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

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

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:

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

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

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

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

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

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 $E[X_i] = r_ip$ and $\variance[X] = r_ip(1 - p) \leq np$. Hence, by the Chebyshev Bound

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

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

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 must $m$ be such that all bins have at most one ball? Applications: Cryptographic Attacks and 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? Application: Assigning jobs to different machines without overloading any machine.

- **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.
Lemma
\[ \Pr[\text{first } m \text{ balls fall in distinct bins}] \leq e^{-m(m-1)/(2n)}. \]

Proof.

- Let \( A_i \) be event that the \( i \)-th ball lands in a bin not containing any of the first \( i - 1 \) balls.
- \( \Pr[\bigcap_{1 \leq i \leq m} A_i] = \Pr[A_1] \Pr[A_2|A_1] \ldots \Pr[A_m|\bigcap_{1 \leq i \leq m-1} A_i] \)
- \( \Pr[A_i|\bigcap_{1 \leq j \leq i-1} A_j] = 1 - (i - 1)/n \)
- Putting it together and using \( \sum_{1 \leq i \leq a} i = (a + 1)a/2 \):

\[
\Pr[\bigcap_{1 \leq i \leq m} A_i] = \prod_{1 \leq i \leq m} \left( 1 - \frac{i - 1}{n} \right) \leq e^{-m(m-1)/(2n)}
\]

With \( n = 365 \) and \( m = 29 \), probability \( < e^{-1} \). Tighter analysis possible.
Coupon Collecting

Suppose you throw $r$ balls into $n$ bins. If each ball is equally likely to land in each bin, how large does $r$ need to be such that a ball lands in every bin with probability at least $1 - 1/n$. We’ll show $r = 2n \ln n$ are sufficient.

- Let $A_i$ be the event that the $i$th bin is empty after $r$ balls are thrown. Then,

  $$\Pr[A_i] = (1 - 1/n)^r = (1 - 1/n)^{2n \ln n} \leq e^{-2 \ln n} = 1/n^2$$

- Then $A_1 \cup A_2 \cup \ldots \cup A_n$ is the event that there is an empty bin:

  $$\Pr[A_1 \cup A_2 \cup \ldots \cup A_n] \leq \Pr[A_1] + \Pr[A_2] + \ldots + \Pr[A_n] = n \times 1/n^2 = 1/n$$
Load Balancing

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

▶ How full is the fullest bin? This has applications to load balancing.
▶ What’s the probability that $k$ or more items land in bin 1?
▶ If $X_1$ is the number of balls that land in bin 1 then $X_1$ is a binomial distribution with $m$ trials and $p = 1/n$.

▶ **Lemma:** $P(X_1 \geq k) \leq \binom{m}{k} p^k$.
▶ If $m/n = 1$ and $k = 2 \log n$,

$$P(X_1 \geq k) \leq \binom{m}{k} p^k \leq \frac{m^k}{k!} \cdot \left(\frac{1}{n}\right)^k = \left(\frac{m}{n}\right)^k / k! = 1/k! \leq 1/2^k = 1/n^2$$

▶ Same analysis applies to $X_2, X_3, \ldots$, i.e., the number of balls in bins 2, 3, \ldots. Hence, no bin has more than $k = 2 \log n$ balls in it with probability at least $1 - 1/n$. 
**Lemma**

Let $X$ be the number of heads observed when we toss $m$ coins each with probability of heads equal to $p$. Then $\Pr[X \geq k] \leq \binom{m}{k} p^k$.

- Let $S_1, S_2, \ldots, S_{\binom{m}{k}}$ be all subsets of $[m]$ with exactly $k$ elements.
  
  \[
P(A_{S_j}) = p^k
  \]
  
  where $A_S$ is the event that for all $i \in S$, the $i$th coin toss is heads.

- Then $A_{S_1} \cup A_{S_2} \cup \ldots \cup A_{S_{\binom{m}{k}}}$ is the event you get $k$ or more heads.

- Hence,

\[
P(k \text{ or more heads}) = P(A_{S_1} \cup A_{S_2} \cup \ldots \cup A_{S_{\binom{m}{k}}}) \leq \sum_{j=1}^{\binom{m}{k}} P(A_{S_j}) = \binom{m}{k} p^k
\]