CMPSCI 711: "Really Advanced Algorithms"

Lecture 4 – Lazy Select and Chernoff Bounds

Andrew McGregor

Last Compiled: February 5, 2009

Outline

Lazy Select

Chernoff Bounds

Set Balancing

Readings

Puzzle

Lazy Select

We have a set S of $n = 2k$ distinct numbers and want to find the k-th smallest element.

Lazy Select

We have a set S of $n = 2k$ distinct numbers and want to find the k-th smallest element.

Algorithm

1. *Let R be a set of $n^{3/4}$ elements chosen uniformly at random with replacement from S.*

Lazy Select

We have a set S of $n = 2k$ distinct numbers and want to find the k-th smallest element.

Algorithm

1. *Let R be a set of $n^{3/4}$ elements chosen uniformly at random with replacement from S.*
2. *Sort R and find a and b such that*

$$\text{rank}_R(a) = kn^{-1/4} - \sqrt{n} \text{ and } \text{rank}_R(b) = kn^{-1/4} + \sqrt{n}$$

where $\text{rank}_X(x) = t$ *if x is the t-th smallest element in X.*

Lazy Select

We have a set S of $n = 2k$ distinct numbers and want to find the k-th smallest element.

Algorithm

1. *Let R be a set of $n^{3/4}$ elements chosen uniformly at random with replacement from S.*
2. *Sort R and find a and b such that*

$$\operatorname{rank}_R(a) = kn^{-1/4} - \sqrt{n} \text{ and } \operatorname{rank}_R(b) = kn^{-1/4} + \sqrt{n}$$

where $\operatorname{rank}_X(x) = t$ *if x is the t-th smallest element in X.*

3. *Compute* $\operatorname{rank}_S(a)$ *and* $\operatorname{rank}_S(b)$*: Output FAIL if*

$$k < \operatorname{rank}_S(a) \quad \text{or} \quad k > \operatorname{rank}_S(b)$$

Lazy Select

We have a set S of $n = 2k$ distinct numbers and want to find the k-th smallest element.

Algorithm

1. *Let R be a set of $n^{3/4}$ elements chosen uniformly at random with replacement from S.*
2. *Sort R and find a and b such that*

$$\mathrm{rank}_R(a) = kn^{-1/4} - \sqrt{n} \text{ and } \mathrm{rank}_R(b) = kn^{-1/4} + \sqrt{n}$$

   where $\mathrm{rank}_X(x) = t$ if x is the t-th smallest element in X.

3. *Compute $\mathrm{rank}_S(a)$ and $\mathrm{rank}_S(b)$: Output FAIL if*

$$k < \mathrm{rank}_S(a) \quad or \quad k > \mathrm{rank}_S(b)$$

4. *Let $P = \{i \in S : a \leq y \leq b\}$: Output FAIL if $|P| \geq 4n^{3/4}$*

Lazy Select

We have a set S of $n = 2k$ distinct numbers and want to find the k-th smallest element.

Algorithm

1. *Let R be a set of $n^{3/4}$ elements chosen uniformly at random with replacement from S.*
2. *Sort R and find a and b such that*

$$\mathrm{rank}_R(a) = kn^{-1/4} - \sqrt{n} \text{ and } \mathrm{rank}_R(b) = kn^{-1/4} + \sqrt{n}$$

   where $\mathrm{rank}_X(x) = t$ if x is the t-th smallest element in X.

3. *Compute $\mathrm{rank}_S(a)$ and $\mathrm{rank}_S(b)$: Output FAIL if*

$$k < \mathrm{rank}_S(a) \;\; or \;\; k > \mathrm{rank}_S(b)$$

4. *Let $P = \{i \in S : a \leq y \leq b\}$: Output FAIL if $|P| \geq 4n^{3/4}$*
5. *Return $(k - \mathrm{rank}_S(a) + 1)$-th smallest element from P*

Lazy Select: Running Time

Theorem

Running time of Lazy Select is $O(n)$.

