2025 国际大学生数学竞赛题目解析

虚空若叶睦

2025 年 8 月 8 日

1. 设实系数多项式 $f \in \mathbb{R}[x]$ 满足 $\deg f \geqslant 2$. 对每个 $x \in \mathbb{R}$, 设 $l_x \subset \mathbb{R}^2$ 表示 f 的图像在点 $(x, f(x))$ 处的切线.

(a) 设 f 的次数是奇数. 证明: $\bigcup_{x\in\mathbb{R}} l_x = \mathbb{R}^2$.

(b) 是否存在一个偶数次数多项式使得上述等号仍然成立?

纯送, 高中生都会证.

证明 (a) 在点 $(x_0, f(x_0))$ 处的切线方程为

$$y = f'(x_0)(x - x_0) + f(x_0). \tag{1}$$

下面只需证明: 对任意 $x, y \in \mathbb{R}$, 方程 (1) 关于 x_0 都有解. 显然, (1) 的右端关于 x_0 是奇数次多项式, 因此 (1) 必有实根, 命题得证.

(b) 不存在. 不妨设 $P(x)$ 是偶数次多项式, 且首项系数为正. 只需证明 y 轴正半轴无法被完全覆盖. 在 (1) 中令 $x = 0$, 可以得到 x_0 处的切线与 y 轴交点的纵坐标为 $f(x_0) - f'(x_0)x_0$. 注意到, 这是一个首项系数为负的偶数次多项式, 因此存在 $M > 0$, 使得

$$f(x_0) - f'(x_0)x_0 < 0, \quad \forall |x_0| \geqslant M.$$

此时切线完全不覆盖 y 轴的正半轴. 另一方面, 当 $x_0 \in [-M, M]$ 时, $f(x_0) - f'(x_0)x_0$ 的值域一定是紧集, 此时切线不能完全覆盖 y 轴的正半轴. 符合要求的偶数次多项式不存在. ■

2. 设 $f:[-1,1]\to\mathbb{R}$ 二阶连续可微, 且 $\int_{-1}^{1} f(x)\mathrm{d}x=0$, $f(1)=f(-1)=1$. 证明:

$$\int_{-1}^{1}(f''(x))^2\mathrm{d}x\geqslant 15,$$

并确定使等号成立的函数,

看起来很像数学一的考研题. 如果对解题技巧没有经验还是有点难度的.

证明 解题的关键在于构造一个特殊的函数 $P(x)$, 使得 $\int_{-1}^{1} f''(x)P(x)\mathrm{d}x$ 是易于计算的. 接着使用 Cauchy 不等式就能得到原题要求的界. 利用分部积分公式,

$$\begin{aligned}\int_{-1}^{1} f''(x)P(x)\mathrm{d}x&=\int_{-1}^{1} P(x)\mathrm{d}f'(x)\\&=\big[f'(x)P(x)\big]_{-1}^{1}-\int_{-1}^{1} f'(x)P'(x)\mathrm{d}x\\&=\big[f'(x)P(x)\big]_{-1}^{1}-\big[f(x)P'(x)\big]_{-1}^{1}+\int_{-1}^{1} f(x)P''(x)\mathrm{d}x.\end{aligned}$$

为了让等式的右端容易计算, 需要取 $P(-1)=P(1)=0$ 且 $P''(x)$ 为定值. 取 $P(x)=x^2-1$, 可以得到

$$\int_{-1}^{1} f''(x)(x^2-1)\mathrm{d}x=-4.$$

于是, 根据 Cauchy 不等式可得

$$16\leqslant\int_{-1}^{1}(f''(x))^2\mathrm{d}x\int_{-1}^{1}(x^2-1)^2\mathrm{d}x=\frac{16}{15}\int_{-1}^{1}(f''(x))^2\mathrm{d}x,$$

从而原不等式得证. 等号成立的条件是 $f''(x)$ 正好和 $1-x^2$ 成正比, 结合条件可以计算出

$$f(x)=\frac{1}{16}(-5x^4+30x^2-9).$$

于是原命题得证. ■

3. 设 $\mathcal{S}$ 表示所有秩为 1, 元素均为 -1 或 $+1$ 的 2025×2025 对称方阵构成的集合. 等可能随机且独立地从 $\mathcal{S}$ 中选取矩阵 A, B. 试求 A, B 可交换, 即 $AB = BA$ 的概率.

这是一道简单的线性代数题目, 可以放到考研里.

