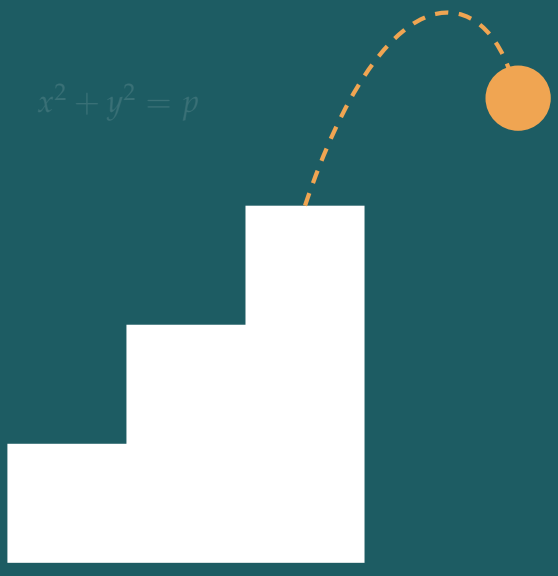
STEP 备试 · 大学数学

# Beyond STEP

STEP I, II 及 III

2018 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

2018 年详细解答

Keith McNulty

**Beyond STEP：以历届 STEP 试题为门，走进大学数学**

**2018 年详细解答**

2018 年 STEP 试卷（STEP I, II 及 III）全部 39 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 曲线之间的面积 纯数学 . . . 3

第 2 题 对数不等式 纯数学 . . . 7

第 3 题 角与双曲线 纯数学 . . . 10

第 4 题 曲线描绘与积分 纯数学 . . . 14

第 5 题 多项式与因式定理 纯数学 . . . 18

第 6 题 中点法则与梯形法则 纯数学 . . . 22

第 7 题 解三次方程 纯数学 . . . 26

第 8 题 耦合微分方程 纯数学 . . . 30

第 9 题 斜面上的运动 力学 . . . 34

第 10 题 连接质点 力学 . . . 38

第 11 题 斜面上的摩擦力 力学 . . . 42

第 12 题 放回抽样 概率与统计 . . . 46

第 13 题 二项分布与贝叶斯定理 概率与统计 . . . 50


## STEP II


第 1 题 多项式的根 纯数学 . . . 54

第 2 题 凹函数 纯数学 . . . 59

第 3 题 旋转对称与积分 纯数学 . . . 63

第 4 题 三角方程 纯数学 . . . 67

第 5 题 级数的积分 纯数学 . . . 71

第 6 题 阶乘与质数 纯数学 . . . 74

第 7 题 几何中的向量 纯数学 . . . 77

第 8 题 伯努利微分方程 纯数学 . . . 80

第 9 题 重力作用下的碰撞 力学 . . . 84

第 10 题 悬链长绳上的运动学 力学 88
第 11 题 摩擦力与力矩 力学 92
第 12 题 最优停止的期望值 概率与统计 96
第 13 题 递推关系 概率与统计 100

## STEP III

第 1 题 二次方程的根与不等式 纯数学 103
第 2 题 埃尔米特多项式 纯数学 107
第 3 题 欧拉微分方程 纯数学 111
第 4 题 双曲线的切线 纯数学 115
第 5 题 算术几何平均不等式 纯数学 119
第 6 题 复数与几何 纯数学 122
第 7 题 棣美弗定理与级数 纯数学 126
第 8 题 小数部分的积分 纯数学 130
第 9 题 碰撞与递推关系 力学 134
第 10 题 复摆 力学 138
第 11 题 松弛绳与抛体 力学 142
第 12 题 顺序统计量 概率与统计 146
第 13 题 瑞利分布 概率与统计 150

详细解答

# STEP I 2018，第 1 题

纯数学 · 曲线之间的面积

**STEP I 2018，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/18-s1-q1

## 关键思路

令三次曲线（cubic）与直线相等，会先得到因式 $x$，余下的是一个平方差（difference of two squares），因此三个交点都能干净利落地求出。之后的一切都是常规的微积分运算，而代换（substitution）$c = b - a$（即 $P$ 的 $x$ 坐标）能令代数运算保持整洁。最后的不等式乍看令人生畏，但只要把两个面积都以 $c$ 表示，它便简化为 $b > a$。

## 解答

### 交点与草图

在交点处，$x(b-x)^2 = a^2x$，所以 $x = 0$（即原点），或

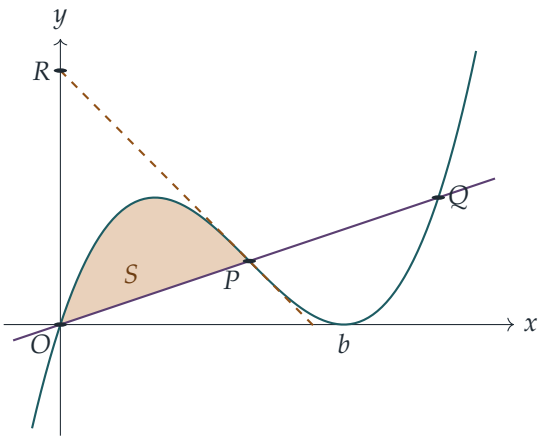
$$(b-x)^2 = a^2 \quad \Longrightarrow \quad b - x = \pm a \quad \Longrightarrow \quad x = b - a \text{ 或 } x = b + a.$$

由于 $a > 0$，$b - a < b + a$，所以

$$P = (b-a,\ a^2(b-a)), \qquad Q = (b+a,\ a^2(b+a)).$$

因为 $0 < a < b$，两点都位于第一象限（first quadrant）。

至于草图：三次曲线 $y = x(b-x)^2$ 在 $O$ 有一个单根，在 $x = b$ 有一个重根（repeated root），曲线在该处与 $x$ 轴相切。当 $x < 0$ 时曲线为负，当 $x > 0$ 时曲线为正（$x = b$ 除外），而当 $x \to \infty$ 时曲线趋向 $+\infty$。其导数（见下文）为 $(b-x)(b-3x)$，所以曲线在 $x = b/3$ 有局部极大值（local maximum），在 $(b, 0)$ 有局部极小值（local minimum）。经过 $O$ 的直线在 $O$ 附近切过上升的一段，再在 $P$ 相交（当 $a < \frac{2}{3}b$ 时 $P$ 位于 $b/3$ 与 $b$ 之间，否则位于 $b/3$ 之前），然后在切点 $x = b$ 之后的 $Q$ 第三次与曲线相交。图中取 $a = 1$、$b = 3$；图中亦显示了 $P$ 处的切线、点 $R$ 以及面积为 $S$ 的区域。




**$P$ 处的切线**

以乘积法则（product rule）求导，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = (b-x)^2 - 2x(b-x) = (b-x)(b-3x).$$

在 $P$，$x = b-a$，所以 $b-x = a$，$b-3x = 3a-2b$。该处的斜率（gradient）为 $m = a(3a-2b)$。

切线（tangent）为 $y = mx+k$，其中 $k$ 的取值须令直线经过 $P$：

$$\begin{aligned} k &= a^2(b-a) - a(3a-2b)(b-a) \\ &= a(b-a)(a-3a+2b) \\ &= 2a(b-a)^2. \end{aligned}$$

因此 $P$ 处的切线为 $y = a(3a-2b)x + 2a(b-a)^2$，与所求相符。特别地，$R = (0,\, 2a(b-a)^2)$，位于 $y$ 轴的正半轴上。

**面积 $S$**

记 $c = b-a > 0$ 为 $P$ 的 $x$ 坐标。在 $0 < x < c$ 上，

$$x(b-x)^2 - a^2x = x\,(b-x-a)(b-x+a) > 0,$$

因为三个因式在该区间上都是正数。所以在 $O$ 与 $P$ 之间曲线位于直线之上，而

$$S = \int_0^c (x(b-x)^2 - a^2x)\,\mathrm{d}x = \int_0^c ((b^2-a^2)x - 2bx^2 + x^3)\,\mathrm{d}x = \frac{(b^2-a^2)c^2}{2} - \frac{2bc^3}{3} + \frac{c^4}{4}.$$

现代入 $b^2-a^2=c(b+a)=c(c+2a)$ 及 $b=c+a$：

$$\begin{aligned}S&=\frac{c^3(c+2a)}{2}-\frac{2c^3(c+a)}{3}+\frac{c^4}{4}\\&=c^3\Big[c(\tfrac12-\tfrac23+\tfrac14)+a(1-\tfrac23)\Big]\\&=c^3\Big(\frac{c}{12}+\frac{a}{3}\Big)=\frac{c^3(c+4a)}{12}.\end{aligned}$$

由于 $c+4a=b+3a$，上式即为

$$S=\frac{1}{12}(b-a)^3(3a+b),$$

与所求相符。

> ? **想一想** 为何不直接把所有东西以 $a$ 和 $b$ 展开?

当然可以，但目标式含有因式 $(b-a)^3$，要从一条关于 $a$ 和 $b$ 的四次式中把它抽出来十分费劲。改写为 $c=b-a$ 后，这个因式会自动出现，因为积分中的每一项本身已带有 $c^2$ 或更高次。

### 三角形 $T$ 与不等式

以 $OR$ 作为三角形 $OPR$ 的底：它位于 $y$ 轴上，长度为 $2ac^2$。垂直高度是 $P$ 到 $y$ 轴的距离，即 $c$。所以

$$T=\tfrac12\cdot 2ac^2\cdot c=ac^3=a(b-a)^3.$$

因此，由于 $(b-a)^3>0$，

$$S-\tfrac13 T=\frac{(b-a)^3}{12}(3a+b-4a)=\frac{(b-a)^4}{12}>0,$$

因为 $b\neq a$。故 $S>\frac13 T$，与所求相符。

## 验算

- 取 $a=1$、$b=3$（即图中情况）：$P=(2,2)$，$Q=(4,4)$，而曲线在 $x=2$ 时给出 $y=2$，在 $x=4$ 时给出 $y=4$。切线公式给出 $y=-3x+8$；斜率公式 $(3-x)(3-3x)$ 在 $x=2$ 时等于 $-3$。直接积分，$S=\int_0^2(x^3-6x^2+8x)\,\mathrm{d}x=4-16+16=4$，与 $\frac{1}{12}\cdot 8\cdot 6=4$ 吻合。此外 $T=8$，而 $4>\frac83$。
- 比值 $S/(\frac13 T)=(3a+b)/(4a)$ 在 $b\to a$ 时趋向 1，所以常数 $\frac13$ 已是最佳可能的。这是合理的：对于抛物线 $y=px-qx^2$，同样的构造给出 $S=qc^3/6$ 及 $T=qc^3/2$，所以 $S=\frac13 T$ 恰好成立；而当 $P$ 非常接近 $O$ 时，两者之间的三次曲线几乎就是一条抛物线。
- 对数以千计随机取得的 $0<a<b$，以电脑检验交点、切线、面积（以数值积分计算）及不等式，结果均与公式一致。

## 常见错误

- 把 $x(b-x)^2 = a^2x$ 除以 $x$ 而没有记下根 $x = 0$，或只取 $b - x = a$ 而遗漏了 $Q$。
- 以直线减曲线来积分，得出负的“面积”：积分前先确认哪个图像在上方。
- 画草图时把三次曲线画成在 $x = b$ 穿过 $x$ 轴，而不是在该处与 $x$ 轴相切。

## STEP 之外

最后一部分中比值的极限为 1，这与阿基米德（Archimedes）的抛物线求积（quadrature of the parabola）一脉相承：对于抛物线，弓形的面积恰好是三角形 $OPR$ 面积的三分之一；而在 $O$ 附近，三次曲线局部上就是一条抛物线，所以不等式在极限下收紧为等式。理解局部行为为何由泰勒（Taylor）展开的头几项决定，并把“几乎是一条抛物线”说得严谨，正是大学一年级分析学（analysis）的课题。Abbott 的 *Understanding Analysis*（《理解分析学》）是一本亲切易读的书，可以看到泰勒定理（Taylor's theorem）的严谨处理。

详细解答

# STEP I 2018，第 2 题

纯数学 · 对数不等式

> **STEP I 2018，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/18-s1-q2

## 关键思路

一切都建基于换底公式（change-of-base rule）：任何底数的对数（logarithm）都可以写成同一底数下两个对数之比，而且 $\frac{1}{\log_b c} = \log_c b$。只要把每个对数都改写成以 ln 表示，每一部分就变成一小段涉及不等式（inequality）的代数运算。题目给出的事实，例如 $\mathrm{e}^2 < 8$，其实只是 $\ln 2$ 和 $\ln 10$ 的界的另一种写法。整个过程中，我们都会用到 ln 是递增（increasing）函数这一事实，因此对两个正数之间的不等式取对数，不等号的方向保持不变。

## 解答

### 开首的结果

根据对数的定义，$x = \log_b c$ 的意思是

$$c = b^x.$$

两边取以 $a$ 为底的对数，并利用 $\log_a(b^x) = x \log_a b$，得

$$\log_a c = x \log_a b.$$

由于 $b \neq 1$（否则 $\log_b$ 没有意义），所以 $\log_a b \neq 0$，可以相除：

$$\frac{\log_a c}{\log_a b} = x = \log_b c,$$

得证。下文会用到两个推论。取 $a = c$ 得 $\log_b c = \dfrac{1}{\log_c b}$；取 $a = \mathrm{e}$ 得 $\log_b c = \dfrac{\ln c}{\ln b}$。

**第 (i) 部分**

根据倒数（reciprocal）法则，$\dfrac{1}{\log_2 \pi} = \log_\pi 2$ 及 $\dfrac{1}{\log_5 \pi} = \log_\pi 5$。所以

$$\frac{1}{\log_2 \pi} + \frac{1}{\log_5 \pi} = \log_\pi 2 + \log_\pi 5 = \log_\pi 10.$$

由于 $\pi > 1$，函数 $\log_\pi$ 是递增的，而 $10 > \pi^2$，因此

$$\log_\pi 10 > \log_\pi \pi^2 = 2,$$

这正是所求的不等式。

**第 (ii) 部分**

把第一个已知事实改写成自然对数（natural logarithm）：

$$\log_2 \frac{\pi}{\mathrm{e}} = \frac{\ln \pi - 1}{\ln 2} > \frac{1}{5} \quad \Longrightarrow \quad \ln \pi > 1 + \frac{\ln 2}{5},$$

这里我们乘以 $\ln 2 > 0$，所以不等号方向不变。

> **想一想** $\mathrm{e}^2 < 8$ 告诉我们甚么?

它是 $\ln 2$ 的一个下界（lower bound）。两边取 ln（由于 ln 是递增的，这样做是容许的），得 $2 < \ln 8 = 3 \ln 2$，所以 $\ln 2 > \frac{2}{3}$。代入，得

$$\ln \pi > 1 + \frac{1}{5} \cdot \frac{2}{3} = 1 + \frac{2}{15} = \frac{17}{15},$$

得证。

**第 (iii) 部分**

由 $\pi^2 < 10$，取对数得 $2 \ln \pi < \ln 10$。所以只需证明 $\ln 10 < \frac{30}{13}$。

把另外两个事实改写。由于 $\mathrm{e}^3 > 20$，我们有 $3 > \ln 20 = \ln 2 + \ln 10$，即

$$\ln 2 < 3 - \ln 10.$$

此外，$\log_{10} 2 = \dfrac{\ln 2}{\ln 10} > \dfrac{3}{10}$，而由于 $\ln 10 > 0$，

$$\ln 10 < \frac{10}{3} \ln 2.$$

第一式是 $\ln 2$ 的一个上界（upper bound），而 $\ln 2$ 在第二式中的系数（coefficient）是正数，所以可以把它代入：

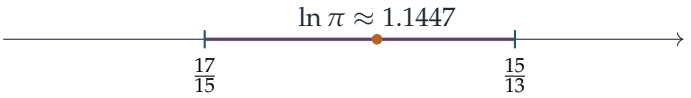
$$\ln 10 < \frac{10}{3} (3 - \ln 10) = 10 - \frac{10}{3} \ln 10.$$

合并 $\ln 10$ 的项，得 $\frac{13}{3}\ln 10 < 10$，所以 $\ln 10 < \frac{30}{13}$。因此

$$\ln \pi < \tfrac{1}{2}\ln 10 < \tfrac{1}{2}\cdot\tfrac{30}{13} = \frac{15}{13},$$

得证。

第 (ii) 和第 (iii) 部分合起来，把 $\ln \pi$ 困在一个狭窄的区间（interval）之内：




## 验算

以数值计算，$\ln \pi \approx 1.1447$，确实介乎 $\frac{17}{15} \approx 1.1333$ 与 $\frac{15}{13} \approx 1.1538$ 之间。在第 (i) 部分，$\log_\pi 10 \approx 2.0115$，只是稍稍大于 2，这与 $\pi^2 \approx 9.87$ 非常接近 10 的情况吻合。与 $\ln 10 \approx 2.3026$ 比较，界 $\ln 10 < \frac{30}{13} \approx 2.3077$ 也是成立的。题目给出的事实也值得核实一下：$\log_2(\pi/\mathrm{e}) \approx 0.209$、$\mathrm{e}^2 \approx 7.39$、$\mathrm{e}^3 \approx 20.09$ 及 $\log_{10} 2 \approx 0.3010$。

## 常见错误

- 把 $\dfrac{1}{\log_2 \pi}$ 写成 $\log_2 \frac{1}{\pi}$。取倒数是把底数和真数互换，而不是把真数取倒数。
- 在第 (iii) 部分，在需要 $\ln 2$ 的上界时错用了下界。只有当每个界都指向同一方向时，才可以把不等式串连起来。
- 在不等式两边相乘或取对数之前，忘记说明 $\ln 2$、$\ln 10$ 和 $\log_\pi$ 的性质良好（为正数，或是递增的）。

## STEP 之外

从寥寥几个粗略的事实推导出数值上的界，让我们稍稍体会到分析学家如何估计常数：分析（analysis）的很大一部分，就是把一个未知的量困在两个你能掌握的估计之间。$\ln 2$ 或 $\pi$ 更精确的界来自级数（series）和积分不等式，例如用矩形来估计 $\ln 2 = \int_1^2 \frac{1}{t}\,\mathrm{d}t$。Abbott 的 *Understanding Analysis*（《理解分析》）是引导读者进入这种思考方式的一本亲切入门书。

详细解答

# STEP I 2018，第 3 题

纯数学 · 角与双曲线

> **STEP I 2018，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/18-s1-q3

## 关键思路

两个角都可以通过正切（tangent）与 $P$ 的坐标联系起来：$RP$ 的斜率给出 $\tan\alpha$，$SP$ 的斜率给出 $\tan\beta$。于是条件 $\beta = 2\alpha$ 就变成 tan 的二倍角恒等式（double-angle identity），消去分母后便得到该曲线。第 (ii) 部分则把同样的代数步骤倒过来进行。难点在于：知道 $\tan\beta = \tan 2\alpha$ 只能确定 $\beta$ 至相差 $\pi$ 的整数倍，因此我们必须留意 $2\alpha$ 位于哪一个象限（quadrant）。

## 解答

### 设置两个角

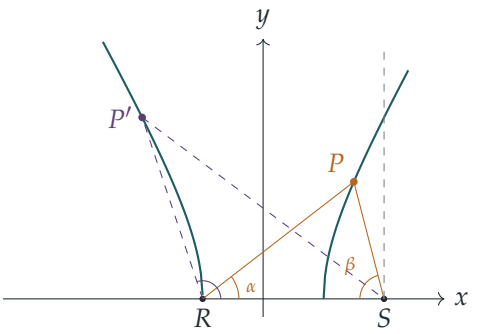
由于 $y > 0$，在 $R$ 处的角 $\alpha$ 是由正 $x$ 方向（沿 $RS$）转到 $RP$ 量度的，所以 $0 < \alpha < \pi$，而且

$$\tan\alpha = \frac{y}{x+a} \qquad (x \neq -a).$$

无论 $P$ 在 $R$ 的右方（$\alpha$ 为锐角，正切为正）还是左方（$\alpha$ 为钝角，正切为负），此式都成立。在 $S$ 处，角 $\beta$ 是由负 $x$ 方向（沿 $SR$）量度到 $SP$ 的。因为 $x < 2a$，点 $P$ 在 $S$ 的左方，所以 $\beta$ 为锐角，而且

$$\tan\beta = \frac{y}{2a-x} > 0, \qquad 0 < \beta < \tfrac{\pi}{2}.$$

当 $x \neq -a$ 时，记 $t = \tan\alpha$。




曲线 $y^2 = 3(x^2 - a^2)$ 在 $y > 0$、$x < 2a$ 的部分（图中取 $a = 1$）由两段组成：右支上 $a < x < 2a$ 的一段弧，以及整条左支的上半部分 $x < -a$。左支上的点 $P'$ 会令 $\alpha$ 成为钝角。

**第 (i) 部分**

设 $\beta = 2\alpha$。由于 $\beta < \frac{\pi}{2}$，我们有 $0 < \alpha < \frac{\pi}{4}$，故 $x > -a$、$0 < t < 1$，而且在 $1 - t^2 \neq 0$ 的情况下可以使用二倍角公式：

$$\frac{y}{2a - x} = \tan 2\alpha = \frac{2t}{1 - t^2} = \frac{2y/(x + a)}{1 - y^2/(x + a)^2} = \frac{2y(x + a)}{(x + a)^2 - y^2}.$$

消去 $y$（因 $y > 0$，这是容许的），再交叉相乘：

$$\begin{aligned}(x + a)^2 - y^2 &= 2(x + a)(2a - x)\\ \iff \quad y^2 &= (x + a)\big[(x + a) - 2(2a - x)\big] = (x + a)(3x - 3a) = 3(x^2 - a^2).\end{aligned}$$

所以 $P$ 位于 $y^2 = 3(x^2 - a^2)$ 上，得证。

**第 (ii) 部分**

现在设 $P$ 位于该曲线上，且 $y > 0$、$x < 2a$。于是 $x^2 - a^2 = \frac{1}{3}y^2 > 0$，故 $|x| > a$：即 $a < x < 2a$ 或 $x < -a$。特别地，$x \neq -a$，所以 $t = \tan\alpha$ 有定义，而且 $t \neq 0$。

第 (i) 部分的步骤可以倒推。曲线方程等价于 $(x + a)^2 - y^2 = 2(x + a)(2a - x)$。两边除以 $(x + a)^2$，并利用 $2a - x = y/\tan\beta$，得

$$1 - t^2 = \frac{2(2a - x)}{x + a} = \frac{2y}{(x + a)\tan\beta} = \frac{2t}{\tan\beta}, \qquad \text{故} \qquad (1 - t^2)\tan\beta = 2t.$$

若 $t^2 = 1$，此式会给出 $t = 0$，这是不可能的；所以 $t^2 \neq 1$，而且

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

**想一想** $\tan\beta = \tan 2\alpha$ 是否意味着 $\beta = 2\alpha$?

不一定：它只表示 $\beta$ 与 $2\alpha$ 相差 $\pi$ 的整数倍。我们利用已知的角的大小范围，按曲线的两段分情况讨论。

情况 $a < x < 2a$。此时 $x + a > 0$，所以 $0 < \alpha < \frac{\pi}{2}$，且 $0 < 2\alpha < \pi$。由于 $\tan 2\alpha = \tan\beta > 0$, $2\alpha$ 位于 $(0, \frac{\pi}{2})$, 与 $\beta$ 在同一区间。由于 tan 在 $(0, \frac{\pi}{2})$ 上是一一对应（one-to-one）的,

$$\beta = 2\alpha.$$

情况 $x < -a$。此时 $x + a < 0$，所以 $\frac{\pi}{2} < \alpha < \pi$，且 $\pi < 2\alpha < 2\pi$。由于 $\tan 2\alpha > 0$, $2\alpha \in (\pi, \frac{3\pi}{2})$，故 $2\alpha - \pi \in (0, \frac{\pi}{2})$，且 $\tan(2\alpha - \pi) = \tan 2\alpha = \tan\beta$。同前，tan 在 $(0, \frac{\pi}{2})$ 上是一一对应的，所以

$$\beta = 2\alpha - \pi.$$

因此可能的关系是 $\beta = 2\alpha$（当 $P$ 在右支上）和 $\beta = 2\alpha - \pi$（当 $P$ 在左支上）。两者都会出现，因为每一段上的每一点都满足所给条件。

## 验算

- 取 $a = 1$ 及 $P = (-2, 3)$, 由于 $3(4-1) = 9$, 此点在曲线上。于是 $\tan\alpha = 3/(-1) = -3$, 故 $\alpha \approx 108.43°$；又 $\tan\beta = \frac{3}{4}$，故 $\beta \approx 36.87°$。确实 $2\alpha - 180° \approx 36.87°$。
- 在右支上当 $x \to 2a$ 时, $y \to 3a$, 所以 $\tan\alpha \to 1$ 而 $\tan\beta \to \infty$: $\alpha \to \frac{\pi}{4}$, $\beta \to \frac{\pi}{2} = 2\alpha$。
- 在左支上远离原点时，$P$ 趋近渐近线（asymptote）$y = -\sqrt{3}\,x$，所以 $\alpha \to \frac{2\pi}{3}$，$\beta \to \frac{\pi}{3} = 2\alpha - \pi$。

我们亦以电脑在每一支上的大量随机点、并取随机的 $a$，验证了这两个关系，结果一致。

## 常见错误

- 写成 $\tan\alpha = y/|x + a|$，或假设 $\alpha$ 为锐角。在左支上 $\alpha$ 是钝角，且 $\tan\alpha < 0$。
- 未检查角的范围，便由 $\tan\beta = \tan 2\alpha$ 断定 $\beta = 2\alpha$。这样会漏掉第二个关系。
- 忘记 $x < 2a$ 和 $y > 0$ 把曲线限制为两段，以及需要 $|x| > a$ 来排除 $x = -a$。

## STEP 之外

这条曲线是一条离心率（eccentricity）为 2、焦点（focus）为 $S$、准线（directrix）为 $x = \frac{a}{2}$ 的双曲线（hyperbola）。第 (i) 部分的性质给出了利用双曲线三等分角（trisect an angle）的经典方法（帕普斯（Pappus）已知晓），从而绕过了仅用尺规不能三等分角这一不可能性。这个不可能性的证明本身，是域论（field theory）入门课程的亮点之一。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）涵盖了这部分内容，而 Needham 的 *Visual Complex Analysis*（《可视化复分析》）则是探讨角与圆锥曲线几何的绝佳读物。

详细解答

# STEP I 2018，第 4 题

纯数学 · 曲线描绘与积分

**STEP I 2018，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/18-s1-q4

## 关键思路

f 的每一部分都由 $\ln x$ 构成，而那个麻烦的因子 $\frac{1}{x}$ 正好是 $\ln x$ 的导数（derivative）。因此代换（substitution）$u = \ln x$ 可把积分变成一个关于 $u$ 的简单有理函数（rational function）的积分。以 $u$ 表示时，把 $x$ 换成 $x^{-1}$ 只是改变 $u$ 的正负号，所以对称性（symmetry）归结为答案是 $u$ 的偶函数（even function）。

## 解答

全文设 $u = \ln x$。由于 $x \neq 1$，所以 $u \neq 0$；而且 $u > 0$ 当且仅当 $x > 1$。

### f 与 f′ 的零点

设

$$g(u) = \frac{(1-u^2)^2}{u} = \frac{1}{u} - 2u + u^3, \qquad \text{从而} \qquad \mathrm{f}(x) = \frac{g(\ln x)}{x}.$$

当 $(\ln x)^2 = 1$ 时，因子 $(1-u^2)^2$ 为零，所以 $\mathrm{f}(x) = 0$。由商法则（quotient rule）及链式法则（chain rule），

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

我们已知 $g(\pm 1) = 0$。此外，由于 $(1-u^2)^2 = (1-u)^2(1+u)^2$ 在 $u = \pm 1$ 处有重因子（repeated factor），它的导数 $-4u(1-u^2)$ 在该处亦为零，因此

$$g'(u) = \frac{-4u(1-u^2)\cdot u - (1-u^2)^2}{u^2} = 0 \quad \text{当} u = \pm 1.$$

所以每当 $(\ln x)^2 = 1$，即在 $x = e$ 及 $x = e^{-1}$ 处，亦有 $\mathrm{f}'(x) = 0$。

### 第 (i) 部分

代入 $u = \ln t$，则 $\mathrm{d}u = \frac{1}{t}\,\mathrm{d}t$。于是

$$\int \mathrm{f}(t)\,\mathrm{d}t = \int \frac{(1-u^2)^2}{u}\,\mathrm{d}u = \int \left(\frac{1}{u} - 2u + u^3\right)\mathrm{d}u = \ln|u| - u^2 + \tfrac{1}{4}u^4 + C.$$

定义 $G(u) = \ln|u| - u^2 + \frac{1}{4}u^4$。F 的两段定义都不会让积分范围跨过 $t = 1$，所以在每一段上 $u$ 的正负号不变，这个原函数（antiderivative）在该处成立。

当 $x > 1$ 时，下限 $t = e$ 给出 $u = 1$，而 $G(1) = 0 - 1 + \frac{1}{4} = -\frac{3}{4}$。当 $0 < x < 1$ 时，下限 $t = e^{-1}$ 给出 $u = -1$，同样有 $G(-1) = -\frac{3}{4}$。因此两种情况得出同一条公式：

$$\boxed{\mathrm{F}(x) = \ln|\ln x| - (\ln x)^2 + \tfrac{1}{4}(\ln x)^4 + \tfrac{3}{4} \qquad (x > 0,\ x \neq 1).}$$

现在 $\ln(x^{-1}) = -\ln x$。上式中每一处 $\ln x$ 都位于 $|\cdot|$ 之内或被提升至偶数次幂，所以

$$\mathrm{F}(x^{-1}) = \ln|-\ln x| - (-\ln x)^2 + \tfrac{1}{4}(-\ln x)^4 + \tfrac{3}{4} = \mathrm{F}(x),$$

得证。（两边的常数 $\frac{3}{4}$ 之所以相同，只因为两个下限 $e$ 和 $e^{-1}$ 本身就是一对倒数。）

### 第 (ii) 部分

> **想一想** F 在哪里递增？在 $x = 1$ 附近会发生甚么事？

**斜率 (gradient)**。$\mathrm{F}'(x) = \mathrm{f}(x)$。因子 $(1-u^2)^2$ 永不为负，且 $x > 0$，所以 $\mathrm{f}(x)$ 与 $u = \ln x$ 同号。因此 F 在 $(0,1)$ 上递减，在 $(1,\infty)$ 上递增。它仅有的驻点（stationary point）位于 $x = e^{-1}$ 及 $x = e$，而按定义在这两点 $\mathrm{F} = 0$。

这两点并非转向点（turning point）。在 $x = e$ 附近，两侧的斜率都是正的；在 $x = e^{-1}$ 附近，两侧的斜率都是负的。由于 $\mathrm{f}' = \mathrm{F}''$ 在该处亦为零，它们是**驻定拐点**（stationary point of inflection）。由此可知

$$\mathrm{F}(x) > 0 \text{ 当 } 0 < x < e^{-1} \text{ 及 } x > e, \qquad \mathrm{F}(x) < 0 \text{ 当 } e^{-1} < x < 1 \text{ 及 } 1 < x < e.$$

**两端的行为**。

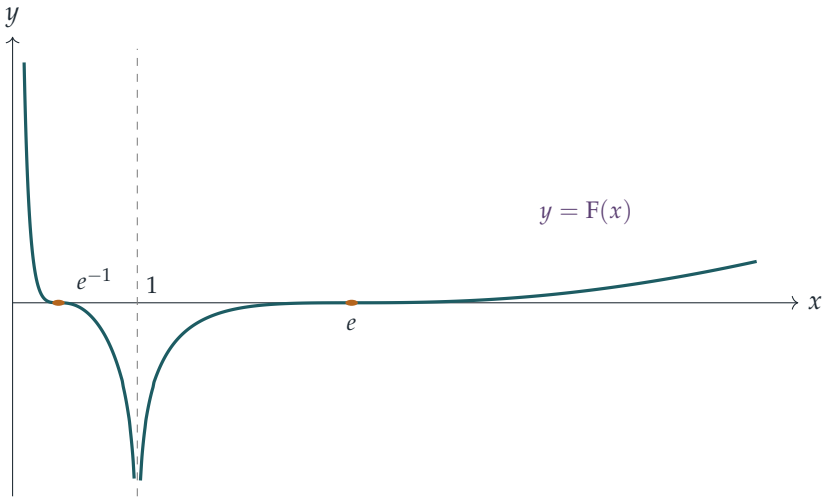
- 当 $x \to 1$ 时，$u \to 0$ 而 $\ln|u| \to -\infty$，其余各项则趋向 $\frac{3}{4}$。所以从两侧都有 $\mathrm{F}(x) \to -\infty$，而 $x = 1$ 是一条垂直渐近线（vertical asymptote）。
- 当 $x \to \infty$ 时，$u \to \infty$，而 $\frac{1}{4}u^4$ 一项占主导，所以 $\mathrm{F}(x) \to \infty$。不过增长非常缓慢：斜率 $\mathrm{f}(x) \approx u^3/x \to 0$。
- 当 $x \to 0^+$ 时，$u \to -\infty$，所以由对称性，同样有 $\mathrm{F}(x) \to \infty$。这里的变化十分陡峭：斜率 $\mathrm{f}(x) \approx u^3/x \to -\infty$。曲线不与 $y$ 轴相交。

**凹性（concavity）（用于更仔细的草图）**。利用 $\mathrm{F}'' = \mathrm{f}' = (g' - g)/x^2$ 并化简，

$$\mathrm{F}''(x) = -\frac{(u^2-1)(u^3-3u^2-u-1)}{u^2x^2}.$$

三次式 $u^3 - 3u^2 - u - 1$ 只有一个实根，位于 $u \approx 3.38$（它在 $u = 1 - \frac{2}{\sqrt{3}}$ 处的局部极大值为负）。对于所有较小的 $u$，它都是负的。所以曲线在 $(0, e^{-1})$ 上向上凹，在 $(e^{-1}, 1)$ 及 $(1, e)$ 上向下凹。从 $e$ 开始曲线向上凹，直至在 $x \approx 29.5$ 处出现另一个拐点，之后向下凹。最后这个拐点远在右方，草图上不必画出。

对称性 $\mathrm{F}(x^{-1}) = \mathrm{F}(x)$ 表示在 $x$ 处的高度与在 $1/x$ 处的高度相同。例如 $\mathrm{F}(2) = \mathrm{F}(\frac{1}{2}) \approx -0.04$ 及 $\mathrm{F}(5) = \mathrm{F}(\frac{1}{5}) \approx 0.31$。因此曲线在 $(1, \infty)$ 上的部分，就是 $(0, 1)$ 上的部分沿水平方向朝渐近线压缩而成。




曲线在 $y$ 轴附近从 $+\infty$ 陡峭地下降，然后逐渐平缓，在 $(e^{-1}, 0)$ 处的驻定拐点触及 $x$ 轴，接着急跌至渐近线 $x = 1$ 处的 $-\infty$。在右边，曲线从 $-\infty$ 回升，在 $(e, 0)$ 处的第二个驻定拐点变得平坦，然后缓慢上升至 $+\infty$。

## 验算

- 对方框内的公式求导：$\frac{\mathrm{d}}{\mathrm{d}x}\left(\ln|u| - u^2 + \frac{1}{4}u^4\right) = \frac{1}{x}\left(\frac{1}{u} - 2u + u^3\right) = \mathrm{f}(x)$，而且公式在 $x = e$ 及 $x = e^{-1}$ 处的值都是 0。
- 数值上，$\int_e^5 \mathrm{f}(t)\,\mathrm{d}t \approx 0.3130$ 及 $\int_{1/e}^{1/5} \mathrm{f}(t)\,\mathrm{d}t \approx 0.3130$，与公式及对称性吻合。
- 正负号亦吻合：在 $(1, e)$ 上函数递增至 0，所以在该处必为负，而事实上 $\mathrm{F}(2) \approx -0.039$。

## 常见错误

- 在 $0 < x < 1$ 时写成 $\ln(\ln x)$，但此时 $\ln x < 0$。$\frac{1}{u}$ 的原函数是 $\ln|u|$，而正是这个绝对值令对称性成立。
- 在 $x = e^{\pm 1}$ 处画成转向点。由于 f 含有平方因子，它在该处不变号，所以这些是驻定拐点。
- 遗漏 $x = 1$ 处的垂直渐近线，或把两支曲线跨过渐近线连接起来。

## STEP 之外

F 由积分定义，而微积分基本定理（fundamental theorem of calculus）指出，在 f 连续的每个区间上都有 $\mathrm{F}' = \mathrm{f}$。这正是两段定义需要各自的基点的原因：f 在 $x = 1$ 处趋向无穷，所以无法用单一个积分跨越该点。对称变换 $x \mapsto x^{-1}$ 在 $u = \ln x$ 之下变成 $u \mapsto -u$，这是把乘法对称转化为加法对称的一个小例子，而这个想法贯穿整个分析学。如欲仔细了解微积分基本定理及瑕积分（improper integral），可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

# STEP I 2018，第 5 题

纯数学 · 多项式与因式定理

> **STEP I 2018，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/18-s1-q5

## 关键思路

“在这些点取值 1”其实是一个关于根（root）的陈述：若 $P(k)=1$，则 $k$ 是 $P(x)-1$ 的根，所以 $x-k$ 是它的因式（factor）。只要减去那个共同的值再因式分解（factorise），题目每一部分都会变得容易。在第 (iii) 部分，因式分解把关于 $S$ 的问题，转化为“把一个小数写成四个相异整数之积”的问题，而真正起决定作用的正是“相异”这个条件。

## 解答

### 第 (i) 部分

根据余式定理（remainder theorem），$P(x)$ 除以 $x-k$ 的余数是 $P(k)$。所以我们需要 $P(1)=P(2)=P(3)=P(4)=1$，即 $1,2,3,4$ 是 $P(x)-1$ 的根。由于 $P(x)-1$ 的次数（degree）最多为 4，而且有这四个相异的根，它必定是 $(x-1)(x-2)(x-3)(x-4)$ 的常数倍。因此最一般的这类多项式是

$$P(x)=1+k(x-1)(x-2)(x-3)(x-4), \qquad k\neq 0,$$

其中需要 $k\neq 0$，次数才会恰好是 4。

### 第 (ii) 部分

与第 (i) 部分完全一样，$1,2,\ldots,N$ 是 $P(x)-1$ 的 $N$ 个相异根，而它的次数为 $N$（因为 $N\geq 1$，减去 1 不会改变次数）。所以

$$P(x)-1=k(x-1)(x-2)\cdots(x-N)$$

其中 $k$ 为某常数；又因为 $P$ 的次数为 $N \geq 1$，所以 $k \neq 0$。于是

$$P(N+1) - 1 = k \cdot N(N-1) \cdots 1 = k\,N! \neq 0,$$

因为 $k$ 和 $N!$ 都不为零。所以 $P(N+1) \neq 1$。

若 $P(N+1) = 2$，则 $k\,N! = 1$，所以 $k = \frac{1}{N!}$。对正整数 $r$，

$$\begin{aligned} P(N+r) &= 1 + \frac{(N+r-1)(N+r-2)\cdots r}{N!} \\ &= 1 + \frac{(N+r-1)!}{(r-1)!\,N!} = 1 + \binom{N+r-1}{N}. \end{aligned}$$

（这个乘积共有 $N$ 个因子，由 $N+r-1$ 递减至 $r$。）

现在我们希望对每个 $N$ 都有 $1 + \binom{N+r-1}{N} = N + r$。试试较小的 $r$。取 $r = 1$，得 $P(N+1) = 2$，它只在 $N = 1$ 时才等于 $N+1$。取 $r = 2$，

$$P(N+2) = 1 + \binom{N+1}{N} = 1 + (N+1) = N+2,$$

所以无论 $N$ 是多少，$r = 2$ 都成立。

### 第 (iii) 部分

四个相异整数 $a, b, c, d$ 是 $S(x) - 2001$ 的根，而它的次数为 4，首项系数（leading coefficient）为 1。所以

$$S(x) = 2001 + (x-a)(x-b)(x-c)(x-d).$$

**(a)** 假设对某整数 $e$ 有 $S(e) = 2018$。那么

$$(e-a)(e-b)(e-c)(e-d) = 17.$$

这四个因子都是整数，而且由于 $a, b, c, d$ 互不相同，它们也互不相同。它们都不是 0，因为乘积不为零。

> **想一想** 四个相异整数为甚么乘不出一个质数？

在相异的非零整数之中，绝对值为 1 的最多只有两个（即 1 和 $-1$）。所以四个因子中至少有两个的绝对值不少于 2，而这两个因子之积的绝对值不少于 4。但这个积整除（divides）17，所以它必须是 $\pm 17$；由于 17 是质数（prime），它不能写成两个绝对值都不少于 2 的整数之积。这个矛盾说明不存在整数 $e$ 使 $S(e) = 2018$。

**(b)** 现在 $S(0) = 2001 + (-a)(-b)(-c)(-d) = 2001 + abcd$，所以 $S(0) = 2017$ 当且仅当

$$abcd = 16.$$

反过来，任何四个乘积为 $abcd = 16$ 的相异整数都给出一个合乎要求的 $S$（上面的式子是首一（monic）且系数为整数），所以我们要数的是由四个相异整数组成、乘积为 16 的集合 $\{a, b, c, d\}$ 有多少个；每个集合都恰好只有一种方式按递增次序排列。