Lazy Select: Running Time

Theorem

Running time of Lazy Select is $O(n)$.

Proof.

- $O(n^{3/4})$ steps to define R.

□

Lazy Select: Running Time

Theorem

Running time of Lazy Select is $O(n)$.

Proof.

- $O(n^{3/4})$ steps to define R.
- $O(n^{3/4} \log n)$ steps to sort R and find a and b.

□

Lazy Select: Running Time

Theorem

Running time of Lazy Select is $O(n)$.

Proof.

- $O(n^{3/4})$ steps to define R.
- $O(n^{3/4} \log n)$ steps to sort R and find a and b.
- $O(n)$ steps to compute $\text{rank}_S(a)$ and $\text{rank}_S(b)$ in S.

$\square$

Lazy Select: Running Time

Theorem

Running time of Lazy Select is $O(n)$.

Proof.

- $O(n^{3/4})$ steps to define R.
- $O(n^{3/4} \log n)$ steps to sort R and find a and b.
- $O(n)$ steps to compute $\text{rank}_S(a)$ and $\text{rank}_S(b)$ in S.
- $O(n^{3/4} \log n)$ steps to sort P and select element.

$\square$

Lazy Select: Probability of Being Correct (1/3)

Theorem

With probability $1 - O(n^{-1/4})$*, algorithm finds the median.*

Lazy Select: Probability of Being Correct (1/3)

Theorem

With probability $1 - O(n^{-1/4})$*, algorithm finds the median.*

Proof.

- If we don't output FAIL, then we get the answer correct.

□

Lazy Select: Probability of Being Correct (1/3)

Theorem

With probability $1 - O(n^{-1/4})$*, algorithm finds the median.*

Proof.

- If we don't output FAIL, then we get the answer correct.
- Only three ways in which we fail and we'll show
  1. $\mathbb{P}[k < \text{rank}_S(a)] \leq O(n^{-1/4})$
  2. $\mathbb{P}[k > \text{rank}_S(b)] \leq O(n^{-1/4})$
  3. $\mathbb{P}\left[|P| \geq 4n^{3/4}\right] \leq O(n^{-1/4})$

□

Lazy Select: Probability of Being Correct (2/3)

Claim

$\mathbb{P}\left[k < \mathrm{rank}_S(a)\right] \leq O(n^{-1/4})$

Lazy Select: Probability of Being Correct (2/3)

Claim

$\mathbb{P}[k < \mathrm{rank}_S(a)] \leq O(n^{-1/4})$

Proof.

- Let u be the k-th smallest element in S

Lazy Select: Probability of Being Correct (2/3)

Claim

$\mathbb{P}[k < \text{rank}_S(a)] \leq O(n^{-1/4})$

Proof.

- Let u be the k-th smallest element in S
- Consider choosing R: Let $X_i = 1$ if i-th sample is $\leq u$ and $X_i = 0$ otherwise. $\mathbb{P}[X_i = 1] = k/n$ and $\mathbb{P}[X_i = 0] = 1 - k/n$

Lazy Select: Probability of Being Correct (2/3)

Claim

$\mathbb{P}[k < \text{rank}_S(a)] \leq O(n^{-1/4})$

Proof.

- Let u be the k-th smallest element in S
- Consider choosing R: Let $X_i = 1$ if i-th sample is $\leq u$ and $X_i = 0$ otherwise. $\mathbb{P}[X_i = 1] = k/n$ and $\mathbb{P}[X_i = 0] = 1 - k/n$
- $X = \sum_{i \in [n^{3/4}]} X_i =$ number of elements in R that are at most u.

Lazy Select: Probability of Being Correct (2/3)

Claim

$\mathbb{P}[k < \mathrm{rank}_S(a)] \leq O(n^{-1/4})$

Proof.

