Analysis in high probability of Randomized QuickSort

RandQuickSort(S) $\qquad |S| = n$, all distinct

if $|S| \leq 1$ then return $\langle S \rangle$

$p = \text{RANDOM}(S)$ \\ pick a "pivot" element uniformly at random from S

$S_1 = \{x \in S \text{ s.t. } x < p\}$ \\ $O(n)$ time

$S_2 = \{x \in S \text{ s.t. } x > p\}$ \\ $O(n)$ time

$Z_1 = \text{RandQuickSort}(S_1)$

$Z_2 = \text{RandQuickSort}(S_2)$

return $\langle Z_1, p, Z_2 \rangle$

Is a LAS VEGAS alg.

Suppose p is always the <u>median</u> of S; then

$$T_{RQS}(n) = \begin{cases} 2T_{RQS}\left(\frac{n}{2}\right) + O(n) & n > 1 \\ 0 & n \leq 1 \end{cases}$$

$$T_{RQS}(n) = O(n \log n)$$

However, p is the median with probability $\frac{1}{n}$, very low

(Can one find the median in linear time? Yes, deterministically $\Rightarrow$ deterministic QS $O(n \log n)$ exists, but the alg. is very complicated and the hidden constant is very high, thus inefficient in practice)

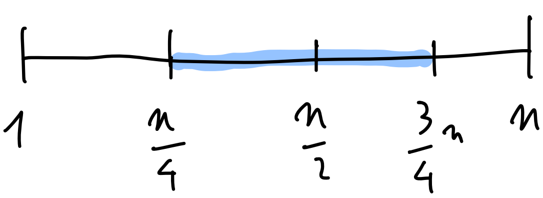
But do we really want <u>exactly</u> the median?

Intuition: if the sizes of S_1 and S_2 are "not too unbalanced", we should be good.

Let's try with a loose request:

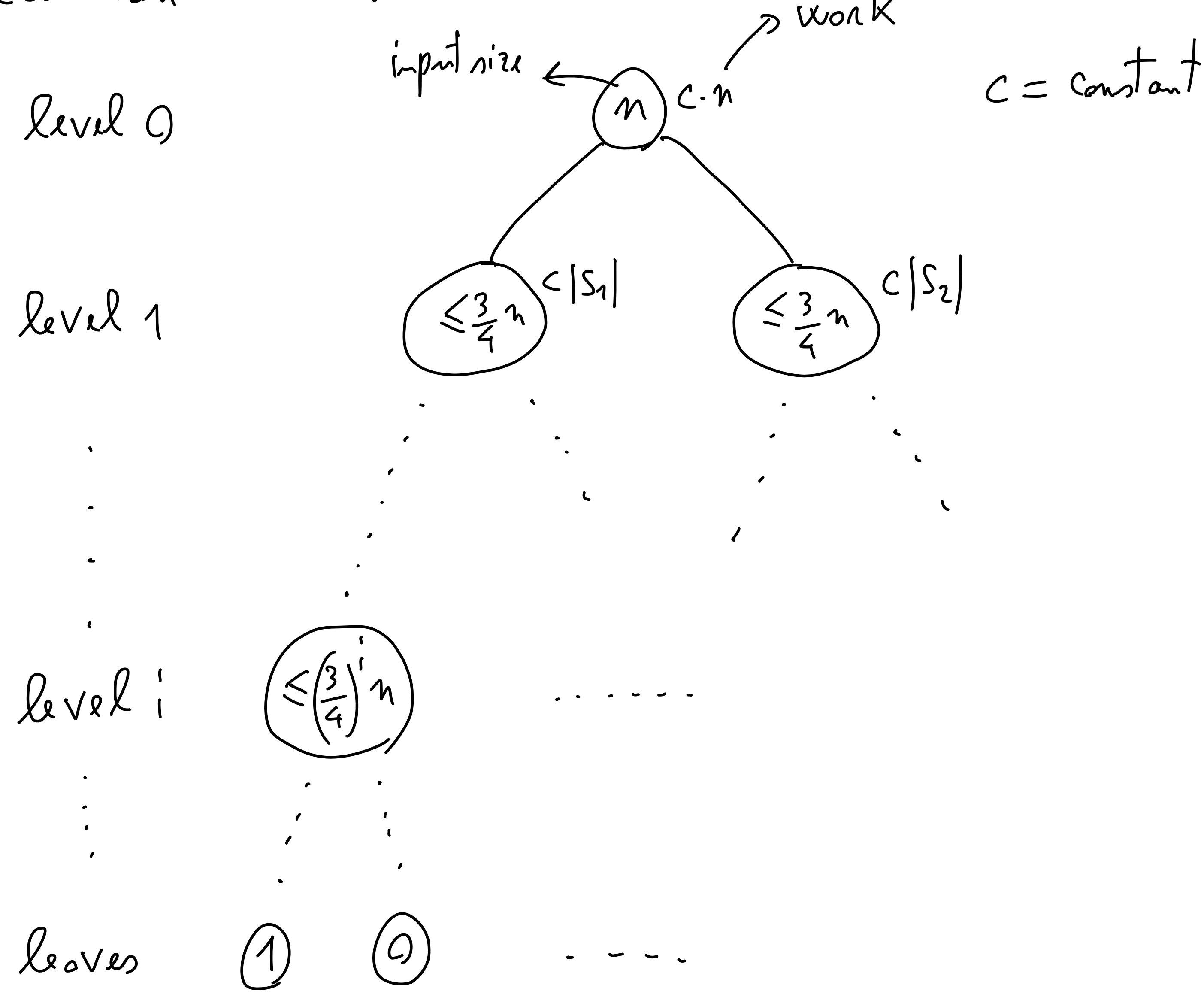
$$\begin{cases} |S_1| \leq \frac{3}{4} n \\ |S_2| \leq \frac{3}{4} n \end{cases}$$

