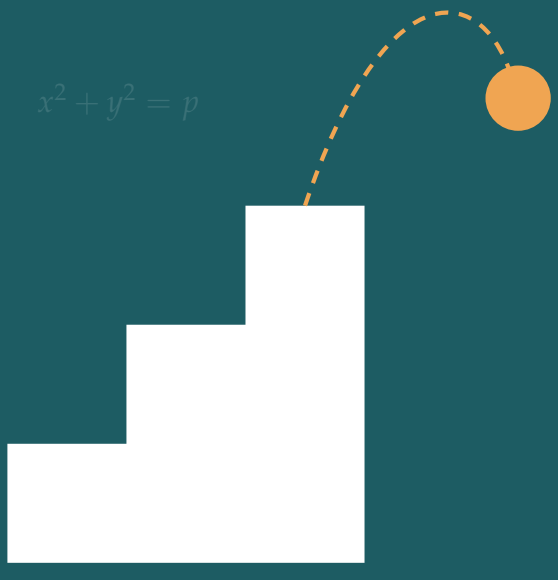
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

1995 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

1995 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

1995 年详细解答

1995 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 多项式不等式 纯数学 . . . 3

第 2 题 代换积分法 纯数学 . . . 7

第 3 题 裂项相消级数 纯数学 . . . 10

第 4 题 棣美弗定理 纯数学 . . . 13

第 5 题 二项式级数与积分 纯数学 . . . 16

第 6 题 伯努利微分方程 纯数学 . . . 20

第 7 题 三角形的向量几何 纯数学 . . . 23

第 8 题 线性微分方程 纯数学 . . . 27

第 9 题 抛体轨道的包络线 力学 . . . 30

第 10 题 弹性绳与反弹 力学 . . . 33

第 11 题 圆柱之间的摩擦力 力学 . . . 37

第 12 题 排列与计数 概率与统计 . . . 41

第 13 题 泊松稀疏化 概率与统计 . . . 44

第 14 题 比例的估计 概率与统计 . . . 47

STEP II

第 1 题 等比级数的求导 纯数学 . . . 51

第 2 题 以递推法计算着色数目 纯数学 . . . 55

第 3 题 逻辑推理与计数 纯数学 . . . 58

第 4 题 沃利斯归约公式 纯数学 . . . 62

第 5 题 最优化与凹性 纯数学 . . . 66

第 6 题 单位根 纯数学 . . . 69

第 7 题 希罗公式 纯数学 . . . 73

第 8 题 逻辑斯谛微分方程 纯数学 . . . 77

第 9 题 摩擦力与平衡 力学 81
第 10 题 碰撞与恢复系数 力学 84
第 11 题 线性空气阻力 力学 87
第 12 题 连续决斗 概率与统计 90
第 13 题 指数分布的最优化 概率与统计 94
第 14 题 假设检验 概率与统计 98

STEP III

第 1 题 联立线性方程 纯数学 102
第 2 题 归约公式 纯数学 106
第 3 题 阻尼谐运动 纯数学 109
第 4 题 三角级数求和 纯数学 112
第 5 题 麦克劳林级数 纯数学 115
第 6 题 轨迹与反演 纯数学 119
第 7 题 群与同构 纯数学 123
第 8 题 平面与球面 纯数学 127
第 9 题 复摆 力学 131
第 10 题 斜面上的抛体 力学 135
第 11 题 相对速度 力学 138
第 12 题 径向对称密度函数 概率与统计 142
第 13 题 有噪声的二元信道 概率与统计 146
第 14 题 矩生成函数 概率与统计 150

详细解答

STEP I 1995，第 1 题

纯数学 · 多项式不等式

> **STEP I 1995，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/95-s1-q1

关键思路

第 (i) 部分的三次式（cubic）可用分组法因式分解（factorise）；一旦写成三个线性因式（linear factor）的乘积，它的正负号便可从数线上直接读出。第 (ii) 部分其实是同一个三次式的化身：每一项的总次数（degree）都是 3，所以除以 y^3 之后，它就变成以 x/y 为变量的第 (i) 部分的三次式。同样的因式分解于是给出三条通过原点的直线，而第 (iii) 部分就是一个正负号表，只不过这次是围绕原点而画，而不是沿着一条直线而画。

解答

第 (i) 部分

把各项两两分组，

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

根（root）是 -1、1 和 4。每个因式都是线性的，只会在自己的根处变号，所以乘积在三个根处都会变号。当 x 是很大的正数时，三个因式全是正数，因此从右边读起，正负号的分布是

x	$x < -1$	$-1 < x < 1$	$1 < x < 4$	$x > 4$
正负号	$-$	$+$	$-$	$+$

把三次式等于零的各个根也包括在内，不等式（inequality）成立的充要条件是

$$-1 \le x \le 1 \quad \text{或} \quad x \ge 4.$$

第 (ii) 部分

设 $f(x,y) = x^3 - 4x^2y - xy^2 + 4y^3$。它就是第 (i) 部分的三次式，只是每一项都用 y 的幂补足至 3 次，而同样的分组法仍然适用：

$$\begin{aligned} f(x,y) &= x^2(x-4y) - y^2(x-4y) \\ &= (x-4y)(x-y)(x+y). \end{aligned}$$

乘积等于零，当且仅当其中一个因式等于零，所以 $f(x,y) = 0$ 在以下三条直线上成立：

$$x = 4y \text{ (即} y = \tfrac{1}{4}x), \qquad y = x, \qquad y = -x,$$

三条直线全都通过原点。

> **想一想** 为甚么 (ii) 的答案会与 (i) 有关?

当 $y \neq 0$ 时，设 $t = x/y$。那么 $f(x,y) = y^3\,(t^3 - 4t^2 - t + 4)$，也就是 y^3 乘以第 (i) 部分的三次式。该三次式在 $t = -1, 1, 4$ 时等于零，即是在 $x = -y$、$x = y$ 和 $x = 4y$ 上等于零。（当 $y = 0$ 时，$f = x^3$ 只在原点等于零，而原点本来就在三条直线上。）

第 (iii) 部分

三条直线把平面分成六个在原点相接的扇形区域（sector）。由正 x 轴起逆时针量度，六条半直线所指的方向是

$$\arctan\tfrac{1}{4} \approx 14^\circ, \quad 45^\circ, \quad 135^\circ, \quad 194^\circ, \quad 225^\circ, \quad 315^\circ.$$

在一个扇形区域之内，f 没有任何因式等于零，所以 f 在该区域内的正负号保持不变。每越过其中一条直线，恰好有一个因式变号（三条直线互不相同，而且在原点以外每次只会越过一条），所以 f 的正负号由一个扇形区域到下一个扇形区域交替变化。

要确定正负号，可取 $(1,0)$：$f(1,0) = 1 > 0$，而 $(1,0)$ 位于 315° 与 14° 之间的扇形区域内。绕一圈下去，正负号因此是

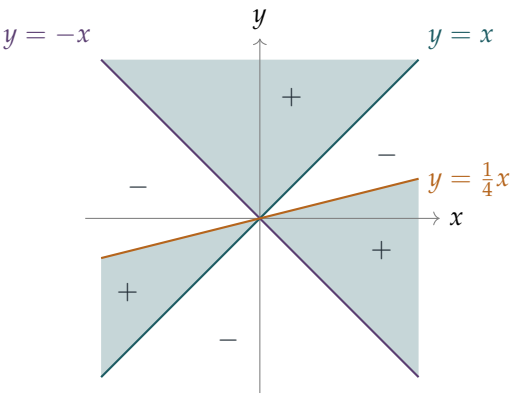
扇形区域（度）	$-45, 14$	$14, 45$	$45, 135$	$135, 194$	$194, 225$	$225, 315$
正负号：f	$+$	$-$	$+$	$-$	$+$	$-$

抽样检查一下：$f(0,1) = 4 > 0$ 属于第三栏，$f(-3,-1) = 8 > 0$ 属于第五栏。

所以 $f(x,y) \geq 0$ 在下图三个有阴影的扇形区域上成立，另外还包括全部三条直线（在直线上 $f = 0$）：

- 包含正 x 轴的扇形区域，介乎 $y = -x$（下方）与 $y = \frac{1}{4}x$（上方）之间，其中 $x > 0$;
- 包含正 y 轴的扇形区域，介乎 $y = x$ 与 $y = -x$ 之间，其中 $y > 0$;

- 第三象限内介乎 $y = x$ 与 $y = \frac{1}{4}x$ 之间的狭窄扇形区域。



想一想 为甚么把图形绕原点旋转半圈之后，图形不变，只是阴影部分与空白部分对调？

f 的每一项都是奇数次 3 次，所以 $f(-x,-y) = -f(x,y)$。把一点绕原点旋转 180°，会令 f 的正负号反转，所以每个有阴影的扇形区域的对面都是一个没有阴影的扇形区域。这是检查任何草图的快捷方法：有阴影和没有阴影的扇形区域必须成对地相对，而图中确实如此。

验算

- 在直线 $y = 1$ 上，$f(x,1)$ 正好是第 (i) 部分的三次式，所以 $y = 1$ 上有阴影的部分应该是 $-1 \le x \le 1$ 和 $x \ge 4$。在草图中，$y = 1$ 在 $y = -x$ 与 $y = x$ 之间（由 $x = -1$ 到 $x = 1$）有阴影，并且在它与 $y = \frac{1}{4}x$ 相交之处，即 $x = 4$，再次进入有阴影的扇形区域。
- 在 x 轴上 $f = x^3$，它非负的充要条件是 $x \ge 0$：正 x 轴有阴影，负的一半则没有。
- 数值上，在 2×10^5 个随机点计算 f 的值，其正负号每次都与六个扇形区域的正负号表吻合；例如 $f(1,0) = 1$、$f(0,-1) = -4$、$f(-3,-1) = 8$。

常见错误

- 除以 y^3 之后，在 $y < 0$ 时仍然保留原来的不等号。由于该处 $y^3 < 0$，条件变成 $t^3 - 4t^2 - t + 4 \le 0$，这正是下半平面看起来与上半平面不同的原因。
- 只在 $y > 0$ 时用第 (i) 部分的答案来画阴影，然后完全忘记了 $y < 0$ 的情况。
- 漏掉直线本身：不等式并非严格不等式，所以直线也属于该区域。

STEP 之外

各项次数全部相同的多项式称为齐次（homogeneous）多项式。它的零点集是若干条通过原点的直线的并集，而它的正负号只取决于方向。齐次多项式是射影几何（projective geometry）的起点，在射影几何中，一个点就是一整条通过原点的直线；当你在线性代数（linear algebra）课程中为圆锥曲线（conics）和二次型（quadratic forms）分类时，它们又会再次出现。Strang 的 *Introduction to Linear Algebra* 涵盖二次型，而 Siklos 的 *Advanced Problems in Mathematics* 则有更多这种风格的 STEP 题目。

详细解答

STEP I 1995，第 2 题

纯数学 · 代换积分法

STEP I 1995，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/95-s1-q2

关键思路

在第 (i) 部分，两个积分（integral）单独来看都不易处理，但两者却是天作之合：相加后得到的分式可约简为 1，相减后得到的分子恰好是分母的导数（derivative）。两个容易的积分，给出关于 S 和 T 的两条联立方程（simultaneous equations）。在第 (ii) 部分，选用代换（substitution）$x = \sin^2 t$，是因为它把 $\frac{1}{x} - 1$ 化为完全平方 $\cot^2 t$，令平方根消失；其后全靠倍角公式（double-angle formulae）便可完成。

解答

第 (i) 部分

把两个被积函数（integrand）相加，

$$S + T = \int \frac{\cos x + \sin x}{\cos x + \sin x}\,\mathrm{d}x = \int 1\,\mathrm{d}x = x + c_1.$$

把两者相减，

$$S - T = \int \frac{\cos x - \sin x}{\cos x + \sin x}\,\mathrm{d}x.$$

分子正是分母的导数，因为 $\frac{\mathrm{d}}{\mathrm{d}x}(\cos x + \sin x) = -\sin x + \cos x$。所以这个积分属于 $\int \frac{f'(x)}{f(x)}\,\mathrm{d}x$ 的形式，而

$$S - T = \ln|\cos x + \sin x| + c_2.$$

把这两条方程相加、相减，再除以二，

$$S = \tfrac{1}{2}(x + \ln|\cos x + \sin x|) + C,$$
$$T = \tfrac{1}{2}(x - \ln|\cos x + \sin x|) + C',$$

其中 C 和 C' 为任意常数（arbitrary constants）。在任何满足 $\cos x + \sin x \neq 0$ 的区间上，被积函数均有定义，以上结果亦在该区间上成立。

第 (ii) 部分

设 I 为所求的积分。取 $x = \sin^2 t$，则 $\mathrm{d}x = 2\sin t \cos t \,\mathrm{d}t$。至于积分上下限（limits），$x = \frac{1}{4}$ 给出 $\sin t = \frac{1}{2}$，所以 $t = \frac{\pi}{6}$；而 $x = \frac{1}{2}$ 给出 $\sin t = \frac{1}{\sqrt{2}}$，所以 $t = \frac{\pi}{4}$。（我们取 t 在 $[\frac{\pi}{6}, \frac{\pi}{4}]$ 内；当 t 走遍这个区间时，x 由 $\frac{1}{4}$ 递增至 $\frac{1}{2}$。）

接着处理平方根：

$$\frac{1}{x} - 1 = \frac{1 - \sin^2 t}{\sin^2 t} = \frac{\cos^2 t}{\sin^2 t}, \qquad \text{所以} \qquad \sqrt{\frac{1}{x} - 1} = \frac{\cos t}{\sin t}.$$

想一想 为甚么平方根是 $\cos t / \sin t$，而不是 $|\cos t / \sin t|$?

一般而言，$\sqrt{\cot^2 t} = |\cot t|$。但当 $\frac{\pi}{6} \leq t \leq \frac{\pi}{4}$ 时，$\sin t$ 和 $\cos t$ 皆为正数，所以 $\cot t > 0$，绝对值（modulus）符号可以略去。这正是我们选取 t 在第一象限（first quadrant）的原因。

代入后，$\sin t$ 互相约去：

$$I = \int_{\pi/6}^{\pi/4} (1 - 4\sin^2 t)\, \frac{\cos t}{\sin t}\, 2\sin t \cos t \,\mathrm{d}t = \int_{\pi/6}^{\pi/4} 2\cos^2 t\,(1 - 4\sin^2 t)\,\mathrm{d}t.$$

现在利用倍角公式，令这个积分变得容易计算。由于 $2\cos^2 t = 1 + \cos 2t$ 及 $8\sin^2 t \cos^2 t = 2\sin^2 2t = 1 - \cos 4t$，

$$2\cos^2 t - 8\sin^2 t \cos^2 t = (1 + \cos 2t) - (1 - \cos 4t) = \cos 2t + \cos 4t.$$

因此

$$\begin{aligned} I &= \int_{\pi/6}^{\pi/4} (\cos 2t + \cos 4t)\,\mathrm{d}t = \left[\tfrac{1}{2}\sin 2t + \tfrac{1}{4}\sin 4t\right]_{\pi/6}^{\pi/4} \\ &= \left(\tfrac{1}{2}\sin\tfrac{\pi}{2} + \tfrac{1}{4}\sin\pi\right) - \left(\tfrac{1}{2}\sin\tfrac{\pi}{3} + \tfrac{1}{4}\sin\tfrac{2\pi}{3}\right) \\ &= \tfrac{1}{2} - \left(\tfrac{\sqrt{3}}{4} + \tfrac{\sqrt{3}}{8}\right), \end{aligned}$$

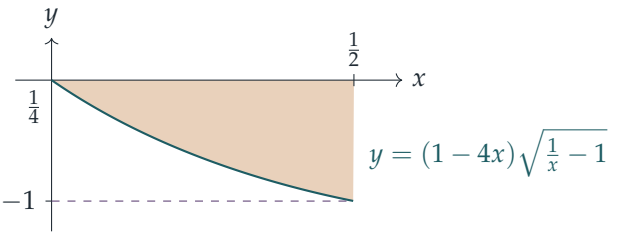
所以

$$\int_{1/4}^{1/2} (1 - 4x)\sqrt{\frac{1}{x} - 1}\,\mathrm{d}x = \frac{1}{2} - \frac{3\sqrt{3}}{8} = \frac{4 - 3\sqrt{3}}{8}.$$

验算

- 正负号：在 $(\frac{1}{4}, \frac{1}{2}]$ 上，$1 - 4x < 0$ 而平方根为正，所以被积函数为负，I 必为负数。果然 $4 - 3\sqrt{3} \approx -1.196$，所以 $I \approx -0.1495$。下图显示被积函数的图像。

- 把辛普森法则(Simpson's rule)直接应用于原来的积分,得出 $-0.149519\ldots$,与 $\frac{1}{2}-\frac{3\sqrt{3}}{8}$ 吻合至小数点后十个位。
- 至于第 (i) 部分,求导验证:$\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{2}\big(x+\ln(\cos x+\sin x)\big)=\frac{1}{2}\Big(1+\frac{\cos x-\sin x}{\cos x+\sin x}\Big)=\frac{\cos x}{\cos x+\sin x}$。例如在 $[0,1]$ 上,S 和 T 的数值均与数值积分(quadrature)的结果相符(0.66168 和 0.33832)。



常见错误

- 忘记把上下限改为 $t=\frac{\pi}{6}$ 和 $t=\frac{\pi}{4}$,或在代回 x 时出错。
- 漏掉 $\mathrm{d}x$ 中的因子 $2\sin t\cos t$,或写成 $\mathrm{d}x=2\sin t\,\mathrm{d}t$。
- 得出正数答案:被积函数在整个区间上都是负数,所以答案若为正值,即表示正负号出了错。

STEP 之外

第 (i) 部分的技巧,是替一个棘手的积分找一个"拍档",使两者之和与差都容易计算。这个技巧一再出现,例如用于 $\int \mathrm{e}^x\cos x\,\mathrm{d}x$ 和 $\int \mathrm{e}^x\sin x\,\mathrm{d}x$,而它的复数形式则是对 $\mathrm{e}^{(1+\mathrm{i})x}$ 积分。第 (ii) 部分是一个普遍原则的小例子:选得恰当的代换,可以把代数形式的被积函数化为三角多项式(trigonometric polynomial),而这正是傅里叶级数(Fourier series)的起点。想看更多这类风格的积分,并附有细致的讨论,可参阅 Siklos 的 *Advanced Problems in Mathematics*(《高等数学难题》)。

详细解答

STEP I 1995，第 3 题

纯数学 · 裂项相消级数

> **STEP I 1995，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/95-s1-q3

关键思路

当一个和式中相邻各项有大部分内容相同时，把每一项写成差 $\mathrm{f}(r)-\mathrm{f}(r-1)$ 的形式，几乎所有东西都会互相抵消：这个和式会裂项相消（telescope）。功夫在于选取 f，使这个差恰好是你想要的那一项；本题已替我们选好了。至于交错级数（alternating series），窍门是把它写成全部立方之和减去偶数立方之和的两倍，于是第 (ii) 部分的结果可以用上两次。

解答

第 (i) 部分

把和式完整写出：

$$\sum_{r=1}^{n}\{\mathrm{f}(r)-\mathrm{f}(r-1)\}=\{\mathrm{f}(1)-\mathrm{f}(0)\}+\{\mathrm{f}(2)-\mathrm{f}(1)\}+\cdots+\{\mathrm{f}(n)-\mathrm{f}(n-1)\}.$$

$\mathrm{f}(1),\mathrm{f}(2),\ldots,\mathrm{f}(n-1)$ 每一个都带正号出现一次、带负号出现一次，所以全部抵消。余下的只有第一个括号中的 $-\mathrm{f}(0)$ 和最后一个括号中的 $\mathrm{f}(n)$，得出 $\mathrm{f}(n)-\mathrm{f}(0)$，正是所求。

如果你想要比“全部抵消”更严谨的说法，可以对 n 使用归纳法（induction）。当 $n=1$ 时，和为 $\mathrm{f}(1)-\mathrm{f}(0)$。若结果对 $n=k$ 成立，则加上下一项后得

$$\mathrm{f}(k)-\mathrm{f}(0)+\mathrm{f}(k+1)-\mathrm{f}(k)=\mathrm{f}(k+1)-\mathrm{f}(0),$$

这就是 $n=k+1$ 时的结果。因此结果对所有 $n\geq 1$ 成立。

第 (ii) 部分

取 $\mathrm{f}(r) = r^2(r+1)^2$，则 $\mathrm{f}(r-1) = (r-1)^2r^2$。两者都含有因子 r^2，所以把它抽出，再运用平方差（difference of two squares）：

$$\begin{aligned}\mathrm{f}(r) - \mathrm{f}(r-1) &= r^2\left[(r+1)^2 - (r-1)^2\right] \\ &= r^2\left[(r+1) + (r-1)\right]\left[(r+1) - (r-1)\right] \\ &= r^2 \cdot 2r \cdot 2 = 4r^3.\end{aligned}$$

现在由 $r = 1$ 加至 n，并应用第 (i) 部分。由于 $\mathrm{f}(0) = 0$，

$$4\sum_{r=1}^{n} r^3 = \mathrm{f}(n) - \mathrm{f}(0) = n^2(n+1)^2, \qquad \text{所以} \qquad \boxed{\sum_{r=1}^{n} r^3 = \tfrac{1}{4}n^2(n+1)^2.}$$

第 (iii) 部分

这个级数（series）共有 $2n+1$ 项：奇数的立方带正号，偶数的立方带负号，最后一项是奇数立方 $(2n+1)^3$。

> **想一想** 只用 $\sum r^3$ 的公式，怎样处理正负交错的符号?

先考虑直到 $(2n+1)^3$ 为止的全部立方之和。当中每个偶数立方都带正号，而我们想要的是负号，所以把每个偶数立方减去两次：

$$1^3 - 2^3 + 3^3 - \cdots + (2n+1)^3 = \sum_{k=1}^{2n+1} k^3 - 2\sum_{m=1}^{n} (2m)^3.$$

偶数立方为 $(2m)^3 = 8m^3$，其中 $m = 1, \ldots, n$，最大的是 $(2n)^3$。运用第 (ii) 部分两次，其中一次以 $2n+1$ 代替 n，

$$\begin{aligned}\sum_{k=1}^{2n+1} k^3 - 16\sum_{m=1}^{n} m^3 &= \tfrac{1}{4}(2n+1)^2(2n+2)^2 - 16 \cdot \tfrac{1}{4}n^2(n+1)^2 \\ &= (2n+1)^2(n+1)^2 - 4n^2(n+1)^2 \\ &= (n+1)^2\left[(2n+1)^2 - (2n)^2\right] \\ &= (n+1)^2(4n+1),\end{aligned}$$

其中第二行用了 $(2n+2)^2 = 4(n+1)^2$，最后一行则用了平方差 $(2n+1)^2 - (2n)^2 = (2n+1+2n) \cdot 1$。因此

$$\boxed{1^3 - 2^3 + 3^3 - 4^3 + \cdots + (2n+1)^3 = (n+1)^2(4n+1).}$$

验算

- 当 $n = 0$ 时，级数只有 $1^3 = 1$，而 $(0+1)^2(0+1) = 1$。
- 当 $n = 1$ 时：$1 - 8 + 27 = 20$，而 $2^2 \times 5 = 20$。
- 当 $n = 2$ 时：$1 - 8 + 27 - 64 + 125 = 81$，而 $3^2 \times 9 = 81$。
- 答案的增长应与 $(2n+1)^3/2$ 相若，即大约是最后一项的一半，因为各项可配对成 $(2m+1)^3 - (2m)^3 \approx 12m^2$。事实上 $(n+1)^2(4n+1) \approx 4n^3$，与 $\sum 12m^2 \approx 4n^3$ 吻合。

我们亦以直接求和的方法，对由 0 至 199 的每一个 n 验证了第 (ii) 和第 (iii) 部分。

常见错误

- 只把偶数立方减去一次，得出 $\sum k^3 - \sum (2m)^3$，这是奇数立方之和，而不是交错和。
- 把偶数立方取到 $(2n+2)^3$，或在全部立方之和的公式中用了 n 而不是 $2n+1$：要仔细数清楚项数。

STEP 之外

第 (i) 部分是微积分基本定理（fundamental theorem of calculus）的离散版本：差 $\mathrm{f}(r) - \mathrm{f}(r-1)$ 扮演导数（derivative）的角色，把它求和便可还原 f，正如对 F' 积分会得出 $F(b) - F(a)$ 一样。这套“有限差分演算”（calculus of finite differences）让你只要找到合适的 f，就能求出任何多项式的和；而同样的裂项相消思想，也是许多级数收敛（convergence）证明的基础。Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）有更多关于裂项相消和这类级数的练习。

详细解答

STEP I 1995，第 4 题

纯数学 · 棣美弗定理

STEP I 1995，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/95-s1-q4

关键思路

棣美弗定理（de Moivre's theorem）把 $\sin 5\theta$ 化为 $(\cos\theta + \mathrm{i}\sin\theta)^5$ 的虚部（imaginary part），而二项式定理（binomial theorem）则把这个幂展开成 $\cos\theta$ 与 $\sin\theta$ 的乘积。虚部的每一项都含有 $\sin\theta$ 的奇数次幂，所以我们可以抽出一个因子，再利用 $\sin^2\theta = 1 - \cos^2\theta$ 把余下的部分化为余弦。接着，角 $\frac{1}{5}\pi$ 之所以特别，是因为在该处 $\sin 5\theta = 0$，于是剩下一条关于 $\cos^2\theta$ 的二次方程（quadratic equation）。

解答

恒等式

设 $c = \cos\theta$ 及 $s = \sin\theta$。由棣美弗定理，

$$\cos 5\theta + \mathrm{i}\sin 5\theta = (c + \mathrm{i}s)^5.$$

由二项式定理，系数为 $1, 5, 10, 10, 5, 1$，

$$(c + \mathrm{i}s)^5 = c^5 + 5c^4(\mathrm{i}s) + 10c^3(\mathrm{i}s)^2 + 10c^2(\mathrm{i}s)^3 + 5c(\mathrm{i}s)^4 + (\mathrm{i}s)^5.$$

由于 $\mathrm{i}^2 = -1$、$\mathrm{i}^3 = -\mathrm{i}$、$\mathrm{i}^4 = 1$ 及 $\mathrm{i}^5 = \mathrm{i}$，虚数项就是含 $\mathrm{i}s$ 奇数次幂的各项：

$$\mathrm{Im}(c + \mathrm{i}s)^5 = 5c^4 s - 10c^2 s^3 + s^5.$$

c 和 s 都是实数，所以比较虚部可得

$$\begin{aligned}\sin 5\theta &= s\left(5c^4 - 10c^2s^2 + s^4\right)\\ &= s\left(5c^4 - 10c^2(1-c^2) + (1-c^2)^2\right)\\ &= s\left(5c^4 - 10c^2 + 10c^4 + 1 - 2c^2 + c^4\right)\\ &= s\left(16c^4 - 12c^2 + 1\right),\end{aligned}$$

其中第二行用了 $s^2 = 1 - c^2$。也就是说，

$$\sin 5\theta = \sin\theta\left(16\cos^4\theta - 12\cos^2\theta + 1\right),$$

这正是所要证明的恒等式（identity）。

求 $\cos^2\frac{1}{5}\pi$ 的值

取 $\theta = \frac{1}{5}\pi$。这时 $\sin 5\theta = \sin\pi = 0$，而 $\sin\frac{1}{5}\pi \neq 0$（因为 $0 < \frac{1}{5}\pi < \pi$）。所以我们可以把恒等式两边除以 $\sin\theta$，于是 $x = \cos^2\frac{1}{5}\pi$ 满足

$$16x^2 - 12x + 1 = 0, \qquad \text{所以} \qquad x = \frac{12 \pm \sqrt{144-64}}{32} = \frac{12 \pm 4\sqrt{5}}{32} = \frac{3 \pm \sqrt{5}}{8}.$$

想一想 这条二次方程有两个根。哪一个才是 $\cos^2\frac{1}{5}\pi$?

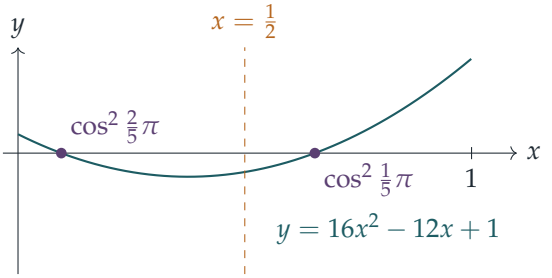
要选出正确的根，我们还需要多一项关于 $\cos\frac{1}{5}\pi$ 的事实。由于 $0 < \frac{1}{5}\pi < \frac{1}{4}\pi$，而 cos 在 $[0, \frac{1}{2}\pi]$ 上递减，我们有 $\cos\frac{1}{5}\pi > \cos\frac{1}{4}\pi = \frac{1}{\sqrt{2}}$，所以

$$\cos^2\tfrac{1}{5}\pi > \tfrac{1}{2}.$$

现在由 $\sqrt{5} < 3$ 得 $\frac{3-\sqrt{5}}{8} < \frac{3}{8} < \frac{1}{2}$，而由 $\sqrt{5} > 1$ 得 $\frac{3+\sqrt{5}}{8} > \frac{1}{2}$。恰好只有一个根大于 $\frac{1}{2}$，所以

$$\cos^2\tfrac{1}{5}\pi = \frac{3+\sqrt{5}}{8}.$$

（另一个根并非多余的假根：当 $\theta = \frac{2}{5}\pi$ 时同样有 $\sin 5\theta = 0$ 及 $\sin\theta \neq 0$，而 $\cos^2\frac{2}{5}\pi = \frac{3-\sqrt{5}}{8}$。）



推导 $\cos\frac{1}{5}\pi$

由于 $0<\frac{1}{5}\pi<\frac{1}{2}\pi$，$\cos\frac{1}{5}\pi>0$，所以我们取正平方根：

$$\cos\tfrac{1}{5}\pi=\sqrt{\frac{3+\sqrt{5}}{8}}=\sqrt{\frac{6+2\sqrt{5}}{16}}=\frac{\sqrt{6+2\sqrt{5}}}{4}.$$

最后，$(1+\sqrt{5})^2=1+2\sqrt{5}+5=6+2\sqrt{5}$ 而且 $1+\sqrt{5}>0$，所以 $\sqrt{6+2\sqrt{5}}=1+\sqrt{5}$，于是

$$\cos\tfrac{1}{5}\pi=\tfrac{1}{4}\left(1+\sqrt{5}\right),$$

这正是所要求的结果。

验算

- 当 $\theta=\frac{1}{2}\pi$ 时，恒等式给出 $\sin\frac{5}{2}\pi=1\cdot(0-0+1)=1$，结果正确。当 θ 很小时：两边都约等于 5θ，因为括号内的式子趋向 $16-12+1=5$。
- 以数值计算，$\cos 36^\circ\approx 0.809017$ 而 $\frac{1}{4}(1+\sqrt{5})\approx 0.809017$；另外 $\cos^2 36^\circ\approx 0.654508=\frac{3+\sqrt{5}}{8}$。我们亦在 10^4 个随机角度上检验过这个恒等式。
- $\frac{1}{4}(1+\sqrt{5})$ 是黄金比（golden ratio）φ 的一半，而黄金比正是正五边形的对角线与边长之比，这与 36° 这个角所暗示的一致。

常见错误

- 忘记 i 的幂所带来的正负号：s^3 项带有 $\mathrm{i}^3=-\mathrm{i}$，因而得到 $-10c^2s^3$。
- 毫无理由地选定二次方程的其中一个根。你需要一个不等式（inequality），例如 $\cos^2\frac{1}{5}\pi>\frac{1}{2}$，才能排除 $\frac{3-\sqrt{5}}{8}$。
- 除以 $\sin\theta$ 时没有指出 $\sin\frac{1}{5}\pi\neq 0$。

STEP 之外

用同样的方法，可以把 $\sin n\theta/\sin\theta$ 和 $\cos n\theta$ 写成 $\cos\theta$ 的多项式，即切比雪夫多项式（Chebyshev polynomials），它们在数值分析（numerical analysis）中随处可见。$\mathrm{e}^{2\pi\mathrm{i}k/5}$ 这些数是五次单位根（fifth roots of unity），而 $\cos\frac{1}{5}\pi$ 只涉及平方根这一事实，正是正五边形可以用尺规作图（ruler-and-compasses construction）的原因；高斯（Gauss）关于哪些正多边形可以尺规作图的定理，是大学本科伽罗瓦理论（Galois theory）的一大亮点。至于棣美弗定理背后的几何，可参阅 Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）。

详细解答

STEP I 1995，第 5 题

纯数学 · 二项式级数与积分

STEP I 1995，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/95-s1-q5

关键思路

对多项式逐项求导（differentiating term by term），可以消去麻烦的 $1/r$ 因子，剩下的几乎就是一个二项展开式（binomial expansion）：系数 $(-1)^{r+1}\binom{n}{r}$ 正好是 $-(1-x)^n$ 去掉常数项后的系数。一旦求得 f'，微积分基本定理（fundamental theorem of calculus）（连同 $\mathrm{f}(0)=0$）便把 f 表示成一个积分，再用代换（substitution）$y=1-t$ 就可化成所要求的形式。那里的被积函数（integrand）是一个等比级数（geometric series），可以立即积分。

解答

导数

记

$$\mathrm{f}(x)=\sum_{r=1}^{n}(-1)^{r+1}\binom{n}{r}\frac{x^r}{r}.$$

这是一个多项式，所以可以逐项求导：

$$\mathrm{f}'(x)=\sum_{r=1}^{n}(-1)^{r+1}\binom{n}{r}x^{r-1}.$$

当 $x \neq 0$ 时，乘以 x 并利用 $(-1)^{r+1}x^r = -(-x)^r$：

$$\begin{aligned} x\,\mathrm{f}'(x) &= -\sum_{r=1}^{n} \binom{n}{r}(-x)^r \\ &= -\Big(\sum_{r=0}^{n} \binom{n}{r}(-x)^r - 1\Big) \\ &= 1-(1-x)^n, \end{aligned}$$

这里用了二项式定理（binomial theorem）。除以 x，得

$$\mathrm{f}'(x) = \frac{1-(1-x)^n}{x} \qquad (x \neq 0),$$

正是所求。（在 $x = 0$ 处，由级数得 $\mathrm{f}'(0) = n$，这也是右边当 $x \to 0$ 时的极限，因为 $1-(1-x)^n = nx + O(x^2)$。）

积分公式

f 的每一项都含有因子 x，所以 $\mathrm{f}(0) = 0$。由微积分基本定理，

$$\mathrm{f}(x) = \mathrm{f}(0) + \int_0^x \mathrm{f}'(t)\,\mathrm{d}t = \int_0^x \frac{1-(1-t)^n}{t}\,\mathrm{d}t.$$

> **想一想** 公式的分母含有 t，积分时穿过 $t = 0$ 是否稳妥?

稳妥。被积函数其实就是多项式 $\mathrm{f}'(t)$；那个分数只是它在 $t \neq 0$ 时的一种简洁写法，而它在 $t = 0$ 处的值就是极限 n。所以这里并没有奇点（singularity），只是公式中有一个可去的缺口。

现在作代换 $y = 1 - t$，于是 $\mathrm{d}t = -\mathrm{d}y$；当 $t = 0$ 时 $y = 1$，当 $t = x$ 时 $y = 1 - x$。于是

$$\begin{aligned} \mathrm{f}(x) &= \int_1^{1-x} \frac{1-y^n}{1-y}\,(-\mathrm{d}y) \\ &= \int_{1-x}^{1} \frac{1-y^n}{1-y}\,\mathrm{d}y, \end{aligned}$$

其中把积分上下限对调，以抵消负号。这就是所要求的结果，而且对任何实数 x 都成立。

$\mathrm{f}(1)$ 的值

取 $x = 1$，则积分下限为 0。当 $y \neq 1$ 时，被积函数是一个等比级数之和，

$$\frac{1-y^n}{1-y} = 1 + y + y^2 + \cdots + y^{n-1},$$

而（一如前述）它在 $y=1$ 处的值是极限 n，所以在整个 $[0,1]$ 上，被积函数就是这个多项式。因此

$$\begin{aligned}\mathrm{f}(1) &= \int_0^1 \left(1+y+\cdots+y^{n-1}\right)\mathrm{d}y \\ &= \left[y+\frac{y^2}{2}+\cdots+\frac{y^n}{n}\right]_0^1 \\ &= 1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n},\end{aligned}$$

正是所求。

留意这个结果的含义：在原来的级数中代入 $x=1$，便得到一个令人惊讶的恒等式

$$\sum_{r=1}^{n}\frac{(-1)^{r+1}}{r}\binom{n}{r} = 1+\frac{1}{2}+\cdots+\frac{1}{n},$$

若要直接看出，绝不容易。

验算

- $n=1$：$\mathrm{f}(x)=x$，而 $\big(1-(1-x)\big)/x = 1 = \mathrm{f}'(x)$；同时 $\mathrm{f}(1)=1$。
- $n=4$：$\mathrm{f}(1) = 4-\frac{6}{2}+\frac{4}{3}-\frac{1}{4} = \frac{25}{12} = 1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}$。
- 我们在 300 个随机有理点上精确地验证了导数公式，又对随机的 $x\in[-1.5,1.5]$ 和 $n\le 15$ 把积分（用辛普森法则（Simpson's rule）计算）与级数比较（相对误差小于 10^{-11}），并对每个 $n\le 40$ 精确地确认了关于 $\mathrm{f}(1)$ 的恒等式。

常见错误

- 与 $(1-x)^n$ 的二项展开式对照时，忘记了 $r=0$ 那一项：由 $r=1$ 开始的和是 $(1-x)^n-1$，而不是 $(1-x)^n$。
- 写出 $\mathrm{f}(x)=\int_0^x \mathrm{f}'(t)\,\mathrm{d}t$ 而没有指出 $\mathrm{f}(0)=0$。
- 作代换 $y=1-t$ 之后把积分上下限的次序弄反：正是 $\mathrm{d}t=-\mathrm{d}y$ 的负号把上下限调回来。

STEP 之外

$H_n = 1+\frac{1}{2}+\cdots+\frac{1}{n}$ 这些数称为调和数（harmonic numbers）；它们的增长速度与 $\ln n$ 相若，而 $H_n-\ln n$ 趋向欧拉常数（Euler's constant）$\gamma\approx 0.5772$。这里所用的技巧，即先求导以消去麻烦的分母，然后再积分回去，正是推导 $\ln(1+x)$ 和 $\tan^{-1}x$ 的幂级数（power

series）时所用的方法；而它对无穷级数是否成立的论证（幂级数的逐项求导），是大学一年级分析学（analysis）的核心课题。Abbott 的 *Understanding Analysis* 对此有严谨的论述。

详细解答

STEP I 1995，第 6 题

纯数学 · 伯努利微分方程

STEP I 1995，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/95-s1-q6

关键思路

两条方程对 y 而言都不是线性的，但各自对 $1/y$ 的某个适当幂次却是线性的。左边含有 y'/y^2，它（除正负号外）正是 $1/y$ 的导数（derivative）；因此设 $u = 1/y$，方程便化为关于 u 的一阶线性方程（first-order linear equation），再用积分因子（integrating factor）即可解决。在第 (ii) 部分，y'/y^3 一项提示我们应改为考虑 $1/y^2$。

解答

第 (i) 部分

设 $u = 1/y$。由链式法则（chain rule），

$$\frac{\mathrm{d}u}{\mathrm{d}x} = -\frac{1}{y^2}\frac{\mathrm{d}y}{\mathrm{d}x},$$

所以方程的第一项恰好是 $-\mathrm{d}u/\mathrm{d}x$，而第二项是 u。方程变成

$$-\frac{\mathrm{d}u}{\mathrm{d}x} + u = \mathrm{e}^{2x}, \qquad \text{即} \qquad \frac{\mathrm{d}u}{\mathrm{d}x} - u = -\mathrm{e}^{2x}.$$

这对 u 而言是线性的。积分因子是 $\mathrm{e}^{\int -1\,\mathrm{d}x} = \mathrm{e}^{-x}$，全式乘以它便得

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(u\,\mathrm{e}^{-x}\right) = \mathrm{e}^{-x}\frac{\mathrm{d}u}{\mathrm{d}x} - \mathrm{e}^{-x}u = -\mathrm{e}^{x}.$$

两边积分，其中 A 为任意常数（arbitrary constant），

$$u\,\mathrm{e}^{-x} = A - \mathrm{e}^{x}, \qquad \text{所以} \qquad u = A\mathrm{e}^{x} - \mathrm{e}^{2x}.$$

由于 $y = 1/u$，

$$y = \frac{1}{A\mathrm{e}^{x} - \mathrm{e}^{2x}},$$

这就是所求的通解（general solution）。（原方程含有 $1/y$，所以自始至终 $y \neq 0$，代换不会遗漏任何解；此公式在任何满足 $Ae^x \neq e^{2x}$ 的区间上都成立，即当 $A > 0$ 时须避开 $x = \ln A$。）

第 (ii) 部分

> **想一想** 这里哪一个代换能担当 $u = 1/y$ 的角色?

