Problem Given n numbers, S,
find median $\quad$ ($(n/2)^{th}$ smallest)

73, 81, 60, 52, 43, 92, 100,

Sorting: $O(n \log n)$ time $\qquad 1 \cdot n \log n + O(n)$ comps.
better?

Known deterministic alg'm:

Blum, Floyd, Pratt, Rivest, Tarjan '73:
$O(n)$ time $\qquad$ "median-of-medians-of-5"

$\sim$ $16n$ comparisons
$\sim$ $5.4n$ "

Schönhage, Paterson, Pippenger '76 $\quad \sim 3n$ comps
Dor, Zwick '95 $\quad \sim 2.95n$

Lower bds:
Bent, John '80 $\quad \sim 2n$
Dor, Zwick '96 $\quad \sim 2.000\,000\,000\,000\,000\,000\,000\,000\,001\ n$

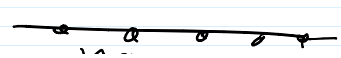
Randomized alg'ms?

quickselect:
pick rand pivot
recurse on one side
$\Rightarrow O(n)$ expected time
if careful, $\leq 3.4n$ comps expected
$\downarrow$
$1.5n$ comps expected

Floyd & Rivest Alg'm ('75)

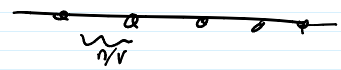
idea - random sampling

1. pick subset $R \subseteq S$ of r random elements $y_1, \ldots, y_r$
2. take $(\frac{r}{2} - c\sqrt{r})$-th smallest a of R



2. take $\left(\frac{r}{2}-c\sqrt{r}\right)$-th smallest a of R

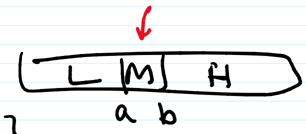
$\left(\frac{r}{2}+c\sqrt{r}\right)$-th smallest b



3. let $L = \{$ all elements of $S \leq a\}$

$M = \{$ all elements of S between a & $b\}$

$H = \{$ all elements of $S \geq b\}$



$$\boxed{\text{check: } |L| \leq \frac{n}{2}, \quad |H| \leq \frac{n}{2}, \quad |M| \leq 4c\frac{n}{\sqrt{r}} \quad (*)}$$

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

Runtime: line 2 $O(r\log r)$ by sorting

line 4 $O\left(\frac{cn}{\sqrt{r}}\log\frac{cn}{\sqrt{r}}\right)$ by sorting

line 3 $2n$ comps naively

or better: $\underbrace{n}_{\text{comp with } a} + \underbrace{|M|+|H|}_{\text{comp with } b} \leq \frac{3n}{2} + 4c\frac{n}{\sqrt{r}}$

$$\Rightarrow 1.5n + O\left(\left(r+\frac{cn}{\sqrt{r}}\right)\log n\right)$$

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

$r = \frac{cn}{\sqrt{r}}$

$r^{3/2} = cn$

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 $E(X) = \mu$, $Var[X] = \sigma^2$

$\left(Var(X) = E\left((X-\mu)^2\right) = E[X^2] - E[X]^2\right)$

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

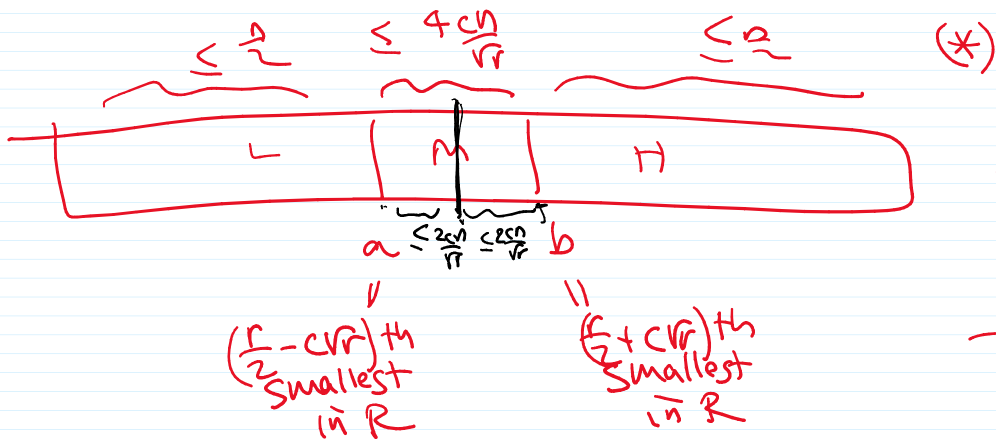
Pf: $\qquad \| $

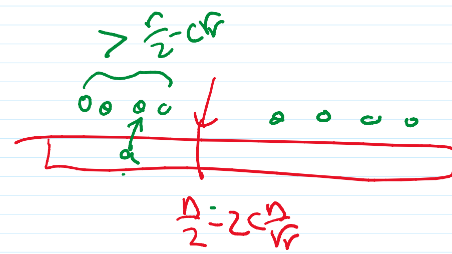
$\Pr\left((X-\mu)^2 \geq c^2\sigma^2\right)$

Pf:

$$\Pr\left((X-\mu)^2 \geq c^2\sigma^2\right) \qquad \square$$

apply Markov





(*) is true if a has rank between $\frac{n}{2} - \frac{2cn}{\sqrt{r}}$ and $\frac{n}{2}$ in S.

