CSE 4502/5717

LECTURE ON

9-22-2022

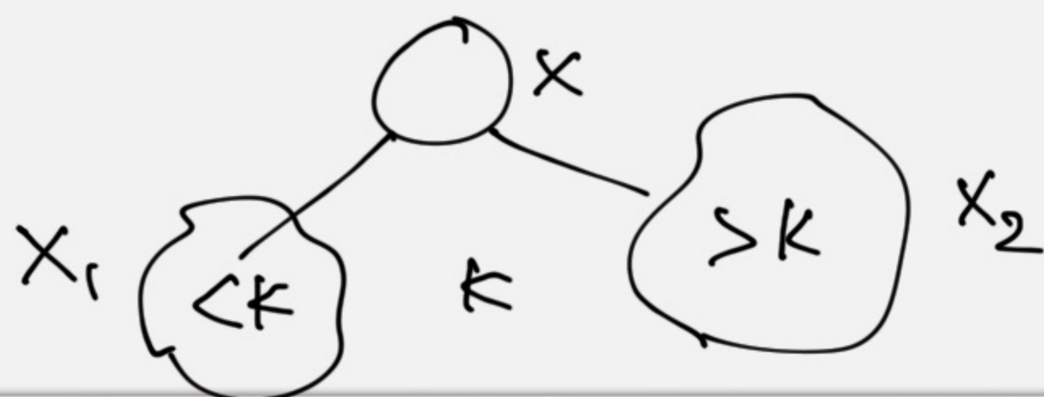
SELECTION: $X = k_1, k_2, \ldots, k_n;\ 1 \le i \le n$

Output: The i^{th} smallest of X;

Quickselect(X, i)

Pick a PIVOT $k \in X$;

Partition X into X_1 and X_2;



CASE 1: If $|X_1| = i-1$ then output k & stop;

CASE 2: If $|X_1| \geq i$ then output QuickSelect(i, X_1);

CASE 3: If $|X_1| + 1 < i$ then output QuickSelect$(X_2, i - |X_1| - 1)$;

WORST CASE RUNTIME = $\Omega(n^2)$

AVERAGE RUN TIME = $O(n)$.

BFPRT:

1. * Partition the input into groups of size 5 EACH; GROUPS → $G_1, G_2, \ldots, G_{\frac{n}{5}}$.
   * FIND the Median M_i of G_i $1 \leq i \leq \frac{n}{5}$;
   * FIND the MEDIAN M of these MEDIANS, RECURSIVELY.

RUNS steps 2 & 3 of QuickSelect with $k = M$.

ANALYSIS:

Let $T(n)$ be the RUN TIME of this algorithm on any input of size n, on any i.

$|X_1| \le \frac{7}{10}n$;

$|X_2| \le \frac{7}{10}n$.

$$T(n) = 9 \cdot \frac{n}{5} + T\left(\frac{n}{5}\right) + n + T\left(\frac{7}{10}n\right) \quad \text{—— (1)}$$

We can solve this by induction.

HYPOTHESIS: $T(n) \le dn$ FOR SOME CONSTANT d.

$$T(n) \le 2.8n + T\left(\frac{n}{5}\right) + T\left(\frac{7}{10}n\right) \quad (1)$$

Base case is EASY.

INDUCTION Step: Assume the hypothesis is true for all inputs of size $\le (n-1)$.

We'll prove it for n.

