1 March 2024 Analyzing Randomized Median

Plan

* Recall Median Finding Algorithm
* Analysis Tool: Expected Running Time
* Announcements
* Expected RT Analysis

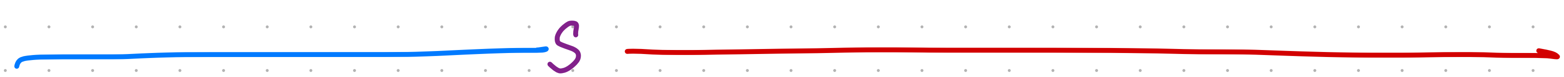
k^{th} element. (a.k.a. k^{th} ORDER STATISTIC)

Given: a list L of n distinct integer

Task: Return the $\underbrace{k^{th} \text{ smallest element}}_{s \in L}$

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

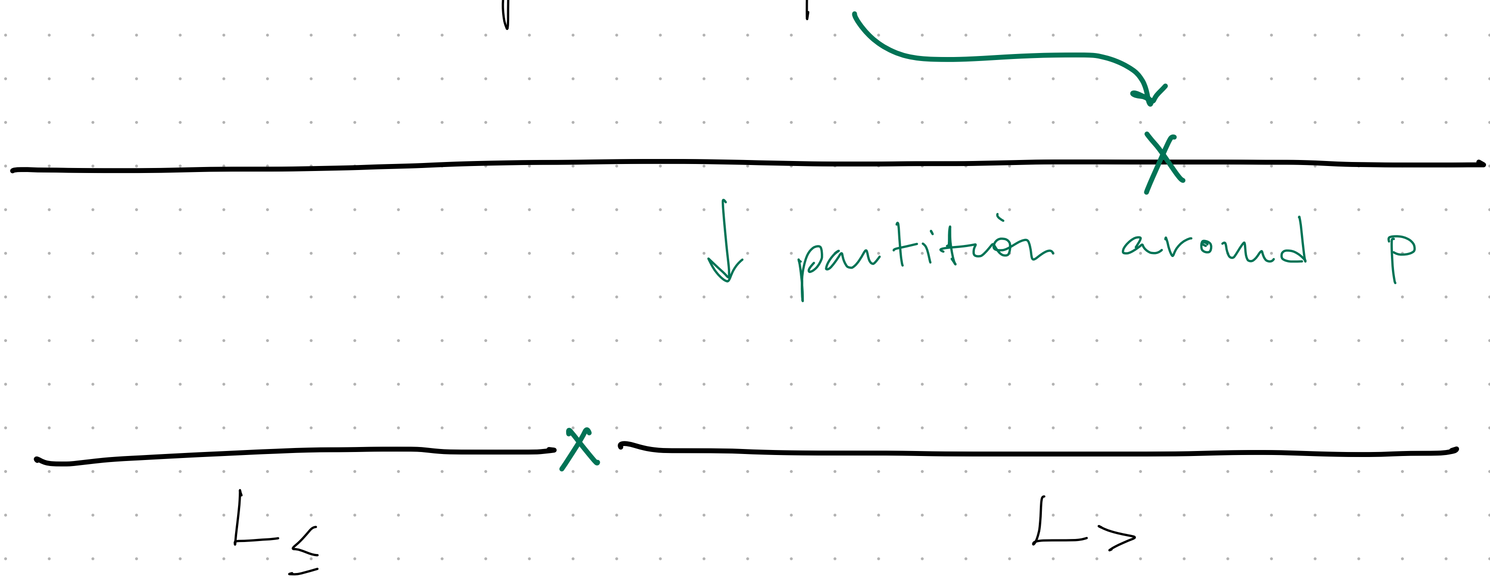
$$\left|\{t \in L : t > s\}\right| = n-k$$



Selection without Sorting

Divide

Choose a "pivot" p.



Conquer

To find k^{th} elem., consider $l = |L_{\le}|$ compared to k.

Select(L, k)

Choose pivot $p \in L$.

$L_{\leq} \leftarrow \langle i \in L : i \leq p \rangle$ // ensure p is final element of $L_{\leq}$

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

let $\ell = |L_{\leq}|$

if $\ell = k$: Return p // pivot was k^{th} elem

if $\ell > k$: Return Select$(L_{\leq}, k)$

else : Return Select$(L_{>}, k - \ell)$

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. $1/2$

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

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

Randomized Algorithms

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

  i.e. draw $Z \xleftarrow[r]{} \{1, \dots, n\}$

  uniformly at random

Randomized Algorithms

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

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

  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 $Z \leftarrow_r \{1, \dots, |L|\}$

$L_{\leq} \leftarrow \langle i \in L : i \leq p \rangle$

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

let $\ell = |L_{\leq}|$

if $\ell = k$: Return p // pivot was k^{th} elem

if $\ell > k$: Return Select $(L_{\leq}, k)$

else : Return Select $(L_{>}, k - \ell)$

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 $Z \leftarrow_r \{1, \ldots, |L|\}$

$L_\leq \leftarrow \langle i \in L : i \leq p \rangle$

$L_> \leftarrow \langle j \in L : j > p \rangle$

let $\ell = |L_\leq|$

if $\ell = k$: Return p // pivot was k^{th} elem

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

else : Return Select$(L_>, k - \ell)$

What is the Expected RT of Randomized Select?