证明 答案是 $\dfrac{1}{2^{2024}}$. 对任意矩阵 $A \in \mathcal{S}$, 由于 A 实对称且秩为 1, 其特征值分解为 $A = cxx^\top$, 其中 $c \neq 0$ 且 $x \in \mathbb{R}^{2025}$. 此时, 可不妨设 $|c| = 1$, 因此 $\mathcal{S}$ 中的矩阵有如下刻画:

$$\mathcal{S} = \left\{ \pm xx^\top : x \in \mathbb{R}^{2025}, x_i \in \{-1, 1\} \right\}.$$

注意到, 对矩阵乘以一个常数不改变其可交换性, 因此只需考察集合

$$\mathcal{S}_1 = \left\{ xx^\top : x \in \mathbb{R}^{2025}, x_i \in \{-1, 1\} \right\}.$$

由于每个 x_i 的取值从 $\{-1, 1\}$ 中选取, x 的构造方法有 2^{2025} 种. 但 x 和 $-x$ 对应于同一个矩阵 $xx^\top$, 故 $\mathcal{S}_1$ 中的矩阵个数为 $|\mathcal{S}_1| = 2^{2024}$.

下面从 $\mathcal{S}_1$ 中独立地选取矩阵 $A = xx^\top$ 和 $B = yy^\top$, 则 $AB = BA$ 的等价条件是

$$x(x^\top y)y^\top = y(y^\top x)x^\top \iff xy^\top = yx^\top \iff \frac{x_i}{y_i} = \frac{x_j}{y_j},$$

即 x, y 两个向量线性相关, 从而一定有 $A = B$. 这就是说, $\mathcal{S}_1$ 中两个矩阵 A, B 可交换当且仅当 $A = B$. 因此 $AB = BA$ 的概率是 $\dfrac{1}{2^{2024}}$. ■

4. 设 a 是正偶数. 试求所有实数 x, 使得对所有正整数 n, 都有

$$\left[\sqrt[a]{n^a+x}\cdot n^{a-1}\right]=n^a+\left[\frac{x}{a}\right].$$

(这里 $[x]$ 表示不超过 x 的最大整数.)

我觉得这个题的难度是一试填空题.

证明 当 $a=2$ 时, x 的取值范围是 $[-1,2)\cup[3,4)$; 当 $a\geqslant 4$ 时, x 的取值范围是 $[-1,a)$.

首先, 当 $n=1$ 时, 定义域要求 $x\geqslant -1$. 对一般的 n, 原等式可以等价写为

$$\left[n^a\left(1+\frac{x}{n^a}\right)^{\frac{1}{a}}\right]=n^a+\left[\frac{x}{a}\right]. \tag{1}$$

根据 Bernoulli 不等式, 有

$$n^a\left(1+\frac{x}{n^a}\right)^{\frac{1}{a}}\leqslant n^a\left(1+\frac{x}{an^a}\right)\leqslant n^a+\frac{x}{a},$$

因此 (1) 成立的充分必要条件是: 对每个 $n\in\mathbb{N}$, 都不存在整数 t 使得

$$n^a\left(\left(1+\frac{x}{n^a}\right)^{\frac{1}{a}}-1\right)<t\leqslant\frac{x}{a}. \tag{2}$$

- 若 $a\geqslant 4$, 我们说明 $x\geqslant a$ 不符合要求. 当 $x\geqslant a$ 时, 一定存在正整数 k 使得 $ka\leqslant x<(k+1)a$. 因此可在 (2) 中取 $t=k$. 此时 (2) 等价于

  $$\left(1+\frac{x}{n^a}\right)^{\frac{1}{a}}<1+\frac{k}{n^a}. \tag{3}$$

  特别的, 在 $n=1$ 时, (3) 转化为

  $$(1+(k+1)a)^{\frac{1}{a}}\leqslant 1+k,$$

  这对一切正整数 k 和正偶数 $a\geqslant 4$ 都成立. 因此 $a\geqslant 4$ 且 $x\geqslant a$ 不符合要求.

- 若 $a=2$, 我们说明 $x\geqslant 4$ 不符合要求. 类似于上面的讨论, 设 $2k\leqslant x<2k+2$, 其中 $k\geqslant 2$. 在 (2) 中取 $t=k$, 则 (2) 等价于

  $$\left(1+\frac{x}{n^2}\right)^{\frac{1}{2}}<1+\frac{k}{n^2}. \tag{4}$$

特别的, 在 $n=1$ 时, (4) 转化为

$$(2k+3)^{\frac{1}{2}} \leqslant 1+k,$$

这对一切正整数 $k \geqslant 2$ 成立. 因此 $a=2$ 且 $x \geqslant 4$ 不合要求.

