《连连看》解题报告

浙江省余姚中学　曹立

1　题意

1.1　题目描述

有一个由 $1, 2, \cdots, n$ 各 2 个组成的随机可重排列 $a_1, a_2, \cdots, a_{2n}$（简记为 a），初始每个位置的值都为未知。考虑以下过程：计数器 cnt 初始为 0。每次随机选择一个未知位置，得到它的值 x 并标记该位置为已知。若没有其他已知位置值为 x，则再随机选择一个位置，得到它的值 y 并标记该位置为已知。如果 $x \neq y$ 则计数器加一。

记 $E(n, m)$ 表示长为 $2(n+m)$ 的满足上述条件的随机排列，其中有 m 个位置已知（且值各不相同），这时执行以上策略直到所有位置均变为已知，最终的 $E[cnt]$。

给定系数 $p_0, p_1, \cdots, p_{n-1}, q_0, q_1, \cdots, q_{m-1}$，求：

$$\left(\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\binom{2i+j}{i}p_iq_jE(i,j)\right) \bmod 998244353$$

1.2　数据范围

子任务编号	$n \leq$	$m \leq$	特殊性质	分值
1	10	1	否	5
2	17	1	否	5
3	2000	2000	否	10
4	5×10^4	5×10^4	是	30
5	2.5×10^5	1	否	10
6	10^5	10^5	否	30
7	2.5×10^5	2.5×10^5	否	10

特殊性质：p 中非零项数与 q 中非零项数的乘积至多为 2500。

对于所有数据，$1 \leq n, m \leq 2.5 \times 10^5, 0 \leq p_i, q_i < 998244353$。

时间限制 2s，空间限制 512MB。

2 解题过程

2.1 初步转化

先考虑求单个 $E(n,m)$。为了方便分析，可以将题意等价转化如下：

可重排列 a 中有 $1\sim n$ 各 2 个，$n+1\sim n+m$ 各 1 个。设定一个指针 r 初始为 1。循环执行如下步骤直到 $r>2n+m$：

case 1. 若 $a_r>n$ 或 a_r 值在 $a_{1\sim r-1}$ 中出现过，则 $r\xleftarrow{+}1$；

case 2. 否则，若 $a_r=a_{r+1}$，则 $r\xleftarrow{+}2$；

case 3. 否则，$r\xleftarrow{+}2, cnt\xleftarrow{+}1$。

$E(n,m)$ 即为所有满足条件的排列 a 对应的 cnt 平均值。

2.2 算法一

考虑搜索。对于只需求 $E(n,0)$ 的情况，直接枚举 a 可以做到搜索量 $(2n-1)!!$，可通过子任务 1。注意到决定着 cnt 的是对 a 作如下映射得到的长与 a 相同的 `01` 序列 b：对于 $a_i=a_j(i<j)$，置 $b_i=\texttt{0}, b_j=\texttt{1}$。因此只需搜索序列 b，并用乘法原理算出可能的原序列 a 数量即可。搜索量为卡特兰数 C_n，可通过子任务 1,2。

2.3 算法二

依照题意可以写出递推式（边界情况略去）：

$$E(n,m)=\frac{1}{2n+m}E(n,m-1)+\frac{2n}{2n+m}\frac{(m+1)E(n-1,m)+(2n-2)E(n-2,m+2)+2n+m-2}{2n+m-1}$$

时间复杂度为 $\mathrm{O}((n+m)^2)$，可通过子任务 1,2,3。

2.4 算法三

在进一步思考当中，首先发现上文中 case 2 妨碍分析，考虑先将它视作也要使 $cnt\xleftarrow{+}1$，然后再减掉。令 $s_{n,m}$ 表示，对于每一个满足条件的排列 a，分别设定指针 r 初始为 1 并循环进行如下步骤直到 $r>2n+m$ 得到的 cnt 之和：

case 1. 若 $a_r>n$ 或 a_r 值在 $a_{1\sim r-1}$ 中出现过，则 $r\xleftarrow{+}1, cnt\xleftarrow{+}1$；

case 2. 否则，$r\xleftarrow{+}2, cnt\xleftarrow{+}1$。

记 a 的数量为 $t_{n,m}=(2n+m)!/2^n$，那么有关系式：

$$E(n,m)=\frac{(2n+m)\cdot t_{n,m}-s_{n,m}}{t_{n,m}}-\frac{n\cdot s_{n-1,m}+t_{n-1,m}}{t_{n,m}} \tag{1}$$

第一项是将 $a_r=a_{r+1}$ 视作也要加 1 的情况下的答案。第二项是多算的部分，意义是将一组 $a_r=a_{r+1}$ 插入 $n-1$ 情况序列的某个空隙中。

