Plan

* Recall Median Finding Algorithm
* Analysis Tool: Expected Running Time
* Announcements
* Expected RT Analysis
Given: a list \( L \) of \( n \) distinct integer

Task: Return the \( k^{th} \) smallest element

\[
\left| \{ r \in L : r \leq s \} \right| = k
\]

\[
\left| \{ t \in L : t > s \} \right| = n - k
\]
Selection without Sorting

Divide

Choose a "pivot" p.

\[ X \rightarrow \text{partition around } p \]

\[ \text{partition around } p \]

Conquer

To find \( k^{th} \) element, consider \( l = \left| L \leq \right| \) compared to \( k \).
\textbf{Select} (L, k)

Choose pivot $p \in L$.

$L_\leq \leftarrow \{i \in L : i \leq p\}$ \hspace{1cm} \// ensure $p$ is final element of $L_\leq$

$L_\succ \leftarrow \{j \in L : j > p\}$

let $l = |L_\leq|

if $l = k$ : Return $p$ \hspace{1cm} \// pivot was $k$th elem

if $l > k$ : Return Select ($L_\leq$, $k$)

else : Return Select ($L_\succ$, $k-l$)
Theorem: For any deterministic pivot selection (that does not depend on L), the worst-case running time of Select is $\Omega(n^2)$.

Idea: Use a RANDOM pivot selection.
Basic Randomness Primitives

* Choose random bit $B \in \{0,1\}$ w.p. $\frac{1}{2}$

* Given $n$, choose $Z \in \{1, \ldots, n\}$ uniformly at random

$$
\forall i \in \{1, \ldots, n\}, \quad \Pr[Z = i] = \frac{1}{n}
$$
Randomized Algorithms

* Algorithms may "flip coins" / "role dice"

i.e. draw $Z \leftarrow \{1, \ldots, n\}$

uniformly at random
Randomized Algorithms

* Algorithms may “flip coins” / “role dice”

i.e. draw $Z \leftarrow \{1, \ldots, n\}$ uniformly at random

* Running Times?

  - Define a Random Variable for the RT
  - Give an upper bound on the Expectation

  $\Rightarrow$ Expected Running Time
**Randomized Select** \((L, k)\)

Choose pivot \(p \leftarrow L[Z]\) for \(Z \leftarrow \{1, \ldots, n\}\)

\(L_L \leftarrow \langle i \in L : i \leq p \rangle\)

\(L_R \leftarrow \langle j \in L : j > p \rangle\)

let \(l = |L_L|\)

if \(l = k\): Return \(p\) // pivot was \(k^{th}\) element

if \(l > k\): Return Select \((L_L, k)\)

else: Return Select \((L_R, k-l)\)

**What is the Expected RT of Randomized Select?**
Announcements

* Prelim 1 Grades returned.
* HW 3 Grades coming soon.
* HW 4 Out today
  2 problem set questions
  1 programming problem
Randomized Select \((L, k)\)

Choose pivot \(p \leftarrow L[Z]\) for \(Z \leftarrow r[1, \ldots, |L|]\)

\(L_\leq \leftarrow \{i \in L : i \leq p\}\)

\(L_\geq \leftarrow \{j \in L : j > p\}\)

let \(l = |L_\leq|\)

if \(l = k\) : Return \(p\) // pivot was \(k^{th}\) element

if \(l > k\) : Return Select \((L_\leq, k)\)

else : Return Select \((L_\geq, k-l)\)

What is the Expected RT of Randomized Select?
Random ≠ Arbitrary

e.g. choosing \textit{random} pivot is very different than choosing \textit{arbitrary} pivot

Adversary is “\textit{oblivious}” to algorithm’s randomness

\begin{itemize}
\item Adversary can anticipate arbitrary decisions
\item Adversary cannot anticipate \textit{random} decisions.
\end{itemize}
**Expected Running Time**

- \( O(1) \) to sample \( Z \) (by assumption)
- \( O(n) \) to partition \( I \) around pivot
- 1 recursive call

\[
T(n) \leq c \cdot n + T(\alpha \cdot n)
\]

for some \( \alpha < 1 \) depending on pivot.
Expected Running Time

- $O(1)$ to sample $Z$ (by assumption)
- $O(n)$ to partition $L$ around pivot
- 1 recursive call

\[ T(n) \leq c \cdot n + T(\alpha \cdot n) \]

for some $\alpha < 1$ depending on pivot.

