Selection

Given n numbers find the i^{th} smallest

3 20 1 5 67 2 40 8 9

$i=1$	1	minimum	$O(n)$
$i=9$	67	maximum	$O(n)$
$i=5$	8	median	
⋮			
$i=3$		arbitrary	

1 2 3 5 8 9 20 40 67

Alg

Sort A
Return A[i]

Median is "not harder than" Sorting.

Sep 24-4:09 PM

Q: Is median (selection) harder than sorting?

i.e. Is it necessary to sort to find the median?

Can do selection (median) in $O(n)$ time.

Sep 24-4:17 PM

3 8 12 20 2 4 7 11 36 1

$n=10$

$i=3$

$p=7$

$L=\{3\ 2\ 4\ 7\ 1\}$

~~$H=\{8\ 12\ 20\ 11\ 36\}$~~

Select$(L,3,5)$ $p=3$ $L=\{3\ 2\ 1\}$ ~~$H=\{4\ 7\}$~~

Select$(L,3,3)$ $p=2$ ~~$L=\{2\ 1\}$~~ $H=\{3\}$

Select$(H,1,1) \to 3$

$$T(n) = T(n/2) + M + O(n)$$

Suppose $M = O(n)$

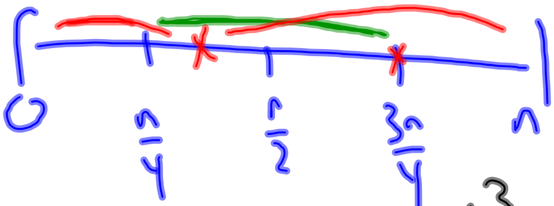
$$= T(n/2) + O(n)$$

$n^{\lg_2 1}$ n $O(n)$.

Suppose, in $O(n)$ time, we can find a number "close" to the median

This is recursive

Suppose we can find a number in the middle half



Running time $T(n) \le O(n) + T(\frac{3}{4}n) = O(n)$

$n \quad n^{\lg 4/3}$!

Sep 24-4:30 PM

dselect.pdf (application/pdf Object) - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.columbia.edu/~cs2035/courses/csor4231.F09/dselect.pdf