现在从算法一中序列 b 的角度来看 $s_{n,m}$ 的定义（对于 a 中 $>n$ 的数，置 b 中对应位置为 `1`）。考虑指针 r 在 b 上运行的过程，每一轮若 $b_r=\texttt{0}$，r 就加 2，否则加 1。

例如 $n=7, m=2, p=[1,2,8,3,4,1,5,9,6,4,2,7,7,6,3,5]$，那么 b 为：

$$\underline{\texttt{00}}\,\underline{\texttt{1}}\,\underline{\texttt{00}}\,\underline{\texttt{1}}\,\underline{\texttt{01}}\,\underline{\texttt{01}}\,\underline{\texttt{1}}\,\underline{\texttt{01}}\,\underline{\texttt{1}}\,\underline{\texttt{1}}\,\underline{\texttt{1}}$$

其中下划线表示每一轮经过的部分。

对于固定的 n, m，设经过 `00` 总次数为 A，`01` 总次数为 B，`1` 总次数为 C。有关系式：

$$\begin{cases} 2A+B &= n \cdot t_{n,m} \\ B+C &= (n+m) \cdot t_{n,m} \\ A+B+C &= s_{n,m} \end{cases}$$

可得：

$$s_{n,m} = \frac{(3n+2m) \cdot t_{n,m} - B}{2}$$

对于一个 b 中的 `1`，若它与上一个 `1` 之间 `0` 的个数（若本身是第一个 `1`，那就是前面所有的 `0` 的个数）为奇数，那它就对 B 贡献 `1`，否则对 C 贡献 `1`。于是令 cm_i 表示所有 a 对应的 b 中长为 i 的极长连续 `0` 的出现次数，则：

$$B = \sum_{1 \le i \le n, 2 \nmid i} cm_i$$

cm_i 不好直接算，令 cl_i 表示所有 a 对应的 b 中连续 i 个 `0` 的出现次数。有：

$$cm_i = cl_i - 2cl_{i+1} + cl_{i+2}$$

计算 cl_i 可以考虑将连续 i 个 `0` 视作整体，与其他数共同参与排列：

$$cl_i = \frac{(2n+m-i+1)!n^{\underline{i}}}{2^{n-i}(i+1)}$$

综合上述四式可得：

$$s_{n,m} = \sum_{i=0}^{n} \frac{(-1)^i (2n+m-i+1)!n^{\underline{i}}}{2^{n-i}(i+1)} - t_{n,m}$$

令：

$$f_{n,m} = \sum_{i=0}^{n} \frac{(-2)^i (2n+m-i+1)!}{(i+1)(n-i)!}$$

代入 (1) 式得：

$$E(n,m) = 2n+m+1 - \frac{n!}{(2n+m)!}(f_{n,m} + 2f_{n-1,m})$$

可以 $\mathrm{O}(n)$ 计算单项，结合算法二可通过子任务 1,2,3,4。

2.5 算法四

考虑一次性对于所有 i 计算 $f_{i,0}$。注意到：

$$\frac{f_{n,0}}{(n+1)!} = [x^{-1}]\frac{\ln(1+2x)}{2(x-x^2)^{n+2}}$$

套用另类拉格朗日反演，其中 $F(x), G(x)$ 常数项均为 0 且 $G(F(x)) = x$：

$$[x^{-1}]H(x)F'(x)F(x)^{-n-1} = [x^n]H(G(x))$$

代入 $F(x)=x-x^2, G(x)=(1-\sqrt{1-4x})/2, H(x)=\ln(1+2x)/(2(1-2x))$ 得：

$$[x^{-1}]\frac{\ln(1+2x)}{2(x-x^2)^{n+1}}=[x^n]\frac{\ln(2-\sqrt{1-4x})}{2\sqrt{1-4x}}$$

可以直接使用多项式开根、求对数等可以做到大常数 $\mathrm{O}(n\log n)$。也可以拆掉 ln（求导再积分），再利用广义二项式定理拆开 $\sqrt{1-4x}$ 和 $1/\sqrt{1-4x}$，做到仅用一次卷积。

事实上可以 $\mathrm{O}(n)$ 线性递推，令 $F(x)$ 为上式右侧的多项式，有：

$$F'(x)=\frac{2F(x)}{1-4x}+\frac{2+\sqrt{1-4x}}{(1-4x)(3+4x)}$$

使用三种方法中的任意一种即可通过子任务 5。

2.6 算法五

同样写出 $f_{n,m}$ 公式的生成函数表达形式：

$$\frac{f_{n,m}}{(n+m+1)!}=[x^n]\frac{\ln(1+2x)}{2x(1-x)^{n+m+2}}$$

代入原问题式（边界 $p_n=0$）：

$$\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\binom{2i+j}{i}p_iq_jE(i,j)=\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}[x^i]\frac{(i+j+1)\left(1-\frac{\ln(1+2x)}{2x}\right)p_iq_j-\frac{\ln(1+2x)}{x}p_{i+1}q_j}{(1-x)^{i+j+2}}$$