& b has rank between $\frac{n}{2}$ and $\frac{n}{2} + \frac{2cn}{\sqrt{r}}$ in S.

Analyze

$$\Pr\left[\operatorname{rank}_S(a) < \frac{n}{2} - \frac{2cn}{\sqrt{r}}\right] = \Pr\left[X \geq \frac{r}{2} - c\sqrt{r}\right]$$

Let $X = \{\#\text{ elems } y_i \in R \text{ with } \operatorname{rank}_S(y_i) < \frac{n}{2} - 2c\frac{n}{\sqrt{r}}\}$.

$R = \{y_1, \dots, y_r\}$

$$X_i = \begin{cases} 1 & \text{if } \operatorname{rank}_S(y_i) < \frac{n}{2} - 2c\frac{n}{\sqrt{r}} \\ 0 & \text{else} \end{cases}$$

$$X = \sum_{i=1}^{r} X_i \qquad E[X_i] = \frac{\frac{n}{2} - \frac{2cn}{\sqrt{r}}}{n} = \frac{1}{2} - \frac{2c}{\sqrt{r}} = p$$

$$\Rightarrow E[X] = pr = \mu$$
$$= \frac{r}{2} - 2c\sqrt{r}$$

$$\operatorname{Var}[X_i] = E(X_i^2) - E[X_i]^2$$
$$\leq E(X_i) = p.$$

$$\operatorname{Var}[X] = \sum^{r} \operatorname{Var}(X_i) \quad \text{by independence}$$

$$\mathrm{Var}[X] = \sum_{i=1}^{r} \mathrm{Var}[X_i] \quad \text{by independence}$$

$$\leq pr = \mu$$

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

$$\Pr\left[X \geq \frac{r}{2} - c\sqrt{r}\right]$$

$$\leq \Pr\left[X \geq \mu + c\sqrt{r}\right]$$

$$\leq \Pr\left[X \geq \mu + c\sigma\right]$$

$$\leq \frac{1}{c^2} \quad \text{by Chebyshev}$$

Similar for the other ways to fail

$\Rightarrow$ error prob $\leq O\left(\frac{1}{c^2}\right)$. $\leftarrow$ with "Chernoff's bd", $e^{-\Theta(c^2)}$

$1.5n + O\left((cn)^{2/3} \log n\right)$ comps.

$+ O\left(\frac{1}{c^2} \cdot n \log n\right)$ $\leftarrow$ switch to sorting if error

Convert to Las Vegas

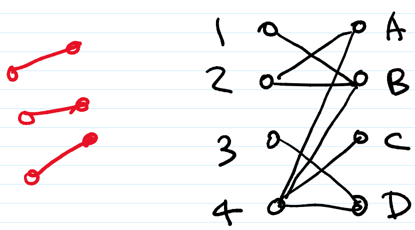
set $c = n^{0.1}$ $\Rightarrow$ $1.5n + o(n)$ comps. expected

OPTIMIZATION PROBLEMS:

Matchings, Flows, & LPs

Problem Given unweighted bipartite graph $G=(V,E)$,
$|V|=n$, $|E|=m$,
$(n \le m \le n^2)$

find matching with max size
subgraph where every vertex $\deg \le 1$.



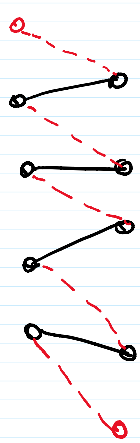
$\{2B, 3D, 4C\}$

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

idea - brute force $\Omega(n!)$
greedy. not work
DP?
D&C?

"Hungarian" Method (König-Egerváry '35 / Edmonds '65)

idea- local improvement
(start with feasible sol'n, modify & improve, ...)



"alternating path"
augmenting

$M = \phi$
repeat: {
find aug path P for M
if not exist, return M
modify M by P.
}

Facts a matching M is optimal $\Leftrightarrow$ it has no augmenting path

can find an augmenting path in $O(m)$ time by BFS/DFS
or decide none exists

iterations $\leq O(n)$

$\Rightarrow$ total time $O(mn)$

$\vdots$