APPLYING the hypothesis on (1),

$$T(n) \le 2.8n + d\frac{n}{5} + d \cdot \frac{7}{10}n.$$

$$= 2.8n + d(0.9n). \quad (2)$$

RHS of ② $\leq dn$ if

$$2.8n + 0.9dn \leq dn$$

$$\Rightarrow \quad 2.8\cancel{n} \leq 0.1d\cancel{n}$$

$$\Rightarrow \quad \boxed{d \geq 28}$$

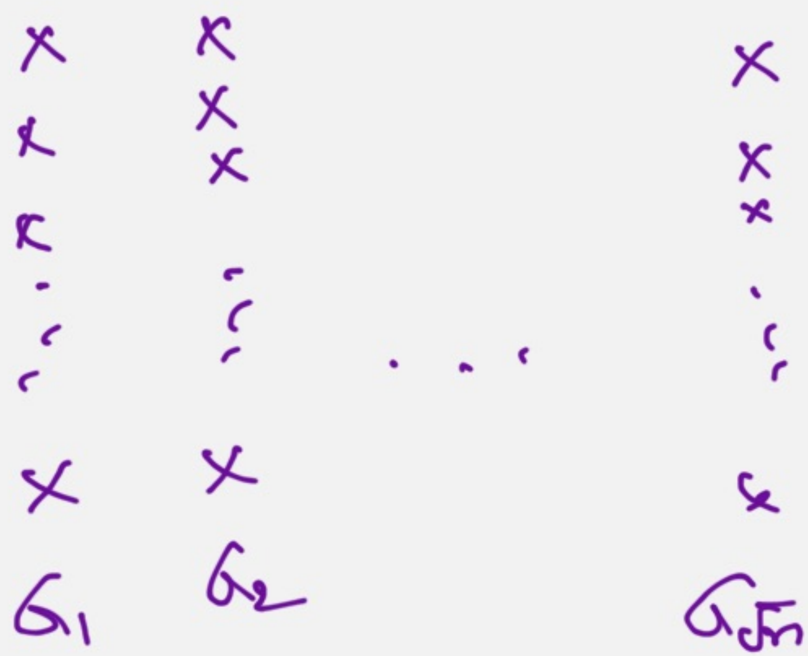
$$\Rightarrow \quad \boxed{T(n) \leq 28n}$$

$\exists$ a RAND. ALG. That makes

$\boxed{n + \text{Min}\{i, (n-i)\}} + o(n)$ COMPARISONS.

HW1: P4: $P = \sqrt{n}$; $T = O(\sqrt{n})$.

HINT:



OUT_OF_CORE SELECTION:

$k_1, k_2, \ldots, k_n$

X

DISK.

Step 1: GROUP X into groups of size 5 EACH; FIND the MEDIAN OF EACH GROUP AND WRITE THESE MEDIANS IN the DISK

$M_1, M_2, \ldots, M_{\frac{n}{5}}$

# of I/O OPERATIONS $= \frac{n}{B}$.

(2) FIND the MEDIAN M of those MEDIANS RECURSIVELY.

(3) Partition X into X_1 & X_2 USING M AS the PIVOT.

I/O OPERATIONS $= \frac{n}{B}$.

(4) IN the WORST CASE RECURSE ON X_1 or X_2.

Let $I(n)$ be the I/O complexity of this alg. on any input of size n and for any i.

$$I(n) \le \frac{n}{B} + I\left(\frac{n}{5}\right) + \frac{n}{B} + I\left(\frac{7}{10}n\right).$$

3

Hypothesis: $I(n) \le d\frac{n}{B}$ for some constant d.

Proof by induction: Base case is easy.

INDUCTION Step: Assume that the hyp. is valid for all inputs of size $\leq (n-1)$.
We'll prove it for n.

$$I(n) \leq 2\frac{n}{B} + I\left(\frac{n}{5}\right) + I\left(\frac{7}{10}n\right).$$

$$\leq 2\frac{n}{B} + d\frac{n}{5B} + d\frac{7}{10}\frac{n}{B}.$$

RHS $\leq d\frac{n}{B}$ IF $\frac{2\cancel{n}}{\cancel{B}} + d\frac{\cancel{n}}{\cancel{B}}(0.2+0.7) \leq d\frac{\cancel{n}}{\cancel{B}}$.

This will happen if $0.1d \geq 2$
$\Rightarrow \boxed{d \geq 20}$.

$$\Rightarrow \quad I(n) \leq 20\,\frac{n}{B} = O\left(\frac{n}{B}\right).$$

<u>DEFN.</u> A BERNOULLI TRIAL HAS TWO POSSIBLE OUTCOMES: SUCCESS, FAILURE.

$\text{Prob}[\text{Success}] = p.$

$\text{Prob}[\text{Failure}] = (1-p).$

A BINOMIAL DISTRIBUTION is the # of SUCCESSES in n INDEPENDENT BERNOULLI TRIALS, DENOTED AS $B(n, p)$.

MEAN $= np$.

MARKOV'S inequality: If X is a +ve Random variable whose MEAN is μ,

$$\text{Prob}[X \geq a\mu] \leq \frac{1}{a} \text{ FOR ANY } a \geq 1.$$

CHERNOFF BOUNDS: If X is $B(n,p)$

then, $\text{Prob.}\left[X > (1+\epsilon)np\right] \leq \exp\left(-\frac{\epsilon^2 np}{3}\right)$

and $\text{Prob}\left[X < (1-\epsilon)np\right] \leq \exp\left(\frac{-\epsilon^2 np}{2}\right)$.

FOR ANY $0 < \epsilon < 1$.

Example:- $n = 1000$; $p = 1/2$.

What is $\text{Prob}\,(X > 600)$?