这可以化成三个形式类似的式子之和，形式如下：

$$[x^{-1}]P\left(\frac{1}{x-x^2}\right)Q\left(\frac{1}{1-x}\right)C(x)$$

其中 P 和 Q 的系数直接由序列 p 和 q 确定，C 是固定的多项式。$Q(1/(1-x))$ 可以直接一次差卷积得到，P 的复合需要多步处理，以 $P(x)=\sum_{i=0}^{n-1}p_ix^{i+1}$ 为例：

$$\begin{aligned}P\left(\frac{1}{x-x^2}\right)&=\frac{\sum_{i=0}^{n-1}p_i(x-x^2)^{n-1-i}}{(x-x^2)^n}\\&=\frac{\bar{P}(x-x^2)}{(1-x)^nx^n}\\&=\frac{\bar{P}\left(\frac{1}{4}-\left(\frac{1}{2}-x\right)^2\right)}{(1-x)^nx^n}\end{aligned}$$

其中 $\bar{P}(x)=\sum_{i=0}^{n-1}p_{n-1-i}x^i$。共需要两次多项式复合一次函数和一次和卷积。

总体来说，需要计算三个 P 复合 $1/(x-x^2)$，两个 Q 复合 $1/(1-x)$，额外再进行三次和卷积。理论上，通过分别分析多个 P 和 Q 的关系，可以达到分别只复合一次，额外再进行两次和卷积。实际数据范围并没有卡这部分的常数。

另一个思路是一次性对所有 i 求出 $g_i=\sum_{j=0}^{m-1}\binom{2i+j}{i}q_jE(i,j)$。先同样卷积算出只与 j 有关的部分，然后仿照算法四套用拉格朗日反演，转化出的式子形如：

$$g_i=[x^i]R\left(\frac{1-\sqrt{1-4x}}{2}\right)C(x)$$

其中 R 只与序列 q 有关，C 是固定的多项式。R 的复合也可以通过分多步（复合 $\sqrt{1-4x}$ 要分开处理奇偶次项）完成，这里不再赘述细节。

使用两种方法中的任意一种即可以 $\mathrm{O}(n\log n)$ 的时间复杂度通过所有子任务，如果实现较差或使用 $\mathrm{O}(n\log^2 n)$ 的暴力复合可能无法通过子任务 7。

3 小结

本题为原创题，灵感来源为 LVII St. Petersburg State University Championship（笔者在 The 2nd Universal Cup Stage 2: SPb 见到）的 L 题 "Memo" Game With a Hint。原题要求的是构造失败（miss）次数较少的翻牌方案，题面中提到 $E(25,0) \approx 14.83$，但笔者发现该值没有简单的低于平方的求法，于是就产生了进一步推导的兴趣。恰巧后面的分析过程比较有意思，所以就出成了题。

本题最巧妙也最困难的点是找到一个能利用乘法原理简单地刻画总循环次数的突破口。由于每个 `01` 序列 b 对应的原排列 a 数量各有不同，故很多直接基于序列 b 的计数就无法进行。笔者也尝试了很多方向，最后回到观察序列本身，寻找对所求值的更简洁的等价表述方式（即"`1` 前连续 `0` 数量的奇偶性"这个部分），才找到了一个可行方向。

命题过程中笔者与深圳市高级中学的林睿同学有讨论。题目最初时，满分所求的是每个 $i \le 10^7$ 的 $E(i,0)$，即正解为子任务 5 的线性做法。后来林睿同学提出可以直接利用 $E(n,m)$ 的前几项解出 $E(n,m)$ 关于 n 或 m 的整式递推式（$E(n,m)$ 存在仅与 $E(n-1,m), E(n-2,m), E(n-3,m)$ 或 $E(n,m-1), E(n,m-2), E(n,m-3)$ 有关的递推式，且系数多项式次数不大），于是可以直接绕过公式推导。

为防止该类做法直接通过本题，笔者只能在输入中给定系数，保证得到非递推的求和式才能往下推，最后用多项式相关算法求解。本题前半部分组合意义的分析和后半部分多项式推导略显割裂，若给大家带来了不好的做题体验，在这里致歉。

4 参考资料

[1] NaCly_Fish,https://www.luogu.com.cn/blog/NaCly-Fish-blog/a-classical-problem