最大。设 $s = \sin\phi$；由于 $0 < \phi < \pi/2$，故 $0 < s < 1$。对正数 $3s$ 和 $1/s$ 应用算术平均—几何平均不等式，

$$3s + \frac{1}{s} \geq 2\sqrt{3s \cdot \frac{1}{s}} = 2\sqrt{3},$$

等号成立当且仅当 $3s = 1/s$，即 $s = 1/\sqrt{3}$。这个值确实落在 $(0,1)$ 之内，所以这个界可由一个真实存在的斜面达到。因此

$$\frac{2V^2}{gR} \geq 2\sqrt{3} \quad\Longrightarrow\quad R \leq \frac{V^2}{\sqrt{3}\,g},$$

等号在 $\sin\phi = 1/\sqrt{3}$ 时成立。所以最大可能射程为 $V^2/(\sqrt{3}\,g)$。

想一想 在最大值处，条件 $\phi + \theta < \pi/2$ 是否成立?

成立，而且事实上任何时候都成立。由 $\tan\theta\tan\phi = \frac{1}{2}$ 及两角皆为锐角，

$$\tan(\phi+\theta) = \frac{\tan\phi + \tan\theta}{1 - \tan\phi\tan\theta} = 2(\tan\phi + \tan\theta) > 0,$$

而由于 $0 < \phi + \theta < \pi$，这就迫使 $\phi + \theta < \pi/2$。在最大值处，$\sin\phi = 1/\sqrt{3}$ 给出 $\tan\phi = 1/\sqrt{2}$，所以 $\tan\theta = 1/\sqrt{2}$ 亦然：最佳情况是 $\theta = \phi \approx 35.3°$。

验算

- 单位：V^2/g 是长度，与射程应有的量纲相符。
- 极限情况：当 $\phi \to 0$ 时，$\operatorname{cosec}\phi \to \infty$ 而 $R \to 0$。在近乎水平的平面上，要垂直落在平面上，唯一的方法是几乎笔直向上发射，射程自然接近零。当 $\phi \to \pi/2$ 时，平面成为一幅铅垂的墙，$\theta \to 0$，且 $2V^2/(gR) \to 4$，所以 $R \to V^2/(2g)$：这正是质点贴着墙笔直向上抛出所达到的高度。两者都合理。
- 数值验算：取 $g = 9.8$、$V = 10$ 及 $\phi = 0.7$，逐步向前推算运动，落点为 $x \approx 5.8561$，与公式吻合。落地时的速度有 $\dot{x} \approx 0$，把它反转后质点便回到原点。对 ϕ 作网格搜索，在 $V = 1$ 时得最大值 $R \approx 0.05891$，等于 $1/(9.8\sqrt{3})$。

常见错误

- 把重力的两个分量弄反了。沿斜面的分量是 $g\sin\phi$，垂直于斜面的分量是 $g\cos\phi$。可用 $\phi = 0$ 检查。
- 说质点必须“水平地”落下，或在“最高点”落下。正确的条件是它垂直撞击平面，即在倾斜坐标轴中 $\dot{x} = 0$。
- 忘记检查 $\sin\phi = 1/\sqrt{3}$ 能否达到。算术平均—几何平均不等式只给出一个界，只有当等号确实可以成立时，它才是最大值。

STEP 之外

质点能沿原路折返，靠的是牛顿力学的时间可逆性（time-reversibility）。若力只取决于位置，把所有速度反转，系统便会沿着自己的历史倒退回去。选取配合几何形状的坐标轴，而不是惯用的水平和铅垂方向，是你在大学力学中会经常用到的习惯。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）对这两个想法都有详细阐述。

详细解答

STEP III 2007，第 11 题

力学 • 滚过边缘

STEP III 2007，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/07-s3-q11

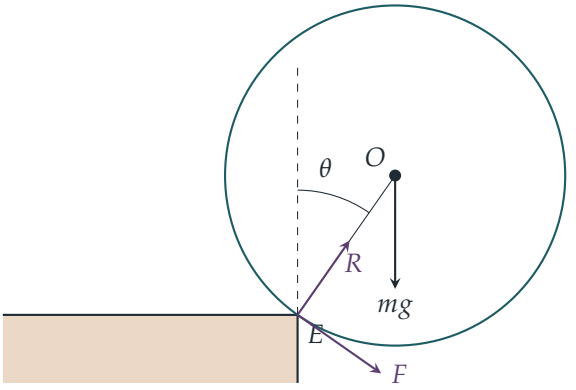
关键思路

轮缘一到达桌边，桌边便成为支点（pivot）：轮缘上与桌边接触的那一点是静止的，所以轮心沿着以桌边为圆心、半径为 a 的圆运动。我们把桌边的接触力分解为沿半径方向的分量 R（法向反作用力，normal reaction）和沿轮缘方向的分量 F（摩擦力，friction），再与质心（centre of mass）的运动作比较。整道题的关键在于质量位于何处。在 (i) 中，全部质量都在轮心，所以轮子对其中心的转动惯量（moment of inertia）为零。在 (ii) 中，质量分布在轮缘上，摩擦力正是由此而生。