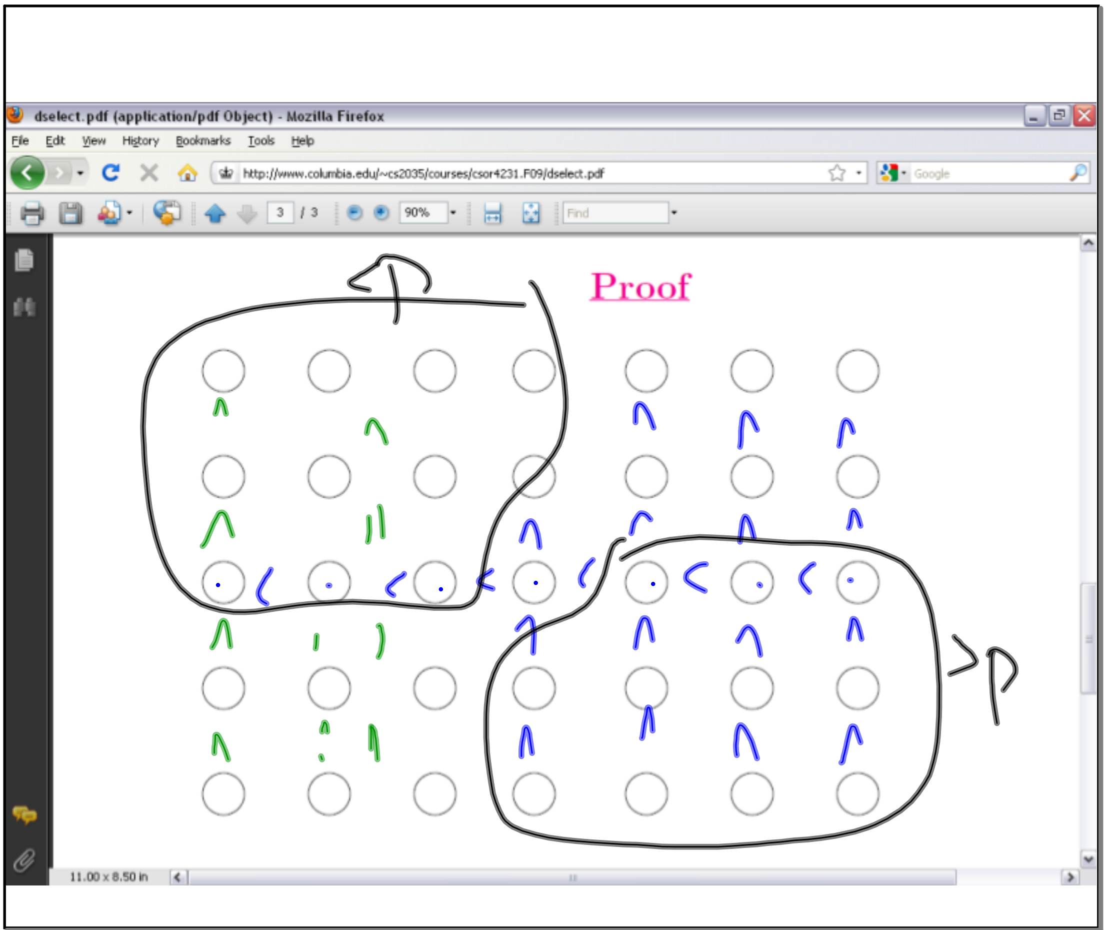
SELECT(A,i,n)

```
if (n = 1)
    return A

Split the items into ⌊n/5⌋ groups 5 (and one more group).
    Call these groups G_1, G_2, ..., G_⌊n/5⌋
Find the median m_i of each G_i
Recursively compute the median of medians,
    p = SELECT({m_1, ..., m_⌊n/5⌋}, ⌊n/10⌋, ⌊n/5⌋)

L = {x ∈ A : x ≤ p}
H = {x ∈ A : x > p}

if i ≤ |L|
    SELECT(L, i, |L|)
else SELECT(H, i − |L|, |H|)
```

Finds a number in the middle "half"

11.00 x 8.50 in

Sep 24-4:39 PM



Sep 24-4:47 PM

dselect.pdf (application/pdf Object) - Mozilla Firefox

http://www.columbia.edu/~cs2035/courses/csor4231.F09/dselect.pdf

SELECT(A,i,n)

if $(n = 1)$ $\quad 1$
  return A

Split the items into $\lfloor n/5 \rfloor$ groups 5 (and one more group).
  Call these groups $G_1, G_2, \ldots, G_{\lfloor n/5 \rfloor}$
Find the median m_i of each G_i $\quad \leftarrow \frac{n}{5}(O(1)) = O(n)$
Recursively compute the median of medians,
  p = SELECT($\{m_1, \ldots, m_{\lfloor n/5 \rfloor}\}, \lfloor n/10 \rfloor, \lfloor n/5 \rfloor$) $\quad \leftarrow T(\frac{n}{5})$

$L = \{x \in A : x \leq p\}$
$H = \{x \in A : x > p\}$ $\quad \leftarrow O(n)$

if $i \leq |L|$
  SELECT($L, i, |L|$)
else SELECT($H, i - |L|, |H|$) $\quad \leftarrow \leq T(\frac{3}{4}n)$

Sep 24-4:49 PM

$$T(n) = T\left(\tfrac{3n}{4}\right) + T\left(\tfrac{n}{5}\right) + \overset{kn}{\Theta(n)}$$

for some k (I can't choose)

<u>Claim</u> $T(n) = O(n)$

$T(n) \le cn$ for some c (I choose)