我们希望新变量的导数能产生 y'/y^3。由于

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{y^2}\right) = -\frac{2}{y^3}\frac{\mathrm{d}y}{\mathrm{d}x},$$

正确的选择是 $v = 1/y^2$。于是第一项是 $-\frac{1}{2}\,\mathrm{d}v/\mathrm{d}x$，第二项是 v，所以

$$-\frac{1}{2}\frac{\mathrm{d}v}{\mathrm{d}x} + v = e^{2x}, \qquad \text{即} \qquad \frac{\mathrm{d}v}{\mathrm{d}x} - 2v = -2e^{2x}.$$

这次积分因子是 e^{-2x}：

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

两边积分，其中 B 为任意常数，

$$v\,e^{-2x} = B - 2x, \qquad \text{所以} \qquad \frac{1}{y^2} = (B - 2x)\,e^{2x}.$$

因此

$$y^2 = \frac{e^{-2x}}{B - 2x}, \qquad y = \pm\frac{e^{-x}}{\sqrt{B - 2x}}.$$

因为 $1/y^2 > 0$，我们需要 $B - 2x > 0$：解只在 $x < \frac{1}{2}B$ 时存在，而正负号则由 y 最初的符号决定（它不可能变号，因为 y 永不为零）。与第 (i) 部分不同，答案并非两个指数函数（exponential）之差：这里右边的 e^{2x} 恰好与积分因子 e^{-2x} 互相抵消，剩下一个常数待积分，这就是线性项 $-2x$ 出现的原因。

验算

- 代回 (ii)：由 $1/y^2 = (B - 2x)e^{2x}$，求导得 $-2y'/y^3 = -2e^{2x} + 2(B - 2x)e^{2x}$，所以 $y'/y^3 = e^{2x} - 1/y^2$，这正是原方程。
- 数值检验：在 (ii) 中取 $y(0) = 1$ 得 $B = 1$，于是公式预测 $y(0.4) = e^{-0.4}/\sqrt{0.2} \approx 1.49888$。由 $x = 0$ 开始以数值方法逐步求解该微分方程（differential equation），所得的值与此相同，准确至小数点后十位。两个通解在许多随机选取的点和常数下，亦都满足各自的方程。

常见错误

- 在 $\mathrm{d}u/\mathrm{d}x$ 中漏掉负号（或第 (ii) 部分的因子 2），令整条线性方程出错。
- 在第 (ii) 部分只写正平方根，或没有留意到解须满足 $x < \frac{1}{2}B$。

STEP 之外

两条方程都是伯努利方程（Bernoulli equation）$y' + P(x)y = Q(x)y^n$ 的例子：除以 y^n 并代入 $u = y^{1-n}$，总能得到一条线性方程。这里第 (i) 部分是 $n = 2$，第 (ii) 部分是 $n = 3$。解可以在有限的 x 处“爆破”（blow up）（例如 $x \to \frac{1}{2}B$ 时），这是非线性方程（nonlinear equation）的典型特征，在常微分方程（ODE）的入门课程中会学到；参见 Boyce and DiPrima, *Elementary Differential Equations*（常微分方程入门教科书）。

详细解答

STEP I 1995，第 7 题

纯数学 · 三角形的向量几何

STEP I 1995，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/95-s1-q7

关键思路

把原点设在与 A、B、C 等距的一点，位置向量（position vector）$\mathbf{a}$、$\mathbf{b}$、$\mathbf{c}$ 的长度便全部相同。整道题就靠这一个事实，因为对于长度相等的向量，$(\mathbf{b}+\mathbf{c})\cdot(\mathbf{c}-\mathbf{b}) = |\mathbf{c}|^2 - |\mathbf{b}|^2 = 0$。一旦发现 Q 的位置向量是简洁漂亮的 $\mathbf{a}+\mathbf{b}+\mathbf{c}$，两部分都只需一两行代数运算。

解答

选取原点

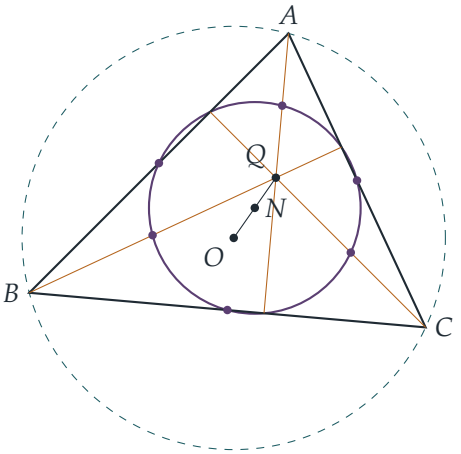
与 A 和 B 等距的点构成 AB 的垂直平分线（perpendicular bisector），这是一条垂直于 AB 的直线；B 和 C 的情况亦然。由于 A、B、C 不共线（collinear），AB 与 BC 不平行，所以它们的垂直平分线也不平行。因此这两条直线恰好相交于一点，该点与 A 和 B 等距，也与 B 和 C 等距，故与三点皆等距。（这一点就是三角形的外心（circumcentre）。）

取这一点为 O，以 $\mathbf{a}$、$\mathbf{b}$、$\mathbf{c}$ 表示 A、B、C 的位置向量，并设 R 为这个共同的距离，于是

$$|\mathbf{a}| = |\mathbf{b}| = |\mathbf{c}| = R.$$

形心（centroid）的位置向量为 $\mathbf{g} = \frac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c})$。由于 $\overrightarrow{GQ} = 2\overrightarrow{OG}$ 即 $\mathbf{q}-\mathbf{g} = 2\mathbf{g}$，

$$\mathbf{q} = 3\mathbf{g} = \mathbf{a}+\mathbf{b}+\mathbf{c}, \qquad \mathbf{n} = \tfrac{1}{2}\mathbf{q} = \tfrac{1}{2}(\mathbf{a}+\mathbf{b}+\mathbf{c}).$$



图中所示为以 O 为圆心、通过 A、B、C 的圆（虚线），相交于 Q 的三条高（altitude），以及以 N 为圆心、半径为 $\frac{1}{2}R$、通过第 (ii) 部分六个中点的圆。

第 (i) 部分

我们有 $\overrightarrow{AQ} = \mathbf{q} - \mathbf{a} = \mathbf{b} + \mathbf{c}$ 及 $\overrightarrow{BC} = \mathbf{c} - \mathbf{b}$，所以

$$\overrightarrow{AQ} \cdot \overrightarrow{BC} = (\mathbf{c} + \mathbf{b}) \cdot (\mathbf{c} - \mathbf{b}) = |\mathbf{c}|^2 - |\mathbf{b}|^2 = R^2 - R^2 = 0,$$

其中用了 $\mathbf{b} \cdot \mathbf{c} = \mathbf{c} \cdot \mathbf{b}$ 来消去交叉项。因此 $\overrightarrow{AQ}$ 垂直于 $\overrightarrow{BC}$。

? **想一想** $\overrightarrow{AQ} \cdot \overrightarrow{BC} = 0$ 真的保证 Q 在由 A 作出的高上吗？如果 $\overrightarrow{AQ}$ 是零向量（zero vector）又如何？

若 $Q \neq A$，直线 AQ 通过 A 且垂直于 BC，所以它正是由 A 作出的高，而 Q 在其上。若 $Q = A$（这恰好在 $\mathbf{b} + \mathbf{c} = \mathbf{0}$ 时发生，即 BC 为直径而 A 处的角为直角），则 Q 显然在由 A 作出的高上。无论哪种情况，Q 都在由 A 作出的高上。

以上论证并没有对 A 另眼相看：表达式 $\mathbf{a} + \mathbf{b} + \mathbf{c}$ 是对称的，所以同理可得

$$\overrightarrow{BQ} \cdot \overrightarrow{CA} = (\mathbf{c} + \mathbf{a}) \cdot (\mathbf{a} - \mathbf{c}) = |\mathbf{a}|^2 - |\mathbf{c}|^2 = 0,$$
$$\overrightarrow{CQ} \cdot \overrightarrow{AB} = (\mathbf{a} + \mathbf{b}) \cdot (\mathbf{b} - \mathbf{a}) = |\mathbf{b}|^2 - |\mathbf{a}|^2 = 0,$$

故 Q 也在由 B 和由 C 作出的高上。所以三条高都通过同一点 Q：它们共点（concurrent）。（Q 就是垂心（orthocentre）。）

第 (ii) 部分

AQ 的中点的位置向量为 $\frac{1}{2}(\mathbf{a} + \mathbf{q}) = \mathbf{a} + \frac{1}{2}(\mathbf{b} + \mathbf{c})$，所以它相对于 N 的位移为

$$\mathbf{a} + \tfrac{1}{2}(\mathbf{b} + \mathbf{c}) - \tfrac{1}{2}(\mathbf{a} + \mathbf{b} + \mathbf{c}) = \tfrac{1}{2}\mathbf{a}.$$

BC 的中点的位置向量为 $\frac{1}{2}(\mathbf{b}+\mathbf{c})$，所以它相对于 N 的位移为

$$\tfrac{1}{2}(\mathbf{b}+\mathbf{c})-\tfrac{1}{2}(\mathbf{a}+\mathbf{b}+\mathbf{c})=-\tfrac{1}{2}\mathbf{a}.$$

两者的长度都是 $\frac{1}{2}|\mathbf{a}|=\frac{1}{2}R$。

根据同样的对称性，BQ 和 CA 的中点相对于 N 的位移为 $\pm\frac{1}{2}\mathbf{b}$，而 CQ 和 AB 的中点则为 $\pm\frac{1}{2}\mathbf{c}$。由于 $|\mathbf{a}|=|\mathbf{b}|=|\mathbf{c}|=R$，六个中点与 N 的距离全部是 $\frac{1}{2}R$，这正是我们要证明的。事实上我们还多证了一点：AQ 的中点和 BC 的中点位于这个圆某条直径的两端，其余两对中点亦然。

验算

- 等边三角形：此时 $\mathbf{a}+\mathbf{b}+\mathbf{c}=\mathbf{0}$，所以 O、G、Q 和 N 全部重合，而六个中点与中心的距离应该都是 $\frac{1}{2}R$。事实上，各边中点与中心的距离为 $R\cos 60^\circ=\frac{1}{2}R$，而 AQ 等的中点则位于半径的正中。
- A 处为直角：此时 $Q=A$，与上面论证所预测的一致，因为两条直角边本身就是通过 A 的高。
- 数值上，我们随机取了 10^5 个三角形，直接计算外心，并确认了 $Q=\mathbf{a}+\mathbf{b}+\mathbf{c}$、三个点积（dot product）全部为零，以及六个中点与 N 的距离全部是 $\frac{1}{2}R$，最大相对误差约为 10^{-13}。

常见错误

- 把 $\overrightarrow{GQ}=2\overrightarrow{OG}$ 读成 $\mathbf{q}=2\mathbf{g}$。它说的是 $\mathbf{q}-\mathbf{g}=2\mathbf{g}$，所以 $\mathbf{q}=3\mathbf{g}$。
- 只证明了一个垂直关系便停笔。要证明共点，Q 必须在全部三条高上；应明确写出同样的计算对 B 和 C 也成立。
- 忘记说明原点为何可以这样选：这一点之所以存在，是因为两条不平行的边的垂直平分线必定相交。

STEP 之外

由于 $\mathbf{q}=3\mathbf{g}$ 及 $\mathbf{n}=\frac{3}{2}\mathbf{g}$，$O$、$G$、$N$、$Q$ 四点全在同一条直线上，这条直线称为欧拉线（*Euler line*），且 $OG:GQ=1:2$。第 (ii) 部分的圆就是九点圆（*nine-point circle*）：它还通过三条高的垂足（图中可见），因为每个垂足对某条直径（例如由 AQ 的中点至 BC 的中点的那一条）所张的角都是直角。巧妙地选取原点，让一个对称的表达式代劳，是向量几何和线性代数（linear algebra）中反复出现的技巧；Strang 的 *Introduction to Linear Algebra*

（线性代数入门教材）进一步发展了向量和点积的工具，而 Siklos 的 *Advanced Problems in Mathematics*（STEP 进阶题集）则有更多同类风格的 STEP 式几何题。

详细解答

STEP I 1995，第 8 题

纯数学 · 线性微分方程

STEP I 1995，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/95-s1-q8

关键思路

未知的是三个函数(function),但在每个固定的 x 处,它们只不过是三个数 $\mathrm{f}(x)$、$\mathrm{g}(x)$、$\mathrm{h}(x)$。把题目给出的三个解代入 $(*)$，便得到关于这三个数的三条线性方程（linear equation），我们把它们当作联立方程（simultaneous equations）来解。至于第二段，$(*)$ 的左边对 y 而言是线性的，所以 x、1 和 x^{-1} 的组合所得的结果，可由每一项各自产生的结果推出。

解答

求 f、g 和 h

把左边 $y'' + \mathrm{f}(x)y' + \mathrm{g}(x)y$ 记作 Ly，于是 $(*)$ 就是 $Ly = \mathrm{h}(x)$。我们把各个解逐一代入。

- $y = 1$：这时 $y' = y'' = 0$，所以 $\mathrm{g}(x) = \mathrm{h}(x)$。
- $y = x$：这时 $y' = 1$，$y'' = 0$，所以 $\mathrm{f}(x) + x\,\mathrm{g}(x) = \mathrm{h}(x)$。
- $y = x^{-1}$：这时 $y' = -x^{-2}$，$y'' = 2x^{-3}$，所以

$$\frac{2}{x^3} - \frac{\mathrm{f}(x)}{x^2} + \frac{\mathrm{g}(x)}{x} = \mathrm{h}(x).$$

在第二条方程中使用 $\mathrm{h} = \mathrm{g}$，得 $\mathrm{f}(x) = (1-x)\,\mathrm{g}(x)$。把这两个结果代入第三条方程，并在两边同乘 x^3（因为 $x > 0$，这样做是容许的）：

$$\begin{aligned} 2 - x(1-x)\,\mathrm{g}(x) + x^2\,\mathrm{g}(x) &= x^3\,\mathrm{g}(x) \\ 2 = \mathrm{g}(x)\left(x^3 - 2x^2 + x\right) &= \mathrm{g}(x)\,x(1-x)^2. \end{aligned}$$

由于 $0 < x < 1$，因子 $x(1-x)^2$ 不等于零，所以可以用它来除。因此

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

在每个 x 处，我们都是一步一步解出这三条方程，过程中没有任何自由选择的余地，所以符合条件的函数就只有这一组。限制 $0 < x < 1$ 使我们避开 $x = 0$ 和 $x = 1$，这些函数在该两处会趋向无穷大。

所以这条方程是

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

$ax + b + c/x$ 何时是解?

算子（operator）L 是线性的：对常数 α, β，有 $L(\alpha u + \beta v) = \alpha\,Lu + \beta\,Lv$，因为求导（differentiation）是线性的，而 f、g 只是分别乘以 y' 和 y。由第一段可知

$$L(x) = L(1) = L(x^{-1}) = \mathrm{h}(x).$$

因此，对于 $y = ax + b + c/x$，

$$Ly = a\,L(x) + b\,L(1) + c\,L(x^{-1}) = (a+b+c)\,\mathrm{h}(x).$$

> ? **想一想** 为甚么我们可以断定 $a+b+c=1$，而不仅仅是“在某个 x 处 $a+b+c=1$”?

我们需要 $(a+b+c)\,\mathrm{h}(x) = \mathrm{h}(x)$ 对所有 $0 < x < 1$ 成立，即 $(a+b+c-1)\,\mathrm{h}(x) = 0$。但在这个区间上 $\mathrm{h}(x) = 2/\big(x(1-x)^2\big) > 0$，所以可以用它来除。因此 $y = ax + b + c/x$ 是 $(*)$ 的解，当且仅当（if and only if）

$$a + b + c = 1.$$

验算

- 取 $(a,b,c) = (1,0,0)$、$(0,1,0)$ 或 $(0,0,1)$，便得回题目给出的三个解，而每一个都满足 $a+b+c=1$。
- 直接验证：取 $x = \frac{1}{2}$，得 $\mathrm{f} = 8$，$\mathrm{g} = \mathrm{h} = 16$。对于 $y = 1/x$，有 $y = 2$，$y' = -4$，$y'' = 16$，而 $16 - 32 + 32 = 16$，符合要求。
- 电脑在 $(0,1)$ 内的 49 个有理数点上精确地检验了全部三个解，并对 2000 组随机的 (a,b,c,x) 确认了 $Ly = (a+b+c)\mathrm{h}$。

常见错误

- 对 $y = x^{-1}$ 写成 $y'' = -2x^{-3}$：两个负号相乘，结果应是 $+2x^{-3}$。
- 答成 $a + b + c = 0$。那是 $ax + b + c/x$ 满足齐次（homogeneous）方程 $Ly = 0$ 的条件；这里右边是 h，不是 0。

STEP 之外

这正是线性微分方程（linear differential equation）的解的结构：$Ly = \mathrm{h}$ 的任意两个解之差都是 $Ly = 0$ 的解。这里 $x - 1$ 和 $x^{-1} - 1$ 是齐次方程的解，而对于二阶方程，两个线性无关（independent）的解便可给出全部的解。所以 $1 + \alpha(x - 1) + \beta(x^{-1} - 1)$ 就是通解（general solution），而这恰好就是满足 $a + b + c = 1$ 的那一族解。朗斯基行列式（Wronskian）可以把“线性无关”的意思说得精确。Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科书）涵盖了以上所有内容。

详细解答

STEP I 1995，第 9 题

力学 · 抛体轨道的包络线

> **STEP I 1995，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/95-s1-q9

关键思路

固定一点 (x, y)，然后问：是否存在某个抛射角（angle of projection），使轨迹（trajectory）经过这一点？把轨迹方程用 $\tan\theta$ 而不是 θ 来表示，它就是一个关于 $\tan\theta$ 的二次方程（quadratic）。一点可以到达，当且仅当这个二次方程有实根（real root），所以整个问题归结为判别式（discriminant）。

解答

模型

我们把粒子视为只受重力作用的质点（point mass）：没有空气阻力（air resistance），g 的大小和方向均不变，路径上亦没有任何障碍。这样，粒子便在包含其初速度的铅垂平面内运动；又因为只有与 O 的水平距离才有关系，由对称性可取 $x \geq 0$，并设粒子朝 x 增加的方向抛出，与水平线所成的仰角为 θ，其中 $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$。（垂直向上抛出，即 $\theta = \frac{\pi}{2}$，留待最后处理。）

轨迹

设 $u = \sqrt{2gh}$，故 $u^2 = 2gh$。加速度恒为向下的 g，经过时间 t 后

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

利用 $t = x/(u\cos\theta)$ 消去时间，并运用 $\sec^2\theta = 1 + \tan^2\theta$，得

$$y = x\tan\theta - \frac{gx^2}{2u^2}\left(1 + \tan^2\theta\right) = x\tan\theta - \frac{x^2}{4h}\left(1 + \tan^2\theta\right),$$

因为 $2u^2 = 4gh$。设 $T = \tan\theta$，并把全式乘以 $4h$，则轨迹经过 (x, y) 当且仅当

$$x^2T^2 - 4hx\,T + \left(x^2 + 4hy\right) = 0. \qquad (*)$$

当 θ 取遍 $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ 时，$T = \tan\theta$ 恰好取每个实数值一次，所以选择一个角度等同于选择一个实数 T。

> **想一想** 为甚么要用 $\tan\theta$ 而不用 θ 来表示轨迹?

若以 θ 表示，方程中混杂了 sin、cos 和 $\sec^2$，处理起来很不方便。利用 $\sec^2\theta = 1 + \tan^2\theta$，一切都变成单一变量 T 的多项式，而“是否存在这样的角度?”就变成“这个二次方程有没有实根?”，这个问题我们可以立即回答。

哪些点可以到达

取 $x > 0$。此时 $(*)$ 是一个真正关于 T 的二次方程，它有实根当且仅当其判别式非负：

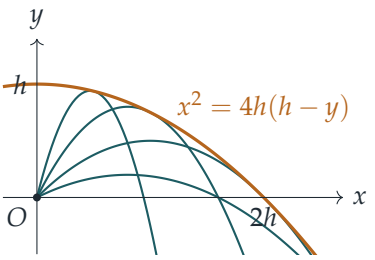
$$\begin{aligned}(4hx)^2 - 4x^2\left(x^2 + 4hy\right) \geq 0 &\iff 4x^2\left(4h^2 - x^2 - 4hy\right) \geq 0\\ &\iff x^2 \leq 4h^2 - 4hy = 4h(h - y),\end{aligned}$$

其中我们除以了 $4x^2 > 0$。因此，若 (x, y) 位于抛物线上方，即 $x^2 > 4h(h - y)$，或等价地 $y > h - \dfrac{x^2}{4h}$，则 $(*)$ 没有实根，没有任何抛射角能使粒子经过 (x, y)。

余下要处理的是 $x = 0$，即通过 O 的铅垂线。这条线上 $y \neq 0$ 的点，只有垂直抛射才能到达（任何其他角度都会令 $x > 0$ 对所有 $t > 0$ 成立）。以速率 u 垂直向上抛出时，粒子上升至高度 $u^2/(2g) = h$ 便不再上升，而这正是抛物线与 $x = 0$ 相交的一点。所以在这条线上，抛物线上方的点同样无法到达。

因此，无论抛射角是多少，粒子永远不能到达抛物线 $x^2 = 4h(h - y)$ 上方的任何一点。■

这个界限是最佳的：当 $x^2 = 4h(h - y)$ 时，判别式为零，恰好有一个角度 $\tan\theta = 2h/x$，其路径刚好与抛物线相切。这条抛物线是所有轨迹的包络线（envelope），常称为安全抛物线（parabola of safety）。



四条速率为 $\sqrt{2gh}$ 的轨迹（青绿色），每条都与边界抛物线（琥珀色）相切一次。

简化假设

- 粒子是一个质点（没有大小，没有自转）。
- 忽略空气阻力（以及重力以外的任何其他力）。
- 重力场是均匀的：g 在整个运动范围内为常数，这亦意味着把地球表面视为平面，并忽略地球的自转。
- 路径不受阻挡（例如不被地面阻挡），所以粒子确实能够到达所讨论的点。

验算

- 在 $x = 0$ 处，抛物线给出 $y = h = u^2/(2g)$，即垂直上抛的最大高度。在 $y = 0$ 上，它给出 $x = 2h = u^2/g$，即大家熟悉的水平地面上的最大射程（maximum range）（在 45° 时取得，此时 $\tan\theta = 2h/x = 1$，与预测相符）。
- 数值上，取 $g = 9.8$ 和 $h = 1.7$，我们对 399 个抛射角（由几乎垂直向下到几乎垂直向上）逐步模拟运动；在所有路径上，高出抛物线的最大高度约为 -5×10^{-13}，即在舍入误差范围内为零；而在固定 x 之下对角度求 y 的最大值，每次都得回 $h - x^2/(4h)$。

常见错误

- 只考虑 θ 在 $\left[0, \frac{\pi}{2}\right]$ 范围内的情况。位于 O 下方的点需要向下抛射（$T < 0$），论证必须涵盖这些情况。
- 除以 x（或 x^2）时，没有另外处理 $x = 0$ 的情况。
- 只证明了某一条特定轨迹位于抛物线之下；命题所说的是*每一个*角度，这正是我们固定一点而让角度变化的原因。

STEP 之外

这里的抛物线是单参数轨迹族的*包络线*：即与族中每一条曲线都相切的曲线。对于曲线族 $F(x, y, T) = 0$，包络线可由 $F = 0$ 与 $\partial F/\partial T = 0$ 消去 T 而求得；对于关于 T 的二次方程，这恰恰就是判别式等于零。包络线还会在光学的焦散线（caustics）、一阶微分方程理论（奇解，singular solutions），以及给定速率的所有抛射体的“安全区域”中再度出现。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门）是进一步学习有空气阻力和无空气阻力的抛体运动的好选择。

详细解答

STEP I 1995，第 10 题

力学 · 弹性绳与反弹

STEP I 1995，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/95-s1-q10

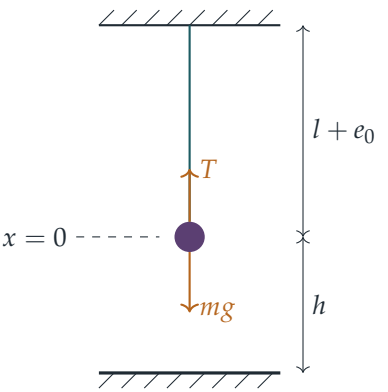
关键思路

若从平衡位置（equilibrium position）量起，重量与张力（tension）中属于平衡状态的那一部分恰好互相抵消，因此唯一未被平衡的力，就是与位移 x 成正比的胡克定律（Hooke's law）回复力（restoring force）。由此可知运动是简谐运动（simple harmonic motion），而它的能量方程正是题目所求的速率公式。其后的一切只是逐项记帐：能量方程负责把球在 $x = 0$ 与表面之间的运动联系起来，而牛顿实验定律（Newton's experimental law）则在每次反弹时把速度反向并按比例缩小。

解答

我们把球当作质点（particle），把绳当作轻而有弹性的绳，并假设它在拉长时服从胡克定律。我们忽略空气阻力，并假设反弹在瞬间完成。全题记

$$\omega^2 = \frac{\lambda}{lm}.$$



平衡时的长度

设平衡时的伸长量（extension）为 e_0，则张力为 $T_0 = \lambda e_0/l$。球处于平衡时，沿铅垂方向分解力，得 $T_0 = mg$，所以

$$\frac{\lambda e_0}{l} = mg \quad \Longrightarrow \quad e_0 = \frac{mgl}{\lambda},$$

而总长度为 $l + e_0 = l(1 + mg/\lambda)$，正如题目所求。

速率公式

设 x 为球在平衡位置以下的距离，并取向下为正。当 $x \geq 0$ 时，伸长量为 $e_0 + x > 0$，所以绳是拉紧的，张力为 $\lambda(e_0 + x)/l$。沿向下方向应用牛顿第二定律（Newton's second law），得

$$m\ddot{x} = mg - \frac{\lambda(e_0 + x)}{l} = -\frac{\lambda x}{l},$$

因为 $\lambda e_0/l = mg$。所以 $\ddot{x} = -\omega^2 x$：这是以 $x = 0$ 为中心的简谐运动。

写成 $\ddot{x} = v\dfrac{\mathrm{d}v}{\mathrm{d}x}$ 并积分，

$$\int v\,\mathrm{d}v = -\omega^2 \int x\,\mathrm{d}x \quad \Longrightarrow \quad v^2 = -\omega^2 x^2 + C.$$

在 $x = 0$ 时速率为 u_0，所以 $C = u_0^2$，而

$$v^2 + \frac{\lambda}{lm}x^2 = u_0^2,$$

正如题目所求。（这等价于能量守恒（conservation of energy）：重力势能与弹性势能合起来就是 $\frac{1}{2} \cdot \frac{\lambda}{l}x^2$ 加一个常数。）

第一次反弹

球能到达表面，当且仅当在 $x = h$ 时 $v^2 > 0$，即 $u_0^2 - \omega^2 h^2 > 0$。这正是题目给出的条件 $\lambda h^2 < lmu_0^2$。所以球以速率 w_0 撞击表面，其中

$$w_0^2 = u_0^2 - \omega^2 h^2.$$

根据牛顿实验定律，球离开表面时向上运动，速率为 ew_0。

球由 $x = h$ 上升到 $x = 0$ 的过程中，绳仍然处于拉长状态，所以同一条运动方程依然成立，而能量积分为 $v^2 + \omega^2 x^2 =$ 常数。唯一不同的是这个常数的值。在 $x = h$ 时 $v^2 = e^2 w_0^2$，所以常数为 $e^2 w_0^2 + \omega^2 h^2$，而在 $x = 0$ 时

$$\begin{aligned} u_1^2 &= e^2 w_0^2 + \omega^2 h^2 \\ &= e^2\left(u_0^2 - \omega^2 h^2\right) + \omega^2 h^2 \\ &= e^2 u_0^2 + \frac{\lambda}{lm}h^2\left(1 - e^2\right), \end{aligned}$$

正如题目所求。这里 u_1 的方向向上。

第二次反弹

> **想一想** 球经过 $x=0$ 之后会升到平衡位置之上，绳可能会变松。球回到 $x=0$ 时，速率还是 u_1 吗?

是的。在 $x=0$ 之上，球在重量和绳的张力作用下运动；若绳变松，则只受重力作用。这两个力都是保守力（conservative force）：绳松弛时，它的弹性势能就是零。因此在 $x=0$ 之上的整段运动中，总机械能守恒。球落回 $x=0$ 时，势能与离开时相同，所以速率同样是 u_1，只是方向变为向下。

所以第二次下降与第一次完全一样，只是以 u_1 取代 u_0。我们仍须确定球能到达表面，而它确实可以，因为

$$u_1^2-\omega^2h^2=e^2\left(u_0^2-\omega^2h^2\right)>0.$$

把 u_0 换成 u_1，重复第一次反弹的论证，

$$\begin{aligned}u_2^2&=e^2u_1^2+\omega^2h^2\left(1-e^2\right)\\&=e^2\left[e^2u_0^2+\omega^2h^2(1-e^2)\right]+\omega^2h^2(1-e^2)\\&=e^4u_0^2+\omega^2h^2(1-e^2)(1+e^2)\\&=e^4u_0^2+\frac{\lambda}{lm}h^2\left(1-e^4\right),\end{aligned}$$

全题至此完成。

验算

- 若 $e=1$，则没有能量损失，两条公式都给出 $u_1=u_2=u_0$，与预期相符。
- 若 $e=0$，球离开表面时处于静止，于是 $u_1^2=\omega^2h^2$：球由振幅（amplitude）h 处从静止开始上升，并以速率 ωh 通过中心，正如简谐运动所预测。
- 有一个简洁的方法看出规律：$u_n^2-\omega^2h^2=e^{2n}\left(u_0^2-\omega^2h^2\right)$，因为每次反弹都把撞击速率的平方乘以 e^2。
- 我们以数值方法逐步计算这个运动（龙格—库塔法（Runge–Kutta），并容许绳变松）。例如，取 $m=1$、$l=1$、$\lambda=20$、$g=9.8$、$u_0=6$、$h=0.3$ 及 $e=0.9$，模拟结果为 $u_1=5.4316$ 和 $u_2=4.9233$，与公式吻合。在这个例子中，绳在两次反弹之间确实会变松。

常见错误

- 从自然长度（natural length）而非平衡位置量度 x。这样重量就不会抵消，公式中会多出一次项。
- 把恢复系数（coefficient of restitution）e 用于 $x = 0$ 处的速率，而不是用于 $x = h$ 处的撞击速率。
- 假设绳在平衡位置之上仍然保持拉紧。其实不一定，但由于能量守恒，这一点并不影响球在 $x = 0$ 处的速率。

STEP 之外

重力在平衡点附近抵消，是一个普遍原理的最初例子：在平衡点附近作线性化（linearise），则小幅振动是简谐运动，其 ω^2 等于“劲度”除以质量。弹跳的球是分段光滑振子（piecewise-smooth oscillator），又称碰撞振子（impact oscillator）的一个例子。当表面本身也在运动时，这类系统的行为可以异常丰富（倍周期（period doubling），甚至混沌（chaos））。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）以大学一年级的程度讲解振子和能量方法。

详细解答

STEP I 1995，第 11 题

力学 · 圆柱之间的摩擦力

STEP I 1995，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/95-s1-q11

关键思路

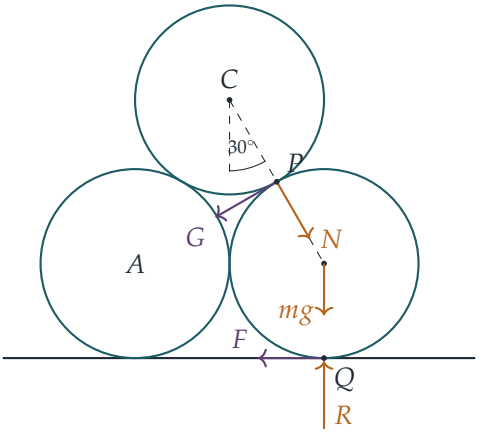
先看底下其中一个圆柱。它的重量和来自顶部圆柱的法向反作用力（normal reaction）都通过它的中心，所以对中心取力矩（moments）时只涉及两个摩擦力（friction），而这两个摩擦力的大小必定相等。接着只要沿水平方向分解（resolve）一次，就能定出圆柱之间接触点上摩擦力与法向反作用力之比，余下的工作就交给三个互相接触的圆的几何（30° 和 60° 的角）来完成。

解答

模型

我们把这些圆柱视为半径为 r 的均匀刚体（rigid body）圆柱，并在垂直于各轴的横截面上考虑问题，因此每个圆柱都是一个圆，其重量作用于圆心。每个接触点的摩擦系数（coefficient of friction）同为 μ。把底下的两个圆柱称为 A（左）和 B（右），顶部的一个称为 C。A 和 B 的中心相距 $2r$，高度同为 r；C 的中心与两者各相距 $2r$，所以三个中心所成的三角形是等边三角形，而由 C 的中心至 B 的中心的直线与铅垂线成 30° 角。

我们暂且假设 A 和 B 之间互不施力；这一点留待最后再作讨论。根据对称性（symmetry），只需考虑 B 以及 C 右边的接触点。



作用于 B 的力有：

- 作用于其中心的重量 mg；
- 在与地面的接触点 Q：向上的法向反作用力 R，以及沿水平方向指向 A 的摩擦力 F；
- 在与 C 的接触点 P：沿连心线的法向反作用力 N（向下向外，与铅垂线成 30° 角），以及沿公切线（common tangent）的摩擦力 G，我们设它的方向为向下并指向 A，即在水平线以下 30°。

假如其中一个摩擦力实际上是朝相反方向作用，算出来的值自然会是负数。

对 B 的中心取力矩

重量通过 B 的中心，N 亦然，因为两圆之间的法向反作用力沿它们的连心线作用。两个摩擦力都沿切线作用，与中心的垂直距离为 r。F 位于底部并指向 A，在图中使 B 顺时针转动；G 位于 P 并沿切线向下指向 A，使它逆时针转动。因此

$$Fr = Gr, \qquad \text{所以} \qquad F = G.$$

对 B 沿水平方向分解

N 有一个向外的水平分量 $N\sin 30^\circ = \frac{1}{2}N$。$F$ 以及 G 的水平分量 $G\cos 30^\circ = \frac{\sqrt{3}}{2}G$ 都指向内。所以

$$\tfrac{1}{2}N = F + \tfrac{\sqrt{3}}{2}G = G\left(1 + \tfrac{\sqrt{3}}{2}\right),$$

由此得

$$\frac{G}{N} = \frac{1}{2+\sqrt{3}} = \frac{2-\sqrt{3}}{(2+\sqrt{3})(2-\sqrt{3})} = 2-\sqrt{3}.$$

这是正数，所以 G 确实沿我们所设的方向作用，$F = G$ 亦然。

要令 P 处的接触点不发生滑动，我们需要 $G \le \mu N$，即

$$\mu \ge 2 - \sqrt{3}.$$

与地面的接触点

我们还须检查与地面的接触点不会要求更大的摩擦力。对 C 而言，来自 B 的力有：N，方向为向上偏左，与铅垂线成 30° 角；以及 G，沿切线向上，与水平线成 30° 角。它们的铅垂分量分别为 $\frac{\sqrt{3}}{2}N$ 和 $\frac{1}{2}G$，而根据对称性，A 的贡献亦相同。对 C 沿铅垂方向分解，

$$2\left(\tfrac{\sqrt{3}}{2}N + \tfrac{1}{2}G\right) = mg \implies N\left(\sqrt{3} + 2 - \sqrt{3}\right) = mg \implies N = \tfrac{1}{2}mg.$$

对整个系统沿铅垂方向分解，系统的重量为 $3mg$，根据对称性由两边平均承担，故 $R = \frac{3}{2}mg$。于是

$$\frac{F}{R} = \frac{(2-\sqrt{3})\cdot\frac{1}{2}mg}{\frac{3}{2}mg} = \frac{2-\sqrt{3}}{3} < 2 - \sqrt{3},$$

所以只要 $\mu \ge 2 - \sqrt{3}$，与地面的接触点同样不会滑动。圆柱之间的接触点才是关键所在。

> ? **想一想** A 和 B 互相接触。为甚么我们可以忽略它们之间的任何力？

它们只能把对方推开，所以假设 A 以大小为 $S \ge 0$ 的力沿连心线把 B 向外推。这个力通过 B 的中心，所以力矩方程仍然给出 $F = G$，但水平方向的分解变成 $\frac{1}{2}N + S = G\left(1 + \frac{\sqrt{3}}{2}\right)$。与此同时，$C$ 的铅垂方向方程仍然成立，把两者联立求解，可知 G/N 随 S 增加而增加（当 $S = 0$ 时其值为 $2 - \sqrt{3}$）。这个问题是静不定（statically indeterminate）的，因为 S 不能由这些方程确定，而当 $S = 0$ 时所需的摩擦力最小。所以摩擦系数的最小可能值是 $2 - \sqrt{3}$，正如题目所求。

验算

- $2 - \sqrt{3} = \tan 15° \approx 0.268$。所以在圆柱之间的接触点，总反作用力与法线恰好成 15° 角，正是几何中 30° 角的一半。这是个令人欣喜的迹象，显示代数运算无误。
- 地面的比值 $(2 - \sqrt{3})/3 \approx 0.089$ 是圆柱之间比值的三分之一，这很合理：地面承受 $\frac{3}{2}mg$，是 N 的三倍，但摩擦力的大小相同。
- 作数值验算，把数值代入 B 和 C 的所有力方程和力矩方程，并对未知的推力 $S \ge 0$ 求最小值，得出最小的 μ 为 0.267949…，与 $2 - \sqrt{3}$ 吻合。