that is, pivot p chosen from



is this good enough for us?

Recursion Tree :



- total work at each level is $\leq c \cdot n$
- depth of the recursion tree $= \min \left\{ \text{integers } i \text{ s.t. } \left(\frac{3}{4}\right)^i n \leq 1 \right\} = \lceil \log_{4/3} n \rceil = O(\log n)$

$$\left(\frac{3}{4}\right)^i n \leq 1 \iff \left(\frac{3}{4}\right)^i \leq \frac{1}{n} \iff \left(\frac{4}{3}\right)^i \geq n$$

$$\iff \log_{4/3} \left(\frac{4}{3}\right)^i \geq \log_{4/3} n \iff i \geq \log_{4/3} n$$

$\Rightarrow T_{RQS}(n) = O(n \log n)$

that is, it's not necessary that S_1 and S_2 be perfectly balanced. I have $\simeq \frac{n}{2}$ "good" choices for the pivot p.

Analysis

hope: depth of the recursion tree $= O(\log n)$ w.h.p.
that is, all the $\leq n$ distinct root-leaf paths have $O(\log n)$ length w.h.p.

Event "lucky choice of the pivot": pivot chosen between the $\left(\frac{n}{4}+1\right)$-th order statistic and the $\left(\frac{3}{4}n\right)$-th order statistic

$$\Pr(\text{"lucky choice"}) = \frac{\frac{3}{4}n - \left(\frac{n}{4}+1\right)+1}{n} = \frac{1}{2}$$

Fix one root-leaf path P:

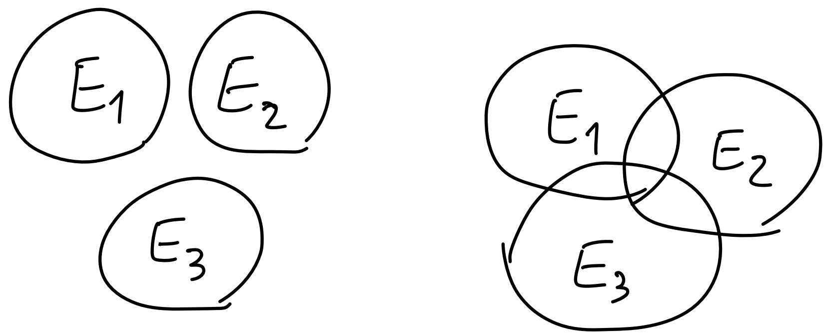
Lemma: $\Pr\left(|P| > \underbrace{a}_{\text{constant}} \cdot \log_{4/3} n\right) < \frac{1}{n^3}$

if this is true we're done:

Lemma (Union bound): for any random events $E_1, E_2, \ldots, E_k$:

$$\Pr\left(E_1 \cup E_2 \cup \cdots \cup E_k\right) \leq \Pr(E_1) + \Pr(E_2) + \ldots + \Pr(E_k)$$