$a, b, c, d$ 每一个都整除 16，所以其绝对值是 $1, 2, 4, 8, 16$ 之一，而四个绝对值之积为 16。由于这些整数互不相同，每个绝对值最多出现两次（作为 $\pm m$）。把绝对值写成幂 $2^p$，四个指数之和为 4，且每个指数最多用两次，可能的情况是

$$\{1,1,2,8\}, \qquad \{1,1,4,4\}, \qquad \{1,2,2,4\}.$$

（若用上 16，其余三个便必须是 1、1、1，这是不容许的；而 $\{2,2,2,2\}$ 把 2 用了四次。）现在选取正负号，使乘积为 $+16$：

- $\{1,1,2,8\}$：那一对是 $1, -1$，贡献 $-1$，所以 2、8 之中恰好有一个是负数。这给出 $\{-2,-1,1,8\}$ 和 $\{-8,-1,1,2\}$。
- $\{1,1,4,4\}$：两对都是 $\pm$，而 $(1)(-1)(4)(-4) = 16$。这给出 $\{-4,-1,1,4\}$。
- $\{1,2,2,4\}$：那一对 $2, -2$ 贡献 $-4$，所以 1、4 之中恰好有一个是负数。这给出 $\{-4,-2,1,2\}$ 和 $\{-2,-1,2,4\}$。

所以选取 $a < b < c < d$ 共有 **5** 种方法。

## 验算

- 在第 (ii) 部分取 $N = 1$，则 $P(x) = x$，所以 $P(N + r) = 1 + r$ 与 $1 + \binom{r}{1}$ 相符，而且 $P(3) = 3$。取 $N = 2$，则 $P(x) = 1 + \frac{1}{2}(x-1)(x-2)$ 给出 $P(4) = 4$，与 $r = 2$ 的结论一致。
- 对于 (iii)(b)，把五个集合逐一乘出来：例如 $(-2)(-1)(2)(4) = 16$。用电脑搜索 $[-16, 16]$ 内所有由相异整数组成的四元组，找到的正是这五个。
- 对于 (iii)(a)，留意 $2018 - 2001 = 17$ 是质数；若把 2001 换成另一个值，使差变成像 $4 = (-2)(-1)(1)(2)$ 这样的数，这样的 $e$ 便可能存在，所以质数性确实是关键。

## 常见错误

- 在第 (i) 部分忘记 $k \neq 0$，或只写 $P(x) = (x-1)(x-2)(x-3)(x-4) + 1$，而这并不是最一般的多项式。
- 在第 (iii)(b) 部分只数正因数，或忘记 1 和 $-1$ 可以同时使用。
- 写成 $S(0) = 2001 - abcd$：其实共有四个负号，所以乘积是 $+abcd$。

## STEP 之外

“一个 $N$ 次多项式由它在 $N+1$ 个点上的值所决定”这个想法，是拉格朗日（Lagrange）插值（Lagrange interpolation）和有限差分（finite differences）的基础；第 (ii) 部分其实就是换了个样子的多项式 $1+\binom{x-1}{N}$。第 (iii) 部分是数论（number theory）中一个标准工具的小例子：若 $P$ 的系数都是整数，则 $m-n$ 整除 $P(m)-P(n)$，这大大限制了多项式在整数上能取的值。Davenport 的 *The Higher Arithmetic*（《高等算术》）是进一步阅读这类整除论证的好起点。

详细解答

# STEP I 2018，第 6 题

纯数学 · 中点法则与梯形法则

**STEP I 2018，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/18-s1-q6

## 关键思路

两个近似值都是等距角的正弦之和，而题目给出的积化和差恒等式（product-to-sum identity）只要乘上一个选得恰当的 $2\sin\theta$，就能令这类和式裂项相消（telescope）。中点法则（midpoint rule）在 $\frac{\pi}{2n}$ 的奇数倍处取样，梯形法则（trapezium rule）则在偶数倍处取样，所以取 $\theta = \frac{\pi}{2n}$ 后，每一部分都变成裂项相消的和式。一旦求得 $A_n$ 和 $B_n$ 的闭式（closed form），第 (iii) 部分就只是半角公式（half-angle formulae）的运用。

## 解答

### 预备恒等式

在恒等式中取 $P = \theta$ 及 $Q = (2k-1)\theta$：

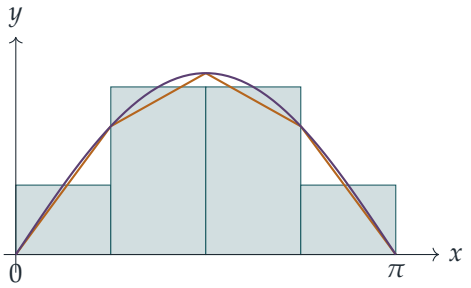
$$2\sin\theta\,\sin(2k-1)\theta = \cos(2k-2)\theta - \cos 2k\theta.$$

由 $k = 1$ 加到 $k = n$，每一项中减去的 $\cos 2k\theta$ 都在下一项中加回，因此和式裂项相消：

$$\begin{aligned} 2\sin\theta\sum_{k=1}^{n}\sin(2k-1)\theta &= (\cos 0 - \cos 2\theta) + (\cos 2\theta - \cos 4\theta) + \cdots \\ &\quad + \big(\cos(2n-2)\theta - \cos 2n\theta\big) \\ &= 1 - \cos 2n\theta, \end{aligned}$$

得证。

以下各部分中，记条宽为 $h = \frac{\pi}{n}$，并记 $\theta = \frac{\pi}{2n} = \frac{h}{2}$。




当 $n = 4$ 时：由于 $\sin x$ 在 $[0, \pi]$ 上是凹函数（concave），中点矩形（阴影部分）高估了面积，而梯形的弦（粗线）则低估了面积。

**第 (i) 部分**

第 $k$ 条由 $(k-1)h$ 延伸至 $kh$，所以其中点位于 $x = (k - \frac{1}{2})h = (2k-1)\theta$，高度为 $\sin(2k-1)\theta$。因此

$$A_n = h \sum_{k=1}^{n} \sin(2k-1)\theta.$$

乘以 $2\sin\theta$ 并利用恒等式，再加上 $2n\theta = \pi$，得

$$2\sin\theta\, A_n = h\,(1 - \cos\pi) = 2h.$$

两边除以 2，得 $A_n \sin\left(\frac{\pi}{2n}\right) = \frac{\pi}{n}$，得证。

**第 (ii) 部分**

纵坐标（ordinates）位于 $x = kh = 2k\theta$，其中 $k = 0, 1, \ldots, n$，而两端的纵坐标为 $\sin 0 = \sin\pi = 0$。所以梯形法则给出

$$B_n = \frac{h}{2}\left(\sin 0 + 2\sum_{k=1}^{n-1} \sin 2k\theta + \sin\pi\right) = h\sum_{k=1}^{n-1} \sin 2k\theta.$$

现在这些是 $\theta$ 的偶数倍，所以我们再用一次恒等式，这次取 $P = \theta$、$Q = 2k\theta$：

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

由 $k = 1$ 加到 $n-1$，同样裂项相消：

$$2\sin\theta \sum_{k=1}^{n-1} \sin 2k\theta = \cos\theta - \cos(2n-1)\theta.$$

? **想一想** 怎样化简 $\cos(2n-1)\theta$?

由于 $2n\theta = \pi$，我们有 $(2n-1)\theta = \pi - \theta$，而 $\cos(\pi-\theta) = -\cos\theta$。所以右方等于 $2\cos\theta$，从而

$$2\sin\theta\, B_n = 2h\cos\theta, \qquad \text{即} \qquad B_n \sin\left(\frac{\pi}{2n}\right) = \frac{\pi}{n}\cos\left(\frac{\pi}{2n}\right).$$

（当 $n=1$ 时和式为空，$B_1 = 0$；由于 $\cos\frac{\pi}{2} = 0$，公式依然成立。）

**第 (iii) 部分**

由于 $0 < \theta \le \frac{\pi}{2}$，$\sin\theta \neq 0$，所以由第 (i) 及 (ii) 部分得

$$A_n = \frac{h}{\sin\theta}, \qquad B_n = \frac{h\cos\theta}{\sin\theta}.$$

把 $n$ 换成 $2n$，$h$ 和 $\theta$ 都减半：

$$A_{2n} = \frac{h/2}{\sin(\theta/2)}, \qquad B_{2n} = \frac{(h/2)\cos(\theta/2)}{\sin(\theta/2)}.$$

第一个结果利用 $1+\cos\theta = 2\cos^2\frac{\theta}{2}$ 及 $\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$：

$$\tfrac{1}{2}(A_n + B_n) = \frac{h}{2}\cdot\frac{1+\cos\theta}{\sin\theta} = \frac{h}{2}\cdot\frac{2\cos^2(\theta/2)}{2\sin(\theta/2)\cos(\theta/2)} = \frac{(h/2)\cos(\theta/2)}{\sin(\theta/2)} = B_{2n}.$$

第二个结果利用同一条 $\sin\theta$ 的二倍角公式（double-angle formula）（以及 $\cos\frac{\theta}{2} \neq 0$）：

$$A_n B_{2n} = \frac{h}{2\sin(\theta/2)\cos(\theta/2)}\cdot\frac{(h/2)\cos(\theta/2)}{\sin(\theta/2)} = \frac{h^2}{4\sin^2(\theta/2)} = \left(\frac{h/2}{\sin(\theta/2)}\right)^2 = A_{2n}^2.$$

两个结果均已证明。

## 验算

- $n=1$：一个宽 $\pi$、高 $\sin\frac{\pi}{2} = 1$ 的中点矩形给出 $A_1 = \pi$，与 $\pi/\sin\frac{\pi}{2}$ 相符；而如前所述 $B_1 = 0$。
- $n=2$：$A_2 = \frac{\pi}{2}(\sin\frac{\pi}{4} + \sin\frac{3\pi}{4}) = \frac{\pi}{\sqrt{2}} \approx 2.2214$，$B_2 = \frac{\pi}{2} \approx 1.5708$；于是 $\frac{1}{2}(A_1+B_1) = \frac{\pi}{2} = B_2$，且 $A_1 B_2 = \frac{\pi^2}{2} = A_2^2$。
- $n=10$：直接求和得 $A_{10} \approx 2.00825$ 及 $B_{10} \approx 1.98352$，分别位于真实面积 2 的两侧。当 $n\to\infty$ 时，$\frac{\pi}{n}\big/\sin\frac{\pi}{2n} \to 2$ 且 $\cos\frac{\pi}{2n} \to 1$，所以两者都趋向 2，正合预期。四个结果亦已用电脑就 200 以内的每个 $n$ 逐一验证。

## 常见错误

- 把中点放在 $kh$ 而不是 $(k-\frac{1}{2})h$，这样中点法则就变成了在梯形法则节点上的求和。

- 忘记梯形法则中两端的纵坐标要减半。这里两端纵坐标为零，所以 $B_n$ 恰好是 $h$ 乘以内部纵坐标之和。
- 没有化简 $\cos(2n-1)\theta$，而不懂利用 $(2n-1)\theta = \pi - \theta$。

## STEP 之外

第 (iii) 部分表明 $B_{2n}$ 是 $A_n$ 与 $B_n$ 的算术平均数（arithmetic mean），而 $A_{2n}$ 是 $A_n$ 与 $B_{2n}$ 的几何平均数（geometric mean）。取倒数后，这就变成先取调和平均数（harmonic mean）、再取几何平均数的模式，正是阿基米德（Archimedes）用圆的外切及内接多边形夹逼 $\pi$ 的方法，而这种迭代称为 Borchardt 算法（Borchardt's algorithm）。更一般地，比较中点法则与梯形法则的误差（中点法则的误差约为梯形法则误差的一半，且符号相反），便引出辛普森法则（Simpson's rule）和理察森外推法（Richardson extrapolation），两者都是数值分析（numerical analysis）的核心概念。关于这些法则背后的误差分析，可参阅 Bartle 与 Sherbert 所著 *Introduction to Real Analysis*（《实分析导论》）中有关黎曼积分（Riemann integral）的章节。

详细解答

# STEP I 2018，第 7 题

纯数学 · 解三次方程

**STEP I 2018，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/18-s1-q7

## 关键思路

这个代换（substitution）是一个默比乌斯变换（Möbius transformation），经过精心选取，使得去分母之后 $z^2$ 项和 $z$ 项互相抵消，只剩下 $z^3$ 和常数项。形如 $az^3+b=0$ 的三次方程（cubic equation）很容易解，所以整个任务就是把一般的既约三次方程（reduced cubic）$x^3-3cx+d=0$ 化成这个特殊形式。要做到这一点，就要由两个数的和与积求出这两个数，这归结为一个二次方程（quadratic equation）；而条件 $d^2>4c^3$ 正是来自这个二次方程有相异实根（distinct real roots）的条件。

## 解答

### 第 (i) 部分

设 $x=\dfrac{pz+q}{z+1}$，并把方程两边乘以 $(z+1)^3$：

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

我们需要用到 $(pz+q)(z+1)^2=pz^3+(2p+q)z^2+(p+2q)z+q$。现在按 $z$ 的幂次合并同类项。

$$\begin{aligned}
z^3:&\quad p^3-3p^2q+p^2q+pq^2=p(p^2-2pq+q^2)=p(p-q)^2,\\
z^2:&\quad 3p^2q-3pq(2p+q)+3pq(p+q)=3p^2q-6p^2q-3pq^2+3p^2q+3pq^2=0,\\
z^1:&\quad 3pq^2-3pq(p+2q)+3pq(p+q)=3pq^2-3p^2q-6pq^2+3p^2q+3pq^2=0,\\
z^0:&\quad q^3-3pq^2+p^2q+pq^2=q(q^2-2pq+p^2)=q(p-q)^2.
\end{aligned}$$

因此方程化为 $az^3+b=0$，其中

$$a=p(p-q)^2, \qquad b=q(p-q)^2.$$

由于 $p\neq q$，我们可以除以 $(p-q)^2$，得到更简单的形式 $pz^3+q=0$。

对于每个 $x\neq p$，这个代换都可以反解为 $z=\dfrac{x-q}{p-x}$；而 $x=p$ 只有在 $p(p-q)^2=0$ 时才是该三次方程的根。此外，$z=-1$ 永远不是 $pz^3+q=0$ 的解，因为那需要 $p=q$。所以三次方程的根与 $pz^3+q=0$ 的根恰好一一对应（当 $p=0$ 时，三次方程就是 $x^3=0$）。

**第 (ii) 部分**

比较 $x^3-3cx+d$ 与 $x^3-3pqx+pq(p+q)$，我们需要

$$pq=c, \qquad pq(p+q)=d, \qquad \text{即} \qquad pq=c, \quad p+q=\frac{d}{c},$$

由于 $c\neq 0$，除以 $c$ 是容许的。和为 $\frac{d}{c}$、积为 $c$ 的两个数，就是

$$t^2-\frac{d}{c}t+c=0.$$

的两个根。这个二次方程有相异实根，当且仅当其判别式（discriminant）为正：

$$\frac{d^2}{c^2}-4c>0 \iff d^2>4c^3,$$

这里乘以了 $c^2>0$。所以当 $d^2>4c^3$ 时，取 $p$ 和 $q$ 为这个二次方程的两个根，就能把方程写成所需的形式，而且 $p$、$q$ 是相异的实数。

**第 (iii) 部分**

这里 $c=-2$ 而 $d=-2$，所以 $d^2=4>-32=4c^3$，第 (ii) 部分适用。我们需要 $p+q=\frac{d}{c}=1$ 及 $pq=c=-2$，所以 $p$、$q$ 是 $t^2-t-2=(t-2)(t+1)=0$ 的根。取 $p=2$、$q=-1$。由第 (i) 部分，方程变为

$$2z^3-1=0, \qquad z^3=\tfrac{1}{2},$$

它唯一的实根是 $z=2^{-1/3}$。记 $u=2^{1/3}$，则 $z=\frac{1}{u}$。于是

$$x=\frac{2z-1}{z+1}=\frac{2-u}{1+u}.$$

? **想一想** 分母含有立方根，这个式子能否再化简?

可以。利用 $1+u^3=(1+u)(1-u+u^2)$ 进行有理化（rationalise），由于 $u^3=2$，这个式子等于 3：

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

所以实根是

$$x = 2^{2/3} - 2^{1/3} = \sqrt[3]{4} - \sqrt[3]{2}.$$

这是唯一的实根：$x^3 + 6x - 2$ 的导数（derivative）是 $3x^2 + 6 > 0$，所以它严格递增。$z$ 的另外两个值是复数，给出两个复根。（若改为取 $p = -1$、$q = 2$，则得到 $z^3 = 2$，结果是同一个根。）

### 第 (iv) 部分

由于 $x = p$ 给出 $p^3 - 3p^3 + 2p^3 = 0$，所以 $x - p$ 是一个因式（factor），而

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

所以根是 $x = p$（重根）和 $x = -2p$。

留意这正是第 (i) 部分中 $q = p$ 的形式，也恰好是第 (ii) 部分排除了的情况。现在设 $d^2 = 4c^3$，且 $c, d \neq 0$。那么 $c^3 = \frac{1}{4}d^2 > 0$，所以 $c > 0$。我们要找 $p$ 使得 $p^2 = c$ 及 $2p^3 = d$。把第二式除以第一式，启发我们取

$$p = \frac{d}{2c},$$

而这确实可行：$p^2 = \dfrac{d^2}{4c^2} = \dfrac{4c^3}{4c^2} = c$，于是 $2p^3 = 2p \cdot p^2 = 2 \cdot \dfrac{d}{2c} \cdot c = d$。所以 $x^3 - 3cx + d = x^3 - 3p^2x + 2p^3$，其根为

$$x = \frac{d}{2c}\text{（二重）} \qquad \text{及} \qquad x = -\frac{d}{c}.$$

## 验算

- 第 (iii) 部分：$\sqrt[3]{4} - \sqrt[3]{2} \approx 0.32748$，而 $x^3 + 6x - 2$ 在这个值的取值准确到小数点后十五位都是 0。卡尔达诺公式（Cardano's formula）给出同一个根：$\sqrt[3]{1+3} + \sqrt[3]{1-3}$。
- 第 (iv) 部分：取 $c = 1$、$d = 2$，三次方程为 $x^3 - 3x + 2 = (x-1)^2(x+2)$，与二重根 $\frac{d}{2c} = 1$ 及单根 $-\frac{d}{c} = -2$ 吻合。
- 第 (i) 部分：取 $p = 1$、$q = 0$，三次方程为 $x^3 = 0$，代换给出 $x = \frac{z}{z+1}$，所以 $x^3(z+1)^3 = z^3$，与 $a = 1$、$b = 0$ 吻合。

## 常见错误

- 在第 (ii) 部分忘记 $c$ 可能是负数：不等式乘以 $c^2$ 是安全的，但乘以 $c$ 就不是。
- 在第 (iii) 部分把 $\frac{1}{2}$ 的三个立方根全部当作实根。只有 $z = 2^{-1/3}$ 是实数。
- 在第 (iv) 部分取 $p = \sqrt{c}$。当 $d < 0$ 时符号就错了；$p = \frac{d}{2c}$ 则会自动给出正确的符号。

## STEP 之外

这是推导三次方程卡尔达诺公式的一条巧妙途径。二次方程 $t^2 - \frac{d}{c}t + c = 0$ 扮演了“预解式”（resolvent）的角色，而 $d^2 - 4c^3$ 的符号（相差一个因子）就是三次方程的判别式，它决定方程有一个实根还是三个实根。为何三次和四次方程可以用根式求解，而一般的五次方程却不能，正是伽罗瓦理论（Galois theory）的主题。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）是开始阅读这方面内容的好起点。

详细解答

# STEP I 2018，第 8 题

纯数学 · 耦合微分方程

> **STEP I 2018，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/18-s1-q8

## 关键思路

两条导数（derivative）规则是刻意设计成互相抵消的：每当出现 $\mathrm{s}'$，就会带来 $\mathrm{c}^2$；每当出现 $\mathrm{c}'$，就会带来 $-\mathrm{s}^2$。因此，贯穿全题最自然的做法，是对乘积和商求导，观察各项如何合并，然后利用 (i) 的恒等式（identity）把所有东西化成单一个函数。至于积分（integral），要把每个被积函数看成“某东西乘以 $\mathrm{s}^2$”或“某东西乘以 $\mathrm{c}^2$”，因为这些正正是我们已知的导数。最后两部分中关于 $u$ 的积分，其实是同样的积分换了个样子。

## 解答

全题中，在不会引起混淆时，以 s 和 c 分别表示 $\mathrm{s}(x)$ 和 $\mathrm{c}(x)$。

### 第 (i) 部分

由连锁法则（chain rule），

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{s}^3+\mathrm{c}^3\right)=3\mathrm{s}^2\,\mathrm{s}'+3\mathrm{c}^2\,\mathrm{c}'=3\mathrm{s}^2\mathrm{c}^2-3\mathrm{c}^2\mathrm{s}^2=0,$$

所以 $\mathrm{s}^3+\mathrm{c}^3$ 是常数。它在 $x=0$ 的值是 $0^3+1^3=1$，因此

$$\mathrm{s}(x)^3+\mathrm{c}(x)^3=1 \quad \text{对所有} x,$$

即为所求。这个恒等式是全题其余部分的引擎：它让我们可以把 $\mathrm{s}^3$ 换成 $1-\mathrm{c}^3$，反之亦然。

**第 (ii) 部分**

由乘积法则（product rule），

$$\frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{sc}) = \mathrm{s}'\mathrm{c} + \mathrm{s}\mathrm{c}' = \mathrm{c}^2 \cdot \mathrm{c} + \mathrm{s} \cdot (-\mathrm{s}^2) = \mathrm{c}^3 - \mathrm{s}^3.$$

利用 (i)，$\mathrm{s}^3 = 1 - \mathrm{c}^3$，所以上式等于 $\mathrm{c}^3 - (1 - \mathrm{c}^3) = 2\mathrm{c}^3 - 1$，正是我们想要的结果。

至于商（在 $\mathrm{c}(x) \neq 0$ 之处），商法则（quotient rule）给出

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{s}}{\mathrm{c}}\right) = \frac{\mathrm{s}'\mathrm{c} - \mathrm{s}\mathrm{c}'}{\mathrm{c}^2} = \frac{\mathrm{c}^3 + \mathrm{s}^3}{\mathrm{c}^2} = \frac{1}{\mathrm{c}(x)^2},$$

同样是用了 (i)。留意这与 $\frac{\mathrm{d}}{\mathrm{d}x}\tan x = \frac{1}{\cos^2 x}$ 之间令人愉快的对应。

**第 (iii) 部分**

由于 $\mathrm{c}' = -\mathrm{s}^2$，第一个积分可以立即写出：

$$\int \mathrm{s}(x)^2 \,\mathrm{d}x = -\mathrm{c}(x) + C.$$

? **想一想** 怎样改写 $\mathrm{s}^5$，才能令已知的导数 $\mathrm{c}' = -\mathrm{s}^2$ 出现?

分出一个因子 $\mathrm{s}^2$，再对余下部分使用 (i)：$\mathrm{s}^5 = \mathrm{s}^3 \cdot \mathrm{s}^2 = (1 - \mathrm{c}^3)\,\mathrm{s}^2 = -(1 - \mathrm{c}^3)\,\mathrm{c}'$。这是 $g(\mathrm{c})\,\mathrm{c}'$ 的形式，可用逆连锁法则（reverse chain rule）积分：

$$\int \mathrm{s}(x)^5 \,\mathrm{d}x = -\int (1 - \mathrm{c}^3)\,\mathrm{c}' \,\mathrm{d}x = -\mathrm{c}(x) + \tfrac{1}{4}\mathrm{c}(x)^4 + C.$$

（求导回去验证：$-\mathrm{c}' + \mathrm{c}^3\mathrm{c}' = -\mathrm{c}'(1 - \mathrm{c}^3) = \mathrm{s}^2 \cdot \mathrm{s}^3$，结果理应如此。）

**第 (iv) 部分**

设 $u = \mathrm{s}(x)$，于是 $x = \mathrm{s}^{-1}(u)$，且 $\mathrm{d}u = \mathrm{s}'(x)\,\mathrm{d}x = \mathrm{c}(x)^2\,\mathrm{d}x$。由 (i)，$\mathrm{c}^3 = 1 - u^3$，所以 $\mathrm{c} = (1 - u^3)^{1/3}$（取实立方根，无论 c 的正负号如何都成立），因此

$$(1 - u^3)^{2/3} = \mathrm{c}(x)^2.$$

于是

$$\int \frac{1}{(1 - u^3)^{2/3}}\,\mathrm{d}u = \int \frac{1}{\mathrm{c}^2}\,\mathrm{c}^2\,\mathrm{d}x = \int 1\,\mathrm{d}x = x + C = \mathrm{s}^{-1}(u) + C,$$

这正是我们要证明的。

**第 (v) 部分**

使用同一个代换（substitution），其中 $(1-u^3)^{1/3}=\mathrm{c}$，且 $\mathrm{d}u=\mathrm{c}^2\,\mathrm{d}x$。

对于第一个积分，$(1-u^3)^{4/3}=\mathrm{c}^4$，所以由 (ii) 中关于商的结果，

$$\int\frac{\mathrm{d}u}{(1-u^3)^{4/3}}=\int\frac{\mathrm{c}^2}{\mathrm{c}^4}\,\mathrm{d}x=\int\frac{\mathrm{d}x}{\mathrm{c}^2}=\frac{\mathrm{s}(x)}{\mathrm{c}(x)}+C=\frac{u}{(1-u^3)^{1/3}}+C.$$

对于第二个积分，

$$\int(1-u^3)^{1/3}\,\mathrm{d}u=\int\mathrm{c}\cdot\mathrm{c}^2\,\mathrm{d}x=\int\mathrm{c}^3\,\mathrm{d}x.$$

(ii) 中关于乘积的结果说明 $\mathrm{c}^3=\frac{1}{2}\left(1+\frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{sc})\right)$，所以

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

回到 $u$，其中 $x=\mathrm{s}^{-1}(u)$，且 $\mathrm{sc}=u(1-u^3)^{1/3}$：

$$\int(1-u^3)^{1/3}\,\mathrm{d}u=\tfrac{1}{2}\left(\mathrm{s}^{-1}(u)+u(1-u^3)^{1/3}\right)+C.$$

## 验算

- (v) 的答案可以通过对 $u$ 求导来验证。例如，

$$\frac{\mathrm{d}}{\mathrm{d}u}\,\frac{u}{(1-u^3)^{1/3}}=\frac{1}{(1-u^3)^{1/3}}+\frac{u^3}{(1-u^3)^{4/3}}=\frac{1}{(1-u^3)^{4/3}}.$$

- 以数值方法解微分方程（differential equation）至 $x=1.2$，得 $\mathrm{s}(1.2)\approx0.9411$ 及 $\mathrm{c}(1.2)\approx0.5500$，而确实 $0.9411^3+0.5500^3\approx1$。设 $U=\mathrm{s}(1.2)$，数值积分给出 $\int_0^U(1-u^3)^{-2/3}\,\mathrm{d}u\approx1.2000$，与 (iv) 的预测一致；又 $\int_0^U(1-u^3)^{1/3}\,\mathrm{d}u\approx0.8588=\frac{1}{2}(1.2+0.9411\times0.5500)$，与 (v) 吻合。
- 数值上，$\int_0^{1.2}\mathrm{s}^2\,\mathrm{d}x\approx0.4500=1-\mathrm{c}(1.2)$，与 (iii) 一致。

## 常见错误

- 写成 $\int\mathrm{s}^2\,\mathrm{d}x=+\mathrm{c}+C$：$\mathrm{c}'=-\mathrm{s}^2$ 中的负号很容易遗漏。
- 在 (iv) 和 (v) 中，忘记把 $\mathrm{d}u$ 转换成 $\mathrm{c}^2\,\mathrm{d}x$，或者最终答案仍以 $x$ 而非 $u$ 表示。
- 在 (v) 中试图直接积分 $\mathrm{c}^3$，而没有察觉 (ii) 早已把它表示为一个导数加一个常数。

## STEP 之外

对曲线 $x^3+y^3=1$ 而言, s, c 这一对函数的角色, 就如 sin 和 cos 之于圆 $x^2+y^2=1$; 这类“广义三角函数”(generalised trigonometric functions) 与椭圆函数 (elliptic functions) 是近亲, 后者正是通过反演类似 (iv) 中的积分而产生的, 就如 $\sin^{-1}u=\int(1-u^2)^{-1/2}\,\mathrm{d}u$ 一样。以带初始条件的微分方程来定义函数, 再由此推导出它们的全部性质, 是大学一年级的标准技巧; 而题目中假设这些函数是“唯一确定”的, 背后就是常微分方程 (ODE) 的存在唯一性定理 (existence and uniqueness theorem)。Boyce 和 DiPrima 的 *Elementary Differential Equations*（《基础微分方程》）既涵盖相关理论, 也有许多耦合方程组的例子。

详细解答

# STEP I 2018，第 9 题

力学 · 斜面上的运动

**STEP I 2018，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：

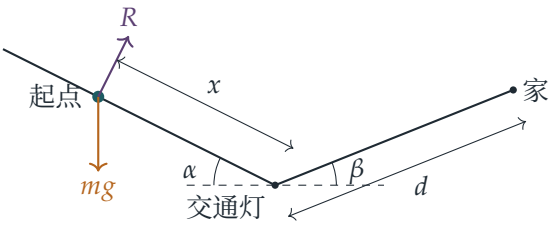
step.maths.org/questions/18-s1-q9

## 关键思路

由于没有阻力，小型赛车（go-kart）在每段直路上的加速度都是常数：沿第一段斜坡向下的加速度为 $g\sin\alpha$，沿第二段斜坡向上则有大小为 $g\sin\beta$ 的减速度。因此每一段都是一个等加速度（constant acceleration）问题，而两段之间由到达交通灯时的速率连接起来。能量（或 $v^2 = u^2 + 2as$）决定我到底能否到家；时间则由每一段上的 $s = ut + \frac{1}{2}at^2$ 得出，其余就是以 $k$ 表示的细心代数运算。

## 解答

把小型赛车连同驾驶者模拟为一个质点（particle），沿每条最大斜率直线（line of greatest slope）运动，没有摩擦力，也没有空气阻力。所受的力只有重量 $mg$ 和法向反作用力（normal reaction）$R$，而后者不做功。



### 到达家门

沿第一段斜坡分解力，重量沿斜坡向下的分量为 $mg\sin\alpha$，所以加速度为 $g\sin\alpha$。由静止出发并走过距离 $x$ 后，到达交通灯时的速率 $V$ 满足

$$V^2 = 2gx\sin\alpha.$$

由于速率没有突变，我以速率 $V$ 开始驶上第二段，此时重量沿斜坡产生大小为 $g\sin\beta$ 的减速度。在交通灯以外距离 $s$ 处，我的速率 $v$ 满足 $v^2 = V^2 - 2gs\sin\beta$。要到达家门，$v^2$ 必须一直保持非负，直至 $s = d$，即

$$2gx\sin\alpha - 2gd\sin\beta \geq 0 \quad \Longrightarrow \quad x\sin\alpha \geq d\sin\beta,$$

这正是所求。（等价地说：根据能量守恒（conservation of energy），我永远不能升到高于起点的高度，而起点比交通灯高 $x\sin\alpha$，家则比交通灯高 $d\sin\beta$。）

### 总时间

为使代数简洁，设 $c = \left(\frac{1}{2}g\sin\alpha\right)^{1/2}$，则 $g\sin\alpha = 2c^2$；又因 $\sin\beta = \sin\alpha/k$，故 $g\sin\beta = 2c^2/k$。

*第一段*。由 $x = \frac{1}{2}(g\sin\alpha)t_1^2 = c^2t_1^2$ 可知到达交通灯的时间为 $t_1 = \sqrt{x}/c$，而该处的速率为 $V = (g\sin\alpha)t_1 = 2c\sqrt{x}$。

*第二段*。从交通灯起量度 $s$，有 $s = Vt - \frac{1}{2}(g\sin\beta)t^2$，所以到达 $s = d$ 的时间 $t_2$ 满足

$$\frac{c^2}{k}t_2^2 - 2c\sqrt{x}\,t_2 + d = 0.$$

**想一想** 这个二次方程有两个根。哪一个才是我到达的时间?

较大的根是：假如斜坡无限长，质点停下后沿斜坡滑回，再次经过 $s = d$ 的时间。我是在第一次到达 $s = d$ 时到家，所以我们要取较小的根：

$$t_2 = \frac{2c\sqrt{x} - \sqrt{4c^2x - 4c^2d/k}}{2c^2/k} = \frac{k}{c}\left(\sqrt{x} - \sqrt{x - \tfrac{d}{k}}\right) = \frac{1}{c}\left(k\sqrt{x} - \sqrt{k^2x - kd}\right),$$

这里因 $k > 0$ 而用了 $k\sqrt{x - d/k} = \sqrt{k^2x - kd}$。平方根之所以是实数，正是因为 $x \geq d/k$，也就是第一部分的条件 $x\sin\alpha \geq d\sin\beta$。

相加得 $T = t_1 + t_2$，于是

$$\left(\frac{g\sin\alpha}{2}\right)^{1/2} T = cT = (1+k)\sqrt{x} - \sqrt{k^2x - kd},$$

这就是所求的结果。

### 使 $T$ 最小

由于 $c$ 是正常数，我们在 $x \geq d/k$ 上求 $f(x) = (1+k)\sqrt{x} - \sqrt{k^2x - kd}$ 的最小值。求导（differentiating）得

$$f'(x) = \frac{1+k}{2\sqrt{x}} - \frac{k^2}{2\sqrt{k^2x - kd}}.$$

令 $f'(x) = 0$，得 $(1+k)\sqrt{k^2x - kd} = k^2\sqrt{x}$。两边皆为正，所以平方不会失去任何信息：

$$\begin{aligned}(1+k)^2(k^2x - kd) &= k^4x \\ x\left[(1+k)^2k^2 - k^4\right] &= (1+k)^2kd \\ xk^2(1+2k) &= (1+k)^2kd,\end{aligned}$$

这里用了 $(1+k)^2 - k^2 = 1 + 2k$。除以 $k^2(1+2k) > 0$，得

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

这是一个容许的起点：$(1+k)^2 = 1 + 2k + k^2 > 1 + 2k$，所以这个 $x$ 严格大于 $d/k$，我确实能到家。

存在最佳的 $x$ 是合理的：起点太近，我在山坡最后一段就会慢得像爬行；起点太远，我又要在第一段斜坡上花很长时间。

## 验算

- 在最佳点，$\sqrt{k^2x - kd} = k^2\sqrt{x}/(1+k)$，所以 $cT = \sqrt{x}\,\dfrac{1+2k}{1+k} = \sqrt{\dfrac{(1+2k)d}{k}}$，最短时间的形式简洁得令人愉快。
- 取 $\alpha = \beta$，即 $k = 1$。此时最佳点为 $x = \frac{4}{3}d$，公式给出 $cT = 2\sqrt{4d/3} - \sqrt{d/3} = \sqrt{3d}$，与上一项验算吻合。
- 在临界值 $x = d/k$ 时，我只是刚好慢慢挪到门口，且 $f'(x) \to -\infty$，所以把起点稍为后移必定能节省时间。
- 以细小的时间间隔逐步数值推进运动，并对多组随机角度和距离进行测试，结果与 $T$ 的公式在五位有效数字内一致；以电脑搜索 $x$，找到的最小值亦正是上面所得的值。

## 常见错误

- 为 $t_2$ 取了二次方程的较大根：那是滑回下坡时经过家门的时间。
- 忘记交通灯处的速率并非已知，必须由第一段延续过来。

► 遗漏条件 $x \geq d/k$，而正是这个条件使 $\sqrt{k^2x - kd}$ 为实数。

## STEP 之外

最短时间的简洁答案暗示了一个更广阔的主题：选择路径或起点以使行程时间最短，正是*变分法*（calculus of variations）的研究对象；其最著名的问题——最速降线（brachistochrone）问题——就是要找出在重力作用下下降最快的曲线。在大学一年级的力学课程中，你亦会接触到这个想法：对于保守力（conservative forces），单靠能量就能告诉你质点可以到达哪里，而要求时间则必须对运动方程（equation of motion）积分。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对这两个想法都有细致的阐述。

详细解答

# STEP I 2018，第 10 题

力学 · 连接质点

**STEP I 2018，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：

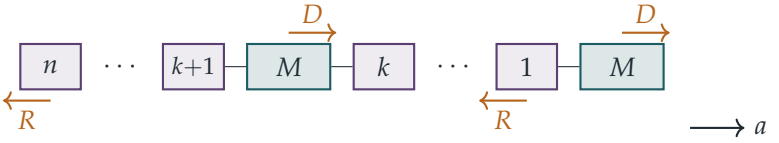
step.maths.org/questions/18-s1-q10

## 关键思路

每节车辆的加速度（acceleration）都相同，所以我们先由整列火车求出加速度一次，再对火车中最方便的那一部分应用牛顿第二定律（Newton's second law）：在某个连接器（coupling）处把火车“切开”，该连接器的张力（tension）就会显露为一个外力（external force）。令后面各部分变得容易的技巧，是用两种方法求最前方的张力 $T$：一次从前方机车入手，一次从它后面的所有车辆入手。这样做会把 $T$ 与单一个量 $R+ma$ 联系起来，而这个量同时支配著其他每一个连接器。

## 解答

**模型。** 把每辆机车和每节车厢视为沿水平路轨运动的质点（particle），并把连接器视为可以受拉或受压（compression）的轻杆（light rods）。机车不受阻力（resistance），而每节车辆所受的铅垂方向的力互相平衡，所以只需考虑水平方向的力。加速度 $a$ 以运动方向为正。由前至后把车厢编号为 $1,2,\ldots,n$，使车厢 $1,\ldots,k$ 位于两辆机车之间，而 $k+1,\ldots,n$ 位于第二辆机车之后（其中 $1\le k\le n-1$）。




（每节车厢都受一个向后的阻力 $R$；图中只画出其中两个。）

**加速度。** 整列火车的质量为 $nm+2M$，而连接器的力属于内力，所以

$$(nm+2M)a = 2D - nR, \qquad a = \frac{2D-nR}{nm+2M} > 0,$$

由于 $2D > nR$，所以加速度为正。特别地，$R+ma>0$，这个事实我们会反复用到。

### 第 (i) 部分

前方机车受向前的力 $D$，以及其连接器向后的张力 $T$，所以

$$D - T = Ma.$$

因此

$$\begin{aligned} T = D - \frac{M(2D-nR)}{nm+2M} &= \frac{D(nm+2M) - 2MD + nMR}{nm+2M} \\ &= \frac{n(mD+MR)}{nm+2M}, \end{aligned}$$

即为所求。

**$T$ 的一个更简洁形式。** 现在改为考虑前方机车后面的所有车辆：一辆机车和 $n$ 节车厢，总质量为 $M+nm$。它们受 $T$ 和第二辆机车的 $D$ 向前拉动，并受阻力 $nR$：

$$T + D - nR = (M+nm)a.$$

从此式减去 $D - T = Ma$ 便可消去 $D$：

$$2T - nR = nma, \qquad \text{所以} \qquad T = \tfrac{1}{2}n\,(R+ma).$$

### 第 (ii) 部分

除最前方的连接器外，其他每个连接器都属于以下两类之一。

**第二辆机车前面的连接器。** 对于 $1 \le j \le k$，设 $T_j$ 为紧接车厢 $j$ 后方的连接器的张力（因此 $T_k$ 是车厢 $k$ 与第二辆机车之间的连接器）。前方机车连同车厢 $1,\dots,j$ 的质量为 $M+jm$，并受 $T_j$ 向后拉：

$$D - jR - T_j = (M+jm)a \quad \Longrightarrow \quad T_j = (D-Ma) - j(R+ma) = T - j(R+ma).$$

由于 $R+ma>0$ 且 $j \ge 1$，无论 $k$ 是多少，这类连接器全部都有 $T_j < T$。

**第二辆机车后面的连接器。** 对于 $k+1 \le j \le n$，设 $S_j$ 为紧接车厢 $j$ 前方的连接器的张力。最简单的做法是考虑它后面的车辆：车厢 $j,\dots,n$，共 $n-j+1$ 节，只受 $S_j$ 向前拉：

$$S_j - (n-j+1)R = (n-j+1)ma \quad \Longrightarrow \quad S_j = (n-j+1)(R+ma).$$

**想一想** 这些后方连接器中，哪一个的张力最大？它与 $T$ 相比又如何？

张力 $S_j$ 随 $j$ 增大而减小，所以最大的是 $S_{k+1} = (n-k)(R+ma)$，位于紧接第二辆机车后方的连接器。将它与 $T = \frac{1}{2}n(R+ma)$ 比较，并除以 $R+ma > 0$，得

$$S_{k+1} < T \iff n-k < \tfrac{1}{2}n \iff k > \tfrac{1}{2}n.$$

所以若 $k > \frac{1}{2}n$，每个后方连接器的张力至多为 $S_{k+1} < T$，而每个前方连接器都有 $T_j < T$。因此 $T$ 大于其他任何连接器的张力，即为所求。（这个条件也是必要的：若 $k \le \frac{1}{2}n$，则 $S_{k+1} \ge T$。）

### 第 (iii) 部分

考虑车厢 $k$ 与第二辆机车之间的连接器。由第 (ii) 部分及 $T = \frac{1}{2}n(R+ma)$，

$$T_k = T - k(R+ma) = \left(\tfrac{1}{2}n - k\right)(R+ma).$$

若 $k > \frac{1}{2}n$，第一个因子为负而第二个因子为正，所以 $T_k < 0$：这个连接器处于受压状态。从物理角度看，第二辆机车正在推动它前面的 $k$ 节车厢（超过整列火车的一半），同时前方机车在拉动它们。

## 验算

- 取 $n=5$、$k=3$、$M=10$、$m=2$、$D=20$、$R=1$。则 $a = \frac{35}{30} = \frac{7}{6}$，公式给出 $T = \frac{5(40+10)}{30} = \frac{25}{3}$，与 $D - Ma = 20 - \frac{35}{3}$ 相符。沿火车由前往后，各张力依次为 $\frac{25}{3}, 5, \frac{5}{3}, -\frac{5}{3}$（车厢 3 至机车），然后是 $\frac{20}{3}, \frac{10}{3}$。所以 $T$ 最大，而且恰好有一个连接器受压，与预测一致。
- 若 $M=0$，机车就只是力，此时 $T = D$，这正是无质量的前方机车应有的结果。
- 每个张力都是一个力乘以一个无量纲（dimensionless）比值，所以单位正确。

## 常见错误

- 把阻力也加在机车上：题目只说车厢受阻力。
- 忘记第二辆机车的驱动力作用于任何包含它的火车部分。
- 假设每个连接器都必定受拉；连接器是一根杆，所以负张力是完全合乎物理的推力。

## STEP 之外

把一个系统分拆成若干部分，并为每一部分分别画出受力图（free-body diagram），其中内力（internal forces）以大小相等、方向相反的成对形式出现，这是连接体力学的基础。在大学一年级的课程中，这会发展为桁架（truss）和框架（framework）的分析：你要同时求出多根构件的内力，而有些构件会原来是受压的支柱（strut），而不是受拉的拉杆（tie）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）细致地讨论了受力图和牛顿第三定律（Newton's third law）。

详细解答

# STEP I 2018，第 11 题

力学 · 斜面上的摩擦力

**STEP I 2018，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/18-s1-q11

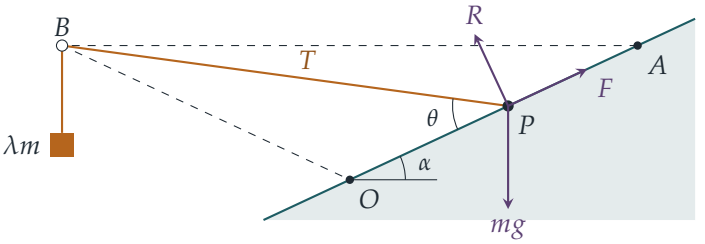
## 关键思路

先从几何入手：$OA = OB = d$，所以三角形 $OAB$ 是等腰三角形（isosceles triangle）；又因为 $AB$ 是水平的，它的两个底角都是 $\alpha$。绳子由 $P$ 往回越过 $O$ 的一侧伸向 $B$，因此它一方面把 $P$ 沿斜面向下拉，另一方面把 $P$ 拉离斜面。于是摩擦力（friction）必须把质点托住，而法向反作用力（normal reaction）则因张力（tension）而减小。每个平衡（equilibrium）条件都是一条关于 $\theta$ 的不等式，而且必须对第 (i) 部分所得范围内的每一个 $\theta$ 都成立，所以我们要找出最不利的位置。

## 解答

把 $P$ 和悬挂的物体视为质点（particle），绳子视为轻而不可伸长（light and inextensible），滑轮视为光滑（smooth），因此整条绳子的张力处处相同。悬挂的质点处于平衡，所以

$$T = \lambda mg.$$



### 第 (i) 部分

把 $O$ 放在原点，$x$ 轴取水平方向。于是 $A = (d\cos\alpha,\ d\sin\alpha)$。点 $B$ 与 $A$ 等高，且与 $O$ 的距离为 $d$，所以 $B = (\pm d\cos\alpha,\ d\sin\alpha)$。取正号得到的就是 $A$ 本身，因此

$$B = (-d\cos\alpha,\ d\sin\alpha).$$

此点位于斜面所在直线的上方，并在 $O$ 相对于 $A$ 的另一侧。在三角形 $OAB$ 中，$OA = OB = d$；又因为 $AB$ 水平而 $AO$ 与水平线成角 $\alpha$，所以 $\angle OAB = \alpha$。因此 $\angle OBA = \alpha$ 亦成立，而 $\angle AOB = 180^\circ - 2\alpha$；由于 $\alpha < 45^\circ$，这是一个钝角。

设 $P$ 位于 $OA$ 上。绳子沿 $PB$ 伸展，它与 $L$ 所成的锐角为 $\theta = \angle BPO$。这个角是锐角，因为它至多是 $2\alpha < 90^\circ$，下面就来证明。它是三角形 $PAB$ 的一个外角（exterior angle），所以

$$\theta = \angle BPO = \angle PAB + \angle PBA = \alpha + \angle PBA.$$

当 $P$ 由 $O$ 移向 $A$ 时，射线 $BP$ 由 $BO$ 转向 $BA$，所以 $\angle PBA$ 由 $\angle OBA = \alpha$ 递减至 0。因此

$$\alpha \le \theta \le 2\alpha,$$

当 $P$ 在 $O$ 时 $\theta = 2\alpha$，而当 $P$ 趋近 $A$ 时 $\theta \to \alpha$。

留意这对各力意味着甚么。从沿斜面向下的方向 $PO$ 量起，绳子成锐角 $\theta$，并指向离开斜面的一方。所以张力有沿斜面向下的分量 $T\cos\theta$，以及垂直于斜面、把 $P$ 拉离斜面的分量 $T\sin\theta$。

### 第 (ii) 部分

沿垂直于斜面的方向分解（resolving）：

$$R = mg\cos\alpha - T\sin\theta = mg(\cos\alpha - \lambda\sin\theta).$$

> ? **想一想** 在这一部分中，既然我们不知道 $\mu$，摩擦力去了哪里?