常见错误

- 对地面接触点或对 P 取力矩，而不是对中心取力矩。这样做也行得通，但会引入 N，令代数运算较为繁复；对中心取力矩则可同时消去 mg 和 N。
- 只检查地面。地面接触点所需的摩擦力小得多，所以会得出错误的临界值。
- 对 C 沿铅垂方向分解时，忘记了根据牛顿第三定律（Newton's third law），作用于 C 的摩擦力是沿切线向上的。

STEP 之外

底下两个圆柱之间的力 S 无法由平衡方程求出：这个结构是静不定的，而题目很合理地要求在 S 的所有可能取值之中求最小的摩擦。在工程学上，这类系统要靠建立弹性形变的模型来解决，而同样的问题亦见于颗粒物料的力学，例如叠起的喉管或沙堆。若想细读有关静力学、摩擦和刚体平衡的严谨论述，可参阅 Kleppner and Kolenkow, *An Introduction to Mechanics*（力学入门教科书）。

详细解答

STEP I 1995，第 12 题

概率与统计 · 排列与计数

STEP I 1995，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/95-s1-q12

关键思路

n 名学生的每一种排列（arrangement）出现的机会均等，所以每个概率（probability）都等于有利排列的数目除以 $n!$。更好的做法是利用对称性（symmetry）：我们只需关心曲棍球员占据了哪些位置，而 $\binom{n}{r}$ 个由 r 个位置组成的集合，每一个出现的机会都相同。于是每一部分都变成一个关于选取位置的计数问题。至于最后一部分，题目中关于空隙的提示正是标准的“隔板法”（stars and bars）技巧：先把非曲棍球员排成一列，曲棍球员就必须分别放进他们之间不同的空隙。

解答

以下一律把 r 名曲棍球员记作 H，其余 $n-r$ 名学生记作 N。把学生视为可区分的，则全部 $n!$ 种排序出现的机会均等。由于 H 所占的每一个 r 个位置的集合，都对应相同数目的排序，即 $r!\,(n-r)!$ 种，所以 H 所占位置的集合，等可能地是 $\binom{n}{r}$ 个可能集合中的任何一个。以下哪一种观点较方便，我们就用哪一种。

第 (i) 部分

先填两端的位置。左端的学生是曲棍球员的概率为 $\frac{r}{n}$。在此条件下，右端的学生是余下 $n-1$ 名学生的其中一人，各人机会均等，而当中有 $r-1$ 人打曲棍球。所以

$$\mathrm{P}(\text{两端都是曲棍球员}) = \frac{r}{n}\cdot\frac{r-1}{n-1} = \frac{r(r-1)}{n(n-1)}.$$

（要这个概率不为零，需要 $r \geq 2$，而题目已保证这一点。）

第 (ii) 部分

曲棍球员全部站在一起，当且仅当他们的位置构成连续 r 个位置的一段。这样的一段可以由第 $1, 2, \ldots, n-r+1$ 个位置开始，所以在 $\binom{n}{r}$ 个位置集合之中，有 $n-r+1$ 个是有利的：

$$\mathrm{P}(\text{全部站在一起}) = \frac{n-r+1}{\binom{n}{r}} = \frac{r!\,(n-r+1)!}{n!}.$$

另一个等价的做法，是把 r 名曲棍球员黏合成一个“超级学生”：$n-r+1$ 个单位共有 $(n-r+1)!$ 种排序，而该段内部有 $r!$ 种排序，得出相同的答案。作为验算，当 $r=n$ 时这个概率为 1，正如所料。

第 (iii) 部分

先让 $n-r$ 名非曲棍球员排成一列。他们造成 $n-r+1$ 个空隙：第一人之前有一个，每对相邻的人之间各有一个，最后一人之后有一个。



没有两名曲棍球员相邻，当且仅当每个空隙最多放入一名曲棍球员。（若两个 H 在同一个空隙，他们便会相邻；若他们在不同的空隙，则至少有一个 N 把他们隔开。）因此，有利的排列对应于从 $n-r+1$ 个空隙中选出 r 个，每个选中的空隙放一个 H。

以位置来计数：一个有利的 H 位置集合，完全取决于用了哪 r 个空隙，所以共有 $\binom{n-r+1}{r}$ 个。因此

$$\mathrm{P}(\text{没有两人相邻}) = \frac{\binom{n-r+1}{r}}{\binom{n}{r}} = \frac{(n-r+1)!\,(n-r)!}{(n-2r+1)!\,n!}.$$

（直接数学生的排序也得出同样的结果：N 的排序有 $(n-r)!$ 种，然后把 r 个互不相同的 H 放进不同的空隙，有 $(n-r+1)(n-r)\cdots(n-2r+2) = \frac{(n-r+1)!}{(n-2r+1)!}$ 种方法，再全部除以 $n!$。）

> **想一想** 在甚么情况下，不可能把所有曲棍球员都分隔开？

我们需要 r 个不同的空隙，而空隙只有 $n-r+1$ 个。所以：

- 若 $r \le n-r+1$，即 $n \ge 2r-1$，概率就是上面那个非零的值；
- 若 $r > n-r+1$，即 $n \le 2r-2$（非曲棍球员太少，不足以把他们隔开），概率为 0。

在临界情况 $n = 2r-1$，排列必然是 $HNHN\cdots NH$ 的模式，而公式给出 $\frac{r!\,(r-1)!}{(2r-1)!} = 1\big/\binom{2r-1}{r}$，即 H 恰好落在那 r 个相间位置的机会。

验算

- 取 $n = 10, r = 3$。则 (i) 给出 $\frac{6}{90} = \frac{1}{15}$；(ii) 给出 $\frac{8}{120} = \frac{1}{15}$；(iii) 给出 $\binom{8}{3}/\binom{10}{3} = \frac{56}{120} = \frac{7}{15}$。三者都与电脑枚举 120 个可能位置集合所得的结果吻合，而且各公式已用这个方法对所有 $2 \le r \le n \le 14$ 验证过（当 $n \le 7$ 时，更与全部 $n!$ 种排序对照过）。
- 当 $r = 2$ 时，第 (ii) 和第 (iii) 部分是互补事件（complementary events）（两名球员要么相邻，要么不相邻），而的确有 $\frac{2}{n} + \frac{(n-1)(n-2)}{n(n-1)} = \frac{2}{n} + \frac{n-2}{n} = 1$。

常见错误

- 只数出 $n - r - 1$ 个空隙（即严格位于非曲棍球员之间的那些），而忘记了队列的两端。
- 混用模型：分子数的是 H 的位置，分母却用学生的全部排序（$n!$）。选定一种观点，然后贯彻到底。
- 把答案非零的条件写成 $n \ge 2r$；其实 $n = 2r - 1$ 的情况（交替排列，首尾都是 H）同样可行。

STEP 之外

空隙论证就是“隔板法”的想法，是计数组合学（enumerative combinatorics）中的常用工具；同一个数目 $\binom{n-r+1}{r}$ 也回答了“$\{1, \ldots, n\}$ 有多少个 r 元子集不含两个连续整数”这个问题，而把它对 r 求和就得到斐波那契数（Fibonacci numbers）。把人的排序换成位置的集合这一步，让人初尝利用对称性来缩小样本空间（sample space）的滋味，这种手法在大学概率论中经常用到。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门）中，关于计数和概率的朴素定义（naive definition of probability）的章节写得十分出色。

详细解答

STEP I 1995，第 13 题

概率与统计 · 泊松稀疏化

STEP I 1995，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/95-s1-q13

关键思路

第 (i) 部分是泊松分布（Poisson distribution）的*稀疏化*（thinning）性质：先以癌细胞的真实数目为条件，则被标记的细胞数目服从二项分布（binomial distribution）。再对未知的真实数目求和，结果恰好化简成一个指数级数（exponential series）。至于第 (ii) 部分，“第二个标记落在第 k 块玻片上”这个事件，取决于头 $k-1$ 块玻片上合共有多少个标记（零个或一个），而第 (i) 部分连同独立泊松变量的可加性（additivity），正好告诉我们这个总数的分布。

解答

全文记 $q = 1-p$ 及 $\lambda = \mu p$。我们假设不同玻片上的癌细胞数目互相独立（independent）（这是对“随机”一词最自然的理解），并假设科学家每次能否察觉细胞亦互相独立，且与细胞数目独立。

第 (i) 部分

在一块玻片上，设 N 为癌细胞的数目，M 为她标记的数目，故 $N \sim \mathrm{Po}(\mu)$。给定 $N = n$，这 n 个细胞各自独立地以概率 p 被发现，所以 $M \mid N = n \sim \mathrm{B}(n, p)$。对于 $m = 0, 1, 2, \ldots$，由全概率公式（law of total probability）（只有 $n \geq m$ 的项有贡献），

$$\begin{aligned}\mathrm{P}(M = m) &= \sum_{n=m}^{\infty} \mathrm{e}^{-\mu} \frac{\mu^n}{n!} \binom{n}{m} p^m q^{n-m} = \mathrm{e}^{-\mu} \frac{(\mu p)^m}{m!} \sum_{n=m}^{\infty} \frac{(\mu q)^{n-m}}{(n-m)!} \\ &= \mathrm{e}^{-\mu} \frac{(\mu p)^m}{m!} \mathrm{e}^{\mu q} = \mathrm{e}^{-\mu p} \frac{(\mu p)^m}{m!},\end{aligned}$$

其中先用 $\binom{n}{m}/n! = 1/\big(m!\,(n-m)!\big)$，再以 $j = n-m$ 套用指数级数，最后用 $-\mu+\mu q = -\mu p$。因此 $M \sim \mathrm{Po}(\mu p)$，正如所求。

第 (ii) 部分

设 M_i 为第 i 块玻片上被标记的细胞数目。由第 (i) 部分及各玻片之间的独立性，$M_1, M_2, \ldots$ 是互相独立的 $\mathrm{Po}(\lambda)$ 变量。设 $S = M_1 + \cdots + M_{k-1}$ 为头 $k-1$ 块玻片上的标记总数；独立泊松变量之和仍服从泊松分布，其均值为各均值之和，所以 $S \sim \mathrm{Po}\big((k-1)\lambda\big)$，而且 S 与 M_k 独立。

> **想一想** 要令第二个标记落在第 k 块玻片上，究竟必须发生甚么事?

在第 k 块玻片之前，她最多只作了一个标记；而到第 k 块玻片完结时，她至少已作了两个标记。所以有两种互斥（disjoint）的情况：



$S = 0$ 且 $M_k \geq 2$，或 $S = 1$ 且 $M_k \geq 1$

（当 $k = 1$ 时，前面没有玻片，$S = 0$ 自动成立，以下公式依然适用。）利用 S 与 M_k 的独立性：

$$\mathrm{P}(S=0,\ M_k \geq 2) = \mathrm{e}^{-(k-1)\lambda}\big(1 - \mathrm{e}^{-\lambda} - \lambda \mathrm{e}^{-\lambda}\big),$$
$$\mathrm{P}(S=1,\ M_k \geq 1) = (k-1)\lambda\,\mathrm{e}^{-(k-1)\lambda}\big(1 - \mathrm{e}^{-\lambda}\big).$$

两式相加，

$$Q = \mathrm{e}^{-(k-1)\lambda}\Big\{\big(1+(k-1)\lambda\big)\big(1-\mathrm{e}^{-\lambda}\big) - \lambda\mathrm{e}^{-\lambda}\Big\}.$$

为了化成题目要求的形式，写 $1+(k-1)\lambda = (1+k\lambda) - \lambda$：

$$\begin{aligned}\big(1+(k-1)\lambda\big)\big(1-\mathrm{e}^{-\lambda}\big) - \lambda\mathrm{e}^{-\lambda} &= (1+k\lambda)\big(1-\mathrm{e}^{-\lambda}\big) - \lambda + \lambda\mathrm{e}^{-\lambda} - \lambda\mathrm{e}^{-\lambda} \\ &= (1+k\lambda)\big(1-\mathrm{e}^{-\lambda}\big) - \lambda.\end{aligned}$$

代入 $\lambda = \mu p$ 即得

$$Q = \mathrm{e}^{-\mu p(k-1)}\Big\{(1+k\mu p)\big(1-\mathrm{e}^{-\mu p}\big) - \mu p\Big\},$$

这正是题目所述的结果。

验算

▶ 当 $k = 1$ 时，公式给出 $1 - \mathrm{e}^{-\lambda} - \lambda\mathrm{e}^{-\lambda} = \mathrm{P}(M_1 \geq 2)$，理应如此。

- 对 $k = 1, 2, \ldots$ 求和，各概率之和应为 1（她最终会标记第二个细胞的概率为 1）；取若干个 λ 的值作数值求和，结果确认了这一点。
- 以 $\mu = 1.6$、$p = 0.45$（故 $\lambda = 0.72$）进行模拟，在 200 000 次试验中，第二个标记落在第 2 块玻片上的约占 26.0%；公式则给出 $Q \approx 0.2591$。

常见错误

- 遗漏了“第 k 块玻片之前有一个标记，第 k 块玻片上至少有一个”这种情况，或误把 $S = 1, M_k = 0$ 也计算在内。
- 逐块玻片地以“第一块有标记的玻片”来论证，这样做会变得十分繁复；头 $k - 1$ 块玻片上的总数 S 才是最有效率的切入对象。
- 没有说明在哪里用到独立性（各玻片之间的独立性，以及 S 与 M_k 之间的独立性）。

STEP 之外

稀疏化是泊松过程（Poisson process）的其中一面：若事件按泊松过程发生，而每个事件各自独立地以概率 p 被保留，则被保留的事件仍然构成一个泊松过程，其发生率（rate）乘以 p。这样看来，第 (ii) 部分就是“等待第二个事件发生所需时间”的离散版本，而在连续时间下，这个等待时间服从伽玛分布（gamma distribution）。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率论入门教材）详细讲述了泊松过程及这些等待时间的分布。

详细解答

STEP I 1995，第 14 题

概率与统计 • 比例的估计

STEP I 1995，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/95-s1-q14

关键思路

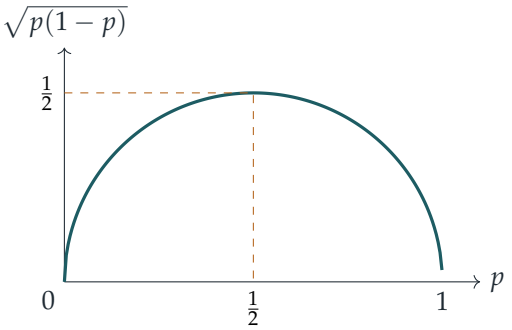
样本比例（sample proportion）就是一个二项（binomial）计数除以 n。当 n 很大时，它近似服从正态分布（normal distribution），平均值为 p，方差（variance）为 $p(1-p)/n$。麻烦在于，这个方差偏偏取决于我们不知道的那个 p。第 (i) 部分正好解决这个问题：它给出标准差（standard deviation）最坏情况下的值，所以按最坏情况选定的样本大小，无论 p 是多少都适用。第 (iii) 部分接着问：如果我们对 p 已经有粗略的认识，情况会有甚么不同。

解答

第 (i) 部分

用配方法（completing the square），设 $f(p) = p(1-p) = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2$。所以在 $[0,1]$ 上 $f(p) \le \frac{1}{4}$，等号仅在 $p = \frac{1}{2}$ 时成立。平方根是递增函数，因此

$$\max_{0 \le p \le 1} \sqrt{p(1-p)} = \sqrt{\tfrac{1}{4}} = \tfrac{1}{2}, \qquad \text{取于} p = \tfrac{1}{2}.$$



第 (ii) 部分

设 X 为样本中女性的人数。每个被抽中的人是女性的概率为 p。我们假设总体(population)很大，因此这 n 次抽选实际上可视为互相独立（从庞大的总体中不放回抽样，与放回抽样几乎没有分别）。于是

$$X \sim \mathrm{B}(n,p), \qquad \hat{p} = \frac{X}{n}, \qquad \mathrm{E}(\hat{p}) = p, \qquad \mathrm{Var}(\hat{p}) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

当 n 很大时，由中心极限定理（central limit theorem，即二项分布的正态近似）可得

$$\hat{p} \approx \mathrm{N}\left(p, \frac{p(1-p)}{n}\right), \qquad \text{所以} \qquad Z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \approx \mathrm{N}(0,1).$$

我们要求 $\mathrm{P}\left(|\hat{p}-p| > 0.01\right) < 0.01$，即

$$\mathrm{P}\left(|Z| > \frac{0.01\sqrt{n}}{\sqrt{p(1-p)}}\right) < 0.01.$$

由对称性，这 1% 平均分到两边尾部，每边 0.5%。由于 $\Phi(2.576) = 0.995$，我们需要

$$\frac{0.01\sqrt{n}}{\sqrt{p(1-p)}} > 2.576 \quad \Longleftrightarrow \quad \sqrt{n} > 257.6\sqrt{p(1-p)}.$$

? **想一想** 我们不知道 p，那么怎样从一个含有它的不等式去选定 n 呢?

这正是第 (i) 部分派上用场的地方。无论 p 是多少，都有 $\sqrt{p(1-p)} \leq \frac{1}{2}$。所以只要 $\sqrt{n} > 257.6 \times \frac{1}{2} = 128.8$，不等式便对任何 p 都成立。这个条件也是必要的，因为当 $p = \frac{1}{2}$ 时（而女性所占的比例确实接近这个值），上界可以取到。因此我们需要

$$n > 128.8^2 = 16\,589.44,$$

所求的样本大小为

$$\boxed{n = 16\,590} \qquad (\text{约}16\,600).$$

若采用更精确的百分点 $z = 2.5758$，则得 $n > 16\,587.2$。无论用哪个值，答案都是“约 16 600”，而第三位有效数字则视乎所用的统计表而定。

第 (iii) 部分

一般而言，所需的条件是

$$n > \left(\frac{2.576}{0.01}\right)^2 p(1-p) \approx 66\,360\, p(1-p).$$

16 590 这个数值来自最坏情况 $p = \frac{1}{2}$。如果我们知道 p 离 $\frac{1}{2}$ 相当远，便可以为 $p(1-p)$ 采用一个较小的上界。

(a) 左撇子。大约 10% 的人是左撇子。如果可以稳妥地假设 $p \le 0.15$，那么 $p(1-p) \le 0.15 \times 0.85 = 0.1275$, 因为 $p(1-p)$ 在 $[0, \frac{1}{2}]$ 上递增。由此得 $n > 8461$。如果恰好 $p = 0.1$, 则 $n > 66\,360 \times 0.09 \approx 5970$，即大约 6000。无论哪种情况，所需样本都远小于一无所知时所需的 16 590。正态近似仍然相当准确，因为 np 数以百计。

(b) 百万富翁。这时 p 非常小，也许介乎 0.001 至 0.01 之间。按字面计算，$p \le 0.01$ 给出 $p(1-p) \le 0.0099$ 及 $n > 657$。但这样做并没有抓住重点：

- 当 p 本身只有大约 0.001 时，0.01 的误差毫无用处：估计值 $\hat{p} = 0$ 与真实值的差距已经在 0.01 之内。我们真正想要的是相对准确度，例如误差不超过 1% 的 p。这样便需要 $n > 2.576^2(1-p)/(p \times 0.01^2)$，当 $p = 0.001$ 时约为 6.6×10^7。这比先前的数字大得多。
- 当 p 极小而 n 只有数百时，np 小于 1。这时二项分布非常偏斜，正态近似不再有效；以泊松近似（Poisson approximation）作为模型会更合适。
- 实际上，百万富翁很难接触得到，而且人们未必会如实说出自己的财富，所以能否做到随机抽样、受访者是否诚实作答，两者都成疑问。

因此，就左撇子而言，粗略知道 p 的大小可以让我们采用较小的样本。就百万富翁而言，任何合理的准确度要求都需要大得多的样本，或者要改用完全不同的抽样方法。

验算

- 当 $p = \frac{1}{2}$ 而 $n = 16\,590$ 时，正态近似给出 $\mathrm{P}(|\hat{p} - p| > 0.01) = 2\big(1 - \Phi(2.5762)\big) \approx 0.00999$，刚好略低于 1%。
- 把 $n = 16\,590$、$p = \frac{1}{2}$ 时的精确二项概率加起来，结果约为 0.0102。稍微超出的部分源于二项分布的离散性（这里的正态近似没有作连续性修正（continuity correction））。当 $p = 0.3$ 时，精确值约为 0.005，绰绰有余地符合要求，正如第 (i) 部分所预示。
- 当 $p = 0.1$ 而 $n = 6000$ 时，精确的二项概率约为 0.0098，印证了 (iii)(a) 中的数字。

常见错误

- 用了 1.96 或 2.326 而不是 2.576。这 1% 是分到两边尾部的，所以我们需要 $\Phi(z) = 0.995$。
- 把 n 留作以 p 表示的式子。题目要的是一个对任何 p 都适用的确定数值，这正是设置第 (i) 部分的原因。
- 把“$\sqrt{p(1-p)} \leq \frac{1}{2}$”理解成得出 $n > 128.8$。必须超过 128.8 的是 $\sqrt{n}$，而不是 n。

STEP 之外

民意调查中所报的“误差范围”（margin of error），背后就是这个标准计算。一项访问约 1000 人的民调，在 95% 置信水平下的误差大约为 ± 3 个百分点，用的正是第 (i) 部分的最坏情况论证。到了大学，这会引伸出置信区间（confidence interval），以及如何估计罕见比例的问题；在那种情况下，泊松方法和精确二项方法会取代正态近似。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门）对中心极限定理和这些近似方法有十分清晰易读的讲解。

详细解答

STEP II 1995，第 1 题

纯数学 · 等比级数的求导

STEP II 1995，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/95-s2-q1

关键思路

有限等比级数（geometric series）有闭合形式（closed form），而两个 x 的函数之间的恒等式（identity）可以求导，从而得出新的恒等式。对 $\sum x^k$ 求导会带出一个因子 k，这正好把等比级数变成 $\sum kx^{k-1}$；再代入 $x=-1$，便得到正负相间的符号。至于平方的情况，最简洁的做法是让答案的形式告诉我们 A 和 B 必须是甚么，然后用归纳法（induction）证明。

解答

第 (i) 部分

设 $S = 1 + x + x^2 + \cdots + x^n$。则

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

因为 $x, x^2, \ldots, x^n$ 每一个幂都以正号和负号各出现一次，互相抵消。若 $x \neq 1$，则 $1 - x \neq 0$，所以可以除以它：

$$1 + x + x^2 + \cdots + x^n = \frac{1 - x^{n+1}}{1 - x},$$

这正是所求。

第 (ii) 部分

(i) 的两边都是在 $x \neq 1$ 上可导（differentiable）的 x 的函数，而且在该处相等，所以它们的导数也相等。左边是

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(1 + x + \cdots + x^n\right) = 1 + 2x + 3x^2 + \cdots + nx^{n-1}.$$

右边则用商法则（quotient rule），得

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

现在取 $x=-1$，这是容许的，因为 $-1\neq 1$。左边变成 $1-2+3-4+\cdots+(-1)^{n-1}n$，因为 kx^{k-1} 这一项变成 $(-1)^{k-1}k$。右边变成

$$\frac{1-(n+1)(-1)^n+n(-1)^{n+1}}{4}=\frac{1-(-1)^n(2n+1)}{4},$$

当中用到 $(-1)^{n+1}=-(-1)^n$。

- 若 n 是偶数，则 $(-1)^n=1$，其值为 $\dfrac{1-2n-1}{4}=-\dfrac{n}{2}$。
- 若 n 是奇数，则 $(-1)^n=-1$，其值为 $\dfrac{1+2n+1}{4}=\dfrac{n+1}{2}$。

这就是我们要证明的结果。

第 (iii) 部分

设 $T_n=1^2-2^2+3^2-\cdots+(-1)^{n-1}n^2$，并假设 $T_n=(-1)^{n-1}f(n)$，其中 $f(n)=An^2+Bn$。

> **想一想** 在未知道公式是否成立之前，我们怎样才能求出 A 和 B?

利用我们对 T_n 唯一确知的事实：相邻两个和之差就是最后一项，即 $T_n-T_{n-1}=(-1)^{n-1}n^2$。由于 $T_{n-1}=(-1)^{n-2}f(n-1)=-(-1)^{n-1}f(n-1)$，上式可写成

$$(-1)^{n-1}\big(f(n)+f(n-1)\big)=(-1)^{n-1}n^2,\qquad\text{所以}\qquad f(n)+f(n-1)=n^2.$$

代入 $f(n)=An^2+Bn$，得

$$\begin{aligned}f(n)+f(n-1)&=An^2+Bn+A(n-1)^2+B(n-1)\\&=2An^2+(2B-2A)n+(A-B).\end{aligned}$$

要使这个式子对每一个 n 都等于 n^2，我们需要 $2A=1$ 和 $B=A$，即

$$A=B=\tfrac{1}{2},\qquad T_n=(-1)^{n-1}\frac{n(n+1)}{2}.$$

（这些值是唯一可能的：$n=1$ 给出 $A+B=1$，而 $n=2$ 给出 $-(4A+2B)=-3$，其唯一解是 $A=B=\frac{1}{2}$。）

用归纳法证明。当 $n=1$ 时，$T_1=1=(-1)^0\cdot\frac{1\cdot 2}{2}$。假设对某个 $n\geq 2$ 有 $T_{n-1}=$

$(-1)^{n-2}\frac{(n-1)n}{2}$。则

$$T_n = T_{n-1} + (-1)^{n-1}n^2 = (-1)^{n-1}\left(n^2 - \frac{n(n-1)}{2}\right)$$
$$= (-1)^{n-1}\frac{n(n+1)}{2},$$

这正是 n 时的公式。因此公式对所有 $n \geq 1$ 都成立，其中 $A = B = \frac{1}{2}$。

另一做法是把各项配对：$(2m-1)^2 - (2m)^2 = -(4m-1)$。当 $n = 2m$ 时，由此得 $T_n = -\sum_{j=1}^{m}(4j-1) = -m(2m+1) = -\frac{n(n+1)}{2}$；当 $n = 2m+1$ 时，再加上 $(2m+1)^2$，便得 $(2m+1)(m+1) = \frac{n(n+1)}{2}$。（也可以把 (ii) 的恒等式乘以 x 后再求导一次，但代数运算会繁复得多。）

验算

当 $n = 4$ 时：$1-2+3-4 = -2 = -\frac{4}{2}$，以及 $1-4+9-16 = -10 = -\frac{4\cdot5}{2}$。当 $n = 5$ 时：$1-2+3-4+5 = 3 = \frac{6}{2}$，以及 $-10+25 = 15 = \frac{5\cdot6}{2}$。用电脑对 n 直至 500 验算全部三个恒等式（例如 $T_{10} = -55$ 和 $T_{11} = 66$），结果完全吻合。

还可留意 $|T_n| = 1+2+\cdots+n$：平方的交错和（alternating sum），其大小竟与首 n 个正整数的普通和相同。这个巧合令人意外，而配对的论证正好解释了它。

常见错误

- 忘记了 $x = -1$ 之所以是合法的取值，全因 (i) 的恒等式对所有 $x \neq 1$ 成立；在 $x = 1$ 时，右边没有定义。
- 处理 $(-1)^{n+1}$ 时弄错符号：它等于 $-(-1)^n$，而不是 $(-1)^n$。
- 由 $n = 1$ 和 $n = 2$ 求出 A 和 B 后便停下来。这只说明了这些是唯一可能的值；要靠归纳法（或配对）才能说明公式确实对每一个 n 都成立。

STEP 之外

对 $\sum x^k$ 逐项求导,是一个威力强大的想法的有限版本:幂级数(power series)在其收敛半径(radius of convergence)之内可以逐项求导, 由此得出当 $|x| < 1$ 时 $\sum kx^{k-1} = (1-x)^{-2}$。在 $x = -1$ 时，无穷级数 $1-2+3-4+\cdots$ 发散（diverges），但公式却暗示其值为 $\frac{1}{4}$；事实上，(ii) 的公式可以写成 $\frac{1}{4} - (-1)^n\frac{2n+1}{4}$，所以部分和（partial sums）以 $\frac{1}{4}$ 为中心对称地摆动。为这类发散级数赋予一个值，正是阿贝尔求和法（Abel summation）背后的想法。Abbott 的 *Understanding Analysis*（实分析入门教材）对幂级数的逐项求导有严

谨的论述。

详细解答

STEP II 1995，第 2 题

纯数学 · 以递推法计算着色数目

> **STEP II 1995，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/95-s2-q2

本题在 *Beyond STEP* 第 9.3 节为篱笆桩着色中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/fence-posts.html

关键思路

把以红色开头的涂色方法，按最后一根柱子的颜色分类。直线上的每一种涂色方法都可以逐根柱子加上去而得出，而每根新柱子恰好有两种可用的颜色，所以总数很容易计算；要用递推关系（recurrence）处理的，是“以红色结尾”与“不以红色结尾”之间的划分。至于圆形，不过是一条两端颜色也必须不同的直线而已。

解答

为何 $r_{n+1} = s_n$

取 $n+1$ 根柱子的一种合规涂色方法，其中第一根和最后一根柱子都是红色。第 n 根柱子与第 $n+1$ 根柱子相邻，而后者是红色，所以第 n 根柱子不是红色。因此，删去第 $n+1$ 根柱子后，余下的是 n 根柱子的一种合规涂色方法，其中第一根柱子是红色而最后一根不是红色：这正是 s_n 种涂色方法之一。

反过来，给定 n 根柱子的 s_n 种涂色方法之一，由于第 n 根柱子不是红色，我们可以在末端加上一根红色柱子，从而得到一种计入 r_{n+1} 的涂色方法。这两个操作互为逆操作，所以它们给出一个一一对应（one-to-one correspondence），于是

$$r_{n+1} = s_n.$$

求 $r_n + s_n$

每一种第一根柱子为红色的合规涂色方法，最后一根柱子不是红色就是另一种颜色，所以 $r_n + s_n$ 就是第一根柱子为红色的所有合规涂色方法的总数。按次序为柱子涂色。第一根是红色（一种选择），其后每根柱子可以涂上除前一根柱子的颜色以外的任何颜色，因此不论之前怎样涂，每次都有 2 种选择。故此

$$r_n + s_n = 2^{n-1}.$$

求 $r_{n+1} + r_n$

代入 $s_n = r_{n+1}$，得

$$r_{n+1} + r_n = 2^{n-1} \qquad (n \geq 1).$$

归纳法

设 $P(n)$ 为命题 $r_n = \frac{1}{3}\left(2^{n-1} + 2(-1)^{n-1}\right)$。（这里用的是数学归纳法（induction）。）

基础情况（*base case*）。只有一根柱子并涂上红色时，这根柱子既是第一根也是最后一根，所以 $r_1 = 1$。公式给出 $\frac{1}{3}(1+2) = 1$，所以 $P(1)$ 成立。

归纳步骤（*inductive step*）。假设 $P(n)$ 对某个 $n \geq 1$ 成立。由上一个结果，

$$\begin{aligned} r_{n+1} &= 2^{n-1} - r_n \\ &= \frac{3 \cdot 2^{n-1} - 2^{n-1} - 2(-1)^{n-1}}{3} \\ &= \frac{2^n + 2(-1)^n}{3}, \end{aligned}$$

其中用到 $2 \cdot 2^{n-1} = 2^n$ 和 $-(-1)^{n-1} = (-1)^n$。这正是 $P(n+1)$。根据归纳法，$P(n)$ 对所有 $n \geq 1$ 成立。

> ? **想一想** 为甚么不由 $n = 2$ 开始归纳，或者使用两个基础情况?

递推关系 $r_{n+1} = 2^{n-1} - r_n$ 只用前一项来表示每一项，所以一个基础情况已经足够。若是涉及前两项的递推关系，就需要两个。

围成圆形的柱子

把柱子沿圆周编号为 $1, 2, \ldots, n$，于是第 n 根柱子现在也与第 1 根柱子相邻。一种合规的圆形涂色方法，正正就是直线 $1, 2, \ldots, n$ 的一种合规涂色方法，并且额外要求第 n 根柱子与第 1 根柱子颜色不同。

> ? **想一想** 如果第一根柱子不固定为红色，我们怎样数算涂色方法?

利用对称性（symmetry）。把两种颜色的名称对调，会把合规涂色方法变成合规涂色方法，所以第 1 根柱子不论选哪一种颜色，涂色方法的数目都相同。当第 1 根柱子为红色时，最后一根柱子与第一根颜色不同的直线涂色方法，正是 s_n 所计算的那些。所以圆形涂色方法的数目是

$$\begin{aligned} 3s_n &= 3\left(2^{n-1} - r_n\right) \\ &= 3 \cdot 2^{n-1} - 2^{n-1} - 2(-1)^{n-1} \\ &= 2^n + 2(-1)^n. \end{aligned}$$

这里的柱子是可以区分的（它们竖立在固定的位置），所以两种只相差一个旋转的涂色方法会分开计算。

验算

- 用手计算小的情况：$r_2 = 0$（两根相邻的柱子不能同为红色）和 $r_3 = 2$（RWR、RBR），与 $\frac{1}{3}(2-2) = 0$ 和 $\frac{1}{3}(4+2) = 2$ 吻合。
- 三根柱子围成三角形时必须用齐三种颜色，共有 $3! = 6$ 种方法，而 $2^3 + 2(-1)^3 = 6$。四根柱子围成正方形则有 $16 + 2 = 18$ 种。
- 以穷举法（brute force）数算 $n \leq 12$ 的所有涂色方法，结果与上面每一条公式都一致；例如 $r_{10} = 170$、$s_{10} = 342$，而 10 根柱子的圆形涂色方法有 1026 种。

常见错误

- 以 r_n 或 s_n 作为圆形的答案，忘记了第 1 根柱子不一定是红色；因子 3 正是用来补回另外两种颜色。
- 除以 n 来消去旋转。柱子的位置是固定的，所以旋转后的涂色方法是不同的涂法。

STEP 之外

圆形的答案 $2^n + 2(-1)^n$ 是圈图（cycle graph）C_n 的色多项式（chromatic polynomial）在 $k = 3$ 时的值；一般而言，使用 k 种颜色时，这个多项式是 $(k-1)^n + (-1)^n(k-1)$，而同样的“先排成直线，再把两端接合”论证就可以证明它。色多项式以及像 $r_{n+1} + r_n = 2^{n-1}$ 这样的递推关系，都是组合数学（combinatorics）或图论（graph theory）入门课程的标准内容。想做更多同类题目，可参看 Siklos 的 *Advanced Problems in Mathematics*（进阶数学难题集）。

详细解答

STEP II 1995，第 3 题

纯数学 · 逻辑推理与计数

> **STEP II 1995，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/95-s2-q3

关键思路

这其实是一道乔装了的逻辑谜题，而入手的方法是计数（counting）。每个赛段都派出 $5+4+3+2+1=15$ 分，所以五位车手的得分之和是一个已知的总数。Arouet 的得分非常接近满分，几乎足以确定他所有赛段的成绩。其后，Eluard 的赛果把 Eluard 的得分从上下两面夹住，而 Camus“四次名次相同”这个条件，只剩下几种情况需要检查。

解答

我们以 A, B, C, D, E 表示各车手的总得分。已知 $A > B > C > D > E$ 及 $A = 24$。

总分与 Arouet 的各赛段成绩

每个赛段派出 15 分，所以五个赛段合共

$$A+B+C+D+E = 75, \qquad \text{因此} \qquad B+C+D+E = 51.$$

Arouet 最多可得 $5 \times 5 = 25$ 分。得到 24 分即是刚好失去一分，所以 Arouet 在四个赛段得第一名，在一个赛段得第二名。Eluard 赢了第 5 赛段，所以 Arouet 的第二名必定出现在第 5 赛段，而 Arouet 赢了第 1–4 赛段。

为何 Eluard 得 11 分

下界（*lower bound*）。Eluard 在第 4 赛段得 3 分，在第 5 赛段得 5 分，而在其余三个赛段每段最少得 1 分。所以 $E \geq 3+5+3 = 11$。

上界（*upper bound*）。各人得分是互不相同的整数，且 $B > C > D > E$，所以 $D \geq E+1$、

$C \geq E+2$ 及 $B \geq E+3$。因此

$$51 = B+C+D+E \geq 4E+6, \qquad \text{所以} \qquad E \leq \tfrac{45}{4} < 12.$$

结合两个界，得 $E = 11$。下界中等号成立，亦告诉我们 Eluard 在第 1、2 及 3 赛段每段都是*最后一名*。

留给 Barthes、Camus 和 Diderot 的名次

我们现在知道 Arouet 和 Eluard 在每个赛段的名次，所以余下的名次由 Barthes、Camus 和 Diderot 三人瓜分：

赛段	Arouet、Eluard	留给 B、C、D 的名次	
1, 2, 3	第 1、第 5	第 2、第 3、第 4	（4、3、2 分）
4	第 1、第 3	第 2、第 4、第 5	（4、2、1 分）
5	第 2、第 1	第 3、第 4、第 5	（3、2、1 分）

此外 $B+C+D = 51-11 = 40$，且 $B > C > D > 11$，所以 $D \geq 12$ 及 $C \geq 13$。

? **想一想** Camus 有四次取得相同的名次。根据上表，那可以是哪些名次?

答案是第 2、第 3 或第 4 名。Camus 只能在第 4 和第 5 赛段取得第五名，那只有两个赛段，不是四个。我们逐一考虑这三种可能。

Camus 四次得第 2 名。第二名只在第 1–4 赛段可供取得，所以 Camus 在这四个赛段全部得第二名。于是 Barthes 在第 1–3 赛段每段最多得 3 分，在第 4 赛段最多得 2 分，在第 5 赛段最多得 3 分。所以 $B \leq 9+2+3 = 14$。但 $C \geq 4 \times 4 = 16 > B$，出现矛盾（contradiction）。

Camus 四次得第 4 名。第四名在每个赛段都值 2 分，所以 Camus 的最佳总分是 $4 \times 2 + 4 = 12$（余下的一个赛段得第二名）。这与 $C \geq 13$ 矛盾。

Camus 四次得第 3 名。第三名在第 1、2、3 及 5 赛段可供取得，每次值 3 分，所以这就是 Camus 的四次第三名。在第 4 赛段 Camus 得 $x \in \{4,2,1\}$ 分，而 $C = 12+x$。

- $x = 4$：$C = 16$，所以 $B \geq 17$ 及 $D = 40-B-C \leq 7$。这与 $D \geq 12$ 矛盾。
- $x = 2$：$C = 14$，所以 $B \geq 15$ 及 $D \leq 11$。同样出现矛盾。
- $x = 1$：$C = 13$，所以 $B+D = 27$，其中 $B \geq 14$ 及 $12 \leq D \leq 12$。因此 $D = 12$ 及 $B = 15$。

所以各人得分为

$$\boxed{B = 15, \quad C = 13, \quad D = 12} \qquad (\text{其中} A = 24,\ E = 11).$$

Barthes 在最后赛段的名次

> **想一想** 我们知道总分，但不知道每个赛段的赛果。现有的资料足以确定第 5 赛段的结果吗?

