Randomized Selection

Same start as for deterministic selection

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

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

$p = \text{MEDIAN}(A)$

$L = \{x \in A : x \leq p\}$
$H = \{x \in A : x > p\}$

if $i \leq |L|$
    then $\text{SELECT}(L, i, |L|)$
    else $\text{SELECT}(H, i - |L|, |H|)$

Choose pivot p randomly.

Randomized Selection

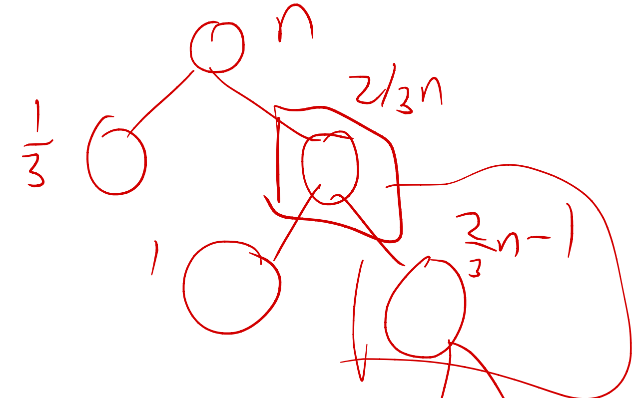
Same start as for deterministic selection

```
SELECT(A,i,n)

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

p = ~~MEDIAN~~(A)        complicated deterministic
                         middle half(A)

    like something "near" the middle.

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

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

Choose pivot p randomly.

A [2, 7, 10, 1, 4, 6]

Randomized Selection



Same start as for deterministic selection

Select**(A,i,n)**

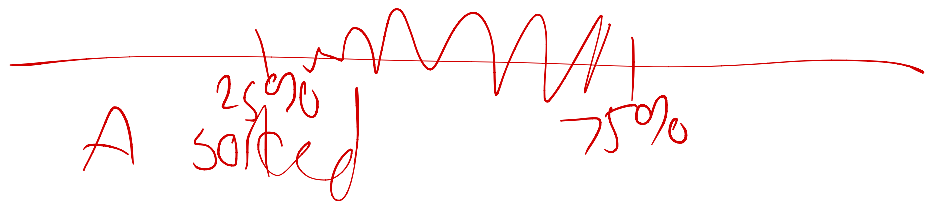
if $(n = 1)$ O(1)
    then return $A[1]$ median

$p = A[\textsc{Random}(1, n)]$ O(1)

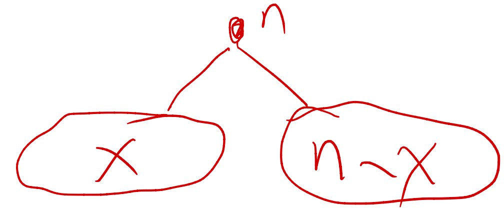
$L = \{x \in A : x \leq p\}$ O(n)
$H = \{x \in A : x > p\}$

if $i \leq |L|$ T(x) or T(n-x)
    then Select$(L, i, |L|)$
    else Select$(H, i - |L|, |H|)$

Pr(A[p] in middle half of a sorted A) = 1/2



expected running time



Analysis

$\frac{1}{n}$

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

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

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

if $x < \frac{n}{2}$ $\max(x, n-x) = n-x$

if $x > \frac{n}{2}$ $\max(x, n-x) = x$

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**

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

Analysis

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

We pulled out the $T(n)$ terms to emphasize them. We might be worried about having $T(n)$ on the right side of the equation, so we will bring it over the left-hand side and obtain

$$\left(1-\frac{2}{n}\right)T(n) \le \left(\frac{2}{n}\sum_{x=0}^{n/2-1} T(n/2+x)\right) + O(n)\ .$$

We now multiply both sides of the inequality by $n/(n-2)$ to obtain:

$$T(n) \le \left(\frac{2}{n-2}\sum_{x=0}^{n/2-1} T(n/2+x)\right) + kn^2/(n-2)\ .$$

We have replaced the $O(n)$ by kn for some constant k before multiplying by $n/(n-2)$. We do this because we will need to for the proof by induction below.

We now have a recurrence in a nice form. $T(n)$ is on the left, and the right has terms of the form $T(x)$ for $x < n$. We can therefore "guess" that $T(n) = O(n)$ and try to prove it. More precisely, we will prove by induction that $T(n) \le cn$ for some c. Since the recurrence is in the stated form, we can substitute in on the right hand side and obtain

Analysis

$\forall n \geq 14$

induction

$T(n) \leq cn$

$$
\begin{aligned}
T(n) &\leq \left(\frac{2}{n-2}\sum_{x=0}^{n/2-1} \boxed{T(n/2+x)}\right) + kn^2/(n-2) \\
&\leq \left(\frac{2}{n-2}\sum_{x=0}^{n/2-1} c(n/2+x)\right) + kn^2/(n-2) \\
&= \left(\frac{2c}{n-2}\right)\left((n/2)(n/2) + (n/2-1)(n/2)/2\right) + kn^2/(n-2) \\
&= \left(\frac{2c}{n-2}\right)\left(3n^2/8 - n/4\right) + kn^2/(n-2) \\
&= \left(\frac{c}{n-2}\right)\left(3n^2/4 - n/\cancel{8}\right) + kn^2/(n-2) \\
&= \frac{1}{n-2}\left((3c/4+k)n^2 - (c/\cancel{8})n\right) \\
&= \frac{n}{n-2}\left((3c/4+k)n - (c/\cancel{8})\right)
\end{aligned}
$$

$\frac{n}{2} + (\frac{n}{2}+1) + \cdots + (n-1)$

$3, 2, \frac{5}{3}, \frac{6}{4}, \frac{7}{5}, \ldots \frac{100}{98},$

$\leq cn$?

Looking at this last term, we see that the leading $n/(n-2)$ is slightly larger than 1, so we can upper bound it by, say $7/6$ for $n \geq 14$ (there are many possible choices of upper bounds.) Our goal, remember, is to show that the term multiplying the n is at most c, and as we will see, this suffices.

So we get

$$T(n) \leq (7/6)\left((3c/4+k)n - (c/8)\right) \ .$$

Analysis

$$T(n) \leq (7/6)\left((3c/4 + k)n - (c/\overset{2}{\cancel{8}})\right) \ .$$

If the right hand side is at most cn we are done. Whether it is will depend on the relative values of c and k. Let's write the constraint we want

$$(7/6)\left(\boxed{(3c/4 + k)}n - (c/\overset{2}{\cancel{8}})\right) \leq cn$$

and solve for c in terms of k. We get

$$(7c/8 + 7k/6 - c)n \leq 7c/48$$

or
$(7k/6 - c/8)n \leq 7c/48.$
Clearly, if $7k/6 - c/8 < 0$ this will hold. So we just choose c sufficiently larger than k, e.g. $c = 28k/3$ and we are done.

$$\frac{21}{24}c + \frac{70}{6} \leq c \qquad c = 10$$