Lecture 6 — Order Statistics, Median ①

[redacted] 6.046

given n elements in array
find kth smallest element (elt. of rank k)

Naive algorithm: sort A, return A[k] $\Theta(n \lg n)$

$k=1$: minimum } easy in $\Theta(n)$
$k=n$: maximum

$k = \lfloor \frac{n+1}{2} \rfloor$ or $\lceil \frac{n+1}{2} \rceil$: median ← harder

Randomized divide & conquer

Rand-Select (A, p, q, i) : // ith smallest in A[p..q]

```
if $p = q$
  then return $A[p]$
$r \leftarrow$ Rand-Partition $(A, p, q)$
$k \leftarrow r - p + 1$    // $k = $ rank $(A[r])$ in $A[p..q]$
                        ↳ # of elts $\le A[r]$ including $A[r]$
if $i = k$
  then return $A[r]$
if $i < k$
  then return Rand-Select$(A, p, r-1, i)$
else [redacted] return Rand-Select$(A, r+1, q, i-k)$
```

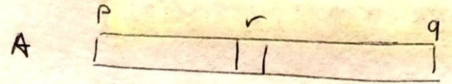
↖ left part recursion, rank remains the same

↖ right part recursion, → rank is offset

Rand-Partition
as subroutine of randomized quicksort:
- pick a random elt x
- swap with the first
- partition s.t.

$A = \overset{p}{[\ \le x\ }\ \overset{r}{\boxed{x}}\ \overset{q}{\ \ge x\ ]}$ $A[r] = x$

- return index of x



if $i<k$ Rand-Select $(A, p, r-1, i)$

if $i>k$ Rand-Select $(A, r+1, q, i-k)$

Ex: $i = 7$

A =

6	10	13	5	8	3	2	11

↑ pivot

2	5	3	6	8	13	10	11

↑ p ↑ r=4 ↑ q

$k=4$ $\underbrace{\qquad}_{\text{rank } 7-4=3}$

Intuition for analysis

(assume elts are distinct)

lucky case : 1/10 : 9/10 $\quad T(n) \le T(\frac{9}{10}n) + \Theta(n) = \Theta(n)$ worst case

case 3 $\quad n^{\log_{10/9} 1} = n^{0}$

unlucky case : 0 : n-1

$$T(n) = T(n-1) + \Theta(n^{*}) = \Theta(n^{2})$$ arithmetic series

↖ much worse than sorting and picking the right element

Analysis of expected time :

- let $T(n)$ be the random variable for running time of Rand-Select on input of size n, assuming random numbers for pivots are independent
- define indicator r.v. X_k for $k = 0, 1, \ldots, n-1$

$$X_k = \begin{cases} 1 & \text{if Partition generates } k : n-k-1 \text{ split} \\ 0 & \text{otherwise} \end{cases}$$

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

$$= \sum_{k=0}^{n-1} X_k \left[T(\max\{k, n-k-1\}) + \Theta(n) \right]$$

$$E[T(n)] = E\left[\sum_{k=0}^{n-1} X_k \left[T(\max\{k, n-k-1\}) + \Theta(n) \right] \right]$$

= [redacted] $\displaystyle\sum_{k=0}^{n-1} E\Big[\underbrace{X_k}_{\text{a random choice of a split (pivot)}} \underbrace{\left[T(\max\{k, n-k-1\}) + \Theta(n) \right]}_{\text{random choices in the recursive call}} \Big]$

random choices in the recursive call [redacted] are independent as the random generator generates [redacted] successive coinflips

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

$$= \frac{1}{n} \sum_{k=0}^{n-1} E\left[T(\max\{k, n-k-1\}) \right] + \underbrace{\frac{1}{n} \sum_{k=0}^{n-1} \Theta(n)}_{\Theta(n} \le \frac{2}{n} \sum_{k=\lfloor \frac{n}{2} \rfloor}^{n-1} E[T(k)] + \Theta(n)$$

lecture 6

②

Claim: $E[T(n)] \le cn$ for suff. large const. $c > 0$

Proof: Substitution method

Assume true for $< n$ ← IH

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

↑ each $k < n$, apply IH

$$\le \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} ck + \Theta(n) \quad \text{by IH}$$