足够，方法是用奇偶性论证（parity argument）。Camus 在第 1、2、3 及 5 赛段得第三名，在第 4 赛段得第五名，于是 Barthes 和 Diderot 在第 1–4 赛段每段都瓜分值 4 分和 2 分的名次。在第 5 赛段，他们瓜分值 2 分和 1 分的名次（第 4 及第 5 名）。假设 Diderot 在第 1–4 赛段中有 k 段取得值 4 分的名次，并在第 5 赛段得 $s \in \{1, 2\}$ 分。那么

$$D = 4 \times 2 + 2k + s = 8 + 2k + s = 12.$$

所以 $s = 4 - 2k$ 是偶数，由此必有 $s = 2$（及 $k = 1$）。Diderot 在最后赛段得第四名，所以 **Barthes 在最后赛段得最后一名（第五名）**。

这样的赛事确实存在。例如，取第 1 赛段：A、D、C、B、E；第 2 及第 3 赛段：A、B、C、D、E；第 4 赛段：A、B、E、D、C；第 5 赛段：E、A、C、D、B。这样得出 $24, 15, 13, 12, 11$，符合要求。

验算

- 总分吻合：$24 + 15 + 13 + 12 + 11 = 75 = 5 \times 15$。
- 在上述例子的赛事中，Camus 在四个赛段得第三名，而 Eluard 在第 4 赛段得第三名，在第 5 赛段得第一名。
- 用电脑搜索全部 120^5 组可能的赛段赛果，只保留符合所有条件的，结果只找到得分 $(24, 15, 13, 12, 11)$。在每一种情况下，Barthes 在第 5 赛段都是第五名。（符合条件的赛事恰好有四个，视乎 Diderot 在第 1–4 赛段中哪一段得第二名而定。）

常见错误

- 只证明了 $E \geq 11$。题目问的是为何 Eluard 必定得 11 分，所以由 $B + C + D + E = 51$ 及得分互不相同而得出的上界是必不可少的。
- 假设 Camus 是在四个连续赛段取得相同名次，或者遗漏了第四名的情况。
- 求出了得分，却没有证明最后赛段的名次是必然的。只举出一个赛事例子并不足够；奇偶性论证说明了在每一个可能的赛事中，Barthes 在第 5 赛段都是最后一名。

STEP 之外

这个论证有三步：守恒定律（conservation law，即总分固定）、极端推理（extremal reasoning，24 分几乎是最高分），以及奇偶性论证。这三者在组合数学（combinatorics）中一再出现。从上下两面夹住一个量，直至它的值被迫确定下来，也是分析学（analysis）和数学奥林匹克解题的基本技巧。想做更多同类的谜题，可试试 Siklos 的 *Advanced Problems in Mathematics*。Engel 的 *Problem-Solving Strategies* 则有整章篇幅讨论不变量原理（invariance principle）和极端原理（extremal principle）。

详细解答

STEP II 1995，第 4 题

纯数学 · 沃利斯归约公式

> **STEP II 1995，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/95-s2-q4

本题在 *Beyond STEP* 第 4.4 节沃利斯、*Beta* 与二分之一的阶乘中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch04/wallis-beta.html

关键思路

利用分部积分法（integration by parts），可把 $\cos^2 t$ 化为 $1-\sin^2 t$，从而得到一条把 u_n 与 u_{n-2} 联系起来的归约公式（reduction formula）。把这条公式乘以 u_{n-1}，可见 nu_nu_{n-1} 不随 n 增大而改变，所以它等于 $n=1$ 时的值。余下的就是夹逼（squeeze）：由于各 u_n 递减，乘积 u_nu_{n-1} 介乎 u_n^2 与 u_{n-1}^2 之间。

解答

归约公式

设 $n \geq 2$。我们把积分写成 $u_n = \int_0^{\pi/2} \sin t \, \sin^{n-1} t \, \mathrm{d}t$，然后分部积分：对 $\sin^{n-1} t$ 求导，并把 $\sin t$ 积分为 $-\cos t$：

$$\begin{aligned} u_n &= \Big[-\cos t \, \sin^{n-1} t\Big]_0^{\pi/2} + (n-1)\int_0^{\pi/2} \cos^2 t \, \sin^{n-2} t \, \mathrm{d}t \\ &= 0 + (n-1)\int_0^{\pi/2} \big(1-\sin^2 t\big)\sin^{n-2} t \, \mathrm{d}t \\ &= (n-1)\big(u_{n-2} - u_n\big). \end{aligned}$$

边界项等于零，因为在上限有 $\cos\frac{\pi}{2} = 0$，在下限有 $\sin 0 = 0$（这里用到 $n-1 \geq 1$）。把含 u_n 的项合并，得

$$nu_n = (n-1)u_{n-2} \qquad (n \geq 2).$$

两边同乘以 u_{n-1}，得

$$nu_nu_{n-1} = (n-1)u_{n-1}u_{n-2} \qquad (n \geq 2),$$

这正是所需的结果。

恒定的乘积

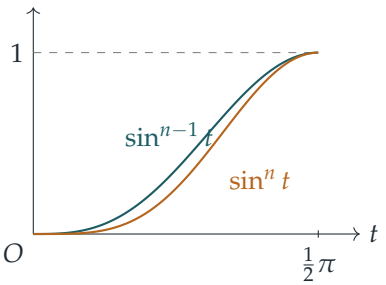
对 $n \geq 1$，记 $p_n = nu_nu_{n-1}$。上面的结果说明，对每个 $n \geq 2$ 都有 $p_n = p_{n-1}$，所以对每个 $n \geq 1$ 都有 $p_n = p_1$。现在

$$u_0 = \int_0^{\pi/2} 1\,\mathrm{d}t = \tfrac{1}{2}\pi, \qquad u_1 = \int_0^{\pi/2} \sin t\,\mathrm{d}t = \Big[-\cos t\Big]_0^{\pi/2} = 1,$$

所以 $p_1 = 1 \cdot u_1u_0 = \frac{1}{2}\pi$。因此

$$nu_nu_{n-1} = \tfrac{1}{2}\pi \qquad \text{对所有} n \geq 1.$$

草图与 $0 < u_n < u_{n-1}$



两条曲线在 $t = 0$ 时都由 0 出发，一直递增，并在 $t = \frac{1}{2}\pi$ 处达到 1（草图所示为 $n = 4$ 的情况；当 $n = 1$ 时，上方的曲线是直线 $y = 1$）。对 $0 < t < \frac{1}{2}\pi$，我们有 $0 < \sin t < 1$，所以

$$0 < \sin^n t = \sin t \cdot \sin^{n-1} t < \sin^{n-1} t.$$

因此，除了两个端点之外，$\sin^n t$ 的图像严格位于 $\sin^{n-1} t$ 的图像之下。两个函数都是连续（continuous）的，所以它们之下的面积满足

$$0 < u_n < u_{n-1} \qquad (n \geq 1).$$

想一想 两条曲线在端点处相接，为甚么不等式仍然是严格的?

差 $\sin^{n-1} t - \sin^n t$ 是连续的，永不为负，而且在（例如）$t = \frac{1}{4}\pi$ 处为正。一个连续函数若在某一点为正，则在该点附近的一个小区间上也为正，所以它围出的面积为正。在两个孤立点处相接，并不会改变这个结论。同样的论证可证明 $u_n > 0$。

双边界限

设 $n \geq 1$。由于 $u_n > 0$，把 $u_n < u_{n-1}$ 乘以 nu_n，得 $nu_n^2 < nu_n u_{n-1}$。由于 $u_{n-1} > 0$，乘以 nu_{n-1}，得 $nu_n u_{n-1} < nu_{n-1}^2$。中间的量是 $\frac{1}{2}\pi$，所以

$$nu_n^2 < \tfrac{1}{2}\pi < nu_{n-1}^2 \qquad (n \geq 1).$$

夹逼与极限

右边的不等式对每个 $n \geq 1$ 都成立，所以我们可以用 $n+1$ 代替 n：

$$\tfrac{1}{2}\pi < (n+1)u_n^2, \qquad \text{所以} \qquad nu_n^2 > \frac{n}{n+1}\cdot\tfrac{1}{2}\pi.$$

把它与左边的不等式结合，得

$$\left(\frac{n}{n+1}\right)\tfrac{1}{2}\pi < nu_n^2 < \tfrac{1}{2}\pi.$$

当 $n \to \infty$ 时，$\frac{n}{n+1} = 1 - \frac{1}{n+1} \to 1$，所以下界趋向 $\frac{1}{2}\pi$。上界本身就是 $\frac{1}{2}\pi$。根据夹逼定理（squeeze theorem，又称 sandwich theorem），

$$nu_n^2 \to \tfrac{1}{2}\pi \quad \text{当} n \to \infty.$$

验算

- 归约公式给出 $u_2 = \frac{1}{2}u_0 = \frac{1}{4}\pi$，这与 $\int_0^{\pi/2} \sin^2 t\,\mathrm{d}t = \frac{1}{4}\pi$ 吻合（$\sin^2$ 的平均值是 $\frac{1}{2}$）。于是 $2u_2u_1 = \frac{1}{2}\pi$，与预期一致。
- 以数值方法（辛普森法则，Simpson's rule）计算，$u_{10} \approx 0.38656$，即 $\frac{63}{512}\pi$。所以 $10u_{10}^2 \approx 1.4943$，介乎 $\frac{10}{11}\cdot\frac{1}{2}\pi \approx 1.4280$ 与 $\frac{1}{2}\pi \approx 1.5708$ 之间。
- 把递推关系（recursion）一直算到 $n = 10^5$，得 $nu_n^2 \approx 1.570788$，已经非常接近 $\frac{1}{2}\pi$。

常见错误

- 忘记归约公式需要 $n \geq 2$，所以常数必须用 u_0 和 u_1 来确定，而不能只用 u_1 和 u_2。
- 单凭 $\sin^n t \leq \sin^{n-1} t$ 就断言 $u_n < u_{n-1}$。这样只能得到 $\leq$；要得到严格不等式，被积函数的不等式必须在某个区间上严格成立。
- 试图直接由 $\frac{1}{2}\pi < nu_{n-1}^2$ 得出下界。这条不等式限制的是 u_{n-1}，而不是 u_n，所以要先把下标移位。

STEP 之外

这些积分称为沃利斯（Wallis）积分。把 u_{2m} 和 u_{2m+1} 写成乘积，再利用 $nu_n^2 \to \frac{1}{2}\pi$，便可得出沃利斯乘积 $\frac{\pi}{2} = \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdots$。同一个估计也确定了斯特灵（Stirling）公式 $n! \sim \sqrt{2\pi n}\,(n/\mathrm{e})^n$ 中的常数 $\sqrt{2\pi}$。这里所用的夹逼论证，是大学一年级分析学的标准工具。Abbott 的 *Understanding Analysis* 对此有细致的阐述。

详细解答

STEP II 1995，第 5 题

纯数学 · 最优化与凹性

STEP II 1995，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：

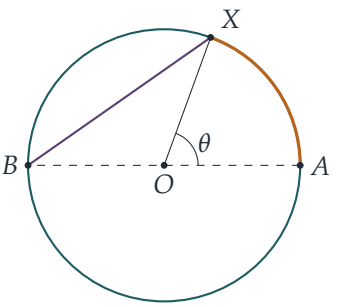
step.maths.org/questions/95-s2-q5

关键思路

选择 (i) 和 (iii) 正是 (ii) 这一族路线的两个极端：$\theta \to 0$ 表示直接游过对岸，$\theta \to \pi$ 表示全程跑步。因此我们把总时间写成 θ 在 $[0, \pi]$ 上的函数，再问它的最小值在哪里。出人意表的是，这个函数是凹函数（concave function），所以它唯一的驻点（stationary point）是极大点，而最佳路线永远在其中一端。

解答

取泳池的半径为 1 个长度单位，她的游泳速率为 1，于是她跑步的速率为 k。这样时间便以（半径）/（游泳速率）为单位，这并不影响任何比较。



三个时间

(i) 她游过直径 AB，长度为 2，需时 2。

(iii) 她跑过一个半圆，长度为 π，需时 π/k。

(ii) 她跑过弧 AX,长度为 θ(半径为 1),需时 θ/k。由于 $\angle AOB = \pi$,我们有 $\angle XOB = \pi - \theta$。三角形 XOB 是等腰三角形，其中 $OX = OB = 1$，所以弦（chord）长为

$$XB = 2\sin\tfrac{\pi-\theta}{2} = 2\cos\tfrac{\theta}{2}.$$

（另一个等价的做法：因为 AB 是直径，所以 $\angle AXB = \frac{\pi}{2}$；又由等腰三角形 AOX 得 $\angle XAB = \frac{\pi-\theta}{2}$，所以 $XB = 2\sin\angle XAB$。）她的总时间为

$$T(\theta) = \frac{\theta}{k} + 2\cos\frac{\theta}{2}, \qquad 0 < \theta < \pi.$$

这条公式在两端同样成立：$T(0) = 2$ 是 (i) 的时间，而 $T(\pi) = \pi/k$ 是 (iii) 的时间。因此，三个选择全部可以用 $[0, \pi]$ 上的同一个连续函数来描述。

T 在哪里最小?

求导数，得

$$T'(\theta) = \frac{1}{k} - \sin\frac{\theta}{2}, \qquad T''(\theta) = -\tfrac{1}{2}\cos\frac{\theta}{2}.$$

当 $0 < \theta < \pi$ 时，我们有 $0 < \frac{\theta}{2} < \frac{\pi}{2}$，所以 $\cos\frac{\theta}{2} > 0$，从而 $T''(\theta) < 0$：函数 T 在 $[0, \pi]$ 上是严格凹（strictly concave）的。

> **想一想** 若 $k > 1$，在 $\sin\frac{\theta}{2} = \frac{1}{k}$ 处有一个驻点。那不就是最佳的下水位置吗?

不是：由于在该处 $T'' < 0$，它是一个极大点，是最差的下水位置。更一般地说，闭区间上的严格凹函数，其图像严格位于连接两端函数值的弦之上，所以对每一个满足 $0 < \theta < \pi$ 的 θ，

$$T(\theta) > \Big(1 - \frac{\theta}{\pi}\Big)T(0) + \frac{\theta}{\pi}\,T(\pi) \geq \min\big(T(0),\, T(\pi)\big).$$

所以无论 k 取甚么值，把 X 选在中途的任何位置，都及不上 (i) 和 (iii) 之中较快的那一个。（若 $k \leq 1$，我们可以更直接地看出这一点：这时在 $(0, \pi)$ 上 $T'(\theta) > 0$，因为 $\sin\frac{\theta}{2} < 1 \leq \frac{1}{k}$，所以 T 是递增的，立即下水游泳便是最佳选择。）

比较 (i) 和 (iii)

余下只需比较 $T(0) = 2$ 与 $T(\pi) = \pi/k$:

- 若 $k < \frac{\pi}{2}$，则 $\pi/k > 2$，所以她应该**直接游过去**，即选择 (i)；
- 若 $k > \frac{\pi}{2}$，则 $\pi/k < 2$，所以她应该**全程绕着池边跑**，即选择 (iii)；
- 若 $k = \frac{\pi}{2}$，(i) 和 (iii) 所需时间相同，两者皆为最佳。

背后的物理原因很简单：半圆的长度是直径的 $\frac{\pi}{2}$ 倍。

最后，选择 (ii) 是她永远不会采用的：对每一个 $k > 0$ 和每一个 $\theta \in (0, \pi)$，它都严格地慢于 (i) 和 (iii) 之中较快的那一个。

验算

取 $k = 3$，跑步需时 $\pi/3 \approx 1.047$，胜过游泳的 2；驻点位于 $\theta = 2\arcsin\frac{1}{3} \approx 0.680$，该处 $T \approx 2.112$，比两端都慢，正是极大点应有的表现。用电脑在 e^{-4} 与 e^{4} 之间随机抽取数千个 k，并在细密的 θ 网格上计算 T，从未发现区间内部有任何值低于 $\min(2, \pi/k)$，同时确认了分界点在 $k = \pi/2 \approx 1.571$。

常见错误

- 令 $T'(\theta) = 0$ 后，没有检查驻点的性质便宣称所得结果是最优解：这里它其实是极大点。
- 用了弧 XB 而不是弦，或者把弦长算成 $2\sin\frac{\theta}{2}$（那是 AX，不是 XB）。
- 忘记了 $k \le 1$ 的情况，这时根本没有驻点。

STEP 之外

这里起作用的原理是：区间上的凹函数必在端点取得最小值；在更高维度，这变成以下事实：线性（或凹）目标函数在多边形上的最小值必在顶点取得，而这正是线性规划（linear programming）的基础。这道题也是斯涅尔定律（Snell's law）和费马（Fermat）最短时间原理的近亲，不过在那些情况下，答案通常是内点，因为时间函数反而是凸（convex）的。想做更多这类风格的题目，可以试试 Siklos 的 *Advanced Problems in Mathematics*（进阶数学难题集）。

详细解答

STEP II 1995，第 6 题

纯数学 · 单位根

STEP II 1995，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/95-s2-q6

关键思路

七次单位根（roots of unity）是 $1, \alpha, \alpha^2, \ldots, \alpha^6$，它们的和为零。把六个非实数的根分成两组，每组三个：$\{\alpha, \alpha^2, \alpha^4\}$ 和 $\{\alpha^3, \alpha^5, \alpha^6\}$。第二组正好是第一组各数的共轭复数（complex conjugate），所以两组的和互为共轭。因此它们的和与积都是实数，而第一部分的结果把这两个数变成二次式的系数。再取该根的实部和虚部，便得到所求的三角和。

解答

根与系数

若 u 和 v 是 $z^2 + az + b = 0$ 的根，则首一（monic）二次式 $z^2 + az + b$ 有因式 $z - u$ 和 $z - v$，所以对所有 z 都有

$$z^2 + az + b = (z - u)(z - v) = z^2 - (u + v)z + uv$$

。比较 z 的系数和常数项，得 $a = -u - v$ 及 $b = uv$，这正是所要证明的。

单位根

先提醒一点。试卷上印的是 $z^6 - 1 = 0$，但这是 $z^7 - 1 = 0$ 的误印。事实上，$\alpha^6 = \cos(12\pi/7) + \mathrm{i}\sin(12\pi/7) \neq 1$，而题目其余部分显然都与七次有关。我们以 $z^7 - 1 = 0$ 来解题。

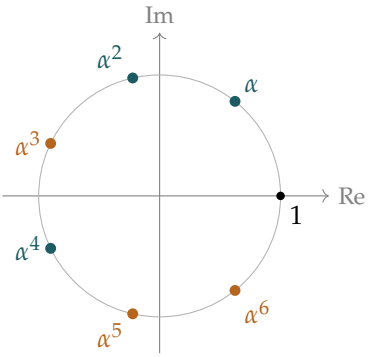
由棣美弗定理（de Moivre's theorem），

$$\alpha^7 = \cos 2\pi + \mathrm{i}\sin 2\pi = 1,$$

所以 α 是 $z^7-1=0$ 的根。更一般地，$\alpha^k=\cos(2k\pi/7)+\mathrm{i}\sin(2k\pi/7)$，而且 $(\alpha^k)^7=(\alpha^7)^k=1$。以下七个数

$$1,\ \alpha,\ \alpha^2,\ \alpha^3,\ \alpha^4,\ \alpha^5,\ \alpha^6$$

互不相同，因为它们的辐角（argument）$2k\pi/7$（$k=0,\dots,6$）在 $[0,2\pi)$ 内互不相同。7 次多项式最多有 7 个根，所以这些就是 $z^7-1=0$ 的全部根。它们位于单位圆上一个正七边形的顶点。



整个解答全靠两个事实。第一，各根之和为零：$z^7-1=(z-1)(z^6+z^5+\cdots+z+1)$，而 $\alpha\neq1$，所以

$$1+\alpha+\alpha^2+\alpha^3+\alpha^4+\alpha^5+\alpha^6=0. \qquad (*)$$

（等价的说法是：z^7-1 中 z^6 的系数为零。）第二，由于 $|\alpha|=1$，我们有 $\overline{\alpha^k}=\alpha^{-k}=\alpha^{7-k}$。所以 $\alpha^6=\overline{\alpha}$、$\alpha^5=\overline{\alpha^2}$ 及 $\alpha^3=\overline{\alpha^4}$。

求 A 和 B

设

$$p=\alpha+\alpha^2+\alpha^4,\qquad q=\alpha^3+\alpha^5+\alpha^6.$$

由上面的共轭配对，$q=\overline{p}$。这正是对另一个根最自然的猜想：一个实系数二次式若有非实数根 p，则 $\overline{p}$ 也必定是它的根。

> **想一想** 我们怎样知道 p 不是实数？假如它是实数，“另一个根”就不一定是 $\overline{p}$ 了。

p 的虚部是 $\sin\frac{2\pi}{7}+\sin\frac{4\pi}{7}+\sin\frac{8\pi}{7}$，我们会在最后证明它是正数。所以 p 不是实数，而两个根就是 p 和 $\overline{p}=q$。（无论如何，只要找到实数 A 和 B，使 p 和 q 为根，这个二次式便已确定。）

由 $(*)$，

$$p+q=\alpha+\alpha^2+\alpha^3+\alpha^4+\alpha^5+\alpha^6=-1.$$

至于乘积，先展开，再利用 $\alpha^7 = 1$ 把各幂次化简：

$$\begin{aligned} pq &= (\alpha + \alpha^2 + \alpha^4)(\alpha^3 + \alpha^5 + \alpha^6) \\ &= \alpha^4 + \alpha^6 + \alpha^7 + \alpha^5 + \alpha^7 + \alpha^8 + \alpha^7 + \alpha^9 + \alpha^{10} \\ &= \alpha^4 + \alpha^6 + 1 + \alpha^5 + 1 + \alpha + 1 + \alpha^2 + \alpha^3 \\ &= 3 + (\alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5 + \alpha^6) = 3 - 1 = 2. \end{aligned}$$

留意在这九项之中，除了三个 1 之外，$\alpha, \ldots, \alpha^6$ 每个都恰好出现一次。由第一部分，$A = -(p+q)$ 及 $B = pq$，所以

$$A = 1, \qquad B = 2,$$

而 p 和 q 是 $z^2 + z + 2 = 0$ 的根，即

$$z = \frac{-1 \pm \mathrm{i}\sqrt{7}}{2}.$$

三角和

取 $p = \alpha + \alpha^2 + \alpha^4$ 的实部和虚部，并利用 $\alpha^k = \cos(2k\pi/7) + \mathrm{i}\sin(2k\pi/7)$，得

$$\operatorname{Re} p = \cos\frac{2\pi}{7} + \cos\frac{4\pi}{7} + \cos\frac{8\pi}{7},$$
$$\operatorname{Im} p = \sin\frac{2\pi}{7} + \sin\frac{4\pi}{7} + \sin\frac{8\pi}{7}.$$

$z^2 + z + 2 = 0$ 的两个根的实部都是 $-\frac{1}{2}$，所以

$$\cos\frac{2\pi}{7} + \cos\frac{4\pi}{7} + \cos\frac{8\pi}{7} = -\frac{1}{2},$$

这正是所要证明的。（另一做法：由 $p + \overline{p} = -1$ 可直接得出 $2\operatorname{Re} p = -1$。）

虚部是 $\pm\frac{\sqrt{7}}{2}$，我们必须判断是哪一个。

> **想一想** 三个正弦值之中两个是正数，一个是负数。我们怎样才能肯定哪一方占优？

把负的那一项与其中一个正项比较。由于 $\frac{8\pi}{7} = \pi + \frac{\pi}{7}$，我们有 $\sin\frac{8\pi}{7} = -\sin\frac{\pi}{7}$。此外，$\sin\frac{4\pi}{7} = \sin\left(\pi - \frac{4\pi}{7}\right) = \sin\frac{3\pi}{7}$。正弦函数在 $[0, \frac{\pi}{2}]$ 上严格递增（strictly increasing），而 $0 < \frac{\pi}{7} < \frac{3\pi}{7} < \frac{\pi}{2}$，所以 $\sin\frac{3\pi}{7} > \sin\frac{\pi}{7}$。因此

$$\sin\frac{2\pi}{7} + \sin\frac{4\pi}{7} + \sin\frac{8\pi}{7} = \underbrace{\sin\frac{2\pi}{7}}_{>0} + \underbrace{\left(\sin\frac{3\pi}{7} - \sin\frac{\pi}{7}\right)}_{>0} > 0,$$

从而

$$\sin\frac{2\pi}{7} + \sin\frac{4\pi}{7} + \sin\frac{8\pi}{7} = \frac{\sqrt{7}}{2}.$$

这亦确认了 $p = \frac{-1+\mathrm{i}\sqrt{7}}{2}$ 不是实数，正如我们在上面所用到的。

验算

用计算机（以弧度计算），$\cos\frac{2\pi}{7}\approx 0.6235$、$\cos\frac{4\pi}{7}\approx -0.2225$ 及 $\cos\frac{8\pi}{7}\approx -0.9010$，三者相加得 -0.5000。三个正弦值约为 0.7818、0.9749 和 -0.4339，相加得 1.3229。这与 $\frac{\sqrt{7}}{2}\approx 1.3229$ 吻合。你也可以直接验证 $p^2+p+2=0$：取 $p=\frac{-1+\mathrm{i}\sqrt{7}}{2}$，则 $p^2=\frac{-6-2\mathrm{i}\sqrt{7}}{4}$，由此即得 $p^2+p+2=0$。

常见错误

- 照试卷所印使用 $z^6-1=0$。这样 α 便不是根，而“和为零”这个事实 $(*)$ 亦不成立。
- 展开 pq 时忘记用 $\alpha^7=1$ 化简幂次，或数错了等于 1 的那三项。
- 声称正弦之和是正数，“因为有两项是正数”。你需要作出比较，例如 $\sin\frac{3\pi}{7}>\sin\frac{\pi}{7}$。

STEP 之外

$\{1,2,4\}$ 与 $\{3,5,6\}$ 这个分法并非随意。$1,2,4$ 这三个数正是模 7 的非零平方数，而 $2p+1=\mathrm{i}\sqrt{7}$ 是一个*高斯和*（Gauss sum）。高斯（Gauss）证明了：对于形如 $4k+3$ 的质数 ℓ，α^r 对所有二次剩余（quadratic residue）r 求和，结果是 $\frac{-1+\mathrm{i}\sqrt{\ell}}{2}$。这是两套理论的第一步：一是关于哪些正多边形可以用尺规作图的理论（伽罗瓦理论，Galois theory），二是二次互反律（quadratic reciprocity）。Davenport 的 *The Higher Arithmetic* 讲述二次剩余，而 Needham 的 *Visual Complex Analysis* 则从几何角度看单位根。

详细解答

STEP II 1995，第 7 题

纯数学 · 希罗公式

> **STEP II 1995，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/95-s2-q7

关键思路

内切圆（incircle）是贯穿全题的主角。由圆心到各切点的半径都垂直于三角形的边，正好是把 ABC 铺满的三个小三角形的现成的高，由此便得 (i)。第 (ii) 部分是希罗公式（Heron's formula）的经典推导：写出 $\sin^2\alpha = 1-\cos^2\alpha$，用余弦定理（cosine rule）消去角度，然后利用平方差（difference of two squares）分解因式，前后两次。至于球体，则要看一个垂直截面。球体与洞口各边的接触点，原来正正就是内切圆与三边相切的点。

解答

第 (i) 部分

设内切圆与 BC、CA 和 AB 分别相切于 P、Q 和 R'。通过切点的半径垂直于切线（tangent），所以 $IP \perp BC$、$IQ \perp CA$ 及 $IR' \perp AB$，而且每一条的长度都是 r。因此三角形 BIC 的底为 a，垂直高度为 r，其余两个三角形亦同理：

$$[BIC] = \tfrac{1}{2}ar, \qquad [CIA] = \tfrac{1}{2}br, \qquad [AIB] = \tfrac{1}{2}cr.$$

由于 I 位于三角形之内，这三个三角形恰好拼成 ABC，互不重叠。相加得

$$\Delta = \tfrac{1}{2}r(a+b+c) = rs,$$

正是所求。

第 (ii) 部分

把 $\Delta = \tfrac{1}{2}bc\sin\alpha$ 平方，并利用 $\sin^2\alpha = 1-\cos^2\alpha$，

$$\Delta^2 = \tfrac{1}{4}b^2c^2(1-\cos^2\alpha) = \tfrac{1}{16}\left[4b^2c^2 - (2bc\cos\alpha)^2\right],$$

这就是第一个结果。方括号内是一个平方差：

$$16\Delta^2 = (2bc - 2bc\cos\alpha)(2bc + 2bc\cos\alpha).$$

根据余弦定理，$2bc\cos\alpha = b^2 + c^2 - a^2$。把它代入每个因式，

$$2bc - 2bc\cos\alpha = a^2 - (b^2 - 2bc + c^2) = a^2 - (b-c)^2,$$
$$2bc + 2bc\cos\alpha = (b^2 + 2bc + c^2) - a^2 = (b+c)^2 - a^2,$$

所以

$$\Delta^2 = \tfrac{1}{16}\left[a^2 - (b-c)^2\right]\left[(b+c)^2 - a^2\right],$$

即第二个结果。每个方括号又是一个平方差：

$$16\Delta^2 = (a-b+c)(a+b-c)(b+c-a)(b+c+a).$$

现在引入 s。由于 $a+b+c = 2s$，

$$b+c-a = 2(s-a), \qquad a-b+c = 2(s-b), \qquad a+b-c = 2(s-c),$$

所以该乘积为 $2s\cdot 2(s-a)\cdot 2(s-b)\cdot 2(s-c) = 16\,s(s-a)(s-b)(s-c)$。因此 $\Delta^2 = s(s-a)(s-b)(s-c)$。由于 $\Delta > 0$，我们取正平方根：

$$\Delta = \sqrt{s(s-a)(s-b)(s-c)}.$$

第 (iii) 部分

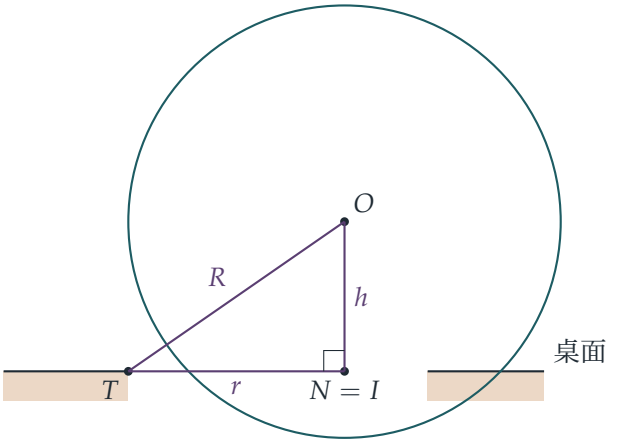
设 O 为球心，h 为球心高出桌面的高度。设 N 为由 O 到桌面所在平面的垂足（foot of the perpendicular），所以 $ON = h$。球体搁在洞口的三条边上。假设它与边 BC 相切于点 T。

> ? **想一想** 为甚么 N 必定是内心（incentre）I?

在球体与直线相切的点上，半径垂直于该直线，所以 $OT \perp BC$ 且 $OT = R$。此外，ON 垂直于整个桌面平面，所以 $ON \perp BC$。于是 BC 垂直于平面 ONT 内两条不平行的直线，因而垂直于该平面，当然也垂直于 NT。所以 NT 就是由 N 到 BC 的垂直距离，而在直角三角形 ONT 中运用毕氏定理（Pythagoras' theorem），得

$$NT^2 = R^2 - h^2.$$

对边 CA 和 AB 作同样的论证，可知 N 与三条边的距离同为 $\sqrt{R^2 - h^2}$。点 N 位于三角形之内（球体坐落在洞口之上），而三角形内与三边等距的点只有内心一个。所以 $N = I$，这个共同的距离就是 r，而各接触点正是内切圆与三边相切之处。这些点都在边本身之上，而不是在边的延线上，所以球体确实是搁在三条边上。



上图显示通过 O 和 T 的垂直截面；边 BC 穿过 T，垂直于纸面。在三角形 ONT 中运用毕氏定理，现在得到 $h^2 = R^2 - r^2$。由 (i) 和 (ii)，

$$r = \frac{\Delta}{s} = \frac{\sqrt{s(s-a)(s-b)(s-c)}}{s}, \qquad \text{所以} \qquad r^2 = \frac{(s-a)(s-b)(s-c)}{s}.$$

因此球心高出桌面的高度为

$$h = \sqrt{R^2 - \frac{(s-a)(s-b)(s-c)}{s}}$$

。这需要 $R > r$：不比内切圆大的球体会直接穿过洞口掉下去。

验算

取 3–4–5 三角形：$s = 6$，希罗公式给出 $\Delta = \sqrt{6 \cdot 3 \cdot 2 \cdot 1} = 6$，与 $\frac{1}{2} \cdot 3 \cdot 4$ 吻合。于是 $r = \Delta/s = 1$，你可以用直角三角形的公式 $r = \frac{1}{2}(3 + 4 - 5)$ 加以核实。当 $R = 2$ 时，公式给出 $h = \sqrt{4 - 1} = \sqrt{3} \approx 1.732$。当 $R \to r^+$ 时，$h \to 0$：球体不断下沉，直至其赤道与桌面齐平，并搁在内切圆上。作为更全面的检验，我们以数值方法测试了 18,816 个随机三角形。在每一个情况中，置于内心上方高度 h 处的球心，与三条边所在直线的距离都恰好是 R，而且最近的点都落在边之内。

常见错误

- 取 $h = \sqrt{R^2 - d^2}$，其中 d 是由 I 到某个顶点的距离。球体接触的是边，而不是角。
- 毫无理由地假设球心位于内心的正上方。这需要用上述的垂直论证，或者至少要给出清晰的对称理由。
- 在 (ii) 中把因式的配对弄错：$b + c - a = 2(s - a)$，而不是 $2(s - b)$ 或 $2(s - c)$。

STEP 之外

希罗公式属于一系列关于面积和体积的对称公式。婆罗摩笈多（Brahmagupta）公式对圆内接四边形（cyclic quadrilateral）起着同样的作用，而凯莱—门格行列式（Cayley–Menger determinant）则可由四面体的六条棱长求出其体积。“半径 ⊥ 切线，所以垂足就是内心”这一步，是三垂线定理（theorem of the three perpendiculars）的三维版本，在向量几何中很值得掌握。想找更多这类题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

STEP II 1995，第 8 题

纯数学 · 逻辑斯谛微分方程

STEP II 1995，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/95-s2-q8

本题在 *Beyond STEP* 第 8.3 节收获种群中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch08/harvesting.html

关键思路

这条增长定律是一条逻辑斯谛方程（logistic equation），属于可分离变量（separable）的类型：利用部分分式（partial fractions），积分便化为两个对数。至于第二部分，把新方程右边配方（completing the square）后可见，平移 $y = x - K + \alpha$ 会把新方程变成完全相同的逻辑斯谛方程，只是以 α 取代了 K。因此我们可以直接沿用第一部分的答案，无须再做任何积分。

解答

没有移除时的种群

我们取 $K > 0$（种群由 $x = K$ 开始，故必有 $K > 0$）。方程为

$$\frac{\mathrm{d}x}{\mathrm{d}t} = (2K - x)x.$$

由 $x = K$ 出发，变化率为 $K^2 > 0$。解永远不会到达 $x = 0$ 或 $x = 2K$，因为这两个常数函数本身就是解，而这条方程的解不可能互相穿越。所以自始至终 $0 < x < 2K$，我们可以除以 $(2K - x)x$。写成部分分式，

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

于是分离变量后得

$$\frac{1}{2K}\int\left(\frac{1}{x}+\frac{1}{2K-x}\right)\mathrm{d}x=\int\mathrm{d}t,$$
$$\frac{1}{2K}\Big(\ln x-\ln(2K-x)\Big)=t+C.$$

由于 x 和 $2K-x$ 均为正数，无须加上绝对值符号。在 $t=0$ 时，$x=2K-x=K$，故左边为 0，而 $C=0$。因此

$$\ln\frac{x}{2K-x}=2Kt\quad\Longrightarrow\quad x=(2K-x)\,\mathrm{e}^{2Kt}\quad\Longrightarrow\quad x=\frac{2K\mathrm{e}^{2Kt}}{1+\mathrm{e}^{2Kt}}=\frac{2K}{1+\mathrm{e}^{-2Kt}}.$$

为了配合题目要求的形式，写 $2=(1+\mathrm{e}^{-2Kt})+(1-\mathrm{e}^{-2Kt})$：

$$x(t)=\frac{K(1+\mathrm{e}^{-2Kt})+K(1-\mathrm{e}^{-2Kt})}{1+\mathrm{e}^{-2Kt}}=K+K\,\frac{1-\mathrm{e}^{-2Kt}}{1+\mathrm{e}^{-2Kt}},$$

这正是所需的结果。（等价地，$x=K+K\tanh Kt$。）

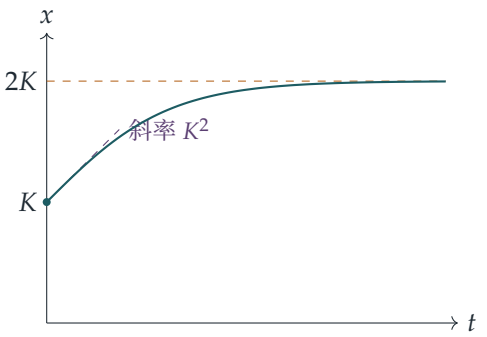
草图。 当 $t\geq 0$ 时，$\mathrm{e}^{-2Kt}\in(0,1]$，所以 $K\leq x<2K$。于是 $\dot{x}=(2K-x)x>0$，故 x 递增。把方程再求导一次，

$$\ddot{x}=(2K-2x)\,\dot{x}\leq 0\quad\text{因为}x\geq K,$$

所以图像是凹向下的（concave），在 $t=0$ 处的斜率为 K^2（该处 $\ddot{x}=0$：时间的原点正是整条 S 形逻辑斯谛曲线的拐点（point of inflection））。当 $t\to\infty$ 时，$\mathrm{e}^{-2Kt}\to 0$，所以

$$x(t)\to 2K.$$

种群数量在 $2K$ 处趋于平稳，从下方逼近它，但永远不会到达。



以速率 L 移除

每小时移除 L 微克，即是从变化率中减去 L：

$$\frac{\mathrm{d}x}{\mathrm{d}t}=(2K-x)x-L.$$

把右边配方：

$$(2K-x)x-L=-(x^2-2Kx)-L=K^2-L-(x-K)^2=\alpha^2-(x-K)^2,$$

其中 $\alpha=\sqrt{K^2-L}>0$，因为 $K^2>L$。

> ? **想一想** 为甚么 $y=x-K+\alpha$ 是合适的代换?