$T(n)$ is Random, so we analyze $E[T(n)]$. 
Theorem. Randomized Select runs in Expected $O(n)$ time.

Proof Strategy.

- Give an expression $T(n)$ 

  $\Rightarrow$ $T(n)$ upper bounds running time on EVERY list of $n$ integers 

  $\Rightarrow$ $T(n)$ depends on randomness of alg.

- Give upper bound for $E[T(n)]$
Expected Running Time Analysis

1. Define a set of "good" pivots.
   \[ \Rightarrow \text{Reduce the problem size significantly} \]

2. Show "good" pivots occur regularly in expectation.

3. By linearity of expectation

   Expected running time bounded in terms of expected number of pivots.
Step 1

A "good" pivot is one where

$$|L| \geq \frac{n}{4} \quad \text{and} \quad |I| \geq \frac{n}{4}$$

i.e. a relatively balanced split.
Step 1

A "good" pivot is one where

\[ |L_<| \geq \frac{n}{4} \quad \text{and} \quad |L>| \geq \frac{n}{4} \]

i.e. a relatively balanced split.

Claim: If we select a good pivot, then the instance size drops by a factor \( \alpha = \frac{3}{4} \).
Step 1

A "good" pivot is one where

$$|L| \geq \frac{n}{4} \quad \text{and} \quad |L^c| \leq \frac{n}{4}$$

i.e. a relatively balanced split.

$$T(n) \leq c \cdot n + T\left(\frac{3n}{4}\right)$$
Step 1

A "good" pivot is one where

\[ |L_L| \geq \frac{n}{4} \quad \text{and} \quad |L_R| \geq \frac{n}{4} \]

i.e. a relatively balanced split.

\[
T(n) \leq c \cdot n + T\left(\frac{3n}{4}\right)
\]

\[
c \cdot \frac{3n}{4} + T\left(\frac{9n}{16}\right)
\]

\[
c \cdot \frac{9n}{16} + T\left(\frac{27n}{64}\right)
\]
Step 1

A "good" pivot is one where

\[ |L| \geq \frac{n}{4} \quad \text{and} \quad |R| \geq \frac{n}{4}. \]

i.e. a relatively balanced split.

\[
T(n) \leq c \cdot n + T\left(\frac{3n}{4}\right) \\
= c \cdot n + c \cdot n \cdot \left(\frac{3}{4}\right) + cn \left(\frac{3}{4}\right)^2 + \cdots \\
\leq cn \cdot \sum_{j=0}^{\infty} \left(\frac{3}{4}\right)^j
\]
Step 1

A "good" pivot is one where

\[ \left| L_\leq \right| \geq \frac{n}{4} \quad \text{and} \quad \left| L_\geq \right| \geq \frac{n}{4} \]

i.e. a relatively balanced split.

\[ T(n) \leq c \cdot n + T \left( \frac{3n}{4} \right) \]

\[ = c \cdot n + c \cdot n \cdot \left( \frac{3}{4} \right) + c \cdot n \cdot \left( \frac{3}{4} \right)^2 + \cdots \]

\[ \leq c \cdot n \cdot \sum_{j=0}^{\infty} \left( \frac{3}{4} \right)^j \]

Geometric Series

For \( x < 1 \):

\[ \sum_{j=0}^{\infty} x^j = \frac{1}{1-x} \]
Step 1

A "good" pivot is one where

\[ |L| \geq \frac{n}{4} \quad \text{and} \quad |L'| \geq \frac{n}{4}. \]

i.e. a relatively balanced split.

