NOIP2018 模拟赛 题解

测试时间：2018 年 10 月 21 日

中间值(median)

按照 NOIP 的惯例，此题是一道送分题。

不难想到二分答案。只需验证某个数是否合法。可以采用 lower_bound，upper_bound 做到$O(n\log^2 n)$，我佛系卡了一波 70。至于怎么卡的可以参见 data-max-Baoli70 程序。

实际上我们可以进行一波分类讨论。可以通过当前数的位置得到这个数在另一个序列的期望位置。假设当前的数为x，期望位置的数为y，下一个数为z，那么$z \le x \le y$时x就是答案，否则比较一下大小，往两边跳。

这种方法要特判很多种情况。事实上，我们还有一种较为简便，普适性更强的方法。假设当前要取的是区间的第 k 大，将 k 折半，放在两个区间的对应位置s,t上，比较$a[s],b[t]$，不妨设$a[s]<b[t]$，那么答案可以化归至区间$[l_1,s-1],[l_2,r_2]$的第 $\frac{k}{2}$ 大数（因为a序列比$a[s]$小的那些数一定可以全部舍去），递归即可，可以参见 data_max_median2，鸣谢 xxcc 提供代码。

事实上，这两种方法都可以拓展到任意区间长度的第k大问题上，做法一样，只是为了标题的统一性（出题人比较懒），所以出了中间值。

最小值(min)

对于30%的数据，可以想到一个 Dp 方程：

$$dp[i]=\max_{j=0}^{i-1} dp[j]+f(\min_{x=j+1}^{i} a_x),dp[0]=0$$

其中$dp[i]$表示分割$[1,i]$的最大答案。

显然可以采用一个单调递增的栈来维护 $g_x = \min_{x=j}^{i} a_x$。具体地，单调栈中的元素 $l_1, l_2 \cdots l_m$ 表示 $g_{l_i} \neq g_{l_i - 1}$ 的每个 l_i（就是最小值变化的转折点），那么有 $\forall x \in [l_i, l_{i+1} - 1], g_x$ 相同，此时 dp 值最大的那个点一定最优秀，于是维护 $h_i = \max_{x=l_i}^{l_i+1} dp[x]$，表示每个取到最小值元素对应区间的最优答案。

这样的话，每一次的答案就是 $\max h_i + f(g_{l_i})$，采用一棵线段树或者可删除堆维护单调栈即可。

最大值(max)

题目大意：n 个位置，每个位置上有若干个有一定概率出现的数，给定若干个询问区间，每次询问区间内每个位置上所有数最小值的最大值的期望之和。

对于10% 的数据，暴力枚举每个数出现或不出现即可。复杂度 $O(nq2^m)$。

对于 30% 的数据，考虑一个随机变量 x，其期望值为 $E(x) = \sum_{i=1}^{\infty} P(x \geq i)$（整数概率公式），实际就是看到最大值的最小值就二分答案。

于是我们可以枚举 x，单独考虑每个询问。

即求 $P(\max_{i=l}^{r} v_i \geq x)$

$\Leftrightarrow 1 - P(\max_{i=l}^{r} v_i \leq x - 1)$

$\Leftrightarrow 1 - P(\forall i \in [l, r] v_i \leq x - 1)$

$\Leftrightarrow 1 - \prod_{i=l}^{r} P(v_i \leq x - 1)$

$\Leftrightarrow 1 - \prod_{i=l}^{r} 1 - P(v_i \geq x)$

于是我们只需要求出每个位置的最小值大于等于 x 的概率即可。这等价于所有小于 x 的数都不出现且至少出现一个大于等于 x 的数。

所以我们可以把小于 x 的数不选的概率乘起来再减去所有数都

不选的概率，就是我们所要的$P(v_i \geq x)$。

同时发现，只有当x取到所给数的某个权值的时候，概率才会有所变化，其余情况概率不变，可以直接加起来。这样可以顺利拿到30分。

对于40%的数据，考虑某次更改的贡献，$1-P(v_i \geq y) \to 1-P(v_i \geq x)$。事实上对于某个询问，外面的1-可以最后统计，仅仅考虑$q=\prod_{i=l}^{r} 1-P(v_i \geq x)$的变化。我们可以简单地 $q \to q \times \frac{1-P(v_i \geq x)}{1-P(v_i \geq y)}$将

这样可做到$O(\log n)$修改某个询问的答案。

特殊地，我们发现$1-P(v_i \geq y)$可能等于零，但是同时也发现$P(v_i \geq x)$随着x的变大而递减，于是有$1-P(v_i \geq x)=0 \Rightarrow 1-P(v_i \geq y)=0(y \leq x)$。于是我们从大到小枚举权值即可解决这个问题。这样子的复杂度是$O(mq)$，可以得到40分。

到目前，还有一个条件没用，就是区间互不包含。

对于数据点5,6，发现一个点仅仅会被1个或2个区间包含，用邻接表处理出每个点对应要处理的区间即可拿到60分。

对于100%的数据，按区间左端点排序（出题人比较懒并没有打乱顺序），可以得到每个点被包含的区间也是一个区间。这样可以实现点和区间的转化。于是问题被转化为将一个区间乘上某一个数。采用线段树打标记维护，每次加上全局的答案即可，复杂度$O(m\log q)$，可以通过全部数据。