- 若 $a=2$, 我们说明 $2 \leqslant x < 3$ 不符合要求. 类似于上面的讨论, 在 (2) 中取 $t=1$, 则 (2) 等价于

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

特别地, 在 $n=1$ 时, 恒有

$$(1+x)^{\frac{1}{2}} < 4^{\frac{1}{2}} = 2,$$

因此 $a=2$ 且 $2 \leqslant x < 3$ 不合要求. ■

最后, 对剩下的情况, 可以利用 (2) 验证是符合要求的.

5. 对正整数 n, 设 $[n]=\{1,2,\cdots,n\}$. 记 S_n 是所有从 $[n]$ 到 $[n]$ 的双射构成的集合, T_n 是所有从 $[n]$ 到 $[n]$ 的映射构成的集合. 对映射 $\tau\in T_n$, 记它的阶 $\operatorname{ord}(\tau)$ 为集合 $\{\mathrm{id},\tau,\tau\circ\tau,\tau\circ\tau\circ\tau,\cdots\}$ 中不同的映射个数, 这里 id 表示恒等映射, $\circ$ 表示复合. 设

$$g(n)=\max_{\tau\in S_n}\operatorname{ord}(\tau),\qquad f(n)=\max_{\tau\in T_n}\operatorname{ord}(\tau).$$

证明: 对充分大的 n, 有 $f(n)<g(n)+n^{0.501}$.

看一眼就知道就不是初等方法能解决的问题. 这东西能放在大学生数学竞赛?

证明 先介绍一些解析数论的背景知识.

- **素数计数函数 (prime-counting function)**

  对任意正整数 n, 定义 $\pi(n)$ 为不超过 n 的素数个数. 素数定理说明了:

  $$\pi(n)\sim\frac{n}{\log n},\qquad \text{当 } n\to\infty \text{ 时}.$$

  这里 $\sim$ 的含义是渐进等价, 即左右两端的商在 $n\to\infty$ 时的极限为 1.

- **Landau 函数 (Landau's function)**

  对任意正整数 n, 定义 $g(n)$ 是对称群 S_n 中的元素的最大阶. 等价地说, $g(n)$ 是 n 的任意划分的最小公倍数的最大可能值:

  $$g(n)=\max\left\{\operatorname{lcm}(c_1,c_2,\cdots,c_k):\sum_{i=1}^{k}c_i=n,c_i\in\mathbb{N}\right\}.$$

  这一等价性成立的原因是: 对称群 S_n 中的任意置换 τ 都可以划分成若干个有向圈, 若设这些有向圈的长度分别为 $c_1,c_2,\cdots,c_k$, 则 τ 的阶 (即使得 $\tau^m=\mathrm{id}$ 的最小正整数 m) 恰好是 $\operatorname{lcm}(c_1,c_2,\cdots,c_k)$. 显然 $g(n)$ 单调递增. Landau 在 1902 年证明了

  $$\log g(n)\sim\sqrt{n\log n},\qquad \text{当 } n\to\infty \text{ 时}.$$

- **素数 Ω 函数 (prime Omega function)**

  对任意正整数 n, 令 $\Omega(n)$ 是 n 的素因子分解中素数的个数 (计算重数). 即当 $n=p_1^{\alpha_1}\cdots p_k^{\alpha_k}$ 时, $\Omega(n)=\alpha_1+\cdots+\alpha_k$. 容易看出

  $$\Omega(n)\leqslant\log_2 n\leqslant 2\log n,\qquad \text{对任意正整数 } n.$$

回到原题. 由条件可知 $g(n)$ 是 Landau 函数, 下面计算 $f(n)$ 的值. 对每个从 $[n]$ 到 $[n]$ 的映射 $\tau \in T_n$, 可以作出一个有 n 个顶点的有向图 G_n, 其中从 $k \in [n]$ 到 $\tau(k) \in [n]$ 恰有一条有向边, $k = 1, \cdots, n$. 此时, G_n 可以被划分为一些连通分支, 其中每个连通分支包含一个有向圈, 和一些指向该圈的树, 如下图所示.