\[ T(n) \leq c \cdot n + T\left( \frac{3}{4} n \right) \]
\[ = c \cdot n + c \cdot n \cdot \left( \frac{3}{4} \right) + c \cdot n \cdot \left( \frac{3}{4} \right)^2 + \cdots \]
\[ \leq c \cdot n \cdot \sum_{j=0}^{\infty} \left( \frac{3}{4} \right)^j \]
\[ = 4 \cdot c \cdot n \]
\[ = O(n). \]
Expected Running Time Analysis

1. Define a set of “good” pivots.
   \[ \Rightarrow \text{Reduce the problem size significantly} \quad \alpha = \frac{3}{4} \]

2. Show “good” pivots occur regularly in expectation.

3. By linearity of expectation.
   
   Expected running time bounded in terms of expected number of pivots.
Step 2

Every time we select a pivot $p$, what is the probability that $p$ is "good"?

Consider the sorted list.
Which elements result in $L_<$ and $L_>$ each with $\frac{n}{4}$ elements?
Step 2

Every time we select a pivot \( p \), what is the probability that \( p \) is "good"?

Consider the sorted list.
Which elements result in \( L_{\leq} \) and \( L_{>\} \) each with \( n/4 \) elements?

\[
L \rightarrow Pr[ p \text{ is "good"}] = \frac{\# "good"}{\# \text{choices}} = \frac{3^{n/4} - n/4}{n} = \frac{1}{2}
\]
Step 2 contd.

What is the expected number of pivot selections until we select a good pivot?

\[ X = \text{number of pivot selections until good} \]

\[ X \text{ is a Geometric Random Variable} \]
Geometric Random Variable

$X$ represents the number of trials before success.

Each trial succeeds with probability $p$.

$$\Pr[X = k] = (1-p)^k \cdot p.$$
Geometric Random Variable

$X$ represents # of trials before success.

each trial succeeds with probability $p$.

What is the expectation of a geometric RV?

Geometric distribution is "memoryless"

$$E[X | X > i] = i + E[X]$$
**Geometric Random Variable**

$X$ represents the number of trials before success. Each trial succeeds with probability $p$.

What is the expectation of a geometric RV?

*Geometric distribution is "memoryless"*

$$E[X] = \Pr[X=1] + E[X | X > 1] \cdot \Pr[X \neq 1]$$
Geometric Random Variable

\(X\) represents \# of trials before success.

Each trial succeeds with probability \(p\).

What is the expectation of a geometric RV?

Geometric distribution is "memoryless".

\[
\mathbb{E}[X] = \Pr[X=1] + \mathbb{E}[X|X>1] \cdot \Pr[X \neq 1]
\]

\[
= p + (1 + \mathbb{E}[X])(1-p)
\]

\[
\Rightarrow p \cdot \mathbb{E}[X] = 1 \quad \Rightarrow \mathbb{E}[X] = \frac{1}{p}.
\]
What is the expected number of pivot selections until we select a good pivot?

\[ X = \text{number of pivot selections until good pivot} \]

\[ X \text{ is a Geometric Random Variable with } p = \frac{1}{2} \]

\[ E[X] = \frac{1}{\Pr[\text{"good pivot"]} = 2 \]
Expected Running Time Analysis

1. Define a set of “good” pivots.
   - Reduce the problem size significantly $\alpha = \frac{3}{4}$

2. Show “good” pivots occur regularly in expectation.

3. By linearity of expectation
   
   Expected running time bounded in terms of expected number of pivots.
At every "good" pivot, instance size drops by $3/4$ factor.

**Recurrence (Intuition)**

* Assume "bad" pivots make no progress
  \[ T(n) \leq c \cdot n + T(n) \]

* "good" pivots get
  \[ T(n) \leq c \cdot n + T\left(\frac{3n}{4}\right) \]
Let $X_j$ be geometric RV for $p=\frac{1}{2}$, representing number of pivots from $j^{th}$ until $(j+1)^{th}$ good pivot.

