Randomized Selection

Algorithm RSelect(S, k):

// S is a sequence of numbers, $1 \leq k \leq |S|$,
// returns the k-th smallest element in S

if $|S| = 1$: return the only element in S

if $|S| \geq 2$: p = uniformly random element in S;

by scanning S, divide it into

$$\underbrace{\boxed{\quad < p \quad}}_{S_<}\underbrace{\boxed{\qquad = p \qquad}}_{S_=}\underbrace{\boxed{\quad > p \quad}}_{S_>}$$

if $k \leq |S_<|$: RSelect$(S_<, k)$

if $|S_<| + 1 \leq k \leq |S_<| + |S_=|$: return p

if $k \geq |S_<| + |S_=| + 1$:

RSelect$(S_>, k - |S_<| - |S_=|)$

RS 2

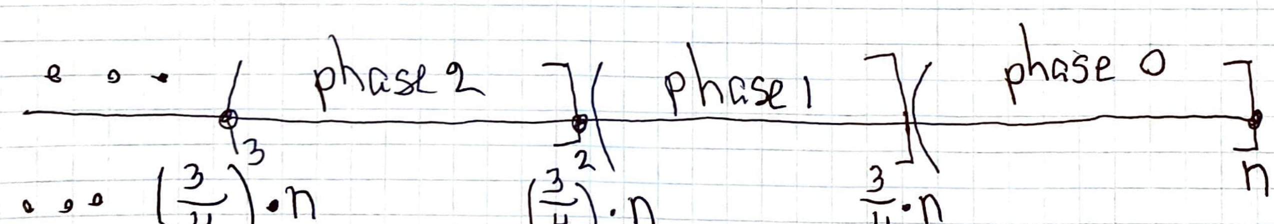
Let n be a large integer, let S be a sequence of length n. Define random variable

T = time when running RSelect(S,k).

We are going to show: $E(T) = O(n)$.

When we run RSelect(S,k), recursive calls are generated on smaller and smaller sequences.

For $i = 0,1,2,\ldots$: a call is in <u>phase</u> i if the length of the sequence in this call is $\leq \left(\frac{3}{4}\right)^i \cdot n$ and $> \left(\frac{3}{4}\right)^{i+1} \cdot n$



Initially: phase 0, because $|S| = n$

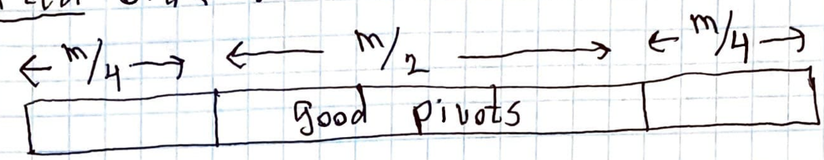
During recursive calls, we either stay in the same phase or move to a phase with a larger index. Which case happens depends on the pivot p (which is randomly chosen).

Consider a call in phase i.

m = length of the sequence in this call.

$$\left(\tfrac{3}{4}\right)^{i+1} \cdot n < m \le \left(\tfrac{3}{4}\right)^{i} \cdot n$$

For purpose of analysis, consider these m numbers in <u>sorted</u> order :



pivot p is <u>good</u> if p is in the middle half

* $\Pr(\text{pivot } p \text{ is good}) = 1/2$
* Assume pivot p is good. If there is a next call, then it is on a sequence of length

$$\le m - m/4 = \frac{3}{4} \cdot m \le \left(\frac{3}{4}\right)^{i+1} \cdot n$$

$\therefore$ this next call is in a phase $\geq i+1$

RS4

From COMP 2804:

experiment → success with probability α
experiment → failure with probability $1-\alpha$

Repeat experiment until success for the first time.

Then: $E(\#\text{times}) = 1/\alpha$

Define random variable

X_i = # calls in phase i

Then: $E(X_i) \leq 2$ [Exercise: Why $\leq$? Why not $=$?]

* time for one call in phase i (excluding the recursive call)

$$\leq cm \leq c\left(3/4\right)^i \cdot n$$

for some constant c.

R55

Remember: T = time when running
RSelect(S,k), where $|S| = n$.

$$T \leq \sum_{i \geq 0} c\left(\tfrac{3}{4}\right)^i \cdot n \cdot X_i$$

Using linearity of expectation:

$$E(T) \leq \sum_{i \geq 0} c\left(\tfrac{3}{4}\right)^i \cdot n \cdot E(X_i)$$

$$\boxed{\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}}$$

$$\leq 2cn \sum_{i=0}^{\infty} \left(\tfrac{3}{4}\right)^i$$

$$= 2cn \cdot \frac{1}{1-3/4}$$

$$= 8cn$$

$$= O(n)$$