《数列重排》命题报告

重庆市巴蜀中学 郭雨豪

2021 年 10 月 3 日

目录

1 题目大意 **3**

2 数据范围 **4**

3 解题过程 **5**

3.1 简单做法 5

3.1.1 暴力做法 5

3.1.2 针对子任务 4 的做法 5

3.2 题目分析 5

3.2.1 初步分析 5

3.2.2 针对子任务 5 的做法 5

3.2.3 进一步分析 6

3.2.4 针对子任务 6 的做法 7

3.3 标准解法 7

3.3.1 标准解法 1 7

3.3.2 标准解法 2 8

3.4 总结 8

4 参考文献 **8**

1 题目大意

定义一个数列区间的 mex 为区间中最小的没有出现过的自然数，对于一个数列 p，定义 $g(p,k)$ 为 p 中 $\text{mex} \ge k$ 的子区间数量。

给定 n 个小于 m 的自然数和一个区间 $[l,r]$，令 $f(k)$ 表示这 n 个数构成的数列 p 可以通过重排列得到的所有的 p' 中 $g(p',k)$ 的最大值，对于每一个 $k \in [l,r]$，求出 $f(k)$。

令 a_i 表示数字 i 出现的次数，保证存在正整数 X，使得 $\forall\, i < m, a_i \in \{X, X+1\}$。

2 数据范围

本题采用子任务评测。

<table>
<tr><th>子任务编号</th><th>分值</th><th>n</th><th>m</th><th>特殊性质</th></tr>
<tr><td>1</td><td>5</td><td>≤ 9</td><td>≤ 9</td><td rowspan="3">无</td></tr>
<tr><td>2</td><td rowspan="2">15</td><td>≤ 200</td><td>≤ 200</td></tr>
<tr><td>3</td><td>$\le 5 \times 10^3$</td><td>$\le 5 \times 10^3$</td></tr>
<tr><td>4</td><td>5</td><td rowspan="6">$\le 10^9$</td><td>≤ 2</td><td>$l = 0, r = 1$</td></tr>
<tr><td>5</td><td rowspan="2">10</td><td rowspan="3">$\le 10^6$</td><td>$l = m, r = m$</td></tr>
<tr><td>6</td><td>$a_i = 1$</td></tr>
<tr><td>7</td><td rowspan="2">15</td><td>$r - l + 1 \le 10^4$</td></tr>
<tr><td>8</td><td>$\le 2 \times 10^6$</td><td rowspan="2">无</td></tr>
<tr><td>9</td><td>10</td><td>$\le 10^7$</td></tr>
</table>

对于所有数据，满足 $1 \le n \le 10^9, 1 \le m \le 10^7, 0 \le l, r \le m$。

时间限制：$600ms$

空间限制：$256MB$

3 解题过程

令 $S = r - l + 1$。

3.1 简单做法

3.1.1 暴力做法

$O(n!)$ 枚举最终所有数的排列，求出每个区间的 mex，时间复杂度 $O(n!n^2)$。

也可以枚举 k，在求区间 mex 时使用双指针维护合法区间，时间复杂度为 $O(n!nS)$

两种算法或一些复杂度更高但常数更小的算法均可以通过子任务 1，期望得分 5。

3.1.2 针对子任务 4 的做法

子任务 4 中，数只包含 $0,1$，并且 $k \in \{0,1\}$，当 $k=0$ 时任何一个重排列的任何一个区间都是合法的，当 $k=1$ 时可以容斥，看作最小化全 1 的连续段个数，根据均值不等式可以转化为将所有的 1 平均的分配在 0 之中，在题目的条件下一定每一段长度均为 1，时间复杂度 $O(1)$，结合子任务 1 的做法可获得 10 分。

3.2 题目分析

3.2.1 初步分析

可以发现，当 k 固定时，所有 $\geq k$ 是等价的，所有 $< k$ 的数数字之间是等价的，这启示我们可以分为两部分考虑，在子任务 5 中，$k = m$，也就是所有数 $< m$，可以从这里入手。

3.2.2 针对子任务 5 的做法

可以直接构造，构造一个 $[0, m-1]$ 的排列，这个排列中出现次数为 $X+1$ 的数靠前，出现次数为 X 的数靠后，将 X 个这样的排列拼接在一起，最后会剩下一些出现次数为 $X+1$ 的数，将这些数按照最初排列中的顺序依次排列在这个数列的末尾。容易发现，此时任意一个长度 $\geq m$ 的区间 mex $\geq m$，无论怎样排列不可能更优，时间复杂度 $O(m)$，结合之前的算法期望得分 20 分。

3.2.3 进一步分析

子任务 5 的做法已经是正解很重要的一部分，但现在我们并不能证明子任务 5 的构造拓展到一般情况时的正确性，这个正确性可以从以下的一些基于贪心的性质中推出。

我们将每一个 $\geq k$ 的数一个视为黑球，其他的每一个数视为一个白球，白球上写着球所代表的数，黑球完全等价，我们称两个球之间有贡献，当且仅当这两个球之间（包括边界）每一个编号的白球都出现，最后的答案即为有贡献的球对数。

性质 1　若一个黑球和最靠左的球无贡献，可以将这个黑球移至最靠左的位置，若一个黑球和最靠右的球无贡献，可以将这个黑球移至最靠右的位置。

证明 1　黑球的移动不会影响其他球对是否产生贡献，若这个黑球和最靠左的位置无贡献，一定和其他左方的球无贡献，但移至更靠左的位置时可能和一些在右方的点产生贡献，所以这样的移动一定不劣，向右移同理。