右边 $\alpha^2-(x-K)^2$ 在 $x=K\pm\alpha$ 处为零，这就是两个平衡（equilibrium）种群数量。平移 $K-\alpha$ 会把较低的平衡点移到 0，而原来的逻辑斯谛方程的较低平衡点也正在那里。由 $x-K=y-\alpha$ 及 $\dot{y}=\dot{x}$，

$$\frac{\mathrm{d}y}{\mathrm{d}t}=\alpha^2-(y-\alpha)^2=2\alpha y-y^2=(2\alpha-y)\,y.$$

这就是第一条方程，只是以 α 取代了 K。初始条件（initial condition）同样吻合：$x(0)=K$ 给出 $y(0)=\alpha$，正好对应原来的 $x(0)=K$。由于 $\alpha>0$，第一部分的论证可以一字不改地套用，于是

$$y(t)=\frac{2\alpha}{1+\mathrm{e}^{-2\alpha t}},\qquad \text{所以}\qquad x(t)=K-\alpha+\frac{2\alpha}{1+\mathrm{e}^{-2\alpha t}}.$$

当 $t\to\infty$ 时，$\mathrm{e}^{-2\alpha t}\to 0$，所以 $y\to 2\alpha$，而

$$x(t)\to K-\alpha+2\alpha=K+\alpha,$$

这正是所求。当 $L>0$ 时，$\alpha<K$，所以被收割的种群会稳定在较低的水平 $K+\alpha<2K$，一如所料。

验算

- 在 $t=0$ 时，两条公式都给出 $x=K$。设 $L=0$ 则 $\alpha=K$，第二条公式便化为 $2K/(1+\mathrm{e}^{-2Kt})$，即第一条公式。
- 对 $x=K+K\tanh Kt$ 求导，得 $\dot{x}=K^2\operatorname{sech}^2 Kt=K^2(1-\tanh^2 Kt)=(K-K\tanh Kt)(K+K\tanh Kt)=(2K-x)x$。
- 数值上，取 $K=1$，$x(1)=2/(1+\mathrm{e}^{-2})\approx 1.7616$；取 $K=1$ 及 $L=0.75$（故 $\alpha=\frac{1}{2}$），$x(1)\approx 1.2311$。两者都与直接以数值方法（龙格—库塔法，Runge–Kutta）求解微分方程所得的结果吻合，而且第二个解稳定在 $1.5=K+\alpha$。

常见错误

- 在部分分式中漏掉因子 $\frac{1}{2K}$，结果得出 e^{-t} 而不是 e^{-2Kt}。

- 在第二部分写成 $+L$，或者忘记积分常数（constant of integration）必须重新求出。这里的代换令 $y(0) = \alpha$，所以无须新的常数。

STEP 之外

第二部分让我们初次接触*平衡点的稳定性*（stability of equilibria）。方程 $\dot{x} = \alpha^2 - (x - K)^2$ 的平衡点在 $K \pm \alpha$：上面的一个吸引附近的解，下面的一个则排斥它们。只要看右边的正负号便可看出这一点，完全不用解方程。这种相线（phase line）方法，以及 $L > K^2$ 时的情况（没有平衡点，种群在有限时间内灭绝），可参阅 Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教科书）。

详细解答

STEP II 1995，第 9 题

力学 · 摩擦力与平衡

STEP II 1995，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：

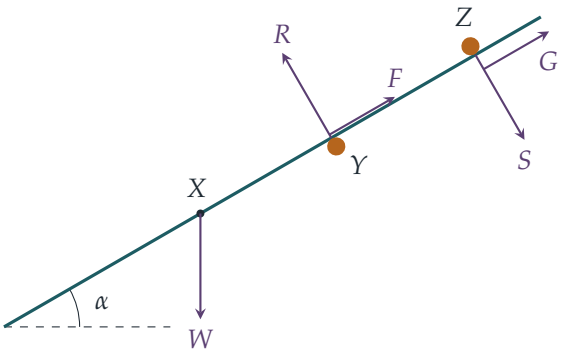
step.maths.org/questions/95-s2-q9

关键思路

杆沿着一条最大斜率线（line of greatest slope）放置，所以很自然会沿杆的方向和垂直于杆的方向分解（resolve）各力。杆搁在其上的那根横杆把它向上推，使它离开平面；杆从其下穿过的那根横杆则只能把它向下推。对 Y 取力矩（moment），便可知道杆的中心必须在哪里；然后令两个摩擦力（friction）都沿斜面向上并达到极限值（limiting value），就得到 μ 的最小值。

解答

我们把杆看成一根重量为 W 的均匀刚性杆（uniform rigid rod），把两根横杆看成固定的细杆，各自对杆施加一个垂直于杆的法向反作用力（normal reaction）和一个沿杆方向的摩擦力。由于杆与两根水平横杆都垂直，它便沿着两根横杆所在平面的一条最大斜率线放置，因此它与水平面的夹角为 α。设 R 为在 Y 处的反作用力（把杆推离平面，因为杆搁在这根横杆之上），S 为在 Z 处的反作用力（把杆推向平面，因为杆在这根横杆之下），其中 $R, S \geq 0$。设 F 和 G 分别为在 Y 和 Z 处的摩擦力，以沿斜面向上为正。



中心必须在哪里

沿垂直于杆的方向分解，并对 Y 取力矩（距离沿杆量度）：

$$R = S + W\cos\alpha,$$
$$S\,d = W\cos\alpha \cdot XY.$$

S 对 Y 的力矩，必须由重量分量 $W\cos\alpha$ 的力矩抵消，而这个分量与 S 一样指向平面。两个方向相同的平行力，只有分别作用在 Y 的两侧时，它们对 Y 的力矩才可以互相抵消。由于 Z 在 Y 的斜面上方，中心 X 必须在 Y 处或在 Y 的下方，于是

$$XZ = XY + YZ, \qquad \text{所以} \qquad l = XY + d \geq d.$$

（假如 X 在 Y 与 Z 之间，或在 Z 之外，力矩方程便会迫使 $S < 0$：在 Z 处的横杆就要把杆拉住，但它做不到，杆便会转动而离开这根横杆。）等式 $l = d$ 是临界情况，此时 X 位于 Y 而 $S = 0$。

想一想 试卷上写的是 $l \leq d$。为甚么我们得到相反的不等式?

试卷上的不等式（inequality）是印刷错误：题目本身接着取 $l = 2d$，而这只有在 $l \geq d$ 时才可能。上面的论证说明，平衡（equilibrium）所要求的正是 $l \geq d$，我们以下就用这个条件。

μ 的最小值

以 $XY = l - d$ 代入，两条方程给出

$$S = \frac{W\cos\alpha\,(l-d)}{d}, \qquad R = \frac{W\cos\alpha\,l}{d}, \qquad \text{所以} \qquad R + S = \frac{W\cos\alpha\,(2l-d)}{d}.$$

沿杆的方向分解，摩擦力必须承托重量沿斜面向下的分量：

$$F + G = W\sin\alpha.$$

摩擦定律要求 $|F| \le \mu R$ 及 $|G| \le \mu S$，所以 $F+G \le \mu(R+S)$。因此平衡需要

$$W\sin\alpha \le \mu\,\frac{W\cos\alpha\,(2l-d)}{d}, \qquad \text{即} \qquad \mu \ge \frac{d\tan\alpha}{2l-d}.$$

反过来，若 μ 等于这个值，则 $F=\mu R$ 和 $G=\mu S$（两者都沿斜面向上，两者都达到极限值）满足每一条方程，所以杆可以处于平衡。因此最小的可能值是

$$\mu_{\min} = \frac{d\tan\alpha}{2l-d}.$$

当 $l=2d$ 时，上式变成 $\mu_{\min} = \dfrac{d\tan\alpha}{3d} = \frac{1}{3}\tan\alpha$，所以 $\mu \ge \frac{1}{3}\tan\alpha$，正是所要证明的结果。

验算

- 当 $l=d$ 时，中心位于较低的横杆之上，$S=0$，我们得到 $\mu \ge \tan\alpha$：这正是单一质点（particle）静止在粗糙斜面上的熟悉条件。
- 当 $l\to\infty$ 时，反作用力无限增大，所以只需要极少的摩擦，而的确有 $\mu_{\min}\to 0$。
- 数值上，取 $\alpha=0.7$ 及 $l=2d$，公式给出 $0.28076\ldots$，即 $\frac{1}{3}\tan 0.7$。我们亦用电脑核对了对随机选取的点所列的力方程和力矩方程，并验证了平衡在略高于 $\mu_{\min}$ 时成立，在略低于它时则不成立。

常见错误

- 把两个反作用力取成同一方向。杆从其下穿过的那根横杆，只能把杆推向平面。
- 假设只有一根横杆处的摩擦达到极限。要令 μ 尽可能小，两根横杆都必须提供最大的摩擦力，方向沿斜面向上。

STEP 之外

这是一道带有不等式约束的静力学（statics）小问题：接触力必须满足一组线性方程，同时满足摩擦不等式，而 μ 的最小值就是可行集（feasible set）刚好变成非空的那一点。同样的结构，即线性方程加上线性不等式，正是线性规划（linear programming）以及工程学中摩擦接触分析的基础。想进一步了解刚体静力学和摩擦，可参阅 Kleppner and Kolenkow, *An Introduction to Mechanics*（一本大学力学入门教科书）。

详细解答

STEP II 1995，第 10 题

力学 · 碰撞与恢复系数

STEP II 1995，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/95-s2-q10

关键思路

每一次碰撞都是一个由两条方程组成的问题：动量守恒（conservation of momentum）和牛顿实验定律（Newton's experimental law）。依次处理两次碰撞，令第二个球的最终速度等于 u，然后解出 m_1；不等式则来自一个简单的事实：质量必须是正数。至于最后一部分，留意碰撞只取决于相对速度，所以只要在一个以速率 v 运动的参考系（frame）中观察这个运动，问题便化为第一部分。

解答

我们把各球视为质点（particle），在光滑桌面上沿同一直线运动，因此水平方向的力就只有球与球之间的冲量（impulse）。取第一个球的运动方向为正方向。



第一次碰撞

设第一次碰撞刚结束时，第一个球和第二个球的速度分别为 v_1 和 w。由动量守恒和牛顿定律（分离速率等于接近速率的 e 倍）可得

$$m_1 v_1 + m_2 w = m_1 u, \qquad w - v_1 = eu.$$

把 $v_1 = w - eu$ 代入第一条方程，得 $(m_1 + m_2)w = m_1(1+e)u$，所以

$$w = \frac{m_1(1+e)u}{m_1 + m_2}, \qquad v_1 = \frac{(m_1 - em_2)u}{m_1 + m_2}.$$

第二次碰撞

现在第二个球以速度 w 撞向静止的第三个球。这与刚才的情况完全相同，只是 m_1, m_2, u 换成了 m_2, m_3, w，所以第二个球碰撞后的速度是

$$\frac{(m_2 - em_3)w}{m_2 + m_3}.$$

题目告诉我们这个速度等于 u。代入 w 的表达式，并约去 $u \neq 0$，得

$$\frac{(m_2 - em_3)\, m_1(1+e)}{(m_2+m_3)(m_1+m_2)} = 1, \qquad \text{所以} \qquad m_1(1+e)(m_2 - em_3) = (m_1+m_2)(m_2+m_3).$$

把含 m_1 的项合并，

$$m_1\big[(1+e)m_2 - e(1+e)m_3 - m_2 - m_3\big] = m_2(m_2+m_3),$$

而方括号内的式子可化简为 $em_2 - m_3(1+e+e^2)$。因此

$$\boxed{m_1 = \frac{m_2(m_2+m_3)}{em_2 - m_3(1+e+e^2)}}.$$

想一想 我们手上只有一条等式，不等式从何而来?

答案来自物理：质量是正数。分子 $m_2(m_2+m_3)$ 是正数，而我们用来相除的那个方括号不可能是零（否则前一条等式就会变成 $0 = m_2(m_2+m_3)$）。所以 $m_1 > 0$ 迫使分母为正数：

$$em_2 - m_3(1+e+e^2) > 0, \qquad \text{即} \qquad m_2 e > m_3(1+e+e^2),$$

这正是所要证明的结果。

由于最后一部分要数算碰撞的次数，值得确认一下之后再没有其他事情发生。第二次碰撞后，第二个球以速度 u 运动，而第一个球的速度是 $v_1 = u - \frac{m_2(1+e)u}{m_1+m_2} < u$，所以第一个球永远不会再追上第二个球。第三个球则以相对速率 $ew > 0$ 远离第二个球。因此碰撞恰好发生两次。

最后一部分

现在三个球都有共同的速度 v，只有第一个球例外，它以 $u > v$ 的速度运动。我们在一个沿该直线以恒定速度 v 运动的参考系中观察这个运动。在这个参考系中：

- 第一个球以 $u - v > 0$ 的速度朝第二个球运动，而第二个球和第三个球都是静止的；
- 每次碰撞中动量仍然守恒，因为把每个速度都减去 v，动量方程两边的改变量相同，都是 $(m_a + m_b)v$；

- 牛顿定律保持不变，因为它只涉及速度之差。

所以在这个运动的参考系中，情况与第一部分完全一样，只是 u 换成了 $u-v$，而各质量仍满足同一关系式（该关系式根本不含 u）。根据第一部分，两次碰撞之后，第二个球相对于这个参考系的速度是 $u-v$。回到桌面的参考系，我们把 v 加回去：第二个球在两次碰撞之后的速度是

$$\boxed{u}.$$

条件 $u>v$ 保证了第一次碰撞确实会发生；至于 $v>0$ 这个假设，论证中其实用不着。

验算

取 $e=\frac{1}{2}$、$m_3=1$ 和 $m_2=4$；则 $em_2=2>\frac{7}{4}=m_3(1+e+e^2)$，而公式给出 $m_1=\frac{4\cdot5}{1/4}=80$。取 $u=7$，第一次碰撞后第二个球的速度是 $w=\frac{80\cdot\frac{3}{2}\cdot7}{84}=10$，第二次碰撞后它的速度则是 $\frac{(4-\frac{1}{2})\cdot10}{5}=7=u$，与预期相符。我们亦用电脑以分数作精确运算，对超过 10^4 组随机选取、符合条件的质量和 e 值模拟了碰撞过程，每一次都确认碰撞恰好发生两次、第二个球最终的速度是 u，而且把共同速度 $v<u$ 加到第二个球和第三个球上之后，最终速度同样是 u。

常见错误

- 把牛顿定律的方向写反，例如写成 $v_1-w=eu$；在前面的球，最终的速度必定不小于后面的球。
- 在最后一部分从头再计算各次碰撞，而没有利用转换参考系的方法，题目其实已明确提示可以这样做。
- 忘记在最后把 v 加回去，以 $u-v$ 作为答案。

STEP 之外

最后一部分是伽利略不变性（Galilean invariance）的一个例子：在每一个以恒定速度运动的参考系中，碰撞定律的形式都相同，所以你可以选择令问题最容易处理的参考系。在大学的力学课程中，最常用的选择是质心参考系（centre-of-mass frame），在这个参考系中总动量为零，而一次碰撞只不过是把各速度反向，再乘以 e。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）在处理碰撞时大量运用了这个想法。

详细解答

STEP II 1995，第 11 题

力学 · 线性空气阻力

STEP II 1995，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：

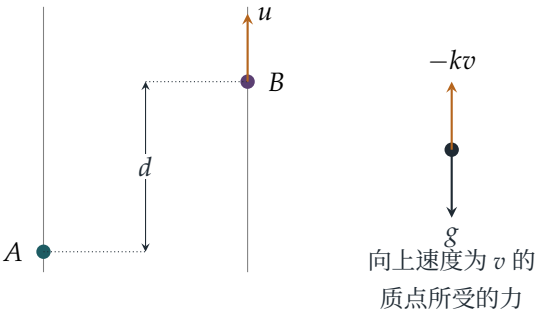
step.maths.org/questions/95-s2-q11

关键思路

由于阻力与速度成线性关系，运动方程（equation of motion）是线性的，而且无论质点正在上升还是下降，同一条方程都成立。因此把两条运动方程相减：重力项互相抵消，相对速度（relative velocity）纯粹按指数衰减。把它积分一次便直接得到两者的距离，完全不必知道任何一个质点身在何处。

解答

把每个物体视为在铅垂线上运动的单位质量质点（particle），受到重量 g 以及大小为 $k|v|$、方向与运动相反的阻力（resistance）作用。取向上为正方向，以 A 为原点，并设质点 1（由 A 出发）和质点 2（由 B 出发）在时间 t 的高度为 x_1、x_2，速度为 v_1、v_2。



运动方程

对于向上速度为 v 的质点，阻力为 $-kv$：若 $v>0$（上升），阻力向下；若 $v<0$（下降），阻力向上，两种情况都正好是“与运动方向相反”。因此，对单位质量应用牛顿第二定律（Newton's second law），每个质点都有

$$\frac{\mathrm{d}v_1}{\mathrm{d}t} = -g - kv_1, \qquad \frac{\mathrm{d}v_2}{\mathrm{d}t} = -g - kv_2,$$

对所有 $t \geq 0$ 均成立。

> **想一想** 质点 2 先上升，停下来，然后下降。难道不需要为每个阶段各写一条方程吗?

假如阻力与速率的平方成正比，那就需要：这时的力是 $-kv|v|$，当 v 变号时，它的形式会随之改变。但这里的力是 $-kv$，这单一的表达式在 v 反向时已自动反向，所以一条方程便涵盖整个运动。

相对运动

设 $w = v_2 - v_1$ 为质点 2 相对于质点 1 的速度。把两条方程相减，重量项互相抵消：

$$\frac{\mathrm{d}w}{\mathrm{d}t} = -kw, \qquad w(0) = u - 0 = u,$$

所以 $w = u\,\mathrm{e}^{-kt}$。现在设 $s = x_2 - x_1$，由于 B 在 A 上方高度 d 处，故 $s(0) = d$。于是 $\mathrm{d}s/\mathrm{d}t = w$，而

$$s = d + \int_0^t u\,\mathrm{e}^{-k\tau}\,\mathrm{d}\tau = d + \frac{u}{k}\bigl(1 - \mathrm{e}^{-kt}\bigr).$$

由于 $u \geq 0$ 及 $1 - \mathrm{e}^{-kt} \geq 0$，我们有 $s \geq d > 0$：质点 2 永远在质点 1 之上，两个质点永不相遇。因此它们在时间 t 的距离是

$$\boxed{d + \frac{u}{k}\bigl(1 - \mathrm{e}^{-kt}\bigr)}$$

当 $t \to \infty$ 时，因为 $k > 0$，所以 $\mathrm{e}^{-kt} \to 0$，距离趋向 $d + \dfrac{u}{k}$，正如题目所求。

个别质点的运动（如果你宁愿走远路）

分别解每条方程（它是一阶线性方程（first-order linear），也是可分离变量（separable）的方程），得 $v = -\frac{g}{k} + \bigl(v(0) + \frac{g}{k}\bigr)\mathrm{e}^{-kt}$，再积分一次，

$$x_1 = -\frac{gt}{k} + \frac{g}{k^2}\bigl(1 - \mathrm{e}^{-kt}\bigr),$$

$$x_2 = d - \frac{gt}{k} + \frac{1}{k}\Bigl(u + \frac{g}{k}\Bigr)\bigl(1 - \mathrm{e}^{-kt}\bigr).$$

两式相减，gt/k 和 g/k^2 的项互相抵消，得回同一个 s。两个质点的速度都趋近终端速度（terminal velocity）$-g/k$，这就是它们的距离最终稳定下来的原因。

验算

- 在 $t = 0$ 时距离为 d；而当 t 很小时，距离约为 $d + ut$，与没有阻力时的情况一样（那时两个质点在重力下以相同方式下坠）。
- 若 $u = 0$，两个质点同时由静止释放，运动完全相同，一直相距 d，与公式吻合。
- 单位：k 的单位是 s^{-1}（每单位质量、每单位速度的力），所以 u/k 是长度。
- 数值上，以龙格－库塔法（Runge–Kutta method）对随机选取的 k、u、d、t 逐步求解两条运动方程，结果与公式吻合至约 10^{-12}。取 $k = 0.5$、$u = 10$、$d = 3$，距离在 $t = 1$ 时为 10.87，在 $t = 20$ 时为 22.999，趋近 $d + u/k = 23$。

常见错误

- 把质点 2 的运动分成“上升”和“下降”两个阶段，并为阻力配上不同的正负号：这里并无必要，而且经常引致正负号出错。
- 当 v 已经是（负值的）向上速度时，仍把下降质点的阻力写成 $+kv$，这样便把正负号重复计算了。

STEP 之外

把方程相减这一招之所以奏效，是因为运动方程是线性的：两个解之差满足齐次方程（homogeneous equation）$\dot{w} = -kw$。这就是叠加原理（superposition principle），它贯穿整套线性微分方程（linear differential equations）理论，其中每个解都是“特解加余函数”（particular solution plus complementary function）。若阻力与速率的平方成正比（quadratic drag），这个原理便不再成立，两者的距离会有截然不同的表现。Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教材）阐述线性理论；Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教材）则讨论两种阻力下的运动。

详细解答

STEP II 1995，第 12 题

概率与统计 · 连续决斗

STEP II 1995，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/95-s2-q12

本题在 *Beyond STEP* 第 12.13 节三人决斗中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/truel.html

关键思路

在一场持续到有人被击中为止的决斗中，每当双方都掷不中，游戏便“重新开始”。因此所求的概率（probability）满足一条以它本身表示的简单方程，根本不必求无穷级数（infinite series）之和（当然你也可以这样做）。至于三人决斗，Algernon 第一掷的每个选择都很快变成二人决斗，而第一个结果正好用来处理这些情况。出人意表的是，Algernon 的最佳一着竟是谁也不击中。

解答

二人决斗

设 P 为 Percy 先掷时他胜出的概率。考虑第一对投掷：

- Percy 第一掷便击中（概率为 p），他立即胜出；
- Percy 掷不中，然后 Rodney 击中（概率为 $(1-p)r$），Percy 落败；
- 两人都掷不中（概率为 $(1-p)(1-r)$），于是回到起点，仍由 Percy 先掷。各次投掷互相独立（independent），所以由此开始 Percy 胜出的概率仍然是 P。

因此

$$P = p + (1-p)(1-r)P \quad\Longrightarrow\quad P\big(1-(1-p)(1-r)\big) = p \quad\Longrightarrow\quad P = \frac{p}{p+r-pr},$$

这是因为 $1-(1-p)(1-r)=p+r-pr$，而由于 $p>0$，此式为正数。这就是所要求的结果。

? **想一想** 游戏是否必定会结束，使得 Percy 落败确实等同于 Rodney 胜出?

是的。首 n 回合全部以双方掷不中告终的概率是 $\big((1-p)(1-r)\big)^n$，由于 $p>0$，它趋向 0。换一个做法，求等比级数（geometric series）$\sum_{k\geq 0} p\big((1-p)(1-r)\big)^k$ 之和，也会得到相同的答案。

在三人决斗中，A、B 和 C 击中的概率分别为 $\frac{2}{5}$、$\frac{3}{5}$ 和 1。由 A 先掷的 A、B 二人决斗会出现两次，所以我们现在先用上面的结果，取 $p=\frac{2}{5}$、$r=\frac{3}{5}$ 把它算出来：

$$\frac{\frac{2}{5}}{\frac{2}{5}+\frac{3}{5}-\frac{6}{25}}=\frac{\frac{2}{5}}{\frac{19}{25}}=\frac{10}{19}.$$

第 (i) 部分

若 Algernon 击中 Cuthbert，便只剩下 Algernon 和 Bertie，接着轮到 Bertie 掷，目标是 Algernon。只有当 Bertie 掷不中（概率为 $\frac{2}{5}$）时，Algernon 才能留在场上。之后便是由 A 先掷的 A、B 二人决斗，Algernon 胜出的概率为 $\frac{10}{19}$。所以

$$\mathrm{P}(\text{A 胜出})=\frac{2}{5}\cdot\frac{10}{19}=\frac{4}{19}.$$

第 (ii) 部分

若 Algernon 击中 Bertie, 接着轮到 Cuthbert 掷, 而他只剩 Algernon 一个目标。Cuthbert 从不失手，所以 Algernon 胜出的机会是 0。

第 (iii) 部分

若 Algernon 掷不中，三人仍然全部在场。接着轮到 Bertie 掷，他瞄准 Cuthbert。

- Bertie 有 $\frac{3}{5}$ 的概率击中 Cuthbert。这时剩下 Algernon 和 Bertie，轮到 Algernon 掷，Algernon 胜出的概率为 $\frac{10}{19}$。
- Bertie 有 $\frac{2}{5}$ 的概率掷不中。Cuthbert 随即瞄准 Bertie，并且必定击中他。这时剩下 Algernon 和 Cuthbert, 轮到 Algernon 掷。Algernon 这一掷必须击中, 否则 Cuthbert 便会击中他。所以他胜出的概率为 $\frac{2}{5}$。

因此

$$\mathrm{P}(\text{A 胜出})=\frac{3}{5}\cdot\frac{10}{19}+\frac{2}{5}\cdot\frac{2}{5}=\frac{6}{19}+\frac{4}{25}=\frac{150+76}{475}=\frac{226}{475}.$$

无论这次掷不中是怎样造成的，是故意还是瞄准后失手，结果都一样，因为 Bertie 和

Cuthbert 其后的打法并无分别。

Algernon 的最佳策略

> **想一想** Algernon 不能选择击中某人，只能选择瞄准某人。每个选择实际上给他带来甚么?

若他瞄准某人，他击中的概率为 $\frac{2}{5}$，掷不中的概率为 $\frac{3}{5}$，而掷不中便归入第 (iii) 部分的情况。所以：

$$\begin{aligned}
\text{瞄准 Cuthbert：}&\quad \frac{2}{5}\cdot\frac{4}{19}+\frac{3}{5}\cdot\frac{226}{475}=\frac{200+678}{2375}=\frac{878}{2375}\approx 0.370,\\
\text{瞄准 Bertie：}&\quad \frac{2}{5}\cdot 0+\frac{3}{5}\cdot\frac{226}{475}=\frac{678}{2375}\approx 0.285,\\
\text{故意掷不中：}&\quad \frac{226}{475}=\frac{1130}{2375}\approx 0.476.
\end{aligned}$$

即使必定击中 Cuthbert，也只能得到 $\frac{4}{19}=\frac{100}{475}$，比 $\frac{226}{475}$ 还要小。击中任何一个对手都只会降低他的胜算。因此 Algernon 应该**第一掷故意掷不中**。这样 Bertie 和 Cuthbert 便会互相攻击，而 Algernon 可以对幸存的一方先掷。采用这个策略，他胜出的概率为 $\frac{226}{475}$，正如题目所求。

验算

- 若 $r=0$，决斗公式给出 1；若 $p=1$，它也给出 1，两者都合乎预期。
- 三个选择的排序为 $0.476>0.370>0.285$。以蒙地卡罗模拟 (Monte Carlo simulation) 模拟整场三人决斗（每种策略 3×10^5 局），得出 0.476、0.369 和 0.284，与精确值吻合。
- 以第 (iii) 部分作合理性检查：$\frac{6}{19}\approx 0.316$ 与 $\frac{4}{25}=0.16$ 相加得 0.476。

常见错误

- 弄错第一掷之后的投掷次序。在第 (i) 部分，接着掷的是 Bertie，所以要待 Bertie 掷不中之后，决斗公式才适用。
- 把“瞄准 Cuthbert”当作“击中 Cuthbert”。第 (i)–(iii) 部分都是以结果为条件的，制订策略时必须分别以 $\frac{2}{5}$ 和 $\frac{3}{5}$ 为它们加权。

STEP 之外

把一个概率写成走了一步之后以它本身表示的式子，这种方法称为首步分析（first-step analysis）。它是研究马可夫链（Markov chain）的基本工具，可把到达概率（hitting probability）和期望到达时间（expected hitting time）表示为线性方程组的解。“三人决斗”（truel）是博弈论（game theory）中一道著名的难题，当中最弱的参与者往往置身事外才最有利。想进一步了解，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门），或 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论导引）。

详细解答

STEP II 1995，第 13 题

概率与统计 · 泊松分布的最优化

> **STEP II 1995，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/95-s2-q13

关键思路

尽管题库把它归入另一个类别，这其实是一道披着商业外衣的概率题：缺席乘客的人数是一个二项（binomial）计数，试验次数很多而成功概率极小，这正是泊松分布（Poisson distribution）登场的时候。之后，要对整数 k 求最优值，窍门不在微积分，而在一阶差分（first difference）$u_{k+1}-u_k$：只要多卖一张机票能令期望利润（expected profit）增加，就继续卖下去。一旦证明了这个差分是递减的，最佳的 k 就是它首次不再为正的地方。

解答

泊松模型

假设 $N+k$ 名持票人各自独立地以概率 ϵ 缺席。那么缺席人数 R 服从二项分布 $\mathrm{B}(N+k,\epsilon)$，其平均值为 $(N+k)\epsilon$。由于 k 与 N 相比非常小，这个平均值近似等于 $N\epsilon=\lambda$。试验次数很多而成功概率极小的二项分布，可以用平均值相同的泊松分布很好地近似：对固定的 r，当 $n\to\infty$ 而 $n\epsilon=\lambda$ 保持不变时，

$$\binom{n}{r}\epsilon^r(1-\epsilon)^{n-r}=\frac{n(n-1)\cdots(n-r+1)}{n^r}\cdot\frac{\lambda^r}{r!}\left(1-\frac{\lambda}{n}\right)^{n-r}\ \to\ \frac{\lambda^r}{r!}\,\mathrm{e}^{-\lambda},$$

因为第一个因子趋向 1，而 $(1-\lambda/n)^n\to\mathrm{e}^{-\lambda}$。所以近似地有 $R\sim\mathrm{Po}(\lambda)$。从现在起，我们把这个分布当作精确成立，并记

$$p_r=\mathrm{P}(R=r)=\frac{\lambda^r}{r!}\,\mathrm{e}^{-\lambda}.$$

利润

航空公司售出 $N+k$ 张机票，每张利润为 $£A$，所以无论发生甚么事，这部分都是 $A(N+k)$。若有 r 名乘客缺席，则有 $N+k-r$ 人到场，争夺 N 个座位。

- 若 $r \geq k$，到场的人最多只有 N 个，人人都可以登机：无人需要赔偿，而 $\max(0, k-r) = 0$。
- 若 $r < k$，则有 $k-r$ 人无法登机，每人令公司损失 $£B$，而 $\max(0, k-r) = k-r$。

两种情况下，以英镑计算的利润都是 $A(N+k) - B\max(0, k-r)$。

期望利润 $u_0, \ldots, u_3$

取期望值时，只有 $r = 0, 1, \ldots, k-1$ 对罚款有贡献，所以

$$u_k = A(N+k) - B\sum_{r=0}^{k-1}(k-r)\,p_r.$$

代入 $p_0 = \mathrm{e}^{-\lambda}$、$p_1 = \lambda\mathrm{e}^{-\lambda}$ 和 $p_2 = \frac{1}{2}\lambda^2\mathrm{e}^{-\lambda}$：

$$\begin{aligned}
u_0 &= AN,\\
u_1 &= A(N+1) - B\mathrm{e}^{-\lambda},\\
u_2 &= A(N+2) - B\mathrm{e}^{-\lambda}(2+\lambda),\\
u_3 &= A(N+3) - B\mathrm{e}^{-\lambda}\left(3 + 2\lambda + \tfrac{1}{2}\lambda^2\right).
\end{aligned}$$

差分 v_k

把一般公式相减，

$$\begin{aligned}
v_k = u_{k+1} - u_k &= A - B\left[\sum_{r=0}^{k}(k+1-r)\,p_r - \sum_{r=0}^{k-1}(k-r)\,p_r\right]\\
&= A - B\left[p_k + \sum_{r=0}^{k-1}\big((k+1-r) - (k-r)\big)p_r\right]\\
&= A - B\sum_{r=0}^{k} p_r,
\end{aligned}$$

其中第一个和式里 $r = k$ 的一项 $(k+1-k)p_k = p_k$，在第二个和式中没有对应的项。用文字来说，就是 $v_k = A - B\,\mathrm{P}(R \leq k)$。这是合理的：第 $(k+1)$ 张额外机票必定赚得 $£A$，而它恰好在该乘客最终被拒登机时才多花 $£B$，这种情况在 $R \leq k$ 时发生。

想一想 为甚么方括号内的式子会化简成单纯的概率之和？

对每个 $r \le k-1$，当 k 增加一时，p_r 的系数恰好增加一（在该结果下多了一名被拒登机的乘客），而结果 $r = k$ 则开始要付一笔赔偿。所以由 0 到 k 的每个 r 都各贡献一次 p_r。

于是

$$v_k - v_{k+1} = B\,p_{k+1} = B\,\frac{\lambda^{k+1}}{(k+1)!}\,\mathrm{e}^{-\lambda} > 0,$$

因为 $B > 0$ 且 $\lambda > 0$。所以对每个 k 都有 $v_k > v_{k+1}$：差分严格递减。

最后，$\sum_{r=0}^{k} p_r$ 是 $\mathrm{e}^{-\lambda}\sum_{r\ge 0}\lambda^r/r! = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = 1$ 的部分和（partial sum），所以当 $k \to \infty$ 时它趋向 1，而

$$v_k \to A - B \qquad \text{当} k \to \infty,$$

正如题目所求。由于 $B > A$，这个极限是负数。

给 Fly By Night 的建议

记 $F(k) = \mathrm{P}(R \le k) = \sum_{r=0}^{k} p_r$，则 $v_k = A - BF(k)$，而

$$u_{k+1} > u_k \iff v_k > 0 \iff F(k) < \frac{A}{B}.$$

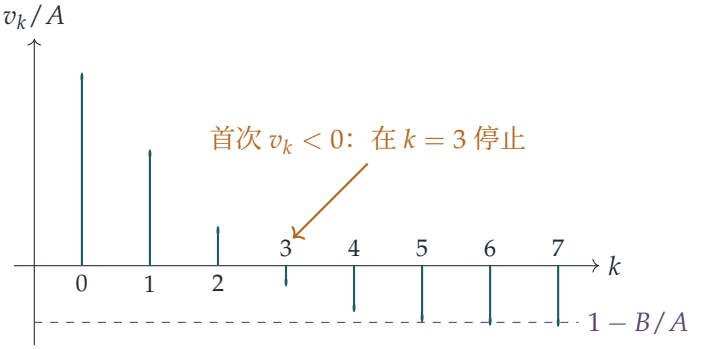
因为 v_k 严格递减而极限为负，数列 $v_0, v_1, v_2, \ldots$ 在某一点之前为正，其后为负（或有一次为零）。所以 u_k 先升后降，其最大值出现在第一个令下一张机票不再有利可图的 k：

$$\boxed{\text{选取最小的}k\text{ 使得}F(k) = \sum_{r=0}^{k}\frac{\lambda^r}{r!}\mathrm{e}^{-\lambda} \;\ge\; \frac{A}{B}.}$$

特别地，若 $\mathrm{e}^{-\lambda} \ge A/B$（乘客非常可靠，或罚款非常重），则 $v_0 \le 0$，航空公司完全不应超卖：$k = 0$。若恰好有 $F(k) = A/B$，则 $v_k = 0$，而 k 与 $k+1$ 给出相同的期望利润。由于 $A/B < 1$ 且 $F(k) \to 1$，这样的 k 必定存在。

所以建议是：只要最多有 k 人缺席的概率仍小于机票利润与赔偿额之比 A/B，就继续多卖机票；到了第一个令这个概率达到 A/B 的 k 就停止。

例如，取 $\lambda = 2$ 和 $B = 1.25A$（即 $A/B = 0.8$），可得 $F(0) \approx 0.135$、$F(1) \approx 0.406$、$F(2) \approx 0.677$ 和 $F(3) \approx 0.857$，所以最佳选择是 $k = 3$。下图显示这个情况下的 v_k/A。



验算

- 由一般公式，$u_1 - u_0 = A - B\mathrm{e}^{-\lambda} = A - Bp_0$ 及 $u_2 - u_1 = A - B\mathrm{e}^{-\lambda}(1+\lambda) = A - B(p_0 + p_1)$，与 v_0 和 v_1 吻合。
- 极限情况：若 $B \to A$，则 $A/B \to 1$，k 无止境地增大（拒载几乎没有代价）；若 $B \to \infty$，则 $A/B \to 0$，$k = 0$。
- 数值上，我们对数百组随机选取的 N, A, B, λ，直接由利润公式计算 u_k，并确认了 $u_0, \ldots, u_3$ 和 v_k 的公式、$v_k - v_{k+1} = Bp_{k+1}$，以及令 u_k 最大的 k 正是满足 $F(k) \geq A/B$ 的最小 k。当 $N = 400$、$k = 3$、$\epsilon = 0.01$ 时，有 $r = 3$ 人缺席的二项概率和泊松概率分别是 0.194 和 0.195。

常见错误

- 把 $r \geq k$ 的项也计入罚款的和式：这些项毫无贡献，因为此时每个到场的人都有座位。
- 把 u_k 对 k 求导。这个变量是整数，差分 v_k 才是合适的工具。
- 把答案写成 “$v_k = 0$”。一般而言没有任何 k 能令 v_k 恰好为零；答案是第一个令 $v_k \leq 0$ 的 k。

STEP 之外

“差分递减，所以在它变号之处停止”这个论证，是凹函数（concave function）在导数为零处取得最大值的离散版本。“当 $\mathrm{P}(R \leq k)$ 达到 A/B 时停止”这条法则，就是运筹学（operations research）中经典的报童（newsvendor，又称临界分位数 critical fractile）法则，而航空公司超卖机票正是它的一个标准应用。上文严格证明了的二项分布的泊松极限，是大学概率课程中最早遇到的极限定理（limit theorem）之一。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门教材）既涵盖泊松近似，也讨论像 $\max(0, k - R)$ 这类函数的期望值。

详细解答

STEP II 1995，第 14 题

概率与统计 · 假设检验

> **STEP II 1995，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/95-s2-q14

关键思路

把密度写成 $x^2\phi(x)$，其中 ϕ 是标准正态密度（standard normal density）。这样一来，我们需要的每一个积分都变成标准正态变量 Z 的矩（moment），题目中关于 $\mathrm{E}(Z^4)$ 的提示正是用在这里。要判断数据是否“动摇我们的信心”，我们要问：假如我们的假设（hypothesis）（$\mu = 0$）成立，出现至少这样极端的结果的概率有多大？如果这个概率极小，而把 μ 移动一下就能令数据变得平平无奇，那么数据就对我们的假设不利。

解答

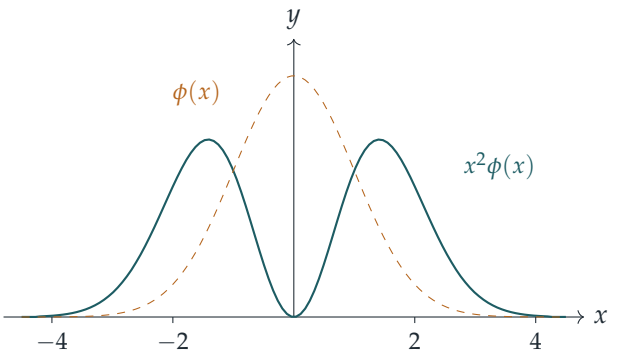
求 A

用分部积分法（integration by parts），取 $u = x$ 及 $\mathrm{d}v = x\mathrm{e}^{-x^2/2}\,\mathrm{d}x$，于是 $v = -\mathrm{e}^{-x^2/2}$：

$$\int_{-\infty}^{\infty} x^2\mathrm{e}^{-x^2/2}\,\mathrm{d}x = \Big[-x\mathrm{e}^{-x^2/2}\Big]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\,\mathrm{d}x = 0 + \sqrt{2\pi},$$

这里用了标准正态积分。总概率必须等于 1，所以

$$A = \frac{1}{\sqrt{2\pi}}, \qquad \text{即} \qquad \mathrm{f}(x) = x^2\phi(x), \quad \phi(x) = \tfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}.$$



这个密度是对称的，在 $x = \pm\sqrt{2}$ 处有两个峰，而在 $x = 0$ 处为零。由对称性可知 $\mathrm{E}(X) = 0$，并且

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^4\phi(x)\,\mathrm{d}x = \mathrm{E}(Z^4) = 3, \qquad \text{所以} \qquad \mathrm{Var}(X) = 3.$$

第 (i) 部分

在我们的假设之下，出现像 $a = 87.3$ 这样偏远的结果，可能性有多大？像之前一样用分部积分法（这次 $v = -\phi(x)$，因为 $\phi'(x) = -x\phi(x)$），

$$\mathrm{P}(X \geq a) = \int_a^{\infty} x \cdot x\phi(x)\,\mathrm{d}x = a\phi(a) + \mathrm{P}(Z \geq a).$$

当 $a > 0$ 时，我们还有 $\mathrm{P}(Z \geq a) = \int_a^{\infty} \phi(x)\,\mathrm{d}x \leq \int_a^{\infty} \frac{x}{a}\phi(x)\,\mathrm{d}x = \frac{\phi(a)}{a}$，因为在积分范围内 $x/a \geq 1$。因此，由对称性，

$$\mathrm{P}(|X| \geq 87.3) \leq 2\left(a + \frac{1}{a}\right)\phi(a) \approx 70\,\mathrm{e}^{-3810.6} \approx 10^{-1653}.$$

这个数值小得荒谬：假如我们的假设成立，我们基本上永远不会见到这样的数值。另一方面，如果 μ 大约是 85 或 89（使得 87.3 正好落在 $\mathrm{f}(x - \mu)$ 的其中一个峰上），那么 87.3 这个结果就完全平常。所以，单单一个 87.3 的结果，就应该彻底动摇我们的信心。唯一的出路是怀疑结果本身（例如记录出错），而明智的科学家在放弃假设之前，会先查证这一点。

? **想一想** 每一个个别数值的概率都是零，只凭一次观测就下判断，岂不是不公平？

重点不在于恰好取得 87.3 这个数值的概率，而在于取得一个离我们的假设所集中的位置至少这么远的数值的概率。同样地，在我们的假设之下，87.3 处的密度大约是某个合适的备择假设之下该处密度的 10^{-1650} 倍。无论从哪个角度看，观测结果极端到这个程度，一次就绰绰有余。

第 (ii) 部分

设 $X_1, \ldots, X_{1000}$ 为各个独立的结果，$\bar{X}$ 为它们的平均值。在我们的假设之下，每个 X_i 的平均值为 0，方差（variance）为 3，所以

$$\mathrm{E}(\bar{X}) = 0, \qquad \mathrm{Var}(\bar{X}) = \frac{3}{1000}, \qquad \mathrm{sd}(\bar{X}) = \sqrt{0.003} \approx 0.0548.$$

方差是有限的，而 1000 又够大，所以根据中心极限定理（central limit theorem），$\bar{X}$ 近似服从 $\mathrm{N}(0, 0.003)$，尽管 f 本身看起来一点也不像正态密度。观测所得的平均值与 0 相距

$$\frac{0.23}{0.0548} \approx 4.2$$

个标准差（standard deviation），而

$$\mathrm{P}(|\bar{X}| \geq 0.23) \approx \mathrm{P}(|Z| \geq 4.2) \approx 3 \times 10^{-5}.$$

所以，假如我们的假设成立，把整个实验重复十万次，样本平均值离 0 这么远的情况大约只会出现三次。而在 $\mu \approx 0.23$ 的备择假设之下，同样的数据就完全典型。这同样应该严重动摇我们的信心，只是不及 (i) 那样具压倒性。在任何常用的显著性水平（significance level）下，例如 1% 甚至 0.1%，我们都会否定 $\mu = 0$。

> **想一想** 我们可否改用切比雪夫不等式（Chebyshev's inequality），从而避开中心极限定理?