我们根本不需要它。无论 $\mu$ 是多少，$P$ 都必须与斜面保持接触，所以需要 $R \ge 0$。（事实上 $R > 0$: 若 $R = 0$, 就没有摩擦力, 沿斜面方向便没有任何力可以平衡 $mg\sin\alpha + T\cos\theta > 0$。）因此，要在每一点都能平衡，必须有

$$\lambda\sin\theta \le \cos\alpha \quad \text{对所有出现的}\theta\ \text{成立.}$$

由于 sin 在 $[\alpha, 2\alpha] \subset [0^\circ, 90^\circ)$ 上递增，$\sin\theta$ 的最大值出现在 $\theta = 2\alpha$（或当 $P$ 趋近 $O$ 时任意接近它）。因此 $\lambda\sin 2\alpha \le \cos\alpha$，即 $2\lambda\sin\alpha\cos\alpha \le \cos\alpha$。两边除以 $\cos\alpha > 0$，得

$$2\lambda\sin\alpha \le 1,$$

即为所求。

**第 (iii) 部分**

重力和张力都把 $P$ 沿斜面向下拉，所以摩擦力 $F$ 沿斜面向上作用。沿斜面方向分解，

$$F = mg\sin\alpha + T\cos\theta = mg(\sin\alpha + \lambda\cos\theta).$$

在此位置能够平衡，当且仅当 $F \le \mu R$（由于 $F > 0$，这同时迫使 $R > 0$）。设 $\mu = \tan\beta$，条件为

$$\sin\alpha + \lambda\cos\theta \le \frac{\sin\beta}{\cos\beta}(\cos\alpha - \lambda\sin\theta).$$

两边乘以 $\cos\beta > 0$，并把含 $\lambda$ 的项归在一起：

$$\lambda(\cos\theta\cos\beta + \sin\theta\sin\beta) \le \sin\beta\cos\alpha - \cos\beta\sin\alpha,$$
$$\lambda\cos(\beta - \theta) \le \sin(\beta - \alpha).$$

这里 $0 < \beta < 90^\circ$ 且 $0 < \theta \le 2\alpha < 90^\circ$，所以 $-90^\circ < \beta - \theta < 90^\circ$，从而 $\cos(\beta - \theta) > 0$。因此，在角度为 $\theta$ 的位置能够平衡，恰好当

$$\lambda \le \frac{\sin(\beta - \alpha)}{\cos(\beta - \theta)}.$$

要使此式在每一个位置都成立，它必须在 $\cos(\beta - \theta)$ 最大之处成立，也就是在范围 $\alpha \le \theta \le 2\alpha$ 容许下 $\theta$ 最接近 $\beta$ 之处。

*情况 $\beta \ge 2\alpha$。*范围内每一个 $\theta$ 都不超过 $\beta$，所以 $0 \le \beta - 2\alpha \le \beta - \theta \le \beta - \alpha < 90^\circ$。余弦在 $[0^\circ, 90^\circ)$ 上递减，所以 $\cos(\beta - \theta)$ 在 $\theta = 2\alpha$ 处最大，即位于 $O$（或紧邻 $O$）的位置。因此一个必要条件是

$$\lambda \le \frac{\sin(\beta - \alpha)}{\cos(\beta - 2\alpha)}.$$

这也是充分条件，因为在其他每一个位置，对 $\lambda$ 的限制都较宽松。

*情况 $\alpha \le \beta \le 2\alpha$。*此时 $\theta = \beta$ 落在范围之内：由第 (i) 部分，$P$ 在 $L$ 上介于 $O$ 与 $A$ 之间的某个位置恰好给出这个角（若 $\beta = \alpha$，则是当 $P \to A$ 时趋近）。在那里 $\cos(\beta - \theta) = 1$，即其最大可能值，条件变为

$$\lambda \le \sin(\beta - \alpha).$$

在该位置，绳子垂直于总接触力；当摩擦力达到极限时，总接触力与法线成角 $\beta$。若 $\beta = \alpha$，这迫使 $\lambda = 0$：此时质点单凭自身重量就已刚好处于极限平衡（limiting equilibrium），任何绳子张力都会令它滑动。

## 验算

► 当 $\beta = 2\alpha$ 时两个答案一致：此时 $\cos(\beta - 2\alpha) = 1$，两者都给出 $\lambda \le \sin\alpha$。

- 当 $\lambda = 0$ 时，条件变为 $\sin(\beta - \alpha) \geq 0$，即 $\mu \geq \tan\alpha$，这正是物块能静止于粗糙斜面上的熟悉条件。
- 对于 $\alpha = 20°$、$\beta = 50°$，界限为 $\sin 30° / \cos 10° \approx 0.508$。一项电脑验算把 $P$ 放在 $OA$ 上的许多点，直接分解各力，结果发现对于数以千计的随机 $\alpha$ 和 $\beta$，使各处皆能平衡的最大 $\lambda$ 都与两条公式吻合。它亦证实了 $\alpha \leq \theta \leq 2\alpha$。

## 常见错误

- 把 $B$ 放在 $O$ 与 $A$ 同一侧。由于 $OB = OA$ 而 $B \neq A$，$B$ 位于 $O$ 的另一侧，所以张力把 $P$ 沿斜面向下拉，而不是向上。
- 忘记张力会减小法向反作用力，从而减少可用的摩擦力。
- 在第 (iii) 部分的第二种情况中只检查两端的位置。最不利的位置是 $\theta = \beta$ 之处，它位于内部。

## STEP 之外

组合式 $\lambda\cos(\beta - \theta) \leq \sin(\beta - \alpha)$ 是摩擦角（angle of friction）的代数形式：总接触力必须落在与法线成角 $\beta$ 的范围之内，即所谓的摩擦锥（cone of friction）。只要从这个锥的角度思考，许多摩擦问题都能化为一行几何。要求“对每一个位置”都能平衡，其实是一个小型的最优化问题；工程师按最坏情况进行设计时，也是用同样的方式推理。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（《力学导论》）细致地处理摩擦力和受力图，并附有大量具挑战性的习题。

详细解答

# STEP I 2018，第 12 题

概率与统计 · 放回抽样

> **STEP I 2018，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/18-s1-q12

## 关键思路

把每个正面次数写成两个指示变量（indicator variable）之和，每次抛掷对应一个。无论各次抛掷是否独立（independent），和的期望值（expectation）都等于期望值之和，所以平均值很容易求得。方差（variance）则取决于一个数，就是两次抛掷皆出现正面的概率，而这正是放回抽样（drawing with replacement）与不放回抽样（drawing without replacement）的分别所在。

## 解答

全文中，设 $p = \frac{1}{3}(p_1 + p_2 + p_3)$ 及

$$q = p_1p_2 + p_2p_3 + p_3p_1.$$

我们假设每枚硬币被选中的机会相等，而且硬币选定后，各次抛掷互相独立。

### 第 (i) 部分

设 $C_k$ 为选中第 $k$ 枚硬币的事件，故 $\mathrm{P}(C_k) = \frac{1}{3}$。由全概率定律（law of total probability），

$$\mathrm{P}(\text{正面}) = \sum_{k=1}^{3} \mathrm{P}(C_k)\,\mathrm{P}(\text{正面} \mid C_k) = \tfrac{1}{3}(p_1 + p_2 + p_3) = p.$$

### 第 (ii) 部分

设 $X_1$ 和 $X_2$ 分别为第一次和第二次抛掷出现正面的指示变量，故 $N_1 = X_1 + X_2$。由于第一枚硬币会被放回，第二次抛掷的情况与第一次完全相同，由第 (i) 部分得 $\mathrm{P}(X_1 = 1) =$

$\mathrm{P}(X_2 = 1) = p$。因此

$$\mathrm{E}(N_1) = \mathrm{E}(X_1) + \mathrm{E}(X_2) = 2p.$$

求方差需要 $\mathrm{P}(X_1 = 1,\, X_2 = 1)$。两次选择互相独立，9 个有序硬币对 $(j,k)$ 的概率各为 $\frac{1}{9}$，而给定该硬币对时，两次皆为正面的机会是 $p_j p_k$。所以

$$\mathrm{P}(X_1 = 1,\, X_2 = 1) = \frac{1}{9}\sum_{j=1}^{3}\sum_{k=1}^{3} p_j p_k = \left(\frac{p_1 + p_2 + p_3}{3}\right)^2 = p^2 = \mathrm{P}(X_1 = 1)\,\mathrm{P}(X_2 = 1).$$

同样的因式分解对其他结果亦成立，因此 $X_1$ 与 $X_2$ 互相独立，且 $N_1 \sim \mathrm{B}(2, p)$。也可以直接计算：由于 $X_i^2 = X_i$，

$$\begin{aligned}\mathrm{E}(N_1^2) &= \mathrm{E}(X_1) + \mathrm{E}(X_2) + 2\,\mathrm{E}(X_1X_2) = 2p + 2p^2,\\ \mathrm{Var}(N_1) &= 2p + 2p^2 - (2p)^2 = 2p(1-p),\end{aligned}$$

正如所求。

**第 (iii) 部分**

现在以不放回方式先后选出两枚硬币，设 $Y_1$、$Y_2$ 分别为第一枚和第二枚选出的硬币出现正面的指示变量，故 $N_2 = Y_1 + Y_2$。每枚硬币成为第一枚被选中的机会相等，由对称性，每枚成为第二枚被选中的机会亦相等，所以与第 (i) 部分一样，$\mathrm{P}(Y_1 = 1) = \mathrm{P}(Y_2 = 1) = p$，且

$$\mathrm{E}(N_2) = 2p.$$

? **想一想** 可以沿用第 (ii) 部分中 P(两次皆正面) 的计算吗?

不可以：同一枚硬币不能再被选中两次。现在有 6 个机会相等的有序对 $(j,k)$，其中 $j \neq k$，而每个无序对 $\{j,k\}$ 出现两次，所以

$$\mathrm{P}(Y_1 = 1,\, Y_2 = 1) = \frac{1}{6}\sum_{j\neq k} p_j p_k = \frac{2q}{6} = \frac{q}{3}.$$

一般而言这并不等于 $p^2$，故 $Y_1$ 与 $Y_2$ 并不独立。再次利用 $Y_i^2 = Y_i$，

$$\begin{aligned}\mathrm{E}(N_2^2) &= \mathrm{E}(Y_1) + \mathrm{E}(Y_2) + 2\,\mathrm{E}(Y_1Y_2) = 2p + \tfrac{2}{3}q,\\ \mathrm{Var}(N_2) &= 2p + \tfrac{2}{3}(p_1p_2 + p_2p_3 + p_3p_1) - 4p^2.\end{aligned}$$

**第 (iv) 部分**

两式相减，

$$\mathrm{Var}(N_1) - \mathrm{Var}(N_2) = \left(2p - 2p^2\right) - \left(2p + \tfrac{2}{3}q - 4p^2\right) = 2p^2 - \tfrac{2}{3}q.$$

而 $9p^2 = (p_1+p_2+p_3)^2 = p_1^2+p_2^2+p_3^2+2q$，所以

$$\begin{aligned} 2p^2 - \tfrac{2}{3}q &= \tfrac{2}{9}\left(p_1^2+p_2^2+p_3^2+2q\right) - \tfrac{6}{9}q = \tfrac{2}{9}\left(p_1^2+p_2^2+p_3^2-p_1p_2-p_2p_3-p_3p_1\right) \\ &= \tfrac{1}{9}\Big[(p_1-p_2)^2+(p_2-p_3)^2+(p_3-p_1)^2\Big], \end{aligned}$$

最后一步成立，是因为展开这三个平方会得到 $2(p_1^2+p_2^2+p_3^2)-2q$。

实数的平方和是非负的，所以 $\mathrm{Var}(N_2) \le \mathrm{Var}(N_1)$。等号成立当且仅当三个平方皆为零，即当且仅当 $p_1 = p_2 = p_3$。证明完毕。

作为副产品，我们得到第 (iii) 部分答案的一个简洁形式：

$$\mathrm{Var}(N_2) = 2p(1-p) - \tfrac{1}{9}\Big[(p_1-p_2)^2+(p_2-p_3)^2+(p_3-p_1)^2\Big].$$

## 验算

- 若 $p_1 = p_2 = p_3 = p$，每枚硬币都一样，那么硬币怎样选出并无分别：两个次数都服从 $\mathrm{B}(2,p)$，方差相等，与等号成立的情况吻合。
- 取 $p_1 = \frac{1}{2}$、$p_2 = \frac{1}{3}$、$p_3 = \frac{1}{6}$，则 $p = \frac{1}{3}$，$q = \frac{11}{36}$。于是 $\mathrm{Var}(N_1) = \frac{4}{9}$，$\mathrm{Var}(N_2) = \frac{2}{3} + \frac{11}{54} - \frac{4}{9} = \frac{23}{54}$，少了 $\frac{1}{54}$，与 $\frac{1}{9}\left(\frac{1}{36}+\frac{1}{36}+\frac{1}{9}\right) = \frac{1}{54}$ 一致。亲手列出所有结果亦得出相同数值。
- 极端情况 $p_1 = p_2 = 1$、$p_3 = 0$：在第 (iii) 部分中，除非第 1 和第 2 枚硬币同时被选中，否则选出的两枚硬币中恰有一枚必定是正面，因此 $N_2$ 以概率 $\frac{1}{3}$ 取值 2，其余情况取值 1，方差为 $\frac{2}{9}$。公式给出 $2p + \frac{2}{3}q - 4p^2 = \frac{4}{3} + \frac{2}{3} - \frac{16}{9} = \frac{2}{9}$。

## 常见错误

- 把第 (iii) 部分的两次抛掷当作独立，于是又得到 $2p(1-p)$；第 (iv) 部分的重点正正在于它们并不独立。
- 把 $\mathrm{Var}(N_1)$ 写成各枚硬币方差 $2p_k(1-p_k)$ 的平均；这忽略了随机选择硬币所带来的额外变异，而且在第 (ii) 部分中，每次抛掷可能用不同的硬币。

## STEP 之外

第 (ii) 与第 (iii) 部分的分别，就是放回抽样与不放回抽样的分别。不放回抽样使各次抽取呈负相关（negatively correlated），从而降低方差；同样的效应在抽样调查中产生“有限总体校正”（finite population correction），亦解释了为何超几何分布（hypergeometric distribution）的方差比具相同平均值的二项分布（binomial distribution）小。第 (iv) 部

分的代数运算就是恒等式 $\sum p_k^2 - \sum_{j<k} p_j p_k = \frac{1}{2}\sum_{j<k}(p_j - p_k)^2$，它与“方差永不为负”这个事实关系密切。Blitzstein 和 Hwang 的 *Introduction to Probability*（《概率导论》）正是以这种精神讲解指示变量、协方差（covariance）和超几何分布。

详细解答

# STEP I 2018，第 13 题

概率与统计 • 二项分布与贝叶斯定理

> **STEP I 2018，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/18-s1-q13

## 关键思路

一切都建基于一个简单的计算：若考生作答 $k$ 题，其中答对 $j$ 题，则得分为 $2j-(k-j)=3j-k$。因此合格即表示至少答对 $(k+5)/3$ 题，而对每个 $k$，这就准确告诉我们要把哪些二项式项相加。在第 (iii) 部分，巧妙的做法是把每一题看成有三种结果（答对、答错、未作答），这样整个问题就化为单一的多项计数（multinomial count）。

## 解答

全文设 $p=\frac{1}{n}$ 为随机选出的答案正确的概率（probability），并设 $q=1-p$。由于 $n\geq 2$，故 $0<p\leq\frac{1}{2}$。

### 第 (i) 部分

设 $P_k$ 为 A 恰好作答 $k$ 题时合格的概率。她的各次选择互相独立（independent），所以她答对的题数 $J$ 服从二项分布（binomial distribution）$\mathrm{B}(k,p)$。她的得分为 $3J-k$，所以她合格当且仅当

$$3J-k\geq 5 \iff J\geq\frac{k+5}{3}.$$

逐一考虑 $k$ 的各个值：

- $k=0,1,2$：分别需要 $J\geq\frac{5}{3},2,\frac{7}{3}$，即 $J\geq 2$、$J\geq 2$ 及 $J\geq 3$，但 $J\leq k$，故这些都不可能（最高得分为 $2k\leq 4$）。所以 $P_0=P_1=P_2=0$。
- $k=3$：需要 $J\geq\frac{8}{3}$，即 $J=3$，所以 $P_3=p^3$。

▶ $k=4$：需要 $J \geq 3$，所以

$$P_4 = \binom{4}{3} p^3 q + p^4 = 4p^3(1-p) + p^4 = p^3(4-3p).$$

▶ $k=5$：需要 $J \geq \frac{10}{3}$，即 $J=4$ 或 5，所以

$$P_5 = \binom{5}{4} p^4 q + p^5 = 5p^4(1-p) + p^5 = p^4(5-4p).$$

现在把 $P_4$ 与它的两个对手比较。由于 $p<1$，有 $4-3p>1$，所以 $P_4 > p^3 = P_3$。此外

$$P_4 - P_5 = p^3\big[(4-3p) - p(5-4p)\big] = p^3\big(4p^2 - 8p + 4\big) = 4p^3(1-p)^2 > 0,$$

因为 $0<p<1$。又因 $P_4>0$，亦胜过 $P_0 = P_1 = P_2 = 0$，所以 $P_4$ 严格最大，故 A 应选择 $k=4$，得证。

**想一想** 作答全部五题，岂不是必然有更多得分机会?

在这里并非如此。作答五题时，她需要答对四题，因为第五题一旦答错便会扣一分，而答错的机会远比答对大。作答四题时，答对三题便已足够。因式分解（factorisation）$4p^3(1-p)^2$ 显示，这个优势对每个 $n \geq 2$ 都成立，而不仅限于 $n$ 很大的情况。

**第 (ii) 部分**

设 $K$ 为 B 作答的题数，所以对 $k=0,\ldots,5$ 有 $\mathrm{P}(K=k)=\frac{1}{6}$；并设 $S$ 为 B 合格这一事件。在 $K=k$ 的条件下，B 的处境与 A 完全相同，所以 $\mathrm{P}(S \mid K=k) = P_k$。由贝叶斯定理（Bayes' theorem），

$$\mathrm{P}(K=4 \mid S) = \frac{\frac{1}{6}P_4}{\frac{1}{6}(P_0+P_1+\cdots+P_5)} = \frac{p^3(4-3p)}{p^3 + p^3(4-3p) + p^4(5-4p)}.$$

约去 $p^3$（它不为零）后，分母变为 $1+4-3p+5p-4p^2 = 5+2p-4p^2$，所以

$$\mathrm{P}(K=4 \mid S) = \frac{4-3p}{5+2p-4p^2}.$$

代入 $p=\frac{1}{n}$，并把分子和分母同乘以 $n^2$：

$$\mathrm{P}(K=4 \mid S) = \frac{4n^2-3n}{5n^2+2n-4} = \frac{n(4n-3)}{5n^2+2n-4}.$$

### 第 (iii) 部分

对每一题而言，C 掷硬币的结果与她所选的答案互相独立，亦与其他题目的情况独立。所以每一题各自独立地有三种结果之一：

$$\begin{aligned} \text{答对 } (+2): &\quad \frac{n}{n+1}\cdot\frac{1}{n} = \frac{1}{n+1}, \\ \text{答错 } (-1): &\quad \frac{n}{n+1}\cdot\frac{n-1}{n} = \frac{n-1}{n+1}, \\ \text{未作答 } (0): &\quad \frac{1}{n+1}. \end{aligned}$$

设她答对 $c$ 题、答错 $w$ 题，其中 $c+w\le 5$。她合格当 $2c-w\ge 5$。这迫使 $c\ge 3$，然后：

- $c=3$：需要 $w\le 1$；
- $c=4$：由于 $c+w\le 5$，$w\le 1$ 自动成立；
- $c=5$：$w=0$。

把 $c$ 题答对、$w$ 题答错、$5-c-w$ 题未作答分配到各题的方法数，就是多项系数（multinomial coefficient）$\frac{5!}{c!\,w!\,(5-c-w)!}$。由于答对与未作答的概率同为 $\frac{1}{n+1}$，每一种排列的概率都是 $\frac{(n-1)^w}{(n+1)^5}$。把五种情况相加：

| $(c,w)$ | 排列数 | 概率 $\times(n+1)^5$ |
|---|---|---|
| $(3,0)$ | 10 | 10 |
| $(3,1)$ | 20 | $20(n-1)$ |
| $(4,0)$ | 5 | 5 |
| $(4,1)$ | 5 | $5(n-1)$ |
| $(5,0)$ | 1 | 1 |

总和为 $16+25(n-1)=25n-9$，所以

$$\mathrm{P}(\text{C 合格}) = \frac{25n-9}{(n+1)^5}.$$

## 验算

- 当 $n=2$：$P_3=\frac{1}{8}$，$P_4=\frac{5}{16}$，$P_5=\frac{3}{16}$，所以 $k=4$ 胜出。在 (ii) 中，公式给出 $\frac{2\cdot 5}{20}=\frac{1}{2}$，与直接计算 $\frac{5/16}{10/16}$ 的结果吻合。
- 对于 (iii)，当 $n=3$ 时答案为 $\frac{66}{1024}\approx 0.0645$，而以电脑模拟 200 000 次测验得出 0.0644。三条公式亦与电脑对 $n=2,\dots,29$ 作精确穷举的结果一致。

- 当 $n \to \infty$，(ii) 的答案趋向 $\frac{4}{5}$：当猜题几乎毫无希望时，合格几乎总是来自作答四题（需答对三题），而非作答五题（需答对四题）或作答三题（需三题全对，概率为 $p^3$，远小于 $4p^3$）。

## 常见错误

- 忘记作答五题时答对三题只得 $6-2=4$ 分，并不合格：需要答对四题。
- 在 (ii) 中，分母漏了 $P_3$：作答三题而三题全对，也是合格的一种途径。
- 在 (iii) 中，把问题看成“以概率 $\binom{5}{k}h^k(1-h)^{5-k}$ 作答 $k$ 题”，然后套用第 (i) 部分，这样做是可行的，但很容易漏掉某一项；三种结果的计数方法较为简洁。

## STEP 之外

第 (iii) 部分是稀疏化（thinning）的一个例子：每一题以某个概率被保留，而各结果仍然互相独立，所以答对、答错及跳过的题数服从多项分布（multinomial distribution）。第 (ii) 部分则是贝叶斯更新（Bayesian updating），这是现代统计学很大部分背后的思想。两者在 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率论导引》）以及 Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率论入门》）中都有深入阐述。

详细解答

# STEP II 2018，第 1 题

纯数学 · 多项式的根

**STEP II 2018，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/18-s2-q1

## 关键思路

系数由前往后读和由后往前读都一样，所以把方程除以 $x^2$，就可以把这个四次方程（quartic）化为关于 $t = x + \frac{1}{x}$ 的二次方程（quadratic）。每一个 $t$ 值都对应四次方程的一对根 $k$、$\frac{1}{k}$，而实数 $x$ 只能给出 $|t| \geq 2$，且 $t = \pm 2$ 当且仅当 $x = \pm 1$。于是，数相异根的个数，就归结为：哪些 $t$ 值会出现，以及其中是否有一个等于 $\pm 2$。

## 解答

### 根成倒数（reciprocal）对出现

设 $P(x) = x^4 + ax^3 + bx^2 + ax + 1$。由于 $P(0) = 1 \neq 0$，任何根 $k$ 都不为零，而且

$$k^4 P\left(\frac{1}{k}\right) = 1 + ak + bk^2 + ak^3 + k^4 = P(k) = 0.$$

因为 $k^4 \neq 0$，得 $P(\frac{1}{k}) = 0$，所以 $k^{-1}$ 也是根。

### 代换（substitution）$t = x + \frac{1}{x}$

把 $P(x) = 0$ 除以 $x^2$（因 $x \neq 0$，这是可以的），并利用 $x^2 + \frac{1}{x^2} = t^2 - 2$：

$$\left(x^2 + \frac{1}{x^2}\right) + a\left(x + \frac{1}{x}\right) + b = 0 \iff t^2 + at + (b-2) = 0.$$

设 $t_1$、$t_2$ 为这个二次方程的根，则 $t_1 + t_2 = -a$，$t_1 t_2 = b - 2$。展开可知

$$P(x) = (x^2 - t_1 x + 1)(x^2 - t_2 x + 1),$$

因为右方等于 $x^4 - (t_1 + t_2)x^3 + (2 + t_1 t_2)x^2 - (t_1 + t_2)x + 1$。所以 $P$ 的根就是 $x^2 - t_1 x + 1$ 的根连同 $x^2 - t_2 x + 1$ 的根。

现在运用“所有根都是实数”这个假设。每个因式 $x^2 - t_i x + 1$ 都有两个实根 $k$ 和 $\frac{1}{k}$（它们的积是 1），而 $t_i = k + \frac{1}{k}$ 是实数。对于实数 $k \neq 0$，$k + \frac{1}{k}$ 与 $k$ 同号，而且

$$\left|k + \tfrac{1}{k}\right| - 2 = \frac{(|k| - 1)^2}{|k|} \geq 0,$$

等号只在 $k = \pm 1$ 时成立。因此：

- 若 $|t_i| > 2$，因式 $x^2 - t_i x + 1$ 有两个相异实根，两者都不等于 $\pm 1$；
- 若 $t_i = 2$，该因式为 $(x-1)^2$；若 $t_i = -2$，则为 $(x+1)^2$。

此外，不同的 $t$ 值不可能有公共根 $x$，因为 $x$ 决定了 $t = x + \frac{1}{x}$。

### 第 (i) 部分

若只有一个相异根 $k$，则 $\frac{1}{k} = k$，所以 $k = \pm 1$，且 $P(x) = (x \mp 1)^4$。展开得

$$(x-1)^4 = x^4 - 4x^3 + 6x^2 - 4x + 1, \qquad (x+1)^4 = x^4 + 4x^3 + 6x^2 + 4x + 1,$$

两者都具有所需的形式。所以所求的值为 $a = -4$, $b = 6$ 及 $a = 4$, $b = 6$。

### 第 (ii) 部分

逐一考虑 $t_1$、$t_2$ 的各种可能（两者都是实数，且 $|t_i| \geq 2$）：

- $t_1 \neq t_2$，且两者都有 $|t_i| > 2$：四个相异根。
- $t_1 = t_2$：四次式是完全平方（perfect square）$(x^2 - t_1 x + 1)^2$，所以若 $|t_1| > 2$，有两个相异根；若 $t_1 = \pm 2$，则只有一个。
- $\{t_1, t_2\} = \{2, -2\}$：$P(x) = (x-1)^2(x+1)^2$，两个相异根。
- 其中一个 $t_i$ 为 $\pm 2$，另一个的绝对值大于 2：根为 $\pm 1$（二重）及另外两个根，即三个相异根。

**?** **想一想** 为甚么其中一个 $t$ 值必定是 $\pm 2$?

四个根中只有三个相异，表示有某个根是重根（repeated root）；上面的列表显示，在出现三个相异根的情况下，这只会在最后一种情况发生。说得更直接一点：根分成一对对互为倒数的根，而相异根的个数是奇数，就迫使某个根等于自己的倒数，即 $\pm 1$。

所以 $t = 2$ 或 $t = -2$ 是 $t^2 + at + b - 2 = 0$ 的根：

$$\begin{aligned} t = 2: &\quad 4 + 2a + b - 2 = 0 \iff b = -2a - 2, \\ t = -2: &\quad 4 - 2a + b - 2 = 0 \iff b = 2a - 2. \end{aligned}$$

因此 $b = 2a - 2$ 或 $b = -2a - 2$，得证。

### 第 (iii) 部分

当 $b = 2a - 2$ 时，关于 $t$ 的二次方程为

$$t^2 + at + 2a - 4 = (t + 2)(t + a - 2),$$

所以 $t_1 = -2$，$t_2 = 2 - a$。前者给出 $(x + 1)^2$，后者给出 $x^2 - (2 - a)x + 1 = 0$，所以

$$x = -1\text{（二重）}, \qquad x = \frac{2 - a \pm \sqrt{a^2 - 4a}}{2}.$$

由于题目告诉我们所有根都是实数，这些根都是实数；这相当于 $a^2 - 4a = a(a - 4) \geq 0$，即 $a \leq 0$ 或 $a \geq 4$。

### 最后部分

由 (ii)，要有三个相异根，就需要 $b = 2a - 2$ 或 $b = -2a - 2$。反过来，假设 $b = 2a - 2$（且所有根都是实数）。由 (iii)，$t$ 的值为 $-2$ 及 $2 - a$；根据上面的情况列表，除非 $2 - a = \pm 2$，否则恰好有三个相异根：

- $a = 0$，$b = -2$：$t = \pm 2$，$P(x) = (x^2 - 1)^2$，只有两个相异根；
- $a = 4$，$b = 6$：$t = -2$ 出现两次，$P(x) = (x + 1)^4$，只有一个根。

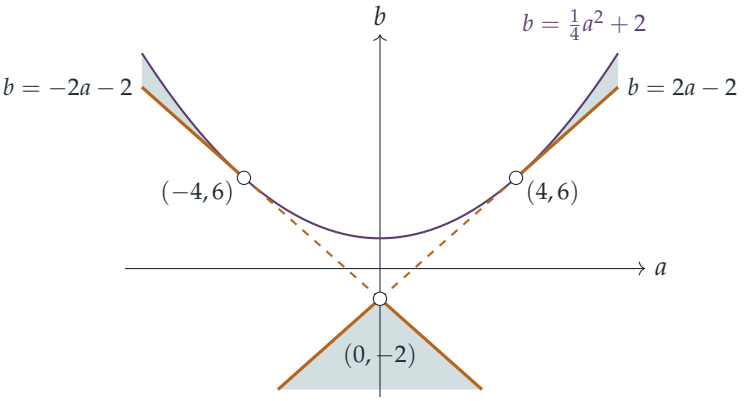
把 $x$ 换成 $-x$，$P$ 会变成同样形式、但 $a$ 换成 $-a$ 的多项式，这会把两条直线互换。所以当 $b = -2a - 2$ 时，$t$ 的值为 2 及 $-a - 2$，例外情况为 $a = 0$，$b = -2$ 及 $a = -4$，$b = 6$。

因此，对于使所有根都是实数的实数 $a$、$b$，$(*)$ 恰好有三个相异根当且仅当

$$b = 2a - 2 \text{ 或 } b = -2a - 2, \qquad \text{且} (a, b) \neq (0, -2),\ (4, 6),\ (-4, 6).$$

注意 $(4, -10)$ 和 $(-4, -10)$ 是容许的：例如 $a = 4$，$b = -10$ 给出 $t = 2$ 及 $t = -6$，所以有三个相异根。

下图显示 $(a, b)$ 平面。阴影区域是所有根都是实数的地方：在两条直线之下（一个 $t$ 值最多为 $-2$，另一个至少为 2），或在 $|a| \geq 4$ 时位于直线与抛物线 $b = \frac{1}{4}a^2 + 2$ 之间（两个 $t$ 值在同一侧）。抛物线是 $t_1 = t_2$ 的地方，它与两条直线相切于 $(\pm 4, 6)$。答案就是两条直线在阴影区域边界上的实线部分，再除去三个空心圆点。




（直线的虚线部分，即 $0 < |a| < 4$，位于所有根都是实数的区域之外，这就是 (iii) 需要 $a \leq 0$ 或 $a \geq 4$ 的原因。）

## 验算

- 取 $a = -4$，$b = -10$（在 $b = 2a - 2$ 上，且容许）。由 (iii) 得 $x = -1, -1, 3 \pm 2\sqrt{2}$，而确实 $(x+1)^2(x^2 - 6x + 1) = x^4 - 4x^3 - 10x^2 - 4x + 1$。根 $3 \pm 2\sqrt{2} \approx 5.828,\ 0.172$ 互为倒数，正如预期。
- 在被排除的点 $a = 0$，$b = -2$，四次式为 $x^4 - 2x^2 + 1 = (x^2 - 1)^2$，只有两个相异根。
- 我们用电脑对步长为 $\frac{1}{4}$、$|a| \leq 10$、$|b| \leq 15$ 的网格上每一点 $(a, b)$ 以数值方法求根；在 1579 个所有根都是实数的网格点中，恰好是上述条件所预测的 88 个点给出三个相异根。

## 常见错误

- 忘记例外的点：单凭 $b = 2a - 2$ 是必要但不充分的，因为 $(0, -2)$ 和 $(4, 6)$ 都在这条直线上。
- 假设重根必定是 $\pm 1$：二重的一对根 $k, k, \frac{1}{k}, \frac{1}{k}$ 也是可能的，但这只给出两个相异根，而不是三个。

## STEP 之外

系数正读反读都一样的多项式称为*回文多项式*（palindromic，又称倒数多项式，reciprocal），而 $t = x + \frac{1}{x}$ 这个技巧可以把任何偶数次的回文多项式的次数减半。同样的代换配合 $x = e^{i\theta}$，会得到 $t = 2\cos\theta$；正是用这个方法可以求出 $2\cos(2\pi/n)$ 的极小多项

式（minimal polynomial），高斯（Gauss）处理正 17 边形时也是用这个方法。描述多项式的根在何处重合（正如上图 $(a, b)$ 平面中的曲线所做的）这个想法，引出了判别式（discriminant）。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）是进一步阅读多项式及其根的好起点。

详细解答

# STEP II 2018，第 2 题

纯数学 · 凹函数

**STEP II 2018，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/18-s2-q2

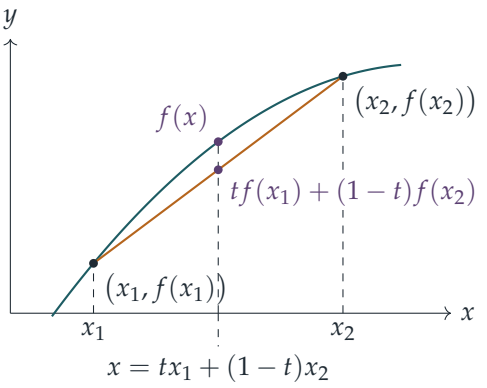
## 关键思路

当 $t$ 由 0 变到 1 时，点 $tx_1 + (1-t)x_2$ 由 $x_2$ 移到 $x_1$，而 $tf(x_1) + (1-t)f(x_2)$ 就是弦（chord）在该点上方的高度。所以凹性（concavity）的意思是：每一条弦都位于曲线之上或之下——准确来说，在曲线上或曲线下方。掌握了这条两点不等式之后，只要用两次就得到三点的版本；再把它应用于适当的凹函数（concave function），并利用 $A + B + C = \pi$，两条三角形不等式便随之而来。

## 解答

### 示意图

设 $x = tx_1 + (1-t)x_2$。那么点 $\big(x,\ tf(x_1) + (1-t)f(x_2)\big)$ 把由 $\big(x_2, f(x_2)\big)$ 到 $\big(x_1, f(x_1)\big)$ 的弦按 $(1-t) : t$ 的比分割，所以它位于弦上、正好在 $x$ 的正上方（或正下方）。定义所说的是：曲线在 $x$ 处的高度至少与弦一样高。




### 为何 $f'' < 0$ 推出凹性

若 $x_1 = x_2$，或 $t = 0$ 或 $t = 1$，不等式取等号，所以可设 $0 < t < 1$，并不失一般性（without loss of generality）设 $x_1 < x_2$（对调 $x_1, x_2$ 只不过是对调 $t$ 与 $1 - t$）。于是 $x = tx_1 + (1-t)x_2$ 满足 $x_1 < x < x_2$，且

$$x - x_1 = (1-t)(x_2 - x_1), \qquad x_2 - x = t(x_2 - x_1).$$

由中值定理（mean value theorem），存在 $c_1 \in (x_1, x)$ 及 $c_2 \in (x, x_2)$，使得

$$\frac{f(x) - f(x_1)}{x - x_1} = f'(c_1), \qquad \frac{f(x_2) - f(x)}{x_2 - x} = f'(c_2).$$

