

## Problem

Given  $n$  numbers,

find median  $((n/2)^{\text{th}}$  smallest)

more generally, select  $k^{\text{th}}$  smallest

81, 42, 61, 50, 83, 4, 21, 95, 25

Sorting  $\downarrow n \log n$  comps.

Known deterministic algms:

Blum, Floyd, Pratt, Rivest, Tarjan '73

$O(n)$  time

"median-of-medians-of-5"

$\sim 16n$  comparisons

$\sim 5.4n$  "

Schonhage, Paterson Pippenger '76  $\sim 3n$

Dor, Zwick '95  $\sim 2.95n$

Lower bd  $> (2+e)n$

Randomized algms?

Method 1: rand. quickselect (variant of quicksort)

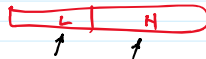
select( $\{a_1, \dots, a_n\}, k$ ): // return  $k^{\text{th}}$  smallest ( $k \in \{1, \dots, n\}$ )

1. if  $n=1$  return  $a_1$

2. pick "pivot"  $x$  AT RANDOM from  $\{a_1, \dots, a_n\}$

3.  $L = \{a_i: a_i \leq x\}$

$H = \{a_i: a_i > x\}$



4. if  $k \leq |L|$  then select( $L, k$ )

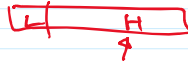
else select( $H, k - |L|$ )

Worst case:

$|L|=1$  or  $n-1$

$\Rightarrow T(n) \leq T(n-1) + O(n)$

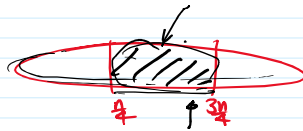
$\Rightarrow O(n^2)$



rand. analysis:

$|L|$  is equally likely to be  $1, 2, \dots, n$

$\Pr\left[\frac{n}{4} \leq |L| \leq \frac{3n}{4}\right] \geq \frac{1}{2}$



$\Rightarrow T(n) \leq T\left(\frac{3n}{4}\right) + 2 \times O(n)$

$\Rightarrow T(n) \leq O\left(2\left(1 + \frac{3}{4} + \left(\frac{3}{4}\right)^2 + \dots\right)n\right)$

$= O(n)$  expected

with more careful analysis,  
(expected # comps  $\leq 3.4n$ ).

Method 2: Floyd-Rivest Alg'm ('75)

(Monte-Carlo)

ideally  
 $T(n) = T\left(\frac{n}{2}\right) + O(n)$   
 $\Rightarrow O(n)$

## (Monte-Carlo)

idea - random sampling

1. pick subset  $R$  of  $r$  rand. elements  $y_1, \dots, y_r$
2. take  $(\frac{r}{2} - c\sqrt{r})$ -th smallest  $a$  of  $R$   $\leftarrow$   
 $(\frac{r}{2} + c\sqrt{r})$ -th smallest  $b$  of  $R$   $\leftarrow$
3. let  $L = \{\text{all elements} \leq a\}$   
 $M = \{\text{all elements between } a \text{ \& } b\}$   
 $H = \{\text{all elements} > b\}$



check:  $|L| \leq \frac{n}{2}$ ,  $|H| \leq \frac{n}{2}$ ,  $|M| \leq \frac{4cn}{\sqrt{r}}$   $\leftarrow (*)$

4. return  $(\frac{n}{2} - |L|)$ -th smallest of  $M$ .

Runtime: line 2  $O(r \log r)$  time (no recursion needed)

line 4  $O(\frac{cn}{\sqrt{r}} \log \frac{cn}{\sqrt{r}})$  time

line 3  ~~$2n$~~  comps  
 $n + \frac{n}{2} + \frac{4cn}{\sqrt{r}}$  comps

Set  $r = (cn)^{2/3} \Rightarrow \underline{\underline{2n + O((cn)^{2/3} \log n)}}$  comps  
 $\underline{\underline{1.5n}}$

$$\begin{aligned} r &= \frac{cn}{\sqrt{r}} \\ r^{3/2} &= cn \end{aligned}$$

## Error Analysis:

Recall Markov's ineq:

$$X \geq 0, E[X] = \mu \Rightarrow \Pr[X \geq c\mu] \leq \frac{1}{c}$$

Chebyshev's ineq:  $E[X] = \mu$ ,  $\text{Var}[X] = \sigma^2$

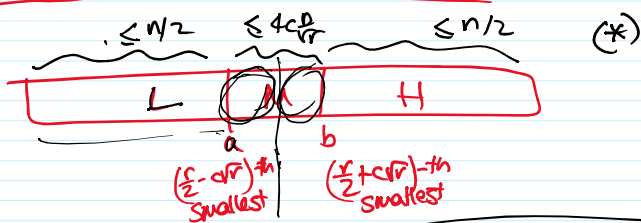
$$\uparrow E[(X - \mu)^2]$$

$$\Rightarrow \Pr[|X - \mu| \geq c\sigma] \leq \frac{1}{c^2}$$

Pf:

$$\Pr\left[\frac{1}{c^2} \leq \frac{(X - \mu)^2}{\sigma^2} \leq 1\right] \leq \frac{1}{c^2}$$