切比雪夫不等式给出 $\mathrm{P}(|\bar{X}| \geq 0.23) \leq 0.003/0.23^2 \approx 0.057$，当中不涉及任何近似。这个结果有参考价值，但不足以下定论，因为这样只能在大约 6% 的水平上排除如此极端的结果。中心极限定理给出的数字要精确得多；而且有 1000 个观测值，分布的尾部又像 $\mathrm{e}^{-x^2/2}$ 那样衰减，这个近似是可以信赖的。

验算

- 辛普森法则（Simpson's rule）给出 $\int x^2 \mathrm{e}^{-x^2/2}\,\mathrm{d}x = 2.506628\ldots = \sqrt{2\pi}$，以及 $\int x^2 \mathrm{f}(x)\,\mathrm{d}x = 3.000$。
- 尾部公式 $\mathrm{P}(|X| \geq a) = 2(a\phi(a) + \mathrm{P}(Z \geq a))$ 在 $a = 3$ 时与数值积分的结果吻合（两者皆为 0.029291）。
- $|X|$ 的密度，与由三个独立 $\mathrm{N}(0,1)$ 变量组成的向量的长度的密度相同。用这个方法模拟 X，每次取 1000 个数值求平均，重复 4000 次，结果约有 4.9% 的平均值落在 2 sd 之外，接近正态分布所预测的 4.6%，而且没有一个超出 0.23。

常见错误

- 因为指数部分看似标准正态分布，就取 $\mathrm{Var}(X) = 1$。额外的 x^2 把概率质量向外推，令方差变成 3。
- 在 (ii) 中，对单一个 X 而不是对 $\bar{X}$ 作正态检验，或者忘记把方差除以 1000。
- 在 (i) 中，辩称“一次观测证明不了甚么”。结果极端到这个程度，一次观测就足以说明问题。

STEP 之外

这是从基本原理出发看假设检验（hypothesis testing）。大学一年级的统计学课程会用 p 值（p-value）、似然比（likelihood ratio）（奈曼–皮尔逊引理，Neyman–Pearson lemma）和中心极限定理把它形式化。顺带一提，$|X|$ 服从自由度（degrees of freedom）为 3 的卡分布（chi distribution），亦即气体中分子速率的麦克斯韦–玻尔兹曼分布（Maxwell–Boltzmann distribution）。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）涵盖了这里用到的极限定理。

详细解答

STEP III 1995，第 1 题

纯数学 · 联立线性方程

STEP III 1995，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/95-s3-q1

关键思路

第二条和第三条方程的 y 项和 z 项完全相同，所以两式相减便可一次过消去两个未知数，只剩下一条关于 x 的方程。这道题的重点，大部分在于找出 a 和 b 取哪些值时，这一步或其后的某次除法会失效。系数的行列式（determinant）能预先告诉我们这些值在哪里，所以我们先把它算出来，然后逐一处理每个例外情况。

解答

把方程编号为

$$a^2x + ay + z = a^2 \tag{1}$$

$$ax + y + bz = 1 \tag{2}$$

$$a^2bx + y + bz = b. \tag{3}$$

哪里会出问题?

把系数矩阵（coefficient matrix）的行列式沿第一行展开，

$$\begin{vmatrix} a^2 & a & 1 \\ a & 1 & b \\ a^2b & 1 & b \end{vmatrix} = a^2(b-b) - a(ab - a^2b^2) + (a - a^2b)$$

$$= a^3b^2 - 2a^2b + a = a(ab-1)^2.$$

所以除非 $a = 0$ 或 $ab = 1$，否则方程组恰有一个解。在这两种情况下，方程组要么无解，要么有无限多个解，我们必须回到方程本身才能判断是哪一种。

一般情况：$a \neq 0$ 且 $ab \neq 1$

由 (3) 减去 (2)，便可消去 y 和 z：

$$(a^2b - a)x = b - 1, \qquad \text{即} \qquad a(ab-1)\,x = b - 1.$$

由于 $a \neq 0$ 且 $ab \neq 1$，我们可以相除，得

$$x = \frac{1-b}{a(1-ab)}.$$

接着，由 (1) 得 $z = a^2 - a^2x - ay$。把它代入 (2)：

$$ax + y + a^2b - a^2bx - aby = 1 \quad \Longrightarrow \quad (1-ab)\,y = (1 - a^2b) - ax(1-ab).$$

除以 $1 - ab \neq 0$，并利用 $ax = \dfrac{1-b}{1-ab}$，

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

最后，

$$\begin{aligned} z = a^2 - a(ax) - ay &= a^2 - \frac{a(1-b)}{1-ab} - \frac{ab(1-a^2)}{1-ab} \\ &= \frac{a^2 - a^3b - a + ab - ab + a^3b}{1-ab} = \frac{a(a-1)}{1-ab}. \end{aligned}$$

所以当 $a \neq 0$ 且 $ab \neq 1$ 时，唯一解为

$$\boxed{x = \frac{1-b}{a(1-ab)}, \qquad y = \frac{b(1-a^2)}{1-ab}, \qquad z = \frac{a(a-1)}{1-ab}.}$$

$a = 0$ 的情况

方程变成

$$z = 0, \qquad y + bz = 1, \qquad y + bz = b,$$

所以 $z = 0$，$y = 1$，而最后一条方程则表示 $1 = b$。因此：

- 若 $a = 0$ 且 $b \neq 1$，方程组**无解**；
- 若 $a = 0$ 且 $b = 1$，解为 $x = \lambda$，$y = 1$，$z = 0$，其中 λ 为任意实数：这些解构成一条直线，因为 x 已不再出现在任何一条方程之中。

$ab = 1$ 的情况（故 $a \neq 0$）

? **想一想** 当 $ab = 1$ 时，方程 $a(ab-1)x = b-1$ 的左方等于 0。这告诉我们甚么？

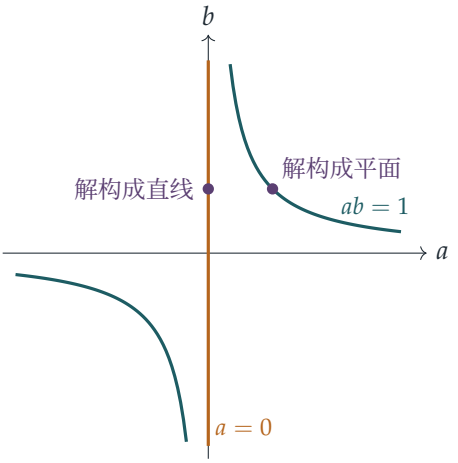
由 (3) 减去 (2) 这一步仍然成立；它现在变成 $0 = b-1$。所以若 $b \neq 1$，方程组便不相容（inconsistent）。在 $ab = 1$ 之下，条件 $b = 1$ 必然令 $a = 1$。因此：

- 若 $ab = 1$ 且 $a \neq 1$，方程组**无解**；
- 若 $a = b = 1$，三条方程全都变成 $x + y + z = 1$，所以解为 $x = \lambda, y = \mu, z = 1 - \lambda - \mu$，其中 λ, μ 为任意实数：这些解构成一整个平面。

这与行列式中的平方因子 $(ab-1)^2$ 吻合。在 $a = b = 1$ 时，矩阵的秩（rank）一直降至 1，所以有两个自由参数；而在 $a = 0$、$b = 1$ 时，矩阵的秩为 2，只有一个自由参数。

总结

下图所示为 (a, b) 平面。在直线 $a = 0$ 和双曲线（hyperbola）$ab = 1$ 以外，方程组有唯一解。在这两条线上，除了两个标示出来的点之外，方程组无解。



- $a \neq 0$，$ab \neq 1$：上面方框内的唯一解。
- $a = 0$，$b = 1$：$(x, y, z) = (\lambda, 1, 0)$，$\lambda \in \mathbb{R}$。
- $a = b = 1$：$(x, y, z) = (\lambda, \mu, 1 - \lambda - \mu)$，$\lambda, \mu \in \mathbb{R}$。
- $a = 0$ 而 $b \neq 1$，或 $ab = 1$ 而 $a \neq 1$：无解。

验算

取 $a=2$，$b=3$。公式给出 $x=\frac{1}{5}$，$y=\frac{9}{5}$，$z=-\frac{2}{5}$，而确实有 $4\cdot\frac{1}{5}+2\cdot\frac{9}{5}-\frac{2}{5}=4=a^2$，$\frac{2}{5}+\frac{9}{5}-\frac{6}{5}=1$ 以及 $\frac{12}{5}+\frac{9}{5}-\frac{6}{5}=3=b$。另一个快捷的测试是取 $b=1$，$a\neq 0,1$：公式给出 $x=0$，$y=1+a$，$z=-a$，而这些值满足全部三条方程。我们亦以精确的有理数运算，对 2000 组随机选取的 (a,b) 检验过行列式和一般公式。

常见错误

- 未有先把 $a=0$ 和 $ab=1$ 分开处理，便除以 $a(ab-1)$。题目说“对所有可能的值”，意思就是特殊情况正是这道题的重点所在。
- 声称“行列式为零，所以无解”。行列式为零只排除了唯一解的可能；在这道题中，有两组特殊的参数值会给出无限多个解。
- 忽略了 $ab=1$ 与 $b=1$ 同时成立必然令 $a=1$，所以双曲线上只有一个特殊点。

STEP 之外

这是一个小例子，说明矩阵的秩如何支配一个线性方程组（linear system）。方程组相容，当且仅当系数矩阵与增广矩阵（augmented matrix）的秩相同；这时所有解构成一个维数（dimension）为 $3-$ 秩 的集合。这就是为甚么我们得到的是单独一点、一条直线或一个平面。高斯消去法（Gaussian elimination）令这一切变得有系统，而 Strang 的 *Introduction to Linear Algebra*（线性代数入门）则以大量例题详细阐述这个课题。

详细解答

STEP III 1995，第 2 题

纯数学 · 归约公式

STEP III 1995，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/95-s3-q2

关键思路

因子 $\sqrt{x}$ 和 $\sqrt{a-x}$ 都提示我们使用代换（substitution）$x = a\sin^2\theta$，这样两者会同时化为 $\sqrt{a}\sin\theta$ 和 $\sqrt{a}\cos\theta$。（求 I_0 还有一条巧妙的几何捷径。）至于归约公式（reduction formula），可用分部积分法（integration by parts），使 x 的幂次降一。这时 $a-x$ 的幂次会升一，我们再利用 $a-x$ 把多出的一个因子拆开。这样便会重新得到 I_n 本身，于是可以合并同类项。

全题设 $a > 0$。

解答

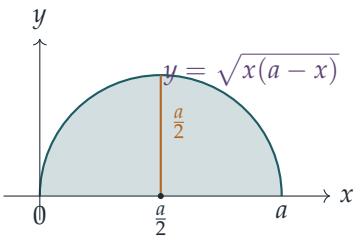
求 I_0

我们要求 $\mathrm{I}_0 = \int_0^a x^{1/2}(a-x)^{1/2}\,\mathrm{d}x$。设 $x = a\sin^2\theta$，则 $\mathrm{d}x = 2a\sin\theta\cos\theta\,\mathrm{d}\theta$。当 x 由 0 变到 a 时，θ 由 0 变到 $\frac{\pi}{2}$，而在这个范围内 $\sin\theta, \cos\theta \geq 0$。所以 $\sqrt{x} = \sqrt{a}\sin\theta$ 及 $\sqrt{a-x} = \sqrt{a}\cos\theta$，不会有正负号的问题。于是

$$\begin{aligned}\mathrm{I}_0 &= \int_0^{\pi/2} (\sqrt{a}\sin\theta)(\sqrt{a}\cos\theta)\,2a\sin\theta\cos\theta\,\mathrm{d}\theta = 2a^2\int_0^{\pi/2}\sin^2\theta\cos^2\theta\,\mathrm{d}\theta \\ &= \frac{a^2}{2}\int_0^{\pi/2}\sin^2 2\theta\,\mathrm{d}\theta = \frac{a^2}{4}\int_0^{\pi/2}(1-\cos 4\theta)\,\mathrm{d}\theta = \frac{a^2}{4}\cdot\frac{\pi}{2} = \frac{\pi a^2}{8},\end{aligned}$$

这正是所求的结果。这里用到了 $\sin\theta\cos\theta = \frac{1}{2}\sin 2\theta$ 和 $\sin^2 2\theta = \frac{1}{2}(1-\cos 4\theta)$，而 $\cos 4\theta$ 在 $[0, \frac{\pi}{2}]$ 上的积分为零。

一幅图。若 $y = \sqrt{x(a-x)}$，则 $y^2 = ax - x^2$，整理后得 $\left(x - \frac{a}{2}\right)^2 + y^2 = \left(\frac{a}{2}\right)^2$，其中 $y \geq 0$。所以 I_0 就是半径为 $\frac{a}{2}$ 的半圆的面积，即 $\frac{1}{2}\pi\left(\frac{a}{2}\right)^2 = \frac{\pi a^2}{8}$。



归约公式

取 $n \geq 1$。以 $u = x^{n+\frac{1}{2}}$ 和 $\mathrm{d}v = (a-x)^{1/2}\,\mathrm{d}x$ 作分部积分。则 $\mathrm{d}u = \left(n+\frac{1}{2}\right)x^{n-\frac{1}{2}}\,\mathrm{d}x$ 及 $v = -\frac{2}{3}(a-x)^{3/2}$。由此得

$$\mathrm{I}_n = \left[-\tfrac{2}{3}x^{n+\frac{1}{2}}(a-x)^{3/2}\right]_0^a + \frac{2}{3}\left(n+\frac{1}{2}\right)\int_0^a x^{n-\frac{1}{2}}(a-x)^{3/2}\,\mathrm{d}x.$$

边界项在两端都等于零：在 $x = a$ 处是因为有因子 $(a-x)^{3/2}$，在 $x = 0$ 处则是因为有 $x^{n+\frac{1}{2}}$。留意 $\frac{2}{3}\left(n+\frac{1}{2}\right) = \frac{2n+1}{3}$。

想一想 新的积分含有 $(a-x)^{3/2}$，并不是 I_{n-1} 的形式。我们该怎样处理它?

把它写成 $(a-x)^{3/2} = (a-x)\,(a-x)^{1/2}$，然后展开第一个因子：

$$\int_0^a x^{n-\frac{1}{2}}(a-x)^{3/2}\,\mathrm{d}x = a\int_0^a x^{n-\frac{1}{2}}(a-x)^{1/2}\,\mathrm{d}x - \int_0^a x^{n+\frac{1}{2}}(a-x)^{1/2}\,\mathrm{d}x = a\,\mathrm{I}_{n-1} - \mathrm{I}_n.$$

所以

$$\mathrm{I}_n = \frac{2n+1}{3}\left(a\,\mathrm{I}_{n-1} - \mathrm{I}_n\right) \quad\Longrightarrow\quad (2n+4)\,\mathrm{I}_n = (2n+1)\,a\,\mathrm{I}_{n-1},$$

这是两边乘以 3 再合并 I_n 各项后所得的（$3+2n+1 = 2n+4$）。这就是所求的结果。

计算 I_n

依次对 $k = n, n-1, \ldots, 1$ 应用 $\mathrm{I}_k = \dfrac{(2k+1)a}{2k+4}\,\mathrm{I}_{k-1}$，最后到达 I_0，得

$$\mathrm{I}_n = a^n\,\frac{\prod_{k=1}^n (2k+1)}{\prod_{k=1}^n (2k+4)}\cdot\frac{\pi a^2}{8}.$$

现在整理这两个乘积。分子是奇数的乘积 $3\cdot 5\cdots(2n+1)$。我们补上缺少的偶数 $2\cdot 4\cdots(2n) = 2^n n!$，得

$$\prod_{k=1}^{n}(2k+1) = \frac{(2n+1)!}{2^n\,n!}.$$

分母是 $\prod_{k=1}^n 2(k+2) = 2^n\cdot 3\cdot 4\cdots(n+2) = 2^n\,\dfrac{(n+2)!}{2}$。因此

$$\mathrm{I}_n = \frac{\pi a^{n+2}}{8}\cdot\frac{(2n+1)!}{2^n\,n!}\cdot\frac{2}{2^n\,(n+2)!} = \boxed{\frac{\pi\,a^{n+2}\,(2n+1)!}{4^{n+1}\,n!\,(n+2)!}}$$

当 $n = 0$ 时，这个公式同样给出正确的值 $\frac{\pi a^2}{8}$。

验算

- $n = 1$：递推关系（recurrence）给出 $6\mathrm{I}_1 = 3a\mathrm{I}_0$，所以 $\mathrm{I}_1 = \frac{\pi a^3}{16}$。公式则给出 $\pi a^3 \cdot 3!/(16 \cdot 1! \cdot 3!) = \frac{\pi a^3}{16}$，两者吻合。
- 对称性：把 x 换成 $a - x$，可见 $\mathrm{I}_1 = \int_0^a x^{1/2}(a-x)^{3/2}\,\mathrm{d}x$ 亦成立。把两个表达式相加，得 $2\mathrm{I}_1 = a\mathrm{I}_0$，这就独立地验证了 $\mathrm{I}_1 = \frac{\pi a^3}{16}$。
- 取 $a = 1$ 作数值计算，用森逊法则（Simpson's rule）或中点法则作数值积分，得 $\mathrm{I}_0 \approx 0.392699 = \pi/8$ 及 $\mathrm{I}_3 \approx 0.085903$。公式给出 $\pi \cdot 7!/(4^4 \cdot 3! \cdot 5!) = 35\pi/1280 \approx 0.085903$。

常见错误

- 弄错 v 的正负号：$\int (a-x)^{1/2}\,\mathrm{d}x = -\frac{2}{3}(a-x)^{3/2}$，其中的负号来自链式法则（chain rule）。
- 一心想直接得出 I_{n-1}，却没有留意到 I_n 在右边再次出现，必须把它合并。
- 在 $3 \cdot 4 \cdots (n+2) = (n+2)!/2$ 中漏掉因子 2。

STEP 之外

这些积分都是贝塔函数（Beta function）$B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}\,\mathrm{d}t$ 的特殊情况。代入 $x = at$ 可得 $\mathrm{I}_n = a^{n+2}B\left(n+\frac{3}{2},\frac{3}{2}\right)$。我们证明的递推关系，正是恒等式 $B(p+1,q) = \frac{p}{p+q}B(p,q)$ 的一个特例。在大学一年级的课程中，你会遇到公式 $B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$，它解释了 π 的来历：$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$。Siklos 的 *Advanced Problems in Mathematics* 一书中，还有更多这类风格的归约公式题目。

详细解答

STEP III 1995，第 3 题

纯数学 · 阻尼谐运动

STEP III 1995，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/95-s3-q3

关键思路

试代入 $x = \mathrm{e}^{\lambda t}$：微分方程（differential equation）便化为一条关于 λ 的二次方程，而题目的三种情况，正好对应其判别式（discriminant）的三种可能（两个相异实根、一个重根、复数根）。在有阻尼而仍然振荡的情况下，驻点（stationary point）每隔半个周期出现一次，时间间隔相等；每经过这样的一段时间，振荡因子只是变号，而指数因子则按相同的比例缩小。这就是相继极值之比为常数的原因。

解答

代入 $x = \mathrm{e}^{\lambda t}$，得辅助方程（auxiliary equation）

$$\lambda^2 + 2k\lambda + 1 = 0, \qquad \lambda = -k \pm \sqrt{k^2 - 1}.$$

以下 A 和 B 均为任意常数。

第 (i) 部分： $k > 1$

两根为相异实数，故

$$x = A\,\mathrm{e}^{(-k+\sqrt{k^2-1})\,t} + B\,\mathrm{e}^{(-k-\sqrt{k^2-1})\,t}.$$

（由于 $\sqrt{k^2-1} < k$，两个指数均为负数：这是过阻尼（overdamping），没有振荡。）

第 (ii) 部分： $k = 1$

辅助方程为 $(\lambda+1)^2 = 0$，有重根（repeated root）$\lambda = -1$。第二个独立解是 $t\,\mathrm{e}^{-t}$，故

$$x = (A + Bt)\,\mathrm{e}^{-t}.$$

（验证：若 $x = te^{-t}$，则 $x' = (1-t)e^{-t}$ 及 $x'' = (t-2)e^{-t}$，而 $(t-2)+2(1-t)+t = 0$。）

第 (iii) 部分： $0 < k < 1$

此时 $k^2 - 1 < 0$，故两根为 $\lambda = -k \pm i\omega$，其中

$$\omega = \sqrt{1-k^2} > 0.$$

取 $e^{(-k\pm i\omega)t}$ 的实组合，得

$$x = e^{-kt}\big(A\cos\omega t + B\sin\omega t\big).$$

相继的极值

由 $x(0) = 0$ 得 $A = 0$，故 $x = Be^{-kt}\sin\omega t$，且 $B \neq 0$（否则 $x \equiv 0$，没有极大值或极小值可言）。求导得

$$\frac{dx}{dt} = Be^{-kt}\big(\omega\cos\omega t - k\sin\omega t\big).$$

由于 $k^2 + \omega^2 = 1$，可设 $k = \sin\varphi$、$\omega = \cos\varphi$，其中 $0 < \varphi < \frac{\pi}{2}$，于是

$$\omega\cos\omega t - k\sin\omega t = \cos(\omega t + \varphi).$$

因此 $\dfrac{dx}{dt} = Be^{-kt}\cos(\omega t + \varphi)$，它等于零当且仅当 $\omega t + \varphi$ 是 $\frac{\pi}{2}$ 的奇数倍，并且在每个这样的时刻都会变号。所以（在 $t > 0$ 时）驻点位于

$$t_n = t_1 + \frac{(n-1)\pi}{\omega}, \qquad n = 1, 2, 3, \ldots,$$

彼此相隔 π/ω，间距相等；又因为 dx/dt 在每个驻点都变号，它们确实是极大值与极小值交替出现。

想一想 不必明确求出 t_1，怎样比较 $x(t_{n+1})$ 与 $x(t_n)$？

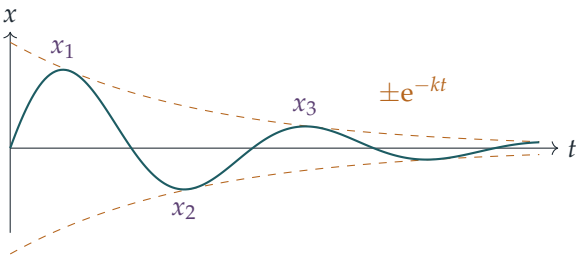
时间推前 π/ω，ωt 便增加 π，而 $\sin(\theta + \pi) = -\sin\theta$。所以

$$x_{n+1} = Be^{-k(t_n+\pi/\omega)}\sin(\omega t_n + \pi) = -e^{-k\pi/\omega}\,Be^{-kt_n}\sin\omega t_n = -e^{-k\pi/\omega}\,x_n.$$

此外 $x_n \neq 0$，因为若 $\sin\omega t_n = 0$，则必有 $\cos\omega t_n = \pm 1$，于是 $dx/dt = \pm B\omega e^{-kt_n} \neq 0$。因此

$$\left|\frac{x_{n+1}}{x_n}\right| = \alpha, \qquad \text{其中}\alpha = e^{-k\pi/\omega} = \exp\left(-\frac{k\pi}{\sqrt{1-k^2}}\right),$$

此值与 n 无关，正是所要证明的结果。（式中的负号印证了极大值与极小值交替出现。）



k^2 的公式

取对数，得 $\ln\alpha = -\dfrac{k\pi}{\sqrt{1-k^2}}$，故

$$(\ln\alpha)^2 = \frac{k^2\pi^2}{1-k^2} \implies (\ln\alpha)^2 - k^2(\ln\alpha)^2 = k^2\pi^2 \implies k^2 = \frac{(\ln\alpha)^2}{\pi^2 + (\ln\alpha)^2},$$

这就是所求的结果。

验算

- 当 $k \to 0$（没有阻尼）时，$\alpha \to 1$ 而 $\ln\alpha \to 0$，公式给出 $k^2 \to 0$，与预期相符：各个极值的大小全都相同。
- 当 $k \to 1^-$ 时，$\alpha \to 0$ 而 $(\ln\alpha)^2 \to \infty$，故 $k^2 \to 1$：在临界阻尼（critical damping）下，运动不再往回摆动。
- 数值上，取 $k = 0.5$ 及 $x(0) = 0$，把方程逐步向前数值求解，得出相继极值之比为 0.163034，与 $\mathrm{e}^{-0.5\pi/\sqrt{0.75}} \approx 0.163034$ 吻合；而且 $(\ln\alpha)^2/(\pi^2 + (\ln\alpha)^2) = 0.25 = k^2$。

常见错误

- 把一个极大值与下一个极小值之间的时间当作一整个周期 $2\pi/\omega$，而不是半个周期 π/ω；这样会得出 α^2 而不是 α。
- 以为极值出现在 $\sin\omega t = \pm 1$ 之处。其实不然：衰减因子令极值提早出现，移到 $\tan\omega t = \omega/k$ 之处。上面的论证自始至终不需要知道极值的确切位置，只需要知道它们的间距。

STEP 之外

$-\ln\alpha$ 这个量称为*对数衰减率*（logarithmic decrement）：工程师从记录下来的振荡中量度它，再用本题的公式把它换算成阻尼 k。Boyce 与 DiPrima 所著的 *Elementary Differential Equations*（初等微分方程教科书）详细讨论了过阻尼、临界阻尼和欠阻尼三种情况。

详细解答

STEP III 1995，第 4 题

纯数学 · 三角级数求和

STEP III 1995，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/95-s3-q4

关键思路

余弦之和是复指数（complex exponential）之和的实部（real part），而 $\sum e^{\mathrm{i}k\theta}$ 是一个等比级数（geometric series）。先把它求和，再从分子和分母各抽出一个“半角”（half-angle）因子 $e^{\mathrm{i}\theta/2}$，就能把难以处理的 $e^{\mathrm{i}m\theta}-1$ 项化为正弦，分母中的 $\sin(\frac{1}{2}\theta)$ 正是由此而来。至于不等式，诀窍是改写 $\mathrm{C}_n(\theta)$，使 $\frac{1}{2}$ 明显地分离出来。

解答

和式 C_n 与 S_n

设 $z=e^{\mathrm{i}\theta}$。由于 $0<\theta<2\pi$，我们有 $z\neq 1$，故等比级数求和公式适用：

$$\mathrm{C}_n(\theta)+\mathrm{i}\,\mathrm{S}_n(\theta)=\sum_{k=0}^{n}e^{\mathrm{i}k\theta}=\frac{e^{\mathrm{i}(n+1)\theta}-1}{e^{\mathrm{i}\theta}-1}.$$

现在把 $e^{\mathrm{i}\phi}-1=e^{\mathrm{i}\phi/2}\bigl(e^{\mathrm{i}\phi/2}-e^{-\mathrm{i}\phi/2}\bigr)=2\mathrm{i}\,e^{\mathrm{i}\phi/2}\sin(\frac{1}{2}\phi)$ 用于分子（取 $\phi=(n+1)\theta$）和分母（取 $\phi=\theta$）：

$$\mathrm{C}_n(\theta)+\mathrm{i}\,\mathrm{S}_n(\theta)=\frac{2\mathrm{i}\,e^{\mathrm{i}(n+1)\theta/2}\sin\bigl(\frac{1}{2}(n+1)\theta\bigr)}{2\mathrm{i}\,e^{\mathrm{i}\theta/2}\sin(\frac{1}{2}\theta)}=e^{\mathrm{i}n\theta/2}\,\frac{\sin\bigl(\frac{1}{2}(n+1)\theta\bigr)}{\sin(\frac{1}{2}\theta)}.$$

因为 $0<\frac{1}{2}\theta<\pi$，分母不为零。这个分数是实数，所以取实部和虚部（利用 $e^{\mathrm{i}n\theta/2}=\cos\frac{1}{2}n\theta+\mathrm{i}\sin\frac{1}{2}n\theta$）便得

$$\mathrm{C}_n(\theta)=\frac{\cos(\frac{1}{2}n\theta)\sin\bigl(\frac{1}{2}(n+1)\theta\bigr)}{\sin(\frac{1}{2}\theta)},\qquad \mathrm{S}_n(\theta)=\frac{\sin(\frac{1}{2}n\theta)\sin\bigl(\frac{1}{2}(n+1)\theta\bigr)}{\sin(\frac{1}{2}\theta)}.$$

第一式就是题目要求的结果，第二式则是 S_n 的相应表达式。

想一想 有没有不用复数也能看出这个结果的方法?

有：把每一项乘以 $2\sin(\frac{1}{2}\theta)$，再用积化和差公式（product-to-sum formula）

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

由 $k=0$ 加到 n，各项裂项相消（telescope），只剩下最后一项和最初一项：

$$2\sin(\tfrac{1}{2}\theta)\,\mathrm{C}_n(\theta) = \sin\big((n+\tfrac{1}{2})\theta\big) - \sin(-\tfrac{1}{2}\theta) = \sin\big((n+\tfrac{1}{2})\theta\big) + \sin(\tfrac{1}{2}\theta). \qquad (*)$$

根据和差化积公式（sum-to-product formula）$\sin A + \sin B = 2\sin\frac{1}{2}(A+B)\cos\frac{1}{2}(A-B)$，取 $A=(n+\frac{1}{2})\theta$，$B=\frac{1}{2}\theta$，右边等于 $2\sin\big(\frac{1}{2}(n+1)\theta\big)\cos(\frac{1}{2}n\theta)$，与上面求得的结果一致。

不等式

式 $(*)$ 正是我们想要的形式。两边除以 $2\sin(\frac{1}{2}\theta)\neq 0$，得

$$\mathrm{C}_n(\theta) = \frac{1}{2} + \frac{\sin\big((n+\frac{1}{2})\theta\big)}{2\sin(\frac{1}{2}\theta)}.$$

（若改由题目给出的公式出发，可用 $2\cos(\frac{1}{2}n\theta)\sin\big(\frac{1}{2}(n+1)\theta\big) = \sin\big((n+\frac{1}{2})\theta\big) + \sin(\frac{1}{2}\theta)$ 得到同一行。）因此

$$\left|\mathrm{C}_n(\theta) - \frac{1}{2}\right| = \frac{\left|\sin\big((n+\frac{1}{2})\theta\big)\right|}{2\sin(\frac{1}{2}\theta)} \le \frac{1}{2\sin(\frac{1}{2}\theta)},$$

其中我们用到了当 $0<\theta<2\pi$ 时 $\sin(\frac{1}{2}\theta)>0$（所以分母不需要加上模（modulus）号），以及 $|\sin x|\le 1$。这就是所求的不等式，而等号成立当且仅当 $\sin\big((n+\frac{1}{2})\theta\big)=\pm 1$。

想一想 为甚么 $\sin(\frac{1}{2}\theta)$ 为正这一点很重要?

假如容许 θ 取 $(0,2\pi)$ 以外的值，$\sin(\frac{1}{2}\theta)$ 便可能是负数，不等式的右边也会变成负数，而模不可能小于负数。在题目给定的范围内，$\frac{1}{2}\theta\in(0,\pi)$，所以这个上界确实是一个正数。

验算

- $n=1$: $\mathrm{C}_1 = 1+\cos\theta = 2\cos^2(\frac{1}{2}\theta)$，而公式给出 $\cos(\frac{1}{2}\theta)\sin\theta/\sin(\frac{1}{2}\theta) = 2\cos^2(\frac{1}{2}\theta)$。同样，$\mathrm{S}_1=\sin\theta$ 与 $\sin(\frac{1}{2}\theta)\sin\theta/\sin(\frac{1}{2}\theta)$ 相符。
- 当 $\theta\to 0$ 时，C_n 的公式趋向 $(n+1)$，正是 $n+1$ 个一之和；上界 $1/(2\sin\frac{1}{2}\theta)$ 则趋向无穷大，这是必然的。

- 数值验算，取 $n=3$, $\theta=1$：直接求和得 $\mathrm{C}_3(1)\approx 0.1342$，与公式相符，而且 $|\mathrm{C}_3-\tfrac{1}{2}|\approx 0.366\le 1/(2\sin\tfrac{1}{2})\approx 1.043$。等号成立的例子有 $n=2$，$\theta=\pi/5$，此时两边都等于 $1.618\ldots$。

常见错误

- 忘记和式由 $k=0$ 开始，所以共有 $n+1$ 项：等比级数的分子是 $e^{\mathrm{i}(n+1)\theta}$，而不是 $e^{\mathrm{i}n\theta}$。
- 没有先把分母化为实数，就对 $\left(e^{\mathrm{i}(n+1)\theta}-1\right)/\left(e^{\mathrm{i}\theta}-1\right)$ 取实部；半角因式分解能干净利落地做到这一点。
- 写下 $|\sin(\tfrac{1}{2}\theta)|$，却没有说明它为何等于 $\sin(\tfrac{1}{2}\theta)$。

STEP 之外

函数 $\mathrm{C}_n(\theta)-\tfrac{1}{2}=\sin\big((n+\tfrac{1}{2})\theta\big)/\big(2\sin\tfrac{1}{2}\theta\big)$ 除了相差一个常数因子外，就是狄利克雷核（Dirichlet kernel），它是傅里叶级数（Fourier series）理论的核心对象：一个函数的第 n 个傅里叶部分和，就是该函数对这个核取的平均。这里证明的上界，加上分子的振荡，是证明傅里叶级数何时收敛的起点（也是用于 $\sum\cos(k\theta)/k$ 这类级数的狄利克雷判别法（Dirichlet's test）的起点）。Abbott 的 *Understanding Analysis* 有一章讲傅里叶级数，写得浅白易读，用的正是这些和式。

详细解答

STEP III 1995，第 5 题

纯数学 · 麦克劳林级数

STEP III 1995，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/95-s3-q5

关键思路

设 $\theta = \sin^{-1} x$，于是 $\mathrm{d}\theta/\mathrm{d}x = 1/\sqrt{1-x^2}$。把 $y^{(1)}$ 乘以 $\sqrt{1-x^2}$，便可消去这个碍事的因子；再求导一次，就得到不含平方根的二阶方程。接着用莱布尼兹法则（Leibniz's rule）把方程化为递推关系（recurrence），而在 $x = 0$ 处，因子 $1 - x^2$ 变成 1，含 x 的项亦随之消失。这样，递推关系便把 0 处的各阶导数（derivative）全部交到我们手上。

解答

微分方程

设 $\theta = \sin^{-1} x$，则 $\sin\theta = x$，且当 $|x| < 1$ 时，$\cos\theta = \sqrt{1-x^2} > 0$，并有 $\dfrac{\mathrm{d}\theta}{\mathrm{d}x} = \dfrac{1}{\sqrt{1-x^2}}$。

于是 $y = \sin^2 m\theta$，由链式法则（chain rule）得

$$y^{(1)} = 2m \sin m\theta \cos m\theta \cdot \frac{1}{\sqrt{1-x^2}} = \frac{m \sin 2m\theta}{\sqrt{1-x^2}}, \qquad \text{所以} \qquad \sqrt{1-x^2}\, y^{(1)} = m \sin 2m\theta.$$

把上式对 x 求导，左边用积法则（product rule）：

$$\sqrt{1-x^2}\, y^{(2)} - \frac{x}{\sqrt{1-x^2}}\, y^{(1)} = 2m^2 \cos 2m\theta \cdot \frac{1}{\sqrt{1-x^2}}.$$

全式乘以 $\sqrt{1-x^2}$，并利用 $\cos 2m\theta = 1 - 2\sin^2 m\theta = 1 - 2y$：

$$(1-x^2)\, y^{(2)} = x\, y^{(1)} + 2m^2(1-2y),$$

这就是所求的方程。

想一想 为甚么不先把 $\sqrt{1-x^2}\,y^{(1)} = m\sin 2m\theta$ 两边平方?

