CompSci 161
Spring 2021 Lecture 06:
Divide and Conquer II:
QuickSort and Order Statistics

2

QuickSort Step 1: Partition

85	24	63	45	17	31	96	50
	L		L	L	L		↑

1. Choose a pivot. } $O(1)$
2. Place that pivot in the right spot. $O(n)$
3. Pivot the rest of the array. $O(n)$

3 ## QuickSort

right spot! :)

85	24	63	45	17	31	96	50
17	24	31	45	50	63	96	~~17~~ 85

4 ## How fast is QuickSort?

- $T(n) = T(\text{lower}) + T(\text{upper}) + \Theta(n)$
- If lower and upper are both size $n/2$?

$$T(n) = 2T(n/2) + \theta(n) \Rightarrow \theta(n \lg n)$$

- What if we select a pivot uniformly at random?
- What if we could find a median in $\Theta(n)$...

5

Average Case Analysis of QuickSort

Suppose

- All permutations equally likely
- All n values are distinct (for simplicity)
- Define $S_1, S_2, \ldots S_n$ as sorted order.

Let $P_{i,j}$ be probability we compare S_i and S_j.

must be chosen as pivot before $\underbrace{S_{i+1} \ldots S_j}_{j-i+1 \text{ items if we incl } S_i, S_j}$

$$P_{i,j} = \frac{\#\text{yes}}{\#\text{total}} = \frac{2}{j-i+1}$$

6

Expected number of comparisons

$X_{i,j}$ = I.R.V.: do i,j get compared?

$$\begin{aligned}
E\Big(\sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{i,j}\Big) &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} E(X_{i,j}) \\
&= \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&= \sum_{i=1}^{n} \sum_{k=2}^{n-i+1} \frac{2}{k} \\
&< \sum_{i=1}^{n} \sum_{k=1}^{n} \frac{2}{k} = 2 \sum_{i=1}^{n} \underbrace{\sum_{k=1}^{n} \frac{1}{k}}_{O(\log n)}
\end{aligned}$$

$O(n \log n)$

7

The Selection Problem

- Given a list S and numeric k
- Want: if we sorted S, what is S_k? ←
- Brute force:
  - Sort S in $\Theta(n \log n)$
  - Return S_k
- Can we do better?

8

Randomized Selection

quickSelect(S, k)

If n is small, brute force and return. ←

Pick a random $x \in S$ and put rest into:

L, elements smaller than x

G, elements greater than x

partition step like QuickSort

if $k \leq |L|$ **then**

return quickSelect(L, K)

else if $k == |L| + 1$ **then**

return x

else

return quickSelect(G, K-(|L|+1))

9

Randomized Selection

- What is the worst-case running time?
- What would cause that bad time?
- Estimate the *expected* running time?
  Hint: on average, the pivot is the median.

$T(n) = T(n/2) + \Theta(n)$ is $\Theta(n)$

10

Deterministic Selection

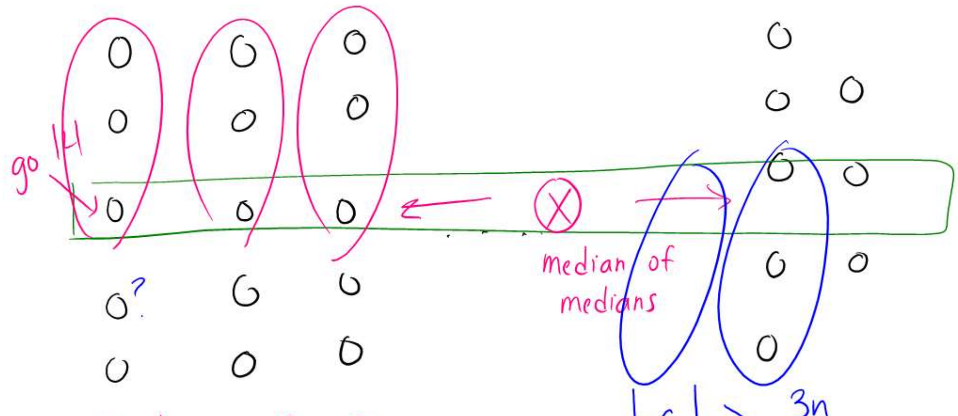
- Instead of picking x at random:
  - Divide S into $g = \lceil n/5 \rceil$ groups
  - Each group has 5 elements (except maybe g^{th})
  - Find median of each group of 5 $\rightarrow O(1)$ each $\times\ \theta(n)$
  - Find median of those medians $\longrightarrow T(n/5)$
  - Let x be that median.
- Let's talk about that pivot.
  - Could it be the smallest?
  - Could it be the largest?
  - How close to median is it?
- Set up a recurrence for this version of selection.

$T(n) = T(n/5) + T(\text{region of recurrence}) + \Theta(n)$

11

Deterministic Selection

Let's visualize: how does pivot compare to list?



$|L| \geq \frac{n}{10} \cdot 3$ $\qquad |G| \geq \frac{3n}{10}$

$\left|\text{"region we recurse into"}\right| \leq \frac{7n}{10} = n - \frac{3n}{10}$

12

Deterministic Selection

- How few elements *must be* smaller than pivot?

  $3n/10$

- How few *must be* non-smaller than pivot?

  $3n/10$

- How many could be in either group?

$$T(n) = T(n/5) + T(7n/10) + \Theta(n)$$

is $\Theta(n)$