Petrozavodsk Summer 2021 Day 2 — Problem G: Permutation CFG

1 题目大意

原题链接：QOJ 题面

给定一个 $1\sim n$ 的排列 π。把每个整数 k 视为一个“可展开的数”，其展开规则为：将 k 替换为集合 $\{1,2,\ldots,k\}$ 按照 π 在该集合上的相对顺序排列得到的长度为 k 的序列。

记号“$\Rightarrow$”表示对一个数执行一次“展开”操作。

例如 $n=4,\ \pi=(1,4,3,2)$ 时

$$1\Rightarrow[1],\quad 2\Rightarrow[1,2],\quad 3\Rightarrow[1,3,2],\quad 4\Rightarrow[1,4,3,2].$$

从单个起始数 n 开始，迭代应用上述规则 s 次，得到一个整数序列。

给出 q 个询问 (k,a)，每次问：最终序列的前 a 个数中，整数 k 出现了多少次？

输入 第一行 n,s,q。第二行给出排列 π。接下来 q 行，每行 k,a。

输出 按输入顺序依次输出每个询问的答案。

2 数据范围

- $2\le n\le 10^5$，$1\le s\le 5$，$1\le q\le 2\times 10^5$；
- $1\le k\le n$，$1\le a\le 10^9$（保证不超过最终长度）；
- 时间限制 7 秒，内存限制 1024 MB。

3 记号与基本性质

记 $E_t(x)$ 为从单个起始数 x 出发进行 t 次展开后得到的序列，

$L_t(x)=|E_t(x)|$ 为其长度。定义：对 $x\ge 1$，$E_0(x)=[x]$，故 $L_0(x)=1$。

约定对一切 $t\ge 0$ 有 $E_t(0)=[]$ 为空序列，于是 $L_t(0)=0$。

记 $\mathbf{1}_{(\cdot)}$ 为指示函数，谓词为真取 1，否则取 0。

记号“$\|$”表示序列拼接：对任意序列 A,B，$A\|B$ 为先接 A 再接 B 的序列；并定义

$$\Big\|_{i=1}^{m}S_i:=S_1\|S_2\|\cdots\|S_m,$$

其中“$\|$”满足结合律。

对 $k\ge 1$，令 $\sigma^{(k)}$ 表示“把排列 π 限制到集合 $\{1,\ldots,k\}$ 并保持相对次序”的排列；于是

$$E_t(k) = E_{t-1}(\sigma_1^{(k)}) \parallel E_{t-1}(\sigma_2^{(k)}) \parallel \cdots \parallel E_{t-1}(\sigma_k^{(k)}),$$

特别地 $E_t(n) = \|_{i=1}^{n} E_{t-1}(\pi_i)$。

引理 1 (长度递推). 对任意 $t \geq 1$, $x \geq 1$，有递推:

$$L_t(x) = L_t(x-1) + L_{t-1}(x).$$

边界条件为:

$$L_t(0) = 0 \ (t \geq 0), \qquad L_0(x) = 1 \ (x \geq 1).$$

从而对一切 $t \geq 0$, $x \geq 1$，

$$L_t(x) = \binom{x+t-1}{t}.$$

证明. 由定义与拼接可得:

$$E_t(x) = \|_{i=1}^{x} E_{t-1}(\sigma_i^{(x)}),$$

从而

$$L_t(x) = \sum_{i=1}^{x} L_{t-1}(\sigma_i^{(x)}),$$

其中 $\{\sigma_i^{(x)}\}$ 是 $\{1, \ldots, x\}$ 的重排，故

$$\sum_{i=1}^{x} L_{t-1}(\sigma_i^{(x)}) = \sum_{i=1}^{x} L_{t-1}(i).$$

这与所给递推、边界等价：将上式与 $x-1$ 的形式相减得 $L_t(x) = L_t(x-1) + L_{t-1}(x)$；显然 $L_t(0) = 0$，且 $t = 0$ 时 $L_0(x) = 1$。

由 $L_t(0) = 0$ 与递推在 $x = 1$ 上的代入得 $L_t(1) = L_{t-1}(1)$，配合 $L_0(1) = 1$ 可知 $L_t(1) = 1$。