平方后得 $(1-x^2)\big(y^{(1)}\big)^2 = 4m^2y(1-y)$，再求导的话，每一项都会带着因子 $2y^{(1)}$。这样我们就要除以 $y^{(1)}$，而它在 $x=0$ 处等于零，偏偏那正是我们稍后最关心的一点。先求导而不先平方，就完全不必做这个除法。

递推关系

利用莱布尼兹法则, 把方程求导 n 次, 其中 $n \geq 1$。由于 $1-x^2$ 只有两个非零的导数 ($-2x$ 和 -2)，而 x 只有一个,

$$\begin{aligned}
\frac{\mathrm{d}^n}{\mathrm{d}x^n}\Big[(1-x^2)y^{(2)}\Big] &= (1-x^2)y^{(n+2)} - 2nx\,y^{(n+1)} - n(n-1)\,y^{(n)},\\
\frac{\mathrm{d}^n}{\mathrm{d}x^n}\Big[x\,y^{(1)}\Big] &= x\,y^{(n+1)} + n\,y^{(n)},\\
\frac{\mathrm{d}^n}{\mathrm{d}x^n}\Big[2m^2(1-2y)\Big] &= -4m^2\,y^{(n)}.
\end{aligned}$$

最后一行需要 $n \geq 1$，因为只有这样，常数 $2m^2$ 求导后才会变成零。令两边相等,

$$(1-x^2)y^{(n+2)} - 2nx\,y^{(n+1)} - n(n-1)y^{(n)} = x\,y^{(n+1)} + n\,y^{(n)} - 4m^2y^{(n)},$$

再合并同类项（共有 $2n+1$ 个 $xy^{(n+1)}$，以及 $n(n-1)+n = n^2$），得

$$(1-x^2)\,y^{(n+2)} = (2n+1)\,x\,y^{(n+1)} + (n^2-4m^2)\,y^{(n)} \qquad (n \geq 1),$$

正如所求。

麦克劳林级数

在 $x=0$ 处有 $\theta = 0$，所以

$$y(0) = \sin^2 0 = 0, \qquad y^{(1)}(0) = m\sin 0 = 0, \qquad y^{(2)}(0) = 0 + 2m^2(1-0) = 2m^2,$$

其中最后一个由微分方程得出。在递推关系中代入 $x=0$，得

$$y^{(n+2)}(0) = (n^2-4m^2)\,y^{(n)}(0) \qquad (n \geq 1).$$

奇数阶导数。由于 $y^{(1)}(0) = 0$，递推关系给出 $y^{(3)}(0) = 0$，继而 $y^{(5)}(0) = 0$，如此类推: 所有奇数阶导数在 0 处都等于零。这是合理的，因为 y 是 x 的偶函数（even function）。

偶数阶导数。取 $n = 2j$:

$$y^{(2j+2)}(0) = (4j^2-4m^2)\,y^{(2j)}(0) = -4(m^2-j^2)\,y^{(2j)}(0) \qquad (j \geq 1).$$

由 $y^{(2)}(0) = 2m^2$ 出发，依次对 $j = 1, 2, \ldots, k-1$ 应用上式，

$$y^{(2k)}(0) = 2m^2 \prod_{j=1}^{k-1} \left(-4(m^2 - j^2)\right) = (-1)^{k-1} 2^{2k-1} \prod_{j=0}^{k-1} (m^2 - j^2),$$

这里我们把最前面的 m^2 并入乘积，作为 $j = 0$ 的因子。（当 $k = 1$ 时，公式给出 $2m^2$，与应有的结果相符。）

因此麦克劳林级数（Maclaurin series）为

$$y = \sin^2\left(m \sin^{-1} x\right) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1} 2^{2k-1}}{(2k)!} m^2(m^2 - 1^2)(m^2 - 2^2) \cdots \left(m^2 - (k-1)^2\right) x^{2k},$$

其首数项为

$$y = m^2 x^2 - \frac{m^2(m^2-1)}{3} x^4 + \frac{2m^2(m^2-1)(m^2-4)}{45} x^6 - \cdots.$$

若 m 是整数，则凡 $k > |m|$ 的项都含有因子 $m^2 - m^2 = 0$，所以级数到此为止，y 是一个 $2|m|$ 次多项式。

验算

- $m = 1$：$y = \sin^2(\sin^{-1} x) = x^2$，而级数给出 x^2，其后各项全为零。$m = 2$：级数给出 $4x^2 - 4x^4$，而确实有 $\sin^2 2\theta = 4 \sin^2\theta(1 - \sin^2\theta) = 4x^2 - 4x^4$。
- $m = \frac{1}{2}$，$x = \frac{1}{2}$：$y = \sin^2(\pi/12) = \frac{1}{2}\left(1 - \frac{\sqrt{3}}{2}\right) \approx 0.066987$。级数为 $\frac{1}{4}x^2 + \frac{1}{16}x^4 + \frac{1}{32}x^6 + \cdots$，其首三项之和为 0.066895，多加几项，差距便会逐渐缩小。
- 我们以数值方法验证过：微分方程在随机选取的 x 和 m 上成立；级数的系数精确地满足该方程；以及当 $|x| \le 0.8$ 时，级数（取至 300 项）与 y 相符，误差约为 10^{-15}。

常见错误

- 在 $n = 0$ 时使用递推关系。它只对 $n \ge 1$ 成立，因为 $n = 0$ 时常数 $2m^2$ 不会消失。$n = 0$ 的情况就是原来的方程，我们正是由它得出 $y^{(2)}(0) = 2m^2$。
- 弄错莱布尼兹法则的系数：$-n(n-1)y^{(n)}$ 这一项来自 $\binom{n}{2} \cdot (-2)$，而不是 $\binom{n}{2} \cdot (-2x)$。
- 写通项时漏掉正负交替的符号。每一步都乘以 $4(j^2 - m^2)$，而当 $m^2 > j^2$ 时这是负数。

STEP 之外

由导数砌出来的麦克劳林级数，未必等于原来的函数，但在这里，当 $|x| < 1$ 时两者确实相等。这个级数的收敛半径（radius of convergence）为 1（当 m 不是整数时，可用比值

审敛法（ratio test）得出），而 $\sin^{-1} x$ 到了 $x = \pm 1$ 就不再性质良好。m 为整数时所得的多项式，是切比雪夫多项式（Chebyshev polynomials）的近亲，因为 $\cos 2m\theta$ 是 $\sin\theta$ 的多项式。想知道这类级数为何真的收敛到原来的函数，可阅读 Abbott 的 *Understanding Analysis* 中关于带余项的泰勒定理（Taylor's theorem with remainder）的部分。Needham 的 *Visual Complex Analysis* 则解释了为何收敛半径取决于 ± 1 处的毛病。

详细解答

STEP III 1995，第 6 题

纯数学 · 轨迹与反演

STEP III 1995，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/95-s3-q6

关键思路

圆 $|z-\mathrm{i}|=1$ 通过原点，所以它的极式（polar form）方程特别简单：模（modulus）是辐角（argument）的单一个三角函数。一旦把 z 写成模—辐角形式，$1/z$ 便立即可得（取模的倒数，并把辐角变号），再把它写成直角坐标形式（Cartesian form），便会发现这个圆变成了一条直线。最后一部分是同一现象的逆向版本，直接设 $w=x+\mathrm{i}y$ 来做最快捷。

解答

z 的模

设 $z=r(\cos\theta+\mathrm{i}\sin\theta)$，其中 $r>0$。则 $z-\mathrm{i}=r\cos\theta+\mathrm{i}(r\sin\theta-1)$，所以

$$\begin{aligned}|z-\mathrm{i}|^2=1 &\iff r^2\cos^2\theta+r^2\sin^2\theta-2r\sin\theta+1=1\\ &\iff r^2-2r\sin\theta=0\\ &\iff r=2\sin\theta,\end{aligned}$$

其中最后一步除以 r，这是容许的，因为 $z\neq 0$。所以

$$|z|=2\sin\theta.$$

由于 $r>0$，必有 $\sin\theta>0$，即 $0<\theta<\pi$：这个圆（圆心 i，半径 1）位于上半平面，只在原点与实轴相切，而原点已被排除。反过来，每一个 $\theta\in(0,\pi)$ 都给出圆上的一点 $z=2\sin\theta\,(\cos\theta+\mathrm{i}\sin\theta)$，所以当 θ 走遍 $(0,\pi)$ 时，z 描出除原点外的整个圆。

? **想一想** 为甚么从图形来看，$|z|=2\sin\theta$ 是“显而易见”的?

由 0 至 2i 的线段是圆的一条直径。根据半圆上的圆周角定理(angle in a semicircle)，以 0、z、2i 为顶点的三角形在 z 处是直角，而在 0 处的角是 $\frac{\pi}{2}-\theta$。所以 $|z| = 2\cos(\frac{\pi}{2}-\theta) = 2\sin\theta$。

$1/z$ 的模和辐角

由于 $z = 2\sin\theta\,(\cos\theta + \mathrm{i}\sin\theta)$，

$$\frac{1}{z} = \frac{1}{2\sin\theta}\big(\cos(-\theta) + \mathrm{i}\sin(-\theta)\big),$$

所以

$$\left|\frac{1}{z}\right| = \frac{1}{2\sin\theta}, \qquad \arg\frac{1}{z} = -\theta.$$

因为 $0 < \theta < \pi$，角 $-\theta$ 位于 $(-\pi, 0)$ 之内，所以它就是辐角主值（principal argument)。

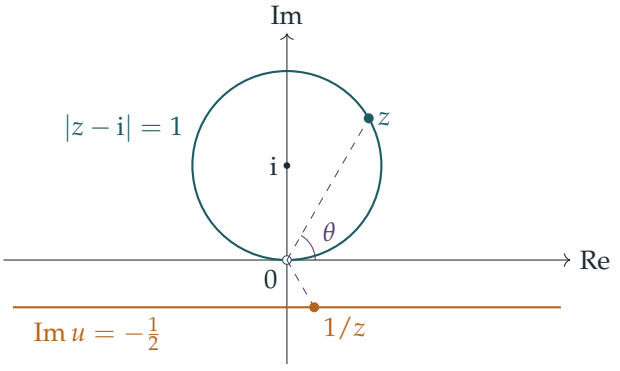
轨迹

要看出 $1/z$ 描出甚么图形，可把它转为直角坐标形式：

$$\frac{1}{z} = \frac{\cos\theta - \mathrm{i}\sin\theta}{2\sin\theta} = \tfrac{1}{2}\cot\theta - \tfrac{1}{2}\mathrm{i}.$$

虚部恒为 $-\frac{1}{2}$，而当 θ 走遍 $(0, \pi)$ 时，实部 $\frac{1}{2}\cot\theta$ 取遍每一个实数值恰好一次（它由 $+\infty$ 递减至 $-\infty$)。所以 $1/z$ 描出整条直线 $\operatorname{Im} u = -\frac{1}{2}$。

因此 z 位于圆心为 i、半径为 1 的圆上（原点除外)，而 $1/z$ 位于通过 $-\frac{1}{2}\mathrm{i}$ 的水平线上。下图画出两者，并标出 $\theta = \frac{\pi}{3}$ 时的一对点：这时 $z = \frac{\sqrt{3}}{2} + \frac{3}{2}\mathrm{i}$，而 $1/z = \frac{1}{2\sqrt{3}} - \frac{1}{2}\mathrm{i}$。这两点分别位于角度为 θ 和 $-\theta$ 的射线上，而它们与 0 的距离相乘得 1。当 z 趋近原点时，$1/z$ 沿着直线走向无穷远；圆的最高点 $z = 2\mathrm{i}$ 则对应点 $-\frac{1}{2}\mathrm{i}$，即直线上最接近原点的一点。



轨迹 C

设 $w = x + \mathrm{i}y \neq 0$（要 $1/w$ 存在，必须有 $w \neq 0$)。则

$$\frac{1}{w} = \frac{\bar{w}}{|w|^2} = \frac{x - \mathrm{i}y}{x^2 + y^2}, \qquad \text{所以} \qquad \operatorname{Re}\frac{1}{w} = \frac{x}{x^2 + y^2}.$$

因此，对于 $w \neq 0$，

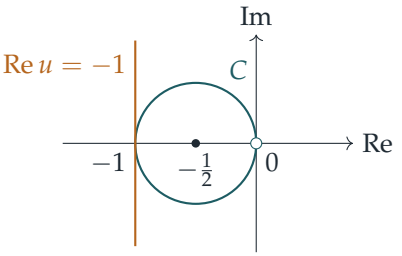
$$\begin{aligned}\operatorname{Re}\frac{1}{w} = -1 &\iff x = -(x^2+y^2)\\ &\iff x^2 + x + y^2 = 0\\ &\iff \left(x+\tfrac{1}{2}\right)^2 + y^2 = \tfrac{1}{4}.\end{aligned}$$

每一步都是可逆的（我们乘以了 x^2+y^2，而它不等于零），所以这是名副其实的“当且仅当”。因此 C 是圆

$$\left|w + \tfrac{1}{2}\right| = \tfrac{1}{2},$$

圆心为 $-\frac{1}{2}$，半径为 $\frac{1}{2}$，但要除去原点。原点确实在这个圆上，但 $1/w$ 在该处没有定义，所以它不属于 C。

这与第一部分互相呼应：在那里，一个通过 0 的圆被 $z \mapsto 1/z$ 映射到一条不通过 0 的直线；在这里，直线 $\operatorname{Re} u = -1$ 被同一个映射（它是自己的逆映射）映射到一个通过 0 的圆。



验算

- 取 $\theta = \frac{\pi}{6}$：则 $|z| = 1$ 且 $z = \frac{\sqrt{3}}{2} + \frac{1}{2}\mathrm{i}$，而确实有 $|z - \mathrm{i}| = \left|\frac{\sqrt{3}}{2} - \frac{1}{2}\mathrm{i}\right| = 1$。它的倒数就是它的共轭复数（conjugate），$\frac{\sqrt{3}}{2} - \frac{1}{2}\mathrm{i} \approx 0.866 - 0.5\mathrm{i}$，其虚部为 $-\frac{1}{2}$，与预期相符。
- $\theta = \frac{\pi}{2}$ 给出 $z = 2\mathrm{i}$，即圆的最高点，而 $1/z = -\frac{1}{2}\mathrm{i}$。
- 对于 C：$w = -1$ 在圆上，而 $1/w = -1$；$w = -\frac{1}{2} + \frac{1}{2}\mathrm{i}$ 在圆上，而 $1/w = -1 - \mathrm{i}$，实部为 -1。

在每条轨迹上各取 10 000 个随机点作数值测试，三个结果全部得到确认。

常见错误

- 忘记范围 $0 < \theta < \pi$：公式 $|z| = 2\sin\theta$ 只在 $\sin\theta > 0$ 时才有意义。

- 把 $1/z$ 的轨迹说成半直线或线段。由于 $\cot\theta$ 在 $(0,\pi)$ 上取遍所有实数值，轨迹是整条直线。
- 把原计数入 C。原点满足圆的方程，但 $1/w$ 在该处并不存在。

STEP 之外

映射 $z \mapsto 1/z$ 是莫比乌斯变换（Möbius transformation）$z \mapsto (az+b)/(cz+d)$ 的一个例子。这类映射的一个核心性质，是它们把“广义圆”（圆和直线）映射到广义圆，而通过原点的圆正正就是被 $1/z$ 映射到直线的那些圆。这是黎曼球面（Riemann sphere）几何的起点；在黎曼球面上，直线只不过是通过无穷远点的圆。Needham 的 *Visual Complex Analysis*（《可视化复分析》）配以图像，把这一切阐述得十分精彩。

详细解答

STEP III 1995，第 7 题

纯数学 · 群与同构

STEP III 1995，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/95-s3-q7

关键思路

结合律（associativity）已经给定，所以对每个集合只需检查三件事：封闭性（closure）、单位元（identity），以及每个元素都有逆元（inverse）。只要其中一项不成立，该集合就不是群（group），而且举出一个明确的反例便已足够。至于同构（isomorphism），则要找出任何同构都必须保持的结构特征：元素的个数，以及各元素的阶（order）。

解答

以下所说的“乘法”，视乎情况而定，分别指复数的普通乘法、模（modulo）n 乘法，或矩阵乘法。

第 (i) 部分

设 $S = \{z \in \mathbb{C} : |z| = 1\}$。

- 封闭性：若 $|z| = |w| = 1$，则 $|zw| = |z||w| = 1$。
- 单位元：$1 \in S$，且 $1z = z1 = z$。
- 逆元：若 $|z| = 1$，则 $z \neq 0$，而 $z^{-1} = \bar{z}$（因为 $z\bar{z} = |z|^2 = 1$），其模（modulus）同样是 1。

所以 S 是一个群（而且是无限群）。

第 (ii) 部分

在模 4 乘法之下，集合 $\{0,1,2,3\}$ 是封闭的，并有单位元 1，但 0 没有逆元，因为对任何 b 都有 $0\times b\equiv 0\not\equiv 1$。所以它不是群。（即使把 0 拿走也无补于事：2 仍然没有逆元，因为 $2b$ 必为偶数；事实上 $2\times 2\equiv 0$，所以 $\{1,2,3\}$ 连封闭性也不具备。）

第 (iii) 部分

利用复角公式（compound-angle formulae），

$$\mathrm{M}(\alpha)\mathrm{M}(\beta)=\begin{pmatrix} c_\alpha c_\beta - s_\alpha s_\beta & -(s_\alpha c_\beta + c_\alpha s_\beta) \\ s_\alpha c_\beta + c_\alpha s_\beta & c_\alpha c_\beta - s_\alpha s_\beta \end{pmatrix}=\mathrm{M}(\alpha+\beta),$$

其中 $c_\alpha=\cos\alpha, s_\alpha=\sin\alpha$，如此类推。由于 cos 和 sin 的周期为 2π，故 $\mathrm{M}(\alpha+\beta)=\mathrm{M}(\gamma)$，其中 γ 是把 $\alpha+\beta$ 约化至 $[0,2\pi)$ 内所得的角。因此：

- 封闭性：集合中两个矩阵的积仍在集合之内。
- 单位元：$\mathrm{M}(0)=I$。
- 逆元：当 $0<\theta<2\pi$ 时，$\mathrm{M}(\theta)\mathrm{M}(2\pi-\theta)=\mathrm{M}(2\pi)=I$；而 $\mathrm{M}(0)$ 的逆元就是它自己。

所以这个集合是一个群。

第 (iv) 部分

模 8 之下，

$$3^2=9\equiv 1,\qquad 5^2=25\equiv 1,\qquad 7^2=49\equiv 1,$$
$$3\times 5=15\equiv 7,\qquad 3\times 7=21\equiv 5,\qquad 5\times 7=35\equiv 3.$$

连同涉及 1 的平凡乘积，这就给出了整个乘法表（乘法满足交换律（commutative）），而每个乘积都在 $\{1,3,5,7\}$ 之内，所以集合是封闭的。单位元是 1，而每个元素的逆元都是它自己。所以这是一个群，阶为 4，其中除 1 以外每个元素的阶都是 2。

第 (v) 部分

单位矩阵 I 的元素都是整数，而且集合是封闭的（整数的积与和仍是整数），但逆元这一项不成立。零矩阵根本没有逆矩阵。即使是可逆矩阵也可能出问题：$A=\begin{pmatrix}2&0\\0&1\end{pmatrix}$ 在实矩阵中的逆矩阵是 $\begin{pmatrix}\frac{1}{2}&0\\0&1\end{pmatrix}$，而逆矩阵是唯一的，所以 A 没有元素全为整数的逆矩阵。所以这个集合不是群。

想一想 单单指出零矩阵真的足够吗？还是需要就所有矩阵说明一下？

只要有一个元素没有逆元，就足以证明该集合不是群。第二个例子则说明，问题并不止于奇异矩阵（singular matrix）。一个整数矩阵有整数逆矩阵，当且仅当它的行列式（determinant）是 ± 1，因为 $\det A \det A^{-1} = 1$，而两个行列式都是整数。若只考虑这些矩阵，便确实得到一个群。

第 (vi) 部分

模 5 之下，2 的各次幂是

$$2^1 = 2, \quad 2^2 = 4, \quad 2^3 = 8 \equiv 3, \quad 2^4 = 16 \equiv 1.$$

所以 $\{1,2,3,4\} = \{2^0, 2^1, 2^2, 2^3\}$，而 $2^a 2^b = 2^{a+b}$，其中指数按模 4 计算。因此集合是封闭的，1 是单位元，而 2^a 的逆元是 2^{4-a}。它是一个群：阶为 4 的循环群（cyclic group），由 2（或由 3）生成。

同构

是群的有 (i)、(iii)、(iv) 和 (vi)，共可组成六对。

(i) 与 (iii) 同构。每个单位复数都可以唯一地写成 $e^{i\theta}$，其中 $0 \le \theta < 2\pi$。定义 $\phi(e^{i\theta}) = M(\theta)$。这是一个双射（bijection），因为 $[0, 2\pi)$ 内不同的 θ 给出不同的矩阵（第一列 $(\cos\theta, \sin\theta)$ 已决定了 θ）。它也保持乘法：$e^{i\alpha}e^{i\beta} = e^{i(\alpha+\beta)}$ 及 $M(\alpha)M(\beta) = M(\alpha+\beta)$，而在两个群中，角都是按模 2π 约化的。从几何角度看，两个群都是平面绕原点的旋转。

(i) 或 (iii) 与 (iv) 或 (vi)：不同构。同构是双射，所以两个群的元素个数必须相同。群 (i) 和 (iii) 是无限的，而 (iv) 和 (vi) 各有四个元素。这样就排除了四对。

想一想 (iv) 和 (vi) 都有四个元素，而且都满足交换律。有甚么可以把它们区分开来？

(iv) 与 (vi)：不同构。同构 ϕ 保持每个元素的阶：$\phi(g)^n = \phi(g^n)$，而它等于单位元当且仅当 g^n 是单位元，因为 ϕ 是单射（injective），并把单位元映到单位元。在 (vi) 中，元素 2 的阶是 4。在 (iv) 中，每个元素 g 都满足 $g^2 = 1$，所以没有元素的阶是 4。因此不存在同构。

总括而言，唯一同构的一对是 (i) 和 (iii)。

验算

我们用电脑对 (iv) 和 (vi) 的乘法表逐一穷举，确认了封闭性、单位元和逆元三项性质。电脑亦发现，模 4 之下 0 和 2 没有逆元，而元素的阶在 (iv) 中是 $1,2,2,2$，在 (vi) 中是 $1,4,4,2$（依次对应 $1,2,3,4$）。它还在 10^4 个随机角上验证了 $\mathrm{M}(\alpha)\mathrm{M}(\beta)=\mathrm{M}(\alpha+\beta)$ 以及与 $\mathrm{e}^{i\theta}$ 的对应关系。(iv) 可以很快用手算验证：任何奇数的平方在模 8 之下都是 1，因为 $(2k+1)^2=4k(k+1)+1$，而 $k(k+1)$ 是偶数。

常见错误

- 说 (iv) 和 (vi) 同构，“因为两者都有四个元素”。元素个数相等是必要条件，但并非充分条件。
- 在 (v) 中只检查 $\det A\neq 0$。逆矩阵的元素也必须全为整数。
- 把 (ii) 误读成模 4 加法。在加法之下，它会是一个阶为 4 的循环群，与 (vi) 同构，但题目指明的是乘法。

STEP 之外

在同构意义下，阶为 4 的群恰好只有两个：循环群 C_4，即 (vi)；以及克莱因四元群（Klein four-group）$C_2\times C_2$，即 (iv)。按元素的阶把小阶数的群分类，是大学群论课程最先处理的课题之一。(i) 和 (iii) 的圆群（circle group）是李群（Lie group）最简单的例子；李群是指同时也是光滑曲线或曲面的群。(v) 中行列式为 ±1 的整数矩阵构成群 $\mathrm{GL}_2(\mathbb{Z})$，它在数论中十分重要。想进一步了解，Allenby 的 *Rings, Fields and Groups* 是一本很好的读物。

详细解答

STEP III 1995，第 8 题

纯数学 · 平面与球面

STEP III 1995，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/95-s3-q8

关键思路

平面截球面时，所得的圆的圆心就是由球心向该平面所作垂线的垂足（foot of the perpendicular）。因此，球心必定位于通过每一个圆的圆心的法线（normal）上，而毕氏定理（Pythagoras）则把球心沿每条法线的距离与球的半径联系起来。题目余下的部分，不过是把这一条向量方程与精心挑选的向量作点积（dot product）而已。

解答

法线与距离

通过 X 而垂直于 π 的法线，就是通过 $\mathbf{x}$ 而方向为 $\mathbf{n}$ 的直线：

$$\mathbf{r} = \mathbf{x} + t\,\mathbf{n}, \qquad t \in \mathbb{R}.$$

它与 π 相交于 $(\mathbf{x}+t\mathbf{n})\cdot\mathbf{n} = p$ 之处，即（因为 $\mathbf{n}\cdot\mathbf{n} = 1$）$t = p - \mathbf{x}\cdot\mathbf{n}$ 之处。所以垂足是 $\mathbf{x} + (p - \mathbf{x}\cdot\mathbf{n})\mathbf{n}$，而由于 $\mathbf{n}$ 是单位向量（unit vector），X 到 π 的垂直距离为

$$|\mathbf{x}\cdot\mathbf{n} - p|.$$

同时通过两个圆的球面

假设有一个球心为 $\mathbf{c}$、半径为 R 的球面包含这两个圆。固定 i。C_i 上的每一点 $\mathbf{y}$ 都满足 $|\mathbf{y} - \mathbf{r}_i| = 1$ 及 $(\mathbf{y} - \mathbf{r}_i)\cdot\mathbf{n}_i = 0$；当 $\mathbf{y}$ 绕 C_i 走一圈时，向量 $\mathbf{u} = \mathbf{y} - \mathbf{r}_i$ 会取遍与 $\mathbf{n}_i$ 垂直的每一个单位向量。

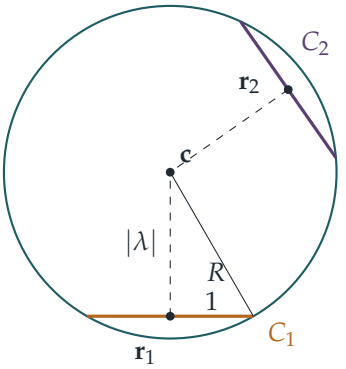
作分解 $\mathbf{c}-\mathbf{r}_i=\mathbf{a}+\mu_i\mathbf{n}_i$，其中 $\mathbf{a}\perp\mathbf{n}_i$。于是

$$\begin{aligned}R^2=|\mathbf{y}-\mathbf{c}|^2&=|\mathbf{u}-\mathbf{a}-\mu_i\mathbf{n}_i|^2\\&=1-2\,\mathbf{u}\cdot\mathbf{a}+|\mathbf{a}|^2+\mu_i^2,\end{aligned}$$

这里用到 $\mathbf{u}\cdot\mathbf{n}_i=\mathbf{a}\cdot\mathbf{n}_i=0$。左边对 C_i 上的每一点都相同，所以 $\mathbf{u}\cdot\mathbf{a}$ 对每一个容许的 $\mathbf{u}$ 都取相同的值。但只要 $\mathbf{u}$ 是容许的，$-\mathbf{u}$ 也是容许的，所以对平面上每一个单位向量 $\mathbf{u}$ 都有 $\mathbf{u}\cdot\mathbf{a}=-\mathbf{u}\cdot\mathbf{a}=0$，而取 $\mathbf{u}$ 与 $\mathbf{a}$ 平行（假如 $\mathbf{a}\neq\mathbf{0}$）便会得出矛盾。因此 $\mathbf{a}=\mathbf{0}$，并且

$$\mathbf{c}=\mathbf{r}_i+\mu_i\mathbf{n}_i,\qquad R^2=1+\mu_i^2.$$

用文字来说：球心位于通过 C_i 的圆心而垂直于 π_i 的法线上，这正是第一部分所求得的直线。



把以上结果应用于 $i=1$ 和 $i=2$，得 $1+\mu_1^2=R^2=1+\mu_2^2$，所以 $\mu_2=\pm\mu_1$。记 $\lambda=\mu_1$，则

$$\mathbf{r}_1+\lambda\mathbf{n}_1=\mathbf{c}=\mathbf{r}_2\pm\lambda\mathbf{n}_2,$$

所以只要存在合适的球面，这样的 λ 就必定存在，这正是所要证明的。

两个必要条件

移项后得

$$\mathbf{r}_1-\mathbf{r}_2=-\lambda\mathbf{n}_1\pm\lambda\mathbf{n}_2.$$

两边与 $\mathbf{n}_1\times\mathbf{n}_2$ 作点积，而这个向量同时垂直于 $\mathbf{n}_1$ 和 $\mathbf{n}_2$：

$$(\mathbf{r}_1-\mathbf{r}_2)\cdot(\mathbf{n}_1\times\mathbf{n}_2)=0.$$

? **想一想** p_1 从何而来？它并没有在向量方程中出现。

圆心 $\mathbf{r}_1$ 位于 π_1 上，所以 $\mathbf{n}_1 \cdot \mathbf{r}_1 = p_1$；同理 $\mathbf{n}_2 \cdot \mathbf{r}_2 = p_2$。记 $k = \mathbf{n}_1 \cdot \mathbf{n}_2$。把 $\mathbf{r}_1 + \lambda \mathbf{n}_1 = \mathbf{r}_2 \pm \lambda \mathbf{n}_2$ 先与 $\mathbf{n}_1$ 作点积，再与 $\mathbf{n}_2$ 作点积，得

$$p_1 + \lambda = \mathbf{n}_1 \cdot \mathbf{r}_2 \pm \lambda k,$$
$$\mathbf{n}_2 \cdot \mathbf{r}_1 + \lambda k = p_2 \pm \lambda.$$

所以

$$p_1 - \mathbf{n}_1 \cdot \mathbf{r}_2 = \lambda(\pm k - 1), \qquad p_2 - \mathbf{n}_2 \cdot \mathbf{r}_1 = \lambda(k \mp 1).$$

取上方的符号时，两式右边都等于 $\lambda(k-1)$；取下方的符号时，它们分别是 $-\lambda(k+1)$ 和 $\lambda(k+1)$。无论哪一种情况，两式左边不是相等就是互为相反数，所以

$$(p_1 - \mathbf{n}_1 \cdot \mathbf{r}_2)^2 = (p_2 - \mathbf{n}_2 \cdot \mathbf{r}_1)^2.$$

几何意义

第一个条件。它说明 $\mathbf{r}_1 - \mathbf{r}_2$、$\mathbf{n}_1$ 和 $\mathbf{n}_2$ 共面（coplanar），即纯量三重积（scalar triple product）等于零。等价地说，通过 C_1 的圆心而垂直于 π_1 的法线，与通过 C_2 的圆心而垂直于 π_2 的法线，位于同一个平面上：它们不是异面直线（skew lines），所以它们或者相交（交于球心），或者平行。

第二个条件。根据第一部分，$|p_1 - \mathbf{n}_1 \cdot \mathbf{r}_2|$ 是 C_2 的圆心到 π_1 的距离，而 $|p_2 - \mathbf{n}_2 \cdot \mathbf{r}_1|$ 是 C_1 的圆心到 π_2 的距离。所以这个条件说明这两个距离相等：每个圆的圆心到另一个圆所在平面的距离，等于另一个圆心到这个圆所在平面的距离。这反映了图形的对称性：两个圆与 $\mathbf{c}$ 的距离同为 $|\lambda|$，所以对某个通过 $\mathbf{c}$ 的合适平面作反射（reflection），便会把两个圆互换，同时把这两个距离互换。

验算

- 取 $\pi_1 : z = 0$ 和 $\pi_2 : z = 2$，两个单位圆的圆心都在 z 轴上。它们位于球心为 $(0, 0, 1)$、半径为 $\sqrt{2}$ 的球面上；这时 $\lambda = 1$，取下方的符号，$\mathbf{n}_1 \times \mathbf{n}_2 = \mathbf{0}$，而第二个条件中的两个距离都等于 2。
- 在数值方面，我们随机构造了 5000 个半径 $R > 1$ 的球面，每个球面配上两个随机平面，各截出一个单位圆，并确认了圆上每一点都在球面上，结果发现两个条件在每一个情况下都成立，误差不超过 10^{-9}，而且两种符号都有出现。至于随机选取、互不相关的两个圆，则每一次都不满足第一个条件。

常见错误

- 未加论证便假设球心位于通过 $\mathbf{r}_i$ 的法线上；这正是整个证明的核心。

- 遗漏了 $\pm$：球心可以位于两个平面各自的法向量所指的同一侧，也可以位于相反的两侧。
- 忘记了由于 C_1 的圆心位于 π_1 上，所以 $\mathbf{n}_1 \cdot \mathbf{r}_1 = p_1$。

STEP 之外

这里的条件只是必要条件（necessary conditions）：要得到完整的“当且仅当”（if and only if），还需要两条法线确实相交于一个与两个圆等距的点。这类题目是踏入二次曲面（quadrics）解析几何（analytic geometry）的第一步，在那里，平面、球面以及它们的交集都用线性代数（linear algebra）来处理；纯量三重积会以量度体积的 3×3 行列式（determinant）的形式再次出现。Strang 的 *Introduction to Linear Algebra*（线性代数入门）对这些概念有细致的阐述。

详细解答

STEP III 1995，第 9 题

力学 · 复摆

STEP III 1995，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：

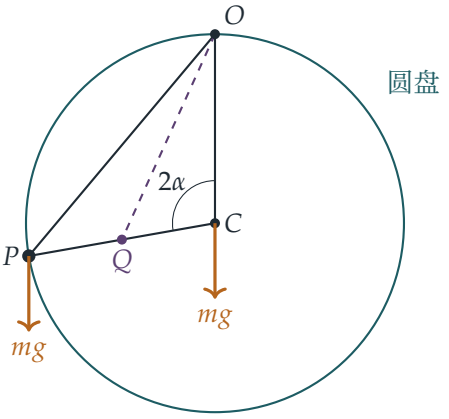
step.maths.org/questions/95-s3-q9

关键思路

圆盘与质点的质量相同，所以整个系统的质心（centre of mass）就是 CP 的中点 Q。差不多所有结论都由此而来：在稳定平衡（stable equilibrium）时，Q 位于 O 的正下方；而第 (iii) 部分的能量方程只需要 Q 的高度。余下的工作，就是等腰三角形 OCP 的几何，以及一个标准的转动惯量（moment of inertia）积分。

解答

我们把圆盘视为刚性薄片（rigid lamina），把 P 视为固定在其边缘上的质点（particle）；转轴是光滑的，所以机械能守恒。全题中 $OC = CP = r$，且 $\angle OCP = 2\alpha$。



第 (i) 部分

由于圆盘（质心为 C）与质点（位于 P）的质量同为 m，系统的质心就是 CP 的中点 Q。作用于系统的力只有重力和转轴上的反作用力，所以系统处于平衡，当且仅当 Q 与 O 在同一铅垂线上；而当 Q 在 O 的下方时，平衡是稳定的（此时任何位移都会令 Q 升高，使势能增加）。因此在稳定平衡时 OQ 是铅垂的，而 β 就是角 $\angle COQ$。

在圆盘所在的平面上取坐标轴，以 C 为原点，并令 O 位于 $(r,0)$；于是 $P = r(\cos 2\alpha, \sin 2\alpha)$，$Q = \frac{r}{2}(\cos 2\alpha, \sin 2\alpha)$。因此

$$\overrightarrow{OC} = (-r,\,0), \qquad \overrightarrow{OQ} = \tfrac{r}{2}\big(-(2-\cos 2\alpha),\ \sin 2\alpha\big).$$

由于 $2-\cos 2\alpha > 0$，向量 $\overrightarrow{OQ}$ 沿 $\overrightarrow{OC}$ 方向的分量为正数 $\frac{r}{2}(2-\cos 2\alpha)$，而与之垂直的分量为 $\frac{r}{2}\sin 2\alpha$，所以 β 是锐角，且

$$\tan\beta = \frac{\frac{r}{2}\sin 2\alpha}{\frac{r}{2}(2-\cos 2\alpha)} = \frac{\sin 2\alpha}{2-\cos 2\alpha},$$

这正是所要证明的结果。

第 (ii) 部分

我们把 $\rho x/r$ 理解为每单位面积的质量。半径介乎 x 与 $x+\mathrm{d}x$ 之间的圆环，面积为 $2\pi x\,\mathrm{d}x$，所以

$$m = \int_0^r \frac{\rho x}{r}\,2\pi x\,\mathrm{d}x = \frac{2\pi\rho r^2}{3}.$$

该圆环上每一点与通过 C 且垂直于圆盘的轴的距离都是 x，所以对该轴的转动惯量为

$$I_C = \int_0^r \frac{\rho x}{r}\,2\pi x\cdot x^2\,\mathrm{d}x = \frac{2\pi\rho r^4}{5} = \frac{3}{5}mr^2,$$

其中用了 $2\pi\rho r^2 = 3m$。通过 O 的轴与这条轴平行（两者都垂直于圆盘所在的铅垂平面），相距 $OC = r$，所以由平行轴定理（parallel axis theorem）得

$$I_O = \frac{3}{5}mr^2 + mr^2 = \frac{8}{5}mr^2.$$

第 (iii) 部分

系统的转动惯量。三角形 OCP 是等腰三角形，顶角为 2α，所以弦长 $OP = 2r\sin\alpha$。质点的贡献是 $m(2r\sin\alpha)^2$，于是

$$I = \frac{8}{5}mr^2 + 4mr^2\sin^2\alpha = \frac{4}{5}mr^2\,(2+5\sin^2\alpha).$$

质心下降的高度。释放时 OQ 是水平的，所以 Q 与 O 同高。在最终位置，P 在 O 的正下方，所以 OP 是铅垂的，而任何一点在 O 以下的深度，就是它沿 OP 方向与 O 的距离。

- P 的深度为 $OP = 2r\sin\alpha$。
- 三角形 OCP 的底角为 $\frac{1}{2}(\pi - 2\alpha) = \frac{\pi}{2} - \alpha$，所以 $\angle COP = \frac{\pi}{2} - \alpha$，而 C 的深度为 $r\cos(\frac{\pi}{2} - \alpha) = r\sin\alpha$。

Q 是 CP 的中点，所以它的深度是两者的平均值，即 $\frac{3}{2}r\sin\alpha$。因此系统（总质量为 $2m$）损失的势能为 $2mg \cdot \frac{3}{2}r\sin\alpha = 3mgr\sin\alpha$。

> **想一想** 系统真的会到达 P 在 O 下方的那个位置吗?

