# Lecture 7 COL 702 Aug 17

Given a set $S$ of $n$ numbers, we want to find an element in $S$ that has rank = $k$ $(1 \le k \le n)$

Selection

5 20 9 15 8

$n = 5$

$k = 2$ 5 [8] 9 15 20

$k$ 2

Do we need to sort to find the rank $k$ element?

Can we do selection in $o(n \log n)$ comparisons

Median finding : $k = \left\lfloor \frac{n}{2} \right\rfloor$ or $\left\lfloor \frac{n}{2} \right\rfloor$

Considered important for even splitting for divide-and-conquer algorithms like sorting

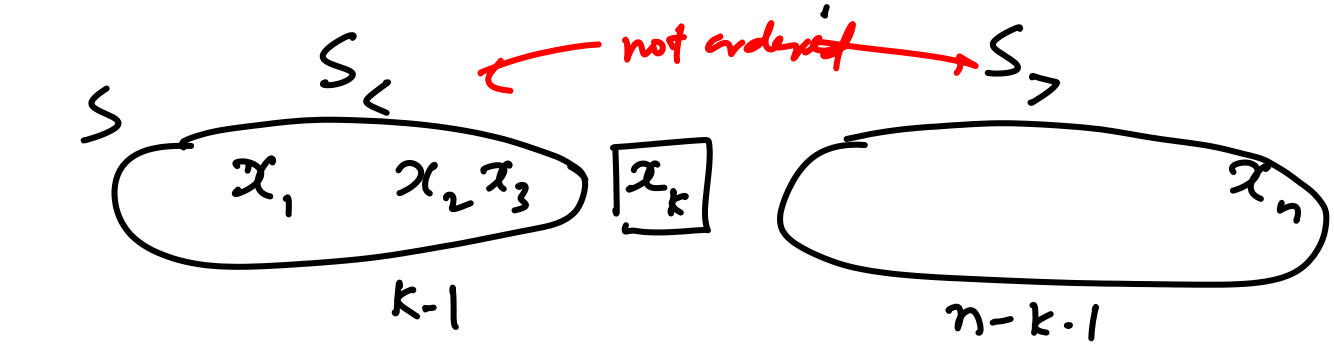
Goal : Design an $O(n)$-time for selection

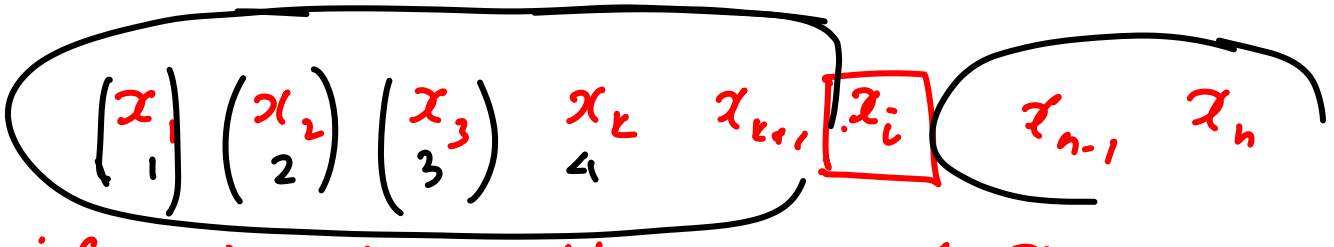
Sol case $k = 1$ / $k = n$

$k = 2$ · $2n$ ? (two scans)

$\Rightarrow$ $O(kn)$ for $k$ scans

Why can't we find minimum faster than $n-1$ comparisons?




if $i > k$ then $x_k < x_i$

$i < k$ then $x_k > x_i$

$S_{<x_i}$ or $S_{>x_i}$ can be generated by comparing all elements with $x_i$

At least one of the subsets $S_{<x_i}$ or $S_{>x_i}$ can be discarded

[We repeat (recursively) the same in the remaining subset with $k = k \,/\, k-i$ until $k = i$

Suppose $i = \frac{n}{2}$

$$n + \frac{n}{2} + \frac{n}{4} \cdots < 2n$$

So find $k^{th}$ ranked element can be reduced to finding median efficiently

Very unbalanced partitioning will lead to $\Omega(n-1 + n-2 + \cdots) = \Omega(n^2)$