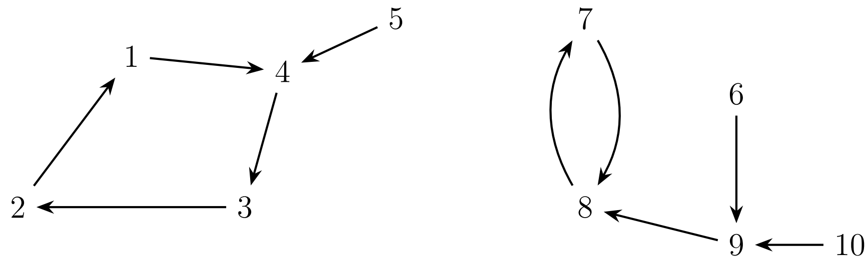

图 1: 这是一个 $n = 10$ 情形下映射 $\tau : [n] \to [n]$ 的例子. 此时 $l = 2, p = 4$.

由于序列 $\{\mathrm{id}, \tau, \tau^2, \tau^3, \cdots\}$ 最终会变得周期, 设 l 是最小的非负整数, 使得存在一个最小的正整数 p 满足 $\tau^l = \tau^{l+p}$.

- l 被称为 τ 的尾长, 它是图中任意一点到达一个有向圈的最大长度.
- p 被称为 τ 的周期, 它是图中所有有向圈的长度的最小公倍数.

根据阶的定义, $\mathrm{ord}(\tau)$ 是集合 $\{\mathrm{id}, \tau, \tau^2, \cdots, \tau^{l+p-1}\}$ 的大小, 因此 $\mathrm{ord}(\tau) = l + p$. 如果 $\tau \in S_n$, 那么 $l = 0$, 从而 $\mathrm{ord}(\tau)$ 与 S_n 置换中的阶的定义等价.

设此时在 $\tau \in T_n$ 中共有 m 个点形成有向圈, 则有 $l \leqslant n - m$ 和 $p \leqslant g(m)$. 因此

$$\mathrm{ord}(\tau) = l + p \leqslant n - m + g(m).$$

因此, 当 $\tau \in T_n$ 时, $\mathrm{ord}(\tau)$ 的最大可能值是

$$f(n) = \max_{1 \leqslant m \leqslant n} \big\{n - m + g(m)\big\}. \tag{1}$$

下面只需证明: 由 (1) 定义的函数 $f(n)$ 满足 $f(n) < g(n) + n^{0.501}$ 对充分大的 n 成立, 即:

$$\text{当 } n \text{ 充分大时,} \quad \text{对任意 } 0 \leqslant k \leqslant n-1, \quad \text{有 } k + g(n-k) < g(n) + n^{0.501}. \tag{2}$$

设 $0 \leqslant k_0 \leqslant n - 1$ 是使得 $k + g(n - k)$ 取最大值的 k, 则 (2) 的一个充分条件是

$$\text{当 } n \text{ 充分大时,} \quad k_0 < n^{0.501}, \quad \text{其中 } k_0 \text{ 使得 } k + g(n-k) \text{ 最大.}$$

更进一步, 我们只需证明:

$$\text{当 } n \text{ 充分大时}, \quad \text{对任意 } n^{0.501} \leqslant k \leqslant n-1, \quad \text{有 } k+g(n-k)<g(n). \tag{3}$$

利用前面介绍的 $\pi(n)$ 和 $\Omega(n)$ 两个函数, 可以得到: 一方面, 当 $k \geqslant n^{0.501}$ 时

$$\pi(k) \geqslant c_1 \frac{k}{\log k} \geqslant c_1 \frac{n^{0.501}}{\log n^{0.501}}, \tag{4}$$

其中 c_1 是不依赖于 n, k 的常数. 另一方面,

$$\Omega(g(n-k)) \leqslant c_2 \log g(n-k) \leqslant c_2 \log g(n) \leqslant c_2 \sqrt{n \log n}. \tag{5}$$

结合 (4)(5) 可知: 当 n 充分大时, 对任意 $n^{0.501} \leqslant k \leqslant n-1$, 都有

$$\pi(k) \geqslant c_1 \frac{n^{0.501}}{\log n^{0.501}} > c_2 \sqrt{n \log n} \geqslant \Omega(g(n-k)).$$

这意味着, 一定存在素数 $p \leqslant k$, 且 p 与 $g(n-k)$ 互素. 于是有不等式

$$g(n) \geqslant \operatorname{lcm}\big(g(n-k), p, k-p\big) \geqslant \operatorname{lcm}(g(n-k), p) = p g(n-k) \geqslant 2 g(n-k).$$

于是从上述可以得到: 当 $n^{0.501} \leqslant k \leqslant n-1$ 时, 有

$$k+g(n-k) \leqslant n+\frac{1}{2} g(n)<g(n), \tag{6}$$

其中最后一个不等号成立是因为 $g(n) \geqslant e^{c_3 \sqrt{n \log n}}$ 对某常数 $c_3>0$ 成立.

