CMPSCI 611: Advanced Algorithms
Lecture 16: Lazy Select

Andrew McGregor
Lazy Select

Next Time: Balls and Bins and Birthdays and Coupons
Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.
Lazy Select: Warm Up

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.

Warm-Up:
Lazy Select: Warm Up

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.

Warm-Up:

1. Assume for a moment, we are given values $a, b \in S$ such that $a < v_2 < b$ and there aren’t too many values in $S$ between $a$ and $b$. 
Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.

**Warm-Up:**

1. Assume for a moment, we are given values $a, b \in S$ such that $a < v_2 < b$ and there aren’t too many values in $S$ between $a$ and $b$.
2. An approach could be:
Lazy Select: Warm Up

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.

Warm-Up:

1. Assume for a moment, we are given values $a, b \in S$ such that $a < v_2 < b$ and there aren’t too many values in $S$ between $a$ and $b$.
2. An approach could be:
   2.1 Take $O(n)$ time to compute the number of elements in $S$ that are less than equal to $a$. Call this number $t$. 

Lazy Select: Warm Up

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.

Warm-Up:

1. Assume for a moment, we are given values $a, b \in S$ such that $a < v_2 < b$ and there aren’t too many values in $S$ between $a$ and $b$.

2. An approach could be:
   
   2.1 Take $O(n)$ time to compute the number of elements in $S$ that are less than equal to $a$. Call this number $t$.
   
   2.2 Let $S' = \{y \in S : a < y < b\}$. Return the $(k - t)$th smallest element in $S'$. This is easier than the original problem since $|S'| \ll |S|$.
Lazy Select: Warm Up

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.

Warm-Up:

1. Assume for a moment, we are given values $a, b \in S$ such that $a < v_2 < b$ and there aren’t too many values in $S$ between $a$ and $b$.
2. An approach could be:
   2.1 Take $O(n)$ time to compute the number of elements in $S$ that are less than equal to $a$. Call this number $t$.
   2.2 Let $S' = \{ y \in S : a < y < b \}$. Return the $(k - t)$th smallest element in $S'$. This is easier than the original problem since $|S'| \ll |S|$.
3. Question: How can we easily compute $a$ and $b$?
Lazy Select

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.
Lazy Select

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.

Algorithm

1. **Finding $a$ and $b$:** Sample each element in $S$ with probability $p = 1/n^{1/4}$. Call the sampled set $R$, sort $R$, and let

   \[
   a = \left(\frac{n^{3/4}}{2} - 5\sqrt{n}\right)\text{th smallest element in } R.
   \]

   \[
   b = \left(\frac{n^{3/4}}{2} + 5\sqrt{n}\right)\text{th smallest element in } R.
   \]
Lazy Select

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.

Algorithm

1. **Finding $a$ and $b$:** Sample each element in $S$ with probability $p = 1/n^{1/4}$. Call the sampled set $R$, sort $R$, and let

   $$a = \left(\frac{n^{3/4}}{2} - 5\sqrt{n}\right)\text{th smallest element in } R.$$  

   $$b = \left(\frac{n^{3/4}}{2} + 5\sqrt{n}\right)\text{th smallest element in } R.$$
Lazy Select

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.

Algorithm

1. **Finding $a$ and $b$:** Sample each element in $S$ with probability $p = 1/n^{1/4}$. Call the sampled set $R$, sort $R$, and let

   $a = (n^{3/4}/2 - 5\sqrt{n})$th smallest element in $R$.

   $b = (n^{3/4}/2 + 5\sqrt{n})$th smallest element in $R$. 


Lazy Select

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.

Algorithm

1. *Finding a and b*: Sample each element in $S$ with probability $p = 1/n^{1/4}$. Call the sampled set $R$, sort $R$, and let

   $a = (n^{3/4}/2 - 5\sqrt{n})$th smallest element in $R$.

   $b = (n^{3/4}/2 + 5\sqrt{n})$th smallest element in $R$.

