

color 解题报告

杭州外国语学校 陈立杰

Contents

1 题目大意	2
2 分析	2
3 80%的算法	2
4 100%的算法	3

1 题目大意

见同文件夹中的题目文件.

2 分析

如果我们想要知道当前一个状态的答案,必须要维护每个颜色的个数,这在 $n \leq 10^{10}$ 的情况下是不可能的.

答案等于,对所有使得某个颜色变成了 n 个的操作序列 $s, \sum \text{Prob}(s) \cdot \text{Length}(s)$. Prob表示该序列出现的概率,Length表示该序列的长度.

我们不妨逆向思考,假定我们知道了是哪个颜色变成了 n 个,来求这种情况下的答案和.

即求对所有使得该给定颜色变成了 n 个的操作序列 $s, \sum \text{Prob}(s) \cdot \text{Length}(s)$.

我们来分析一下如果这个颜色有 i 个,那么下一步的可能性.

- 该颜色损失一个的概率 $sub_i = \frac{i(n-i)}{n(n-1)}$
- 该颜色增加一个的概率 $add_i = \frac{i(n-i)}{n(n-1)}$
- 该颜色数量不变的概率 $stay_i = 1 - sub_i - add_i$

由于 $add_i = sub_i$,不妨令 $a_i = add_i = sub_i, b_i = stay_i$.

注意到这跟其他颜色的情况没有任何关系,只跟该颜色的个数有关.

3 80%的算法

令 f_i 表示 i 个颜色的答案.

那么有些人可能会列出

$$f_i = a_i(f_{i-1} + f_{i+1}) + b_i f_i + 1$$

这是错误的,我们注意到,我们枚举了下一步,就意味让所有的操作序列长度都+1了,但是因为我们只考虑使得该给定颜色变成 n 个的序列,所以实际上Length +1的只有那些序列,那么增加的也只有那些序列的Prob和,也就是该颜色最终能变成 n 个的概率.

不妨令 p_i 表示有 i 个的颜色,最终变成 n 个的概率.

我们可以列出方程 $p_i = a_i(p_{i-1} + p_{i+1}) + b_i p_i$

化简得 $p_i = \frac{p_{i+1} + p_{i-1}}{2}$.

带入 $p_0 = 0, p_n = 1$ 易解得 $p_i = \frac{i}{n}$.

那么正确的式子是

$$f_i = a_i(f_{i-1} + f_{i+1}) + b_i f_i + p_i$$

经过化简可以得到

$$f_{i+1} = 2f_i - f_{i-1} - \frac{n-1}{n-i}$$

我们注意到 $f_0 = 0, f_1 = \frac{(n-1)^2}{n}$.这个的证明较为繁琐,就略去不谈.

那么我们只需要通过一次简单的递推,就能算出所有 f_i 的值.

那么累加所有颜色的结果,就能得到答案.

4 100%的算法

考虑到 n 很大的情况,没有办法使用 $O(n)$ 的递推计算出 f_i 的值.

让我们仔细分析 f_i 这个函数

$$\begin{aligned}
 f_{i+1} &= 2f_i - f_{i-1} - \frac{n-1}{n-i} \\
 f_{i+1} - f_i &= f_i - f_{i-1} - \frac{n-1}{n-i} \\
 f_{i+1} - f_i &= f_1 - f_0 - \sum_{k=1}^i \frac{n-1}{n-k} \\
 f_{i+1} - f_i &= \frac{(n-1)^2}{n} - \sum_{k=1}^i \frac{n-1}{n-k} \\
 f_{i+1} &= (i+1) \frac{(n-1)^2}{n} - \sum_{j=0}^i \sum_{k=1}^j \frac{n-1}{n-k} \\
 f_{i+1} &= (i+1) \frac{(n-1)^2}{n} - \sum_{k=1}^i (i+1-k) \frac{n-1}{n-k}
 \end{aligned}$$

为了方便我们定义函数 g .

$$\begin{aligned}
 (n-1)g_{i+1} &= \sum_{k=1}^i (i+1-k) \frac{n-1}{n-k} \\
 g_{i+1} &= \sum_{k=1}^i (i+1-k) \frac{1}{n-k} \\
 g_{i+1} &= (i+1) \sum_{k=1}^i \frac{1}{n-k} - \sum_{k=1}^i k \frac{1}{n-k}
 \end{aligned}$$

我们再定义调和函数 H_i .

$$H_i = \sum_{k=1}^i \frac{1}{k}$$

$$g_{i+1} = (i+1)(H_{n-1} - H_{n-i-1}) - \sum_{k=1}^i (n - (n-k)) \frac{1}{n-k}$$

$$g_{i+1} = (i+1)(H_{n-1} - H_{n-i-1}) - \sum_{k=1}^i n \frac{1}{n-k} + i$$

$$g_{i+1} = (i+1-n)(H_{n-1} - H_{n-i-1}) + i$$

那么我们只要能快速计算 H_i 的值, f_i 的值也可以快速的计算了.

注意到

$$H_n \sim \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} \quad (1)$$

γ 在计算 $H_{n-1} - H_{n-i-1}$ 时消去了故可以不做考虑.

通过这个公式就能 $O(1)$ 计算出 H_n ,同时精度误差是 $O(\frac{1}{n^4})$ 的,不妨先预处理 $n \leq 10^6$ 的 H_n 值,对于 $n > 10^6$ 的情况,精度就非常高了.