解答

模型与记号。轮缘（rim）、轮辐（spokes）和圆环（hoop）都是轻而刚硬的；轮毂（hub）和两个质点（particles）都视为质点。设 E 为桌边上与轮缘接触的点，O 为轮心，θ 为 EO 与竖直向上方向所成的角，所以在轮子到达桌边的一刻 $\theta = 0$。只要轮缘与 E 保持接触而不滑动，轮缘上位于 E 的那个物质点便是静止的。因此这个刚体（rigid body）绕 E 转动，角速率（angular speed）为 $\dot{\theta}$，而 O 沿着以 E 为圆心、半径为 a 的圆运动，速率为 $a\dot{\theta}$。

在 E 的接触力有一个沿 EO 方向（背离 E）的分量 R，以及一个垂直于 EO 的分量 F，后者以 θ 增加的方向为正。由于 E 静止，接触力不作功，所以能量守恒（conservation of energy）。



图中 R 沿 EO 作用，而 F 则按其正方向绘出。

质心 G 与 E 的距离为 a，其加速度沿运动方向的分量为 $a\ddot{\theta}$，指向 E 的分量为 $a\dot{\theta}^2$。分解作用于整个物体的外力（重力 mg 和接触力）：

$$\text{沿运动方向：}\quad ma\ddot{\theta} = mg\sin\theta + F, \tag{1}$$

$$\text{指向}E\text{：}\quad ma\dot{\theta}^2 = mg\cos\theta - R. \tag{2}$$

在两部分中，质心都位于 O（即轮毂；亦即两个质点的中点），所以 (1) 和 (2) 在两部分都成立。

第 (i) 部分

轮子在桌面上以恒定速率 u 运动，所以在 $\theta = 0$ 时有 $\dot{\theta} = u/a$。唯一的动能是轮毂的动能 $\frac{1}{2}ma^2\dot{\theta}^2$，而轮毂下降了 $a(1-\cos\theta)$，所以

$$a^2\dot{\theta}^2 = u^2 + 2ga(1-\cos\theta).$$

对时间求导，并约去 $2a\dot{\theta}$（它不为零，因为 $\dot{\theta} \geq u/a > 0$）：

$$a\ddot{\theta} = g\sin\theta.$$

代入 (1) 得 $F = 0$：**摩擦力为零**，正如所求。

这个结果还有一个直接的理由。轮缘和轮辐都是轻的，所以作用于它们的各力对 O 的总力矩（moment）必须为零。这些力是在 E 的接触力，以及轮毂在 O 施加的力，而后者对 O 没有力矩。接触力对 O 的力矩是 Fa，所以 $F = 0$。

由 (2)，法向反作用力为

$$R = mg\cos\theta - \frac{m}{a}\left(u^2 + 2ga(1-\cos\theta)\right) = mg(3\cos\theta - 2) - \frac{mu^2}{a}.$$

在 $\theta = 0$ 时，这个值是 $m(g - u^2/a) > 0$，因为 $u^2 < ag$，所以轮子不会立即离开桌边。其后 R 随 θ 增加而减少，当 $R = 0$ 时轮缘便失去接触，即当

$$\cos\theta = \frac{2ag + u^2}{3ag},$$

这个值小于 1，因为 $u^2 < ag$。此时轮毂已下降了

$$a(1 - \cos\theta) = a \cdot \frac{3ag - 2ag - u^2}{3ag} = \frac{ag - u^2}{3g},$$

正如所求。由于自始至终 $F = 0$，不滑动的假设根本无须依靠摩擦力来维持，所以以上分析一直到这一点都成立。

第 (ii) 部分

> ? **想一想** 圆环绕 E 转动时，两个质点在哪里?

