CMPSCI 611: Advanced Algorithms

Lecture 16: Lazy Select

Andrew McGregor
Lazy Select

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

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value.

**Algorithm**

1. *Add each element in $S$ to a set $R$ with probability $p = 1/n^{1/4}$.*
2. *Call this set $R$. Sort $R$ and let*

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

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

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 $\mathbb{V}[X] = r_i p(1 - p) \leq np$. Hence, by the Chebychev Bound

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

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.

- **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.