由 (6) 即可知 (3) 成立, 从而原命题得证. ■

6. 设 $f:(0,+\infty)\to\mathbb{R}$ 是连续可微函数, 实数 $b>a>0$ 满足 $f(a)=f(b)=k$. 证明: 存在一点 $\xi\in(a,b)$ 使得

$$f(\xi)-\xi f'(\xi)=k.$$

实在太简单了, 数学三都不会出这么简单的.

证明 不妨设 $k=0$, 否则可用 $f(x)-k$ 代替讨论. 现考察函数

$$g(x)=\frac{f(x)}{x},$$

则

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

由于 $g(a)=g(b)=0$, 根据 Lagrange 中值定理可知存在 $\xi\in(a,b)$ 使得 $g'(\xi)=0$, 即 $f(\xi)-\xi f'(\xi)=0$. 原命题得证. ■

7. 设正整数集为 $\mathbb{N}$. 试求所有满足如下两个条件的非空子集 $M \subset \mathbb{N}$:

- 若 $x \in M$, 则 $2x \in M$;
- 若 $x, y \in M$, $x+y$ 为偶数, 则 $\dfrac{x+y}{2} \in M$.

非常有趣的组合题, 难度不大, 可以放在一试第 10 题的位置.

证明 所有满足条件的集合为: $\{nt : n \geqslant n_0\}$, 其中 n_0 为正整数, t 为正奇数.

对于一个符合条件的集合 M, 我们证明 M 一定具有上面的形式. 设 $x, x+t$ 是 M 中最小的两个元素, 其中 x, t 均为正整数. 若 t 为偶数, 则由条件可知

$$x + \frac{t}{2} = \frac{x + (x+t)}{2} \in M,$$

这与 $x+t$ 是 M 中次小的元素矛盾! 因此 t 必为奇数. 由归纳法知, 对任何正整数 n 有

$$nx \in M, \qquad n(x+t) \in M. \tag{1}$$

接下来证明: $x + 2t \in M$. 注意到有如下的引理:

若 x, y 为正整数且 $(x, y) = 1$, 则当 $n > xy$ 时, 关于 (a, b) 的不定方程

$$ax + by = n$$

一定有正整数解. 由此可知, 一定存在正整数 n, 使得存在正整数 a, b 满足

$$ax + b(x+t) = x + 2^n t. \tag{2}$$

根据 (1), 我们有 $ax \in M$ 和 $b(x+t) \in M$. 由条件和 (2) 即可知 $x + 2^n t \in M$. 于是有

$$x + 2^{n-1}t = \frac{x + (x + 2^n t)}{2} \in M \Longrightarrow \cdots \Longrightarrow x + 2t = \frac{x + (x + 2^2 t)}{2} \in M,$$

从而 $x + 2t \in M$. 由于 $x, x+t, x+2t$ 构成等差数列, 故根据归纳法可以得到: 对任意非负整数 n, 都有 $x + nt \in M$. 下面证明: M 中不会有不是 $x + nt$ 形式的元素.

否则, 假设 y 是 M 中最小的, 不是 $x + nt$ 形式的元素, 则 $y > x + t$, 因为 $x + t$ 是次小的元素. 此时, x 和 $x+t$ 两个元素之中必有一个与 y 同奇偶, 因此一定有

$$\frac{x+y}{2} \in M \quad 或 \quad \frac{x+y+t}{2} \in M. \tag{3}$$

无论 (3) 中的哪种情况出现, 都有 $\dfrac{x+y}{2} < \dfrac{x+y+t}{2} < y$, 于是根据 y 的最小性, $\dfrac{x+y}{2}$ 或 $\dfrac{x+y+t}{2}$ 一定是形如 $x+nt$ 的元素, 其中 n 是正整数. 这样, y 本身也是形如 $x+nt$ 的元素, 矛盾! 故 M 中的元素必有 $x+nt$ 的形式, 其中 n 为非负整数.