在 $\theta = 0$ 时，通过两个质点的直径是竖直的，所以下方的质点位于接触点 E。其后圆环作为刚体绕着位于 E 的物质点转动，而这个物质点正是该质点，所以该质点一直留在 E。另一个质点则一直留在直径的另一端，沿 EO 方向与 E 相距 $2a$。

因此对 E 的转动惯量是 $\frac{m}{2}(2a)^2 = 2ma^2$。圆环在桌面上以恒定速率 u 滚动（没有任何力作功），所以在 $\theta = 0$ 时 $\dot\theta = u/a$。运动中的质点比 E 高出 $2a\cos\theta$，由能量守恒得

$$\tfrac{1}{2}(2ma^2)\dot\theta^2 + \tfrac{m}{2}g(2a\cos\theta) = \tfrac{1}{2}(2ma^2)\frac{u^2}{a^2} + \tfrac{m}{2}g(2a), \qquad \text{所以} \qquad a^2\dot\theta^2 = u^2 + ga(1 - \cos\theta).$$

求导并约去 $2a\dot\theta \neq 0$，得 $a\ddot\theta = \frac{1}{2}g\sin\theta$。于是由 (1) 得

$$F = ma\ddot\theta - mg\sin\theta = -\tfrac{1}{2}mg\sin\theta.$$

所以对每一个 $\theta > 0$，**摩擦力都不为零**。（它的大小是 $\frac{1}{2}mg\sin\theta$，方向与 O 的运动方向相反。）与 (i) 不同，现在轮缘带有质量，所以圆环需要一个对 O 的力矩（torque），才能使它的转动跟得上 O 的运动，而只有摩擦力能够提供这个力矩。

由 (2)，

$$R = mg\cos\theta - \frac{m}{a}\left(u^2 + ga(1 - \cos\theta)\right) = mg(2\cos\theta - 1) - \frac{mu^2}{a}.$$

在 $\theta = 0$ 时，$R = m(g - u^2/a) > 0$ 而 $F = 0$，所以越过桌边的运动确实是在不滑动的情况下开始的。假如圆环一直保持接触而不滑动，R 便会在以下角度降至零：

$$\cos\theta_1 = \frac{ag + u^2}{2ag} < 1,$$

这里用到了 $u^2 < ag$。所以 $\theta_1 > 0$，而在该处 $|F| = \frac{1}{2}mg\sin\theta_1 > 0$。

想一想 为甚么这样就必定会在失去接触之前滑动?

设摩擦系数（coefficient of friction）为 μ，不滑动的条件是 $|F| \le \mu R$。当 $\theta \to \theta_1$（从下方趋近）时，$R \to 0^+$ 但 $|F| \to \frac{1}{2}mg\sin\theta_1 > 0$，所以 $|F|/R \to \infty$。因此无论 μ 取甚么（有限）值，总有一个角 $\theta < \theta_1$，在 R 仍然为正的时候，$|F|$ 就必须超过 μR。所以圆环**在与桌子失去接触之前便会滑动**。

验算

- 在 (i) 中，若 $u^2 \to ag$，下降的距离趋向 0：此时轮毂在 $\theta = 0$ 已处于即将飞离的临界状态，该处 $g = u^2/a$。若 $u \to 0$，下降的距离是 $a/3$，这正是质点从光滑球面顶点由静止开始滑下的那个熟悉结果。
- 取 $a = 1\,\mathrm{m}$、$u = 2\,\mathrm{m\,s^{-1}}$ 和 $g = 9.81\,\mathrm{m\,s^{-2}}$，以数值方法逐步求解方程，在 (i) 中得出下降距离为 0.1974 m，与 $(ag - u^2)/(3g)$ 吻合。在 (ii) 中我们得出 $\cos\theta_1 = 0.7039$，而且 $|F|/R$ 在临近 $R = 0$ 之前已超过 10^4。
- 在 (ii) 中直接由两个质点的加速度计算接触力（$m\ddot{\mathbf{r}}_G$ 减去重力），所得的 F 和 R 都与上面相同。

常见错误

- 在 (ii) 中使用对 O 的滚动动能 $\frac{1}{2}mv^2 + \frac{1}{2}I\omega^2$，却忘记了一旦开始绕支点转动，其中一个质点便固定在 E。对 E 来计算会简洁得多。
- 在桌边沿水平和竖直方向分解力。在边角处，“法向”是沿半径 EO 的方向，而不是竖直方向。
- 仅仅因为 $F \neq 0$ 便断定“它会滑动”。论证需要的是 $R \to 0$ 而 $|F|$ 始终不趋近零。

STEP 之外

两部分的对比在于质量的分布：同样的质心运动，会因转动惯量不同而需要不同的接触力，因为角动量（angular momentum）方程和线动量（linear momentum）方程必须同时满足。在大学第一门力学课程中，这就发展成刚体动力学（rigid-body dynamics）：物体的运动由 $\mathbf{F} = m\ddot{\mathbf{r}}_G$ 连同对 G 的力矩方程决定，而接触约束（滚动、滑动、失去接触）则要用不等式 $R \ge 0$ 和 $|F| \le \mu R$ 来检验。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教材）对这两条方程都有透彻的讲解。

