Lecture 8 Median Problem.

$A = \{a_1, \cdots, a_n\}$ $a_1 \le \cdots a_n$.

median is $a_{\lceil \frac{n}{2} \rceil} = m$

Input: unsorted array

Goal: Find the median.

Best known det. alg. $O(n)$ time.

Today: Simple randomized algorithm: $O(n)$ time.

Assumptions:

- All elements are distinct
- n is odd
- We can sample from A in constant time.

Idea:

Sample to find elements l and u. s.t.

1. $l \le m \le u$
2. \# of elements of A whose value between l and u is small

$$\rule{3cm}{0.4pt}\ \boxed{l}\ \underbrace{—\ m\ —}_{\text{sort}}\ \boxed{u}\ \rule{3cm}{0.4pt}$$

Randomized Median Alg (RMA)

Input: unsorted array A of $\{a_1 \cdots a_n\}$

Output: median.

$t = \lceil n^{\frac{3}{4}} \rceil$

$O(n^{\frac{3}{4}})$ 1. Let R be an array $r_1 \cdots r_t$, where each number r_i is choosen u.i.r. (uniformly, independently, at random). with replacement.

$O(n^{\frac{3}{4}} \lg n^{\frac{3}{4}})$ 2. Sort R

3. Let l be $\lfloor \frac{n^{\frac{3}{4}}}{2} - \sqrt{n} \rfloor$ smallest ele of R.

4. Let u be $\lceil \frac{n^{\frac{3}{4}}}{2} + \sqrt{n} \rceil$ smallest ele of R.

$O(n)$ 5. By using Partition from Quicksort

Compute. $C = \{ a \in A \mid l \le a \le u \}$

n_l = # elements $< l$ = $|\{ a \in A \mid a < l \}|$

n_u = # elements $> u$ = $|\{ a \in A \mid a > u \}|$

6. If $n_l > \frac{n}{2}$ or $n_u > \frac{n}{2}$. fail

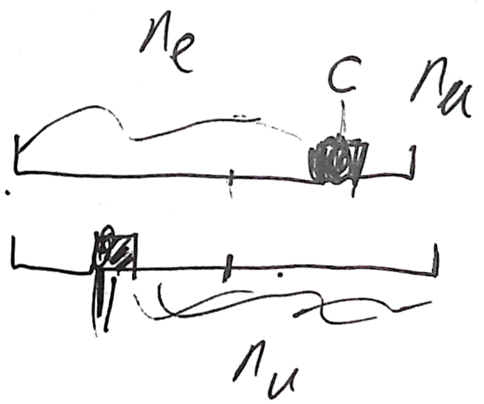
7. if $|C| \le 4n^{\frac{3}{4}}$, sort C. o/w fail.

8. output the $\lfloor \frac{n}{2} \rfloor - n_l + 1$ th element of C.

Run time: $O(n)$.

Thm 1

RMA output $\begin{cases} \text{fail} \\ \text{median} \end{cases}$ of small prob.



<u>Thm2</u> RMA fails ~~at~~ with prob $\leq n^{-\frac{1}{4}}$

<u>Bad events x3</u>

$Y_1 = |\{r \in R \mid r \leq m\}|$ $\quad \mathcal{E}_1 : Y_1 < \frac{n^{\frac{3}{4}}}{2} - \sqrt{n}.$

$Y_2 = |\{r \in R \mid r \geq m\}|$ $\quad \mathcal{E}_2 : Y_2 < \frac{n^{\frac{3}{4}}}{2} - \sqrt{n}.$

$\mathcal{E}_3 : |C| > 4n^{\frac{3}{4}}$

<u>Lemma 1</u>. $\Pr[\mathcal{E}_1] \leq \frac{1}{4} \frac{1}{n^{\frac{1}{4}}}$

<u>Lemma 2</u>. $\Pr[\mathcal{E}_2] \leq \frac{1}{4} \frac{1}{n^{\frac{1}{4}}}$

<u>Lemma 3</u> $\Pr[\mathcal{E}_3] \leq \frac{1}{2} \cdot \frac{1}{n^{\frac{1}{4}}}$

<u>Pf 1</u> $\forall i \in [\cancel{t}]$. $X_i = \begin{cases} 1 & \text{if } r_i \leq m \\ 0 & \text{o.w.} \end{cases}$

$$Y_{\ell} = \sum_{i \in [\cancel{t}]}^{\cancel{t}} X_i$$

$$\Pr[X_i = 1] = p = \cancel{\frac{n+1}{2n} = \frac{1}{2} + \frac{1}{2n}} = \frac{\frac{n+1}{2}}{n} = \frac{n+1}{2n} = \frac{1}{2} + \frac{1}{2n}$$

Chebyshev $\quad Y_1$ is a binomial distribution $\sim Bin(n^{\frac{3}{4}}, p)$

$$E[Y_1] = n^{\frac{3}{4}} \cdot p = n^{\frac{3}{4}}\left(\frac{1}{2} + \frac{1}{2n}\right) = \frac{n^{\frac{3}{4}}}{2} + \frac{1}{2n^{\frac{1}{4}}}$$

$$Var[Y_1] = \cancel{p}\, n^{\frac{3}{4}} p(1-p) = n^{\frac{3}{4}}\left(\frac{1}{2} + \frac{1}{2n}\right)\left(\frac{1}{2} - \frac{1}{2n}\right) \leq \frac{n^{\frac{3}{4}}}{4}$$

$$\Pr[\mathcal{E}_1] = \Pr\left[Y_1 < \frac{n^{\frac{3}{4}}}{2} - \sqrt{n}\right]$$

$$= \Pr\left[|Y_1 - E[Y_1]| > \sqrt{n}\right] \leq \frac{Var[Y_1]}{n} \leq \frac{1}{4n^{\frac{1}{4}}}$$