From Markov's inequality, the probability that the # iterations exceed \( 2 \cdot n \) is \( \leq \frac{1}{2} \) (use \( k = 2 \)).

Alternatively, the prob that we fail in consecutive \( n \) iterations:

\[
\leq \left(1 - \frac{1}{n}\right)^n \leq e^{-\frac{1}{2}} \leq \frac{1}{2}
\]

\((1 + x \leq e^x \text{ for any } x)\)

\(\Rightarrow\) with 50% likelihood, we will succeed within \( O(n) \) iterations i.e. about \( O(n^2) \) comparisons.

\[\frac{X_1}{X_2} / \frac{X_3}{X_4} / ... / \frac{X_n}{X_n} \]

discard

remaining

Revise the value of \( k \) (in this case \( \frac{n}{4} \))
Define the elements with ranks \( r \in \left[ \frac{n}{4}, \frac{3n}{4} \right] \) as "good" elements. Since they can be used to prune at least \( \frac{n}{4} \) elements for the next round.

Observation: If we pick a "good" splitter every time, then there are at most \( \log_{4/3} n \) iterations.

\[ \implies \text{Total # comparisons} \]
\[ n + \frac{3n}{4} + (\frac{3}{4})^2 n + \cdots - n = O(n) \]

Prove that picking a good element is
\[ \frac{n}{2n} = \frac{1}{2} \]

\[ \implies \text{Let } Y \text{ represent the # trials before picking a "good" element} \]
\[ E[Y] = \frac{1}{1} = 2, \quad E[Y_i] = 2 \]

Let \( Y_i \) represent the number of trials in recursive level \( i \).

In 1st level there are \( n \) elements.

2nd level: \( \frac{3n}{4} \)

In general: \( \left(\frac{3}{4}\right)^i \cdot n \)

Overall, the number of comparisons can be bounded by:

\[ \sum_{i=1}^{\infty} \left(\frac{3}{4}\right)^i \cdot n \cdot Y_i \]

Comparisons

Total number of comparisons:

\[ E[T] = E \left[ \sum_{i=1}^{\infty} \left(\frac{3}{4}\right)^i \cdot n \cdot Y_i \right] = \sum_i E \left[ \left(\frac{3}{4}\right)^i \cdot n \cdot Y_i \right] \]

Linearity property of expectation.

For any r.v. \( X_1, X_2 \), not necessarily independent:

\[ E[X_1 + X_2] = E[X_1] + E[X_2] \]
\[
\sum_i E \left[ \left( \frac{3}{4} \right)^i n \cdot Y_i \right] \\
= \sum_i n \cdot \left( \frac{3}{4} \right)^i \cdot E[Y_i] \\
= n \sum_i \left( \frac{3}{4} \right)^i \cdot 2 \\
= 2n \sum_i \left( \frac{3}{4} \right)^i \\
= O(n)
\]

The expected number of comparisons over all the iterations is \( \leq c n \).

From Markov's inequality the prob.
that we exceed \( k \) comparisons is \( \leq \frac{1}{k} \).

The expected running time does not depend on the input distribution (i.e. not averaged over input) for randomized algorithms.
They are independent I input distribution, i.e. worst case input.

The averaging is done over random choices made inside the algorithm (not controlled by anyone – depends on the random no. generated).

To make the selection algo deterministic, we would like to pick a “good” element with certainty.

Claim: The "median of medians" is a "good element."