详细解答

STEP III 2007，第 12 题

概率与统计 • 随机和

STEP III 2007，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/07-s3-q12

关键思路

Y 是一个项数随机的和，所以不能直接套用固定项数的变量之和的法则。窍门是以 N 为条件（condition）：一旦知道 $N = k$，Y 就是 k 个独立（independent）变量的普通和，其平均值（mean）和方差（variance）我们都知道。然后再对 N 的分布取平均，这正是题目第一部分要求 $\mathrm{E}(N)$ 和 $\mathrm{E}(N^2)$ 的原因。

解答

N 的分布

我们假设该数是均匀地随机选取的，因此对 $k = 1, 2, \ldots, 2n-1$ 有 $\mathrm{P}(N = k) = \frac{1}{2n-1}$。利用 $\sum_{k=1}^{m} k = \frac{1}{2}m(m+1)$ 和 $\sum_{k=1}^{m} k^2 = \frac{1}{6}m(m+1)(2m+1)$，并取 $m = 2n-1$：

$$\mathrm{E}(N) = \frac{1}{2n-1} \cdot \frac{(2n-1)(2n)}{2} = n,$$
$$\mathrm{E}(N^2) = \frac{1}{2n-1} \cdot \frac{(2n-1)(2n)(4n-1)}{6} = \frac{n(4n-1)}{3}.$$

（平均值 n 亦可由对称性直接看出：n 正是 $1, \ldots, 2n-1$ 的中间一项。）以下会用到

$$\mathrm{Var}(N) = \frac{n(4n-1)}{3} - n^2 = \frac{n^2 - n}{3} = \frac{n(n-1)}{3}.$$

以 N 为条件

固定 k，并假设 $N = k$。由于各 X_i 与 N 独立，以 $N = k$ 为条件并不会改变它们的分布，所以在已知 $N = k$ 的情况下，$Y = X_1 + \cdots + X_k$，即 k 个独立变量之和，每个变量的平

均值为 μ，方差为 σ^2。因此

$$\mathrm{E}(Y \mid N=k) = k\mu, \qquad \mathrm{Var}(Y \mid N=k) = k\sigma^2,$$

其中第二式用到各 X_i 互相独立（方差可以相加）。于是

$$\mathrm{E}(Y^2 \mid N=k) = \mathrm{Var}(Y \mid N=k) + \big(\mathrm{E}(Y \mid N=k)\big)^2 = k\sigma^2 + k^2\mu^2.$$

想一想 怎样由条件期望值（conditional expectation）回到普通的期望值?

根据全期望定律（law of total expectation，又称分割定理 partition theorem）：$\mathrm{E}(g(Y,N)) = \sum_k \mathrm{P}(N=k)\,\mathrm{E}(g(Y,N) \mid N=k)$。在每一项之内，$N$ 只不过是数字 k。

Y 的平均值

$$\mathrm{E}(Y) = \sum_k \mathrm{P}(N=k)\,k\mu = \mu\,\mathrm{E}(N) = n\mu,$$

正是所求。

协方差

已知 $N=k$ 时，$YN = kY$，所以 $\mathrm{E}(YN \mid N=k) = k\cdot k\mu = k^2\mu$。因此 $\mathrm{E}(YN) = \mu\,\mathrm{E}(N^2)$，而协方差（covariance）为

$$\begin{aligned}\mathrm{Cov}(Y,N) &= \mathrm{E}(YN) - \mathrm{E}(Y)\,\mathrm{E}(N) = \mu\,\mathrm{E}(N^2) - n\mu\cdot n \\ &= \mu\,\mathrm{Var}(N) = \tfrac{1}{3}n(n-1)\mu,\end{aligned}$$

这就是所要求的结果。

Y 的方差

把 $\mathrm{E}(Y^2 \mid N=k) = k\sigma^2 + k^2\mu^2$ 对 N 取平均：

$$\mathrm{E}(Y^2) = \sigma^2\,\mathrm{E}(N) + \mu^2\,\mathrm{E}(N^2) = n\sigma^2 + \frac{n(4n-1)}{3}\mu^2.$$

因此

$$\mathrm{Var}(Y) = n\sigma^2 + \frac{n(4n-1)}{3}\mu^2 - n^2\mu^2 = n\sigma^2 + \frac{n(n-1)}{3}\mu^2.$$

这两项各有自然的含义：$n\sigma^2$ 是试验次数固定时 Y 的平均离散程度，而 $\frac{1}{3}n(n-1)\mu^2 = \mu^2\,\mathrm{Var}(N)$ 则是因为不知道试验次数而额外产生的离散程度。