Proof by picture:



$\Pr\left(\text{all root-leaf paths have length} \leq a \cdot \log_{4/3} n\right) =$

$1 - \Pr\left(\exists \text{ path} > a \cdot \log_{4/3} n\right) \geq \ldots$

E_i = path P_i has length $> a \cdot \log_{4/3} n$

$$\Pr\left(\exists \text{ path} > a \cdot \log_{4/3} n\right) = \Pr\left(\bigcup_{i=1}^{n} E_i\right)$$

$$\leq \sum_{i=1}^{n} \Pr(E_i) \qquad \text{(Union bound)}$$

$$< n \cdot \frac{1}{n^3} \qquad \text{(Lemma)}$$

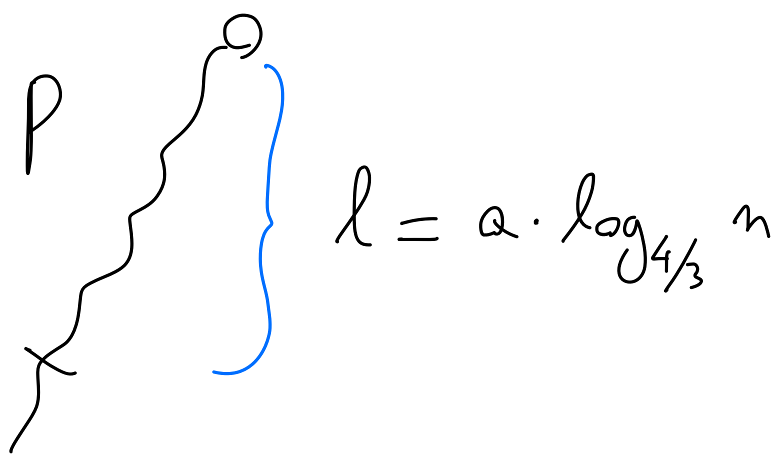
$$= \frac{1}{n^2}$$

$$\ldots \geq 1 - \frac{1}{n^2}$$

$$\Rightarrow T_{RQS}(n) = O(n \log n) \quad \text{w.h.p.}$$

It remains to prove that

$$\Pr\left(|P| > \underbrace{a}_{\text{constant}} \cdot \log_{4/3} n\right) < \frac{1}{n^3}$$



E = "in the first $\ell = a \cdot \log_{4/3} n$ vertices of P there have been $< \log_{4/3} n$ lucky choices"

$X_i \quad 1 \leq i \leq \ell = a \cdot \log_{4/3} n$

$X_i = 1$ if at the i-th vertex of P there is a lucky choice of the pivot

$Pr(X_1 = 1) = \frac{1}{2} \quad \forall i$

X_i are independent

$Pr\left(\sum_{i=1}^{\ell} X_i < \log_{4/3} n\right)$ to bound

$$X = \sum_{i=1}^{\ell} X_i$$

$$\mu = E[X] = E\left[\sum_{i=1}^{\ell} X_i\right] = \sum_{i=1}^{\ell} E[X_i] =$$

$$= \sum_{i=1}^{\ell} \frac{1}{2} = \frac{\ell}{2} = \frac{a}{2} \cdot \log_{4/3} n$$

Let's apply this:

$$\Pr\left(X < (1-\delta)\mu\right) < e^{-\frac{\mu\delta^2}{2}} \qquad 0 < \delta \leq 1$$

$$\downarrow$$

$$(1-\delta)\mu = \log_{4/3} n$$

$$(1-\delta)\frac{a}{2}\log_{4/3} n = \log_{4/3} n$$

one possible choice is $a = 8$

$\delta = 3/4$

$$\Pr\left(X < \log_{4/3} n\right) < e^{-\frac{8}{4}\log_{4/3} n \cdot \frac{9}{16}}$$

$$= e^{-\log_{4/3} \cdot \frac{9}{8}}$$

$$< e^{-\log_{4/3} n}$$

$$= e^{-\frac{\ln n}{\ln 4/3}}$$

$$= \left(e^{-\ln n}\right)^{1/\ln 4/3}$$

$$= \left(\frac{1}{n}\right)^{1/\ln 4/3} \quad 1/\ln 4/3 \simeq 3.47$$

$$< \frac{1}{n^3}$$