Random $\neq$ Arbitrary

e.g. choosing random pivot is very different than choosing arbitrary pivot

Adversary is "oblivious" to algorithm's randomness

↳ Adversary can anticipate arbitrary decisions

↳ Adversary cannot anticipate random decisions.

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.

Expected Running Time

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

$$T(n) \le c \cdot n + T(\alpha \cdot n)$$

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

$T(n)$ is Random, so we analyze $\mathbb{E}[T(n)]$

Theorem. Randomized Select runs in Expected $O(n)$ time.

Proof Strategy.

- Give an expression $T(n)$
  - ↳ $T(n)$ upper bounds running time on EVERY list of n integers
  - ↳ $T(n)$ depends on randomness of alg.
- Give upper bound for $\mathbb{E}[T(n)]$

Expected Running Time Analysis

① Define a set of "good" pivots.

↳ Reduce the problem size significantly

② Show "good" pivots occur regularly in expectation

③ By linearity of expectation

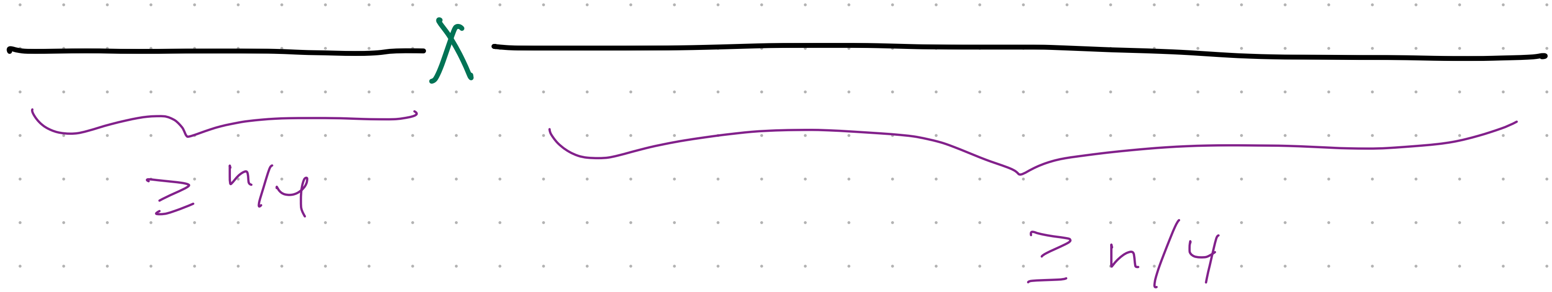
Expected running time bounded in terms of expected number of pivots.

Step 1

a "good" pivot is one where

$$|L_{\leq}| \geq \frac{n}{4} \quad \text{and} \quad |L_{>}| \geq \frac{n}{4}.$$

i.e. a relatively balanced split.



Step 1

a "good" pivot is one where

$$|L_{\leq}| \geq \frac{n}{4} \quad \text{and} \quad |L_{>}| \geq \frac{n}{4}.$$

i.e. a relatively balanced split.

Claim. If we select a good pivot, then the instance size drops by a factor $\alpha = 3/4$.

Step 1

a "good" pivot is one where

$$|L_{\leq}| \geq \frac{n}{4} \quad \text{and} \quad |L_{>}| \geq \frac{n}{4}.$$

i.e. a relatively balanced split.

$$T(n) \leq c \cdot n + T(3n/4)$$

Step ①

a "good" pivot is one where

$$|L_{\le}| \ge \frac{n}{4} \quad \text{and} \quad |L_{>}| \ge \frac{n}{4}.$$

i.e. a relatively balanced split.

$$T(n) \le c \cdot n + \underbrace{T(3n/4)}$$

$$c \cdot 3n/4 + \underbrace{T(9n/16)}$$

$$c \cdot 9n/16 + \underbrace{T(27n/64)}$$

$$\vdots$$

Step ①

a "good" pivot is one where

$$|L_{\leq}| \geq \frac{n}{4} \quad \text{and} \quad |L_{>}| \geq \frac{n}{4}.$$

i.e. a relatively balanced split.

$$T(n) \leq c \cdot n + T(3n/4)$$

$$= c \cdot n + c \cdot n \cdot (3/4) + cn\,(3/4)^2 + \cdots$$

$$\leq cn \cdot \sum_{j=0}^{\infty} (3/4)^j$$

Step 1

a "good" pivot is one where

$$|L_{\leq}| \geq \frac{n}{4} \quad \text{and} \quad |L_{>}| \geq \frac{n}{4}.$$

i.e. a relatively balanced split.