验算

- 若 $n = 1$，则必有 $N = 1$，各公式化为 $\mathrm{E}(Y) = \mu$、$\mathrm{Cov}(Y, N) = 0$ 和 $\mathrm{Var}(Y) = \sigma^2$，与单独一个 X_1 的情况相符。
- 若 $\mu = 0$，则额外的一项消失：项数不再影响平均值，所以它既不会增加离散程度，也不会与 Y 相关。
- 取 $n = 2$，并设每个 X_i 各以概率 $\frac{1}{2}$ 取 0 或 1。那么 $N \in \{1, 2, 3\}$，而 $Y \mid N \sim \mathrm{B}(N, \frac{1}{2})$；公式给出 $\mathrm{Var}(Y) = 2 \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{1}{4} = \frac{2}{3}$。我们已逐一列举 (N, Y) 的联合分布（joint distribution），就 $n = 1, \ldots, 6$ 以及 X_i 的几种不同分布，核实了这个数值和所有结果。

常见错误

- 只写 $\mathrm{Var}(Y) = \mathrm{E}(N)\sigma^2$，忘记了 N 本身也会变动。第二项才是整题的重点。
- 把 Y 和 N 当作独立，得出 $\mathrm{Cov}(Y, N) = 0$。当 $\mu > 0$ 时，试验次数越多，和就越大，所以两者是正相关的。

STEP 之外

Y 是一个随机和（random sum，即复合分布 compound distribution），而方差的结果是全方差定律（law of total variance）的一个特例：$\mathrm{Var}(Y) = \mathrm{E}(\mathrm{Var}(Y \mid N)) + \mathrm{Var}(\mathrm{E}(Y \mid N))$，在这里即 $\sigma^2\mathrm{E}(N) + \mu^2\mathrm{Var}(N)$。随机和可用作保险索偿和分支过程（branching processes）的模型，而用概率生成函数（probability generating functions）处理则十分简洁：$G_Y(s) = G_N(G_X(s))$。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）正是以这种方法处理随机和与分支过程。

详细解答

STEP III 2007，第 13 题

概率与统计 · 随机漫步的期望值

STEP III 2007，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：

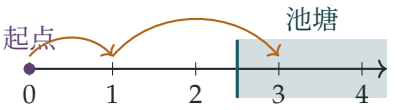
step.maths.org/questions/07-s3-q13

关键思路

那半米的作用，是消除“刚好落在池边”的任何疑问：青蛙由距离池边 $(n-\frac{1}{2})\,\mathrm{m}$ 处出发，牠身处池中，当且仅当牠已跳的总距离至少为 n。因此整个问题的核心，就是由 1 和 2 组成的累计总和首次达到 n 的时刻。当 n 较小时，我们可以直接列出所有可能情况；至于一般公式，u_n 的三个值可给出三条方程，用以求出三个常数。

解答

设 S_j 为首 j 次跳跃所跳的总距离（以米为单位），并设 T 为到达池塘所需的跳跃次数。青蛙由距离池边 $(n-\frac{1}{2})\,\mathrm{m}$ 处出发，跳了 j 次后身处池中，当且仅当 $S_j > n-\frac{1}{2}$，亦即 $S_j \geq n$，因为 S_j 是整数。所以 $p_n(j) = \mathrm{P}(T=j)$，其中 T 是首个满足 $S_j \geq n$ 的 j。



图中所示为 $n=3$ 的情况（池边距离起点 2.5 m），以及“短跳、长跳”这个跳跃序列，青蛙跳两次便到达池塘。

第 (i) 部分

当 $n=2$ 时，池边距离起点 1.5 m。若第一跳是长跳（概率（probability）为 q），青蛙便直接跳进池中，所以 $p_2(1)=q$。否则第一跳是短跳（概率为 p），青蛙距离池边 0.5 m，而

第二跳无论长短都会把牠带进池中。各次跳跃互相独立（independent），所以

$$p_2(2) = p \times 1 = p,$$

这正是所要证明的结果。（此外，当 $j \geq 3$ 时 $p_2(j) = 0$，而 $p + q = 1$ 则确认了这个分布（distribution）是完整的。）

第 (ii) 部分

由距离池边 $\frac{1}{2}$ m 处出发，任何一跳都会落入池中，所以 $u_1 = 1$。

利用第 (i) 部分，

$$u_2 = 1 \cdot p_2(1) + 2 \cdot p_2(2) = q + 2p = 2 - q.$$

当 $n = 3$ 时，我们需要 $S_j \geq 3$。

- 只跳一次永远不够，因为 $S_1 \leq 2$：$p_3(1) = 0$。
- 跳两次仍不够，只会在两次都是短跳时发生（$S_2 = 2$）；其余每一种组合都有 $S_2 \geq 3$。所以 $p_3(2) = 1 - p^2$。
- 两次短跳之后，青蛙距离池边 0.5 m，第三跳必定落入池中：$p_3(3) = p^2$。