由于在 $(a, b)$ 上 $f'' < 0$，$f'$ 在该区间上严格递减（strictly decreasing），而 $c_1 < c_2$，所以 $f'(c_1) > f'(c_2)$。代入上面的长度，并约去正因子 $x_2 - x_1$，得

$$\frac{f(x) - f(x_1)}{1 - t} > \frac{f(x_2) - f(x)}{t}.$$

两边乘以 $t(1-t) > 0$，得 $tf(x) - tf(x_1) > (1-t)f(x_2) - (1-t)f(x)$，即

$$f(x) > tf(x_1) + (1-t)f(x_2).$$

所以 $f$ 在 $(a, b)$ 上是凹函数。用文字来说：曲线的斜率不断下降，所以由 $x_1$ 到 $x$ 的斜率大于由 $x$ 到 $x_2$ 的斜率，曲线便向上鼓起，高于弦。

### 第 (i) 部分

**想一想** 哪两个点、用甚么权重，加权平均后会得到 $\frac{1}{3}(u + v + w)$?

把 $v$ 和 $w$ 归为一组：取 $x_1 = u$、$x_2 = \frac{1}{2}(v + w)$ 及 $t = \frac{1}{3}$。由于 $x_2$ 介于 $v$ 与 $w$ 之间，它位于 $(a, b)$ 内，而且

$$tx_1 + (1-t)x_2 = \tfrac{1}{3}u + \tfrac{2}{3} \cdot \tfrac{1}{2}(v + w) = \tfrac{1}{3}(u + v + w).$$

由定义得

$$f\Big(\frac{u+v+w}{3}\Big) \geq \tfrac{1}{3}f(u) + \tfrac{2}{3}f\Big(\frac{v+w}{2}\Big).$$

现在再用一次定义，取 $x_1 = v$、$x_2 = w$、$t = \frac{1}{2}$：$f\big(\frac{1}{2}(v+w)\big) \geq \frac{1}{2}f(v) + \frac{1}{2}f(w)$。由于这一项的系数 $\frac{2}{3}$ 是正数，代入后不等号的方向不变：

$$f\Big(\frac{u+v+w}{3}\Big) \geq \tfrac{1}{3}f(u) + \tfrac{2}{3}\cdot\tfrac{1}{2}\big(f(v)+f(w)\big) = \frac{f(u)+f(v)+f(w)}{3},$$

这就是所求的结果。

### 第 (ii) 部分

在 $0 < x < \pi$ 上取 $f(x) = \sin x$。此时 $f''(x) = -\sin x < 0$，所以 sin 在 $(0,\pi)$ 上是凹函数。三角形的内角满足 $0 < A, B, C < \pi$ 及 $A + B + C = \pi$，因此在第 (i) 部分中取 $u = A$、$v = B$、$w = C$，得

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

所以 $\sin A + \sin B + \sin C \leq \dfrac{3\sqrt{3}}{2}$。

### 第 (iii) 部分

当 $0 < x < \pi$ 时 $\sin x > 0$，所以 $g(x) = \ln(\sin x)$ 在该区间上有定义，且

$$g'(x) = \frac{\cos x}{\sin x} = \cot x, \qquad g''(x) = -\operatorname{cosec}^2 x < 0.$$

所以 $g$ 在 $(0,\pi)$ 上是凹函数，由第 (i) 部分得

$$\frac{\ln\sin A + \ln\sin B + \ln\sin C}{3} \leq \ln\sin\frac{\pi}{3} = \ln\frac{\sqrt{3}}{2}.$$

两边乘以 3 并合并对数，得 $\ln(\sin A \sin B \sin C) \leq \ln\big(\frac{\sqrt{3}}{2}\big)^3$。由于 exp 是递增函数，

$$\sin A \times \sin B \times \sin C \leq \Big(\frac{\sqrt{3}}{2}\Big)^3 = \frac{3\sqrt{3}}{8}.$$

## 验算

两个上界都由等边三角形（equilateral triangle）取到：$3\sin\frac{\pi}{3} = \frac{3\sqrt{3}}{2} \approx 2.598$，以及 $\sin^3\frac{\pi}{3} = \frac{3\sqrt{3}}{8} \approx 0.6495$。等腰直角三角形给出的和是 $1+\sqrt{2} \approx 2.414$，积是 $\frac{1}{2}$，两者都低于上界。用电脑检验 200 000 个随机三角形，没有一个的和超过 2.598076，也没有一个的积超过 0.649519。非常扁平的三角形（$A, B \to 0$，$C \to \pi$）会令两者都趋向 0，正如预期。

## 常见错误

- 在第 (i) 部分，若取 $t = \frac{1}{3}$、$x_1 = u$、$x_2 = v$，得到的是 $\frac{1}{3}u + \frac{2}{3}v$，并非三个数的平均值；第二个点必须是 $v$ 与 $w$ 的中点。
- 在第 (iii) 部分，不能直接把第 (ii) 部分的不等式套用到乘积上；要靠对数把乘积化为和。另外也要检查 $\ln \sin x$ 有定义，这需要 $\sin x > 0$。
- 只说 $f'' < 0$“意味着曲线向下弯”，本身并不构成弦不等式的证明；中值定理的论证才令它变得严谨。

## STEP 之外

第 (i) 部分是琴生不等式（Jensen's inequality）的三点情况：对凹函数 $f$ 及总和为 1 的权重 $t_i \geq 0$，有 $f\left(\sum t_i x_i\right) \geq \sum t_i f(x_i)$，其证明正是用归纳法（induction），做法与这里完全一样。取 $f = \ln$，便得到算术－几何平均不等式（AM–GM inequality）；在概率论中，它说的是对凹函数 $f$，有 $f(\mathrm{E}[X]) \geq \mathrm{E}[f(X)]$。凸函数（convex function）与凹函数，以及第二段背后的中值定理，在 Bartle 与 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）中有详细而严谨的讨论。

详细解答

# STEP II 2018，第 3 题

纯数学 · 旋转对称与积分

**STEP II 2018，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/18-s2-q3

## 关键思路

绕点 $(a,b)$ 作半周旋转（half-turn），会把 $(x,y)$ 映到 $(2a-x,\,2b-y)$。因此，具有这种对称性的曲线会把每个 $x$ 值与其反射值 $2a-x$ 配对，并使两个对应的高度之和为 $2b$。对正切函数而言，最自然的配对是 $x \leftrightarrow \frac{1}{2}\pi - x$，因为 $\tan(\frac{1}{2}\pi - x) = 1/\tan x$。单凭这一条恒等式，就能得出曲线的对称性；之后，对于关于中心对称的区间上的积分，根本不用真正积分也能求出结果。

## 解答

### 第 (i) 部分

设 $f(x) = (1+\tan x)^{-1}$，利用链式法则（chain rule），

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

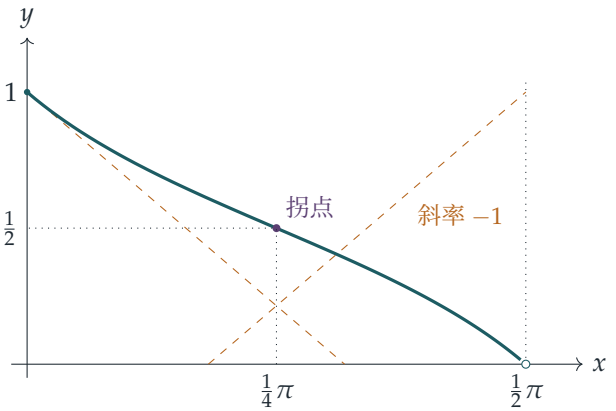
即为所求。（把 $1+\tan x$ 乘以 $\cos x$，正是令整个式子变得简洁的一步。）

至于值域（range），由 $0 \le x < \frac{1}{2}\pi$ 得 $0 \le 2x < \pi$，所以 $\sin 2x$ 取遍 $[0,1]$ 内的每个值：在 $x=0$ 时取值 0，在 $x=\frac{1}{4}\pi$ 时取值 1。因此 $1 \le 1+\sin 2x \le 2$。取倒数（由于各项皆为正数，不等号方向会反转），再取负号，得

$$-1 \le f'(x) \le -\tfrac{1}{2}.$$

两个端点都能取到：$f'(0) = -1$ 及 $f'(\frac{1}{4}\pi) = -\frac{1}{2}$。

关于草图：$f(0)=1$，$f(\frac{1}{4}\pi)=\frac{1}{2}$，而由于 $\tan x\to\infty$，当 $x\to\frac{1}{2}\pi$ 时 $f(x)\to 0$。曲线处处递减，斜率介乎 $-1$ 与 $-\frac{1}{2}$ 之间。曲线在 $x=0$ 处最陡（斜率 $-1$），在 $x=\frac{1}{4}\pi$ 处最平缓，而当 $x\to\frac{1}{2}\pi$ 时又再变陡，斜率趋向 $-1$。所以 $f'$ 在 $\frac{1}{4}\pi$ 处取得极大值，该点是一个拐点（point of inflection）：曲线在 $[0,\frac{1}{4}\pi)$ 上向上凹（concave up），在 $(\frac{1}{4}\pi,\frac{1}{2}\pi)$ 上向下凹（concave down）。（事实上 $f''(x)=2\cos 2x/(1+\sin 2x)^2$，它在 $\frac{1}{4}\pi$ 处变号。）




两条虚线的斜率均为 $-1$：一条是曲线在 $(0,1)$ 处的切线，另一条是当 $x\to\frac{1}{2}\pi$ 时曲线趋近的方向。曲线看来在绕 $(\frac{1}{4}\pi,\frac{1}{2})$ 作半周旋转下是对称的，而这正是本题余下部分的主题。

**第 (ii) 部分**

绕 $(a,b)$ 旋转 $\pi$ 会把 $(x,y)$ 映到 $(2a-x,\,2b-y)$，因为 $(a,b)$ 是一点与其像的中点。曲线关于 $(a,b)$ 具有二阶旋转对称（rotational symmetry of order 2），当且仅当这个旋转把曲线映成它自己。（要做到这一点，定义域（domain）必须关于 $x=a$ 对称，使得只要 $x$ 在定义域内，$2a-x$ 亦在定义域内。）

（$\Rightarrow$）假设曲线具有这种对称性。对定义域内任意 $x$，点 $(x,g(x))$ 在曲线上，所以其像 $(2a-x,\,2b-g(x))$ 亦在曲线上。由于曲线是函数的图像，曲线上 $x$ 坐标为 $2a-x$ 的点只有 $(2a-x,\,g(2a-x))$。所以 $g(2a-x)=2b-g(x)$，即

$$g(x)+g(2a-x)=2b.$$

（$\Leftarrow$）假设对定义域内每个 $x$ 都有 $g(x)+g(2a-x)=2b$。曲线上一般点 $(x,g(x))$ 的像是

$$(2a-x,\,2b-g(x))=(2a-x,\,g(2a-x)),$$

它在曲线上。所以这个旋转把曲线映入其自身；又由于旋转两次即得恒等变换，它把曲线映成其自身。在恒等式中代入 $x=a$，得 $g(a)=b$，与 $(a,b)$ 在曲线上一致。

至于积分, 关于原点对称即 $a=b=0$, 所以 $g(-x)=-g(x)$: $g$ 是奇函数 (odd function)。在左半部分作代换（substitution）$x=-u$,

$$\int_{-1}^{0} g(x)\,\mathrm{d}x = \int_{0}^{1} g(-u)\,\mathrm{d}u = -\int_{0}^{1} g(u)\,\mathrm{d}u,$$

所以两半互相抵消，得

$$\int_{-1}^{1} g(x)\,\mathrm{d}x = 0.$$

**第 (iii) 部分**

设 $h(x)=\dfrac{1}{1+\tan^k x}$，其中 $0<x<\frac{1}{2}\pi$。参考 (i) 的草图，尝试取 $a=\frac{1}{4}\pi$、$b=\frac{1}{2}$。定义域 $(0,\frac{1}{2}\pi)$ 关于 $\frac{1}{4}\pi$ 对称，且 $2a-x=\frac{1}{2}\pi-x$。

**想一想** 怎样化简 $h(\frac{1}{2}\pi-x)$?

利用 $\tan(\frac{1}{2}\pi-x)=\cot x=1/\tan x$，此式在 $0<x<\frac{1}{2}\pi$ 时成立，而此时 $\tan x>0$。于是，分子分母同乘以 $\tan^k x$,

$$h(\tfrac{1}{2}\pi-x)=\frac{1}{1+\tan^{-k}x}=\frac{\tan^k x}{\tan^k x+1},$$

所以

$$h(x)+h(\tfrac{1}{2}\pi-x)=\frac{1+\tan^k x}{1+\tan^k x}=1=2\cdot\tfrac{1}{2}.$$

由于 $\tan\frac{1}{4}\pi=1$ 给出 $h(\frac{1}{4}\pi)=\frac{1}{2}$，点 $(\frac{1}{4}\pi,\frac{1}{2})$ 在曲线上。由 (ii)，无论正常数 $k$ 取何值，曲线都关于 $(\frac{1}{4}\pi,\frac{1}{2})$ 具有二阶旋转对称。

至于积分，设 $I=\displaystyle\int_{\pi/6}^{\pi/3} h(x)\,\mathrm{d}x$。区间 $[\frac{1}{6}\pi,\frac{1}{3}\pi]$ 关于 $\frac{1}{4}\pi$ 对称，所以代换 $x=\frac{1}{2}\pi-u$ 把它映成其自身（两端互换，正好抵消 $\mathrm{d}x=-\mathrm{d}u$ 带来的负号）:

$$\begin{aligned} I &= \int_{\pi/6}^{\pi/3} h(\tfrac{1}{2}\pi-u)\,\mathrm{d}u = \int_{\pi/6}^{\pi/3}\big(1-h(u)\big)\,\mathrm{d}u \\ &= \left(\tfrac{1}{3}\pi-\tfrac{1}{6}\pi\right)-I. \end{aligned}$$

因此 $2I=\frac{1}{6}\pi$，而

$$\int_{\pi/6}^{\pi/3}\frac{1}{1+\tan^k x}\,\mathrm{d}x=\frac{\pi}{12}.$$

从几何角度看，在矩形 $[\frac{1}{6}\pi,\frac{1}{3}\pi]\times[0,1]$ 内，半周旋转把曲线下方的区域与曲线上方的区域互换，所以面积是矩形面积的一半，即 $\frac{1}{2}\cdot\frac{1}{6}\pi$。这正是把 (ii) 的论证中的 $g$ 换成 $h-\frac{1}{2}$。

## 验算

- 取 $k = 1$ 及 $x = \frac{1}{6}\pi$：$h = 1/(1 + \frac{1}{\sqrt{3}}) = \frac{1}{2}(3 - \sqrt{3}) \approx 0.634$；取 $x = \frac{1}{3}\pi$：$h = 1/(1 + \sqrt{3}) = \frac{1}{2}(\sqrt{3} - 1) \approx 0.366$。两者之和为 1，与对称性所言相符。
- 答案与 $k$ 无关。当 $k \to \infty$ 时，$h$ 在 $x < \frac{1}{4}\pi$ 时接近 1，在 $x > \frac{1}{4}\pi$ 时接近 0，所以积分趋向区间 $[\frac{1}{6}\pi, \frac{1}{4}\pi]$ 的长度，即 $\frac{1}{12}\pi$，结果一致。
- 对 $k = 0.3$、1、2、3.7 及 10 作数值积分，每次都得 0.2617994，与 $\pi/12 \approx 0.2617994$ 吻合。

## 常见错误

- 在 (i) 中把值域写成 $-1 < f'(x) \leq -\frac{1}{2}$：值 $-1$ 在 $x = 0$ 处取到，而 $x = 0$ 在定义域内。
- 在 (ii) 中只证明一个方向。“当且仅当”需要同时证明对称性 $\Rightarrow$ 恒等式，以及恒等式 $\Rightarrow$ 对称性。
- 在 (iii) 中尝试直接积分 $1/(1 + \tan^k x)$。对一般的 $k$，根本找不到初等原函数（elementary antiderivative），而对称性才是整道题的关键。

## STEP 之外

(iii) 中的技巧,即作代换 $x \mapsto a + b - x$ 后把积分与自身相加,常被称为“King 法则”(King's rule)。它是利用定义域的对称性来简化积分的一个特例；这个想法在群论（group theory）和物理学中占有核心地位，因为问题的对称性可以预言哪些积分为零（如 (ii)），或取简单的值。同样的推理可求得对每个 $k > 0$ 都有 $\int_0^{\pi/2} \frac{\mathrm{d}x}{1+\tan^k x} = \frac{1}{4}\pi$。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶问题》）收录了更多可以用这类对称论证解决的积分。

详细解答

# STEP II 2018，第 4 题

纯数学 · 三角方程

**STEP II 2018，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/18-s2-q4

## 关键思路

题目给出的恒等式（identity）把余弦之和化为乘积，而解一条等于零的方程时，我们正需要乘积。在 (i) 中，把头尾两项配对、中间两项配对，立即得到一个公因子（common factor）。在 (ii) 和 (iii) 中，有用的做法是把一切以两角的半和与半差表示；这样一来，(ii) 可以因式分解（factorise），而 (iii) 则变成两个平方之和等于零。

## 解答

### 第 (i) 部分

把第一项与最后一项配对，中间两项配对。由恒等式，

$$\cos x + \cos 4x = 2\cos\tfrac{5x}{2}\cos\tfrac{3x}{2},$$
$$3(\cos 2x + \cos 3x) = 6\cos\tfrac{5x}{2}\cos\tfrac{x}{2}.$$

（由于余弦是偶函数（even function），我们可写 $\cos(-\frac{3x}{2}) = \cos\frac{3x}{2}$。）因此左边为

$$2\cos\tfrac{5x}{2}\left(\cos\tfrac{3x}{2} + 3\cos\tfrac{x}{2}\right).$$

设 $c = \cos\frac{x}{2}$，由三倍角公式（triple-angle formula）$\cos 3\theta = 4\cos^3\theta - 3\cos\theta$ 得 $\cos\frac{3x}{2} = 4c^3 - 3c$，所以括号内正好是 $4c^3$。因此方程为

$$8\cos\tfrac{5x}{2}\cos^3\tfrac{x}{2} = 0.$$

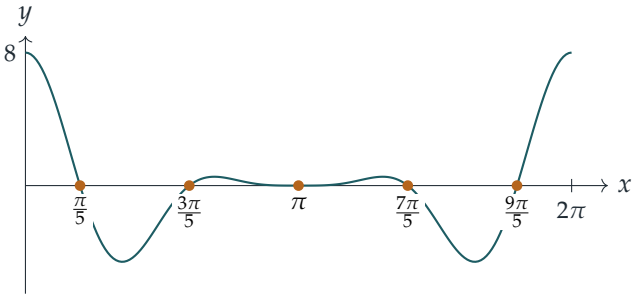
第一个因子。$\cos\frac{5x}{2} = 0$ 即 $\frac{5x}{2} = \frac{\pi}{2} + k\pi$，其中 $k$ 为整数，亦即 $x = \frac{(2k+1)\pi}{5}$。要 $0 \le x \le 2\pi$，须 $0 \le 2k+1 \le 10$，所以 $k = 0, 1, 2, 3, 4$。

第二个因子。$\cos\frac{x}{2}=0$ 且 $0\le\frac{x}{2}\le\pi$，只得 $\frac{x}{2}=\frac{\pi}{2}$，即 $x=\pi$，而这已在上面的列表中。

因此解为

$$x=\frac{\pi}{5},\ \frac{3\pi}{5},\ \pi,\ \frac{7\pi}{5},\ \frac{9\pi}{5}.$$

左边的图像显示了这些解；在 $x=\pi$ 处，$\cos\frac{5x}{2}$ 和 $\cos^3\frac{x}{2}$ 同时为零，所以曲线在该处只是触及横轴，而不是穿过它。




**第 (ii) 部分**

对头两项使用恒等式，并利用 $\cos 2x=2\cos^2 x-1$，

$$\cos(x+y)+\cos(x-y)-\cos 2x=2\cos x\cos y-2\cos^2 x+1.$$

令它等于 1，得

$$2\cos x\,(\cos y-\cos x)=0,$$

所以 $\cos x=0$ 或 $\cos y=\cos x$。

- 若 $\cos x=0$ 且 $0\le x\le\pi$，则 $x=\frac{\pi}{2}$（此时任何 $y$ 皆可）。
- 若 $\cos y=\cos x$，则 $x=y$，因为余弦在 $[0,\pi]$ 上严格递减（strictly decreasing），所以在该区间内每个值只取一次。

所以 $x=y$ 或 $x=\frac{\pi}{2}$，得证。

**第 (iii) 部分**

设 $s=\frac{1}{2}(x+y)$ 及 $d=\frac{1}{2}(x-y)$。由恒等式得 $\cos x+\cos y=2\cos s\cos d$，而 $\cos(x+y)=\cos 2s=2\cos^2 s-1$。方程变为

$$2\cos s\cos d-2\cos^2 s+1=\tfrac{3}{2}\iff 4\cos^2 s-4\cos s\cos d+1=0.$$

? **想一想** 一条方程，两个未知数：它怎能同时确定 $x$ 和 $y$?

对 $\cos s$ 配方（complete the square）。由于 $1 = \cos^2 d + \sin^2 d$，

$$4\cos^2 s - 4\cos s\cos d + \cos^2 d + \sin^2 d = (2\cos s - \cos d)^2 + \sin^2 d = 0.$$

两个实数的平方之和为零，当且仅当两者皆为零，于是得到两条方程：

$$\sin d = 0 \qquad 及 \qquad 2\cos s = \cos d.$$

由于 $x, y \in [0, \pi]$，有 $-\frac{\pi}{2} \le d \le \frac{\pi}{2}$，而 sin 在此区间内唯一的零点是 $d = 0$。所以 $x = y$，于是 $s = x$ 且 $\cos d = 1$，得 $\cos x = \frac{1}{2}$。由 $0 \le x \le \pi$，即 $x = \frac{\pi}{3}$。

所以唯一的解是

$$x = y = \frac{\pi}{3}.$$

确实，$\cos\frac{\pi}{3} + \cos\frac{\pi}{3} - \cos\frac{2\pi}{3} = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2}$。

同样的代数运算还说明了更多：对所有 $x$、$y$，

$$\tfrac{3}{2} - \big(\cos x + \cos y - \cos(x+y)\big) = \tfrac{1}{2}\big[(2\cos s - \cos d)^2 + \sin^2 d\big] \ge 0,$$

所以 $\frac{3}{2}$ 是左边所能取的最大值，而这条方程问的正是此最大值在何处取得。

## 验算

- 在 (i) 中，当 $x = \frac{\pi}{5}$：由于 $\cos(\pi - \theta) = -\cos\theta$，有 $\cos\frac{\pi}{5} + \cos\frac{4\pi}{5} = 0$ 及 $\cos\frac{2\pi}{5} + \cos\frac{3\pi}{5} = 0$。当 $x = 0$，左边为 $1 + 3 + 3 + 1 = 8$，与 $8\cos 0\cos^3 0$ 吻合。
- 用电脑在 $[0, 2\pi]$ 上以细密网格搜索，恰好找到上述五个零点；在 $[0, \pi]^2$ 上以网格搜索，找到 $\cos x + \cos y - \cos(x+y)$ 的最大值为 1.5，且只在 $x = y \approx 1.0472 = \frac{\pi}{3}$ 处取得。
- 在 (ii) 中，$x = \frac{\pi}{2}$ 对每个 $y$ 都给出 $\cos(\frac{\pi}{2} + y) + \cos(\frac{\pi}{2} - y) - \cos\pi = -\sin y + \sin y + 1 = 1$。

## 常见错误

- 把 $x = \pi$ 列了两次，或者漏掉了它：在 (i) 中它来自两个因子，但只算一个解。
- 在 (ii) 中，除以 $\cos x$，因而失去 $x = \frac{\pi}{2}$。
- 在 (iii) 中，停在关于 $\cos s$ 的二次方程，而没有利用判别式（discriminant）（或平方和）迫使 $\cos^2 d = 1$。

## STEP 之外

(i) 中的系数 $1,3,3,1$ 并非巧合：左边是 $e^{ix}(1+e^{ix})^3$ 的实部（real part），而由于 $1+e^{ix} = 2\cos\frac{x}{2}\,e^{ix/2}$，一行便得到 $8\cos^3\frac{x}{2}\cos\frac{5x}{2}$。第 (iii) 部分其实是经典不等式（inequality）$\cos A+\cos B+\cos C \le \frac{3}{2}$ 的变装版本，其中 $A$、$B$、$C$ 为三角形的三个角（令 $A=x$、$B=y$、$C=\pi-x-y$），等号只在等边三角形（equilateral triangle）时成立；像这样求二元函数的最大值，是大学一年级多变量微积分（multivariable calculus）的标准课题。Needham 的 *Visual Complex Analysis*（《可视化复分析》）展示了复指数如何令这类三角恒等式几乎变得自动成立。

详细解答

# STEP II 2018，第 5 题

纯数学 · 级数的积分

**STEP II 2018，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/18-s2-q5

## 关键思路

每一部分都是先把函数化成幂级数（power series），再逐项（term by term）积分。在 (i) 中，这只不过是对一个几何级数（geometric series）积分。在 (ii) 中，把指数函数展开后，每一项都变成形如 $\int_0^\infty x^m \mathrm{e}^{-x}\,\mathrm{d}x$ 的积分，而这个积分是一个阶乘（factorial）；阶乘互相约去后，剩下的正好是 (i) 中的级数。至于 (iii)，分母中的 $\ln x$ 就是线索：代换（substitution）$x = \mathrm{e}^{-t}$ 会把它变成 $-t$，并把整个积分变成两个 (ii) 中的积分。

## 解答

### 第 (i) 部分

当 $|x| < 1$ 时，二项式展开（binomial expansion）（即公比为 $-x$ 的几何级数）为

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

将两边由 0 积分至 $x$。左边，由于 $\ln 1 = 0$，所以 $\int_0^x \frac{\mathrm{d}u}{1+u} = \ln(1+x)$。右边逐项积分，得

$$\begin{aligned}\ln(1+x) &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} \qquad (\text{将} n+1 \text{ 换成} n) \\ &= -\sum_{n=1}^{\infty} \frac{(-1)^n x^n}{n} = -\sum_{n=1}^{\infty} \frac{(-x)^n}{n},\end{aligned}$$

这正是 $|x| < 1$ 时所要求的结果。

**第 (ii) 部分**

对所有 $x$，

$$\mathrm{e}^{-ax} = \sum_{n=0}^{\infty} \frac{(-ax)^n}{n!} = 1 - ax + \frac{a^2x^2}{2!} - \frac{a^3x^3}{3!} + \cdots .$$

$n = 0$ 的一项是 1，所以

$$\frac{1 - \mathrm{e}^{-ax}}{x} = -\sum_{n=1}^{\infty} \frac{(-a)^n x^{n-1}}{n!}.$$

在这里除以 $x$ 并无问题：余下的每一项都含有因子 $x$，所以被积函数（integrand）在 $x = 0$ 处完全正常（它趋向 $a$）。乘以 $\mathrm{e}^{-x}$ 再逐项积分（题目说可以不理会收敛（convergence）问题），得

$$\int_0^{\infty} \frac{(1 - \mathrm{e}^{-ax})\,\mathrm{e}^{-x}}{x}\,\mathrm{d}x = -\sum_{n=1}^{\infty} \frac{(-a)^n}{n!} I_{n-1}, \qquad \text{其中} I_m = \int_0^{\infty} x^m \mathrm{e}^{-x}\,\mathrm{d}x.$$

**想一想** $I_m$ 是甚么？

取 $u = x^m$ 及 $\mathrm{d}v = \mathrm{e}^{-x}\,\mathrm{d}x$ 进行分部积分（integration by parts）。当 $m \geq 1$ 时，

$$I_m = \Big[ -x^m \mathrm{e}^{-x} \Big]_0^{\infty} + m \int_0^{\infty} x^{m-1} \mathrm{e}^{-x}\,\mathrm{d}x = m I_{m-1},$$

因为当 $x \to \infty$ 时 $x^m \mathrm{e}^{-x} \to 0$（指数函数占上风），而在 $x = 0$ 处它等于 0。由于 $I_0 = \int_0^{\infty} \mathrm{e}^{-x}\,\mathrm{d}x = 1$，可得 $I_m = m!$。

所以 $I_{n-1} = (n-1)!$，而 $\dfrac{(n-1)!}{n!} = \dfrac{1}{n}$。因此

$$\int_0^{\infty} \frac{(1 - \mathrm{e}^{-ax})\,\mathrm{e}^{-x}}{x}\,\mathrm{d}x = -\sum_{n=1}^{\infty} \frac{(-a)^n}{n} = \ln(1 + a),$$

这里用了第 (i) 部分并取 $x = a$；由于 $|a| < 1$，这是容许的。这正是所要求的结果。

**第 (iii) 部分**

代入 $x = \mathrm{e}^{-t}$，使得 $\ln x = -t$ 及 $\mathrm{d}x = -\mathrm{e}^{-t}\,\mathrm{d}t$。当 $x$ 由 0 变到 1 时，$t$ 由 $\infty$ 减至 0。此外 $x^p = \mathrm{e}^{-pt}$ 及 $x^q = \mathrm{e}^{-qt}$。因此

$$\int_0^1 \frac{x^p - x^q}{\ln x}\,\mathrm{d}x = \int_{\infty}^0 \frac{\mathrm{e}^{-pt} - \mathrm{e}^{-qt}}{-t}\left(-\mathrm{e}^{-t}\right)\mathrm{d}t = \int_0^{\infty} \frac{\left(\mathrm{e}^{-qt} - \mathrm{e}^{-pt}\right)\mathrm{e}^{-t}}{t}\,\mathrm{d}t,$$

其中两个负号互相抵消，而对调上下限则令 $p$ 和 $q$ 互换。现在写成 $\mathrm{e}^{-qt} - \mathrm{e}^{-pt} = \left(1 - \mathrm{e}^{-pt}\right) - \left(1 - \mathrm{e}^{-qt}\right)$，并把积分拆开。每一部分都是 (ii) 中的形式，分别取 $a = p$ 和 $a = q$，而两者都满足 $|a| < 1$。所以

$$\int_0^1 \frac{x^p - x^q}{\ln x}\,\mathrm{d}x = \ln(1 + p) - \ln(1 + q) = \ln\left(\frac{1 + p}{1 + q}\right).$$

把积分拆开是合理的，因为两个积分各自都收敛：由 (ii)，每一个都有有限的值。

## 验算

- 若 $p=q$，被积函数为零，而 $\ln\frac{1+p}{1+p}=0$ 亦确实如此。
- 互换 $p$ 和 $q$，被积函数和答案都会变号。
- 符号检查：当 $0<x<1$ 时，$\ln x<0$；若 $p<q$，则 $x^p>x^q$，所以被积函数为负，而 $\ln\frac{1+p}{1+q}$ 亦为负。
- 取 $p=\frac{1}{2}$、$q=0$，答案为 $\ln\frac{3}{2}\approx 0.405465$。以数值方法计算 $\int_0^1 \frac{\sqrt{x}-1}{\ln x}\,\mathrm{d}x$ 得到相同的值；随机在 $(-1,1)$ 中选取 $p$ 和 $q$，结果与公式吻合至小数点后六位。

## 常见错误

- 忘记 $\mathrm{e}^{-ax}$ 中 $n=0$ 的一项会与 1 抵消，结果试图积分 $\mathrm{e}^{-x}/x$，而这个积分在 0 处发散。
- 在 (iii) 的代换中弄错符号：要追踪三个负号（分别来自 $\ln x$、$\mathrm{d}x$ 及对调上下限），弄错其中一个就会得到 $\ln\frac{1+q}{1+p}$。
- 写成 $I_{n-1}=n!$ 而不是 $(n-1)!$；这样阶乘就不再约成 $\frac{1}{n}$。

## STEP 之外

积分 $I_m=\int_0^\infty x^m\mathrm{e}^{-x}\,\mathrm{d}x=m!$ 就是伽玛函数（gamma function）$\Gamma(m+1)$，它把阶乘推广到非整数。(iii) 中的积分是“积分号下求导”（differentiating under the integral sign）的经典例子：对 $p$ 求导可以消去麻烦的 $\ln x$，只剩下 $\int_0^1 x^p\,\mathrm{d}x=\frac{1}{p+1}$。事实上，只要 $p,q>-1$，结果都成立。题目叫我们不必理会的“交换求和与积分次序”，在数学分析课程中是以一致收敛（uniform convergence）或控制收敛（dominated convergence）来证明其合理性的。如想了解这些步骤背后的分析，可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

# STEP II 2018，第 6 题

纯数学 · 阶乘与质数

**STEP II 2018，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/18-s2-q6

## 关键思路

两部分都取决于同一个事实：质数（prime）$p$ 整除（divides）$k!$ 当且仅当 $p \le k$。在 (i) 中，一旦 $n \ge 5$，阶乘（factorial）便是 5 的倍数，所以 $n! + 5$ 也是。在 (ii) 中，左方只由不超过 $2n-1$ 的数构成，因此任何大于 $2n-1$ 的质数都不能整除它。如果 $m!$ 大得足以包含这样一个质数，方程便不成立。题中的两个定理，作用正是令 $m$ 足够大，并提供那个质数。

## 解答

### 第 (i) 部分

当 $n \ge 5$ 时，5 是 $n!$ 的其中一个因数，所以 $5 \mid n!$，从而 $5 \mid n! + 5$。但 $n! + 5 > 5$，所以 $n! + 5$ 有一个严格介乎 1 与其本身之间的因数 5，因此不是质数。故此只需检查 $n \le 4$：

$$1! + 5 = 6, \qquad 2! + 5 = 7, \qquad 3! + 5 = 11, \qquad 4! + 5 = 29.$$

6 不是质数，其余三个都是。解为

$$(n, p) = (2, 7),\ (3, 11),\ (4, 29).$$

### 第 (ii) 部分

记 $P_n = 1! \times 3! \times \cdots \times (2n-1)!$。这些阶乘中的每一个因数，都是介乎 1 与 $2n-1$ 之间的整数。因此若质数 $q$ 整除 $P_n$，由于 $q$ 是质数，它必整除其中一个整数，所以 $q \le 2n-1$。这就得出关键事实：

$$\text{没有质数} q > 2n-1 \text{ 整除} P_n. \qquad (*)$$

另一方面，每个质数 $q \le m$ 都整除 $m!$。所以若 $P_n = m!$，则 $m$ 小于每一个大于 $2n-1$ 的质数。

**$n$ 较大时。** 假设 $n \ge 7$ 且 $P_n = m!$。由定理 1，$m! = P_n > (4n)!$，所以 $m > 4n$（阶乘是递增的）。由定理 2，存在质数 $q$ 使得 $2n < q < 4n$。于是 $q < 4n < m$，所以 $q \mid m! = P_n$。但 $q > 2n > 2n-1$，这与 $(*)$ 矛盾（contradiction）。因此 $n \ge 7$ 时没有解。

> **想一想** 为甚么我们不能对所有 $n$ 都直接引用这两个定理?

定理 1 只声称对 $n \ge 7$ 成立，而事实上它在 $n = 6$ 时并不成立。所以 $n = 1, \ldots, 6$ 需要逐一亲手检查。我们再次利用 $(*)$ 来简化计算。

**$n$ 较小时。**

- $n = 1$：$P_1 = 1 = 1!$，所以 $m = 1$。
- $n = 2$：$P_2 = 1 \times 6 = 6 = 3!$，所以 $m = 3$。
- $n = 3$：$P_3 = 6 \times 120 = 720 = 6!$，所以 $m = 6$。
- $n = 4$：$P_4 = 720 \times 5040$。由于 $720 = 6!$ 且 $5040 = 7 \times 8 \times 9 \times 10$，得 $P_4 = 6! \times 7 \times 8 \times 9 \times 10 = 10!$，所以 $m = 10$。
- $n = 5$：$P_5 = P_4 \times 9! = 10! \times 9! > 10!$，所以 $m \ge 11$。于是质数 11 整除 $m!$，但 $11 > 9 = 2n-1$，与 $(*)$ 矛盾。无解。
- $n = 6$：$P_6 = 10! \times 9! \times 11! > 12!$（因为 $10! \times 9! > 12$），所以 $m \ge 13$。于是质数 13 整除 $m!$，但 $13 > 11 = 2n-1$，与 $(*)$ 矛盾。无解。

在每个有解的情况中，$m$ 都是唯一的，因为正整数的阶乘从 $2!$ 起严格递增（而 $1! = 1 < 2!$）。所以解为

$$(n, m) = (1, 1),\ (2, 3),\ (3, 6),\ (4, 10).$$

## 验算

恒等式 $1!\,3!\,5!\,7! = 10!$ 很容易验证：两边都等于 3 628 800。以电脑在 $n \le 59$ 及 $m < 400$ 范围内搜索，恰好找到上述四组解，并确认 $n! + 5$ 只在 $n = 2, 3, 4$ 时是质数（搜索范围 $n \le 30$）。搜索亦确认定理 1 在 $n = 6$ 时不成立，这正是为何确实需要亲手检查 $n = 5$ 和 $n = 6$。

## 常见错误

- 在第 (ii) 部分遗漏 $n = 1$：此时乘积只有单一因数 $1!$，所以 $(1,1)$ 是一个解。
- 对 $n = 5$ 或 $n = 6$ 使用定理 1。定理只对 $n \geq 7$ 陈述，所以这两个情况需要各自的论证。
- 未检查 $q \leq m$ 便声称“$q \mid m!$”。这正是我们需要由定理 1 得出 $m > 4n$ 的原因。

## STEP 之外

定理 2 就是伯特兰公设（Bertrand's postulate）：对每个 $n \geq 1$，都存在质数 $p$ 使得 $n < p \leq 2n$。切比雪夫（Chebyshev）于 1852 年证明了它，后来艾狄胥（Erdős）利用中心二项式系数（central binomial coefficient）$\binom{2n}{n}$ 给出了一个优美的初等证明。像 $n! + 1 = m^2$ 这类方程（布罗卡问题，Brocard's problem）至今仍未解决。Davenport 的 *The Higher Arithmetic*（《高等算术》）是深入学习质数与阶乘算术的好起点。

详细解答

# STEP II 2018，第 7 题

纯数学 · 几何中的向量

**STEP II 2018，第 7 题**

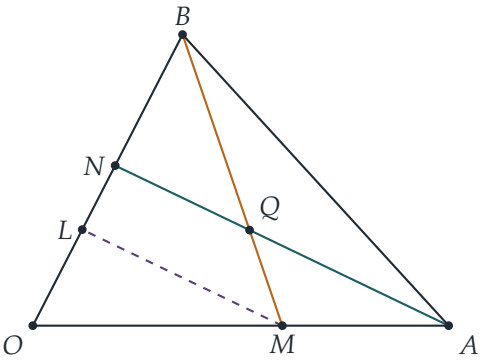
继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/18-s2-q7

## 关键思路

由于 $M$ 在 $OA$ 上，而 $N$ 在 $OB$ 上，我们可以把 $\mathbf{m}$ 写成 $\mathbf{a}$ 的倍数，把 $\mathbf{n}$ 写成 $\mathbf{b}$ 的倍数。然后点 $Q$ 可以从两条路径到达：一次沿 $BM$，一次沿 $AN$。每条路径都把 $\mathbf{q}$ 表示为 $\mathbf{a}$ 与 $\mathbf{b}$ 的组合。由于 $\mathbf{a}$ 与 $\mathbf{b}$ 不平行（parallel），我们可以比较系数（compare coefficients），所有结果便由此而得。

## 解答




**求 m**

设 $\mathbf{m} = m\,\mathbf{a}$ 及 $\mathbf{n} = n\,\mathbf{b}$，其中 $0 < m < 1$ 及 $0 < n < 1$，因为 $M$ 和 $N$ 都在三角形的边上。

沿 $BM$。$Q$ 位于 $M$ 与 $B$ 之间，且 $MQ : QB = \mu : 1$，所以 $Q$ 位于由 $M$ 到 $B$ 全程的 $\frac{\mu}{1+\mu}$

处：

$$\mathbf{q} = \mathbf{m} + \frac{\mu}{1+\mu}(\mathbf{b} - \mathbf{m}) = \frac{1}{1+\mu}\,\mathbf{m} + \frac{\mu}{1+\mu}\,\mathbf{b} = \frac{m}{1+\mu}\,\mathbf{a} + \frac{\mu}{1+\mu}\,\mathbf{b}.$$

沿 $AN$。同理，$NQ : QA = \nu : 1$，所以

$$\mathbf{q} = \mathbf{n} + \frac{\nu}{1+\nu}(\mathbf{a} - \mathbf{n}) = \frac{\nu}{1+\nu}\,\mathbf{a} + \frac{1}{1+\nu}\,\mathbf{n} = \frac{\nu}{1+\nu}\,\mathbf{a} + \frac{n}{1+\nu}\,\mathbf{b}.$$

**想一想** 为甚么我们可以分别令 **a** 的系数相等，以及 **b** 的系数相等?

$O$、$A$、$B$ 是一个真正三角形的顶点，所以 **a** 与 **b** 都非零且不平行。若 $\alpha\mathbf{a} + \beta\mathbf{b} = \mathbf{0}$ 而 $\beta \neq 0$，则 **b** 会是 **a** 的倍数；所以 $\beta = 0$，继而 $\alpha = 0$。因此平面上每个向量都只有唯一一种表示为 **a** 与 **b** 组合的方式。

把 **q** 的两个表达式相减并比较系数：

$$\frac{m}{1+\mu} = \frac{\nu}{1+\nu}, \qquad \frac{\mu}{1+\mu} = \frac{n}{1+\nu}.$$