特别的, $2x \in M$ 意味着 $2x = x + nt$ 对某个正整数 n 成立, 从而 $t|x$. 设 $x = n_0 t$, 于是

$$M = \big\{nt : n \geqslant n_0\big\}, \text{ 其中 } n_0 \text{ 为正整数}, t \text{ 为正奇数}. \qquad \blacksquare$$

8. 对一个实矩阵 $A \in M_n(\mathbb{R})$, 记 A^{R} 为把它逆时针旋转 90° 得到的矩阵. 例如,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}^{\mathsf{R}} = \begin{bmatrix} 3 & 6 & 9 \\ 2 & 5 & 8 \\ 1 & 4 & 7 \end{bmatrix}.$$

证明: 若 $A = A^{\mathsf{R}}$, 则对 A 的每个特征值 λ, 有 $\operatorname{Re}\lambda = 0$ 或 $\operatorname{Im}\lambda = 0$.

很不常规的线性代数题目, 需要一点几何直觉.

证明 先证一个引理: 考虑 $\mathbb{R}^2$ 上一个以 0 为中心的正方形 S, 则 S 关于 x 轴的轴对称图形和关于 $y = -x$ 的轴对称图像完全重合. (这是什么玩意? 难道不是初中生就会证的东西?)

设 S 上的一个顶点是 $z = a + bi$. 它关于 x 轴的对称点是 $z_1 = a - bi$, 关于 $y = -x$ 的对称点是 $z_2 = -b - ai$. 注意到 $z_2 = iz_1$, 故 z_1, z_2 在同一个正方形上, 命题得证.

回到原题, 定义矩阵

$$J = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix} \in M_n(\mathbb{R}).$$

由于 $A = A^{\mathsf{R}} = A^{\mathsf{RR}} = A^{\mathsf{RRR}}$, 故有

$$JA = A^\top. \tag{1}$$

前者的作用是把 A 的所有行沿着 x 轴作对称, 后者的作用是把 A 沿着 $y = -x$ 作对称. 根据引理, 它们应该相等. 于是根据 (1) 有

$$A = JA^\top \Longrightarrow A^2 = AJA^\top,$$

从而 A^2 为实对称矩阵, 于是 A^2 的特征值全部为实数. 因此, A 的特征值 λ 必然满足 $\operatorname{Re}\lambda = 0$ 或 $\operatorname{Im}\lambda = 0$. ■

9. 设 n 是正整数. 考虑如下的随机过程, 它会生成 n 个不同的正整数:

首先, 选取 X_1, 它的选取服从对所有正整数 i, $\mathbb{P}(X_1 = i) = 2^{-i}$. 对 $1 \leqslant j \leqslant n-1$, 若 $X_1, \cdots, X_j$ 已选定, 将剩余的正整数按剩下排列成 $n_1 < n_2 < \cdots$, 然后选取 X_{j+1}, 它的选取服从对所有正整数 i, $\mathbb{P}(X_{j+1} = n_i) = 2^{-i}$.

记 $Y_n = \max\{X_1, \cdots, X_n\}$. 证明:

$$\mathbb{E}[Y_n] = \sum_{i=1}^{n} \frac{2^i}{2^i - 1}.$$

这是一道难度中等的概率题, 兼具了概率思想和计算技巧. 可以作为概率论期末考试的压轴题. 如果在每步完成以后直接考察正整数列 $n_1, n_2, \cdots$ 的排列, 将会非常复杂. 更好的做法是固定所有正整数 $1, 2, \cdots$ 的位置: 每当正整数 k 被取走, 那么 k 处的概率值变为 0, k 之后的所有正整数被取走的概率翻倍.

证明 将所有正整数 $1, 2, \cdots$ 写成一列, 其中正整数 i 上的标签为 2^{-i}, 表示它被选择的初始概率. 每当一个正整数 k 被选中时, 将 k 上的标签改为 0, 将比 k 大的所有正整数上的标签改为原来的两倍. 这样进行 n 步后取中的随机变量 $X_1, \cdots, X_n$ 恰好就是条件中所描述的随机变量. 下面来计算 $Y_n = \max\{X_1, \cdots, X_n\}$ 的分布.