因此

$$\begin{aligned} u_3 &= 2(1 - p^2) + 3p^2 = 2 + p^2 \\ &= 2 + (1 - q)^2 = 3 - 2q + q^2, \end{aligned}$$

这正是所要证明的结果。

? **想一想** 有没有更快的方法看出这些值？这个方法能否推广至所有 n？

有：以第一跳的结果作条件（conditioning）来考虑。第一跳之后，青蛙的处境与原来相同，只是 n 减少了 1 或 2，所以当 $n \geq 2$ 时

$$u_n = 1 + p\,u_{n-1} + q\,u_{n-2}, \qquad u_0 = 0,$$

其中 $u_0 = 0$ 对应由 1.5 m 处出发而第一跳是长跳的情况。由此再次得出 $u_2 = 1 + p$ 和 $u_3 = 1 + p(1 + p) + q = 2 + p^2$。这也解释了第 (iii) 部分的形式为何正确：辅助方程（auxiliary equation）$\lambda^2 - p\lambda - q = 0$ 可因式分解为 $(\lambda - 1)(\lambda + q) = 0$，由此得出 $(-q)^{n-1}$ 这一项。而重根 1（源自那个“+1”）则解释了 $B + Cn$ 的出现。

第 (iii) 部分

把 $n = 1, 2, 3$ 代入 $u_n = A(-q)^{n-1} + B + Cn$，并利用第 (ii) 部分：

$$A + B + C = 1, \tag{1}$$

$$-qA + B + 2C = 2 - q, \tag{2}$$

$$q^2A + B + 3C = 3 - 2q + q^2. \tag{3}$$

两式相减以消去 B：

$$(2) - (1): \quad -(1+q)A + C = 1 - q, \tag{4}$$

$$(3) - (2): \quad q(1+q)A + C = 1 - q + q^2. \tag{5}$$

现在 $q \times (4) + (5)$ 可消去 A：

$$(1+q)C = q - q^2 + 1 - q + q^2 = 1, \qquad \text{所以} \qquad C = \frac{1}{1+q},$$

这正是所要证明的结果。然后由 (4) 得

$$(1+q)A = \frac{1}{1+q} - (1-q) = \frac{1-(1-q^2)}{1+q} = \frac{q^2}{1+q}, \qquad \text{所以} \qquad A = \frac{q^2}{(1+q)^2},$$

再由 (1) 得

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

所以

$$u_n = \frac{q^2(-q)^{n-1} + q + (1+q)n}{(1+q)^2}.$$

n 很大时。 由于 $0 \le q \le 1$，$A(-q)^{n-1}$ 和 B 这两项都是有界（bounded）的（各自的绝对值最多为 $\frac{1}{4}$），而 Cn 则无止境地增大。因此当 n 很大时

$$u_n \approx Cn = \frac{n}{1+q} = \frac{n}{p+2q},$$

这里用到 $1 = p + q$，故 $1 + q = p + 2q$。

为何这在意料之中。 每一跳的平均长度是 $1 \cdot p + 2 \cdot q = p + 2q$ 米。青蛙须跳过大约 n 米，而根据大数定律（law of large numbers），当 N 很大时，N 次跳跃的总长度接近 $N(p+2q)$。所以所需的跳跃次数接近 $n/(p+2q)$。这个近似相当准确，因为起点那半米的偏移，以及终点处多跳出的距离（青蛙落点最多超出池边 1.5 m）都是有界的。它们只会令次数改变一个有界的量，与 n 相比微不足道。这个有界的修正，正正就是 $A(-q)^{n-1} + B$ 这部分。

验算

- 在公式中取 $n = 0$，得 $\left(q^2 \cdot (-1/q) + q\right)/(1+q)^2 = 0 = u_0$，与应有的结果相符，而且公式满足上面的递推关系（recurrence）。
- $q = 0$（所有跳跃都是短跳）：$u_n = n$，这是正确的，因为青蛙恰好需要 n 次单位长度的跳跃。
- 以分数精确枚举所有跳跃序列，结果在 $n = 1, \ldots, 12$ 及若干个 p 值下均与公式吻合。以 $q = 0.7$、$n = 200$ 进行模拟，得出平均跳跃次数为 117.90，而公式给出 117.89，$n/(p+2q) \approx 117.65$。