<u>Pf</u>

$$T(n) = T\left(\tfrac{3n}{4}\right) + T\left(\tfrac{n}{5}\right) + kn$$

$$\le \tfrac{3}{4}cn + \tfrac{1}{5}cn + kn$$

$$= \tfrac{19}{20}cn + kn = cn - \tfrac{1}{20}cn + kn$$

need $-\frac{1}{20}cn + kn \le 0$

$\iff \left(-\frac{c}{20} + k\right)n \le 0$

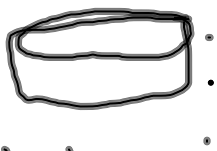
$\iff -\frac{c}{20} + k \le 0$

$\iff c \ge 20k$

set $c = 20k$

$\le cn$ ∎

Sep 24-4:44 PM

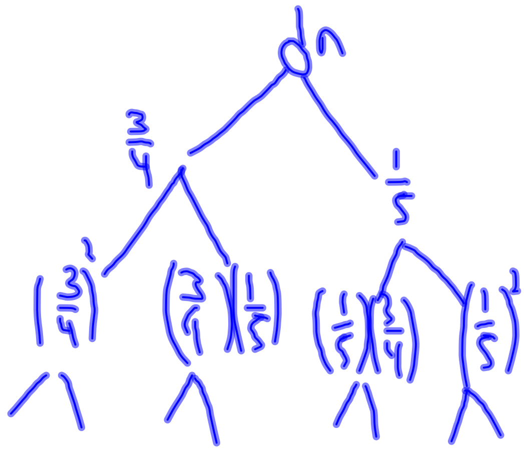


$\frac{2}{3} : \frac{1}{2}$ $\quad \frac{1}{3}$ $\quad \frac{3n}{4}$ $\quad \frac{n}{5}$

$T\left(\frac{2n}{3}\right)$ $\quad T\left(\frac{n}{3}\right)$

Sep 24-4:55 PM

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



Work

n

$\left(\frac{3}{4}+\frac{1}{5}\right)n = \frac{19}{20}n$

$\left(\frac{3}{4}+\frac{1}{5}\right)^2 = \left(\frac{19}{20}\right)^2 n$

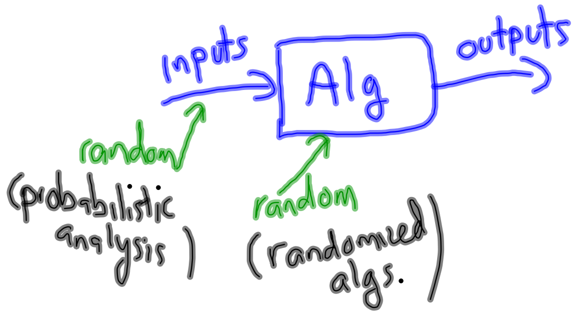
$\left(\frac{3}{4}+\frac{1}{5}\right)^3 = \left(\frac{19}{20}\right)^3 n$

$$\sum \text{work} \leq n\left(1 + \frac{19}{20} + \frac{19}{20}^2 + \frac{19}{20}^3 + \cdots\right) = O(n)$$

Sep 24-4:57 PM

Randomization in Algorithms

Randomization is a tool for designing good algorithms



analyze: running time
correctness

Las Vegas - always correct, running time is random

Monte Carlo - may return incorrect answer, but running time is deterministic

Sep 24-5:02 PM

Hiring Problem

best = 0; hired = i
for i = 1 to n
    interview candidate i, give score
    if (score > best)
        fire hired
        hired = i
        best = score

how many people are hired

Sep 24-5:10 PM

How many hirings   Not n/2.

Random variable X
which takes on values from some set, each w/ a probability.

$$E[X] = \sum_{\text{values } x} \Pr(X=x) \cdot x$$

X = roll of a die

	1	2	3	4	5	6
Pr	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E[X] = \tfrac{1}{6}(1) + \tfrac{1}{6}(2) + \tfrac{1}{6}(3) + \tfrac{1}{6}(4) + \tfrac{1}{6}(5) + \tfrac{1}{6}(6)$$
$$= 3.5$$

$$E[X] = \sum_{\text{outcomes } x \text{ of } X} \Pr(X=x) \cdot \text{value}(x)$$

Sep 24-5:17 PM

Expected # of hirings
(assume all orderings of candidates are equally likely)

$n!$ orderings $\pi_1, \pi_2, \ldots \pi_{n!}$

H = # hirings

$h(\pi_i)$ = # hirings for permutation π_i

$$E[h] = \sum_{\text{all } \pi_i} \frac{1}{n!} h(\pi_i)$$

Sep 24-5:23 PM