- Let u be the k-th smallest element in S
- Consider choosing R: Let $X_i = 1$ if i-th sample is $\leq u$ and $X_i = 0$ otherwise. $\mathbb{P}[X_i = 1] = k/n$ and $\mathbb{P}[X_i = 0] = 1 - k/n$
- $X = \sum_{i \in [n^{3/4}]} X_i =$ number of elements in R that are at most u.
- $k < \mathrm{rank}_S(a)$ implies $X < kn^{-1/4} - \sqrt{n}$

Lazy Select: Probability of Being Correct (2/3)

Claim

$\mathbb{P}[k < \text{rank}_S(a)] \leq O(n^{-1/4})$

Proof.

- Let u be the k-th smallest element in S
- Consider choosing R: Let $X_i = 1$ if i-th sample is $\leq u$ and $X_i = 0$ otherwise. $\mathbb{P}[X_i = 1] = k/n$ and $\mathbb{P}[X_i = 0] = 1 - k/n$
- $X = \sum_{i \in [n^{3/4}]} X_i =$ number of elements in R that are at most u.
- $k < \text{rank}_S(a)$ implies $X < kn^{-1/4} - \sqrt{n}$
- X has binomial distribution:

$$\mathbb{E}[X] = kn^{-1/4} \text{ and } \mathbb{V}[X] = n^{3/4}(k/n)(1 - k/n) = n^{3/4}/4$$

Lazy Select: Probability of Being Correct (2/3)

Claim

$\mathbb{P}[k < \text{rank}_S(a)] \leq O(n^{-1/4})$

Proof.

- Let u be the k-th smallest element in S
- Consider choosing R: Let $X_i = 1$ if i-th sample is $\leq u$ and $X_i = 0$ otherwise. $\mathbb{P}[X_i = 1] = k/n$ and $\mathbb{P}[X_i = 0] = 1 - k/n$
- $X = \sum_{i \in [n^{3/4}]} X_i =$ number of elements in R that are at most u.
- $k < \text{rank}_S(a)$ implies $X < kn^{-1/4} - \sqrt{n}$
- X has binomial distribution:

$$\mathbb{E}[X] = kn^{-1/4} \text{ and } \mathbb{V}[X] = n^{3/4}(k/n)(1 - k/n) = n^{3/4}/4$$

- Apply Chebyshev bound: $\mathbb{P}\left[X < kn^{-1/4} - \sqrt{n}\right]$ is at most

$$\mathbb{P}\left[|X - \mathbb{E}[X]| < \sqrt{n}\right] \leq \mathbb{P}\left[|X - \mathbb{E}[X]| < 2n^{1/8}\sigma_X\right] = O(n^{-1/4})$$

Lazy Select: Probability of Being Correct (3/3)

Claim

$\mathbb{P}\left[|P| \geq 4n^{3/4}\right] \leq O(n^{-1/4})$

Lazy Select: Probability of Being Correct (3/3)

Claim

$\mathbb{P}\left[|P| \geq 4n^{3/4}\right] \leq O(n^{-1/4})$

Proof.

- If $|P| \geq 4n^{3/4}$ then either

$$\text{rank}_S(a) \leq k - 2n^{3/4} \quad \text{or} \quad \text{rank}_S(b) \geq k + 2n^{3/4} - 1$$

□

Lazy Select: Probability of Being Correct (3/3)

Claim

$\mathbb{P}\left[|P| \geq 4n^{3/4}\right] \leq O(n^{-1/4})$

Proof.

- If $|P| \geq 4n^{3/4}$ then either

$$\text{rank}_S(a) \leq k - 2n^{3/4} \quad \text{or} \quad \text{rank}_S(b) \geq k + 2n^{3/4} - 1$$

- To bound

$$\mathbb{P}\left[\text{rank}_S(a) \leq k - 2n^{3/4}\right] \quad \text{and} \quad \mathbb{P}\left[\text{rank}_S(b) \geq k + 2n^{3/4} - 1\right]$$

  define X_i and use Chebyshev along the same lines as the previous claim.