常见错误

- 把“落入池中”当作恰好 $S_j = n$。由于有那半米，青蛙可以跳过头，凡是满足 $S_j \geq n$ 的序列都要计算在内。
- 写成 $p_3(2) = 2pq + q^2$ 时，忘记了长跳在先的情况。事实上 $pq + qp + q^2 = 1 - p^2$ 已涵盖两种次序。

STEP 之外

u_n 的递推关系是一条更新方程（renewal equation）。结果 $u_n \sim n/\mu$（其中 μ 为每步的平均长度）就是基本更新定理（elementary renewal theorem）。由沃尔德恒等式（Wald's identity）$\mathrm{E}(S_T) = \mu\,\mathrm{E}(T)$ 可得出一个精确的版本：这里 S_T 不是 n 就是 $n+1$，所以 $(1+q)u_n = n + \mathrm{P}(\text{跳过头})$。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）以同样的思路讲述随机漫步（random walk）和母函数（generating function）。

详细解答

STEP III 2007，第 14 题

概率与统计 · 几何概率

STEP III 2007，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/07-s3-q14

关键思路

包围所有洞孔的最小同心圆（或正方形），由最远的那个洞孔决定，所以我们要找的是最大值的分布，而 $P(\max \le r)$ 不过是单次投掷的概率（probability）的 n 次方。在圆盘上，一个洞孔与 O 的距离的平方 R^2 原来在 $[0,1]$ 上服从均匀分布（uniform distribution），于是一切都变得简洁。在正方形上，合适的“距离”是 $\max(|x|,|y|)$，它的表现与圆盘上的半径完全一样。这个巧合也令第 (iii) 部分几乎不费吹灰之力就解决了。

解答

第 (i) 部分

设 n 个洞孔与 O 的距离为 $R_1,\dots,R_n$。由于各次投掷互相独立，这些距离亦互相独立（independent）。对单次投掷及 $0 \le r \le 1$，洞孔与 O 的距离不超过 r，当且仅当它落在半径为 r 的圆盘内，所以

$$P(R_i \le r) = \frac{\pi r^2}{\pi \cdot 1^2} = r^2.$$

设 $U_i = R_i^2$。那么对 $0 \le u \le 1$，$P(U_i \le u) = P(R_i \le \sqrt{u}) = u$，所以每个 U_i 都在 $[0,1]$ 上服从均匀分布，而且各 U_i 互相独立。

以 O 为圆心并包围所有洞孔的最小圆，半径为 $M = \max_i R_i$，面积为 $\pi M^2 = \pi \max_i U_i$。记 $V = \max_i U_i$。由独立性，

$$P(V \le v) = P(U_1 \le v) \cdots P(U_n \le v) = v^n \qquad (0 \le v \le 1),$$

所以 V 的概率密度函数（density）为 nv^{n-1}，而

$$E(V) = \int_0^1 v \cdot nv^{n-1}\,\mathrm{d}v = \frac{n}{n+1}.$$

因此，包围全部 n 个洞孔的最小圆的期望（expected）面积为

$$\boxed{\frac{n\pi}{n+1}}.$$

至于包围最接近 O 的 $n-1$ 个洞孔的圆，其半径是第二大的距离，所以面积为 πW，其中 W 是 $U_1,\ldots,U_n$ 之中第二大的值。

> **想一想** 怎样求出第二大的值的分布?

$W \le w$ 当且仅当各 U_i 之中最多只有一个超过 w。要么没有任何一个超过（概率为 w^n），要么恰好有一个超过：选出是哪一个有 n 种方法，得出概率 $n(1-w)w^{n-1}$。所以

$$P(W \le w) = w^n + n(1-w)w^{n-1} = nw^{n-1} - (n-1)w^n.$$

我们不必求导，而可以利用 $E(W) = \int_0^1 P(W > w)\,\mathrm{d}w$，这条公式对任何取值于 $[0,1]$ 的随机变量（random variable）都成立：

$$E(W) = \int_0^1 \left(1 - nw^{n-1} + (n-1)w^n\right)\mathrm{d}w = 1 - 1 + \frac{n-1}{n+1} = \frac{n-1}{n+1}.$$

所以，包围最近的 $n-1$ 个洞孔的圆的期望面积为

$$\boxed{\frac{(n-1)\pi}{n+1}}.$$

（这符合一个漂亮的规律：n 个独立的均匀分布点把 $[0,1]$ 分成 $n+1$ 段间隙，每段的期望长度同为 $\frac{1}{n+1}$，所以第 k 小的点的平均值为 $\frac{k}{n+1}$。）