系统在 Q 与 O 同高时由静止释放，其后便像复摆（compound pendulum）一样，在 OQ 水平的两个位置之间摆动，并经过 Q 低于 O 所在水平的每一个位置。在最终位置，Q 的深度为 $\frac{3}{2}r\sin\alpha > 0$，所以系统确实会到达该位置（而“$C$ 起初比 P 低”这个条件，只是决定了摆动的方向）。

能量。由静止开始，设角速率（angular speed）为 ω，则

$$\tfrac{1}{2}I\omega^2 = 3mgr\sin\alpha \quad\Longrightarrow\quad \omega^2 = \frac{6gr\sin\alpha}{\frac{4}{5}r^2(2+5\sin^2\alpha)} = \frac{15g\sin\alpha}{2r(2+5\sin^2\alpha)}.$$

C 沿着以 O 为圆心、半径为 $OC = r$ 的圆运动，所以 $v = r\omega$，且

$$v^2 = \frac{15gr\sin\alpha}{2(2+5\sin^2\alpha)},$$

这正是所要证明的结果。

极大值。设 $s = \sin\alpha$；由于 $0 < \alpha \le \pi/2$，故 $s \in (0,1]$。令 $f(s) = \dfrac{s}{2+5s^2}$，则

$$f'(s) = \frac{(2+5s^2) - 10s^2}{(2+5s^2)^2} = \frac{2-5s^2}{(2+5s^2)^2},$$

当 $s < \sqrt{2/5}$ 时为正，当 $s > \sqrt{2/5}$ 时为负。由于 $\sqrt{2/5} < 1$ 在容许的范围之内，f 在该处取得极大值，$f = \dfrac{\sqrt{2/5}}{4} = \dfrac{\sqrt{10}}{20}$。因此

$$v_{\max}^2 = \frac{15gr}{2}\cdot\frac{\sqrt{10}}{20} = \frac{3\sqrt{10}}{8}gr \approx 1.19\,gr, \qquad \text{当}\ \sin\alpha = \sqrt{2/5}.$$

验算

- 若 $\alpha \to 0$，P 就落在 O 上，只有圆盘悬挂着，OC 是铅垂的：果然 $\tan\beta \to 0$。若 $\alpha = \pi/2$，P 与 O 位于直径的两端，Q 在 OC 上，同样有 $\beta = 0$。
- 取 $\alpha = \pi/2$：$I = \frac{28}{5}mr^2$，质量 $2m$ 下降了 $\frac{3}{2}r$，所以 $v^2 = r^2\cdot 3mgr\cdot\frac{10}{28mr^2} = \frac{15}{14}gr$，与公式吻合。

- 我们就几个不同的 α 值，由释放位置开始对运动方程作数值积分（RK4）；例如当 $\alpha = 0.7$、$r = 1.3$、$g = 9.81$ 时，模拟结果为 $v^2 = 15.1206$，与公式吻合至六位有效数字。直接搜索亦确认极大值为 $3\sqrt{10}/8 \approx 1.1859$（以 gr 为单位）。

常见错误

- I_C 用了均匀圆盘的数值 $\frac{1}{2}mr^2$；这里的密度向外递增，令 I_C 增大至 $\frac{3}{5}mr^2$。
- 在第 (iii) 部分忘记了质点对转动惯量的贡献 $m \cdot OP^2$。
- 答了 ω^2 而不是 v^2：题目问的是 C 的速率，即 $r\omega$。

STEP 之外

绕水平轴转动的刚体（rigid body）就是*复摆*：它的表现就像一个摆锤位于质心的单摆（simple pendulum），只是惯量由 Md^2 换成 I。在第 (i) 部分的平衡位置附近作小幅振动时，周期为 $2\pi\sqrt{I/(Mgd)}$，其中 $d = OQ$；而“等效单摆”的长度 $I/(Md)$ 则引出了 Kater 的可倒摆（reversible pendulum），这种摆曾用来精确量度 g。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）正是以这种思路处理刚体转动和平行轴定理。

详细解答

STEP III 1995，第 10 题

力学 · 斜面上的抛体

STEP III 1995，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/95-s3-q10

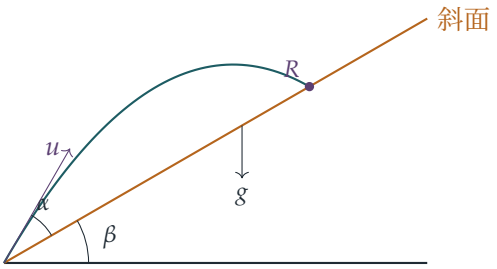
关键思路

在斜面（inclined plane）上，最自然的坐标轴不是水平和垂直方向，而是沿着斜面和垂直于斜面的方向。在这组坐标轴下，重力有两个恒定的分量，飞行时间（time of flight）单凭垂直于斜面的运动便可求得，而射程（range）则是三角函数的乘积。利用积化和差公式（factor formula）把这个乘积化为两个正弦之差，最大值便一目了然，完全不需要微积分。

解答

建立模型

把炮弹视为只受重力作用而自由运动的质点（particle），不计空气阻力，从斜面底部沿最大斜率线（line of greatest slope）向上发射。设发射方向与斜面的夹角为 α，则它相对于水平面的仰角（angle of elevation）为 $\theta = \alpha + \beta$；我们需要 $0 < \alpha < \frac{\pi}{2} - \beta$，炮弹才会真正沿斜面向上飞行。



取 x-轴沿最大斜率线向上，y-轴垂直于斜面并指向离开斜面的一方。把加速度 g（铅垂向

下）沿这两个方向分解，得

$$\ddot{x} = -g\sin\beta, \qquad \ddot{y} = -g\cos\beta,$$

而初速度沿斜面的分量为 $u\cos\alpha$，垂直于斜面的分量为 $u\sin\alpha$。因此在时间 t 时

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

飞行时间与射程

当 $t > 0$ 而再次有 $y = 0$ 时，炮弹着地：

$$T = \frac{2u\sin\alpha}{g\cos\beta}.$$

于是炮弹沿斜面向上所走的距离为

$$\begin{aligned} R = x(T) &= \frac{2u^2\sin\alpha\cos\alpha}{g\cos\beta} - \tfrac{1}{2}g\sin\beta\cdot\frac{4u^2\sin^2\alpha}{g^2\cos^2\beta} \\ &= \frac{2u^2\sin\alpha}{g\cos^2\beta}\big(\cos\alpha\cos\beta - \sin\alpha\sin\beta\big) \\ &= \frac{2u^2\sin\alpha\cos(\alpha+\beta)}{g\cos^2\beta}. \end{aligned}$$

想一想 怎样才能不用求导而求出 $\sin\alpha\cos(\alpha+\beta)$ 的最大值?

在积化和差公式 $2\sin A\cos B = \sin(A+B) + \sin(A-B)$ 中取 $A = \alpha$ 和 $B = \alpha+\beta$，得

$$2\sin\alpha\cos(\alpha+\beta) = \sin(2\alpha+\beta) - \sin\beta,$$

所以

$$R = \frac{u^2\big(\sin(2\alpha+\beta) - \sin\beta\big)}{g\cos^2\beta}.$$

只有 $\sin(2\alpha+\beta)$ 这一项与 α 有关，而它最大只能是 1。当 α 取遍 $(0, \frac{\pi}{2}-\beta)$ 时，角 $2\alpha+\beta$ 取遍 $(\beta, \pi-\beta)$，而这个区间包含 $\frac{\pi}{2}$；所以最大值 1 确实可以取到，而且恰好在 $2\alpha+\beta = \frac{\pi}{2}$ 时取到，即

$$\alpha = \frac{\pi}{4} - \frac{\beta}{2}, \qquad \text{等价地} \qquad \theta = \alpha+\beta = \frac{\pi}{4} + \frac{\beta}{2}.$$

因此大炮应以高于水平面 $\frac{\pi}{4} + \frac{\beta}{2}$ 的角度发射：发射方向平分斜面与铅垂线之间的夹角。

最大射程

取 $\sin(2\alpha+\beta) = 1$，并利用 $\cos^2\beta = 1 - \sin^2\beta = (1-\sin\beta)(1+\sin\beta)$，

$$R_{\max} = \frac{u^2(1-\sin\beta)}{g(1-\sin\beta)(1+\sin\beta)} = \frac{u^2}{g(1+\sin\beta)},$$

其中由于 $\beta < \frac{\pi}{2}$，约去 $1-\sin\beta$ 是合理的。这就是所求的距离。

验算

- 平地，即 $\beta = 0$：最佳角度是 $45°$，射程为 u^2/g，正是大家熟悉的结果。
- 当 $\beta \to \frac{\pi}{2}$（一个铅垂的“斜面”）时，最佳方向趋向垂直向上，而 $R_{\max} \to u^2/(2g)$，即垂直向上发射的炮弹所能达到的最大高度。
- 数值上，取 $u = 20\,\mathrm{m\,s^{-1}}$、$g = 9.81\,\mathrm{m\,s^{-2}}$ 和 $\beta = 30°$，公式给出在仰角 $60°$ 时 $R_{\max} \approx 27.18\,\mathrm{m}$。在水平和垂直坐标下逐步计算炮弹的运动，直至它穿过斜面，对随机选取的 u, β, α 都能重现 $R(\alpha)$；再对 α 作精细搜索，最大值确实出现在 $\frac{\pi}{4} - \frac{\beta}{2}$。

常见错误

- 混淆了与水平面的夹角和与斜面的夹角；答案 $\frac{\pi}{4} + \frac{\beta}{2}$ 是从水平面量起的。
- 遗漏了沿斜面方向的分量 $-g\sin\beta$，结果得到的只是比例出错的平地射程公式。

STEP 之外

这个问题有一个优雅的几何看法：从同一点以速率 u 发射的所有轨迹，都位于一条“安全抛物线”（parabola of safety，即这些轨迹的包络线（envelope））之内，其焦点（focus）就在发射点，而任何斜面上的最大射程，就在斜面与这条抛物线相交之处。利用焦点–准线（directrix）性质，一行便可得出 $R_{\max} = u^2/(g(1+\sin\beta))$，同时解释了为何最佳方向平分斜面与铅垂线之间的夹角。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门）以向量（vector）方法处理抛体（projectile）问题，是进一步学习的好选择。

详细解答

STEP III 1995，第 11 题

力学 · 相对速度

> **STEP III 1995，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/95-s3-q11

关键思路

改用船的参考系（frame of reference）来看。相对于船，飞机必须沿西北方向的直线笔直飞出，再沿同一直线笔直飞回。空气带着飞机一起移动，所以在这个参考系中，飞机的速度等于它的空速（airspeed）向量再加上一股“水流”：风的速度减去船的速度。一旦知道这股水流，每一程都是一道渡河问题。我们选取航向（heading），使合速度（resultant velocity）指向这条直线的方向，然后由来回所需的时间定出距离。

解答

全题取 x 轴向东、y 轴向北，速率以节（knot，即每小时海里数）为单位，于是距离便以海里（nautical mile）为单位。船的速度是 $(-V,0)$。若空气以速度 $\mathbf{w}$ 移动，而飞机相对于空气的速度是 $\mathbf{a}$，其中 $|\mathbf{a}| = kV$，则飞机相对于船的速度（相对速度，relative velocity）是

$$\mathbf{a}+\mathbf{c}, \qquad \mathbf{c} = \mathbf{w} - (-V,0) = \mathbf{w} + (V,0).$$

飞机一直保持在船的西北方，所以 $\mathbf{a}+\mathbf{c}$ 在去程必须指向 $\mathbf{u} = \frac{1}{\sqrt{2}}(-1,1)$ 的方向，在回程则必须指向 $-\mathbf{u}$ 的方向。

有风时：去程

风是由北向南吹的，所以 $\mathbf{w} = (0,-V)$，而

$$\mathbf{c} = (V,-V) = -\sqrt{2}\,V\,\mathbf{u}.$$

这股水流的速率是 $\sqrt{2}\,V$，方向为正东南，恰好沿着巡逻线。因此机师只需把机头对准这条直线。去程（向西北）时她正面逆流而飞；回程时则顺流而飞：

$$\text{去程速率} = (k-\sqrt{2})V, \qquad \text{回程速率} = (k+\sqrt{2})V.$$

这里需要 $k > \sqrt{2}$：否则飞机永远无法沿西北方向的直线离开船。

若飞机飞到距离船 d_w 之处，而每次来回需时 $\frac{1}{2}$ 小时，则

$$\frac{d_w}{(k-\sqrt{2})V} + \frac{d_w}{(k+\sqrt{2})V} = \frac{1}{2}$$
$$\frac{d_w}{V}\cdot\frac{2k}{k^2-2} = \frac{1}{2},$$

所以

$$d_w = \frac{(k^2-2)V}{4k}.$$

因此去程需时

$$\frac{d_w}{(k-\sqrt{2})V} = \frac{k^2-2}{4k(k-\sqrt{2})} = \frac{k+\sqrt{2}}{4k}\ \text{小时} = 15\left(1+\frac{\sqrt{2}}{k}\right)\ \text{分钟}.$$

这个时间必定多于 15 分钟，正合预期：去程是逆流而行的。

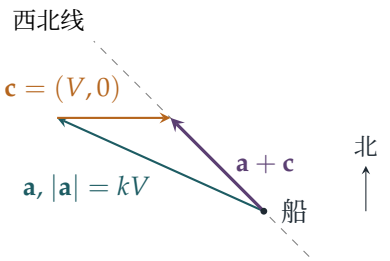
想一想 在无风的情况下，飞机为甚么不能同样只把机头对准西北？

无风时，水流不再沿着巡逻线。假如机师把机头对准西北，水流中向东的部分便会把飞机带离这条直线。她必须把机头稍为偏向迎着水流的一方，而这会耗去她一部分空速。

无风时

现在 $\mathbf{w} = \mathbf{0}$，所以 $\mathbf{c} = (V, 0)$：相对于船，空气以速率 V 向东流动。把 $\mathbf{c}$ 分解为沿直线的分量（component）和垂直于直线的分量，前者用 $\mathbf{u}$，后者用 $\mathbf{v} = \frac{1}{\sqrt{2}}(1,1)$：

$$\mathbf{c} = -\frac{V}{\sqrt{2}}\,\mathbf{u} + \frac{V}{\sqrt{2}}\,\mathbf{v}.$$



合速度不能有 $\mathbf{v}$ 分量，所以 $\mathbf{a}$ 的 $\mathbf{v}$ 分量必须是 $-V/\sqrt{2}$。由于 $|\mathbf{a}| = kV$，它沿行进方向的分量便是

$$\sqrt{k^2V^2 - \tfrac{1}{2}V^2} = V\sqrt{k^2 - \tfrac{1}{2}}.$$

沿着这条直线，水流在去程的贡献是 $-V/\sqrt{2}$，在回程则是 $+V/\sqrt{2}$（因为 $\mathbf{c}\cdot(-\mathbf{u}) = V/\sqrt{2}$）。设 $A = V\sqrt{k^2 - \tfrac{1}{2}}$ 及 $B = V/\sqrt{2}$，则

$$\text{去程速率} = A - B, \qquad \text{回程速率} = A + B.$$

去程速率是正数，因为 $A^2 - B^2 = V^2(k^2 - 1) > 0$。

来回同样需时半小时，于是

$$\frac{d_c}{A-B} + \frac{d_c}{A+B} = \frac{2A\,d_c}{A^2-B^2} = \frac{1}{2} \quad\Longrightarrow\quad d_c = \frac{A^2-B^2}{4A} = \frac{V(k^2-1)}{4\sqrt{k^2-\tfrac{1}{2}}}.$$

比值

两式相除，

$$\frac{d_w}{d_c} = \frac{(k^2-2)V}{4k}\cdot\frac{4\sqrt{k^2-\tfrac{1}{2}}}{V(k^2-1)} = \frac{(k^2-2)\sqrt{k^2-\tfrac{1}{2}}}{k(k^2-1)}.$$

由于 $\sqrt{k^2 - \tfrac{1}{2}} = \tfrac{1}{2}\sqrt{4k^2-2}$，

$$\frac{d_w}{d_c} = \frac{k^2-2}{2k(k^2-1)}\sqrt{4k^2-2},$$

这正是所求的结果。

验算

- 当 $k \to \infty$ 时，风和船几乎无关重要。去程时间趋向 15 分钟，而 $d_w/d_c \to 1$（比值的分子和分母都以 $2k^3$ 的速度增长）。
- 当 $k \to \sqrt{2}^{+}$ 时，有风时的去程速率降至零，所以 $d_w \to 0$，比值 $\to 0$，与公式相符。
- 当 $k = 2$ 时，去程需时 $15(1 + 1/\sqrt{2}) \approx 25.6$ 分钟，而 $d_w/d_c = \tfrac{2}{12}\sqrt{14} \approx 0.624$。我们直接解出航向，并逐步模拟整个运动，以数值方法确认了这个结果。

常见错误

- 把风向弄反：“由北向南”的风是*朝着*南方吹的。

- 把船的速度加上去，而不是减去。相对于一艘向西航行的船，静止的空气是向东漂移的。
- 在无风时把机头对准西北，并以 $kV \pm V/\sqrt{2}$ 作为速率。必须先抵销横向的水流。

STEP 之外

这是一次伽利略参考系变换（Galilean change of reference frame）：在彼此作匀速相对运动的参考系中量度所得的速度，只相差一个常向量。迈克生—莫雷（Michelson–Morley）的“河中小船”论证，背后也是同一套计算。在那个论证中，横渡来回与顺逆流来回所需的时间并不相同，情况与本题如出一辙，而这正是后来被狭义相对论（special relativity）推翻的经典计算。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教材）正是以这种思路先讲相对运动，再讲相对论。

详细解答

STEP III 1995，第 12 题

概率与统计 • 径向对称密度函数

STEP III 1995，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/95-s3-q12

本题在 *Beyond STEP* 第 12.22 节没有平均值的分布中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/cauchy.html

关键思路

径向对称（radially symmetrical）的密度只通过与原点的距离 $r = \sqrt{u^2+v^2}$ 而取决于 (u,v)，所以绕原点作任何旋转之后，它看起来都一样。特别地，点落在以原点为顶点的扇形内的概率，只取决于该扇形的角度。事件 $U/V < k$ 是由直线 $u = kv$ 与 u 轴割出的一对相对的扇形，所以它的概率不过是一个角度除以 2π。

解答

联合密度与径向对称

X 和 Y 各自的密度为 $\phi(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$。由于 X 与 Y 独立（independent），它们的联合密度（joint density）是边际密度（marginal density）的乘积：

$$f(x,y) = \phi(x)\,\phi(y) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\cdot\frac{1}{\sqrt{2\pi}}e^{-y^2/2} = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)},$$

对所有实数 x 和 y 成立，这正是所求。

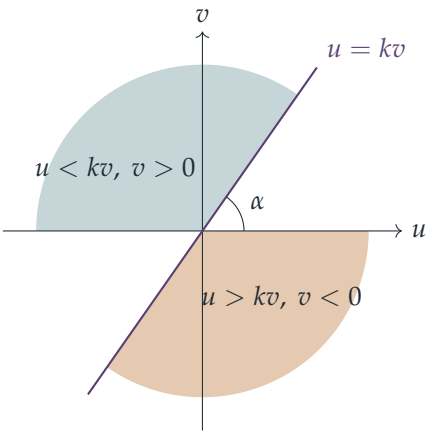
说这个密度是*径向对称*的，意思是它在点 (x,y) 的值只取决于该点与原点的距离 $r = \sqrt{x^2+y^2}$，而与该点的方向无关。这里 $f(x,y) = \frac{1}{2\pi}e^{-r^2/2}$。换一个等价的说法：把平面绕原点作任何旋转，密度都不变，而它的等高线是以原点为圆心的圆。

关于 $\mathrm{P}(U/V < k)$ 的恒等式

设 U 和 V 具有径向对称的联合密度 $h(u,v) = \rho\left(\sqrt{u^2+v^2}\right)$。直线 $v=0$ 的概率为零（它的面积为零），所以我们可以不理会 $V=0$ 的情况。把不等式 $U/V < k$ 乘以 V，并在 $V<0$ 时把不等号反转，得

$$\left\{\frac{U}{V} < k\right\} = \{U < kV,\ V > 0\} \cup \{U > kV,\ V < 0\},$$

而右边的两个事件是互斥（disjoint）的。



如图所示（图中取 $k>0$），通过原点的直线 $u=kv$ 与 u 轴把平面分成四个扇形，其中两个涂上阴影的扇形合起来就是事件 $U/V<k$。映射 $(u,v)\mapsto(-u,-v)$ 是绕原点旋转 π，它把区域 $\{u<kv,\ v>0\}$ 恰好映到 $\{u>kv,\ v<0\}$ 上（把两个不等式都乘以 -1 即可）。径向对称的密度满足 $h(-u,-v)=h(u,v)$，所以这两个区域的概率相同。因此

$$\mathrm{P}\left(\frac{U}{V} < k\right) = 2\,\mathrm{P}(U < kV,\ V > 0),$$

这就是我们要证明的结果。

U/V 的密度

采用极坐标（polar coordinates）$u = r\cos\theta$，$v = r\sin\theta$。区域 $\{u<kv,\ v>0\}$ 是扇形 $\alpha<\theta<\pi$，其中 $\alpha\in(0,\pi)$ 是射线 $u=kv$，$v>0$ 与正 u 轴所成的角，所以 $\cot\alpha = k$。（事实上，当 $0<\theta<\pi$ 时有 $\sin\theta>0$，于是 $u<kv$ 变成 $\cot\theta<k$；由于 cot 在 $(0,\pi)$ 上递减，这就是 $\theta>\alpha$。）

> **想一想** 在不知道函数 ρ 的情况下，(U,V) 落在以原点为顶点、角度为 β 的扇形内的概率是多少？

在极坐标中，面积元素是 $r\,\mathrm{d}r\,\mathrm{d}\theta$，所以扇形 $\theta_1 < \theta < \theta_1 + \beta$ 的概率为

$$\int_{\theta_1}^{\theta_1+\beta}\int_0^\infty \rho(r)\, r\,\mathrm{d}r\,\mathrm{d}\theta = \beta\int_0^\infty \rho(r)\, r\,\mathrm{d}r = \frac{\beta}{2\pi},$$

因为取 $\beta = 2\pi$ 便得到整个平面，所以 $2\pi\int_0^\infty \rho(r)\, r\,\mathrm{d}r = 1$。换句话说，无论 ρ 是甚么，(U, V) 的方向都是均匀分布（uniformly distributed）的。

我们的扇形的角度是 $\pi - \alpha$，所以

$$\mathrm{G}(k) = \mathrm{P}\Big(\frac{U}{V} < k\Big) = 2\cdot\frac{\pi-\alpha}{2\pi} = 1 - \frac{\alpha}{\pi}.$$

由于 $\cot\alpha = k$ 且 $\alpha \in (0, \pi)$，我们有 $\alpha = \frac{\pi}{2} - \arctan k$（两边都在 $(0,\pi)$ 内，而且余切都等于 k，因为 $\cot(\frac{\pi}{2} - \psi) = \tan\psi$）。因此

$$\mathrm{G}(k) = \frac{1}{2} + \frac{1}{\pi}\arctan k, \qquad -\infty < k < \infty.$$

它由 0 上升至 1，正如分布函数（distribution function）所必须的那样。求导便得到 U/V 的密度：

$$\mathrm{g}(k) = \mathrm{G}'(k) = \frac{1}{\pi(1+k^2)}, \qquad -\infty < k < \infty,$$

这正是所求。特别地，这个结果适用于第一部分中两个独立标准正态变量（standard normal）之比 X/Y。

验算

- $\mathrm{G}(0) = \frac{1}{2}$：由对称性，$U/V$ 取正值与取负值的可能性相同。
- $\int_{-\infty}^{\infty} \frac{\mathrm{d}k}{\pi(1+k^2)} = \frac{1}{\pi}\big[\arctan k\big]_{-\infty}^{\infty} = 1$。
- 用三个不同的径向对称分布（两个独立的标准正态变量、单位圆盘上的均匀分布，以及与 e^{-r} 成正比的密度）各取 2×10^5 个点进行模拟，每次都得到 $\mathrm{P}(U/V < 0.7) \approx 0.694$，与 $\frac{1}{2} + \frac{1}{\pi}\arctan 0.7 \approx 0.6944$ 吻合；对于若干个 k 值，恒等式两边的差距都在抽样误差之内。

常见错误

- 写成 $\{U/V < k\} = \{U < kV\}$：当 $V < 0$ 时，乘以 V 会令不等号反转。
- 不加细想便取 $\alpha = \arctan(1/k)$：当 $k < 0$ 时它落在 $(-\frac{\pi}{2}, 0)$ 内，而我们需要的角却在 $(0,\pi)$ 内。写成 $\alpha = \frac{\pi}{2} - \arctan k$ 则对每个 k 都适用。

STEP 之外

密度 $\frac{1}{\pi(1+k^2)}$ 就是标准柯西分布（*Cauchy distribution*）。它以没有平均值而闻名（$|k|\,\mathrm{g}(k)$ 的积分发散），所以大数定律（law of large numbers）对它并不成立：n 个独立柯西变量的平均值，与其中任何一个变量具有相同的柯西分布。这里的极坐标论证，让你初尝以雅可比行列式（Jacobian）变换联合密度的滋味，这是大学一年级的标准技巧。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）对两者都有论述。

详细解答

STEP III 1995，第 13 题

概率与统计 · 有噪声的二元信道

STEP III 1995，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：

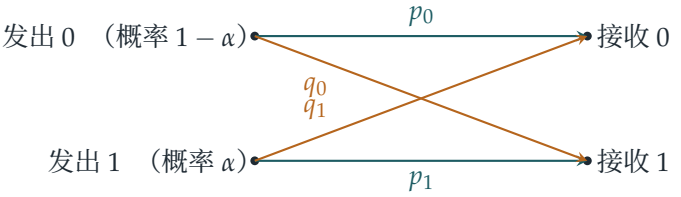
step.maths.org/questions/95-s3-q13

关键思路

一切都由逐个数字考虑而来。每个数字都经过一条“二元信道”（binary channel）：它以 0 或 1 发出，然后要么原样保留，要么被翻转。一旦弄清楚单一数字的情况，整段信息不过是许多独立的复本，所以概率相乘即可。至于增强站，窍门在于留意到：一个数字经过两段电缆后正确到达，当且仅当它被翻转了偶数次。

解答

全篇以 $N = 10^k$ 表示信息的长度。我们按照题目的原意，假设原信息的各个数字互相独立（independent）（每个数字是 1 的概率为 α），并假设电缆对每个数字的处理与其他数字无关。单一数字的情况如下图所示。



整段信息正确接收

设 C 为某个指定数字被正确接收的事件。这个数字要么是保留下来的 0，要么是保留下来的 1，所以

$$\mathrm{P}(C) = (1-\alpha)p_0 + \alpha p_1.$$

由于 N 个数字互相独立，

$$\mathrm{P}(\text{整段信息正确}) = \left[(1-\alpha)p_0 + \alpha p_1\right]^{10^k}.$$

概率 β

一个数字被接收为 1，要么它原本是 1 而且保留下来，要么它原本是 0 而被翻转：

$$\beta = \alpha p_1 + (1-\alpha)q_0.$$

现在

$$\begin{aligned}\beta = \alpha &\iff \alpha p_1 + (1-\alpha)q_0 = \alpha \iff (1-\alpha)q_0 = \alpha(1-p_1) = \alpha q_1\\ &\iff q_0 = \alpha(q_0+q_1) \iff \alpha = \frac{q_0}{q_0+q_1},\end{aligned}$$

其中最后一步除以 q_0+q_1，而除非电缆完美无误（$q_0 = q_1 = 0$，此时 $\beta = \alpha$ 显然成立），否则这个数必为正数。每一步都可以逆推，所以这就是所要求的“当且仅当”（if and only if）。用文字来说：电缆不改变 1 所占的比例，恰好就是 $0 \to 1$ 的流量（即 $(1-\alpha)q_0$）与 $1 \to 0$ 的流量（即 αq_1）互相抵销之时。

全是零的信息

假设上述条件成立，而且接收到的每个数字都是 0。由于各个位置上的（发出数字，接收数字）这些数对互相独立，整段信息正确的条件概率（conditional probability）就是各个单一数字的概率 $\mathrm{P}(\text{发出0} \mid \text{接收0})$ 的乘积。

> ? **想一想** 怎样把“接收到 0”转化为一句关于发出了甚么的陈述?

用贝叶斯定理（Bayes' theorem）。一个数字被接收为 0 的概率是 $1-\beta$，而条件说明 $\beta = \alpha$，所以 $1-\beta = 1-\alpha$。因此

$$\mathrm{P}(\text{发出0} \mid \text{接收0}) = \frac{(1-\alpha)p_0}{1-\beta} = \frac{(1-\alpha)p_0}{1-\alpha} = p_0,$$

于是

$$\mathrm{P}(\text{信息正确} \mid \text{接收到的全是零}) = p_0^{10^k}.$$

一半一半的机会

当 $q_0 = q_1 = q$ 且 $\alpha = \frac{1}{2}$ 时，每个数字正确的概率是 $\frac{1}{2}(1-q) + \frac{1}{2}(1-q) = 1-q$。对于一段有 10^6 个数字的信息，我们需要

$$(1-q)^{10^6} = \tfrac{1}{2} \quad\Longrightarrow\quad 10^6 \ln(1-q) = -\ln 2.$$

由于 q 极小，$\ln(1-q)\approx -q$，得出

$$q \approx \frac{\ln 2}{10^6} \approx 6.93\times 10^{-7}.$$

加设增强站

长度减半会使 q 除以 $2^2=4$，因为 q 与长度的平方成正比。所以每半段电缆翻转一个数字的概率是 $q/4$，而且与另一半段无关（增强站只是把收到的东西原样再发出）。

> **想一想** “两半段都正确传送”是数字正确到达的唯一途径吗?

不是：一个在两半段都被翻转的数字会被翻回原状。所以

$$\mathrm{P}(\text{数字正确}) = \left(1-\frac{q}{4}\right)^2 + \left(\frac{q}{4}\right)^2 = 1-\frac{q}{2}+\frac{q^2}{8} \approx 1-\frac{q}{2},$$

这里略去了 q^2 阶的项。（在这里，双重翻转正是这类项之一，所以它对答案没有影响，但明白个中原因仍然值得。）整段信息正确的概率现在是

$$\left(1-\frac{q}{2}\right)^{10^6}.$$

在同一阶的近似下，$(1-\frac{q}{2})^2 \approx 1-q$，所以

$$\left(1-\frac{q}{2}\right)^{10^6} \approx (1-q)^{10^6/2} = \left[(1-q)^{10^6}\right]^{1/2} = \left(\tfrac{1}{2}\right)^{1/2} = \frac{1}{\sqrt{2}} \approx 0.707.$$

所以增强站把信息完全正确的机会由 $\frac{1}{2}$ 提高到约 0.71。留意我们在这里根本用不着 q 的数值：原来的条件 $(1-q)^{10^6}=\frac{1}{2}$ 已经包办了一切。同样的想法配合 $e^{-10^6 q}=\frac{1}{2}$，可得 $e^{-10^6 q/2}=1/\sqrt{2}$。

验算

- 把长度为 4 的所有发出信息和接收信息逐一准确列举（取 $p_0=0.9$，$p_1=0.8$），可以重现 $[(1-\alpha)p_0+\alpha p_1]^4$，以及在 α 满足该条件时的条件概率 $p_0^4=0.6561$。
- 准确解出 $(1-q)^{10^6}=\frac{1}{2}$ 可得 $q=6.9315\times 10^{-7}$，与 $\ln 2/10^6$ 吻合至五位有效数字。
- 取这个 q 值，增强站情况下的准确概率 $\left[(1-\frac{q}{4})^2+(\frac{q}{4})^2\right]^{10^6}$ 是 0.707107，与 $1/\sqrt{2}$ 相符。
- 合理性检查：若 $q_0=q_1$，条件给出 $\alpha=\frac{1}{2}$，正如对称性所要求。

常见错误

- 把每半段的出错概率当作 $q/2$。出错概率与长度的平方成正比，所以应是 $q/4$。

- 在“全是零”的部分只答 $(1-\alpha)^N$ 或单单一个 p_0：我们需要的是每个数字的条件概率，再取 10^k 次方。

STEP 之外

这就是信息论（information theory）中的二元信道。当 $q_0 = q_1$ 时，它就是二元对称信道（binary symmetric channel），是香农（Shannon）理论的起点——这套理论研究有噪声的信道能够传送多少信息；它也是纠错码（error-correcting codes）的起点，纠错码加入冗余，使偶尔出现的翻转可以被侦测和修复，而不仅仅是变得较为罕见。增强站的计算也是一条两步的马可夫链（Markov chain）：翻转概率的复合，就是把转移矩阵（transition matrices）相乘。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）以恰到好处的程度讲解条件概率和马可夫链。

详细解答

STEP III 1995，第 14 题

概率与统计 · 矩生成函数

> **STEP III 1995，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/95-s3-q14

关键思路

矩母函数（moment generating function，简称 MGF）能把独立随机变量（independent random variables）之和化为乘积：若 T_1 与 T_2 独立，则 $\mathrm{E}\left[\mathrm{e}^{\theta(T_1+T_2)}\right] = \mathrm{E}\left[\mathrm{e}^{\theta T_1}\right]\mathrm{E}\left[\mathrm{e}^{\theta T_2}\right]$。因此我们先求出一个 MGF，把它自乘若干次，再认出所得结果正是同一族中某个密度函数（density）的 MGF。由于 MGF 决定其分布，认出它便已足够。

解答

T 的 MGF

对实数 θ，

$$M_T(\theta) = \mathrm{E}\left[\mathrm{e}^{\theta T}\right] = \int_0^\infty t\,\mathrm{e}^{-(1-\theta)t}\,\mathrm{d}t.$$

这个积分只在 $\theta < 1$ 时收敛（否则被积函数不趋向 0），所以下文一律假设 $\theta < 1$。我们会反复用到以下积分：

$$\int_0^\infty t^k \mathrm{e}^{-\lambda t}\,\mathrm{d}t = \frac{k!}{\lambda^{k+1}} \qquad (\lambda > 0,\ k = 0, 1, 2, \ldots),$$

这可由分部积分法（integration by parts）连做 k 次得出（或代入 $s = \lambda t$，再利用 $\int_0^\infty s^k \mathrm{e}^{-s}\,\mathrm{d}s = k!$）。取 $k = 1$ 及 $\lambda = 1 - \theta > 0$，得

$$\boxed{M_T(\theta) = \frac{1}{(1-\theta)^2}}, \qquad \theta < 1.$$

简单核对一下：$M_T(0) = 1$，这是每个 MGF 都必须满足的。

U 的 MGF

设 T_1 和 T_2 为做这两题所用的时间，则 $U = T_1 + T_2$，其中 T_1 和 T_2 的密度函数均为 f，而且互相独立（题目已说明）。于是 $\mathrm{e}^{\theta T_1}$ 与 $\mathrm{e}^{\theta T_2}$ 亦互相独立，所以它们乘积的期望值（expectation）等于各自期望值的乘积：

$$M_U(\theta) = \mathrm{E}\left[\mathrm{e}^{\theta T_1}\mathrm{e}^{\theta T_2}\right] = M_T(\theta)^2 = \frac{1}{(1-\theta)^4}, \qquad \theta < 1.$$

U 的密度函数

现在计算题目所给的密度函数 $\mathrm{g}(u) = \frac{1}{6}u^3\mathrm{e}^{-u}$ 的 MGF。在上述积分中取 $k = 3$ 及 $\lambda = 1-\theta$，

$$\int_0^\infty \mathrm{e}^{\theta u}\,\tfrac{1}{6}u^3\mathrm{e}^{-u}\,\mathrm{d}u = \frac{1}{6}\cdot\frac{3!}{(1-\theta)^4} = \frac{1}{(1-\theta)^4}.$$

（取 $\theta = 0$ 可确认 g 的积分为 1，所以它确实是一个密度函数。）

> **想一想** 我们已证明 g 与 U 有相同的 MGF。为甚么这就表示 U 的密度函数是 g?

这里用到 MGF 的唯一性定理（uniqueness theorem）：若两个随机变量的 MGF 在包含 $\theta = 0$ 的某个开区间上存在而且相等，则两者有相同的分布。这里两个 MGF 对所有 $\theta < 1$ 都相等，当中包括区间 $(-1, 1)$。所以 U 的密度函数为

$$\mathrm{g}(u) = \frac{1}{6}u^3\mathrm{e}^{-u}, \qquad u \geq 0,$$

正如题目所求。

如果你不想使用唯一性定理，卷积（convolution）公式可直接给出同一结果。由于 $T_1, T_2 \geq 0$，对 $u \geq 0$，

$$\begin{aligned}\mathrm{g}(u) &= \int_0^u \mathrm{f}(s)\,\mathrm{f}(u-s)\,\mathrm{d}s = \mathrm{e}^{-u}\int_0^u s(u-s)\,\mathrm{d}s \\ &= \mathrm{e}^{-u}\left[\tfrac{1}{2}us^2 - \tfrac{1}{3}s^3\right]_0^u = \frac{1}{6}u^3\mathrm{e}^{-u}.\end{aligned}$$

做 n 题所用的总时间

设 $U_n = T_1 + \cdots + T_n$，其中各 T_i 互相独立，密度函数均为 f。同样的乘积论证给出

$$M_{U_n}(\theta) = M_T(\theta)^n = \frac{1}{(1-\theta)^{2n}}, \qquad \theta < 1.$$

由 $n = 2$ 的规律可以猜想密度函数为 $\mathrm{h}_n(u) = c\,u^{2n-1}\mathrm{e}^{-u}$。利用上述积分并取 $k = 2n-1$，它的 MGF 是 $c\,(2n-1)!/(1-\theta)^{2n}$。这恰好在 $c = 1/(2n-1)!$ 时等于 M_{U_n}。再次根据

唯一性，做 n 题所用的总时间的密度函数为

$$\boxed{\mathrm{h}_n(u) = \frac{u^{2n-1}\mathrm{e}^{-u}}{(2n-1)!}, \qquad u \geq 0.}$$

当 $n = 1$ 时这就是 f，当 $n = 2$ 时这就是 g，与预期一致。

验算

- $\mathrm{E}[T] = M_T'(0) = 2$，与 $\int_0^\infty t^2\mathrm{e}^{-t}\,\mathrm{d}t = 2$ 吻合。所以 $\mathrm{E}[U_n] = 2n$。由 h_n 亦得出同一数值：$\int_0^\infty u^{2n}\mathrm{e}^{-u}\,\mathrm{d}u/(2n-1)! = (2n)!/(2n-1)! = 2n$。
- 数值上，用森逊法则（Simpson's rule）可得 $\int_0^\infty \mathrm{e}^{0.5t}\,t\mathrm{e}^{-t}\,\mathrm{d}t = 4.000$，与 $1/(1-0.5)^2$ 吻合。模拟 2×10^5 次，每次从 f 抽取三个时间并求和，得出 $\mathrm{P}(U_3 \leq 5) \approx 0.386$，而由 h_3 算得 $\int_0^5 u^5\mathrm{e}^{-u}/120\,\mathrm{d}u \approx 0.384$。

常见错误

- 忘记 $\theta < 1$ 这个限制：当 $\theta \geq 1$ 时 MGF 并不存在。
- 把 MGF 相加而不是相乘。U 的 MGF 是 $M_T(\theta)^2$，而不是 $2M_T(\theta)$。（$2T$ 的 MGF 则是 $M_T(2\theta)$，又是另一回事。）
- 弄错归一化常数（normalising constant）中的阶乘。u 的幂是 $2n-1$，所以常数是 $1/(2n-1)!$，而不是 $1/(2n)!$。

STEP 之外

密度函数 $t\mathrm{e}^{-t}$ 是形状参数（shape）为 2、速率参数（rate）为 1 的伽玛分布（gamma distribution），而 h_n 则是形状参数为 $2n$ 的伽玛分布。这道题说明：速率相同的独立伽玛变量相加时，形状参数亦相加。特别地，每个 T 的表现就如两个独立指数分布（exponential）时间之和，而伽玛分布正是这样作为泊松过程（Poisson process）中第 k 个事件的等候时间而出现的。在大学里，你会遇到特征函数（characteristic function）$\mathrm{E}[\mathrm{e}^{\mathrm{i}\theta X}]$，它必定存在，并可用来证明中心极限定理（central limit theorem）。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）对母函数（generating functions）及这些概念有清晰的阐述。

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

1995 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。