设

$$F(t, x) = \binom{x+t-1}{t}.$$

下面对 t 归纳证明 $L_t(x) = F(t, x)$ 对所有 $x \geq 1$ 成立。

基例 $t = 0$: $L_0(x) = 1 = F(0, x)$。

归纳步: 设对 $t-1$ 已成立。利用递推，

$$L_t(x) = L_t(x-1) + L_{t-1}(x) = F(t, x-1) + F(t-1, x).$$

将 F 写成组合数：

$$F(t,x-1)+F(t-1,x)=\binom{x+t-2}{t}+\binom{x+t-2}{t-1}.$$

由帕斯卡恒等式 $\binom{a}{b}=\binom{a-1}{b}+\binom{a-1}{b-1}$（取 $a=x+t-1,\ b=t$）可得

$$\binom{x+t-2}{t}+\binom{x+t-2}{t-1}=\binom{x+t-1}{t}.$$

再将结果写回 F，得

$$F(t,x-1)+F(t-1,x)=F(t,x).$$

故 $L_t(x)=F(t,x)$。 □

引理 2 (目标值出现次数)**.** 设 $t\ge 0,\ x\ge 0,\ y\ge 1$。记 $C_t(x\leftarrow y)$ 为 $E_t(x)$ 中数字 y 的出现次数，约定对一切 $t\ge 0$ 有 $C_t(0\leftarrow y)=0$。则

$$C_t(x\leftarrow y)=\begin{cases}\mathbf{1}_{(x=y)}, & t=0,\\ 0, & t\ge 1 \text{ 且 } x<y,\\ L_{t-1}(x-y+1)=\binom{x-y+t-1}{t-1}, & t\ge 1 \text{ 且 } x\ge y.\end{cases}$$

证明. 由拼接：

$$E_t(x)=\|_{i=1}^{x}E_{t-1}(\sigma_i^{(x)}),$$

故

$$C_t(x\leftarrow y)=\sum_{i=1}^{x}C_{t-1}(\sigma_i^{(x)}\leftarrow y)=\sum_{i=1}^{x}C_{t-1}(i\leftarrow y).$$

当 $i<y$ 时，$E_{t-1}(i)$ 的元素均不超过 i，因此 $C_{t-1}(i\leftarrow y)=0$，故下限可改为 y。于是对 $x\ge y$ 有递推

$$C_t(x\leftarrow y)=C_t(x-1\leftarrow y)+C_{t-1}(x\leftarrow y),$$

且边界为

$$C_t(y-1\leftarrow y)=0\ (t\ge 0),\qquad C_0(x\leftarrow y)=\mathbf{1}_{(x=y)}.$$

设

$$G_t(x)=\begin{cases}\mathbf{1}_{(x=y)}, & t=0,\\ 0, & t\ge 1 \text{ 且 } x<y,\\ \binom{x-y+t-1}{t-1}, & t\ge 1 \text{ 且 } x\ge y,\end{cases}$$

对 t 归纳可得 $C_t=G_t$：基例 $t=0$ 成立；归纳步对 $t\ge 1$ 与 $x\ge y$，

$$G_t(x-1)+G_{t-1}(x)=\binom{x-y+t-2}{t-1}+\binom{x-y+t-2}{t-2}=\binom{x-y+t-1}{t-1}=G_t(x).$$

从而结论成立。 □

4 详细做法

约定 任意一层递归中，所有“线段树上二分”“区间求和”等操作，均在**该层起始数**所对应的数值前缀版本 R_{x_0} 上进行。这里 R_x 定义为：仅把排列 π 中“数值 $\le x$” 的元素，按其原排列位置插入线段树（其余位置权为 0）。**特别地，约定 R_0 为全零版本**（所有位置权值为 0）。这与从起始数 x 展开时，一级只会出现 $\le x$ 的数相一致。

4.1 结构

设排列位置为叶节点索引，叶上存放其值 v。在线段树每个节点维护两类可加型向量：

$$\big(L_0(v), L_1(v), \dots, L_s(v)\big), \qquad \big(v^0, v^1, \dots, v^s\big),$$

其中 $L_t(v) = \binom{v+t-1}{t}$（见引理 1）。按数值从小到大（$1 \to n$）依次插入，得到持久化版本链 $\{R_x\}_{x=0}^n$。

基于版本 R_x，对任意区间 $[l,r]$ 与整数 $d \ge 0$、$u \in \{0,1,\dots,s\}$，定义

$$\mathrm{PowSum}_d(R_x;[l,r]) := \sum_{i=l}^{r} \mathbf{1}_{(\pi_i \le x)}\,(\pi_i)^d, \qquad \mathrm{LenSum}_u(R_x;[l,r]) := \sum_{i=l}^{r} \mathbf{1}_{(\pi_i \le x)}\,L_u(\pi_i),$$

