Same start as for deterministic selection

SELECT(**A,i,n**)

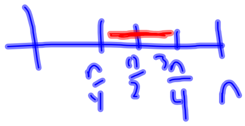
```
if (n = 1)
    then return A[1]

p = A[RANDOM(1, n)]


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

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

$\frac{1}{2}$ the time
p is in the middle half



Oct 1-4:09 PM

rselect.pdf (application/pdf Object) - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.columbia.edu/~cs2035/courses/csor4231.F09/rselect.pdf
2 / 6 100% Find

SELECT(**A,i,n**)

```
if (n = 1)
    then return A[1]

p = A[RANDOM(1, n)]


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

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

first or last is bad

p minimum
$|L| = 1 \quad |H| = n-1$

p max
$|L| = n \quad |H| = 0$

worst-case running time
∞ running time.

11.00 x 8.50 in

Oct 1-4:21 PM

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

File Edit View History Bookmarks Tools Help

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

3 / 6 100%

Analysis

T(n) is expected running time.

$T(n) = \sum_{x=1}^{n}$ Pr(partition is x smallest)·(Running time when partition is x sm

Using x and $n-x$ as an upper bound of the sizes of the two sides:

$$\begin{aligned} T(n) &\leq \sum_{x=1}^{n} \frac{1}{n}((T(x) \text{ or } T(n-x)) + O(n)) \\ &\leq \sum_{x=1}^{n} \frac{1}{n}(T(\max\{x, n-x\}) + O(n)) \\ &\leq \left(\frac{1}{n}\right) \sum_{x=1}^{n} (T(\max\{x, n-x\})) + O(n) \end{aligned}$$

We now rewrite the max term. Notice that as x goes from 1 to n, the term $\max\{x, n-x\}$ takes on the values $n-1, n-2, n-3, \ldots, n/2, n/2, n/2+1, n/2+2, \ldots, n-1, n$. As an overestimate, we say that it takes all the values between $n/2$ and n twice. Thus we substitute and obtain

11.00 x 8.50 in

Oct 1-4:23 PM

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

File Edit View History Bookmarks Tools Help

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

3 / 6 100%

$(n) = \sum_{x=1}^{n}$ Pr(partition is x smallest)·(Running time when partition is x smal

Using x and $n-x$ as an upper bound of the sizes of the two sides:

$n-1$, $n-2$, $\vdots$, $n/2$ $\qquad$ n, $n-1$, $n-2$, $n/2+1$

$$\begin{aligned} T(n) &\leq \sum_{x=1}^{n} \frac{1}{n}((T(x) \text{ or } T(n-x)) + O(n)) \\ &\leq \sum_{x=1}^{n} \frac{1}{n}(T(\max\{x, n-x\}) + O(n)) \\ &\leq \left(\frac{1}{n}\right) \sum_{x=1}^{n} (T(\max\{x, n-x\})) + O(n) \end{aligned}$$

We now rewrite the max term. Notice that as x goes from 1 to n, the erm $\max\{x, n-x\}$ takes on the values $n-1, n-2, n-3, \ldots, n/2, n/2, n/2+$ $n/2+2, \ldots, n-1, n$. As an overestimate, we say that it takes all the values etween $n/2$ and n twice. Thus we substitute and obtain

11.00 x 8.50 in

Oct 1-4:29 PM

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

File Edit View History Bookmarks Tools Help

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

Using x and $n-x$ as an upper bound of the sizes of the two sides:

$$\begin{aligned} T(n) &\leq \sum_{x=1}^{n} \frac{1}{n}((T(x) \text{ or } T(n-x)) + O(n)) \\ &\leq \sum_{x=1}^{n} \frac{1}{n}(T(\max\{x, n-x\}) + O(n)) \\ &\leq \left(\frac{1}{n}\right) \sum_{x=1}^{n} (T(\max\{x, n-x\})) + O(n) \end{aligned}$$

We now rewrite the max term. Notice that as x goes from 1 to n, the term $\max\{x, n-x\}$ takes on the values $n-1, n-2, n-3, \ldots, n/2, n/2, n/2+1, n/2+2, \ldots, n-1, n$. As an overestimate, we say that it takes all the values between $n/2$ and n twice. Thus we substitute and obtain

$$\begin{aligned} T(n) &\leq \left(\frac{2}{n} \sum_{x=0}^{n/2} T(n/2+x)\right) + O(n) \\ &= \frac{2}{n} T(n) + \left(\frac{2}{n} \sum_{x=0}^{n/2-1} T(n/2+x)\right) + O(n) \end{aligned}$$

11.00 x 8.50 in

Oct 1-4:44 PM

QuickSort

2 5 7 1 4 9 8 3 6

L | 2 | 5 | 1 | 4 | 3 | 7 | 9 | 8 | 6 | H

1 2 3 4 5 6 7 8 9

Oct 1-4:46 PM

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

File Edit View History Bookmarks Tools Help

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

```
        QUICKSORT(A, p, q − 1)
        QUICKSORT(A, q + 1, r)

PARTITION(A, p, r)

y ← RANDOM(p, r)
Exchange A[y] and A[r]
x ← A[r]
i ← p − 1
for j ← p to r − 1
    do if A[j] ≤ x
        then i ← i + 1
            exchange A[i] ↔ A[j]
exchange A[i + 1] ↔ A[r]
return i + 1
```