第一式给出 $m = \dfrac{(1+\mu)\nu}{1+\nu}$，即

$$\mathbf{m} = \frac{(1+\mu)\nu}{1+\nu}\,\mathbf{a},$$

正如所求。第二式以角色互换的同一论证给出

$$\mathbf{n} = \frac{\mu(1+\nu)}{1+\mu}\,\mathbf{b}.$$

**求 $\lambda$**

现在 $\mathbf{l} = \lambda\mathbf{b}$，所以

$$\overrightarrow{ML} = \lambda\mathbf{b} - m\,\mathbf{a}, \qquad \overrightarrow{AN} = n\,\mathbf{b} - \mathbf{a}.$$

两者平行当且仅当对某个纯量（scalar）$k$ 有 $\overrightarrow{ML} = k\,\overrightarrow{AN}$。比较 **a** 的系数（理由同前）得 $k = m$，再由 **b** 的系数得 $\lambda = kn = mn$。（从几何角度看：$ML \parallel AN$ 使三角形 $OML$ 与 $OAN$ 相似（similar），所以 $OL/ON = OM/OA$。）因此

$$\lambda = mn = \frac{(1+\mu)\nu}{1+\nu} \cdot \frac{\mu(1+\nu)}{1+\mu} = \mu\nu.$$

**条件 $\mu\nu < 1$**

$$m < 1 \iff (1+\mu)\nu < 1+\nu \iff \mu\nu < 1,$$
$$n < 1 \iff \mu(1+\nu) < 1+\mu \iff \mu\nu < 1,$$

这里利用了 $1+\nu>0$ 及 $1+\mu>0$ 来消去分母。所以，在 $\mu$ 和 $\nu$ 均为正数的前提下，条件 $\mu\nu<1$ 正正就是 $M$ 严格位于 $O$ 与 $A$ 之间、$N$ 严格位于 $O$ 与 $B$ 之间所需的条件。换句话说，直线 $AN$ 与 $BM$（$Q$ 按这些比把它们分割）在三角形内部相交。这条件亦表示 $\lambda<1$：经过 $M$ 而平行于 $AN$ 的直线与 $OB$ 相交于 $O$ 与 $B$ 之间，而由于 $M$ 在边 $OA$ 上，这是必然的。若 $\mu\nu=1$，则 $M=A$ 及 $N=B$，整个图形退化；若 $\mu\nu>1$，则 $M$ 和 $N$ 位于延长边上 $A$ 和 $B$ 之外。

## 验算

- 中线。若 $M$ 和 $N$ 是中点，则 $Q$ 是重心（centroid），它把每条中线（median）分成由顶点起计 $2:1$ 的比。所以 $\mu=\nu=\frac{1}{2}$，而公式给出 $m=\frac{(3/2)(1/2)}{3/2}=\frac{1}{2}$，与预期相符。于是 $\lambda=\frac{1}{4}$：在三角形 $OAN$ 中应用中点定理（midpoint theorem），$L$ 是 $ON$ 的中点，即位于 $OB$ 全长的四分之一处。
- 对称性。互换 $A\leftrightarrow B$ 会互换 $M\leftrightarrow N$ 及 $\mu\leftrightarrow\nu$，而 $m$ 与 $n$ 的公式亦正确地互换。答案 $\lambda=\mu\nu$ 同样是对称的。
- 以电脑对约 $10^5$ 个随机三角形进行验算，在边上随机取 $M$ 和 $N$，量度 $\mu$ 和 $\nu$，并确认 $m$、$n$ 及 $\lambda=\mu\nu$ 均准确至 $10^{-12}$ 以内。例如 $m=0.6$ 及 $n=0.55$（如图所示）给出 $\lambda=0.33$。

## 常见错误

- 把分点比（section ratio）的方向弄反：$Q$ 位于由 $M$ 到 $B$ 全程的 $\frac{\mu}{1+\mu}$ 处，而不是由 $B$ 到 $M$。
- 令系数相等而不说明原因。这一步依赖 **a** 与 **b** 不平行。
- 以为“三角形是锐角三角形”这个条件会起作用。其实不会：论证对任何三角形都成立。

## STEP 之外

比较系数之所以可行，是因为 **a** 与 **b** 线性无关（linearly independent），所以它们构成平面的一组基（basis）。这是大学第一个线性代数（linear algebra）课程的核心概念。简洁的答案 $\lambda=\mu\nu$ 是一个比例结果，与塞瓦定理（Ceva's theorem）及梅涅劳斯定理（Menelaus' theorem）同属一类，而这些定理用重心坐标（barycentric coordinates）表达最为简洁。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）对基与坐标有细致的阐述。

详细解答

# STEP II 2018，第 8 题

纯数学 · 伯努利微分方程

**STEP II 2018，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/18-s2-q8

## 关键思路

两条方程都是伯努利方程（Bernoulli equation）：右方出现 $y$ 的幂，破坏了线性（linearity），但只要选取 $y$ 的适当幂作为新变量，方程便会变成线性。设 $y = v^2$（或 $y = v^3$），公因子 $v$（或 $v^2$）可以约去，余下的是一条常系数一阶线性方程（first-order linear equation），可用积分因子（integrating factor）或分离变量（separating variables）求解。至于作图，比较两条曲线最简洁的方法，是把两者都以同一个指数 $u = e^{-t/6}$ 表示。

## 解答

题目写的是 $y_1(x)$ 和 $y_2(x)$，但自变量（independent variable）自始至终都是 $t$，所以我们写成 $y_1(t)$ 和 $y_2(t)$。

### 第 (i) 部分

设 $y = v^2$，其中 $v = \sqrt{y} \geq 0$。于是 $\dfrac{\mathrm{d}y}{\mathrm{d}t} = 2v\dfrac{\mathrm{d}v}{\mathrm{d}t}$ 及 $y^{1/2} = v$，方程变为

$$2v\frac{\mathrm{d}v}{\mathrm{d}t} = \alpha v - \beta v^2.$$

其中一个解是 $v \equiv 0$，即常数解（constant solution）$y \equiv 0$。在 $v \neq 0$ 之处，我们可以除以 $v$：

$$\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{\beta}{2}v = \frac{\alpha}{2}.$$

这是线性方程，积分因子为 $e^{\beta t/2}$：

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(ve^{\beta t/2}\right) = \frac{\alpha}{2}e^{\beta t/2} \quad \Longrightarrow \quad ve^{\beta t/2} = \frac{\alpha}{\beta}e^{\beta t/2} + C.$$

因此通解（general solution）为

$$\sqrt{y} = \frac{\alpha}{\beta} + Ce^{-\beta t/2}, \qquad \text{即} \qquad y = \left(\frac{\alpha}{\beta} + Ce^{-\beta t/2}\right)^2,$$

其中 $C$ 为常数（只取括号内非负的范围，因为 $v = \sqrt{y} \geq 0$），另外还有常数解 $y \equiv 0$。取 $C = 0$ 则得到另一个常数解 $y \equiv \alpha^2/\beta^2$。

要 $y_1(0) = 0$，便需要 $v(0) = 0$，故 $C = -\alpha/\beta$，而

$$\boxed{y_1(t) = \frac{\alpha^2}{\beta^2}\left(1 - e^{-\beta t/2}\right)^2}$$

这个解不是常数，而且当 $t \geq 0$ 时括号 $1 - e^{-\beta t/2}$ 非负，所以它确实是一个解。（在 $t > 0$ 时 $v > 0$，所以除以 $v$ 是合理的。）

**第 (ii) 部分**

现在幂是 $\frac{2}{3}$，所以我们希望 $y^{2/3}$ 与 $y$ 有公因子。设 $y = v^3$，即 $v = y^{1/3} \geq 0$。于是 $\dfrac{\mathrm{d}y}{\mathrm{d}t} = 3v^2\dfrac{\mathrm{d}v}{\mathrm{d}t}$ 及 $y^{2/3} = v^2$，得

$$3v^2\frac{\mathrm{d}v}{\mathrm{d}t} = \alpha v^2 - \beta v^3.$$

**想一想** 你会怎样想到代换 $v = y^{1/3}$?

在第 (i) 部分，$v = y^{1-\frac{1}{2}}$ 令两边都有公因子 $v$。对伯努利方程 $\dot{y} = \alpha y^n - \beta y$ 而言，代换 $v = y^{1-n}$ 总是奏效，而这里 $1 - n = \frac{1}{3}$。

除了 $v \equiv 0$（即常数解 $y \equiv 0$）之外，除以 $v^2$：

$$\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{\beta}{3}v = \frac{\alpha}{3}.$$

以积分因子 $e^{\beta t/3}$，与之前完全一样，得

$$y^{1/3} = \frac{\alpha}{\beta} + De^{-\beta t/3}, \qquad \text{即} \qquad y = \left(\frac{\alpha}{\beta} + De^{-\beta t/3}\right)^3,$$

其中 $D$ 为常数（括号内非负）。条件 $y_2(0) = 0$ 迫使 $D = -\alpha/\beta$，故

$$\boxed{y_2(t) = \frac{\alpha^3}{\beta^3}\left(1 - e^{-\beta t/3}\right)^3}$$

**第 (iii) 部分**

当 $\alpha = \beta$ 时，

$$y_1(t) = \left(1 - e^{-t/2}\right)^2, \qquad y_2(t) = \left(1 - e^{-t/3}\right)^3.$$

每条曲线的形状。两条曲线都从 0 开始并且递增，因为 $1-e^{-t/2}$ 和 $1-e^{-t/3}$ 都由 0 增加并趋向 1。当 $t\to\infty$ 时两者都趋向 1，即 (i) 的常数解 $y=\alpha^2/\beta^2=1$ 与 (ii) 的常数解 $y=\alpha^3/\beta^3=1$。求导得

$$y_1'=e^{-t/2}\left(1-e^{-t/2}\right),\qquad y_2'=e^{-t/3}\left(1-e^{-t/3}\right)^2,$$

所以两者在原点的斜率都是零。斜率 $y_1'=w(1-w)$（其中 $w=e^{-t/2}$）在 $w=\frac12$ 时最大，所以 $y_1$ 在 $t=2\ln 2$、$y_1=\frac14$ 处有拐点（point of inflection）。同样，$y_2'=w(1-w)^2$（其中 $w=e^{-t/3}$）在 $w=\frac13$ 时最大，得拐点在 $t=3\ln 3$、$y_2=\frac{8}{27}$。每条曲线都呈 S 形，由原点上升至渐近线（asymptote）$y=1$。

相对位置。设 $u=e^{-t/6}$，则当 $t>0$ 时 $0<u<1$，且 $e^{-t/2}=u^3$、$e^{-t/3}=u^2$。于是

$$y_1=(1-u^3)^2=(1-u)^2(1+u+u^2)^2,\qquad y_2=(1-u^2)^3=(1-u)^3(1+u)^3,$$

因此

$$\frac{y_1}{y_2}=\frac{(1+u+u^2)^2}{(1-u)(1+u)^3}.$$

展开得 $(1+u+u^2)^2=1+2u+3u^2+2u^3+u^4$ 及 $(1-u)(1+u)^3=(1-u^2)(1+u)^2=1+2u-2u^3-u^4$。分子减分母为

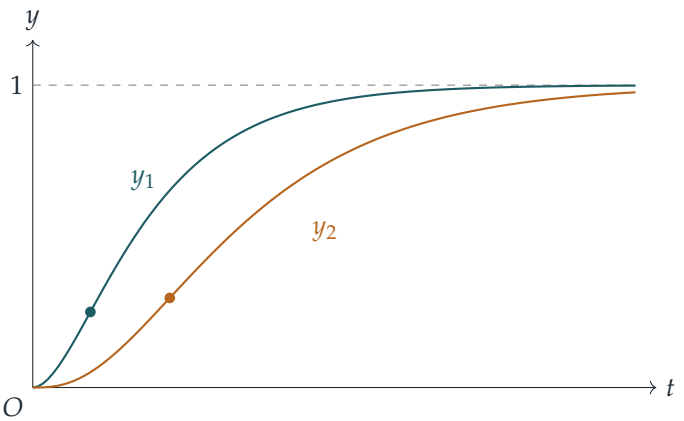
$$3u^2+4u^3+2u^4>0\qquad(0<u<1),$$

而分母为正，所以 $y_1/y_2>1$。因此

$$y_1(t)>y_2(t)\quad\text{对所有}t>0:$$

两条曲线只在原点相交，而在其他地方 $y_1$ 都位于 $y_2$ 之上。

这与两端的表现吻合。在 $t=0$ 附近，$y_1\approx\frac{t^2}{4}$，而 $y_2\approx\frac{t^3}{27}$，后者小得多。当 $t$ 很大时，$1-y_1\approx 2e^{-t/2}$ 而 $1-y_2\approx 3e^{-t/3}$，且 $e^{-t/3}$ 衰减得较慢，所以 $y_2$ 离渐近线较远。




图中的点标示拐点：$y_1$ 上的 $(2\ln 2,\frac14)$ 及 $y_2$ 上的 $(3\ln 3,\frac{8}{27})$。

## 验算

- 代回原方程：对 $y_1$，取 $\alpha = \beta = 1$，则 $y_1^{1/2} - y_1 = (1-w) - (1-w)^2 = w(1-w)$，其中 $w = e^{-t/2}$，这正是 $y_1'$。
- $t$ 很大时：$y_1 \to \alpha^2/\beta^2$，即 $\alpha y^{1/2} = \beta y$ 的非零解；同样 $y_2 \to \alpha^3/\beta^3$，即 $\alpha y^{2/3} = \beta y$ 之处。
- 取 $\alpha = \beta = 1$ 的数值：在 $t = 2$ 时，$y_1 \approx 0.400$、$y_2 \approx 0.115$；在 $t = 10$ 时，$y_1 \approx 0.987$、$y_2 \approx 0.897$。两条公式亦已用电脑对微分方程作逐步数值积分加以验证。

## 常见错误

- 除以 $v$（或 $v^2$）时没有指出 $y \equiv 0$ 也是一个解。正因如此，题目才要求通过原点的非常数解。
- 弄错指数：代换 $y = v^3$ 得出的是 $e^{-\beta t/3}$，而不是 $e^{-\beta t/2}$。
- 单凭斜率判断两条曲线的高低。在两条曲线高度相同、$y \in (0,1)$ 之处，$y^{2/3} > y^{1/2}$，所以 $y_2$ 在那里会上升得较快；但这并不能说明哪条曲线较高，因此需要直接比较。

## STEP 之外

这些是伯努利方程，属大学一年级的标准方程类型，以 $v = y^{1-n}$ 求解。它们亦展示了解的非唯一性（non-uniqueness）：由于 $y^{1/2}$ 和 $y^{2/3}$ 在 $y = 0$ 处斜率无限大，右方在该处不满足利普希茨条件（Lipschitz condition），于是 $y \equiv 0$ 与 $y_1$ 都通过原点。事实上，先停留在 0 直至任意时刻 $t_0$，然后沿着 $y_1(t - t_0)$ 走，便得到一整族解。皮卡–林德勒夫定理（Picard–Lindelöf theorem）准确说明了何时不会出现这种情况。Boyce 与 DiPrima 合著的 *Elementary Differential Equations*（《基础微分方程》）同时涵盖伯努利方程及解的存在性与唯一性。

详细解答

# STEP II 2018，第 9 题

力学 · 重力作用下的碰撞

> **STEP II 2018，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/18-s2-q9

## 关键思路

两粒珠子（bead）的加速度始终相同，都是向下的 $g$，所以两者之间的相对运动是等速的。$A$ 反弹之后，两珠之间的距离以恒定速率缩小，由此即可求出碰撞所需的时间。最后一部分则纯粹是能量与动量的计算：知道 $A$ 撞到地面时的速率，就知道它离开碰撞时的速率，其余交给牛顿实验定律（Newton's experimental law）便可。

## 解答

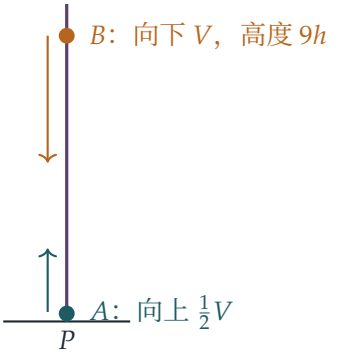
我们把两粒珠子视为质量同为 $m$ 的质点（particle），套在一条光滑（smooth）的竖直金属线上，并忽略空气阻力，因此在两次撞击之间，每粒珠子都只受重力作用而自由运动。取向上为正方向，高度由 $P$ 起量度。

### 直至第一次反弹

$A$ 由静止下落 $8h$，故由能量守恒（conservation of energy）得

$$V^2 = 2g(8h) = 16gh, \qquad V = 4\sqrt{gh}.$$

在同一段时间内，$B$ 亦由静止下落了 $8h$（运动完全相同），所以此刻 $B$ 位于高度 $17h - 8h = 9h$，以速率 $V$ 向下运动。$A$ 从地面以速率 $\frac{1}{2}V$ 向上反弹。




### 碰撞

从 $A$ 离开地面的一刻起，两粒珠子的加速度都是 $-g$，因此 $B$ 相对于 $A$ 的速度一直保持其初始值

$$(-V)-\tfrac{1}{2}V=-\tfrac{3}{2}V.$$

**想一想** 为甚么计算两珠何时相遇时可以不理会重力?

重力令两者的速度改变相同的量 $-gt$，所以在两者之差中互相抵消。于是 $9h$ 的距离以恒定速率 $\frac{3}{2}V$ 缩小，碰撞发生在反弹后

$$\tau=\frac{9h}{\frac{3}{2}V}=\frac{6h}{V}=\frac{6h}{4\sqrt{gh}}=\frac{3}{2}\sqrt{\frac{h}{g}}$$

的时刻。（我们还需要 $A$ 在那时仍在空中；稍后会看到确实如此。）此时 $A$ 的高度为

$$H=\tfrac{1}{2}V\tau-\tfrac{1}{2}g\tau^2=2\sqrt{gh}\cdot\frac{3}{2}\sqrt{\frac{h}{g}}-\frac{1}{2}g\cdot\frac{9h}{4g}=3h-\frac{9}{8}h=\frac{15}{8}h,$$

正如所求。碰撞前一瞬间，

$$\text{速度：}\ A=\tfrac{1}{2}V-g\tau=2\sqrt{gh}-\tfrac{3}{2}\sqrt{gh}=\tfrac{1}{2}\sqrt{gh},$$
$$\text{速度：}\ B=-V-g\tau=-4\sqrt{gh}-\tfrac{3}{2}\sqrt{gh}=-\tfrac{11}{2}\sqrt{gh}.$$

所以 $A$ 仍在上升，而且

$$u_A=\tfrac{1}{2}\sqrt{gh}\,(\text{向上}),\qquad u_B=\tfrac{11}{2}\sqrt{gh}\,(\text{向下}).$$

验算一下：$B$ 的高度为 $9h-V\tau-\frac{1}{2}g\tau^2=9h-6h-\frac{9}{8}h=\frac{15}{8}h$，同样吻合。

### 恢复系数

设 $e$ 为两珠之间的恢复系数（coefficient of restitution），并设 $v_A$、$v_B$ 为两者在碰撞后一瞬间的速度（向上为正）。暂且以 $\sqrt{gh}$ 为单位，于是 $u_A=\frac{1}{2}$ 向上，$u_B=\frac{11}{2}$ 向下。

动量（momentum）（两者质量相等，而在瞬间发生的碰撞中，重力的冲量（impulse）可以忽略）：

$$v_A + v_B = \tfrac{1}{2} - \tfrac{11}{2} = -5.$$

牛顿实验定律：两珠以速率 $\frac{1}{2} + \frac{11}{2} = 6$ 互相接近，碰撞后位于上方的 $B$ 以此速率的 $e$ 倍与 $A$ 分离：

$$v_B - v_A = 6e.$$

两式相减，得 $v_A = -\frac{1}{2}(5+6e)$，所以 $A$ 以速率 $w = \frac{1}{2}(5+6e)\sqrt{gh}$ 向下运动。

现在 $A$ 从高度 $\frac{15}{8}h$ 以初速率 $w$ 下落，并以速率 $V$ 到达地面（它在 $B$ 下方，且向下运动得较快，因为 $v_B = -5 - v_A = \frac{1}{2}(6e-5) > v_A$，所以两珠不会在此之前再次相遇）。由能量守恒，

$$V^2 = w^2 + 2g \cdot \tfrac{15}{8}h \quad \Longrightarrow \quad w^2 = 16gh - \tfrac{15}{4}gh = \tfrac{49}{4}gh,$$

故 $w = \frac{7}{2}\sqrt{gh}$。于是 $\frac{1}{2}(5+6e) = \frac{7}{2}$，得

$$e = \frac{1}{3}.$$

碰撞后，$A$ 以 $\frac{7}{2}\sqrt{gh}$ 向下运动，$B$ 以 $\frac{3}{2}\sqrt{gh}$ 向下运动。

## 验算

- 动能（kinetic energy）：碰撞前总动能为 $\frac{1}{2}m\,gh\,(\frac{1}{4} + \frac{121}{4}) = \frac{61}{4}mgh$；碰撞后为 $\frac{1}{2}m\,gh\,(\frac{49}{4} + \frac{9}{4}) = \frac{29}{4}mgh$。能量有所损失，这正是 $e < 1$ 时应有的情况。
- 若 $e = 1$（且质量相等），两珠会直接交换速度，于是 $A$ 会以 $\frac{11}{2}\sqrt{gh}$ 离开，著地时速率大于 $V$；若 $e = 0$，两珠会以 $\frac{5}{2}\sqrt{gh}$ 一同运动，$A$ 著地时速率会较小。$\frac{1}{3}$ 这个数值合理地介乎两者之间。
- 以电脑逐步模拟整个运动（取 $g = h = 1$），得出碰撞发生在高度 1.875，两者速率分别为 0.5 和 5.5；并且在 $e = \frac{1}{3}$ 时，第二次著地的速率为 $4 = V$。

## 常见错误

- 忘记在 $A$ 下落期间 $B$ 也一直在下落：反弹时 $B$ 位于 $9h$，而不是 $17h$，而且已经以速率 $V$ 运动。
- 牛顿定律中的正负号错误：所有速度都应以同一个正方向表示，并检查得出的分离速率是否为正。

## STEP 之外

在其中一粒下落珠子的参考系（frame）中考虑问题，这个技巧让我们初尝伽利略相对性（Galilean relativity）的滋味：在均匀重力场中，自由下落的参考系完全感受不到重力，而这正是爱因斯坦（Einstein）等效原理（equivalence principle）的出发点。大学一年级力学中的碰撞问题，通常在质心参考系（centre-of-mass frame）中求解；在这个参考系中，恢复系数为 $e$ 的撞击只不过是把两个速度反向，并乘以 $e$。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）对这两个概念都有细致的论述。

详细解答

# STEP II 2018，第 10 题

力学 • 伸长绳上的运动学

**STEP II 2018，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：

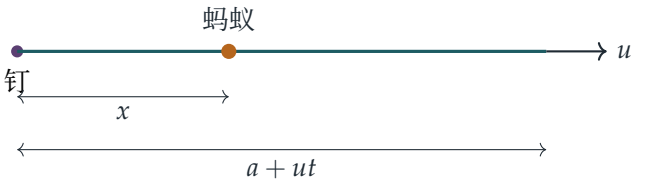
step.maths.org/questions/18-s2-q10

本题在 *Beyond STEP* 第 11.12 节橡皮绳上的蚂蚁中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch11/rubber-rope.html

## 关键思路

由于绳子是均匀（uniform）的，它在每一处都以相同的倍数拉长，所以绳上每一点沿绳的比例位置保持不变。因此，绳上一点的速率与它跟钉的距离成正比。这样，蚂蚁的方程便是线性（linear）的，而最自然要追踪的量是蚂蚁沿绳所处的比例位置 $x/(a+ut)$，而不是 $x$ 本身。

## 解答




### 绳上一点的速率

在时间 $t$，自由端与钉的距离为 $a+ut$，这也就是绳子的长度。绳子是均匀的，所以整条绳都以相同的倍数 $(a+ut)/a$ 被拉长。因此，在 $t=0$ 时与钉相距 $x_0$ 的绳上一点，始终位于沿绳相同的比例位置 $x_0/a$：

$$x = \frac{x_0}{a}(a+ut).$$

固定 $x_0$ 求导，得其速率为 $\dfrac{x_0u}{a}$；以该点目前的位置 $x$ 表示，

$$\text{点的速率，其中该点与钉相距}x = \frac{ux}{a+ut}.$$

作为验证，此速率在钉处为 0，在自由端为 $u$。

### 蚂蚁的方程

蚂蚁相对于钉的速率，等于 $v$ 加上牠脚下绳子的速率，所以

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v + \frac{ux}{a+ut}.$$

利用商法则（quotient rule），再代入这条方程，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{x}{a+ut}\right) &= \frac{1}{a+ut}\frac{\mathrm{d}x}{\mathrm{d}t} - \frac{ux}{(a+ut)^2}\\ &= \frac{1}{a+ut}\left(v + \frac{ux}{a+ut}\right) - \frac{ux}{(a+ut)^2} = \frac{v}{a+ut},\end{aligned}$$

此即所求。

### 到达末端所需的时间

由 0 积分至 $t$，并利用 $t=0$ 时 $x=0$：

$$\frac{x}{a+ut} = \int_0^t \frac{v}{a+us}\,\mathrm{d}s = \frac{v}{u}\ln\left(\frac{a+ut}{a}\right) = \frac{1}{k}\ln\left(1+\frac{ut}{a}\right).$$

左边是蚂蚁已走过的绳长比例，当它等于 1 时，蚂蚁便到达末端。右边由 0 开始持续递增且无上界，所以这情况恰好发生一次，发生于时间 $T$，其中

$$\ln\left(1+\frac{uT}{a}\right) = k \iff 1+\frac{uT}{a} = \mathrm{e}^k \iff uT = a(\mathrm{e}^k - 1),$$

这正是题目给出的结果。

### 回程

现在蚂蚁朝钉的方向走，所以牠自身沿绳的速度为 $-v$，于是

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -v + \frac{ux}{a+ut}.$$

**想一想** 我们需要重做一次商法则的计算吗?

不需要：上面的推导除了 $v$ 的正负号外完全不变，所以

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{x}{a+ut}\right) = -\frac{v}{a+ut}.$$

回程始于 $t = T$，此时蚂蚁在绳的末端，即 $x/(a+ut) = 1$。由 $T$ 积分至 $t$，

$$\frac{x}{a+ut} = 1 - \frac{1}{k}\ln\left(\frac{a+ut}{a+uT}\right).$$

当这个比例为 0 时，蚂蚁便到达钉处；这首先发生于时间 $t_1$，其中

$$\ln\left(\frac{a+ut_1}{a+uT}\right) = k \iff a+ut_1 = (a+uT)\,\mathrm{e}^k.$$

所以回程需时

$$t_1 - T = \frac{(a+uT)(\mathrm{e}^k-1)}{u}.$$

由去程可知 $\mathrm{e}^k - 1 = uT/a$ 及 $a+uT = a\mathrm{e}^k$，所以

$$t_1 - T = \frac{a\mathrm{e}^k}{u}\cdot\frac{uT}{a} = T\mathrm{e}^k.$$

回程需时 $T\mathrm{e}^k$，比去程长 $\mathrm{e}^k$ 倍。

## 验算

- 若 $u \to 0$，绳子不会拉长：$k \to 0$，而 $T = a(\mathrm{e}^k-1)/u \to a/v$，同时回程时间 $T\mathrm{e}^k \to a/v$ 亦然，正合预期。
- 回程时间较长，这是合理的：回程时，蚂蚁脚下的绳子会把牠带离钉。
- 对随机选取的 $a$、$u$ 和 $v$ 值（例如 $a = 1$、$u = v$，因此 $k = 1$ 及 $T = \mathrm{e} - 1 \approx 1.718$），以数值方法逐步推进两条微分方程（differential equation），结果显示蚂蚁在时间 $T$ 到达末端，再经过 $T\mathrm{e} \approx 4.67$ 后回到钉处，与公式吻合。

## 常见错误

- 以为绳上每一点都以速率 $u$ 移动；其实只有自由端是这样。
- 忘记回程始于 $t = T$ 而非 $t = 0$，因此积分常数（constant of integration）会改变。
- 题目要求只以 $T$ 和 $k$ 表示答案，却把答案停留在 $(a+uT)(\mathrm{e}^k-1)/u$。

## STEP 之外

无论绳子拉长得多快，蚂蚁总会到达末端，因为 $\int \mathrm{d}t/(a+ut)$ 发散（diverges）。蚂蚁已走过的距离占绳长的比例，随时间的对数（logarithm）增长；倘若拉长得更快（例如以指数方式），该积分便会收敛（converge），蚂蚁就可能永远到不了。同样的概念——在一个本身正在膨胀的空间中以固定速率前进——也出现在宇宙学（cosmology）中，它决定了遥远星系发出的光能否到达我们。这里通过找出合适的量来解线性方程，其实是积分因子法（integrating-factor method）的一个例子；参见 Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）。

详细解答

# STEP II 2018，第 11 题

力学 · 摩擦力与力矩

**STEP II 2018，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/18-s2-q11

## 关键思路

三条方程正好确定三个未知的接触力：水平方向的牛顿（Newton）第二定律（Newton's second law）、竖直方向没有加速度（acceleration），以及题目给定的条件——对质心（centre of mass）的力矩（moment）互相平衡。只要把两个法向反作用力（normal reaction）都用加速度 $a$ 表示，两种失效情况便各自变成 $a$ 的一个临界值。前轮在其反作用力降至零时离地；后轮在所需的摩擦力（friction）超过其反作用力的 $\mu$ 倍时打滑。哪个临界值较低，哪种情况便先发生。

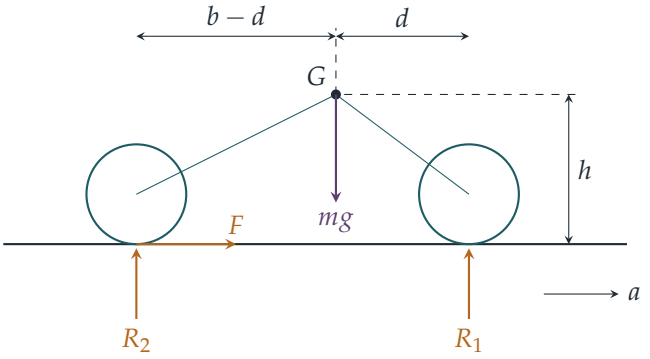
## 解答

### 模型与受力

把电单车视为沿直线运动的刚体（rigid body），向右的加速度为 $a \geq 0$。路面水平，所以质心 $G$ 没有竖直方向的加速度，而车身亦不转动。所受的力为：

- 重力 $mg$，于 $G$ 处向下作用；
- 前轮的法向反作用力 $R_1$（前轮光滑，所以该处没有摩擦力）；
- 后轮的法向反作用力 $R_2$ 及摩擦力 $F$。引擎驱动后轮，所以 $F$ 向前作用，而且是唯一的水平力。

前轮接触点在 $G$ 前方水平距离 $d$ 处，后轮接触点则在 $G$ 后方 $b - d$ 处。两者都在 $G$ 下方距离 $h$ 处。




### 以 $a$ 表示反作用力

水平及竖直方向分别有

$$F = ma, \qquad R_1 + R_2 = mg.$$

现在对 $G$ 取力矩，以逆时针为正（即会令前轮擡起的方向）。重力对 $G$ 没有力矩。$R_1$ 在 $G$ 前方距离 $d$ 处向上作用，产生逆时针力矩 $R_1 d$。$R_2$ 在 $G$ 后方距离 $b-d$ 处向上作用，产生顺时针力矩 $R_2(b-d)$。而 $F$ 沿 $G$ 下方距离 $h$ 的直线向前作用，所以产生逆时针力矩 $Fh$。令总力矩为零：

$$R_1 d + Fh = R_2(b-d).$$

代入 $F = ma$ 及 $R_1 = mg - R_2$：

$$mgd - R_2 d + mah = R_2(b-d) \quad \Longrightarrow \quad R_2 = \frac{m(gd+ah)}{b},$$

于是

$$R_1 = mg - R_2 = \frac{m\big((b-d)g - ah\big)}{b}.$$

随着 $a$ 增大，重量由前轮转移到后轮：$R_1$ 减小，$R_2$ 增大。

### 两种失效各在何时发生？

前轮。只要 $R_1 \geq 0$，前轮便保持贴地，即

$$a \leq \frac{(b-d)g}{h} =: a_{\text{离地}}.$$

后轮。只要 $F \leq \mu R_2$，后轮便不会打滑，即

$$ma \leq \frac{\mu m(gd+ah)}{b} \iff a(b-\mu h) \leq \mu dg.$$

? **想一想** 为甚么我们可以除以 $b-\mu h$ 而不必把不等式（inequality）反向？

这正是给定条件 $\mu < b/h$ 的用途：它表示 $b - \mu h > 0$。因此后轮在以下情况下保持抓地：

$$a \le \frac{\mu d g}{b - \mu h} =: a_{\text{打滑}}.$$

（若没有这个条件，可提供的摩擦力的增长速度至少与所需摩擦力一样快，后轮便永远不会打滑。）

两个临界值都是正数，而当 $a$ 由零开始增大时，后轮先打滑当且仅当 $a_{\text{打滑}} < a_{\text{离地}}$。由于 $b - \mu h > 0$ 且 $h > 0$，我们可以交叉相乘：

$$\begin{aligned}
\frac{\mu d g}{b - \mu h} < \frac{(b-d)g}{h} &\iff \mu d h < (b-d)(b - \mu h) \\
&\iff \mu d h + \mu h (b-d) < b(b-d) \\
&\iff \mu h b < b(b-d) \\
&\iff \mu < \frac{b-d}{h},
\end{aligned}$$

最后一步除以 $bh > 0$。所以若 $(*)$ 成立，后轮会在前轮离地之前打滑，此即所求。（事实上，这串推导显示两个命题是等价的。）

### $(*)$ 成立时的最大加速度

若 $(*)$ 成立，则 $a_{\text{打滑}} < a_{\text{离地}}$。任何不超过 $a_{\text{打滑}}$ 的加速度都同时满足两个条件，所以前轮贴地而后轮抓地。任何更大的加速度都需要 $F > \mu R_2$。因此不打滑时的最大加速度为

$$a_{\max} = \frac{\mu d g}{b - \mu h},$$

此即所求。

### $(*)$ 不成立时

此时 $\mu \ge \dfrac{b-d}{h}$，所以同一串等价推导（把 $<$ 换成 $\ge$）给出 $a_{\text{离地}} \le a_{\text{打滑}}$。因此每个不超过 $a_{\text{离地}}$ 的加速度都令 $R_1 \ge 0$，同时亦满足 $a \le a_{\text{打滑}}$，所以后轮不会打滑。超过 $a_{\text{离地}}$ 便需要 $R_1 < 0$，而只要前轮接触路面，这是不可能的。所以前轮保持接触时的最大加速度为

$$a_{\max} = \frac{(b-d)g}{h}.$$

在这个加速度下 $R_1 = 0$：电单车正处于翘头（wheelie）的临界状态，全部重量都落在后轮上。

## 验算

- 当 $a = 0$：$R_2 = mgd/b$ 及 $R_1 = mg(b-d)/b$，正是按杠杆原理（lever rule）得出的一般静态分配。较接近 $G$ 的车轮承受较多重量。
- 在边界情况 $\mu = (b-d)/h$ 下，有 $b - \mu h = d$，所以 $a_{\text{打滑}} = \mu g = (b-d)g/h = a_{\text{离地}}$。两个答案一致，理应如此。
- 数值例子：$b = 1.5\,\text{m}$，$d = 0.75\,\text{m}$，$h = 0.6\,\text{m}$，$\mu = 0.8$，$g = 9.8\,\text{m s}^{-2}$。此时 $(b-d)/h = 1.25 > \mu$，而 $a_{\text{打滑}} = 5.88/1.02 \approx 5.76\,\text{m s}^{-2}$，远低于 $a_{\text{离地}} = 12.25\,\text{m s}^{-2}$。用电脑对大量随机设置的电单车进行检验，逐步增大 $a$ 并记录哪个条件先失效，结果与两条公式都吻合。

## 常见错误

- 把摩擦力力矩的正负号弄错。在 $G$ 下方向前的摩擦力倾向令前轮擡起，这正是加速会令前轮减载的原因。
- 除以 $b - \mu h$ 却没有说明它为何是正数。这正是条件 $\mu < b/h$ 唯一的作用。
- 对接触点取力矩并令总力矩为零。电单车正在加速，所以力矩只在对 $G$ 时平衡，对其他点则不然。

## STEP 之外

题目给定的力矩条件，是刚体定律的一个特例：对质心的角动量（angular momentum）变化率等于对质心的总转矩（torque）。这里角动量为零，因为电单车不转动。若对任何其他点取力矩，加速中的物体会由 $ma$ 额外产生一个“虚拟”（fictitious）转矩，这便解释了最后一个常见错误。同样的重量转移分析，也是车辆动力学（vehicle dynamics）中刹车与牵引极限的基础。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对刚体动力学有细致的阐述。

详细解答

# STEP II 2018，第 12 题

概率与统计 · 最优停止的期望值

**STEP II 2018，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/18-s2-q12

## 关键思路

目标定为 $h$ 次正面（Heads）时，我只有在游戏结束前达到目标，才会赢得 £$h$，否则分文不得。所以期望收益（expected winnings）就是 $h$ 乘以达到 $h$ 次正面的概率（probability），整道题的重点在于求出这个概率。要在整数 $h$ 上求最大值，不要用微分，而应考虑比值（ratio）$E_{h+1}/E_h$，比较相邻的项。

## 解答

全题设 $q = 1 - p = \frac{1}{N+1}$ 为反面（Tails）的概率，并设 $E_h$ 为目标定为 $h$（其中 $h \geq 1$）时我的期望收益。由于我达到 $h$ 次正面便赢得 £$h$，否则赢得 £0，

$$E_h = h \times \mathrm{P}(\text{我在游戏结束前达到}h\text{ 次正面}).$$

### 第 (i) 部分

在版本 1 中，只要出现一次反面，游戏便结束，所以我达到 $h$ 次正面，当且仅当首 $h$ 次掷币全是正面。各次掷币互相独立（independent），因此这事件的概率为 $p^h$，而

$$E_h = hp^h.$$

要找出最佳的 $h$，可比较相邻的值。由于 $E_h > 0$，

$$\frac{E_{h+1}}{E_h} = \frac{(h+1)p^{h+1}}{hp^h} = \frac{(h+1)N}{h(N+1)},$$

而

$$\frac{E_{h+1}}{E_h} \geq 1 \iff (h+1)N \geq h(N+1) \iff N \geq h,$$

且等号只在 $h = N$ 时成立。所以

$$E_1 < E_2 < \cdots < E_N = E_{N+1} > E_{N+2} > \cdots .$$

这个数列（sequence）在 $h = N$ 之前严格递增，下一个值与它相等，之后便严格递减。所以 $E_h$ 的最大值是 $E_N$，选取 $h = N$ 可使我的期望收益最大，正如所求。（选取 $h = N+1$ 效果同样好。）最大值为 $N\left(\frac{N}{N+1}\right)^N$。

**第 (ii) 部分**

现在游戏只会在连续两次反面时结束。把直至第 $h$ 次正面（包括该次）为止的掷币结果看成 $h$ 个区块，每个区块都以一次正面结尾。如果我安全地达到 $h$ 次正面，那么每个区块必定是

$$\text{H} \quad \text{或} \quad \text{TH},$$

因为连续两次反面会令游戏结束。（一开始就掷出反面是没有问题的：它前面甚么也没有。）反过来，任何由 $h$ 个这类区块组成的序列，都不会含有连续两次反面。序列中每次反面，要么紧接在一次正面之后，要么位于开头，而且必定紧接在一次正面之前。所以我达到 $h$ 次正面，当且仅当掷币结果可分拆为 $h$ 个属于这两类的区块。

? **想一想** 为甚么可以直接把各区块的概率相乘?