2. *Construct $S'$*: $\{y \in S : a < y < b\}$ and let $t$ be the number of values less or equal to $a$ amongst $S$. 
Lazy Select

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value. For the sake of analysis, let $v_2$ be the value that we need to return.

Algorithm

1. **Finding $a$ and $b$:** Sample each element in $S$ with probability $p = 1/n^{1/4}$. Call the sampled set $R$, sort $R$, and let
   
   $$a = \left(\frac{n^{3/4}}{2} - 5\sqrt{n}\right)\text{th smallest element in } R.$$  
   
   $$b = \left(\frac{n^{3/4}}{2} + 5\sqrt{n}\right)\text{th smallest element in } R.$$  

2. **Construct $S'$:** $S' = \{y \in S : a < y < b\}$ and let $t$ be the number of values less or equal to $a$ amongst $S$.
3. **Sort $S'$ and return $(k - t)$th smallest value in $S'$.**
Theorem

Running time of Lazy Select is $O(n)$ if $|R| \leq 2n^{3/4}$ and $|S'| \leq 20n^{3/4}$
Lazy Select: Running Time

Theorem
Running time of Lazy Select is $O(n)$ if $|R| \leq 2n^{3/4}$ and $|S'| \leq 20n^{3/4}$

Proof.
- $O(n)$ steps to define $R$.
Lazy Select: Running Time

Theorem

Running time of Lazy Select is $O(n)$ if $|R| \leq 2n^{3/4}$ and $|S'| \leq 20n^{3/4}$

Proof.

- $O(n)$ steps to define $R$.
- $O(|R| \log |R|)$ steps to sort $R$ and find $a$ and $b$. 
Lazy Select: Running Time

**Theorem**

*Running time of Lazy Select is $O(n)$ if $|R| \leq 2n^{3/4}$ and $|S'| \leq 20n^{3/4}$*

**Proof.**

- $O(n)$ steps to define $R$.
- $O(|R| \log |R|)$ steps to sort $R$ and find $a$ and $b$.
- $O(n)$ steps to compute $S'$ and find $t$. 

□
Lazy Select: Running Time

**Theorem**

Running time of Lazy Select is $O(n)$ if $|R| \leq 2n^{3/4}$ and $|S'| \leq 20n^{3/4}$

**Proof.**

- $O(n)$ steps to define $R$.
- $O(|R| \log |R|)$ steps to sort $R$ and find $a$ and $b$.
- $O(n)$ steps to compute $S'$ and find $t$.
- $O(|S'| \log |S'|)$ steps to sort $|S'|$ and select element.
Correctness Analysis

Let $v_1, v_2, v_3, v_4$ be the values in $S$ of rank

$$r_1 = \frac{n}{2} - 10n^{3/4}, \quad r_2 = \frac{n}{2}, \quad r_3 = \frac{n}{2} + 10n^{3/4}, \quad r_4 = n$$

where the rank of a value is the number of values less or equal to it.
Correctness Analysis

Let $v_1, v_2, v_3, v_4$ be the values in $S$ of rank

$$
\begin{align*}
    r_1 &= \frac{n}{2} - 10n^{3/4} \\
    r_2 &= \frac{n}{2} \\
    r_3 &= \frac{n}{2} + 10n^{3/4} \\
    r_4 &= n
\end{align*}
$$

where the rank of a value is the number of values less or equal to it.

Define $X_i =$ number of values sampled in $R$ less or equal to $v_i$ and note:

$$
X_4 < 2n^{3/4} \Rightarrow |R| < 2n^{3/4}
$$
Correctness Analysis

Let \(v_1, v_2, v_3, v_4\) be the values in \(S\) of rank

\[
\begin{align*}
    r_1 &= n/2 - 10n^{3/4}, \\
    r_2 &= n/2, \\
    r_3 &= n/2 + 10n^{3/4}, \\
    r_4 &= n
\end{align*}
\]

where the rank of a value is the number of values less or equal to it.