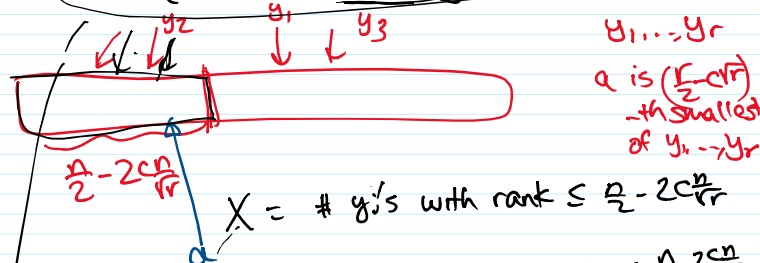
$$\Pr \left( \frac{(X - \mu)}{\sigma} \geq c\sigma \right) \quad \square$$



(\*) is true if  $a$  has rank  $\in \left( \frac{n}{2} - \frac{2cn}{r}, \frac{n}{2} \right]$   
 &  $b$  has rank  $\in \left[ \frac{n}{2}, \frac{n}{2} + \frac{2cn}{r} \right]$

Will analyze  $\Pr \left[ \text{rank}(a) < \frac{n}{2} - \frac{2cn}{r} \right] \leq \dots$

$$\Pr \left[ \left( \# \text{ } y_i \text{'s that have rank} \leq \frac{n}{2} - \frac{2cn}{r} \right) \geq \frac{n}{2} - cn/r \right]$$



for  $i = 1, \dots, r$ ,  $X_i = \begin{cases} 1 & \text{if } y_i \text{ has rank} \leq \frac{n}{2} - \frac{2cn}{r} \\ 0 & \text{else} \end{cases}$

$$X = \sum_{i=1}^r X_i$$

$$\begin{aligned} E[X_i] &= \Pr \left( y_i \text{ has rank} \leq \frac{n}{2} - \frac{2cn}{r} \right) \\ &= \frac{\frac{n}{2} - \frac{2cn}{r}}{n} \\ &= \frac{1}{2} - \frac{2c}{r} \end{aligned}$$

$$\begin{aligned} \text{Var}(X_i) &= E[X_i^2] - E[X_i]^2 \\ X_i \in \{0, 1\} \implies E[X_i^2] &= E[X_i] \\ &\leq E[X_i] \end{aligned}$$

$$\Pr \left[ X \geq \frac{n}{2} - cn/r \right]$$

$$\mu = E[X] = \sum_{i=1}^r E[X_i] = \left( \frac{n}{2} - \frac{2cn}{r} \right)$$

$$\sigma^2 = \text{Var}[X] = \sum_{i=1}^r \text{Var}[X_i] \leq \mu$$

by independence

$$\sigma \leq \sqrt{\mu} \leq \sqrt{r}$$

$$\begin{aligned} \Pr \left[ X - \mu \geq cn/r \right] &\leftarrow \\ &\leq \Pr \left[ X - \mu \geq c\sigma \right] \leftarrow \end{aligned}$$

$$\leq \Pr[X - \mu \geq c\sigma] \leftarrow$$

$$\leq \frac{1}{c^2} \quad \text{by Chebyshev's ineq.}$$

3 other ineqs similar

$$\Rightarrow \Pr[(*) \text{ fails}] \leq \underbrace{4 \cdot \frac{1}{c^2}}_{\leftarrow \text{with Chernoff, } \frac{1}{e^{\Theta(c^2)}}}$$

Rmk. easy to convert to Las Vegas

$$1.5n + O((cn)^{2/3} \log n) + O\left(\frac{1}{c^2}\right) n \log n$$

↑  
if fail, switch to sorting

$$c = n^{0.1}$$

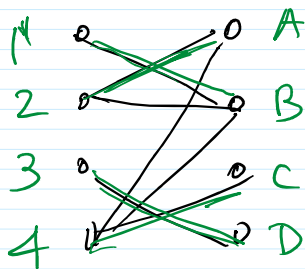
$(1.5n^{1.1} + o(n))$  expected #comps.

## OPTIMIZATION : Matching, Flows, & LP

Problem

Given bipartite graph  $G = (V, E)$ ,  
find matching with max size

↗ subgraph with max deg 1



$\{1B, 3D, \}$   
 $4C, 2A$