每次掷出正面后，游戏的处境与开始时完全相同：没有待定的反面，而且将来的掷币与过去的结果互相独立。所以每个区块，与其他区块独立地，安全的概率为

$$\text{P(H)} + \text{P(TH)} = p + qp = p(1+q) = \frac{N}{N+1} \cdot \frac{N+2}{N+1} = \frac{N(N+2)}{(N+1)^2}.$$

（这两个事件互斥（disjoint），因为它们在第一次掷币已经不同。）因此达到 $h$ 次正面的概率为 $\left(\frac{N(N+2)}{(N+1)^2}\right)^h$，而

$$E_h = \frac{hN^h(N+2)^h}{(N+1)^{2h}},$$

正如所求。

**$N = 2$ 的情况**。此时

$$E_h = h\left(\frac{2 \cdot 4}{9}\right)^h = h\left(\frac{8}{9}\right)^h.$$

与第 (i) 部分一样，

$$\frac{E_{h+1}}{E_h} = \frac{8(h+1)}{9h} \geq 1 \iff 8h + 8 \geq 9h \iff h \leq 8,$$

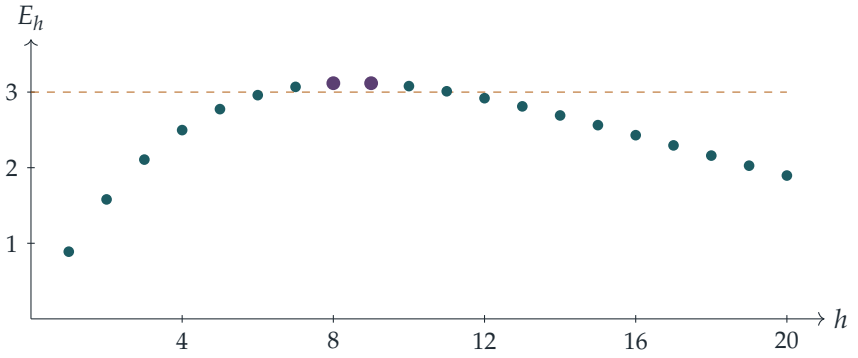
且等号只在 $h = 8$ 时成立。所以 $E_h$ 在 $h = 8$ 之前严格递增，然后 $E_9 = E_8$，之后严格递减。最大值为

$$E_8 = 8\left(\frac{8}{9}\right)^8 = \frac{2^3 \cdot 2^{24}}{3^{16}} = \frac{2^{27}}{3^{16}}.$$

要估算这个值，可取以 3 为底的对数（logarithm），并利用 $\log_3 2 \approx 0.63$：

$$\log_3 E_8 = 27\log_3 2 - 16 \approx 27(0.63) - 16 = 17.01 - 16 = 1.01.$$

所以 $E_8 \approx 3^{1.01}$，非常接近 $3^1 = 3$。最大期望收益约为 £3。




版本 2 中 $N = 2$ 时的值 $E_h = h(8/9)^h$；在 $h = 8, 9$ 处并列的最大值刚好高于 3 的虚线。

## 验算

- 版本 2 只会比版本 1 对玩家更有利，因为每个全是正面的开局在两个版本中都是安全的。事实上 $\frac{N(N+2)}{(N+1)^2} = 1 - \frac{1}{(N+1)^2} > \frac{N}{N+1}$。
- 当 $N = 1$ 时，版本 1 给出 $E_1 = E_2 = \frac{1}{2}$：以概率 $\frac{1}{2}$ 得一次正面，或以概率 $\frac{1}{4}$ 得两次正面。
- 准确值为 $E_8 = 2^{27}/3^{16} = 134\,217\,728/43\,046\,721 \approx 3.118$，所以估计值 £3 相当准确。以电脑模拟（simulation）$2 \times 10^5$ 局 $N = 2$、$h = 8$ 的游戏，所得平均收益约为 £3.11。

## 常见错误

- 在版本 2 中，忘记开头可以有单独一次反面，或误以为第 $h$ 次正面之后还可以出现反面。其实第 $h$ 次正面一出现，我便停止。
- 把 $hp^h$ 微分并令结果等于零。这样得出 $h = -1/\ln p$，它通常不是整数，而要由此找出最佳整数需要小心处理。比值方法更简洁，而且准确。
- 忽略了并列的情况：$h = N$ 与 $h = N + 1$ 给出相同的期望收益，所以题目的“证明”只要求证明 $h = N$ 能达到最大值。

## STEP 之外

这是一个简单的最优停止（*optimal stopping*）问题。这里我预先定下目标，但你可以验证这样做并无损失：每当我手持 $H$ 次正面而没有待定的反面时，多掷一次的增益为正，当且仅当 $H$ 足够小。这类通过比较“现在停止”与“再走一步”来决定何时停止的问题，会引出鞅（martingale）和动态规划（dynamic programming）。第 (ii) 部分的区块论证，让你初尝马尔可夫链（Markov chain）中的更新（renewal）思想。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）是很好的下一步。

详细解答

# STEP II 2018，第 13 题

概率与统计 • 递推关系

**STEP II 2018，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：

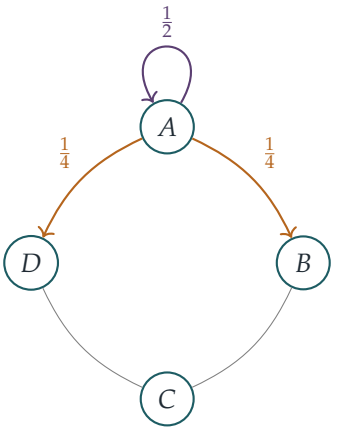
step.maths.org/questions/18-s2-q13

## 关键思路

无论包裹在谁手上，下一下哨声都按固定规则移动它，因此 $A_{n+1}$、$B_{n+1}$、$C_{n+1}$、$D_{n+1}$ 每一个都是 $A_n$、$B_n$、$C_n$、$D_n$ 的固定组合。四条互相耦合的递推关系（recurrence）看似令人生畏，但选对组合（相对两个小孩的和与差），每个组合都各自满足一条只有一行的递推关系，而这些我们可以即时解出。

## 解答

为围成一圈的小孩标上名字。每次吹哨时，持有包裹的人以概率 $\frac{1}{2}$ 保留包裹，并以概率 $\frac{1}{4}$ 传给每一位相邻的小孩；注意，按题目的设置，顺时针与逆时针传递的可能性相同。




图中显示由 $A$ 出发的各种移动；每个小孩处的模式都一样。

**第 (i) 部分**

由 $A$ 开始，吹哨一次后

$$A_1 = \tfrac{1}{2}, \qquad B_1 = \tfrac{1}{4}, \qquad C_1 = 0, \qquad D_1 = \tfrac{1}{4},$$

因为 $C$ 与 $A$ 相隔两个位置，一步之内无法到达。

要在第二次吹哨后落在某个小孩手上，包裹要么本来已在那里而且留下（概率 $\frac{1}{2}$），要么在相邻的小孩手上并被传向正确方向（每个相邻小孩各贡献概率 $\frac{1}{4}$）。所以

$$\begin{aligned}
A_2 &= \tfrac{1}{2}A_1 + \tfrac{1}{4}B_1 + \tfrac{1}{4}D_1 = \tfrac{1}{4} + \tfrac{1}{16} + \tfrac{1}{16} = \tfrac{3}{8},\\
B_2 &= \tfrac{1}{2}B_1 + \tfrac{1}{4}A_1 + \tfrac{1}{4}C_1 = \tfrac{1}{8} + \tfrac{1}{8} + 0 = \tfrac{1}{4},\\
C_2 &= \tfrac{1}{2}C_1 + \tfrac{1}{4}B_1 + \tfrac{1}{4}D_1 = 0 + \tfrac{1}{16} + \tfrac{1}{16} = \tfrac{1}{8},\\
D_2 &= \tfrac{1}{2}D_1 + \tfrac{1}{4}C_1 + \tfrac{1}{4}A_1 = \tfrac{1}{8} + 0 + \tfrac{1}{8} = \tfrac{1}{4}.
\end{aligned}$$

验算一下，$\frac{3}{8} + \frac{1}{4} + \frac{1}{8} + \frac{1}{4} = 1$。

**第 (ii) 部分**

用于 $n = 2$ 的论证对每个 $n \geq 0$ 都成立，依据是全概率定律（law of total probability）（以第 $n$ 次吹哨后谁持有包裹作为条件）：

$$\begin{aligned}
A_{n+1} &= \tfrac{1}{2}A_n + \tfrac{1}{4}(B_n + D_n), & B_{n+1} &= \tfrac{1}{2}B_n + \tfrac{1}{4}(A_n + C_n),\\
C_{n+1} &= \tfrac{1}{2}C_n + \tfrac{1}{4}(B_n + D_n), & D_{n+1} &= \tfrac{1}{2}D_n + \tfrac{1}{4}(A_n + C_n),
\end{aligned}$$

其中 $A_0 = 1$ 及 $B_0 = C_0 = D_0 = 0$。

**求 $B_n$ 和 $D_n$。** 把 $B_{n+1}$ 与 $D_{n+1}$ 的方程相加，

$$B_{n+1} + D_{n+1} = \tfrac{1}{2}(B_n + D_n) + \tfrac{1}{2}(A_n + C_n) = \tfrac{1}{2}(A_n + B_n + C_n + D_n) = \tfrac{1}{2},$$

因为包裹总在某人手上，所以四个概率之和为 1。

? **想一想** 现在已知 $B + D$ 之和，但要怎样把它分配给 $B$ 和 $D$?

改为相减而非相加：$A_n + C_n$ 项互相抵消，剩下

$$B_{n+1} - D_{n+1} = \tfrac{1}{2}(B_n - D_n).$$

由于 $B_0 - D_0 = 0$，可得对每个 $n$ 都有 $B_n - D_n = 0$。（这正是游戏的镜像对称：顺时针与逆时针传递的可能性相同，而 $B$ 和 $D$ 对称地位于 $A$ 的两侧。）结合 $B_n + D_n = \frac{1}{2}$，

$$B_n = D_n = \tfrac{1}{4} \qquad \text{对所有} n \geq 1,$$

当然 $B_0 = D_0 = 0$。这与第 (i) 部分吻合。

**求 $A_n$ 和 $C_n$。** 对 $n \geq 1$，我们现已知道 $A_n + C_n = 1 - (B_n + D_n) = \frac{1}{2}$。把 $A_{n+1}$ 与 $C_{n+1}$ 的方程相减，$B_n + D_n$ 项互相抵消：

$$A_{n+1} - C_{n+1} = \tfrac{1}{2}(A_n - C_n),$$

而此式对所有 $n \geq 0$ 成立。由于 $A_0 - C_0 = 1$，得对所有 $n \geq 0$ 有 $A_n - C_n = \left(\frac{1}{2}\right)^n$。与 $A_n + C_n = \frac{1}{2}$ 联立求解：

$$\boxed{A_n = \frac{1}{4} + \frac{1}{2^{n+1}}, \qquad C_n = \frac{1}{4} - \frac{1}{2^{n+1}} \qquad (n \geq 1).}$$

这些公式在 $n = 0$ 时不成立（会得出 $A_0 = \frac{3}{4}$），因为关系式 $A_n + C_n = \frac{1}{2}$ 要在第一次吹哨之后才开始成立。

## 验算

- $n = 1$：$A_1 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ 及 $C_1 = 0$；$n = 2$：$A_2 = \frac{3}{8}$ 及 $C_2 = \frac{1}{8}$。两者都与第 (i) 部分吻合。
- $n = 3$：公式给出 $(A_3, B_3, C_3, D_3) = (\frac{5}{16}, \frac{1}{4}, \frac{3}{16}, \frac{1}{4})$，其和为 1，并与直接按递推关系逐步计算的结果一致；以电脑模拟 200 000 局游戏，得出 $A_3 \approx 0.313$ 及 $C_3 \approx 0.188$。
- 当 $n \to \infty$ 时，每个概率都趋向 $\frac{1}{4}$：吹哨多次之后，包裹在任何位置的可能性都相同，正如对称性所暗示。

## 常见错误

- 用从 $B$ 传出的概率来写 $B_{n+1}$，而不是用传入 $B$ 的概率。应该问：“包裹在上一刻可能在哪里?”
- 声称 $A_n = \frac{1}{4} + 2^{-(n+1)}$ 对所有 $n \geq 0$ 成立；它在 $n = 0$ 时是错的。
- “凭对称性”假设 $B_n = D_n$ 而不给出任何理由；差的递推关系能令这一点变得严谨。

## STEP 之外

这是一条马可夫链(Markov chain)：一个在四个状态组成的环上进行的随机游走(random walk)，转移矩阵（transition matrix）为 $P$，而 $(A_n, B_n, C_n, D_n) = (1,0,0,0)P^n$。组合 $B_n - D_n$、$A_n - C_n$ 及 $A_n + C_n - B_n - D_n$ 来自 $P$ 的特征向量（eigenvectors）（对应的特征值（eigenvalues）为 $\frac{1}{2}$、$\frac{1}{2}$ 及 0），而以 $2^{-n}$ 的速率收敛（convergence）至均匀分布（uniform distribution），是由第二大的特征值所决定的。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（《概率导论》）对马可夫链及其长期行为有浅白易明的论述。

详细解答

# STEP III 2018，第 1 题

纯数学 · 二次方程的根与不等式

**STEP III 2018，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/18-s3-q1

## 关键思路

韦达定理（Vieta's formulas）把 $u$ 和 $v$ 的两个对称（symmetric）表达式都化为系数 $\alpha$ 和 $\beta$ 的简单函数。一旦固定其中一个表达式，就可以消去 $\alpha$，另一个表达式便成为只含 $\beta$ 的函数，而且正好是第 (i) 部分所描绘的两个函数之一。$u$ 和 $v$ 为实数的条件（判别式（discriminant）$\geq 0$）则告诉我们可以使用曲线的哪一段，其余的就交给草图。

## 解答

### 第 (i) 部分

**曲线** $y = \mathrm{f}(\beta)$。 求导，得

$$\mathrm{f}'(\beta) = 1 + \frac{1}{\beta^2} + \frac{2}{\beta^3} = \frac{\beta^3 + \beta + 2}{\beta^3} = \frac{(\beta+1)(\beta^2 - \beta + 2)}{\beta^3},$$

其中我们看出了三次式的根 $\beta = -1$。二次因式的判别式为 $1 - 8 < 0$，所以它恒为正。因此唯一的驻点（stationary point）位于 $\beta = -1$，此时

$$\mathrm{f}(-1) = -1 + 1 - 1 = -1.$$

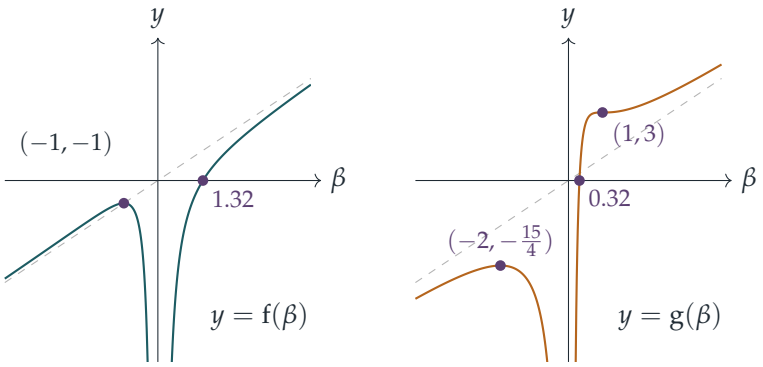
正负号：当 $-1 < \beta < 0$ 时，分子为正而 $\beta^3 < 0$，所以 $\mathrm{f}' < 0$；当 $\beta < -1$ 时两者皆为负，所以 $\mathrm{f}' > 0$。因此 $(-1, -1)$ 是极大值（maximum）点，而且对每个 $\beta < 0$ 都有 $\mathrm{f}(\beta) \leq -1$。当 $\beta > 0$ 时各项皆为正，所以 f 在该处递增（increasing）。

边界的行为：当 $\beta \to 0$（从任一方向）时，$-1/\beta^2$ 一项起主导作用，所以 $\mathrm{f} \to -\infty$。当 $\beta \to \pm\infty$ 时，$\mathrm{f}(\beta) - \beta \to 0$，因此 $y = \beta$ 是斜渐近线（oblique asymptote）。曲线与轴相交一次，交点满足 $\beta^3 - \beta - 1 = 0$，位于 $\beta \approx 1.32$。

**曲线** $y = \mathrm{g}(\beta)$。同样地，

$$\mathrm{g}'(\beta) = 1 - \frac{3}{\beta^2} + \frac{2}{\beta^3} = \frac{\beta^3 - 3\beta + 2}{\beta^3} = \frac{(\beta-1)^2(\beta+2)}{\beta^3}.$$

驻点位于 $\beta = -2$，此时 $\mathrm{g}(-2) = -2 - \frac{3}{2} - \frac{1}{4} = -\frac{15}{4}$；以及位于 $\beta = 1$，此时 $\mathrm{g}(1) = 3$。当 $-2 < \beta < 0$ 时 $\mathrm{g}' < 0$，当 $\beta < -2$ 时 $\mathrm{g}' > 0$，所以 $(-2, -\frac{15}{4})$ 是极大值点，而且对所有 $\beta < 0$ 都有 $\mathrm{g}(\beta) \le -\frac{15}{4}$。当 $\beta > 0$ 时，$\mathrm{g}' \ge 0$，且只在 $\beta = 1$ 时取等号：导数在该处不变号，所以 $(1,3)$ 是驻拐点（stationary point of inflection），而 g 在 $\beta > 0$ 上递增。同样地，当 $\beta \to 0$ 时 $\mathrm{g} \to -\infty$，$y = \beta$ 是渐近线，而且只有一个零点，位于 $\beta \approx 0.32$（即 $\beta^3 + 3\beta - 1 = 0$ 的根）。




虚线是渐近线 $y = \beta$。

### 第 (ii) 部分

由韦达定理，$u + v = -\alpha$ 及 $uv = \beta$（而由于 $\beta \ne 0$，故 $uv \ne 0$）。所以

$$u + v + \frac{1}{uv} = -\alpha + \frac{1}{\beta}, \qquad \frac{1}{u} + \frac{1}{v} + uv = \frac{u+v}{uv} + uv = \beta - \frac{\alpha}{\beta}.$$

### 第 (iii) 部分

条件 $-\alpha + \frac{1}{\beta} = -1$ 给出 $\alpha = 1 + \frac{1}{\beta}$。代入第二个表达式，

$$\frac{1}{u} + \frac{1}{v} + uv = \beta - \frac{1}{\beta}\left(1 + \frac{1}{\beta}\right) = \beta - \frac{1}{\beta} - \frac{1}{\beta^2} = \mathrm{f}(\beta).$$

因此我们须证明对每个容许的 $\beta$ 都有 $\mathrm{f}(\beta) \le -1$。

**想一想** 曲线 $y = \mathrm{f}(\beta)$ 趋向 $+\infty$，所以这个不等式不可能对所有 $\beta$ 成立。是甚么阻止 $\beta$ 变得很大?

是 $u$ 和 $v$ 必须为实数。两根为实数当且仅当 $\alpha^2 \geq 4\beta$，即

$$\left(1+\frac{1}{\beta}\right)^2 \geq 4\beta \iff (\beta+1)^2 \geq 4\beta^3 \iff 4\beta^3-\beta^2-2\beta-1 \leq 0,$$

其中我们乘以了 $\beta^2 > 0$。现在 $\beta = 1$ 是这个三次式的根，而且

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

二次因式的判别式为 $9-16 < 0$，所以它恒为正。因此两根为实数当且仅当 $\beta \leq 1$（且 $\beta \neq 0$）。

- 若 $\beta < 0$：由第 (i) 部分，$f(\beta) \leq f(-1) = -1$。
- 若 $0 < \beta \leq 1$：f 在 $\beta > 0$ 上递增，所以 $f(\beta) \leq f(1) = 1-1-1 = -1$。

在所有情况下都有 $\frac{1}{u}+\frac{1}{v}+uv \leq -1$，命题得证。等号在 $\beta = -1$（$\alpha = 0$，两根为 $\pm 1$）及 $\beta = 1$（$\alpha = 2$，重根（double root）$-1$）时成立。

**第 (iv) 部分**

现在 $-\alpha+\frac{1}{\beta} = 3$，所以 $\alpha = \frac{1}{\beta}-3$，而且

$$\frac{1}{u}+\frac{1}{v}+uv = \beta-\frac{1}{\beta}\left(\frac{1}{\beta}-3\right) = \beta+\frac{3}{\beta}-\frac{1}{\beta^2} = g(\beta).$$

当 $\alpha^2 \geq 4\beta$ 时两根为实数：

$$\left(\frac{1}{\beta}-3\right)^2 \geq 4\beta \iff (1-3\beta)^2 \geq 4\beta^3 \iff 4\beta^3-9\beta^2+6\beta-1 \leq 0.$$

同样，$\beta = 1$ 是一个根，而这次它是重根：

$$4\beta^3-9\beta^2+6\beta-1 = (\beta-1)(4\beta^2-5\beta+1) = (\beta-1)^2(4\beta-1).$$

由于 $(\beta-1)^2 \geq 0$，这个三次式 $\leq 0$ 当且仅当 $\beta \leq \frac{1}{4}$ 或 $\beta = 1$。因此容许的值为 $\beta < 0$、$0 < \beta \leq \frac{1}{4}$，以及孤立的值 $\beta = 1$。

**想一想** 最大值是否就是区间右端的 $g(\frac{1}{4})$?

并不完全是：别遗漏那个孤立点。在三个部分上分别考虑：

- $\beta < 0$：由第 (i) 部分，$g(\beta) \leq -\frac{15}{4}$。
- $0 < \beta \leq \frac{1}{4}$：g 递增，所以 $g(\beta) \leq g(\frac{1}{4}) = \frac{1}{4}+12-16 = -\frac{15}{4}$。
- $\beta = 1$：$g(1) = 3$。

所以 $\frac{1}{u}+\frac{1}{v}+uv$ 的最大值是 $\boxed{3}$，在 $\beta = 1$、$\alpha = -2$，即 $u = v = 1$ 时取得。（事实上两个表达式都等于 $1+1+1 = 3$。）若没有那个重根，就不可能得到任何大于 $-\frac{15}{4}$ 的值。

## 验算

- 把 $u = v = 1$ 代入两个表达式：两者皆为 3，与第 (iv) 部分吻合。代入 $u = v = -1$：两者皆为 $-1$，即第 (iii) 部分取等号的情况。
- 在第 (iv) 部分试 $\beta = \frac{1}{4}$：$\alpha = 1$，而 $x^2 + x + \frac{1}{4} = (x + \frac{1}{2})^2$，所以 $u = v = -\frac{1}{2}$。于是 $u + v + \frac{1}{uv} = -1 + 4 = 3$ 及 $\frac{1}{u} + \frac{1}{v} + uv = -4 + \frac{1}{4} = -\frac{15}{4}$，与所述相符。
- 用电脑在数千个给出实根的随机 $\beta$ 中搜索，在第 (iii) 部分找不到任何大于 $-1$ 的值，而在第 (iv) 部分除 $\beta = 1$ 外也找不到任何大于 $-3.75$ 的值。

## 常见错误

- 忘记判别式条件，因而在第 (iii) 部分得出没有上界的结论。
- 把 $(\beta - 1)^2(4\beta - 1) \leq 0$ 解成只有 $\beta \leq \frac{1}{4}$，遗漏了 $\beta = 1$，结果答案变成 $-\frac{15}{4}$ 而非 3。
- 把不等式 $\alpha^2 \geq 4\beta$ 乘以 $\beta$（它可能是负数），而不是乘以 $\beta^2$。

## STEP 之外

这是一个小型的约束最优化（constrained optimisation）问题：我们在给出实根的系数集合上求函数的最大值，而这个集合除了区间之外，还可以包含孤立点。函数在一个集合上的最大值不一定出现在驻点或区间端点，这正是为何在分析和最优化中，仔细描述可行集（feasible set）如此重要。Abbott 的 *Understanding Analysis*（《理解分析》）对连续函数在一般集合上的极值有严谨的论述。

详细解答

# STEP III 2018，第 2 题

纯数学 · 埃尔米特多项式

**STEP III 2018，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/18-s3-q2

## 关键思路

把定义倒过来看：它说明 $z$ 的 $n$ 阶导数（derivative）等于 $(-1)^n y_n z$。将这式再微分一次，并与 $n+1$ 时的同一个式子比较，一行便得出第 (i) 部分。之后，第 (i) 部分就成为我们微分递推关系（recurrence）的工具，推动第 (ii) 部分的归纳法（induction）。最后的不等式（inequality）则由第 (ii) 部分的恒等式（identity），再用一次非常简短的归纳法得出。

## 解答

全文中 $z = \mathrm{e}^{-x^2}$，故 $z' = -2xz$，而且对每个 $x$ 都有 $z > 0$。将定义移项，得

$$\frac{\mathrm{d}^n z}{\mathrm{d}x^n} = (-1)^n y_n z \qquad (n \geq 0),$$

由于 $y_0 = 1$，此式在 $n = 0$ 时亦成立。

### 第 (i) 部分

利用乘积法则（product rule）及 $z' = -2xz$，将上面列出的等式微分：

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

但以 $n+1$ 代替 $n$ 的同一等式指出，左方等于 $(-1)^{n+1} y_{n+1} z$。令两者相等，再除以永不为零的 $(-1)^n z$，得

$$y_n' - 2x y_n = -y_{n+1}, \qquad \text{即} \qquad \frac{\mathrm{d}y_n}{\mathrm{d}x} = 2x y_n - y_{n+1},$$

此即所求。以上推导并不需要 $n \geq 1$，因此结果在 $n = 0$ 时亦成立；下面会用到这一点。

留待后用：$z' = -2xz$ 及 $z'' = (4x^2 - 2)z$，故 $y_1 = 2x$ 及 $y_2 = 4x^2 - 2$。

**第 (ii) 部分**

**基本情况。** 当 $n = 1$ 时，我们需要 $y_2 = 2xy_1 - 2y_0$，而确实 $2x \cdot 2x - 2 = 4x^2 - 2 = y_2$。

**归纳步骤。** 假设对某个 $k \geq 1$，

$$y_{k+1} = 2xy_k - 2ky_{k-1}. \tag{$*$}$$

我们想得到往后一步的同一关系，所以将 $(*)$ 微分：

$$y'_{k+1} = 2y_k + 2xy'_k - 2ky'_{k-1}.$$

? **想一想** 怎样消去这些导数？当 $k = 1$ 时，第 (i) 部分可以用于 $y_{k-1}$ 吗？

利用第 (i) 部分替换每一个导数。当 $k = 1$ 时，$y'_{k-1}$ 对应 $n = 0$ 的情况，而我们已指出这情况同样成立。于是得

$$2xy_{k+1} - y_{k+2} = 2y_k + 2x\big(2xy_k - y_{k+1}\big) - 2k\big(2xy_{k-1} - y_k\big),$$

所以

$$y_{k+2} = 4xy_{k+1} - 4x^2y_k - 2(k+1)y_k + 4kx\,y_{k-1}.$$

含 $y_{k-1}$ 的项再用一次假设 $(*)$ 消去：$2ky_{k-1} = 2xy_k - y_{k+1}$，故 $4kx\,y_{k-1} = 4x^2y_k - 2xy_{k+1}$。代入后，

$$\begin{aligned} y_{k+2} &= 4xy_{k+1} - 4x^2y_k - 2(k+1)y_k + 4x^2y_k - 2xy_{k+1} \\ &= 2xy_{k+1} - 2(k+1)y_k, \end{aligned}$$

这正是以 $k+1$ 代替 $k$ 的关系。由归纳法，对所有 $n \geq 1$，$y_{n+1} = 2xy_n - 2ny_{n-1}$。

**推论。** 使用递推关系两次，一次在 $n$，一次在 $n+1$，使两个乘积中都出现 $2xy_ny_{n+1}$ 这一项：

$$\begin{aligned} y_{n+1}^2 &= y_{n+1}\big(2xy_n - 2ny_{n-1}\big) = 2xy_ny_{n+1} - 2ny_{n-1}y_{n+1}, \\ y_ny_{n+2} &= y_n\big(2xy_{n+1} - 2(n+1)y_n\big) = 2xy_ny_{n+1} - 2(n+1)y_n^2. \end{aligned}$$

相减后，含 $x$ 的项互相抵消：

$$y_{n+1}^2 - y_ny_{n+2} = 2(n+1)y_n^2 - 2ny_{n-1}y_{n+1} = 2n\big(y_n^2 - y_{n-1}y_{n+1}\big) + 2y_n^2,$$

此即所求的恒等式。

**第 (iii) 部分**

记 $D_n = y_n^2 - y_{n-1}y_{n+1}$。第 (ii) 部分的恒等式可写成

$$D_{n+1} = 2nD_n + 2y_n^2 \qquad (n \geq 1).$$

我们用归纳法证明对所有 $n \geq 1$，$D_n > 0$。当 $n = 1$ 时，

$$D_1 = y_1^2 - y_0y_2 = 4x^2 - (4x^2 - 2) = 2 > 0$$

对每个 $x$ 都成立。若对某个 $n \geq 1$ 有 $D_n > 0$，则由于 $n \geq 1$，$2nD_n > 0$，又 $2y_n^2 \geq 0$，故 $D_{n+1} > 0$。因此对所有 $n \geq 1$ 及所有实数 $x$，$y_n^2 - y_{n-1}y_{n+1} > 0$，此即所求。

注意，要得到严格不等式，我们需要 $D_n$ 严格为正：单靠 $y_n^2$ 是不够的，因为在 $y_n$ 的零点上它可以等于零。

## 验算

- 由递推关系，$y_3 = 2x(4x^2 - 2) - 4 \cdot 2x = 8x^3 - 12x$。直接将 $z'' = (4x^2 - 2)z$ 微分，得 $z''' = (-8x^3 + 12x)z$，与 $-y_3z$ 吻合。
- 第 (i) 部分在 $n = 1$ 时：$y_1' = 2$，而 $2x \cdot 2x - (4x^2 - 2) = 2$。✓
- 第 (iii) 部分在 $n = 2$ 时：$D_2 = (4x^2 - 2)^2 - 2x(8x^3 - 12x) = 8x^2 + 4$，这是正数，并且等于 $2D_1 + 2y_1^2 = 4 + 8x^2$，与恒等式的预测一致。
- 以电脑在多个有理数 $x$ 值（包括接近各 $y_n$ 零点之处）检验递推关系及不等式直至 $n = 14$，结果完全吻合。

## 常见错误

- 忘记归纳步骤要对 $y_{k-1}$ 使用第 (i) 部分，而在 $k = 1$ 时这是指 $y_0$：题目只要求第 (i) 部分对 $n \geq 1$ 成立，所以要另外检查 $n = 0$ 的情况。
- 在第 (iii) 部分只得出 $D_{n+1} \geq 0$：严格不等式来自 $D_n > 0$，而不是来自 $y_n^2$。

## STEP 之外

函数 $y_n$ 就是埃尔米特多项式（Hermite polynomials）$H_n(x)$，而关系式 $\mathrm{d}^n z/\mathrm{d}x^n = (-1)^n H_n z$ 是它们的罗德里格公式（Rodrigues formula）。它们在整条实数线上以 $\mathrm{e}^{-x^2}$ 为权函数正交（orthogonal），并给出量子谐振子（quantum harmonic oscillator）的能量状态。第 (iii) 部分的不等式是一个图兰型不等式（Turán-type inequality），这类结果对

许多经典正交多项式都成立。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）通过级数解处理埃尔米特方程。

详细解答

# STEP III 2018，第 3 题

纯数学 · 欧拉微分方程

**STEP III 2018，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/18-s3-q3

## 关键思路

左边是一条欧拉方程（Euler equation，又称等维方程，equidimensional equation）：每个导数都配上相应次数的 $x$ 幂。把嵌套形式 $(*)$ 展开并比较系数（comparing coefficients），就能确定 $a$、$b$ 和 $c$。方程一旦写成这个形式，便可以通过两次简单的积分求解，每次剥去一层导数；而分情况讨论的需要，正正出现在其中一个积分是 $x^{-1}$ 的积分、因而产生对数的时候。

## 解答

全题我们都在 $x > 0$ 上讨论，使 $x$ 的所有幂都有定义。

**求 $a$、$b$ 和 $c$**

由乘积法则（product rule），

$$(x^c y)' = x^c y' + cx^{c-1}y, \qquad \text{所以} \qquad x^b(x^c y)' = x^{b+c}y' + cx^{b+c-1}y.$$

再求导一次，

$$\left(x^b(x^c y)'\right)' = x^{b+c}y'' + (b+c)x^{b+c-1}y' + cx^{b+c-1}y' + c(b+c-1)x^{b+c-2}y.$$

乘以 $x^a$，并选取 $a+b+c=2$，得

$$x^a\left(x^b(x^c y)'\right)' = x^2y'' + (b+2c)\,xy' + c(b+c-1)\,y.$$

这与题目所给的左边吻合，只要

$$b + 2c = 1 - 2p, \qquad c(b+c-1) = p^2 - q^2.$$

由第一式得 $b+c-1=-2p-c$，于是第二式变为 $-c^2-2pc=p^2-q^2$，即 $(c+p)^2=q^2$。取 $c+p=q$，

$$c=q-p,\qquad b=1-2p-2c=1-2q,\qquad a=2-b-c=1+p+q.$$

因此方程可写成

$$x^{1+p+q}\left(x^{1-2q}\left(x^{q-p}y\right)'\right)'=\mathrm{f}(x),$$

此即所求。（另一个根 $c+p=-q$ 给出的结果只是把 $q$ 换成 $-q$，两者并无分别，因为方程中只出现 $q^2$。）

### 第 (i) 部分

当 $\mathrm{f}=0$ 时，我们可以除以 $x^{1+p+q}\neq 0$：

$$\left(x^{1-2q}\left(x^{q-p}y\right)'\right)'=0\quad\Longrightarrow\quad x^{1-2q}\left(x^{q-p}y\right)'=A\quad\Longrightarrow\quad\left(x^{q-p}y\right)'=Ax^{2q-1},$$

其中 $A$ 为常数。下一次积分的做法，取决于幂 $2q-1$ 是否等于 $-1$。

> **想一想** 为甚么 $p$ 的值在这里不会引出不同的情况?

我们要积分的指数是 $2q-1$，它与 $p$ 无关；$p$ 只通过因子 $x^{q-p}$ 出现，而这个因子我们最后才除去。所以情况的划分只取决于 $q$。

**情况** $q\neq 0$。积分得 $x^{q-p}y=\dfrac{A}{2q}x^{2q}+B$，再乘以 $x^{p-q}$（并把常数 $\frac{A}{2q}$ 重新记作 $A$），

$$y=Ax^{p+q}+Bx^{p-q}.$$

**情况** $q=0$。此时 $\left(x^{-p}y\right)'=Ax^{-1}$，所以 $x^{-p}y=A\ln x+B$，从而

$$y=x^p\left(A\ln x+B\right).$$

每种情况都有两个任意常数（arbitrary constant），正如二阶方程所应有的。

### 第 (ii) 部分

这是 $q=0$、$\mathrm{f}(x)=x^n$ 的情况，所以 $a=1+p$、$b=1$、$c=-p$，方程为

$$x^{1+p}\left(x\left(x^{-p}y\right)'\right)'=x^n\quad\Longleftrightarrow\quad\left(x\left(x^{-p}y\right)'\right)'=x^{n-p-1}.$$

第一次积分涉及 $x^{n-p-1}$，它恰好在 $n=p$ 时等于 $x^{-1}$。

**情况** $n\neq p$。积分一次，

$$x\left(x^{-p}y\right)'=\frac{x^{n-p}}{n-p}+A\quad\Longrightarrow\quad\left(x^{-p}y\right)'=\frac{x^{n-p-1}}{n-p}+\frac{A}{x}.$$

由于 $n-p\neq 0$，再积分得

$$x^{-p}y=\frac{x^{n-p}}{(n-p)^2}+A\ln x+B,\qquad \text{所以}\qquad y=\frac{x^n}{(n-p)^2}+x^p\,(A\ln x+B).$$

**情况** $n=p$。此时第一次积分便产生对数：

$$x\big(x^{-p}y\big)'=\ln x+A\quad\Longrightarrow\quad \big(x^{-p}y\big)'=\frac{\ln x}{x}+\frac{A}{x}.$$

由于 $\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{2}(\ln x)^2=\frac{\ln x}{x}$，

$$x^{-p}y=\tfrac{1}{2}(\ln x)^2+A\ln x+B,\qquad \text{所以}\qquad y=x^p\left(\tfrac{1}{2}(\ln x)^2+A\ln x+B\right).$$

在两种情况中，含 $A$ 和 $B$ 的部分正是第 (i) 部分中 $q=0$ 时的余函数（complementary function），理应如此。

## 验算

- 取 $p=q=0$：方程为 $x^2y''+xy'=0$，即 $(xy')'=0$，所以 $y=A\ln x+B$，与第 (i) 部分吻合。
- 在齐次方程（homogeneous equation）中试代 $y=x^m$：$m(m-1)+(1-2p)m+p^2-q^2=(m-p)^2-q^2$，所以 $m=p\pm q$，当 $q=0$ 时为重根（repeated root），与上面所得一致。
- 在第 (ii) 部分中取 $n=p=1$，$y=\frac{1}{2}x(\ln x)^2$ 有 $y'=\frac{1}{2}(\ln x)^2+\ln x$ 及 $y''=\frac{\ln x}{x}+\frac{1}{x}$，所以 $x^2y''-xy'+y=x\ln x+x-\frac{1}{2}x(\ln x)^2-x\ln x+\frac{1}{2}x(\ln x)^2=x$，此即所求。
- 所有解亦已用电脑在大量随机的 $p$、$q$、$n$ 和 $x$ 值下代回方程验证；残差（residual）在舍入误差范围内为零（约 $10^{-6}$）。

## 常见错误

- 在 (i) 中遗漏 $q=0$ 的情况，或在 (ii) 中遗漏 $n=p$ 的情况；这些情况下要用 $\int x^{-1}\,\mathrm{d}x=\ln x$ 代替幂法则。
- 在第一步遗漏积分常数：正是它产生第二个线性独立的解。
- 没有说明理由便假设 $a+b+c=2$：这是为了使 $y''$ 乘以 $x^2$ 所必需的。

## STEP 之外

把二阶算子写成一阶算子的复合，称为把算子因式分解（factorising），这正是降阶法（reduction of order）以及量子力学中阶梯算子（ladder operators）背后的想法。代换

$x = \mathrm{e}^t$ 可把任何欧拉方程变成常系数方程，这就解释了重根 $m = p$ 以及 $\ln x$（即 $t$）这个因子的出现。Boyce 和 DiPrima 的 *Elementary Differential Equations*（《基础微分方程》）论述了欧拉方程和降阶法。

详细解答

# STEP III 2018，第 4 题

纯数学 · 双曲线的切线

**STEP III 2018，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/18-s3-q4

## 关键思路

求切线最快的方法是直接用标准结果：$x^2/a^2 - y^2/b^2 = 1$ 在 $(x_1, y_1)$ 处的切线（tangent）是 $xx_1/a^2 - yy_1/b^2 = 1$。之后的一切都只是两条直线相交，所以只需解几组一次联立方程，再整理三角式子。在第 (ii) 部分，垂直条件是 $\sin\theta$ 与 $\sin\phi$ 之间一条涉及 $b$ 的关系式，正正是它让我们能把 $b$ 从式子中消去，最后再以 $a^2 - b^2$ 的形式把它带回来。和差化积公式（sum-to-product formulae）能令最后的代数运算变得简短。

## 解答

### $P$ 点的切线

对 $x^2/a^2 - y^2/b^2 = 1$ 进行隐函数求导（implicit differentiation），得 $\dfrac{2x}{a^2} - \dfrac{2y}{b^2}\dfrac{\mathrm{d}y}{\mathrm{d}x} = 0$，因此在 $(x_1, y_1)$ 处的斜率（gradient）是 $\dfrac{b^2x_1}{a^2y_1}$，而切线为

$$y - y_1 = \frac{b^2x_1}{a^2y_1}(x - x_1) \iff \frac{xx_1}{a^2} - \frac{yy_1}{b^2} = \frac{x_1^2}{a^2} - \frac{y_1^2}{b^2} = 1,$$

这里用到 $(x_1, y_1)$ 在曲线上这一事实。（这个形式同样适用于顶点（vertex），即 $y_1 = 0$ 而切线是铅垂线的情况。）代入 $x_1 = a\sec\theta$、$y_1 = b\tan\theta$：

$$\frac{x\sec\theta}{a} - \frac{y\tan\theta}{b} = 1.$$

两边乘以 $ab\cos\theta$（它不等于零，因为 $\sec\theta$ 存在），得

$$bx - ay\sin\theta = ab\cos\theta,$$

即所求。

**第 (i) 部分**

题目给出的两条直线就是渐近线（asymptote）$y = \pm\frac{b}{a}x$。

在 $S$ 点，$ay = bx$。把 $ay = bx$ 代入切线方程，得 $bx(1-\sin\theta) = ab\cos\theta$，所以

$$S = \left(\frac{a\cos\theta}{1-\sin\theta}, \frac{b\cos\theta}{1-\sin\theta}\right).$$

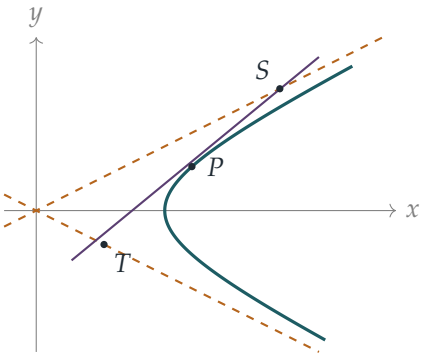
在 $T$ 点，$ay = -bx$，得 $bx(1+\sin\theta) = ab\cos\theta$，所以

$$T = \left(\frac{a\cos\theta}{1+\sin\theta}, -\frac{b\cos\theta}{1+\sin\theta}\right).$$

两个分母都不为零，因为 $\cos\theta \neq 0$ 意味着 $\sin\theta \neq \pm 1$。将两者相加，并利用 $(1-\sin\theta)(1+\sin\theta) = \cos^2\theta$：

$$x_S + x_T = a\cos\theta \cdot \frac{(1+\sin\theta)+(1-\sin\theta)}{\cos^2\theta} = \frac{2a}{\cos\theta} = 2a\sec\theta,$$
$$y_S + y_T = b\cos\theta \cdot \frac{(1+\sin\theta)-(1-\sin\theta)}{\cos^2\theta} = \frac{2b\sin\theta}{\cos\theta} = 2b\tan\theta.$$

因此 $ST$ 的中点（mid-point）是 $(a\sec\theta, b\tan\theta)$：**$ST$ 的中点正是 $P$ 本身**。任一点的切线被两条渐近线所截的线段，都在切点处被平分。