取定正整数 k, 我们来计算 $a_k = \mathbb{P}(Y_n \leqslant k)$. 由于 $X_1, \cdots, X_n$ 是互不相同的正整数, 故当 $k \leqslant n-1$ 时, 总有 $a_k = 0$. 当 $k \geqslant n$ 时, $Y_n \leqslant k$, 要求 $X_1, \cdots, X_n$ 都不超过 k, 即前 n 步选择的数都不超过 k. 注意到, 在第 j 次选取 X_j 后, 比 X_j 大的正整数的标签上的概率都会翻倍, 因此第 j 步选择的 X_j 不超过 k 的概率为

$$1 - 2^{j-1}\left(\frac{1}{2^{k+1}} + \frac{1}{2^{k+2}} + \cdots\right) = 1 - \frac{1}{2^{k-j+1}}.$$

于是前 n 次选中的数都不超过 k 的概率为

$$a_k = \mathbb{P}(Y_n \leqslant k) = \prod_{j=1}^{n}\left(1 - \frac{1}{2^{k-j+1}}\right) = \prod_{j=0}^{n-1}\left(1 - \frac{1}{2^{k-j}}\right).$$

于是

$$\mathbb{E}[Y_n] = 1 + \sum_{k=1}^{\infty}(1 - a_k) = n + \sum_{k=n}^{\infty}(1 - a_k).$$

要证明原等式, 接下来只需证明

$$
\begin{aligned}
&\sum_{k=n}^{\infty}(1-a_k)=\sum_{i=1}^{n}\frac{1}{2^i-1}\\
\iff &\sum_{k=n}^{\infty}\left(1-\prod_{j=0}^{n-1}\left(1-\frac{1}{2^{k-j}}\right)\right)=\sum_{i=1}^{n}\frac{1}{2^i-1}\\
\iff &\sum_{k=1}^{\infty}\left(1-\prod_{j=0}^{n-1}\left(1-\frac{1}{2^{k+n-1-j}}\right)\right)=\sum_{i=1}^{n}\frac{1}{2^i-1}\\
\iff &\sum_{k=1}^{\infty}\left(1-\prod_{j=0}^{n-1}\left(1-\frac{1}{2^{k+j}}\right)\right)=\sum_{i=1}^{n}\frac{1}{2^i-1}.
\end{aligned}
\tag{1}
$$

对 n 使用数学归纳法, 只需证明 (1) 在 n 和 $n+1$ 时作差的结果相等即可:

$$
\sum_{k=1}^{\infty}\frac{1}{2^{k+n}}\prod_{j=0}^{n-1}\left(1-\frac{1}{2^{k+j}}\right)=\frac{1}{2^{n+1}-1}.
\tag{2}
$$

等式 (2) 直接用裂项相消即可证明. 事实上, (2) 等价于

$$
\begin{aligned}
&\sum_{k=1}^{\infty}\left(\frac{2^{n+1}-1}{2^{k+n}}\prod_{j=0}^{n-1}\left(1-\frac{1}{2^{k+j}}\right)\right)=1\\
\iff &\sum_{k=1}^{\infty}\left(\left(\frac{1}{2^{k-1}}-\frac{1}{2^{k+n}}\right)\prod_{j=0}^{n-1}\left(1-\frac{1}{2^{k+j}}\right)\right)=1\\
\iff &\sum_{k=1}^{\infty}\left(\prod_{j=0}^{n}\left(1-\frac{1}{2^{k+j}}\right)-\prod_{j=-1}^{n-1}\left(1-\frac{1}{2^{k+j}}\right)\right)=1\\
\iff &\sum_{k=1}^{\infty}\left(\prod_{j=0}^{n}\left(1-\frac{1}{2^{k+j}}\right)-\prod_{j=0}^{n}\left(1-\frac{1}{2^{k-1+j}}\right)\right)=1\\
\iff &\lim_{k\to\infty}\prod_{j=0}^{n}\left(1-\frac{1}{2^{k+j}}\right)=1.
\end{aligned}
$$

对于给定的 n, 上面的极限显然成立. 因此原命题得证. ■

10. 对正整数 N, 定义 S_N 为使得 $(a^2+a)(b^2+b)$ 是完全平方数的正整数 $1 \leqslant a, b \leqslant N$ 的对数. 证明: 极限

$$\lim_{N\to\infty} \frac{S_N}{N}$$

存在, 并求其值.