□

Lazy Select: Probability of Being Correct (3/3)

Claim

$\mathbb{P}\left[|P| \geq 4n^{3/4}\right] \leq O(n^{-1/4})$

Proof.

- If $|P| \geq 4n^{3/4}$ then either

$$\text{rank}_S(a) \leq k - 2n^{3/4} \quad \text{or} \quad \text{rank}_S(b) \geq k + 2n^{3/4} - 1$$

- To bound

$$\mathbb{P}\left[\text{rank}_S(a) \leq k - 2n^{3/4}\right] \quad \text{and} \quad \mathbb{P}\left[\text{rank}_S(b) \geq k + 2n^{3/4} - 1\right]$$

  define X_i and use Chebyshev along the same lines as the previous claim.

- Apply union bound.

□

Outline

Chernoff Bound: Upper Tail (1/3)

Theorem

Let $X_1, \ldots, X_n$ be independent boolean random variables such that $\mathbb{P}[X_i = 1] = p_i$. Then, for $X = \sum_i X_i$, $\mu = \mathbb{E}[X]$, and $\delta > 0$,

$$\mathbb{P}[X > (1+\delta)\mu] < \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$$

Chernoff Bound: Upper Tail (2/3)

Proof.

- For any $t > 0$: $\mathbb{P}\left[X > (1+\delta)\mu\right] = \mathbb{P}\left[e^{tX} > e^{t(1+\delta)\mu}\right]$

Chernoff Bound: Upper Tail (2/3)

Proof.

- For any $t > 0$: $\mathbb{P}[X > (1+\delta)\mu] = \mathbb{P}\left[e^{tX} > e^{t(1+\delta)\mu}\right]$
- Apply Markov inequality:

$$\mathbb{P}\left[e^{tX} > e^{t(1+\delta)\mu}\right] \geq \mathbb{E}\left[e^{tX}\right]/e^{t(1+\delta)\mu}$$

Chernoff Bound: Upper Tail (2/3)

Proof.

- For any $t > 0$: $\mathbb{P}[X > (1+\delta)\mu] = \mathbb{P}\left[e^{tX} > e^{t(1+\delta)\mu}\right]$
- Apply Markov inequality:

$$\mathbb{P}\left[e^{tX} > e^{t(1+\delta)\mu}\right] \geq \mathbb{E}\left[e^{tX}\right]/e^{t(1+\delta)\mu}$$

- By independence:

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t\sum_i X_i}\right] = \mathbb{E}\left[\prod_i e^{tX_i}\right] = \prod_i \mathbb{E}\left[e^{tX_i}\right]$$

Chernoff Bound: Upper Tail (2/3)

Proof.

- For any $t > 0$: $\mathbb{P}[X > (1+\delta)\mu] = \mathbb{P}\left[e^{tX} > e^{t(1+\delta)\mu}\right]$
- Apply Markov inequality:

$$\mathbb{P}\left[e^{tX} > e^{t(1+\delta)\mu}\right] \geq \mathbb{E}\left[e^{tX}\right] / e^{t(1+\delta)\mu}$$

- By independence:

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t\sum_i X_i}\right] = \mathbb{E}\left[\prod_i e^{tX_i}\right] = \prod_i \mathbb{E}\left[e^{tX_i}\right]$$

- We will prove $\prod_i \mathbb{E}\left[e^{tX_i}\right] \leq e^{(e^t-1)\mu}$ in a sec.

Chernoff Bound: Upper Tail (2/3)

Proof.