Total work upper bounded by

$$T(n) \leq X_0 \cdot cn + X_1 \cdot cn \cdot (\frac{3}{4}) + X_2 \cdot cn(\frac{3}{4})^2 + \ldots$$
Step 3

Let $X_j$ be geometric RV for $p = \frac{1}{2}$, representing number of pivots from $j^{th}$ until $(j+1)^{th}$ good pivot.

Total work upper bounded by

$$T(n) \leq X_0 \cdot cn + X_1 \cdot cn \cdot \left(\frac{3}{4}\right) + X_2 \cdot cn \left(\frac{3}{4}\right)^2 + \cdots$$

$$\leq \sum_{j=0}^{\infty} X_j \cdot cn \cdot \left(\frac{3}{4}\right)^j$$
Step 3

Expected work?
Apply linearity of expectation!

\[ E[T(n)] \leq E \left[ \sum_{j=0}^{\infty} X_j \cdot c_n \cdot (3/4)^j \right] \]
Step 3

Expected work?

Apply linearity of expectation!

\[
E[T(n)] \leq E\left[ \sum_{j=0}^{\infty} X_j \cdot cn \cdot \left(\frac{3}{4}\right)^j \right]
\]

\[
= \sum_{j=0}^{\infty} E[X_j] \cdot cn \cdot \left(\frac{3}{4}\right)^j
\]
Step 3

Expected work?

Apply linearity of expectation!

\[ E[T(n)] \leq E \left[ \sum_{j=0}^{\infty} X_j \cdot cn \cdot (3/4)^j \right] \]

\[ = \sum_{j=0}^{\infty} E[X_j] \cdot cn \cdot (3/4)^j \]

\[ = E[X] \cdot cn \cdot \sum_{j=0}^{\infty} (3/4)^j \]
Step 3

Expected work?

Apply linearity of expectation!

\[
E[T(n)] \leq E \left[ \sum_{j=0}^{\infty} X_j \cdot c_n \cdot (3/4)^j \right]
\]

\[
= \sum_{j=0}^{\infty} E[X_j] \cdot c_n \cdot (3/4)^j
\]

\[
= E[X] \cdot c_n \cdot \sum_{j=0}^{\infty} (3/4)^j
\]

\[
\downarrow
\]

\[
2 \rightarrow 4
\]
Step 3

Expected work?
Apply linearity of expectation!

\[
E[T(n)] \leq E \left[ \sum_{j=0}^{\infty} X_j \cdot cn \cdot (3/4)^j \right]
\]

\[
= \sum_{j=0}^{\infty} E[X_j] \cdot cn \cdot (3/4)^j
\]

\[
= E[X] \cdot cn \cdot \sum_{j=0}^{\infty} (3/4)^j
\]

\[
= 8cn
\]

= \(O(n)\)
Expected Running Time Analysis

1. Define a set of "good" pivots.
   - Reduce the problem size significantly \( \alpha = \frac{3}{4} \)

2. Show "good" pivots occur regularly in expectation

3. By linearity of expectation

   Expected running time bounded in terms of expected number of pivots.
Expected Runtime vs. High Probability?

* Good to have a guarantee of linear time.

What is the probability that Randomized Select runs for longer than $T$ steps?

Markov's Inequality:

$$\Pr[Z > t] \leq \frac{E[Z]}{t}$$
\[
\Pr \left[ R \text{Select runs longer than } 16 \cdot cn \text{ time} \right] \leq \frac{E[T(n)]}{16 \cdot cn} < \frac{8 \cdot cn}{16 \cdot cn} = \frac{1}{2}
\]

Pr [ RSelect runs longer than ] \( \Omega(n \log n) \) \( \leq \frac{1}{n^{100}} \)