性质 2　若两个黑球之间没有贡献，则可以将一个黑球移至另一个黑球的位置。

证明 2　黑球的移动不会影响其他球对是否产生贡献，对于两个黑球，分别计算与其有贡献的球的个数，将小的一方移动到大的一方一定更优。

接下来考虑白球的变化，我们还是将其看做一些白球构成了若干个排列加上最后不到整个的一个排列，其他的白球看成散点，若某一次存在一个可以移动至当前排列末尾的白球则移动，但仍然保留在排列末尾的黑色球连续段在排列末尾，对白球的贡献可以抵消，对黑球的贡献会少去对最后一个连续段的黑球的贡献（根据性质 1 不会损失其它贡献），但由于排列长度增加，会有一个新的位置与末尾的黑球产生新的贡献，所以一定不劣，如果没有找到可以移动至末尾的白球可以将白球内部的排列改变，这样不会影响贡献，但可以将一个白球变为可以放在末尾的情况，可以发现最终白球内部的顺序仍然会像子任务 5 中提到的一样。

根据上述讨论，我们进行一个简单总结：所有白球按照子任务 5 中的方法排列，若两个黑球对应的左右两个相邻白球不同，则两黑球之间白球个数 $\geq k$，这样可以保证任意两个黑球若没有贡献，一定其中没有白球。

令一个黑球的位置为其左边第一个白球的位置，我们希望同一个位置的黑球对尽量少，首先可以让可行的位置尽量多，令白球总数为 A，则经过简单分析可以得出位置总数为 $\frac{A}{k}+1$。

白球内部两两之间的贡献是确定的，由于白球是固定的，可以先考虑白球和黑球之间的贡献，可以发现，除了两个边界的上连续的黑球和除了相邻的 $k-1$ 个白球以外的 $A-k+1$ 个白球有贡献，其它的黑球只和除了相邻的 $2k-2$ 个白球以外 $A-2k+2$ 个白球有贡献，也就是说中间的所有位置是等价的，两边的两个位置是等价的，假如说确定了两边的黑球个数之和，根据均值不等式，那么这两边一定平均分，中间一定平均分其余的黑球。

直接枚举两边的黑球总数后计算，所有贡献可以拆分为白球之间，白球和黑球之间，黑球之间三类，均可以 $O(1)$ 得出答案，总复杂度可以做到 $O(Sn)$，可以通过子任务 $1\sim5$，期望得分 50。

3.2.4 针对子任务 6 的做法

在这个限制下，白球个数恰好 $=k$，可以放黑球的位置只有最前方和最后方两个位置，根据之前的分析平均分一定是最优的，时间复杂度 $O(m)$，结合之前的算法期望得分 60 分。

3.3 标准解法

3.3.1 标准解法 1

令 C 表示两边的黑球个数之和，将问题转化与 C 相关的函数关系，注意函数中下取整函数较多，不方便直接处理，一种简单的分析方法是从调整的角度寻找函数性质，即从 C 的变化时答案的变化量分析函数。

当 C 变大时，可以看作如下过程：在非边界所有黑球的连续段中找到黑球数量最多的一个，将其中的一个黑球移至两边黑球的连续段中黑球更少的一边，假设这个黑球原本所在所在的连续段有 x 个黑球，移到的边界的连续段有 y 个白球，那么答案变化量为 $x-1-(y-1)+k-1$，其中 $k-1$ 来自于对白球的贡献变化，$x-1$ 表示新增了对原来连续段中其它球的贡献，$-(y-1)$ 表示减少了对原来边界连续段上球的贡献，发现 C 增加时，x 不增，y 不降，变化量单调，可以三分解决，时间复杂度 $O(S\log n)$，可以通过子任务 $1\sim7$，但由于三分时计算答案常数非常大可能无法通过子任务 8，期望得分 $75\sim90$。

观察到直接求函数较为复杂，但可以根据上述调整法在极少的计算次数中求出函数在 C 变化为 $C+1$ 时的答案变化量，此时使用二分斜率代替三分可以大量减少常数，可以通过所有子任务，期望得分 90。

我们其实想找的就是第一个斜率 ≤ 0 的位置，斜率为 0 时边界上的数的出现次数的最小值为中间出现数次数的最大值 $+k-1$，可以先给两个边界各自分配 $k-1$ 个黑球，然后把剩下的黑球平均分在每一个位置上，可以发现此时一定斜率 ≤ 0，且倒退一步斜率 ≥ 0，时空复杂度均为 $O(S)$，期望得分 100。

3.3.2 标准解法 2

把最大化收益看成最小化 $\sum$（同一位置上黑球的对数 − 对白球产生的贡献），将每一个位置的代价看做一个函数，可以发现是每一个位置都一个简单二次函数，于是可以贪心地每次选当前函数值最小的一个。并且二次函数只有在边界和不在边界两种，在边界的函数初始对白球会带来额外的贡献，相当于额外加上一个一次函数，无论从函数的意义来看还是从实际意义来看，都可以得出和标准解法 1 类似的结论，前 $2k-2$ 次一定选在两边，之后所有位置等价，平均分即可，时间复杂度 $O(S)$，期望得分 100。

3.4 总结

本题从一个简单的问题入手，从特殊性质构造发现题目算法框架，通过使用贪心分析性质简化问题，再利用拆分贡献后简单的均值不等式以及调整法加快计算，最后用调整法代替复杂的函数分析证明函数的凸性，使用二分斜率代替三分减少计算量，再从函数性质重新入手快速计算，算法知识点非常简单，主要考察了贪心以及调整对题目的性质分析，是一道较为简单的贪心题。

4 参考文献

[1] 李煜东，《算法竞赛进阶指南》