- For any $t > 0$: $\mathbb{P}[X > (1+\delta)\mu] = \mathbb{P}\left[e^{tX} > e^{t(1+\delta)\mu}\right]$
- Apply Markov inequality:

$$\mathbb{P}\left[e^{tX} > e^{t(1+\delta)\mu}\right] \geq \mathbb{E}\left[e^{tX}\right]/e^{t(1+\delta)\mu}$$

- By independence:

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t\sum_i X_i}\right] = \mathbb{E}\left[\prod_i e^{tX_i}\right] = \prod_i \mathbb{E}\left[e^{tX_i}\right]$$

- We will prove $\prod_i \mathbb{E}\left[e^{tX_i}\right] \leq e^{(e^t-1)\mu}$ in a sec.
- For $t = \ln(1+\delta)$:

$$\mathbb{E}\left[e^{tX}\right]/e^{t(1+\delta)\mu} \leq e^{(e^t-1)\mu}/e^{t(1+\delta)\mu} = \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$$

Chernoff Bound: Upper Tail (3/3)

Lemma

$\prod_i \mathbb{E}\left[e^{tX_i}\right] \leq e^{(e^t-1)\mu}$

Proof.

- Using $1 + x \leq e^x$:

$$\mathbb{E}\left[e^{tX_i}\right] = p_i e^t + (1 - p_i) = 1 + p_i(e^t - 1) \leq \exp(p_i(e^t - 1))$$

□

Chernoff Bound: Upper Tail (3/3)

Lemma

$\prod_i \mathbb{E}\left[e^{tX_i}\right] \leq e^{(e^t-1)\mu}$

Proof.

- Using $1 + x \leq e^x$:

$$\mathbb{E}\left[e^{tX_i}\right] = p_i e^t + (1 - p_i) = 1 + p_i(e^t - 1) \leq \exp(p_i(e^t - 1))$$

- Using $\mu = \mathbb{E}\left[\sum_i X_i\right] = \sum_i p_i$:

$$\prod_i \exp(p_i(e^t - 1)) = \exp(\sum_i p_i(e^t - 1)) = \exp((e^t - 1)\mu)$$

□

Chernoff Bound: Upper Tail Simplification

Theorem

Let $X_1, \ldots, X_n$ be independent boolean random variables such that $\mathbb{P}[X_i = 1] = p_i$. Let $X = \sum_i X_i$ and $\mu = \mathbb{E}[X]$.

Chernoff Bound: Upper Tail Simplification

Theorem

Let $X_1, \ldots, X_n$ be independent boolean random variables such that $\mathbb{P}[X_i = 1] = p_i$. Let $X = \sum_i X_i$ and $\mu = \mathbb{E}[X]$.

- *If $\delta > 2e - 1$,*

$$\mathbb{P}[X > (1+\delta)\mu] < 2^{-(1+\delta)\mu}$$

Chernoff Bound: Upper Tail Simplification

Theorem

Let $X_1, \ldots, X_n$ be independent boolean random variables such that $\mathbb{P}[X_i = 1] = p_i$. Let $X = \sum_i X_i$ and $\mu = \mathbb{E}[X]$.

- *If* $\delta > 2e - 1$,

$$\mathbb{P}[X > (1+\delta)\mu] < 2^{-(1+\delta)\mu}$$

- *If* $0 < \delta \leq 2e - 1$,

$$\mathbb{P}[X > (1+\delta)\mu] < e^{-\mu\delta^2/4}$$

Chernoff Bound: Lower Tail (1/2)

Theorem

Let $X_1, \ldots, X_n$ be independent boolean random variables such that $\mathbb{P}[X_i = 1] = p_i$. Then, for $X = \sum_i X_i$, $\mu = \mathbb{E}[X]$, and $1 > \delta > 0$,

$$\mathbb{P}[X < (1-\delta)\mu] < \exp(-\mu\delta^2/2)$$

Chernoff Bound: Lower Tail (2/2)

Proof.

- For any $t > 0$: $\mathbb{P}[X < (1-\delta)\mu] = \mathbb{P}\left[e^{-tX} > e^{-t(1-\delta)\mu}\right]$

$\square$

Chernoff Bound: Lower Tail (2/2)

Proof.