证明 先介绍 Pell 方程的背景知识. 当 d 是非完全平方的正整数时, 整数方程

$$x^2 - dy^2 = 1$$

称为 Pell 方程. 除了平凡解 $(x, y) = (\pm 1, 0)$ 外, Pell 方程存在唯一的最小正整数解 (x_1, y_1), 称为基本解. Pell 方程的第 j 个正整数解由表达式

$$x_j + \sqrt{d} y_j = \left(x_1 + \sqrt{d} y_1\right)^j$$

给出. 下面假设 d 是 4 的倍数, 那么 x_1 一定是奇数. 并且对任意正整数 j, 有

$$x_j \geqslant x_1^j \geqslant 3^j.$$

回到原题. 对任意正整数 n, 定义 $\sigma(n)$ 是其非完全平方的部分:

$$\sigma(n) = p_1 \cdots p_k, \qquad 其中\ p_1, \cdots, p_k\ 是所有使得\ v_p(n)\ 为奇数的素数\ p.$$

于是, $(a^2+a)(b^2+b)$ 是完全平方数当且仅当 $\sigma(a^2+a) = \sigma(b^2+b)$. 这启发我们根据 $\sigma(n^2+n)$ 的值对所有正整数进行等价类的划分. 给定正整数 N 和非完全平方正整数 k, 定义

$$C_k(N) = \{1 \leqslant n \leqslant N : \sigma(n^2+n) = k\}, \qquad c_k(N) = \#C_k(N). \tag{1}$$

于是, $\{1, \cdots, N\} = \bigcup_k C_k(N)$, 从而有

$$N = \sum_k c_k(N), \qquad S_N = \sum_k c_k^2(N). \tag{2}$$

在这里, 对 k 求和的范围是所有非完全平方的正整数. 如果 $n \in C_k(N)$, 那么由 (1) 有

$$\sigma(n^2+n) = k,$$

进而存在正整数 m, 使得 $n^2+n = km^2$, 从而得到

$$(2n+1)^2 - 4km^2 = 1, \tag{3}$$

因此 $(2n+1, m)$ 是 Pell 方程 $x^2 - 4ky^2 = 1$ 的正整数解. 接下来证明两个引理:

引理 1 对任意 k, 有 $c_k(N) \leqslant \log N + 1$.

设 $(x_j, y_j)_{j=1}^{\infty}$ 是 Pell 方程 $x^2 - 4ky^2 = 1$ 的所有正整数解, 则易知 $x_j \geqslant 3^j$. 此时 $c_k(N)$ 计算 $x^2 - 4ky^2 = 1$ 的满足 $1 < x \leqslant 2N+1$ 的正整数解的数量, 因此

$$c_k(N) = \sum_{j:x_j \leqslant 2n+1} 1 \leqslant \sum_{j:3^j \leqslant 2n+1} 1 \leqslant \log_3(2N+1) \leqslant \log_3(3N) \leqslant \log N + 1.$$

引理 2 给定正整数 N, 使得 $c_k(N) > 1$ 的 k 不超过 $\sqrt{N}$ 个.

若 $c_k(N) > 1$, 则 Pell 方程 (3) 至少存在两组正整数解 $(n_1, m_1) < (n_2, m_2)$, 且 $n_1, n_2 \leqslant N$. 不妨设这两组解就是最小的两组解, 从而有

$$(2n_1+1)^2 \leqslant 2n_2 + 1 \leqslant 2N + 1 \Longrightarrow n_1 \leqslant \frac{\sqrt{2N+1} - 1}{2} \leqslant \sqrt{N}.$$

因此我们构造了一个映射: 如果某个 k 使得 $c_k(N) > 1$, 就能找到 Pell 方程 (3) 的最小解 (n_1, m_1), 且 $n_1 \leqslant \sqrt{N}$. 不仅如此, 这样的从 k 到 n_1 的映射一定是单射, 因为 $\sigma(n_1^2 + n_1) = k$. 于是, 使得 $c_k(N) > 1$ 的 k 一定不超过 $\sqrt{N}$ 个.

利用引理 1 和引理 2, 我们可以给出

$$\begin{aligned}
S_N &= \sum_k c_k^2(N) \\
&= \sum_{k:c_k(N)=1} c_k^2(N) + \sum_{k:c_k(N)>1} c_k^2(N) \\
&\leqslant \sum_{k:c_k(N)=1} 1 + (\log N + 1)^2 \sum_{k:c_k(N)>1} 1 \\
&\leqslant N + \sqrt{N}(\log N + 1)^2.
\end{aligned}$$

由于 $S_N \geqslant N$, 故 S_N/N 的极限为 1. ■