Order statistics and Medium :

⇒ Given n element in an array (unsorted). find k^{th} ~~element~~ smallest element (element of rank k).

Solⁿ :-

① Naive Solⁿ :- $n\log n$, $O(n + k\log n)$

② Expected Linear time Solⁿ :-

$K = 1$: minimum

$K = n$: maximum

$K = \left\lfloor \frac{n+1}{2} \right\rfloor$ or $\left\lceil \frac{n+1}{2} \right\rceil$: medium

③ Randomized divide & Conquer :-

i = index ~~i = index of smallest~~

Random-Select (A, p, q, i) // i^{th} smallest in A[p..q]

if p == q then return A[p]

r ← Random partition (A, p, q) // pick random index in b/w p & q

K ← r − p + 1 // K = rank (A[r]) in A[p...q].

if i == k

then return A[r]

if i < k.

A [≤A[r] | | ≥A[r]] (p … r … q)

then

return Random-Select (A, p, r−1, i)

else

return Random-select (A, r+1, q, i−k)

eg :- $i = 7$

$A =$

6	10	13	5	8	3	2	11
↑ Pivot							

$A' =$

2	5	3	6	8	13	10	11
↑ p			↑ r				↑ q

$k = 4$. rank $= 7 - 4$

$= 3$

⇒ Intution for analysis :-

(All elements are distinct)

Lucky case : $\frac{1}{10} : \frac{9}{10}$

$$T(n) = T\left(\frac{9}{10}n\right) + \Theta(n) \leftarrow \text{partitioning}$$

$$= O(n)$$

Unlucky Case :-

$0 : n-1$

every time → pick corners

$$T(n) = T(n-1) + \Theta(n)$$

$$= O(n^2) \quad \text{Arithmatic}$$

Analysis of expected time :-

- Let $T(n)$ be the random variable for running time of Random-Select on input size n.
- Assuming random no's are chosen independently.
- define indicator variable X_K for $k = 0, 1, \dots, n-1$

$$X_K = \begin{cases} 1 & \text{if partition generate } k : n-k-1 \text{ split} \\ 0 & \text{otherwise.} \end{cases}$$

$$T(n) \leq \begin{cases} T(\max\{0, n-1\}) + O(n) & \text{if } 0 : n-1 \text{ split} \\ T(\max\{1, n-1\}) + O(n) & \text{if } 1 : n-2 \text{ split} \\ \vdots & \vdots \\ T(\max\{n-1, 0\}) + O(n) & \text{if } n-1 : 0 \text{ split} \end{cases}$$

$$= \sum_{k=0}^{n-1} X_K \left(T(\max\{K, n-k-1\}) + O(n) \right)$$

$$E[T(n)] = E\left[\sum_{K=0}^{n-1} X_K \left(T(\max\{K, n-K-1\}) + O(n) \right) \right]$$

$$= \sum_{K=0}^{n-1} E\left[X_K \left(T(\max\{K, n-K-1\}) + O(n) \right) \right]$$

independent ∵ random variable are choosen independently

$$= \sum_{K=0}^{n-1} \underbrace{E[X_K]}_{1/n} E\left[T(\max\{K, n-K-1\}) + O(n) \right]$$

$$= \frac{1}{n} \sum_{K=0}^{n-1} E\left[T(\max\{K, n-K-1\}) \right] + \frac{1}{n} \sum_{K=0}^{n-1} O(n) \Rightarrow O(n)$$

$$\leq \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} E[T(k)] + O(n)$$

Claim : $E[T(n)] \leq c \cdot n$ for suffi. large const c.

Proof :- Substitution method

Assume true for $< n$

$$E[T(n)] \leq \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} \underbrace{E[T(k)]}_{\leq ck \text{ by Induction hypothesis}} + O(n)$$

$$\leq \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} c \cdot k + O(n)$$

$$\leq \frac{2c}{n} \underbrace{\sum_{k=\lfloor n/2 \rfloor}^{n-1} k}_{\leq \frac{3}{8} \cdot n^2} + O(n)$$

$$= \underbrace{c \cdot n}_{\text{desired}} - \underbrace{\left(\frac{1}{4} c \cdot n - O(n)\right)}_{\text{residual}}$$

$$= O(n)$$

So Random - Select

$O(n)$ expected running

$O(n^2)$ worst case

(3) Worst Case linear-time order Statistics :

idea : generate good pivot recursively

select (i, n)

(1) Divide the n elements into $\lfloor \frac{n}{5} \rfloor$ group of 5 elements.

~~(2) Recursively select~~

- find Median of each group $\Rightarrow O(n)$

(2) Recursively select the median 'x' of the $\lfloor \frac{n}{5} \rfloor$ group medians (median of medians) $\Rightarrow O(n)$

(3) Partition with 'x' as pivot

Let $k =$ rank (x) $\Rightarrow O(n)$

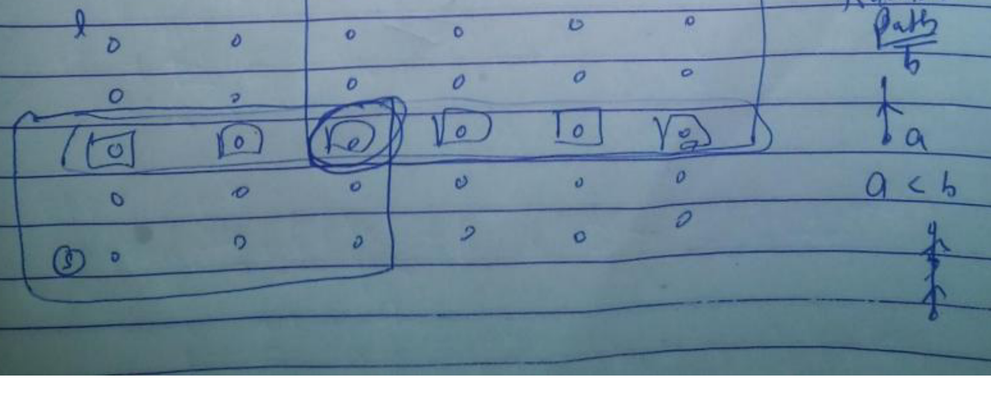
(4) if $i = k$ then return x

if $i < k$ then recursively select i^{th} smallest element in lowest part of array

else

recursively select $(i-k)^{th}$ smallest element in upper part of array $\Rightarrow O(\frac{3n}{4})$

Same randomized select



$$\geq 3\left\lfloor \left\lfloor \frac{n}{5} \right\rfloor / 2 \right\rfloor \text{ elements } \leq X$$

$$\geq \left\lfloor \left\lfloor \frac{n}{5} \right\rfloor / 2 \right\rfloor \text{ group element } \leq X$$

$$\geq 3 \left\lfloor \frac{n}{10} \right\rfloor$$

Simplification: for $n \geq 50$

$$3 \left\lfloor \frac{n}{10} \right\rfloor \geq \frac{n}{4}$$

$$T(n) \leq T\left(\frac{n}{5}\right) + T\left(\frac{3}{4} n\right) + O(n)$$

Claim: $T(n) \leq c \cdot n$

Proof: Substitution

$$T(n) \leq \frac{c}{5} \cdot n + \frac{3}{4} c \cdot n + O(n)$$

$$= \frac{19}{20} c \cdot n + O(n)$$

$$= O(n)$$