MEDIAN

Nguồn: AtCoder Regular Contest 101 - Problem B

Lời giải

Ta nhận xét rằng, với một giá trị x bất kì, nếu có ít nhất $\left\lceil \frac{|M|}{2} \right\rceil$ phần tử của dãy M lớn hơn hoặc bằng x thì trung vị của dãy M cũng sẽ lớn hơn hoặc bằng x và ngược lại. Nói cách khác, giá trị trung vị của dãy M chính là giá trị x lớn nhất sao cho có ít nhất $\left\lceil \frac{|M|}{2} \right\rceil$ phần tử của dãy M có giá trị lớn hơn hoặc bằng x.

Do đó, ta có thể dùng kĩ thuật chặt nhị phân để tìm giá trị x nói trên. Bài toán con ta cần giải quyết là: cho một giá trị x, đếm số cặp chỉ số (l, r) sao cho trung vị của dãy $a_l, a_{l+1}, \ldots, a_r$ lớn hơn hoặc bằng x - nói cách khác, có ít nhất $\left\lceil \frac{r-l+1}{2} \right\rceil$ phần của dãy $a_l, a_{l+1}, \ldots, a_r$ có giá trị lớn hơn hoặc bằng x.

Gọi d_i là số vị trí j từ 1 đến i sao cho $a_j \geq x$. Khi đó, điều kiện để trung vị của dãy $a_l, a_{l+1}, \ldots, a_r$ lớn hơn hoặc bằng x trở thành:

$$d_r - d_l \geq \frac{r-l+1}{2} \Leftrightarrow 2d_{l-1} - (l-1) \leq 2d_r - r$$

Gọi $f_i = 2d_i - i$. Khi đó, bài toán trở thành: đếm số cặp chỉ số (l, r) sao cho $0 \leq l < r \leq N$ và $f_l \leq f_r$. Ta có thể sử dụng cây chỉ mục nhị phân (Binary Indexed Tree, viết tắt là BIT) để giải bài toán này trong $O(N \log N)$.

Độ phức tạp: $O(N \log^2 N)$ với N là số phần tử của dãy A.

Tag: binary search, data structure