Define \(X_i = \) number of values sampled in \(R\) less or equal to \(v_i\) and note:

\[
X_4 < 2n^{3/4} \Rightarrow |R| < 2n^{3/4}
\]

\[
X_2 > n^{3/4}/2 - 5\sqrt{n} \Rightarrow \text{"a" is below median}
\]
Correctness Analysis

Let \( v_1, v_2, v_3, v_4 \) be the values in \( S \) of rank

\[
\begin{align*}
r_1 &= n/2 - 10n^{3/4} , \\
r_2 &= n/2 , \\
r_3 &= n/2 + 10n^{3/4} , \\
r_4 &= n
\end{align*}
\]

where the rank of a value is the number of values less or equal to it.

Define \( X_i = \) number of values sampled in \( R \) less or equal to \( v_i \) and note:

\[
X_4 < 2n^{3/4} \Rightarrow |R| < 2n^{3/4}
\]

\[
X_2 > n^{3/4}/2 - 5\sqrt{n} \Rightarrow \text{“}a\text{” is below median}
\]

\[
X_2 < n^{3/4}/2 + 5\sqrt{n} \Rightarrow \text{“}b\text{” is above median}
\]

If “a” is above \( v_1 \) and “b” is below \( v_3 \) then

\[
|S'| < r_3 - r_1 = 20n^{3/4}.
\]
Correctness Analysis

Let $v_1, v_2, v_3, v_4$ be the values in $S$ of rank

$$r_1 = n/2 - 10n^{3/4}, \quad r_2 = n/2, \quad r_3 = n/2 + 10n^{3/4}, \quad r_4 = n$$

where the rank of a value is the number of values less or equal to it.

Define $X_i =$ number of values sampled in $R$ less or equal to $v_i$ and note:

$$X_4 < 2n^{3/4} \Rightarrow |R| < 2n^{3/4}$$

$$X_2 > n^{3/4}/2 - 5\sqrt{n} \Rightarrow \text{"a" is below median}$$

$$X_2 < n^{3/4}/2 + 5\sqrt{n} \Rightarrow \text{"b" is above median}$$

$$X_1 < n^{3/4}/2 - 5\sqrt{n} \Rightarrow \text{"a" is above } v_1$$
Correctness Analysis

Let $v_1, v_2, v_3, v_4$ be the values in $S$ of rank

$$r_1 = \frac{n}{2} - 10n^{3/4}, \quad r_2 = \frac{n}{2}, \quad r_3 = \frac{n}{2} + 10n^{3/4}, \quad r_4 = n$$

where the rank of a value is the number of values less or equal to it.

Define $X_i = \text{number of values sampled in } R \text{ less or equal to } v_i$ and note:

$X_4 < 2n^{3/4} \Rightarrow |R| < 2n^{3/4}$

$X_2 > n^{3/4}/2 - 5\sqrt{n} \Rightarrow \text{“a” is below median}$

$X_2 < n^{3/4}/2 + 5\sqrt{n} \Rightarrow \text{“b” is above median}$

$X_1 < n^{3/4}/2 - 5\sqrt{n} \Rightarrow \text{“a” is above } v_1$

$X_3 > n^{3/4}/2 + 5\sqrt{n} \Rightarrow \text{“b” is below } v_3$
Correctness Analysis

Let $v_1, v_2, v_3, v_4$ be the values in $S$ of rank

$$r_1 = \frac{n}{2} - 10n^{3/4}, \quad r_2 = \frac{n}{2}, \quad r_3 = \frac{n}{2} + 10n^{3/4}, \quad r_4 = n$$

where the rank of a value is the number of values less or equal to it.