The rank of the partitioning element $x_i$ is key to the success of the approach

Suppose $i = \frac{3n}{4}$

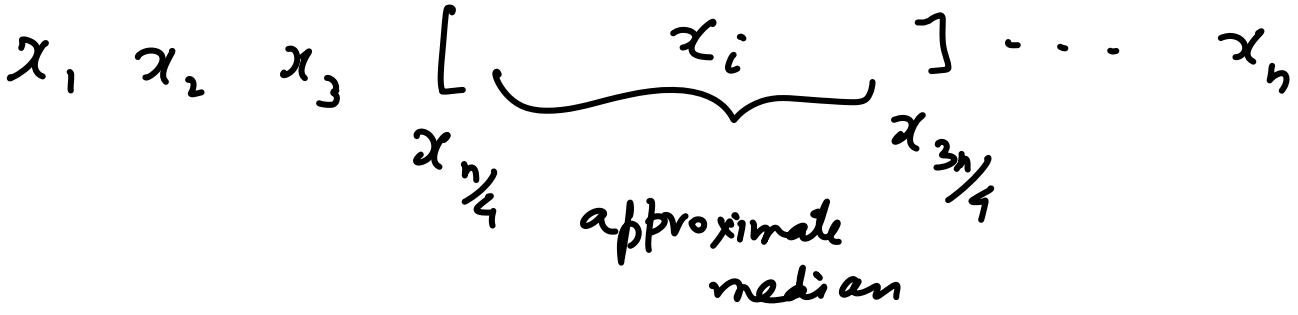

For any rank $\alpha n$, $0 < \alpha < 1$, we will be long $n + \alpha n + \alpha^2 n, \cdots$

$= O(n)$

As long as the rank of the partitioning element is "approx in the middle"

$$[\alpha n, (1-\alpha)n]$$

$\alpha = \frac{1}{4}$ $\left[\frac{n}{4}, \frac{3n}{4}\right]$




Choose $x_i$ uniformly at random from $S$

$\Rightarrow$ Rank of the pivoting element is uniformly distributed in $[1, 2 \ldots n]$

Say $Y$ is the random variable with a value in $[1, 2 \ldots n]$

rank of $x_i$

$$Pr[Y = j] = \frac{1}{n}$$

$1 \le j \le n$

What is the prob that $\frac{n}{4} \le X \le \frac{3n}{4}$

$$= \frac{1}{2}$$

If $Y$ is not in this range we choose another splitter randomly till we succeed

What is the expected no. of times we repeat the above

Let $Z$ be a random variable that has value $k$ with prob $p_k$

Then Expectation of $Z$ : $E[Z] = \sum_k k \cdot p_k$

$Z$ : # of trials

| $Z$ = | 1 | 2 | 3 | 4 | . . |
|---|---|---|---|---|---|
| Prob | $\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{8}$ | | |

$E[Z] = 2 = \sum i \cdot \frac{1}{2^i}$

For a geometric distribution with parameter $p$ the Expectation is $\frac{1}{p}$

For any r.v. $Z_1, Z_2, Z_3, \ldots Z_t$ (not necessarily independent)

$$E[Z_1 + Z_2 + \cdots Z_t] = \sum_{i=1}^{t} E[Z_i]$$

linearity of expectation

What is the expected running time of the selection algorithm based on random choice of splitter

Let $X_1$ be the random variable for the # comparisons in 1st round

$X_i$ : r.v. for # comparisons in $i^{th}$ round

Total cost : $X_1 + X_2 + \cdots X_t$

$$E[\text{Total cost}] = E[X_1 + X_2 + \cdots X_t]$$

$$= \sum_{i=1}^{p} E[X_i]$$

$t = \log_{4/3} n$

$= O(\log n)$

$E[X_1] \leq 2n$

$E[X_2] \leq 2 \times \frac{3}{4} n$

$E[X_i] \leq 2 \times \left(\frac{3}{4}\right)^{i-1} n$

$2n \leq \left(\frac{3}{4}\right)^{i} \leq O(n)$

What does it mean by Expected running time $\leq f(n)$

Markov inequality

$$Pr[X \geq c\,E[X]] \leq \frac{1}{c}$$

for non-neg rand variables