COL 351 Lecture 8 2023/01/19

Topic : Median and Order Statistics

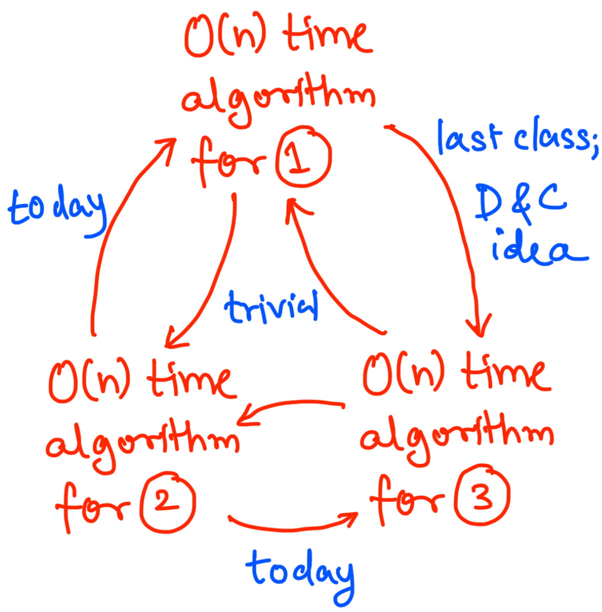
Recap

Given an array A of n (distinct) elements of an ordered set,

1. Find the median of A.

2. Find an α-approx median of A, ie an element which would lie in the middle α fraction of A on sorting. $(\alpha < 1)$

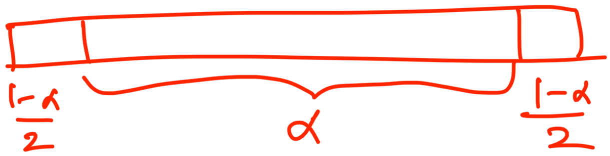
3. Given a k, find the k^{th} smallest element of A. (a.k.a. k^{th} order statistic of A)



Claim: $O(n)$ time algo for α-approx median (for $\alpha < 1$) $\Rightarrow$ $O(n)$ time algo for order statistic.

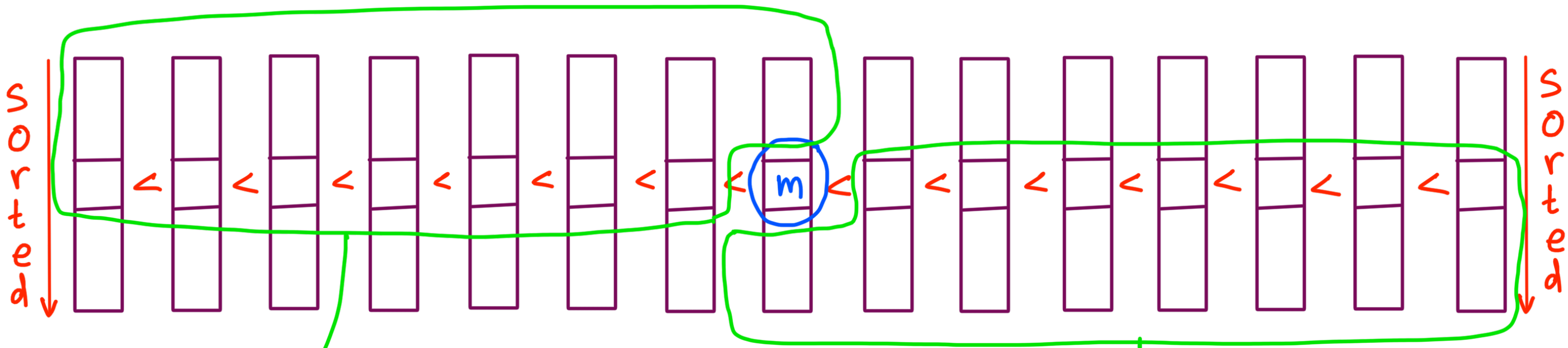
Algorithm

1. Find an α-approx median, m —— $O(1)$, assumed
2. Split A into A_0 : elements of A $\leq m$
   A_1 : elements of A $> m$ $\Big\}$ $O(n)$
3. Find $|A_0|$ —— $O(n)$
4. $T\left(\frac{1+\alpha}{2}\cdot n\right)$ $\Big\{$ If $k \leq |A_0|$, recursively find k^{th} order statistic of A_0 and return it.
   Else ($k > |A_0|$), recursively find $(k-|A_0|)^{th}$ order statistic of A_1 and return it.



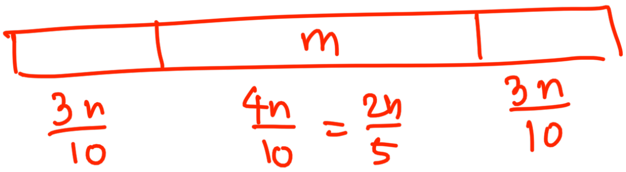
$T(n) = cn + T\left(\frac{1+\alpha}{2}\cdot n\right)$ $T(n) = cn \cdot \frac{2}{1-\alpha}$
$= O(n)$

Finding α- approx median ($\alpha = ?$)



Guaranteed to be $< m$
$\approx \frac{3n}{10}$

Guaranteed to be $> m$
$\approx \frac{3n}{10}$



Claim: Median of group medians is a $\frac{2}{5}$-approx median of the whole array.

orderstat (A, k):

(Part 1: Find α-approx median)

1. Divide A into groups of 5 elements. — $O(n)$
2. Find median of each group and put it in array M — $O(n)$
3. $m \leftarrow$ orderstat $(M, |M|/2)$. — $T(\frac{n}{5})$

(Part 2: D & C step)

4. $A_1 \leftarrow$ array of elements of A that are $\leq m$
   $A_2 \leftarrow$ array of elements of A that are $> m$ $\Big\}$ $O(n)$
5. Find $|A_1|$ — $O(n)$
6. If $k \leq |A_1|$
   return orderstat (A_1, k)
   else
   return orderstat $(A_2, k - |A_1|)$ $\Big\}$ $T(\frac{1+\alpha}{2} \cdot n)$ $= T(\frac{7}{10}n)$ $(\alpha = 2/5)$

$T(n) = T(\frac{n}{5}) + T(\frac{7n}{10}) + cn$ $T(n) = O(n)$ (Refer Tut2 problem 2)

Technical details.

- n need not be a multiple of 5

  Create $g = \lceil \frac{n}{5} \rceil$ groups in step 1.

- α calculation: $\geq \frac{g}{2} - 1$ group medians are $< m$

  $\geq 3(\frac{g}{2} - 1)$ elements $< m$

If n is large enough: $\geq 0.29\, n$ elements $< m$

($\therefore m$ is α-approx median for $\alpha = 0.42$)

In step 6, we recurse on an array of $\leq 0.71 n$ elements in the worst case

For large enough n, in step 3, we recurse on an array of $\lceil \frac{n}{5} \rceil \leq 0.21 n$.

Let n_0 be the largest integer such that

$$\left\lceil \frac{n_0}{5} \right\rceil > 0.21 n_0 \quad \text{OR} \quad 3\left(\frac{1}{2}\left\lceil \frac{n_0}{5} \right\rceil - 1\right) < 0.29 n_0$$

Base case: If $n \leq n_0$, brute force.

$$T(n) \leq T(0.21 n) + T(0.79 n) + cn \quad \text{if } n > n_0$$

$$\leq \text{const} \quad \text{if } n \leq n_0$$

$\therefore T(n) = O(n)$.