**第 (ii) 部分**

**垂直条件**。把 $P$ 点的切线写成 $ay\sin\theta = bx - ab\cos\theta$，当 $\sin\theta \neq 0$ 时其斜率为 $\dfrac{b}{a\sin\theta}$。（若 $\sin\theta = 0$，切线是铅垂线，那么 $Q$ 点的切线就必须是水平线；但由于 $b/(a\sin\phi) \neq 0$，没有任何切线是水平的。所以 $\sin\theta \neq 0$，同理 $\sin\phi \neq 0$。）两条切线互相垂直（perpendicular），需要两斜率之积为 $-1$：

$$\frac{b^2}{a^2\sin\theta\sin\phi} = -1 \qquad \Longleftrightarrow \qquad \sin\theta\sin\phi = -\frac{b^2}{a^2}. \tag{$*$}$$

特别地，$\sin\theta$ 与 $\sin\phi$ 异号，所以 $\sin\phi - \sin\theta \neq 0$。留意 $(*)$ 需要 $b^2/a^2 < 1$，这正是用到 $a > b$ 的地方。

**交点**。 点 $(x,y)$ 满足

$$bx - ay\sin\theta = ab\cos\theta, \qquad bx - ay\sin\phi = ab\cos\phi.$$

以第一式减去第二式，得 $ay(\sin\phi - \sin\theta) = ab(\cos\theta - \cos\phi)$。把第一式乘以 $\sin\phi$、第二式乘以 $\sin\theta$，再相减，得 $bx(\sin\phi - \sin\theta) = ab(\sin\phi\cos\theta - \cos\phi\sin\theta) = ab\sin(\phi - \theta)$。因此

$$x = \frac{a\sin(\phi-\theta)}{\sin\phi - \sin\theta}, \qquad y = \frac{b(\cos\theta - \cos\phi)}{\sin\phi - \sin\theta}.$$

**想一想** 这些式子仍然含有 $b$。怎样才能只用 $a$、$\theta$ 和 $\phi$ 表示 $x^2$ 和 $y^2$？这些分式又能否化简？

条件 $(*)$ 表示 $b^2 = -a^2\sin\theta\sin\phi$，所以取平方就能消去 $b$。为了化简，设 $\alpha = \frac{1}{2}(\phi+\theta)$ 及 $\beta = \frac{1}{2}(\phi-\theta)$。和差化积公式给出

$$\sin\phi - \sin\theta = 2\cos\alpha\sin\beta, \qquad \cos\theta - \cos\phi = 2\sin\alpha\sin\beta, \qquad \sin(\phi-\theta) = 2\sin\beta\cos\beta,$$

而由于 $\sin\phi - \sin\theta \neq 0$，所以 $\cos\alpha\sin\beta \neq 0$。约去 $2\sin\beta$，

$$x = \frac{a\cos\beta}{\cos\alpha}, \qquad y = b\tan\alpha,$$

因此

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

$$y^2 = -a^2\sin\theta\sin\phi\,\tan^2\left(\tfrac{1}{2}(\phi+\theta)\right).$$

等价地，不用半角可写成

$$x^2 = \frac{a^2\sin^2(\phi-\theta)}{(\sin\phi-\sin\theta)^2}, \qquad y^2 = -\frac{a^2\sin\theta\sin\phi\,(\cos\theta-\cos\phi)^2}{(\sin\phi-\sin\theta)^2}.$$

**圆**。 我们还需要以 $\alpha$ 和 $\beta$ 表示 $\sin\theta\sin\phi$。由于 $\theta = \alpha - \beta$ 及 $\phi = \alpha + \beta$，

$$\begin{aligned}\sin\theta\sin\phi &= \sin^2\alpha\cos^2\beta - \cos^2\alpha\sin^2\beta \\ &= (1-\cos^2\alpha)\cos^2\beta - \cos^2\alpha(1-\cos^2\beta) = \cos^2\beta - \cos^2\alpha.\end{aligned}$$

设 $k = \sin\theta\sin\phi = \cos^2\beta - \cos^2\alpha$。于是，利用 $\tan^2\alpha = (1-\cos^2\alpha)/\cos^2\alpha$，

$$\begin{aligned}x^2 + y^2 &= \frac{a^2}{\cos^2\alpha}\left[\cos^2\beta - k\,(1-\cos^2\alpha)\right] \\ &= \frac{a^2}{\cos^2\alpha}\left[(\cos^2\beta - k) + k\cos^2\alpha\right] \\ &= \frac{a^2}{\cos^2\alpha}\left[\cos^2\alpha + k\cos^2\alpha\right] \qquad \text{因为 } \cos^2\beta - k = \cos^2\alpha \\ &= a^2(1+k) = a^2 + a^2\sin\theta\sin\phi = a^2 - b^2,\end{aligned}$$

最后一步用到 $(*)$。所以任意两条互相垂直的切线的交点都在圆 $x^2+y^2=a^2-b^2$ 上，即所求。

## 验算

- 取 $a=2$、$b=1$、$\theta=\frac{\pi}{6}$、$\phi=-\frac{\pi}{6}$。那么 $\sin\theta\sin\phi=-\frac{1}{4}=-b^2/a^2$，所以两切线互相垂直（斜率分别为 1 和 $-1$）。此时 $\alpha=0$、$\beta=-\frac{\pi}{6}$，所以 $x=2\cos\frac{\pi}{6}=\sqrt{3}$、$y=0$，而 $x^2+y^2=3=a^2-b^2$。直接验证：切线 $x-y=\sqrt{3}$ 与 $x+y=\sqrt{3}$ 的确相交于 $(\sqrt{3},0)$。
- 在第 (i) 部分，当 $\theta\to 0$ 时，切线变成铅垂线 $x=a$，它与渐近线相交于 $(a,\pm b)$，中点为 $(a,0)=P$。
- 以上所有公式，包括 (i) 中的中点以及 $x^2+y^2=a^2-b^2$，都已用电脑对数以千计随机选取的 $a>b>0$ 及互相垂直的 $\theta,\phi$ 组合，直接解两条切线方程加以确认。

## 常见错误

- 在 $x^2$ 和 $y^2$ 的式子中保留 $b$：题目要求只用 $a$、$\theta$ 和 $\phi$ 表示，所以必须用 $(*)$ 消去 $b^2$。
- 移项出错，把切线斜率写成 $-b/(a\sin\theta)$；应检查在第一象限的点处斜率是否为正。
- 除以 $\sin\phi-\sin\theta$ 时没有说明它为何不等于零。

## STEP 之外

第 (ii) 部分中的圆是双曲线（hyperbola）的**准圆**（director circle）：即从该点引出的两条切线互相垂直的点的轨迹（locus）。对于椭圆（ellipse），准圆是 $x^2+y^2=a^2+b^2$；而对于 $a<b$ 的双曲线，则根本没有实的准圆，这就是题目坚持 $a>b$ 的原因。第 (i) 部分的平分性质让我们初步窥见圆锥曲线（conics）的射影几何（projective geometry）与仿射几何（affine geometry）；在这些理论中，切线、渐近线以及极点与极线（poles and polars）都以统一的方式处理。若想找完全同类风格的题目，可以试读 Siklos 的 *Advanced Problems in Mathematics*（《数学进阶题》）。

详细解答

# STEP III 2018，第 5 题

纯数学 · 算术几何平均不等式

**STEP III 2018，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/18-s3-q5

## 关键思路

这是用归纳法（induction）证明 AM–GM 不等式（算术－几何平均不等式，AM–GM inequality），但有一个巧妙之处：我们不直接证明 $A_n \geq G_n$，而是证明每加入一个新的数，加权差距 $n(A_n - G_n)$ 都不会减少。当 $n = 1$ 时差距为零，因此它一直保持非负。$\lambda_k$ 的选取使 $G_{k+1}$ 成为 $G_k$ 的一个简单倍数，从而把整个比较化为一个单变量的多项式不等式（polynomial inequality）。

## 解答

### 第 (i) 部分

有两个事实把第 $k$ 步和第 $k+1$ 步联系起来。就和而言，

$$(k+1)A_{k+1} = a_1 + \cdots + a_k + a_{k+1} = kA_k + a_{k+1}.$$

就积而言，$G_k^k = a_1 a_2 \cdots a_k$，而由 $\lambda_k$ 的定义得 $a_{k+1} = G_k \lambda_k^{k+1}$。所以

$$G_{k+1}^{k+1} = G_k^k \, a_{k+1} = G_k^k \cdot G_k \lambda_k^{k+1} = (G_k \lambda_k)^{k+1}, \qquad \text{所以} \qquad G_{k+1} = G_k \lambda_k,$$

这里取正的 $(k+1)$ 次方根，因为 $G_k > 0$ 且 $\lambda_k > 0$。这正是 $\lambda_k$ 要这样定义的原因。

现在比较两边：

$$\begin{aligned}(k+1)(A_{k+1} - G_{k+1}) - k(A_k - G_k) &= kA_k + a_{k+1} - (k+1)G_{k+1} - kA_k + kG_k \\ &= G_k \lambda_k^{k+1} - (k+1)G_k \lambda_k + kG_k \\ &= G_k\left(\lambda_k^{k+1} - (k+1)\lambda_k + k\right).\end{aligned}$$

由于 $G_k > 0$，左边 $\geq 0$ 当且仅当括号内的式子 $\geq 0$。这就是所要求的等价关系。

**第 (ii) 部分**

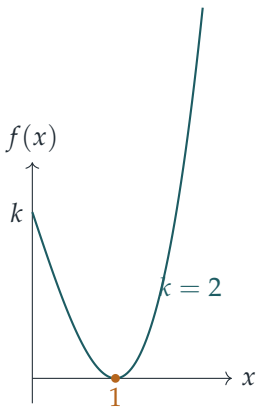
求导数（derivative），

$$f'(x) = (k+1)x^k - (k+1) = (k+1)(x^k - 1).$$

当 $0 < x < 1$ 时有 $x^k < 1$，故 $f'(x) < 0$；当 $x > 1$ 时有 $x^k > 1$，故 $f'(x) > 0$。因此 $f$ 在 $(0, 1]$ 上严格递减（strictly decreasing），在 $[1, \infty)$ 上严格递增，它在 $x > 0$ 上的最小值是

$$f(1) = 1 - (k+1) + k = 0.$$

所以对所有 $x > 0$ 都有 $f(x) \geq 0$。此外，若 $x \neq 1$，则由 1 两侧的严格单调性得 $f(x) > f(1) = 0$。所以 $f(x) = 0$ 仅当 $x = 1$ 时成立，而 $f(1) = 0$ 给出逆命题。




（当 $k = 2$ 时，$f(x) = x^3 - 3x + 2 = (x-1)^2(x+2)$；在 $x = 1$ 的二重根（double root）处，图像刚好与轴相切。）

**第 (iii) 部分**

对每个正整数 $n$，记 $d_n = n(A_n - G_n)$。

**(a)** 由第 (ii) 部分，取 $x = \lambda_k > 0$，得 $\lambda_k^{k+1} - (k+1)\lambda_k + k \geq 0$，所以由第 (i) 部分

$$d_{k+1} \geq d_k \qquad \text{对每个} k \geq 1.$$

又 $A_1 = a_1 = G_1$，所以 $d_1 = 0$。因此

$$0 = d_1 \leq d_2 \leq d_3 \leq \cdots \leq d_n,$$

所以 $d_n \geq 0$，两边除以 $n > 0$，便得对所有 $n$ 都有 $A_n \geq G_n$。

**(b)**

**想一想** 若 $A_n = G_n$，这对整条不等式链中较早的各步（而不只是最后一步）说明了甚么?

若 $A_n = G_n$，则 $d_n = 0$，不等式链变成 $0 = d_1 \le d_2 \le \cdots \le d_n = 0$。一条以 0 开始并以 0 结束的非递减链必然是常数，所以对每个 $k = 1, \ldots, n-1$ 都有 $d_{k+1} = d_k$。由第 (i) 部分的恒等式，这表示 $G_k f(\lambda_k) = 0$，而因为 $G_k > 0$，所以 $f(\lambda_k) = 0$。由第 (ii) 部分得 $\lambda_k = 1$，即对每个 $k = 1, \ldots, n-1$ 都有 $a_{k+1} = G_k$。

现在对 $k$ 用归纳法论证。当 $k = 1$ 时：$a_2 = G_1 = a_1$。若 $a_1 = \cdots = a_k = a$，则 $G_k = (a^k)^{1/k} = a$，所以 $a_{k+1} = G_k = a$。因此 $a_1 = a_2 = \cdots = a_n$，得证。

## 验算

取 $a_1, a_2, a_3 = 1, 4, 16$。则 $A_2 = 2.5, G_2 = 2, A_3 = 7, G_3 = 4$，所以 $d_1, d_2, d_3 = 0, 1, 9$：正如所料是递增的。这里 $\lambda_2 = (16/2)^{1/3} = 2$，而 $G_2 f(\lambda_2) = 2(8 - 6 + 2) = 8 = d_3 - d_2$，这印证了第 (i) 部分的恒等式。用电脑在数以千计的随机正数数列上检验该恒等式及 $d_{k+1} \ge d_k$，并没有发现任何例外。

## 常见错误

- 在第 (i) 部分只证明了一个方向。把差写成 $G_k f(\lambda_k)$ 就能同时得到两个方向，因为 $G_k > 0$。
- 在 (iii)(b) 中只推出 $\lambda_{n-1} = 1$。你需要整条链的每一步都取等号，而这是由链从 $d_1 = 0$ 开始得出的。
- 在应用第 (ii) 部分之前忘记说明为何 $\lambda_k > 0$，而第 (ii) 部分只涵盖 $x > 0$ 的情况。

## STEP 之外

不等式 $x^{k+1} - (k+1)x + k \ge 0$ 是伯努利不等式（Bernoulli's inequality）的特例，而 AM–GM 本身则是琴生不等式（Jensen's inequality）应用于凹函数（concave function）$\ln x$ 的特例。这些凸性不等式是大学一年级分析学的基本工具。尝试 AM–GM 的其他几种证明，包括柯西（Cauchy）的“前进－后退”归纳法，是很好的练习；Bartle 和 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）涵盖了相关的分析学背景。

详细解答

# STEP III 2018，第 6 题

纯数学 · 复数与几何

**STEP III 2018，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/18-s3-q6

## 关键思路

一个复数（complex number）是实数，当且仅当它等于自己的共轭（conjugate），因此三点共线（collinear）的条件可以化为一条联系 $z$ 与 $z^*$ 的方程。在单位圆（unit circle）上有 $a^* = 1/a$，借此可以消去圆上各点的共轭，于是弦（chord）上任意一点 $q$ 都满足一条线性方程。之后每个结果，都是写出两条这样的弦方程，再将它们相加或相减而得。

## 解答

### 第 (i) 部分

由于 $A$、$Q$、$C$ 共线且 $A \neq C$，点 $Q$ 等于 $A$ 加上由 $A$ 到 $C$ 的向量的某个实数倍：对某个实数 $t$，有 $q - a = t(c - a)$。因为 $c \neq a$，可以相除，所以

$$\frac{q-a}{c-a} = t \quad \text{是实数。}$$

实数等于其共轭，故

$$\frac{q-a}{c-a} = \frac{q^*-a^*}{c^*-a^*},$$

其中 $c^* - a^* \neq 0$，因为 $c \neq a$。交叉相乘得 $(c-a)(q^*-a^*) = (c^*-a^*)(q-a)$，即所求。

现设 $aa^* = cc^* = 1$，于是 $a^* = 1/a$，$c^* = 1/c$。那么

$$c^* - a^* = \frac{1}{c} - \frac{1}{a} = \frac{a-c}{ac} = -\frac{c-a}{ac},$$

而上一条方程变成

$$(c-a)\left(q^* - \frac{1}{a}\right) = -\frac{c-a}{ac}(q-a).$$

两边除以 $c-a\neq 0$，再乘以 $-ac$，得

$$-ac\,q^*+c=q-a,\qquad \text{即}\qquad q+ac\,q^*=a+c.$$

这就是弦 $AC$ 的方程：经过 $A$ 和 $C$ 的直线上每一点 $q$ 都满足它。

**第 (ii) 部分**

由于 $A$、$B$、$C$、$D$ 按逆时针次序排列，$AC$ 与 $BD$ 是圆内接（inscribed）凸四边形（convex quadrilateral）的两条对角线（diagonal），因此它们相交于圆内部的一点 $Q$。特别地，$Q$ 与这四点都不同，所以第 (i) 部分适用于直线 $AC$ 和 $BD$：

$$q+ac\,q^*=a+c,$$
$$q+bd\,q^*=b+d.$$

第一式减去第二式，消去 $q$：

$$(ac-bd)\,q^*=(a+c)-(b+d). \tag{1}$$

**想一想** 我们已得到关于 $q^*$ 的方程；怎样才能得到一条关于 $q$ 的方程与之相加？

对 (1) 取共轭，并再次利用 $a^*=1/a$ 等关系。左边变成

$$\Big(\frac{1}{ac}-\frac{1}{bd}\Big)q=\frac{bd-ac}{abcd}\,q,$$

右边变成

$$\frac{1}{a}+\frac{1}{c}-\frac{1}{b}-\frac{1}{d}=\frac{a+c}{ac}-\frac{b+d}{bd}=\frac{bd(a+c)-ac(b+d)}{abcd}.$$

两边乘以 $-abcd$，得

$$(ac-bd)\,q=ac(b+d)-bd(a+c). \tag{2}$$

将 (1) 与 (2) 相加，

$$(ac-bd)(q+q^*)=a+c-b-d+abc+acd-abd-bcd.$$

把右边各项分组，

$$(a-b)+(acd-bcd)+(c-d)+(abc-abd)=(a-b)(1+cd)+(c-d)(1+ab),$$

即得所求结果。

### 第 (iii) 部分

直线 $AB$ 与 $CD$ 是内接四边形的一对对边，所以它们的交点 $P$ 在圆外。因此 $P$ 与 $A$、$B$ 不同，也与 $C$、$D$ 不同。对共线的三点 $A$、$P$、$B$ 应用第 (i) 部分，得 $p + ab\,p^* = a + b$。由于 $p$ 是实数，$p^* = p$，所以

$$p(1 + ab) = a + b.$$

同理，由 $C$、$P$、$D$ 得

$$p(1 + cd) = c + d.$$

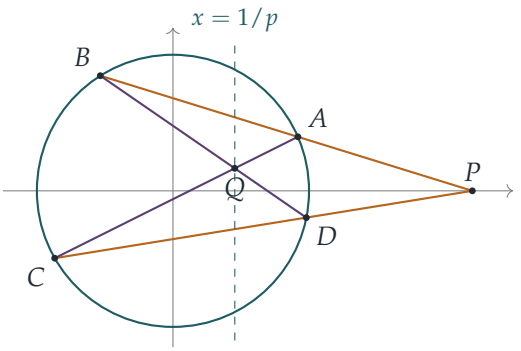
现将第 (ii) 部分的最终结果乘以 $p$，并代入这两条方程：

$$\begin{aligned} p\,(ac - bd)(q + q^*) &= (a - b)\,p(1 + cd) + (c - d)\,p(1 + ab) \\ &= (a - b)(c + d) + (c - d)(a + b) \\ &= (ac + ad - bc - bd) + (ac + bc - ad - bd) \\ &= 2(ac - bd). \end{aligned}$$

由于 $ac - bd \neq 0$，可以两边除以它，得 $p(q + q^*) = 2$，即所求。

事实上，这里 $ac - bd$ 不可能为零。若 $ac = bd$，则 (1) 迫使 $a + c = b + d$ 亦成立，于是 $z^2 - (a+c)z + ac$ 与 $z^2 - (b+d)z + bd$ 是同一个二次式（quadratic）。那么 $\{a, c\} = \{b, d\}$，与各点互不相同矛盾。

由于 $q + q^* = 2\,\mathrm{Re}\,q$，这个结果说明 $\mathrm{Re}\,q = 1/p$。无论经过 $P$ 的两条直线怎样选取，对角线的交点 $Q$ 都落在固定的铅垂线 $x = 1/p$ 上。




## 验算

- 对称情况：取 $A$、$B$ 位于角度 $20°$、$70°$，$C$、$D$ 位于 $-70°$、$-20°$。由关于实轴的反射对称，$P$ 和 $Q$ 都是实数，而结果表明 $pq = 1$。设 $a = \mathrm{e}^{\mathrm{i}\alpha}$，$b = \mathrm{e}^{\mathrm{i}\beta}$，由公式 $p = (a+b)/(1+ab)$ 得 $p = \cos\frac{\alpha-\beta}{2} / \cos\frac{\alpha+\beta}{2} = \cos 25° / \cos 45° \approx 1.2817$。直接求 $AC$ 与实轴的交点作为 $q$：$q \approx 0.7802$，而 $pq \approx 1.0000$。

- 图中取 $p = 2.2$。对角线相交于 $x \approx 0.4545 = 1/2.2$，与预测相符。三条恒等式亦已用电脑在数千个随机构形上验证。

## 常见错误

- 忘记 $ac - bd$ 取共轭后得到的是 $1/(ac) - 1/(bd)$，而不是 $ac - bd$；必须小心消去因子 $abcd$。
- 除以 $c - a$ 或 $ac - bd$ 时没有说明它为何不为零。

## STEP 之外

这里 $Q$ 位于 $P$ 关于该圆的极线（polar）上：对于单位圆外、与圆心距离为 $p$ 的点 $P$，其极线是垂直于 $OP$、与 $O$ 距离为 $1/p$ 的直线。这是射影几何（projective geometry）中布洛卡定理（Brocard's theorem）的一部分，而背后的映射 $P \mapsto 1/p^*$ 正是关于该圆的反演（inversion），它是复分析（complex analysis）和几何中的重要工具。Needham 的 *Visual Complex Analysis*（《可视化复分析》）对反演及阿冈图（Argand diagram）的几何有精彩的阐述。

详细解答

# STEP III 2018，第 7 题

纯数学 · 棣美弗定理与级数

> **STEP III 2018，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/18-s3-q7

本题在 *Beyond STEP* 第 2.4 节 $\pi^2/6$ 中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch02/basel.html

## 关键思路

式子 $\cot\theta + \mathrm{i}$ 其实就是乔装了的 $(\cos\theta + \mathrm{i}\sin\theta)/\sin\theta$，因此棣美弗定理（De Moivre's theorem）可把它的幂化为倍角。改用二项式定理（binomial theorem）展开同样的幂，则得到一个以 $\cot^2\theta$ 为变量的多项式（polynomial），而它的根（roots）可以从 $\sin(2n+1)\theta$ 等于零的位置直接读出。接着，韦达定理（Vieta's formula）不费吹灰之力便给出各根之和；再把 $1/\theta^2$ 夹在 $\cot^2\theta$ 与 $1+\cot^2\theta$ 之间，便能把巴塞尔（Basel）级数（series）困在两个趋向同一极限的界之间。

## 解答

### 第 (i) 部分

由于 $\sin\theta \neq 0$，

$$\cot\theta \pm \mathrm{i} = \frac{\cos\theta \pm \mathrm{i}\sin\theta}{\sin\theta}.$$

根据棣美弗定理，$(\cos\theta \pm \mathrm{i}\sin\theta)^{2n+1} = \cos(2n+1)\theta \pm \mathrm{i}\sin(2n+1)\theta$，所以

$$\frac{(\cot\theta+\mathrm{i})^{2n+1} - (\cot\theta-\mathrm{i})^{2n+1}}{2\mathrm{i}} = \frac{2\mathrm{i}\sin(2n+1)\theta}{2\mathrm{i}\,\sin^{2n+1}\theta} = \frac{\sin(2n+1)\theta}{\sin^{2n+1}\theta},$$

这正是所求的恒等式（identity）。

现在改用二项式定理展开左方。记 $c=\cot\theta$ 及 $N=2n+1$,

$$(c+\mathrm{i})^N-(c-\mathrm{i})^N=\sum_{k=0}^{N}\binom{N}{k}c^{N-k}\,\mathrm{i}^k\big(1-(-1)^k\big).$$

当 $k$ 为偶数时, 括号内的值是 0; 当 $k$ 为奇数时, 则是 2。令 $k=2j+1$, 其中 $j=0,1,\ldots,n$; 那么 $\mathrm{i}^k=\mathrm{i}(-1)^j$，而 $c^{N-k}=c^{2(n-j)}$。除以 2i，并记 $x=\cot^2\theta$，得

$$\frac{\sin(2n+1)\theta}{\sin^{2n+1}\theta}=\sum_{j=0}^{n}(-1)^j\binom{2n+1}{2j+1}x^{n-j}=\binom{2n+1}{1}x^n-\binom{2n+1}{3}x^{n-1}+\cdots+(-1)^n.$$

把这个多项式记作 $P(x)$。它的次数是 $n$，首项系数（leading coefficient）为 $2n+1\neq 0$。取 $\theta_m=\dfrac{m\pi}{2n+1}$, 其中 $m=1,\ldots,n$。那么 $0<\theta_m<\frac{1}{2}\pi$, 故 $\sin\theta_m\neq 0$, 且 $\sin(2n+1)\theta_m=\sin m\pi=0$。因此

$$P\big(\cot^2\theta_m\big)=0,\qquad m=1,2,\ldots,n.$$

**想一想** 我们怎样知道这些就是全部的解，而不只是其中一部分?

在 $(0,\frac{1}{2}\pi)$ 上，$\cot\theta$ 为正而且严格递减（strictly decreasing），所以 $\cot^2\theta$ 亦严格递减。角 $\theta_1<\theta_2<\cdots<\theta_n$ 全都位于这个区间内，故这 $n$ 个数 $\cot^2\theta_m$ 互不相同。一个 $n$ 次多项式最多只有 $n$ 个根，所以它们恰好就是 $P(x)=0$ 的所有解。

### 第 (ii) 部分

$a_nx^n+a_{n-1}x^{n-1}+\cdots$ 的各根之和是 $-a_{n-1}/a_n$。这里 $a_n=\binom{2n+1}{1}$, $a_{n-1}=-\binom{2n+1}{3}$, 所以

$$\sum_{m=1}^{n}\cot^2\left(\frac{m\pi}{2n+1}\right)=\frac{\binom{2n+1}{3}}{\binom{2n+1}{1}}=\frac{(2n+1)(2n)(2n-1)}{6\,(2n+1)}=\frac{n(2n-1)}{3},$$

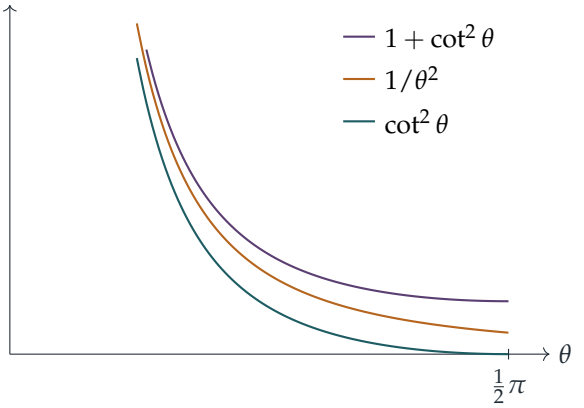
正如所求。

### 第 (iii) 部分

题目给出，当 $0<\theta<\frac{1}{2}\pi$ 时，$0<\sin\theta<\theta<\tan\theta$。三者皆为正数，所以平方后大小次序不变，而取倒数（reciprocal）则令次序倒转:

$$\frac{1}{\tan^2\theta}<\frac{1}{\theta^2}<\frac{1}{\sin^2\theta},\qquad \text{即}\qquad \cot^2\theta<\frac{1}{\theta^2}<1+\cot^2\theta,$$

这里用了 $\operatorname{cosec}^2\theta=1+\cot^2\theta$。




令 $\theta=\theta_m=\dfrac{m\pi}{2n+1}$，使得 $\dfrac{1}{\theta_m^2}=\dfrac{(2n+1)^2}{\pi^2}\cdot\dfrac{1}{m^2}$，然后把 $m=1,\dots,n$ 的不等式相加。记 $S_n=\sum_{m=1}^{n}\frac{1}{m^2}$，并利用第 (ii) 部分，得

$$\frac{n(2n-1)}{3}<\frac{(2n+1)^2}{\pi^2}S_n<n+\frac{n(2n-1)}{3}=\frac{2n(n+1)}{3}.$$

全式乘以 $\pi^2/(2n+1)^2>0$，得

$$\frac{\pi^2}{3}\cdot\frac{n(2n-1)}{(2n+1)^2}<S_n<\frac{\pi^2}{3}\cdot\frac{2n(n+1)}{(2n+1)^2}.$$

把分子分母同除以 $n^2$，可见当 $n\to\infty$ 时两个分数都趋向 $\frac{2}{4}=\frac{1}{2}$，所以两个界都趋向 $\frac{\pi^2}{6}$。根据夹逼定理（squeeze / sandwich theorem），$S_n\to\frac{\pi^2}{6}$，即

$$\sum_{m=1}^{\infty}\frac{1}{m^2}=\frac{\pi^2}{6}.$$

## 验算

- 当 $n=1$ 时，多项式是 $3x-1$，其根为 $\frac{1}{3}=\cot^2\frac{\pi}{3}$，而第 (ii) 部分的公式亦给出 $1\times1/3=\frac{1}{3}$。当 $n=2$ 时，$\cot^2\frac{\pi}{5}+\cot^2\frac{2\pi}{5}$ 应等于 2；数值上 $1.894+0.106=2.000$。
- 取 $n=50$，第 (iii) 部分的界为 $1.5964<S_{50}=1.6251<1.6448$，而 $\pi^2/6\approx1.6449$。两个界互相逼近，这是理所当然的，因为两者之差为 $\pi^2n/(2n+1)^2\approx\pi^2/(4n)$。

## 常见错误

- 证明了每个 $\cot^2\theta_m$ 都是根，却忘记论证它们互不相同——正是这一点才令它们成为全部的解。

- 弄错韦达定理中的正负号：$x^{n-1}$ 的系数是 $-\binom{2n+1}{3}$，所以各根之和是 $+\binom{2n+1}{3}/\binom{2n+1}{1}$。
- 把不等式倒转时没有指出所有量皆为正数，而这正是容许平方及取倒数的前提。

## STEP 之外

欧拉（Euler）在 1734 年首先求得 $\sum 1/m^2 = \pi^2/6$，方法是把 $\sin x/x$ 当作一个“无穷次多项式”来比较系数——与这里所用的韦达思路相同，只是欠缺了夹逼法所提供的严谨性。到了大学，你会通过傅里叶级数（Fourier series）（帕塞瓦尔恒等式（Parseval's identity））再次遇见这个和，亦会以 $\zeta(2)$ 的身份遇见它，即黎曼（Riemann）$\zeta$ 函数（zeta function）的第一个值；所有 $\zeta(2k)$ 都是 $\pi^{2k}$ 的有理数倍，但人们对 $\zeta(3)$ 所知甚少。若想了解极限与夹逼定理背后的分析学，可以读 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

# STEP III 2018，第 8 题

纯数学 • 小数部分的积分

**STEP III 2018，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/18-s3-q8

## 关键思路

被积函数（integrand）中最麻烦的是 $x^{-1}$，所以作代换（substitution）$y = 1/x$：这把 $[0,1]$ 变成 $[1,\infty)$，而后者很自然地可以分拆成一个个单位区间。如果 f 的周期（period）是 1，每个单位区间都可以平移回 $[0,1]$，剩下的有理因子就组成一个伸缩和（telescoping sum）。这样一来，第 (ii) 部分只需看出 $\{x\}$ 和 $\{2x\}$ 的周期都是 1 便可。

## 解答

### 第 (i) 部分

设 $x = 1/y$，于是 $\mathrm{d}x = -\frac{1}{y^2}\,\mathrm{d}y$；当 $x$ 由 0 变到 1 时，$y$ 由 $\infty$ 递减至 1。此外 $1 + x = \frac{y+1}{y}$，所以

$$I = \int_{\infty}^{1} \frac{\mathrm{f}(y)}{(y+1)/y}\left(-\frac{1}{y^2}\right)\mathrm{d}y = \int_{1}^{\infty} \frac{\mathrm{f}(y)}{y(1+y)}\,\mathrm{d}y.$$

把 $[1,\infty)$ 分拆成区间 $[n, n+1]$（其中 $n = 1, 2, 3, \ldots$），便得

$$I = \sum_{n=1}^{\infty} \int_{n}^{n+1} \frac{\mathrm{f}(y)}{y(1+y)}\,\mathrm{d}y,$$

正是所求。

现在假设对所有 $x$ 都有 $\mathrm{f}(x) = \mathrm{f}(x+1)$，因此对每个整数 $n$ 都有 $\mathrm{f}(x+n) = \mathrm{f}(x)$。在第 $n$ 个积分中设 $y = x + n$，于是 $x$ 的范围是 $[0,1]$：

$$\int_{n}^{n+1} \frac{\mathrm{f}(y)}{y(1+y)}\,\mathrm{d}y = \int_{0}^{1} \frac{\mathrm{f}(x)}{(x+n)(x+n+1)}\,\mathrm{d}x.$$

**想一想** 怎样把无限多个这样的积分加起来?

每个积分都有相同的 $\mathrm{f}(x)$ 和积分范围，所以（按题目指示，不理会收敛（convergence）问题）我们可以交换求和与积分的次序，先把有理因子加起来。利用部分分式（partial fractions），

$$\frac{1}{(x+n)(x+n+1)} = \frac{1}{x+n} - \frac{1}{x+n+1},$$

所以这个和会逐项相消：对固定的 $x \in [0,1]$，

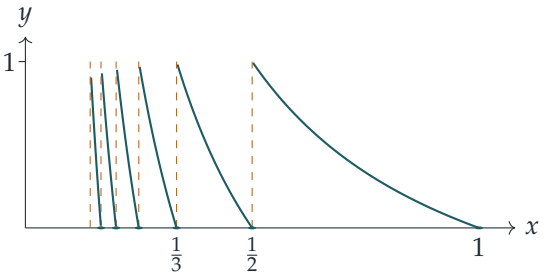
$$\sum_{n=1}^{N}\left(\frac{1}{x+n} - \frac{1}{x+n+1}\right) = \frac{1}{x+1} - \frac{1}{x+N+1} \longrightarrow \frac{1}{1+x} \quad \text{当} N \to \infty.$$

因此

$$I = \int_0^1 \mathrm{f}(x) \sum_{n=1}^{\infty} \frac{1}{(x+n)(x+n+1)}\,\mathrm{d}x = \int_0^1 \frac{\mathrm{f}(x)}{1+x}\,\mathrm{d}x,$$

这就是我们想要的结果。

**第 (ii) 部分**




上面 $\{x^{-1}\}$ 的图像说明了为何直接计算是行不通的：在每个区间 $\left(\frac{1}{n+1}, \frac{1}{n}\right]$ 上，它等于 $\frac{1}{x} - n$，而这样的片段有无限多段，越来越密地挤向 0。第 (i) 部分让我们完全避开了这些麻烦。

*第一个积分*。取 $\mathrm{f}(x) = \{x\}$。把 $x$ 加 1，$\lfloor x \rfloor$ 也会加 1，所以 $\{x+1\} = \{x\}$，即 f 的周期是 1。在 $[0,1)$ 上有 $\{x\} = x$（单一点 $x = 1$ 的函数值不影响积分）。所以由第 (i) 部分，

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

*第二个积分*。取 $\mathrm{f}(x) = \{2x\}$，那么正如所需，$\mathrm{f}(x^{-1}) = \{2x^{-1}\}$。

**想一想** 函数 $\{2x\}$ 的周期是 $\frac{1}{2}$。我们仍然可以用第 (i) 部分吗?

可以：任何周期为 $\frac{1}{2}$ 的函数，周期也是 1，因为 $\{2(x+1)\} = \{2x+2\} = \{2x\}$。所以第 (i) 部分适用。在 $[0, \frac{1}{2})$ 上，$2x \in [0,1)$ 而 $\{2x\} = 2x$；在 $[\frac{1}{2}, 1)$ 上，$2x \in [1,2)$ 而 $\{2x\} = 2x-1$。因此

$$\int_0^1 \frac{\{2x^{-1}\}}{1+x}\,\mathrm{d}x = \int_0^{1/2} \frac{2x}{1+x}\,\mathrm{d}x + \int_{1/2}^1 \frac{2x-1}{1+x}\,\mathrm{d}x.$$

把 $\frac{2x}{1+x} = 2 - \frac{2}{1+x}$ 及 $\frac{2x-1}{1+x} = 2 - \frac{3}{1+x}$ 代入，

$$\int_0^{1/2} \frac{2x}{1+x}\,\mathrm{d}x = 1 - 2\ln\tfrac{3}{2},$$
$$\int_{1/2}^1 \frac{2x-1}{1+x}\,\mathrm{d}x = 1 - 3\ln\frac{2}{3/2} = 1 - 3\ln\tfrac{4}{3}.$$

两者相加，再合并对数（logarithm）项，

$$\begin{aligned}\int_0^1 \frac{\{2x^{-1}\}}{1+x}\,\mathrm{d}x &= 2 - 2\ln 3 + 2\ln 2 - 6\ln 2 + 3\ln 3 \\ &= 2 + \ln 3 - 4\ln 2 = 2 + \ln\tfrac{3}{16}.\end{aligned}$$

## 验算

- 由于 $0 \le \{\cdot\} < 1$，两个答案都应该介乎 0 与 $\int_0^1 \frac{\mathrm{d}x}{1+x} = \ln 2 \approx 0.693$ 之间。我们得到 $1 - \ln 2 \approx 0.3069$ 和 $2 + \ln\frac{3}{16} \approx 0.3260$，都很合理：小数部分（fractional part）的平均值约为 $\frac{1}{2}$，所以答案应大约是 0.693 的一半。
- 用电脑把在片段 $\left(\frac{1}{n+1}, \frac{1}{n}\right]$ 上（第二个积分则用这些片段的两半）的积分逐一加起来，取数以千计的片段，得到 0.306853 和 0.326024，与两个答案吻合。

## 常见错误

- 代换 $y = 1/x$ 后把积分上下限弄反，又漏掉负号；其实 $\mathrm{d}x$ 带来的负号与上下限对调正好互相抵消。
- 以为在整个 $[0,1]$ 上都有 $\{2x\} = 2x$：其实它在 $x = \frac{1}{2}$ 处下跌 1。
- 以为第 (i) 部分要求周期恰好是 1；其实任何周期为 $\frac{1}{k}$（$k$ 为正整数）的函数，周期也是 1。

## STEP 之外

我们在这里不加说明就交换了无穷级数与积分的次序，而这正是在大学里需要严格论证的那类步骤；单调收敛定理（monotone convergence theorem）和控制收敛定理（dom-

inated convergence theorem）等定理说明了何时可以安全地这样做（在这里，当 $f \geq 0$ 时每一项都非负，属于容易的情况）。同样的代换思路可以得出一些关于小数部分的著名积分，例如 $\int_0^1 \{1/x\}\,\mathrm{d}x = 1 - \gamma$，其中 $\gamma$ 是欧拉（Euler）常数。若想严谨地学习极限、级数与积分，可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

# STEP III 2018，第 9 题

力学 · 碰撞与递推关系

**STEP III 2018，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/18-s3-q9

## 关键思路

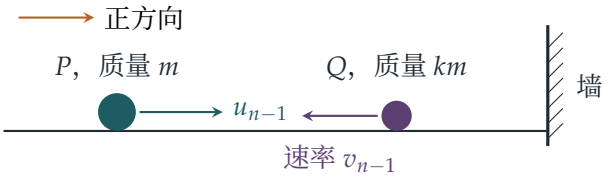
$P$ 与 $Q$ 每次碰撞（collision）都给出两条线性方程：一条来自动量守恒（conservation of momentum），一条来自牛顿碰撞定律（Newton's law of restitution），把碰撞前后的速度联系起来。墙壁只是把 $Q$ 的速度反向，因此某次碰撞前的状态完全由上一次碰撞后的状态决定。这样便得到一对关于 $u_n$ 和 $v_n$ 的一阶递推关系（first-order recurrence），消去 $v$ 后就剩下一条关于 $u_n$ 的二阶递推关系（second-order recurrence）。第 (ii) 部分就是常规工作：把线性递推关系的解配合其首两项。

## 解答

我们把 $P$ 和 $Q$ 视为在光滑（smooth）地面上沿同一直线运动的质点（particle），因此两次碰撞之间没有水平力作用，每个质点都保持等速。全程取朝向墙壁的方向为正方向。

**两次碰撞之间发生甚么。** 第 $n$ 次碰撞后，$Q$ 的速度为 $v_n$（朝向墙壁，因为它接着会撞墙）。与墙的碰撞是完全弹性（perfectly elastic）的，所以 $Q$ 以速度 $-v_n$ 反弹回来，而 $P$ 的速度仍是 $u_n$。因此在第 $n+1$ 次碰撞前一刻，速度分别为 $u_n$（$P$ 的）和 $-v_n$（$Q$ 的）。

只要正确理解 $v_0$，这对第一次碰撞同样成立：碰撞前，$Q$ 以速率 $v_0$ 离开墙壁运动，所以其速度为 $-v_0$，正好是上述模式在 $n=0$ 时的情况。




**第 (i) 部分**

对第 $n$ 次碰撞（$n \geq 1$），由动量守恒（两边除以 $m$）得

$$u_{n-1} - kv_{n-1} = u_n + kv_n. \tag{1}$$

接近速率（speed of approach）为 $u_{n-1} - (-v_{n-1}) = u_{n-1} + v_{n-1}$。碰撞后，较接近墙壁的 $Q$ 朝墙运动得比 $P$ 快，所以分离速率（speed of separation）为 $v_n - u_n$。由牛顿碰撞定律得

$$v_n - u_n = e\,(u_{n-1} + v_{n-1}). \tag{2}$$

把 (2) 中的 $v_n$ 代入 (1)：

$$u_{n-1} - kv_{n-1} = (1+k)u_n + ke\,(u_{n-1} + v_{n-1}),$$

所以

$$k(1+e)\,v_{n-1} = (1-ke)\,u_{n-1} - (1+k)\,u_n \qquad (n \geq 1). \tag{3}$$

这条式把每个 $v$ 以两个相邻的 $u$ 表示，正是消去 $v$ 所需要的。

? **想一想** 方程 (3) 消去了 $v_{n-1}$，但 (2) 还含有 $v_n$。该怎样处理?

