Given a set $S$ of $n$ elements $x_1, x_2, \ldots, x_n$, and an integer $1 \leq K \leq n$, we want to select an element in $S$ with rank $K$.

$$\text{rank}(x, S) = \left| \{ x_i \in S \mid x_i \leq x \} \right|$$

$S = 6, 3, 9, 4, 20 \mid 3, 4, 6, 9, 20$

$x = 3.8, \quad \text{rank}(x, S) = 1$

Select $(S, K)$: returns an element $x_k$ in $S$ with rank $K$.

Assume all elements in $S$ to be distinct.

$$[x_1, 1], (x_2, 2), (x_3, 3), \ldots, (x_n, n)$$

$x_i = x_j$ if $x_i \leq x_j$ and $x_j < x_i$.

$x_i = x_j$ if $x_i = x_j$ and $\text{smaller}(i, j)$.
Sorting (S) vs. Select (S, K)

1. Selection is reducible to Sorting
   \( \xrightarrow{\sim} \)

2. Sorting can be accomplished by multiple invocations of Selection
   \( \times \)

Select (S, K) runs in \( O(n \log n) \) comparisons

Can we select in \( O(n) \) steps?

Suppose \( k = 1 \) or \( k = n \) trivial

\( k = 2 \), \( k = 3 \)

This procedure takes \( O(k \cdot n) \) steps

\( k = \frac{n}{2} \) (median) \( \Omega(n^2) \)

Look at the sorted set \( S \) (we don't)

\( \tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \ldots, \tilde{x}_n \)

\( \tilde{x}_i < \tilde{x}_{i+1} \),
1. Choose an arbitrary element \( r \) from \( S \).
2. Lucky? Find \( \text{rank}(r, S) \).

   Tune \( N \) comparisons.

What is the probability of success?

\[
= \frac{1}{N}
\]

Pick up the \( k \)-th rank element. Using a random choice: every element is picked with equal probability.

Random variables \( X \),

Expectation of \( X \), \( E[X] \)

\( X \) = # times we iterate

\( X \in \{1, 2, 3, \ldots\} \)

Probability distribution of \( X \), say

\[
\text{Prob}[X = i] = p_i
\]

\[
E[X] = \sum_{i \geq 1} i \cdot p_i
\]
Let $P_i$ follow geometric distribution

Fail $i-1$ times and succeed on the $i^{th}$ trial where every trial is "independent"

$$P_i = ? \quad (1 - \frac{1}{n})^{i-1} \times \frac{1}{n}$$

If success prob is $p$ then $i_i = (1-p)^{i-1} \times p$

$$E[x] = ? \quad \frac{1}{p} = \frac{1}{\frac{n}{p}}$$

$$Pr \left[ x > k \cdot E[x] \right] \leq \frac{1}{k}$$

Markov's inequality for non-negative random variables

Proof (by contradiction): Suppose $j$ is the smallest integer such that $j > k \cdot E[x]$

Then $\sum_{t \geq j} t \cdot \frac{1}{p} > \sum_{t \geq j} j \cdot \frac{1}{p} = j \cdot \frac{1}{p}$

$$> k \cdot E[x] \left( \frac{1}{k} \right) > E[x]$$

contradiction