Randomized Algorithms

- Deterministic and (Las Vegas & Monte Carlo) Randomized Algorithms
- Probability Review
- Probabilistic Analysis of deterministic QUICK-SORT Algorithm
- RANDOMIZED-SELECT and RANDOMIZED-QUICK-SORT
- Max-Cut
- Min-Cut
- MAX-3-SAT and Derandomization
- Closest pair
- Hashing, Bloom filters, Streams, Sampling, Reservoir sampling, Sketch

IMDAD ULLAH KHAN

Input: An array S with n distinct numbers and $k \in \mathbb{Z}$ $(1 \leq k \leq n)$

Output: The kth smallest number in S (a number with rank k)

Obvious solutions:

- For $1 \leq i \leq n$, find rank of $S[i]$ in S

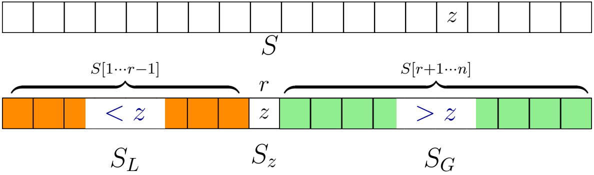
  $\triangleright$ Each find rank takes $O(n)$ time

- Sort S, and return $S[k]$

Input: An array S with n distinct numbers and $k \in \mathbb{Z}$ $(1 \leq k \leq n)$

Output: The kth smallest number in S (a number with rank k)

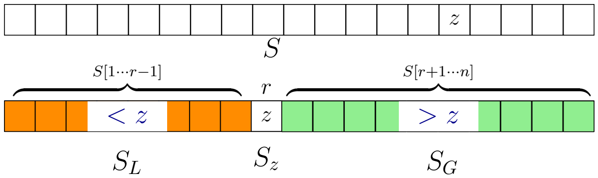
For $z \in S$, partition S into S_L $(< z)$, S_z $(= z)$ and S_G $(> z)$



Input: An array S with n distinct numbers and $k \in \mathbb{Z}$ $(1 \leq k \leq n)$

Output: The kth smallest number in S (a number with rank k)

For $z \in S$, partition S into S_L $(< z)$, S_z $(= z)$ and S_G $(> z)$



The following recurrence gives a clear algorithm (subject to choosing z)

$$\textsc{select}(S,k) = \begin{cases} \textsc{select}(S_L, k) & \text{if } k \leq |S_L| \\ z & \text{if } k = r = |S_L| + 1 \\ \textsc{select}(S_G, k - |S_L| - 1) & \text{if } k > |S_L| + 1 \end{cases}$$

RANDOMIZED-SELECT

For $z \in S$, partition S into S_L $(< z)$, S_z $(= z)$ and S_G $(> z)$

$$\textsc{select}(S,k) = \begin{cases} \textsc{select}(S_L,k) & \text{if } k \leq |S_L| \\ z & \text{if } k = r = |S_L|+1 \\ \textsc{select}(S_G, k-|S_L|-1) & \text{if } k > |S_L|+1 \end{cases}$$

Let $T(n)$ be the runtime of this algorithm on $|S| = n$

$$T(n) = T\left(\textsc{max}\{r, n-r-1\}\right) + \Theta(n)$$

- Worst case $T(n) = \Theta(n^2)$ ▷ imbalanced partition

RANDOMIZED-SELECT

For $z \in S$, partition S into S_L $(< z)$, S_z $(= z)$ and S_G $(> z)$

$$\textsc{select}(S,k) = \begin{cases} \textsc{select}(S_L, k) & \text{if } k \leq |S_L| \\ z & \text{if } k = r = |S_L| + 1 \\ \textsc{select}(S_G, k - |S_L| - 1) & \text{if } k > |S_L| + 1 \end{cases}$$

Let $T(n)$ be the runtime of this algorithm on $|S| = n$

$$T(n) = T\left(\textsc{max}\{r, n - r - 1\}\right) + \Theta(n)$$

- Worst case $T(n) = \Theta(n^2)$ ▷ imbalanced partition
- Choose z at random ▷ want $z \sim \textsc{median}(S)$

RANDOMIZED-SELECT

Let $T(n)$ be the runtime of partition-based SELECT algorithm on $|S| = n$

For random z, $T(n) = T\left(\text{MAX}\{r, n-r-1\}\right) + \Theta(n)$ is a random variable

$$E[T(n)] = n + \sum_{r=1}^{k} T(r)p[rank(z) = r] + \sum_{r=k+1}^{n} T(n-r)p[rank(z) = r]$$

- $Pr[rank(z) = r] \;=\; {}^{1}\!/_{n}$

$$E[T(n)] \;=\; n + \frac{1}{n}\Bigg[\sum_{r=1}^{k} T(r) + \sum_{r=k+1}^{n} T(n-r)\Bigg] \;\le\; n + \frac{2}{n}\Bigg[\sum_{r=\lfloor n/2 \rfloor}^{n} T(r)\Bigg]$$

$E[T(n)] \le cn$ $\quad\triangleright$ Easily proved by induction