方程 (3) 对所有 $n \geq 1$ 成立，所以我们也可以用 $n+1$ 代替 $n$：

$$k(1+e)\,v_n = (1-ke)\,u_n - (1+k)\,u_{n+1}.$$

现在把 (2) 乘以 $k(1+e)$，并利用这两条式替换 $k(1+e)v_n$ 和 $k(1+e)v_{n-1}$：

$$(1-ke)u_n - (1+k)u_{n+1} - k(1+e)u_n = ke(1+e)\,u_{n-1} + e\big[(1-ke)u_{n-1} - (1+k)u_n\big].$$

整理各项。右边 $u_{n-1}$ 的系数为 $ke + ke^2 + e - ke^2 = e(1+k)$。把 $u_n$ 的项移到左边，其系数为

$$1 - ke - k - ke + e + ke = 1 - k + e - ke = (1+e)(1-k).$$

因此

$$(1+k)u_{n+1} - (1-k)(1+e)u_n + e(1+k)u_{n-1} = 0 \qquad (n \geq 1),$$

再以 $n-1$ 代替 $n$，便得到对 $n \geq 2$ 成立的 $(*)$，即所求。

**第 (ii) 部分**

当 $e=\frac{1}{2}$ 及 $k=\frac{1}{34}$ 时，把 $(*)$ 乘以 68 得

$$70u_n-99u_{n-1}+35u_{n-2}=0,$$

其辅助方程（auxiliary equation）$70t^2-99t+35=(10t-7)(7t-5)=0$ 的根为 $\frac{7}{10}$ 和 $\frac{5}{7}$，与题目给出的形式吻合。

要确定 $A$ 和 $B$ 需要两个条件，即首两项。我们已知 $u_0$。至于 $u_1$，在 $(3)$ 中取 $n=1$：

$$(1+k)u_1=(1-ke)u_0-k(1+e)v_0,$$

$$\text{所以}\quad \tfrac{35}{34}u_1=\tfrac{67}{68}u_0-\tfrac{3}{68}v_0,\quad \text{得}\quad u_1=\frac{67u_0-3v_0}{70}.$$

把 $n=0$ 和 $n=1$ 代入给定的形式，

$$A+B=u_0,\qquad \tfrac{7}{10}A+\tfrac{5}{7}B=\frac{67u_0-3v_0}{70}.$$

把第二条方程乘以 70 得 $49A+50B=67u_0-3v_0$。减去第一条方程的 49 倍，

$$B=18u_0-3v_0,\qquad A=u_0-B=3v_0-17u_0.$$

**$n$ 很大时 $u_n$ 的正负号**。由于 $\frac{5}{7}>\frac{7}{10}$（因为 $50>49$），$B$ 项占主导地位。准确地说，

$$u_n=\left(\tfrac{5}{7}\right)^n\left[B+A\left(\tfrac{49}{50}\right)^n\right],$$

而 $\left(\frac{49}{50}\right)^n\to 0$，所以括号趋于 $B$。若 $0<6u_0<v_0$，则 $B=3(6u_0-v_0)<0$，所以对所有足够大的 $n$，括号都是负数。由于 $\left(\frac{5}{7}\right)^n>0$，当 $n$ 很大时 $u_n$ 为负，即所求。（具体而言，一旦 $|A|(\frac{49}{50})^n<|B|$，括号便是负数。）

从物理角度看，$P$ 最终会被撞回、离开墙壁：$Q$ 相对于 $P$ 有足够的初始速率，能不断把它推回去，尽管 $Q$ 轻得多。

## 验算

- 取 $e=1$、$k=1$（质量相等、完全弹性）：$(*)$ 变成 $2u_n+2u_{n-2}=0$，所以 $u_n=-u_{n-2}$。事实上，质量相等的物体作弹性碰撞会互换速度，所以 $u_1=-v_0$ 及 $u_2=-v_1$；然后 $Q$ 带着 $P$ 原来的速度 $u_0$ 撞墙再折返，得 $u_2=-u_0$。
- 在 $n=0$ 时公式给出 $A+B=u_0$，在 $n=1$ 时给出

  $$\tfrac{7}{10}(3v_0-17u_0)+\tfrac{5}{7}(18u_0-3v_0)=\tfrac{1}{70}(67u_0-3v_0),$$

  正如预期。
- 取 $u_0=1$、$v_0=7$，得 $A=4$、$B=-3$。逐次计算每次碰撞，结果与公式完全吻合，而 $u_n$ 首次变为负数是在 $n=15$，此时 $u_{15}\approx -0.00029$。

## 常见错误

- 忘记 $Q$ 在碰撞前一刻的速度是 $-v_{n-1}$ 而不是 $v_{n-1}$：墙壁把它反向了。
- 在碰撞定律方程中把分离速率的正负号弄错；碰撞后 $Q$ 在正方向上位于 $P$ 前面，所以分离速率是 $v_n - u_n$。
- 未从碰撞方程求出 $u_1$ 便使用 $u_1 = \frac{7}{10}A + \frac{5}{7}B$；单靠 $(*)$ 是无法得出它的。

## STEP 之外

方程 (1) 和 (2) 合起来是一个线性映射（linear map），把状态 $(u_{n-1}, v_{n-1})$ 映到 $(u_n, v_n)$，所以整个运动就是反复乘以一个 $2 \times 2$ 矩阵（matrix）。根 $\frac{7}{10}$ 和 $\frac{5}{7}$ 是它的特征值（eigenvalue），长期行为由最大的特征值支配：这正是线性动力系统（linear dynamical systems）与差分方程（difference equations）理论的起点。想进一步了解碰撞力学，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）；至于矩阵的观点，可参阅 Strang 的 *Introduction to Linear Algebra*（线性代数导论）。

详细解答

# STEP III 2018，第 10 题

力学 · 复摆

**STEP III 2018，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/18-s3-q10

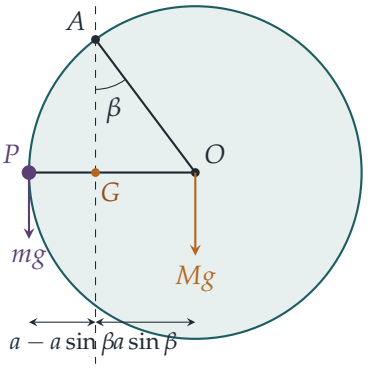
## 关键思路

圆盘与质点合起来构成单一个刚体（rigid body），绕固定轴 $L$ 转动，因此一切都由一个角度决定。在平衡时，整体的质心（centre of mass）必须位于 $A$ 的正下方，这就是关于 $A$ 的力矩（moment）平衡。至于运动，由于轴是光滑（smooth）的，该处的反作用力不做功，所以能量守恒（conservation of energy）。只要把能量写成 $\theta$ 和 $\dot{\theta}$ 的函数，对它求导便得到运动方程（equation of motion），再把 $\sin\theta$ 换成 $\theta$，便得出小幅振动（small oscillations）的结果。

## 解答

我们把圆盘模型化为均匀的刚性薄片（lamina），把质点视为固定在圆盘边缘上的点质量（point mass），并假设轴 $L$ 是光滑的。

**平衡**。 关于 $A$ 有力矩的力只有两个重力：作用于 $O$ 的 $Mg$ 和作用于 $P$ 的 $mg$。$A$ 处的反作用力通过 $A$。由于 $AO$ 与向下的铅垂线成角 $\beta$，$O$ 位于 $A$ 某一侧、水平距离为 $a\sin\beta$ 之处，不妨设在右侧。这样重力 $Mg$ 会令圆盘绕 $A$ 顺时针转动，而只有 $mg$ 能与之平衡，所以 $P$ 必须在通过 $A$ 的铅垂线的左侧。因为 $OP$ 是水平的，长度为 $a$，所以 $P$ 在 $A$ 左方的水平距离为 $a - a\sin\beta$。（这里 $a\sin\beta < a$，所以 $P$ 确实在铅垂线的左侧。）




关于 $A$ 取力矩，

$$Mg\,a\sin\beta = mg\,(a - a\sin\beta) \quad\Longrightarrow\quad \sin\beta = \frac{m}{M+m}.$$

$O$ 和 $P$ 都位于 $A$ 下方、铅垂距离为 $a\cos\beta$ 之处。所以在以 $AP$ 为斜边（hypotenuse）的直角三角形中，水平边为 $a(1-\sin\beta)$，铅垂边为 $a\cos\beta$：

$$\begin{aligned} AP^2 &= a^2(1-\sin\beta)^2 + a^2\cos^2\beta = a^2\big(2-2\sin\beta\big) \\ &= 2a^2\Big(1-\frac{m}{M+m}\Big) = \frac{2Ma^2}{M+m}, \end{aligned}$$

因此 $AP/a = \sqrt{2M/(M+m)}$，正如所求。

值得留意整体质心 $G$ 的位置。它是 $O$ 和 $P$ 的加权平均（weighted average），所以位于线段 $OP$ 上，即在 $A$ 下方深度 $a\cos\beta$ 处。在平衡时它位于 $A$ 的正下方，所以

$$AG = a\cos\beta.$$

**能量守恒**。 圆盘与质点作为一个刚体一起运动。在平衡时 $OP$ 是水平的，而 $OP$ 随刚体一起转动，所以当 $OP$ 与水平线成角 $\theta$ 时，整个刚体已从平衡位置转过了 $\theta$。因此它的角速率（angular speed）为 $|\dot\theta|$。

> ? **想一想** 质点的速率是多少?

刚体上每一点都沿以轴为圆心的圆周运动。质点与 $A$ 的距离恒为 $AP$，所以它的速率是 $AP\,|\dot\theta|$，动能（kinetic energy）为

$$\tfrac12 m\,AP^2\,\dot\theta^2 = \tfrac12 m\cdot 2a^2(1-\sin\beta)\,\dot\theta^2 = (1-\sin\beta)\,ma^2\dot\theta^2,$$

这里用了上面得到的 $AP^2 = 2a^2(1-\sin\beta)$。圆盘的动能是 $\frac12 I\dot\theta^2$。

至于势能（potential energy），可把总质量 $M+m$ 视为集中在 $G$。直线 $AG$ 在平衡时是铅垂的，它同样转过了 $\theta$，所以 $G$ 现在位于 $A$ 下方深度 $a\cos\beta\cos\theta$ 处。以平衡位置为势

能零点，

$$\text{势能} = (M+m)g\big(a\cos\beta - a\cos\beta\cos\theta\big) = (M+m)ga\cos\beta\,(1-\cos\theta).$$

轴是光滑的，而 $A$ 处的反作用力作用在一个不动的点上，所以它不做功。唯一其他的力是重力，而重力是保守力（conservative force）。所以动能加势能是常数：

$$\tfrac{1}{2}I\dot\theta^2 + (1-\sin\beta)\,ma^2\dot\theta^2 + (m+M)ga\cos\beta\,(1-\cos\theta) = \text{常数},$$

这正是所求的结果。

**小幅振动**。 对 $t$ 求导：

$$\dot\theta\Big[\big(I + 2(1-\sin\beta)ma^2\big)\ddot\theta + (M+m)ga\cos\beta\,\sin\theta\Big] = 0.$$

只要 $\dot\theta \neq 0$，方括号内的式子就必须为零；由连续性，它在每一时刻都为零。当 $\theta$ 很小时，$\sin\theta \approx \theta$，所以

$$\ddot\theta = -\omega^2\theta, \qquad \omega^2 = \frac{(M+m)ga\cos\beta}{I + 2(1-\sin\beta)ma^2},$$

这是周期（period）为 $2\pi/\omega$ 的简谐运动（simple harmonic motion）。

现在设 $m = \frac{3}{2}M$。那么 $\sin\beta = \dfrac{3/2}{5/2} = \frac{3}{5}$，又因为 $\beta$ 是锐角，$\cos\beta = \frac{4}{5}$。分子为

$$\tfrac{5}{2}M\cdot ga\cdot\tfrac{4}{5} = 2Mga,$$

而以 $I = \frac{3}{2}Ma^2$ 代入，分母为

$$\tfrac{3}{2}Ma^2 + 2\cdot\tfrac{2}{5}\cdot\tfrac{3}{2}Ma^2 = \tfrac{3}{2}Ma^2 + \tfrac{6}{5}Ma^2 = \tfrac{27}{10}Ma^2.$$

所以 $\omega^2 = \dfrac{20g}{27a}$，周期为

$$\frac{2\pi}{\omega} = 2\pi\sqrt{\frac{27a}{20g}} = 2\pi\cdot\frac{3}{2}\sqrt{\frac{3a}{5g}} = 3\pi\sqrt{\frac{3a}{5g}},$$

这里用了 $\sqrt{27/20} = \sqrt{9\cdot 3/(4\cdot 5)} = \frac{3}{2}\sqrt{3/5}$。这就是所求的周期。

## 验算

- 取 $m = 0$，得 $\sin\beta = 0$ 及 $AP/a = \sqrt{2}$。此时圆盘悬挂时 $O$ 在 $A$ 的正下方，而与 $O$ 同高的边缘点与 $A$ 的距离确实是 $\sqrt{2}\,a$。周期变为 $2\pi\sqrt{(3a/2)/g}$，正是以边缘上一点为支点的圆盘的熟悉结果。

- 动能各项相加得 $\frac{1}{2}(I + m\,AP^2)\dot{\theta}^2$，其中括号内是整个刚体关于 $L$ 的转动惯量（moment of inertia），理应如此。
- 以 $m = \frac{3}{2}M$ 及 $a = 0.5\,\mathrm{m}$ 对该刚体进行逐步的电脑模拟，得到小幅振动周期为 $1.64815\,\mathrm{s}$。公式给出 $3\pi\sqrt{1.5/49.05} \approx 1.64815\,\mathrm{s}$。模拟亦显示，即使摆幅很大，能量表达式仍保持不变。

## 常见错误

- 把 $P$ 放在铅垂线上与 $O$ 同一侧。这样两个重力会令刚体向同一方向转动，不可能平衡。
- 用 $a$ 而不是 $AP$ 作为质点与轴的距离。质点是绕 $A$ 而不是绕 $O$ 作圆周运动。
- 分别计算 $O$ 和 $P$ 的势能，结果在三角运算中迷失方向。改用与 $A$ 相距 $a\cos\beta$ 的整体质心 $G$，计算会简洁得多。

## STEP 之外

这是一个**复摆**（compound pendulum）：任何绕固定轴摆动的刚体，其振动都如同一个长度为 $I_{总}/\big((M+m)h\big)$ 的单摆，其中 $h$ 是轴到质心的距离。我们的能量方法是拉格朗日力学（Lagrangian mechanics）的单变量情况：在拉格朗日力学中，运动方程由以单一广义坐标（generalised coordinate）表示的动能和势能得出。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对刚体转动和物理摆有细致的论述。

详细解答

# STEP III 2018，第 11 题

力学 · 松弛绳与抛体

**STEP III 2018，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：

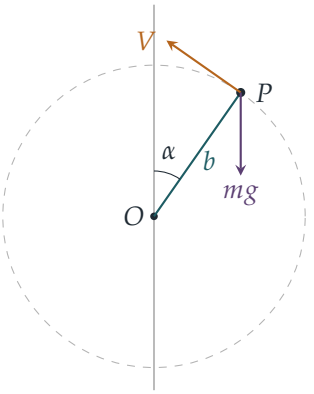
step.maths.org/questions/18-s3-q11

## 关键思路

当绳的张力（tension）降至零的一刻，绳便会变得松弛，所以该瞬间沿径向的运动方程便给出 $V$。此后质点只是一个普通的抛射体（projectile），而当质点与 $O$ 的距离回到 $b$ 时，绳便再次拉紧。若采用以 $O$ 为中心的坐标，这个条件会漂亮地化简，因为抛射体由圆上出发，并沿切线（tangent）方向运动。最后一部分则归结为：在急拉（jerk）的一刻，速度是否正好沿着绳的方向。

## 解答

我们把质点模拟为一个质量为 $m$ 的质点，系于一条轻而不可伸长（inextensible）的绳上，并忽略空气阻力。取 $O$ 为原点，$x$ 为水平方向，$y$ 为铅直向上，并假设质点沿逆时针方向摆动，使绳在向上铅直线右方的点 $P$ 变得松弛。




在 $P$ 点，绳与向上铅直线成角 $\alpha$，所以 $P = (b\sin\alpha,\ b\cos\alpha)$。该处的速度垂直于绳，方

向朝上并越过顶部：

$$\mathbf{v}_P = V(-\cos\alpha,\ \sin\alpha).$$

**绳变得松弛时的速率**

沿绳朝 $O$ 的方向分解。张力 $R$ 朝向 $O$；由于 $P$ 在 $O$ 之上，重力朝向 $O$ 的分量亦为 $mg\cos\alpha$。质点在半径为 $b$ 的圆上运动，所以

$$R + mg\cos\alpha = \frac{mV^2}{b}.$$

当 $R = 0$ 时绳变得松弛，所以

$$V^2 = gb\cos\alpha, \qquad V = \sqrt{gb\cos\alpha}.$$

**绳何时再次拉紧**

绳一旦松弛，质点便是一个抛射体。经过时间 $t$ 后，其位置为

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

当 $x^2 + y^2 = b^2$ 时，绳再次拉紧。展开得

$$\begin{aligned} x^2 + y^2 &= b^2\sin^2\alpha - 2bVt\sin\alpha\cos\alpha + V^2t^2\cos^2\alpha \\ &\quad + b^2\cos^2\alpha + V^2t^2\sin^2\alpha + \tfrac{1}{4}g^2t^4 + 2bVt\sin\alpha\cos\alpha - bgt^2\cos\alpha - gVt^3\sin\alpha \\ &= b^2 + \left(V^2 - gb\cos\alpha\right)t^2 - gVt^3\sin\alpha + \tfrac{1}{4}g^2t^4. \end{aligned}$$

含 $t$ 的项互相抵消，因为质点起初沿切线方向运动；含 $t^2$ 的项亦互相抵消，因为 $V^2 = gb\cos\alpha$。因此

$$x^2 + y^2 - b^2 = gt^3\left(\tfrac{1}{4}gt - V\sin\alpha\right).$$

当 $0 < t < 4V\sin\alpha/g$ 时，此式为负，所以质点严格位于圆内，绳保持松弛。此式在时间 $T$ 再次等于零（其后变为正），其中

$$gT = 4V\sin\alpha,$$

此即所求。

**拉紧前一刻的运动方向**

在时间 $T$，速度的分量为

$$\dot{x} = -V\cos\alpha, \qquad \dot{y} = V\sin\alpha - gT = V\sin\alpha - 4V\sin\alpha = -3V\sin\alpha.$$

所以质点正向左下方运动，而其路径与水平所成的角 $\beta$ 满足

$$\tan\beta = \frac{|\dot{y}|}{|\dot{x}|} = \frac{3V\sin\alpha}{V\cos\alpha} = 3\tan\alpha,$$

此即所求。

### 静止下来

**想一想** 速度要满足甚么条件，质点才会骤然停住?

急拉只会消去沿绳方向的动量（momentum）分量，垂直于绳的分量则不受影响。所以质点静止下来，当且仅当该垂直分量本身已经为零，即当且仅当拉紧前一刻的速度平行于绳 $\overrightarrow{OQ}$，其中 $Q$ 为质点在时间 $T$ 的位置。（速度并非零，因为 $\dot{x} = -V\cos\alpha \neq 0$。）

利用 $V^2 = gb\cos\alpha$ 及 $T = 4V\sin\alpha/g$，

$$\begin{aligned} x(T) &= b\sin\alpha - \frac{4V^2\sin\alpha\cos\alpha}{g} = b\sin\alpha\,(1-4\cos^2\alpha), \\ y(T) &= b\cos\alpha + \frac{4V^2\sin^2\alpha}{g} - \frac{8V^2\sin^2\alpha}{g} = b\cos\alpha\,(1-4\sin^2\alpha). \end{aligned}$$

向量 $(x,y)$ 与 $(\dot{x},\dot{y})$ 平行，当且仅当 $x\dot{y} - y\dot{x} = 0$。除以 $-bV$，得

$$3\sin^2\alpha\,(1-4\cos^2\alpha) - \cos^2\alpha\,(1-4\sin^2\alpha) = 0.$$

设 $S = \sin^2\alpha$，则 $\cos^2\alpha = 1-S$ 及 $1-4\cos^2\alpha = 4S-3$:

$$\begin{aligned} 3S(4S-3) - (1-S)(1-4S) &= 0 \\ 12S^2 - 9S - (1-5S+4S^2) &= 0 \\ 8S^2 - 4S - 1 &= 0, \end{aligned}$$

所以 $S = \frac{1}{16}\left(4 \pm \sqrt{16+32}\right) = \frac{1}{4}\left(1 \pm \sqrt{3}\right)$。由于 $S = \sin^2\alpha > 0$，负根不可能成立，而余下的根位于 $(0,1)$ 之内，所以它对应一个真正的锐角。以上每一步都是可逆的，所以当绳拉紧时质点瞬间静止，当且仅当

$$\sin^2\alpha = \frac{1+\sqrt{3}}{4}.$$

## 验算

- 公式 $V^2 = gb\cos\alpha$ 表明当 $\alpha \to \frac{1}{2}\pi$ 时 $V \to 0$：绳要在 $O$ 的水平高度变得松弛，质点必须在该处已经停下，这是正确的。

- 当 $\sin^2\alpha = \frac{1}{4}(1+\sqrt{3})$ 时,得 $\alpha \approx 55.7°$ 及 $\cos^2\alpha = \frac{1}{4}(3-\sqrt{3})$,所以 $Q = \big(b\sin\alpha\,(\sqrt{3}-2),\ -\sqrt{3}\,b\cos\alpha\big)$，位于 $O$ 的左下方。$OQ$ 的斜率为 $\sqrt{3}\cot\alpha/(2-\sqrt{3})$，而路径的斜率为 $3\tan\alpha$。两者相符，因为

$$3(2-\sqrt{3})\tan^2\alpha = \frac{3(2-\sqrt{3})(1+\sqrt{3})}{3-\sqrt{3}} = \frac{3(\sqrt{3}-1)}{\sqrt{3}(\sqrt{3}-1)} = \sqrt{3}.$$

  以电脑模拟这个运动，三个结果均得到证实。

## 常见错误

- 只用能量来求 $V$：能量把 $V$ 与初速联系起来，但以 $b$、$g$ 及 $\alpha$ 确定 $V$ 的，是张力为零的条件。
- 弄错 $P$ 点的速度方向。速度垂直于 $OP$，所以它与水平（而非铅直）成角 $\alpha$。
- 忘记证明 $T$ 是绳首次拉紧的时间；因式分解（factorisation）$gt^3(\frac{1}{4}gt - V\sin\alpha)$ 正好说明了这一点。

## STEP 之外

绳只能拉、不能推,所以松弛绳问题实际上是一个约束时有时无的受约束系统(constrained system)；同样的单侧约束（one-sided constraint）概念，亦出现于接触力学（contact mechanics），以及带不等式约束的拉格朗日力学（Lagrangian mechanics）之中。骤然的急拉是一种冲击性约束力（impulsive constraint force），在大学一年级课程中会与碰撞及恢复系数（coefficient of restitution）一并讨论。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对圆周运动和冲量都有细致的阐述。

详细解答

# STEP III 2018，第 12 题

概率与统计 · 顺序统计量

**STEP III 2018，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/18-s3-q12

## 关键思路

事件“第 $k$ 小的数不大于 $y$”与事件“$n$ 个数之中至少有 $k$ 个不大于 $y$”是同一回事，而不大于 $y$ 的数的个数服从二项分布（binomial distribution）。由此可以立即得出分布函数（distribution function）。将它求导会得到两个和式，而 (ii) 中的两条二项式恒等式正好令它们相消（telescope）。一旦知道了密度函数（density），由于密度函数的积分等于 1，便可以顺手求出一个积分；而同一个积分（把 $n$ 和 $k$ 各自平移）就给出期望值。

## 解答

把这 $n$ 个数记为 $X_1, \ldots, X_n$。它们互相独立（independent），每个都在 $[0,1]$ 上服从均匀分布（uniform distribution）。

### 第 (i) 部分

固定 $y$，其中 $0 \le y \le 1$，并设 $N$ 为满足 $X_i \le y$ 的 $X_i$ 的个数。每个 $X_i$ 满足 $X_i \le y$ 的概率为 $y$，且与其他各数独立，所以 $N \sim \mathrm{B}(n, y)$。

$Y_k \le y$ 成立，当且仅当这些数之中至少有 $k$ 个不大于 $y$。若 $Y_k \le y$，则 $Y_k$ 以及比它小的 $k-1$ 个数都不大于 $y$，所以 $N \ge k$。反过来，若 $N \ge k$，则最小的 $k$ 个数都不大于 $y$，特别地 $Y_k \le y$。因此

$$\mathrm{P}(Y_k \le y) = \mathrm{P}(N \ge k) = \sum_{m=k}^{n} \binom{n}{m} y^m (1-y)^{n-m},$$

这就是 $(*)$。

**第 (ii) 部分**

把阶乘写出来，对 $1 \le m \le n$，

$$m\binom{n}{m} = \frac{n!}{(m-1)!\,(n-m)!} = n \cdot \frac{(n-1)!}{(m-1)!\,\big((n-1)-(m-1)\big)!} = n\binom{n-1}{m-1},$$

同理，对 $0 \le m \le n-1$，

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

（当 $m = n$ 时，左边为 0；若把 $\binom{n-1}{n}$ 理解为 0，右边亦为 0。）

**密度函数。** $Y_k$ 的密度函数 $f(y)$ 是 $(*)$ 的导数。用乘积法则（product rule）逐项求导，

$$f(y) = \sum_{m=k}^{n} \binom{n}{m} \Big[m\,y^{m-1}(1-y)^{n-m} - (n-m)\,y^{m}(1-y)^{n-m-1}\Big].$$

在第二部分中，由于因子 $n-m$，$m=n$ 的项等于零，所以该和式实际上只到 $m = n-1$。利用上述两条恒等式，

$$f(y) = n\sum_{m=k}^{n} \binom{n-1}{m-1} y^{m-1}(1-y)^{n-m} \;-\; n\sum_{m=k}^{n-1} \binom{n-1}{m} y^{m}(1-y)^{n-1-m}.$$

? **想一想** 这两个和式看起来几乎一样。怎样才能把它们对齐?

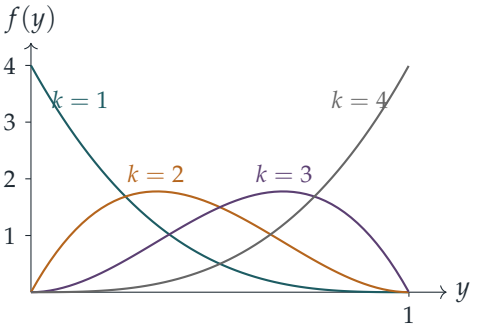
在第二个和式中令 $j = m+1$。它变成

$$n\sum_{j=k+1}^{n} \binom{n-1}{j-1} y^{j-1}(1-y)^{n-j},$$

这与第一个和式逐项相同，只是从 $k+1$ 而不是从 $k$ 开始。除了第一个和式中 $m = k$ 的项之外，其余全部抵消：

$$f(y) = n\binom{n-1}{k-1} y^{k-1}(1-y)^{n-k},$$

正如所求。（若 $k = n$，第二个和式是空和，第一个和式只有 $m = n$ 一项，所以结果依然成立。）




$n = 4$ 时的密度函数：四个数之中第 $k$ 小的数集中在 $k/5$ 附近。

**积分。** 密度函数在其取值范围上的积分等于 1，这里的范围是 $[0,1]$。因此

$$n\binom{n-1}{k-1}\int_0^1 y^{k-1}(1-y)^{n-k}\,\mathrm{d}y = 1,$$

所以

$$\int_0^1 y^{k-1}(1-y)^{n-k}\,\mathrm{d}y = \frac{1}{n\binom{n-1}{k-1}} = \frac{(k-1)!\,(n-k)!}{n!}.$$

此式对所有整数 $1 \le k \le n$ 都成立。

### 第 (iii) 部分

按定义，

$$\mathrm{E}(Y_k) = \int_0^1 y\,f(y)\,\mathrm{d}y = n\binom{n-1}{k-1}\int_0^1 y^k(1-y)^{n-k}\,\mathrm{d}y.$$

> ? **想一想** 这个积分与刚才求得的并不完全一样。取哪些 $n$ 和 $k$ 的值，才能把它化成那个形式?

被积函数（integrand）是 $y^{K-1}(1-y)^{N-K}$，其中 $K = k+1$、$N = n+1$，且 $1 \le K \le N$，所以 (ii) 的公式适用：

$$\int_0^1 y^k(1-y)^{n-k}\,\mathrm{d}y = \frac{k!\,(n-k)!}{(n+1)!}.$$

由于 $n\binom{n-1}{k-1} = \dfrac{n!}{(k-1)!\,(n-k)!}$，

$$\mathrm{E}(Y_k) = \frac{n!}{(k-1)!\,(n-k)!}\cdot\frac{k!\,(n-k)!}{(n+1)!} = \frac{k}{n+1}.$$

## 验算

- 当 $n=1$、$k=1$ 时，只有一个均匀分布的数，而 $\mathrm{E}(Y_1)=\frac{1}{2}$，理应如此。
- 对称性（symmetry）：把每个 $X_i$ 换成 $1-X_i$，第 $k$ 小的数就变成第 $k$ 大的数，所以 $\mathrm{E}(Y_k)+\mathrm{E}(Y_{n+1-k})=1$。确实 $\frac{k}{n+1}+\frac{n+1-k}{n+1}=1$。
- 这 $n$ 个数把 $[0,1]$ 分成 $n+1$ 段间隔，而答案表明每段间隔的平均长度都是 $\frac{1}{n+1}$，从对称性看来非常合理。
- 当 $n=5$、$k=2$ 时：$(*)$ 在 $y=0.3$ 处的值为 0.47178；用电脑模拟 200 000 个样本，得到 0.4726，而 $Y_2$ 的样本平均值为 0.33333，理论值为 $\frac{1}{3}$。

## 常见错误

- 写成 $\mathrm{P}(Y_k\le y)=\binom{n}{k}y^k(1-y)^{n-k}$，即恰好有 $k$ 个值小于 $y$ 的概率，而不是至少有 $k$ 个。
- 忘记第二个和式中 $m=n$ 的项等于零，以致相消看似差了一项而不成立。
- 在 (iii) 中重用积分时只平移 $k$（而没有平移 $n$）。

## STEP 之外

这些 $Y_k$ 是样本的*顺序统计量*（order statistics），而 $Y_k$ 服从参数为 $k$ 和 $n-k+1$ 的贝塔分布（Beta distribution）。(ii) 中的积分就是贝塔函数 $B(a,b)=\int_0^1 y^{a-1}(1-y)^{b-1}\,\mathrm{d}y=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$，你日后会与伽玛函数（Gamma function）一并再遇见它。顺序统计量是样本中位数和无母数统计（non-parametric statistics）的基础。Blitzstein 和 Hwang 的 *Introduction to Probability*（《概率导论》）对顺序统计量和贝塔分布有十分精彩的论述，其中包括贝塔积分的一个“故事式”证明，与这里的论证非常相似。

详细解答

# STEP III 2018，第 13 题

概率与统计 · 泊松分布

> **STEP III 2018，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/18-s3-q13

## 关键思路

把 $t=-1$ 代入概率母函数（probability generating function），会令每个概率带上符号 $(-1)^r$，所以把 $G(1)$ 和 $G(-1)$ 取平均，奇数项便互相抵消，只留下偶数项。用四次单位根（fourth roots of unity）$1,-1,\mathrm{i},-\mathrm{i}$ 作同样的筛选，就能抽出 4 的倍数。每个新的母函数一旦写成闭合形式（closed form），平均值（mean）就只是它在 $t=1$ 的导数，问题便化为比较几个初等函数。

## 解答

记 $p_r=\mathrm{P}(X=r)$，于是 $G(t)=\sum_{r\geq 0}p_r t^r$。

**偶数值。** 把级数在 $t=1$ 和 $t=-1$ 的值相加，

$$G(1)+G(-1)=\sum_{r\geq 0}p_r\left(1+(-1)^r\right)=2\sum_{r\text{ 为偶数}}p_r,$$

因为当 $r$ 为偶数时 $1+(-1)^r$ 等于 2，当 $r$ 为奇数时等于 0。两边除以 2 即得所求结果。（由于 $\sum p_r=1$，两个级数都绝对收敛（converge absolutely）。）

**泊松（Poisson）分布的母函数。** 当 $p_r=\mathrm{e}^{-\lambda}\lambda^r/r!$ 时，

$$G(t)=\mathrm{e}^{-\lambda}\sum_{r\geq 0}\frac{(\lambda t)^r}{r!}=\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda t}=\mathrm{e}^{-\lambda(1-t)},$$

这里用了指数级数（exponential series），它对所有 $t$ 都收敛。

### 第 (i) 部分

$Y$ 的概率总和必须为 1，所以 $k$ 是 $\mathrm{P}(X \text{ 为偶数})$ 的倒数。由第一个结果，

$$\mathrm{P}(X \text{ 为偶数}) = \tfrac{1}{2}\left(1+\mathrm{e}^{-2\lambda}\right) = \mathrm{e}^{-\lambda}\cdot\tfrac{1}{2}\left(\mathrm{e}^{\lambda}+\mathrm{e}^{-\lambda}\right) = \mathrm{e}^{-\lambda}\cosh\lambda, \qquad \text{所以} \qquad k = \frac{\mathrm{e}^{\lambda}}{\cosh\lambda}.$$

$Y$ 的母函数只保留 $G$ 的偶数项：

$$\begin{aligned} G_Y(t) = k\sum_{r\text{ 为偶数}} p_r t^r &= k\cdot\tfrac{1}{2}\left(G(t)+G(-t)\right) \\ &= \frac{\mathrm{e}^{\lambda}}{\cosh\lambda}\cdot\mathrm{e}^{-\lambda}\cdot\tfrac{1}{2}\left(\mathrm{e}^{\lambda t}+\mathrm{e}^{-\lambda t}\right) = \frac{\cosh\lambda t}{\cosh\lambda}, \end{aligned}$$

其中 $\frac{1}{2}\left(G(t)+G(-t)\right)$ 是 $G$ 的偶部（even part），与第一个结果完全一样，只是以 $t$ 代替 1。这正是所求的形式。

平均值为 $G_Y'(1)$：

$$\mathrm{E}(Y) = \frac{\lambda\sinh\lambda}{\cosh\lambda} = \lambda\tanh\lambda.$$

当 $\lambda > 0$ 时有 $\mathrm{e}^{-\lambda} > 0$，所以 $\sinh\lambda = \frac{1}{2}(\mathrm{e}^{\lambda}-\mathrm{e}^{-\lambda}) < \frac{1}{2}(\mathrm{e}^{\lambda}+\mathrm{e}^{-\lambda}) = \cosh\lambda$，从而 $0 < \tanh\lambda < 1$。因此 $\mathrm{E}(Y) < \lambda$，得证。

### 第 (ii) 部分

现在我们只需保留 $r$ 为 4 的倍数的项。

> ? **想一想** 当我们想按 4（而非 2）的倍数筛选时，甚么担当 $t = -1$ 的角色?

用全部四个四次单位根。对整数 $r$，

$$1+(-1)^r+\mathrm{i}^r+(-\mathrm{i})^r = \begin{cases} 4 & \text{若} 4 \mid r, \\ 0 & \text{其他情况.} \end{cases}$$

若 $4 \mid r$，每一项都是 1。若 $r$ 为奇数，各项两两抵消：$(-1)^r = -1$ 且 $(-\mathrm{i})^r = -\mathrm{i}^r$。若 $r \equiv 2 \pmod 4$，各项为 $1+1-1-1 = 0$。所以

$$\sum_{4\mid r}\frac{x^r}{r!} = \tfrac{1}{4}\left(\mathrm{e}^{x}+\mathrm{e}^{-x}+\mathrm{e}^{\mathrm{i}x}+\mathrm{e}^{-\mathrm{i}x}\right) = \tfrac{1}{2}\left(\cosh x+\cos x\right).$$

（也可以直接看出：$\cosh x = \sum_{r\text{ 为偶数}} x^r/r!$，而 $\cos x$ 的项相同，只是符号依次为 $+,-,+,\ldots$，所以两者相加时，$4 \mid r$ 的项加倍，其余的项抵消。）

取 $x = \lambda t$ 并乘以 $\mathrm{e}^{-\lambda}$，再以 $G_Z(1) = 1$ 这个要求定出 $c$，得

$$G_Z(t) = \frac{\cosh\lambda t+\cos\lambda t}{\cosh\lambda+\cos\lambda}.$$

当 $\lambda > 0$ 时分母为正，因为 $\cosh\lambda > 1 \geq -\cos\lambda$。求导后令 $t = 1$，

$$\mathrm{E}(Z) = G_Z'(1) = \frac{\lambda(\sinh\lambda - \sin\lambda)}{\cosh\lambda + \cos\lambda}.$$

由于 $\lambda > 0$ 且分母为正，

$$\begin{aligned}\mathrm{E}(Z) < \lambda &\iff \sinh\lambda - \sin\lambda < \cosh\lambda + \cos\lambda \\ &\iff \sin\lambda + \cos\lambda > \sinh\lambda - \cosh\lambda = -\mathrm{e}^{-\lambda} \\ &\iff \sin\lambda + \cos\lambda + \mathrm{e}^{-\lambda} > 0.\end{aligned}$$

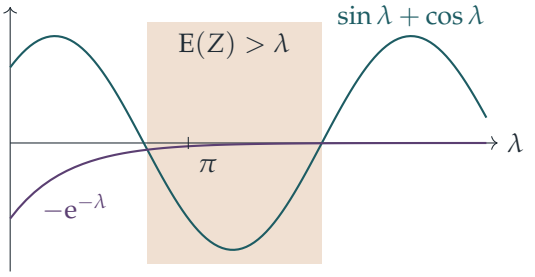
这在 $\lambda = \pi$ 时不成立：此时 $\sin\pi + \cos\pi + \mathrm{e}^{-\pi} = -1 + \mathrm{e}^{-\pi} < 0$，因为 $\mathrm{e}^{-\pi} < 1$。事实上

$$\mathrm{E}(Z)\big|_{\lambda=\pi} = \frac{\pi\sinh\pi}{\cosh\pi - 1} > \pi,$$

因为 $\sinh\pi > \cosh\pi - 1$（等价于 $1 > \mathrm{e}^{-\pi}$）。

所以答案是**否**：$\mathrm{E}(Z) < \lambda$ 并非对所有正数 $\lambda$ 都成立。

若要看得更全面，可写 $\sin\lambda + \cos\lambda = \sqrt{2}\sin(\lambda + \frac{\pi}{4})$。每当这条曲线跌到 $-\mathrm{e}^{-\lambda}$ 以下，条件便不成立。第一个这样的区间大约由 $\lambda \approx 2.42$ 至 $\lambda \approx 5.49$，而随着 $\mathrm{e}^{-\lambda} \to 0$，之后的区间愈来愈接近 $\left(\frac{3\pi}{4} + 2n\pi,\ \frac{7\pi}{4} + 2n\pi\right)$。




## 验算

- 当 $\lambda \to 0$ 时，$\mathrm{E}(Y) = \lambda\tanh\lambda \approx \lambda^2$。这是合理的：$Y$ 几乎总是 0，而 $\mathrm{P}(Y = 2) \approx \lambda^2/2$。
- 当 $\lambda \to \infty$ 时，$\tanh\lambda \to 1$，所以 $\mathrm{E}(Y)$ 略小于 $\lambda$。当 $\lambda$ 很大时，泊松分布很分散，所以舍弃奇数值几乎不会改变平均值。
- 电脑直接对泊松级数求和，核实了每个母函数和两个平均值。例如在 $\lambda = \pi$ 时，得到 $\mathrm{E}(Y) \approx 3.1299 < \pi$ 及 $\mathrm{E}(Z) \approx 3.4254 > \pi$。它亦找到 $\sin\lambda + \cos\lambda + \mathrm{e}^{-\lambda}$ 的变号点位于 $\lambda \approx 2.419$ 和 5.495。

## 常见错误

- 忘记归一化常数（normalising constant）：单独的 $\cosh \lambda t$ 并不是母函数，因为它在 $t = 1$ 的值不是 1。
- 以为 (ii) 的答案既然与 (i) 类似，就必定是“是”。只保留 4 的倍数，可能在 $\lambda$ 以下移除的概率质量比以上的多，例如 $\lambda = \pi$ 时：最接近的容许值 4 在 $\pi$ 之上，而且其概率远大于 0 的概率。
- 把不等式除以 $\cosh \lambda + \cos \lambda$ 时，没有先检查它是否为正。

## STEP 之外

把母函数在单位根上取平均，以抽出某一同余类（residue class）中的项，这方法称为单位根滤波（roots-of-unity filter），又称级数的多重分割（multisection）。它是组合数学（combinatorics）的基本工具，也是离散傅里叶变换（discrete Fourier transform）最简单的情形。像本题这样，把分布以某事件为条件，再重新计算平均值，在概率论中随处可见。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）深入讲解了母函数。

# 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

**2018 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。