$x = 6$

1 2 3 4 5 6 7 8

A[8 3 6 1 4 9 2 5]

3 5 1 4 2 9 8 6

3 5 1 4 2 | 6 | 8 9

Oct 1-4:56 PM

---

if partition is x^{th} smallest $\quad T(n) = T(x) + T(n-x) + O(n)$

intuition

if $x = \frac{n}{2}$ $\quad T(n) = T(\frac{n}{2}) + T(\frac{n}{2}) + O(n)$

$= 2T(\frac{n}{2}) + O(n)$

$= O(n \lg n)$

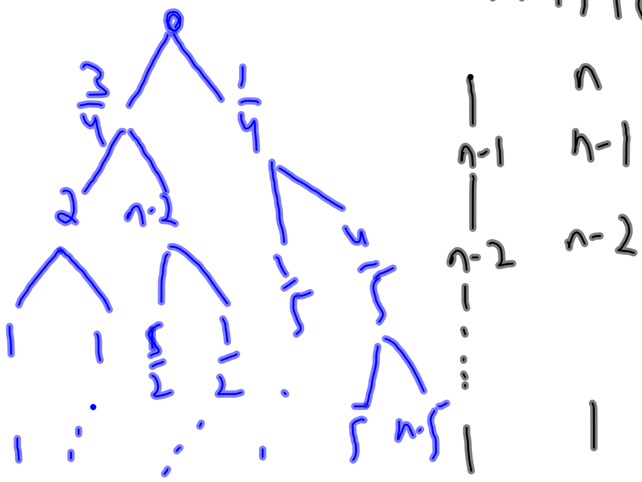
if $x = \frac{n}{10}$ $\quad T(n) = T(\frac{n}{10}) + T(\frac{9n}{10}) + O(n)$

$= O(n \lg n)$

Oct 1-5:00 PM

$x = 1$

$$T(n) = T(1) + T(n-1) + O(n)$$
$$= T(n-1) + O(n)$$



$$\sum_{i=1}^{n} i = O(n^2).$$

Oct 1-5:03 PM

Following selection

$$T(n) = \sum_{i=1}^{n} \frac{1}{n} \cdot \left(T(i) + T(n-i) + O(n) \right)$$

Solve this

⋮

Oct 1-5:05 PM

Count comparisons of array elements

All comparisons are from line 6 of PARTITION

$A[j] \le x$ x is the pivot

∴ all comparisons involve the pivot.

after one run of PARTITION, pivot is fixed in the right place.

∴ each pair of elements is compared at most once.

Oct 1-5:06 PM

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

File Edit View History Bookmarks Tools Help

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

2 / 2   100%   Find

1 $y \leftarrow \text{RANDOM}(p, r)$
2 **Exchange** $A[y]$ **and** $A[r]$
3 $x \leftarrow A[r]$
4 $i \leftarrow p - 1$
5 **for** $j \leftarrow p$ **to** $r - 1$
6 **do if** $A[j] \le x$
7 **then** $i \leftarrow i + 1$
8 **exchange** $A[i] \leftrightarrow A[j]$
9 **exchange** $A[i + 1] \leftrightarrow A[r]$
10 **return** $i + 1$

Loop Invariant At the beginning of each iteration of the for loop in Partition

1. $A[p \ldots i] \le x$
2. $A[i + 1 \ldots j - 1] > x$

11.00 x 8.50 in

Oct 1-5:08 PM

let data be $z_1, \dots z_n$ in sorted order

$$Z_{ij} = \{z_i, z_{i+1}, \dots, z_{j-1}, z_j\}$$

example
recurse on z_{15} z_{78}

Let X_{ij} = i.r.v. for z_i compared to z_j

X = total # comps.

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} \qquad E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[X_{ij}]$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Pr(z_i \text{ is comp. to } z_j).$$

Oct 1-5:08 PM

What is $Pr(z_i \text{ compared to } z_j)$?

When is z_i compared to z_j

- either z_i or z_j is chosen as pivot & the other one (of z_i & z_j) is still in the same set.

$\Leftrightarrow$ in the set of numbers Z_{ij}, either z_i or z_j is the first pivot chosen out of Z_{ij}

$\dots\ z_{10}\ z_{11}\ \underline{z_{12}}\ z_{13}\ z_{14}\ z_{15}\ z_{16}\ \dots$

if z_{12} chosen first out of $Z_{10,16}$ z_{10} & z_{16} are never compared but if z_{16} or z_{16} chosen first then they are compared.

Oct 1-5:15 PM

$\Pr(\text{from } Z_{ij},\ Z_i \text{ or } Z_j \text{ is first chosen})$

$$= \frac{2}{j-i+1}$$

$$E(X) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}$$

$k = j-i+1$

$$= \sum_{i=1}^{n} \sum_{k=2}^{n-i+1} \frac{2}{k} = 2\sum_{i=1}^{n} \ln(n-i+1) \le 2\sum_{i=1}^{n} \ln n$$

$$= O(n \lg n).$$

Oct 1-5:21 PM