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

fact for series... can proved by induction

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

positive for c sufficiently large

[redacted] choose $c \ge 4 \cdot$ constant in $\Theta(n)$

my comment: look at [redacted] c necessary for small n

[redacted]

choose the [redacted] largest

Rand-Select

$T(n) = \Theta(n)$ in expectation

$\Theta(n^2)$ worst case

worst-case linear-time order statistics

(Blum, Floyd, Pratt, Rivest, Tarjan 1973)

- **idea**: generate good pivot recursively

Select(i, n):

1) Divide the n elements into $\lfloor n/5 \rfloor$ groups of 5 elts each

[redacted] Find the median of each group → $\Theta(n)$

2) Recursively Select the median x of the $\lfloor n/5 \rfloor$ group medians → $T(\frac{n}{5})$

3) Partition with x as pivot
Let k = rank(x) ← median of medians } $\Theta(n)$

4) if i = k then return x
if i < k then recursively Select ith smallest elt in the lower part of the array
else ——— (i−k)th smallest elt ——— upper part of the array → $T(\frac{3}{4}n)$

as Rand-Select

$$T(n) \le T(n/5) + T\left(\frac{3}{4}n\right) + \Theta(n)$$

my comment ↑ add up to less than 1, need it for $\Theta(n)$ [redacted]

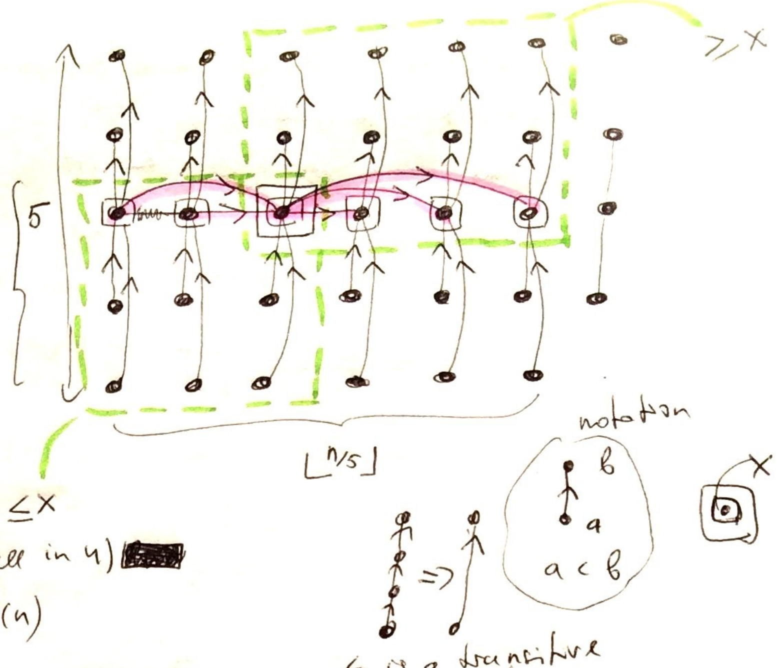
[redacted] (sel: case 3) ← not directly applicable, because of different sizes

$\geq 3\lfloor \lfloor n/5 \rfloor /2 \rfloor$ elts $\leq x$

$\geq x$

$\underbrace{\geq \lfloor \lfloor n/5 \rfloor /2 \rfloor}_{\lfloor n/10 \rfloor}$ group medians $\leq x$

$\Rightarrow \geq 3 \lfloor n/10 \rfloor$ elts $\leq x$



$\leq$ is a transitive relation [redacted]

Simplification

for $n \geq 50$

$3 \lfloor n/10 \rfloor \geq n/4$

my comment

guarantee : in the recursive call in 4) [redacted]

$T(n) \leq T(\frac{n}{5}) + T(\frac{3}{4}n) + \Theta(n)$

Claim $T(n) \leq cn$

Proof : Substitution

$T(n) \leq \frac{c}{5}n + \frac{3}{4}cn + \Theta(n)$ by IH

$= \frac{19}{20}cn + \Theta(n)$

$= cn - (\frac{1}{20}cn - \Theta(n))$

set $c \geq 20$ const in $\Theta(n)$, and check base case if need a larger c

[redacted]

$\leq cn$ for c sufficiently large