Selection

Given n numbers find the kth smallest

3 20 1 5 6 7 2 4 0 8 9

l = 1  1  minimum  O(n)

l = 9  9  maximum  O(n)

l = 5  8  median

l = 3  arbitrary

Alg
Sort A
Return A[i]

Median is "not harder than"

Sorting.

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.
\[ \begin{align*}
3 & \quad 8 & \quad 12 & \quad 20 & \quad 2 & \quad 4 & \quad 7 & \quad 11 & \quad 36 & \quad 1 \\
\text{n=10} & \quad \text{i=3} & \quad p=7 \\
L &= \{3, 24, 7, 1\} \\
H &= \{8, 12, 20, 11, 36\} \\
\text{Select} (L, 3, 5) & \quad p = 3 \\
\text{Select} (L, 3, 3) & \quad p = 2 \\
\text{Select} (H, 1, 1) & \Rightarrow 3
\end{align*} \]

\[ T(n) = T(n/2) + M + O(n) \]

Suppose \( M = O(n) \)

\( n \log n \quad \circ \quad O(n). \)
Suppose, in $O(n)$ time, we can find a number "close" to the median.

Suppose we can find a number in the middle half.

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

---

**Algorithm**

`SELECT(A, i, n)`

1. if $(n = 1)$
   - return $A$
2. Split the items into $\lfloor n/5 \rfloor$ groups 5 (and one more group).
   - Call these groups $G_i, G_2, \ldots, G_{\lfloor n/5 \rfloor}$
3. Find the median $m_i$ of each $G_i$
4. Recursively compute the median of medians,
   - $p = \text{SELECT}(\{m_1, \ldots, m_{\lfloor n/5 \rfloor}\}, \lfloor n/10 \rfloor, \lfloor n/5 \rfloor)$
5. $L = \{x \in A : x \leq p\}$
   - $H = \{x \in A : x > p\}$
6. if $i \leq |L|$
   - \text{SELECT}(L, i, |L|)
   - return $L$
7. else \text{SELECT}(H, i - |L|, |H|)

Finds a number in the middle half.
SELECT(A, i, n)
1 if \( (n = 1) \)
2 \quad \text{return } A
3 Split the items into \([n/5]\) groups 5 (and one more group). 
   Call these groups \( G_1, G_2, \ldots, G_{[n/5]} \)
4 Find the median \( m_i \) of each \( G_i \)
5 Recursively compute the median of medians, 
   \( p = \text{SELECT}\{m_1, \ldots, m_{[n/5]}\}, [n/10], [n/5]\) 
6 \( L = \{x \in A : x \leq p\} \)
7 \( H = \{x \in A : x > p\} \)
8 if \( i \leq |L| \)
9 \quad \text{SELECT}(L, i, |L|)
10 else \text{SELECT}(H, i - |L|, |H|)

\[ \frac{n}{5} \in O(n) \quad \leq T\left(\frac{n}{5}\right) \]
\[ O(n) \quad \leq T\left(\frac{3}{4}n\right) \]
\[ T(n) = T\left(\frac{3n}{4}\right) + T\left(\frac{n}{2}\right) + o(n) \]

\[ \text{Claim: } T(n) = O(n) \]

\[ T(n) \leq cn \text{ for some } c \text{ (I choose)} \]

\[ \text{Pf: } T(n) = T\left(\frac{3n}{4}\right) + T\left(\frac{n}{2}\right) + kn \]

\[ \leq \frac{3}{4}cn + \frac{1}{2}cn + kn \]

\[ = \frac{5}{4}cn + kn = \frac{cn - \frac{1}{2}cn + kn}{\frac{5}{2}} \]

\[ \Rightarrow \frac{1}{2}cn + kn \leq 0 \]

\[ \Rightarrow \left(-\frac{1}{20} + k\right)n \leq 0 \]

\[ \Rightarrow -\frac{1}{20} + k \leq 0 \]

\[ \Rightarrow c \geq 20k \]

\[ \text{Set } c = 20k \]

\[ \leq cn \]
\[ T(n) = T\left(\frac{3n}{4}\right) + T\left(\frac{n}{5}\right) + n \]

\[ \frac{\text{work}}{n} = \left(\frac{3}{4} + \frac{1}{5}\right) n = \frac{19}{20} n \]

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

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

\[ \sum \text{work} \leq n \left(1 + \frac{19}{20} + \frac{19}{20} + \cdots \right) = \Theta(n) \]

---

**Randomization in Algorithms**

Randomization is a tool for designing good algorithms.

Inputs → Alg → Outputs

- Analyze: running-time, correctness
- Random (randomized alg.
- Probabilistic analysis

- Las Vegas: always correct, running-time is random
- Monte Carlo: may return incorrect answer, but running-time is deterministic
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
   fire
   hired = 0

How many people are hired

Random variable $X$
which takes on values from some set,
each with a probability:

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

$X =$ roll of a die $1 2 3 4 5 6$

$\Pr \begin{cases} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{cases}$

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

$$= 3.5$$

$$E[X] = \sum_{outcomes \, x} \Pr(X=x) \cdot value(x)$$
Expected # of hirings

(given all orderings of candidates are equally likely)

\[ n! \text{ orderings } \pi_1, \pi_2, \ldots, \pi_n. \]

\[ H = \# \text{ hirings} \]
\[ h(\pi_i) = \# \text{ hirings for permutations } \pi_i \]

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