二者分别由上述两类向量支持加法维护与区间查询。

4.2 状态

定义：

$$\mathrm{Ans}_t(x_0, m; k) = \text{“}E_t(x_0)\text{ 的前 } m \text{ 个元素中，数字 } k \text{ 的出现次数”}$$

所求为 $\mathrm{Ans}_s(n, a; k)$。

4.3 二分

对 $t \ge 1$，令

$$w_i^{(x_0)} := \mathbf{1}_{(\pi_i \le x_0)} \cdot L_{t-1}(\pi_i), \qquad P_0 := 0, \quad P_j := \sum_{i=1}^{j} w_i^{(x_0)} = \mathrm{LenSum}_{t-1}(R_{x_0};[1,j])\ (j \ge 1).$$

在版本 R_{x_0} 上，对权数组 $\{w_i^{(x_0)}\}$ 的前缀和数组 $\{P_i\}$ 做二分，找最小位置

$$\mathrm{pos}\text{ 使得 } P_{\mathrm{pos}} > m.$$

若不存在，则不存在残块，查询的是整个展开后的序列；此时取

$$p = n, \quad r = 0, \quad x_{\mathrm{next}} = 0.$$

若存在，则令

$$p = \mathrm{pos} - 1, \qquad r = m - P_{\mathrm{pos}-1}, \qquad x_{\mathrm{next}} = \pi_{\mathrm{pos}}.$$

引理 3 (分块唯一性). 对任意固定的 $t \geq 1, x_0 \geq 1$ 与 $m \geq 0$，上述定义的 $w_i^{(x_0)} \geq 0$ 使得前缀和数组 P_j 非减。若存在 pos 使 $P_{\text{pos}} > m$，则必有 $w_{\text{pos}}^{(x_0)} > 0$ 且 $0 \leq r < w_{\text{pos}}^{(x_0)}$。此时由 (p, r) 给出的"$[1, p]$ 内完整块 + 第 pos 块的长度为 r 的前缀"分解是唯一的；若不存在 pos，则唯一地取 $p = n$, $r = 0$, $x_{\text{next}} = 0$。

证明. 由 pos 的最小性有 $P_{\text{pos}-1} \leq m < P_{\text{pos}} = P_{\text{pos}-1} + w_{\text{pos}}^{(x_0)}$，故 $w_{\text{pos}}^{(x_0)} > 0$ 且 $r = m - P_{\text{pos}-1}$ 满足 $0 \leq r < w_{\text{pos}}^{(x_0)}$。设存在另一分解 (p', r') 对应某 pos'。若 $\text{pos}' \leq \text{pos} - 1$，则 $P_{\text{pos}'} \leq P_{\text{pos}-1} \leq m$，与"第 pos' 块被截断"矛盾；若 $\text{pos}' \geq \text{pos} + 1$，则 $P_{\text{pos}-1} < m < P_{\text{pos}} \leq P_{\text{pos}'-1}$，与"前 $\text{pos}' - 1$ 块全取完"矛盾。故 $\text{pos}' = \text{pos}$ 且由 $m = P_{\text{pos}-1} + r = P_{\text{pos}-1} + r'$ 得 $r' = r$。唯一性成立。 □

4.4 完整块

由引理 2，对任意 $t \geq 1$ 与 $x \geq k$，

$$C_t(x \leftarrow k) = \binom{x - k + t - 1}{t - 1} = \sum_{i=0}^{t-1} c_i \, x^i,$$

其中系数 c_i 仅依赖于 t, k。记

$$S_{p,k}^{(x_0)} = \left\{ x \;\middle|\; \text{其在 } \pi \text{ 中的位置} \leq p, \quad k \leq x \leq x_0 \right\},$$

则位置 $[1, p]$ 内完整块的总贡献为

$$\sum_{x \in S_{p,k}^{(x_0)}} C_t(x \leftarrow k) = \sum_{i=0}^{t-1} c_i \sum_{x \in S_{p,k}^{(x_0)}} x^i.$$

为避免在 R_{x_0} 上查询时混入超出阈值的元素，查询时对同一查询参数与同一区间取两个版本的差，即

$$\sum_{x \in S_{p,k}^{(x_0)}} x^i = \text{PowSum}_i(R_{x_0}; [1, p]) - \text{PowSum}_i(R_{k-1}; [1, p]).$$