$$T(n) \leq c \cdot n + T(3n/4)$$

$$= c \cdot n + c \cdot n \cdot (3/4) + cn\,(3/4)^2 + \cdots$$

$$\leq cn \cdot \sum_{j=0}^{\infty} (3/4)^j$$

Geometric Series

For $\gamma < 1$:

$$\sum_{j=0}^{\infty} \gamma^j = \frac{1}{1-\gamma}$$

Step 1

a "good" pivot is one where

$$|L_{\le}| \ge \frac{n}{4} \quad \text{and} \quad |L_{>}| \ge \frac{n}{4}.$$

i.e. a relatively balanced split.

$$T(n) \le c \cdot n + T(3n/4)$$

$$= c \cdot n + c \cdot n \cdot (3/4) + cn\,(3/4)^2 + \cdots$$

$$\le cn \cdot \sum_{j=0}^{\infty} (3/4)^j$$

$$= 4cn$$

$$= O(n).$$

Expected Running Time Analysis

1. Define a set of "good" pivots.
   ↳ Reduce the problem size significantly $\alpha = 3/4$

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 $L_{\leq}$ and $L_{>}$
each w/ $n/4$ elems?

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 w/ $n/4$ elems?

$$\hookrightarrow \Pr[p \text{ is "good"}] = \frac{\# \text{"good"}}{\# \text{choices}}$$

$$= \frac{3n/4 - n/4}{n} = \boxed{\frac{1}{2}}$$

Step 2 cntd.

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

X = number of pivot selections until good.

X is a <u>Geometric Random Variable</u>

Geometric Random Variable

X represents # of independent trials before success.

each trial succeeds
with probability p.

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

Geometric Random Variable

X represents # of $\underbrace{\text{trials}}_{\text{each trial succeeds with probability } p.}$ before success.

each trial succeeds with probability p.

What is the expectation of a geometric RV?

Geometric distribution is "memoryless"

$$\mathbb{E}[X \mid X > i] = i + \mathbb{E}[X]$$

Geometric Random Variable

X represents # of $\underbrace{\text{trials}}_{}$ before success.

each trial succeeds
with probability p.

What is the expectation of a geometric RV?

Geometric distribution is "memoryless"

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

Geometric Random Variable

X represents # of trials before success.

$\underbrace{\text{trials}}_{}$ each trial succeeds with probability p.

What is the expectation of a geometric RV?

Geometric distribution is "memoryless"

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

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

$$\Rightarrow p \cdot \mathbb{E}[X] = 1 \Rightarrow \mathbb{E}[X] = \frac{1}{p}.$$

Step 2 cntd.

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

X = number of pivot selections until good.

X is a <u>Geometric Random Variable</u> wp $1/2$

$$E[X] = \frac{1}{Pr[\text{"good" pivot}]} = 2.$$

Expected Running Time Analysis

1. Define a set of "good" pivots.
   ↳ Reduce the problem size significantly $\alpha = 3/4$

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 3

At every "good" pivot, instance size drops by 3/4 factor.

Recurrence (Intuition)

* Assume "bad" pivots make NO progress

$$T(n) \le c \cdot n + T(n)$$

* "good" pivots get

$$T(n) \le c \cdot n + T(3n/4)$$

Step 3

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

Total work upper bounded by

$$T(n) \leq X_0 \cdot cn + X_1 \cdot cn \cdot (3/4) + X_2 \cdot cn (3/4)^2 + \cdots$$

Step 3

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

Total work upper bounded by

$$T(n) \leq X_0 \cdot cn + X_1 \cdot cn \cdot (3/4) + X_2 \cdot cn (3/4)^2 + \cdots$$

$$\leq \sum_{j=0}^{\infty} X_j \cdot cn \cdot (3/4)^j$$

Step 3

Expected work?

Apply linearity of expectation!

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

Step 3

Expected work?

Apply linearity of expectation!

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

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

Step 3

Expected work?

Apply linearity of expectation!

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

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

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

Step 3

Expected work?

Apply linearity of expectation!

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

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

$$= \underbrace{\mathbb{E}[X]}_{2} \cdot cn \cdot \underbrace{\sum_{j=0}^{\infty} (3/4)^j}_{4}$$

Step 3

Expected work?

Apply linearity of expectation!

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

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

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

$$= 8cn$$

$$= O(n)$$

Expected Running Time Analysis

(1) Define a set of "good" pivots.

↳ Reduce the problem size significantly $\alpha = 3/4$

(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.

$$\Pr[Z > t] \leq \frac{\mathbb{E}[Z]}{t}$$

$$\Pr\left[\begin{array}{c}\text{RSelect runs longer than}\\ 16c\cdot n \text{ time}\end{array}\right] \le \frac{\mathbb{E}[T(n)]}{16\cdot cn} \le \frac{8\cdot cn}{16\cdot cn} = 1/2$$

careful analysis

$\Downarrow$

$$\Pr\left[\begin{array}{c}\text{RSelect runs longer than}\\ \Omega(n\log n)\end{array}\right] \le 1/n^{100}$$