第 (ii) 部分

取通过 Q 且与镖靶各边平行的坐标轴，于是镖靶就是 $-1 \le x, y \le 1$。我们取包围洞孔的正方形，使其各边与镖靶的边平行。以 Q 为中心、半边长为 s 的正方形是 $\{|x| \le s,\ |y| \le s\}$，所以它包含洞孔 (x,y) 当且仅当 $\max(|x|,|y|) \le s$。

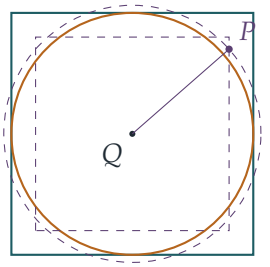
对第 i 个洞孔，设 $S_i = \max(|x_i|,|y_i|)$。对 $0 \le s \le 1$，

$$P(S_i \le s) = \frac{(2s)^2}{2^2} = s^2,$$

与第 (i) 部分的 $P(R_i \le r) = r^2$ 相同。包围所有洞孔的最小正方形，半边长为 $\max_i S_i$，面积为 $(2\max_i S_i)^2 = 4\max_i S_i^2$。与第 (i) 部分完全一样，各 S_i^2 互相独立，并在 $[0,1]$ 上服从均匀分布，所以它们的最大值的期望值为 $\frac{n}{n+1}$。包围所有洞孔的最小正方形的期望面积为

$$\boxed{\frac{4n}{n+1}}.$$

第 (iii) 部分



现在镖靶是正方形，而我们要找以 Q 为圆心并包围所有洞孔的最小圆。洞孔 $P=(x,y)$ 与 Q 的距离为 $\sqrt{x^2+y^2}$，而

$$\sqrt{x^2+y^2} \;\geq\; \max(|x|,|y|) = S,$$

等号仅在 P 位于坐标轴上时成立。（在图中，通过 P 的虚线圆永远不会小于通过 P 的虚线正方形的内切圆。）所以，若 D 是洞孔与 Q 的最大距离，则 $D \geq \max_i S_i$，而包围所有洞孔的圆的面积满足

$$\pi D^2 \;\geq\; \pi \max_i S_i^2.$$

由第 (ii) 部分，$\max_i S_i^2$ 的分布与圆形镖靶上 $\max_i R_i^2$ 的分布完全相同，所以

$$E(\pi D^2) \;\geq\; \pi\, E\big(\max_i S_i^2\big) = \frac{n\pi}{n+1},$$

这正是第 (i) 部分圆形镖靶的答案。这个不等式（inequality）是严格的：没有任何洞孔位于坐标轴上的概率为正，而这时 $D > \max_i S_i$。所以正方形镖靶上的圆，其期望面积**较大**。

直观地看：正方形镖靶包含了圆形镖靶（圆盘就是它的内切圆），而飞镖会散落到四个角落，所以最远的洞孔往往离中心更远。

验算

- 当 $n=1$ 时，公式给出 $\frac{\pi}{2}$ 和 2：均匀分布的 R^2 的平均值为 $\frac{1}{2}$，所以结果正确。当 $n\to\infty$ 时，面积趋向 π 和 4，即整个镖靶，这正合预期。
- 以蒙地卡罗模拟（Monte Carlo simulation）对 $n=2,3,5$ 各进行 2×10^5 次试验，结果与 $\frac{n\pi}{n+1}$、$\frac{(n-1)\pi}{n+1}$ 和 $\frac{4n}{n+1}$ 吻合至约三位有效数字；例如 $n=3$ 时得出 2.3561，而 $\frac{3\pi}{4}\approx 2.3562$。
- 对第 (iii) 部分，以数值积分（quadrature）计算精确分布，在 $n=3$ 时得出正方形镖靶上的期望面积约为 3.257，而圆盘上则为 2.356：一如所言，前者较大。

常见错误

- 先求 $E(M)$ 然后平方：期望面积是 $\pi E(M^2)$，而不是 $\pi\big(E(M)\big)^2$。
- 以为与中心的距离在 $[0,1]$ 上服从均匀分布。命中的概率与面积成正比，因此服从均匀分布的是 R^2 而不是 R；外围的环面积较大。
- 处理第二大的值时，忘记了因子 n，即哪一个洞孔落在圆外的选法数目。

STEP 之外

样本的最大值和第二大的值都是*顺序统计量*（order statistics）。对 $[0,1]$ 上 n 个独立的均匀分布变量，第 k 小的值服从贝塔分布（Beta distribution），平均值为 $\frac{k}{n+1}$。第 (iii) 部分是*随机优势*（stochastic dominance）（耦合，coupling）的一个例子：若一个随机变量总是不小于另一个，则它的期望值亦然，而这个比较无须任何细致的计算。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门）以及 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论）都有讨论顺序统计量和随机变量的变换。

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

2007 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。