4.5 残块

若存在残块，还需统计 $E_{t-1}(x_{\text{next}})$ 的前 r 个元素中 k 的出现次数。此时递归到

$$\text{Ans}_{t-1}(x_{\text{next}}, r; k),$$

若 $x_{\text{next}} < k$ 或 $r = 0$ 则该项为 0。若不存在残块，则无此项。

4.6 递推

综上，对 $t \geq 1$ 有

$$\boxed{\text{Ans}_t(x_0, m; k) = \sum_{x \in S_{p,k}^{(x_0)}} C_t(x \leftarrow k) \; + \; \mathbf{1}_{(r > 0 \text{ 且 } x_{\text{next}} \geq k)} \cdot \text{Ans}_{t-1}(x_{\text{next}}, r; k)}$$

其中 p, r, x_{next} 依上述规则确定。

4.7 边界

$$\mathrm{Ans}_0(x_0,m;k)=\begin{cases}1, & m\ge 1 \text{ 且 } x_0=k,\\ 0, & \text{否则},\end{cases}$$

若进入某层时 $k>x_0$ 或 $r=0$，则直接返回 0。

5 正确性证明

引理 4 (块求和)**.** 设 $t\ge 1$、$S_{p,k}^{(x_0)}=\{x\mid \text{其在排列中的位置}\le p,\quad k\le x\le x_0\}$. 则

$$\sum_{x\in S_{p,k}^{(x_0)}} C_t(x\leftarrow k)=\sum_{i=0}^{t-1} c_i \sum_{x\in S_{p,k}^{(x_0)}} x^i,$$

其中 c_i 由恒等式 $\binom{x-k+t-1}{t-1}=\sum_{i=0}^{t-1} c_i x^i$ 给出。集合 $S_{p,k}^{(x_0)}$ 的逐幂和可通过“版本 R_{x_0} 与版本 R_{k-1} 的差分”在区间 $[1,p]$ 上实现，即 $\mathrm{PowSum}_i(R_{x_0};[1,p])-\mathrm{PowSum}_i(R_{k-1};[1,p])$。

定理 1 (算法正确性)**.** 对任意 $t\ge 1$、$x_0\ge 1$、前缀长度 $m\ge 0$ 与目标数 k，按引理 3 的分块分解，恒有

$$\mathrm{Ans}_t(x_0,m;k)=\sum_{x\in S_{p,k}^{(x_0)}} C_t(x\leftarrow k)\;+\;\mathbf{1}_{(r>0 \text{ 且 } x_{\mathrm{next}}\ge k)}\cdot \mathrm{Ans}_{t-1}(x_{\mathrm{next}},r;k),$$

递归边界为 $t=0$ 或早停情形 $k>x_0$ 或 $r=0$。与定义一致，故成立。特别地，最终答案为 $\mathrm{Ans}_s(n,a;k)$。

6 复杂度分析

构建可持久化结构时，每次插入在 $\mathcal{O}(\log n)$ 条路径上维护 $\Theta(s)$ 个分量，故

$$\text{建树时间 } \mathcal{O}(n\,s\log n),\qquad \text{空间 } \mathcal{O}(n\,s\log n).$$

单次询问：

- 线段树上二分：每层一次 $\mathcal{O}(\log n)$，共 s 层，合计 $\mathcal{O}(s\log n)$；
- 逐幂区间和：每层查询 $\Theta(s)$ 个分量，$\mathcal{O}(s\log n)$；乘以 s 层为 $\mathcal{O}(s^2\log n)$；
- 多项式系数：每层 $\mathcal{O}(s^2)$，共 s 层为 $\mathcal{O}(s^3)$。

总时间

$$\mathcal{O}\big(n\,s\log n+q(s\log n+s^2\log n+s^3)\big).$$

7 参考资料

1. OI-Wiki, 排列组合, https://oi-wiki.org/math/combinatorics/combination.
2. OI-Wiki, 线段树, https://oi-wiki.org/ds/seg.
3. OI-Wiki, 可持久化线段树, https://oi-wiki.org/ds/persistent-seg.
4. Wikipedia, *Pascal's rule*, https://en.wikipedia.org/wiki/Pascal%27s_rule.
5. Wikipedia, *Binomial coefficient*, https://en.wikipedia.org/wiki/Binomial_coefficient.