CSL 356 Aug 2

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

Alternately

The prob that we fail in consecutive n iterations

$$\leq \left(1 - \frac{1}{n}\right)^n \leq \frac{1}{e} = \frac{1}{2.7} < \frac{1}{2}$$

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

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

$$\underbrace{\cancel{\tilde{x}_1}\ \cancel{\tilde{x}_2} \cdot \overset{\tilde{x}_{n/4}}{\boxed{\cancel{\cdot}}}}_{\text{discard}} \cdot\ \underbrace{\boxed{\tilde{x}_{\frac{n}{2}}} \cdots\ \tilde{x}_n}_{\text{remaining}}$$

Revise the value of K (in this case $\frac{n}{4}$)

$\tilde{x}_1 \quad \tilde{x}_2 \quad \tilde{x}_3 \quad \underbrace{\tilde{x}_{n/4} \; \cdots \; \tilde{x}_{n/2} \qquad \tilde{x}_{3n/4}}\; \cdots \; \tilde{x}_n$

Define the elements with ranks $\in \left[\frac{n}{4} \ldots \frac{3n}{4}\right]$ as "good" elements. Since they can used to prune at least $\frac{n}{4}$ elements for the next round.

<u>Observation</u>: $\boxed{\text{If}}$ we pick a "good" splitter every time, then there are at most $\rightarrow \log_{4/3} n$ iterations

$\Rightarrow$ Total # comparisons

$$n + \frac{3n}{4} + \left(\frac{3}{4}\right)^2 n + \cdots$$

$$\sim O(n)$$

Prob of picking a good element is

$$\frac{\frac{n}{2}}{n} = \frac{1}{2}$$

$\Rightarrow$ Let Y represent the # trials before we pick a good element

$E[Y] = \frac{1}{1/2} = 2$. $E[Y_i] = 2$

Let Y_i represent the # trials in recursive level i

In 1^{st} level there are n elements

2^{nd} " " " $\frac{3n}{4}$ "

i^{th} $\left(\frac{3}{4}\right)^i \cdot n$ "

Overall the # comparisons can be bounded by $\sum_{i=1} \left[\left(\frac{3}{4}\right)^{i-1} \cdot n\right] Y_i$ comparisons

Total # comparison $= T$

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

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 # comparisons over all the iterations is $\leq cn$

From Markov's inequality the prob that we exceed $K \cdot cn$ comparisons $\leq \frac{1}{K}$

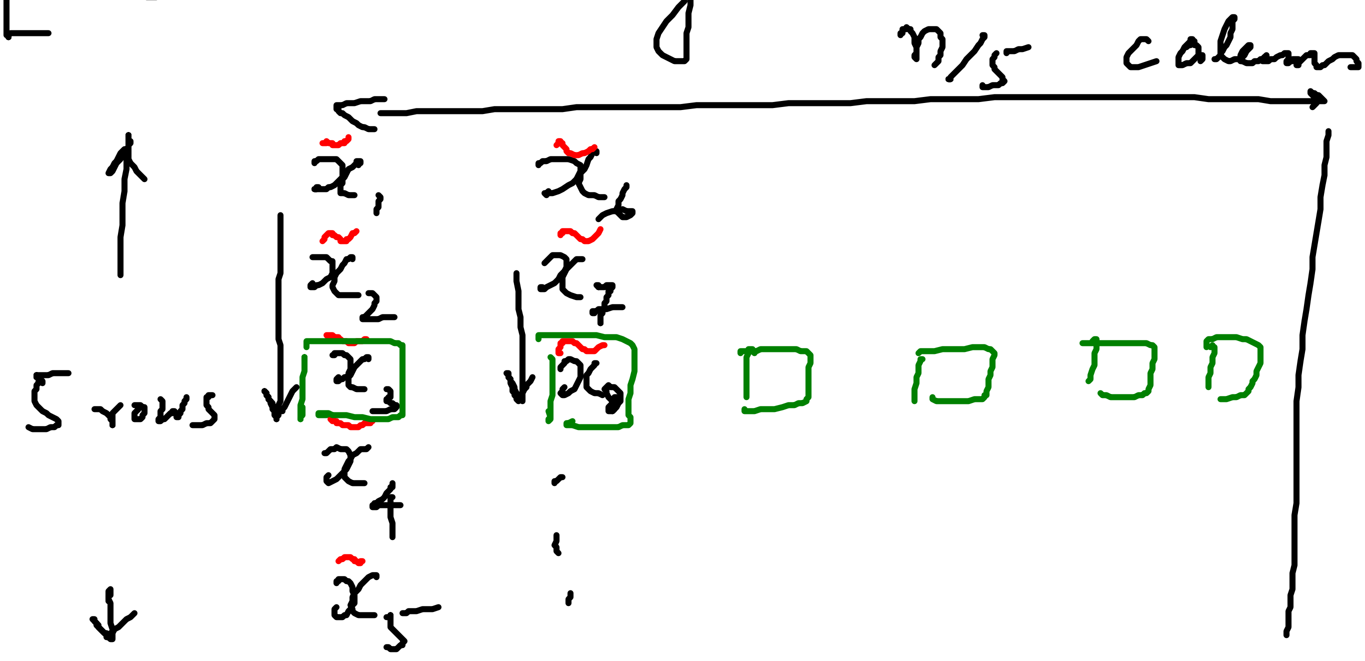
The expected running time <u>does not</u> depend on the input distribution (i.e. not averaged over input)

<u>Randomized algorithms</u>

→ They are independent of 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. generator)

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".