- For any $t > 0$: $\mathbb{P}[X < (1-\delta)\mu] = \mathbb{P}\left[e^{-tX} > e^{-t(1-\delta)\mu}\right]$
- Apply Markov inequality:

$$\mathbb{P}\left[e^{-tX} > e^{-t(1-\delta)\mu}\right] \geq \mathbb{E}\left[e^{-tX}\right] / e^{-t(1-\delta)\mu}$$

$\square$

Chernoff Bound: Lower Tail (2/2)

Proof.

- For any $t > 0$: $\mathbb{P}[X < (1-\delta)\mu] = \mathbb{P}\left[e^{-tX} > e^{-t(1-\delta)\mu}\right]$
- Apply Markov inequality:

$$\mathbb{P}\left[e^{-tX} > e^{-t(1-\delta)\mu}\right] \geq \mathbb{E}\left[e^{-tX}\right] / e^{-t(1-\delta)\mu}$$

- Similarly to before: $\mathbb{E}\left[e^{-tX}\right] = \prod_i \mathbb{E}\left[e^{-tX_i}\right] \leq e^{(e^{-t}-1)\mu}$

□

Chernoff Bound: Lower Tail (2/2)

Proof.

- For any $t > 0$: $\mathbb{P}[X < (1-\delta)\mu] = \mathbb{P}\left[e^{-tX} > e^{-t(1-\delta)\mu}\right]$
- Apply Markov inequality:

$$\mathbb{P}\left[e^{-tX} > e^{-t(1-\delta)\mu}\right] \geq \mathbb{E}\left[e^{-tX}\right] / e^{-t(1-\delta)\mu}$$

- Similarly to before: $\mathbb{E}\left[e^{-tX}\right] = \prod_i \mathbb{E}\left[e^{-tX_i}\right] \leq e^{(e^{-t}-1)\mu}$
- For $t = -\ln(1-\delta)$:

$$\mathbb{E}\left[e^{-tX}\right] / e^{-t(1-\delta)\mu} \leq e^{(e^{-t}-1)\mu} / e^{-t(1-\delta)\mu} = \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu}$$

□

Chernoff Bound: Lower Tail (2/2)

Proof.

- For any $t > 0$: $\mathbb{P}[X < (1-\delta)\mu] = \mathbb{P}\left[e^{-tX} > e^{-t(1-\delta)\mu}\right]$
- Apply Markov inequality:

$$\mathbb{P}\left[e^{-tX} > e^{-t(1-\delta)\mu}\right] \geq \mathbb{E}\left[e^{-tX}\right] / e^{-t(1-\delta)\mu}$$

- Similarly to before: $\mathbb{E}\left[e^{-tX}\right] = \prod_i \mathbb{E}\left[e^{-tX_i}\right] \leq e^{(e^{-t}-1)\mu}$
- For $t = -\ln(1-\delta)$:

$$\mathbb{E}\left[e^{-tX}\right] / e^{-t(1-\delta)\mu} \leq e^{(e^{-t}-1)\mu} / e^{-t(1-\delta)\mu} = \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu}$$

- Simplify using $(1-\delta)^{1-\delta} > \exp(-\delta + \delta^2/2)$ since $\delta \in (0,1)$.

□

Outline

Set Balancing

Let $A_1, \ldots, A_n$ be subsets of $[n]$ such that $|A_i| = n/2$. We want to partition $[n]$ into B and C such that

$$\max_i \left| |A_i \cap B| - |A_i \cap C| \right|$$

is minimized.

Hint: Use $\mathbb{P}\left[|X - \mathbb{E}\left[X\right]| < \delta\mu\right] \leq 2\exp(-\mathbb{E}\left[X\right]\delta^2/4)$.

Outline

Lazy Select

Chernoff Bounds

Set Balancing

Readings

Puzzle

Readings

For next time, please make sure you've read:

- Chapter 3: Moments and Deviations (20 pages)

Outline

(Too) Easy Puzzle

- There are 3 coins in a bag: the first coin has two heads, the second coin has two tails, and the third coin has one head and one tail.
- You draw a coin at random without looking and toss it in the air. It lands heads up.
- What's the probability that the other side of the coin is heads?