Define $X_i =$ number of values sampled in $R$ less or equal to $v_i$ and note:

$$X_4 < 2n^{3/4} \Rightarrow |R| < 2n^{3/4}$$

$$X_2 > n^{3/4}/2 - 5\sqrt{n} \Rightarrow \text{"a" is below median}$$

$$X_2 < n^{3/4}/2 + 5\sqrt{n} \Rightarrow \text{"b" is above median}$$

$$X_1 < n^{3/4}/2 - 5\sqrt{n} \Rightarrow \text{"a" is above } v_1$$

$$X_3 > n^{3/4}/2 + 5\sqrt{n} \Rightarrow \text{"b" is below } v_3$$

If "a" is above $v_1$ and "b" is below $v_3$ then $|S'| < r_3 - r_1 = 20n^{3/4}$. 
Correctness Analysis

Each $X_i$ is a binomial random variable and $E[X_i] = r_i p$ and $\text{Var}[X] = r_i p (1 - p) \leq np$. Hence, by the Chebyshev Bound

$$\mathbb{P} \left[ |X_i - E[X_i]| \geq \sqrt{n} \right] \leq \text{Var}[X_i]/n \leq n^{-1/4}$$
Correctness Analysis

Each $X_i$ is a binomial random variable and $E[X_i] = r_i p$ and $\mathbb{V}[X] = r_i p (1 - p) \leq n p$. Hence, by the Chebyshev Bound

$$\mathbb{P}\left[|X_i - E[X_i]| \geq \sqrt{n}\right] \leq \mathbb{V}[X_i] / n \leq n^{-1/4}$$

i.e.,

$$E[X_i] - \sqrt{n} < X_i < E[X_i] + \sqrt{n}$$

with probability at least $1 - n^{-1/4}$. 
Correctness Analysis

Each $X_i$ is a binomial random variable and $E[X_i] = r_ip$ and $\mathbb{V}[X] = r_ip(1 - p) \leq np$. Hence, by the Chebychev Bound

$$\mathbb{P} \left[ |X_i - E[X_i]| \geq \sqrt{n} \right] \leq \mathbb{V}[X_i] / n \leq n^{-1/4}$$

i.e.,

$$E[X_i] - \sqrt{n} < X_i < E[X_i] + \sqrt{n}$$

with probability at least $1 - n^{-1/4}$.

In particular, with probability at least $1 - 4n^{-1/4}$,

$$X_1 < \frac{n^{3/4}}{2} - 10\sqrt{n} + \sqrt{n} < \frac{n^{3/4}}{2} - 5\sqrt{n}$$

$$\frac{n^{3/4}}{2} - \sqrt{n} < X_2 < \frac{n^{3/4}}{2} + \sqrt{n}$$

$$\frac{n^{3/4}}{2} + 5\sqrt{n} < \frac{n^{3/4}}{2} + 10\sqrt{n} - \sqrt{n} < X_3$$

$$X_4 < n^{3/4} + \sqrt{n} < 2n^{3/4}$$
Lazy Select

Next Time: Balls and Bins and Birthdays and Coupons
Balls and Bins

Throw \( m \) balls into \( n \) bins where each throw is independent.
Balls and Bins

Throw $m$ balls into $n$ bins where each throw is independent.

- **Birthday Paradox**: How large can $m$ be such that all bins have at most one ball? Applications: Picking IDs without coordination in a Multi-Agent System.
Balls and Bins

Throw $m$ balls into $n$ bins where each throw is independent.

▶ **Birthday Paradox**: How large can $m$ be such that all bins have at most one ball? Applications: Picking IDs without coordination in a Multi-Agent System.

▶ **Coupon Collecting**: How large must $m$ be such that all bins get at least one ball?
Balls and Bins

Throw $m$ balls into $n$ bins where each throw is independent.

- **Birthday Paradox**: How large can $m$ be such that all bins have at most one ball? Applications: Picking IDs without coordination in a Multi-Agent System.

- **Coupon Collecting**: How large must $m$ be such that all bins get at least one ball?

- **Load Balancing**: What is the maximum number of balls that fall into the same bin? Application: Assigning jobs to different machines without overloading any machine.