APPLYING MARKOV'S INEQUALITY,

$$\text{Prob.}\left[X > a\cdot 500\right] \le \frac{1}{a}.$$

$$a\cdot 500 = 600 \Rightarrow a = \frac{6}{5}.$$

$$\Rightarrow \text{Prob.}\left[X > 600\right] \le 5/6 \leftarrow (1)$$

Use Chernoff Bounds:

$$(1+\epsilon)500 = 600 \Rightarrow \epsilon = 0.2.$$

$$\Rightarrow \text{Prob.}\left[X > 600\right] \le \exp\left(\frac{-(0.04)500}{3}\right)$$

$$= \exp\left(\frac{-20}{3}\right) = 0.0012726$$

FLOYD and RIVEST (1977):

$X = k_1, k_2, \dots, k_n$; $1 \le i \le n$;

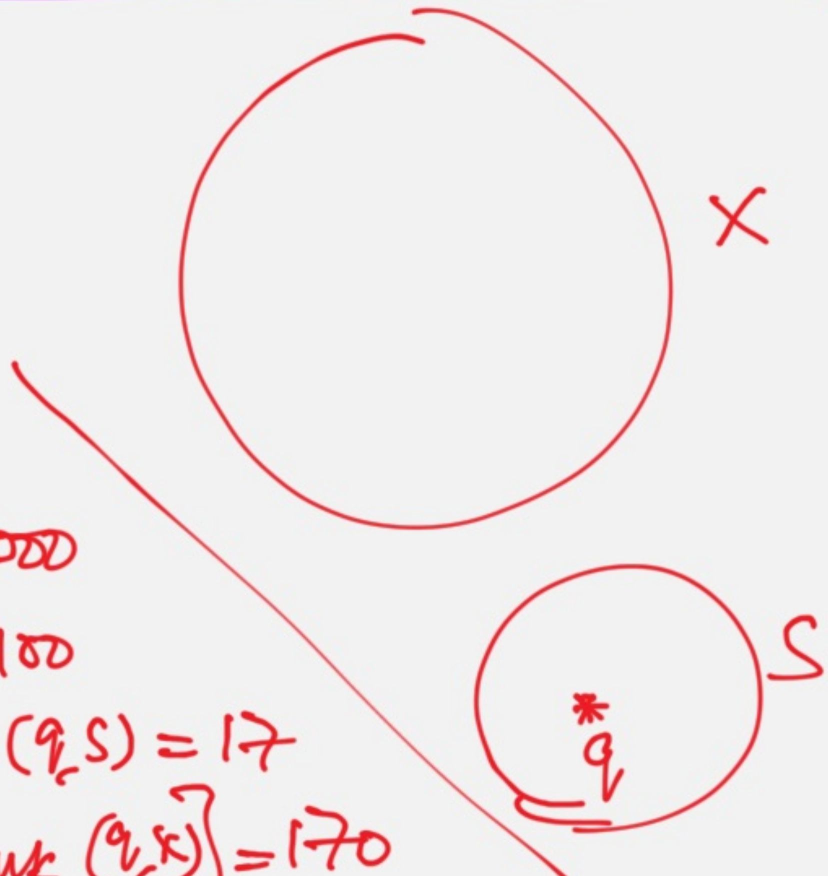
① PICK A RADOM SAMPLE S with $|S| = s$.

② Pick two elements l_1 and l_2 FROM S s.t.

$$\text{Rank}(l_1, S) = i\frac{s}{n} - \delta$$

$$\text{Rank}(l_2, S) = i\frac{s}{n} + \delta$$

$$\boxed{\delta = \sqrt{4 \alpha s \log n}}$$



$$E[\mathrm{Rank}(q, X)] = m \cdot \frac{|X|}{S}$$

$$\mathrm{Rank}(q, S) = m$$

$|X| = 1000$

$|S| = 100$

$\mathrm{Rank}(q, S) = 17$

$E[\mathrm{Rank}(q, X)] = 170$

Theorem: (RAJASEKARAN & REIF 1985): $|X| = n$.

Let S be a Random Sample FROM X.

Let j be the Rank of e in S.

Let r_j be the Rank of e in X.

$|S| = s$.

$$\text{Prob}\left(\left[r_j - j \frac{n}{s} \right] > \sqrt{3\alpha}\, \frac{n}{\sqrt{s}} \sqrt{\log n} \right) \leq n^{-\alpha}.$$

$j\frac{n}{s} - \sqrt{3\alpha}\frac{n}{\sqrt{s}}\sqrt{\log n}$ $\quad E[r_j] = j\frac{n}{s}$ $\quad j\frac{n}{s} + \sqrt{3\alpha}\frac{n}{\sqrt{s}}\sqrt{\log n}$

① $\left|\{ q \in X : l_1 \le q \le l_2 \}\right|$ is "SMALL"

② The i^{th} smallest element of X $\in [l_1, l_2]$ w.h.p.