EXERCISE 3 Consider the following algorithm:

```
        SELECT(S, i)    { 1 ≤ i ≤ |S|         }
        n ← |S|         { selects i-th o.s.   }
(B)     if (n=1) then return S[1]
(D)     p ← RANDOM(S)
        S1 ← {s∈S: s<p}; S2 ← {s∈S: s>p}
        if (|S1| = i-1) then return p
(R+C)   if (|S1| > i-1)
          then return SELECT(S1, i)
          else return SELECT(S2, i-|S1|-1)
```

1. Prove that SELECT(S, i) correctly returns the i-th order-statistic of S (i.e., i-th smallest element)

2. Analyze its average running time

---

1. induzione su n:
   - base: $n=1 \Rightarrow$ OK
   - induzione: $n>1$:
     - $|S_1| = i-1 \Rightarrow S_1$ contiene $x_1, \dots, x_{i-1} \Rightarrow$ subito dopo c'è $x_i = p$
     - $|S_1| > i-1 \Rightarrow S_1$ contiene $x_1, \dots, x_{|S_1|} \Rightarrow x_i = p$ sarà in S_1 con rango i
     - $|S_1| < i-1 \Rightarrow x_i$ sarà in $S_2 \Rightarrow$ rango di x_i in S_2: $i - |S_1| - 1$

2. w.h.p. dà $O(n \log n)$ ⇒ vogliamo tempo lineare
recursion path in worst case:

$$(n_0 = n) - (n_1) - \dots - (n_{\ell-1}) - (n_\ell = 1)$$

(mai $|S_1| = i-1$)

work $\quad n_0 \quad n_1 \quad \dots \quad n_{\ell-1} \quad 1$

(in realtà $n_i - 1$)
(solo confronti fra el.,
non fra dimensioni)

n_i: r.v. ⇒ stesso argomento di quicksort
$\Pr[\text{scelta pivot fortunata}] = \Pr[n_j \leq \frac{3}{4} n_{j-1}] \geq 1/2$

Dividiamo recursion path in segmenti $\sigma_1, \sigma_2, \dots$

$$(n_{\sigma_1} = n) - \dots - (n_{\sigma_2}) - \dots - (n_{\sigma_3}) - \dots$$

qui ho ridotto
dimensione di 3/4 ⇒ $n_{\sigma_i} \leq \frac{3}{4} n_{\sigma_{i-1}}$
k^*: ultimo segmento è $n_{\sigma_{k^*}}$ ⇒ $k^* \leq \lceil \log_{4/3} n \rceil + 1 \left(\Rightarrow n_{\sigma_{k^*}} \leq \left(\frac{3}{4}\right)^{k^*} n \right)$
Tutti nodi in segmento σ_i fanno lavoro $\leq n_{\sigma_i}$

$$\mathbb{E}[T(n)] = \mathbb{E}\left[\sum_{i=1}^{k^*} n_{\sigma_i} \cdot |\sigma_i|\right] = \mathbb{E}\left[\sum_{i=1}^{k^*} \left(\frac{3}{4}\right)^{i-1} n \cdot |\sigma_i|\right] =$$
$$= \sum_{i=1}^{k^*} \mathbb{E}\left[\left(\frac{3}{4}\right)^{i-1} n \, |\sigma_i|\right] = \sum_{i=1}^{k^*} \left(\frac{3}{4}\right)^{i-1} n \, \mathbb{E}[|\sigma_i|]$$

$p \geq 1/2$: scelta fortunata ⇒ avanti in segmento finché non la prendo
⇒ $|\sigma_i| \sim Geom(p \geq 1/2) \Rightarrow \mathbb{E}[|\sigma_i|] = 1/p \leq 2$ ⇒

$$\Rightarrow \mathbb{E}[T(n)] \leq \left(\sum_{i=1}^{k^*} \left(\frac{3}{4}\right)^{i-1}\right) 2n \leq \left(\sum_{i=1}^{\infty} \left(\frac{3}{4}\right)^{i-1}\right) 2n = \left(\sum_{j=0}^{\infty} \left(\frac{3}{4}\right)^{j}\right) 2n = \frac{1}{1-\frac{3}{4}} 2n$$

Whp: $c \log n$ nodi ⇒ in questi livelli non sappiamo dove avremo scelte fortunate ⇒ contributi diverse

REMARK

- Unlike QUICKSORT, we cannot prove that

$$T_{SELECT}(n) = \Theta(n) \text{ w.h.p.!}$$

- High-probability randomized selection is possible:

IDEA: HP_SELECT(S, i)

- based on large sample selection

$\Theta(n^{3/4})$ - pick $\sim n^{3/4}$ random elements

$\Theta(n^{3/4}\log n^{3/4}) = \Theta(n^{3/4}\log n)$ - sort them

$\Theta(1)$ - pick $\underline{a, b}$ of rank $\sim i \cdot n^{-1/4} \pm \sqrt{n}$ in the sequence

$\Longrightarrow$

$\Theta(n)$ w.h.p. Set $S' = \{s \in S : a \leq s \leq b\}$

1) contains i-th o.s.

2) has size $O(n^{3/4})$

$\Longrightarrow$ Sort S' to identify i-th o.s.

$\Theta(n^{3/4}\log n)$

EXERCISE 4 Consider k i.i.d. geometric variables:

$$Z_1, \dots, Z_k \sim Geom(p) \quad 0<p<1$$

$$\forall i: Pr(Z_i = j) = p(1-p)^{j-1}, \; j \in \mathbb{Z}^+$$

Recall that $E[Z_i] = \frac{1}{p}$.

Let $X = \sum_{i=1}^{k} Z_i$.

Then $\mu = E[X] = \sum_{i=1}^{k} E[Z_i] = \frac{k}{p}$.

Discuss how to use Chernoff's bounds to upper bound

$$Pr(X > t\mu) \quad t \in \mathbb{Z}^+$$

(Hint: associate the event to Bernoulli trials)

Colleghiamo X a sequenza di lanci monete

Z_i: # lancio con i-esimo successo dopo $(i-1)$-esimo

$$\left.\begin{array}{l} E_1 = \text{"} X > t\mu \text{"} \\ E_2 = \text{"} <k \text{ successi in } t\mu \text{ lanci"} \end{array}\right\} E_1 \to E_2$$

$Y_1, \dots, Y_{t\mu} \sim Ber(p) \Rightarrow E_2 = \text{"} Y = \sum_{i=1}^{t\mu} Y_i < k \text{"} \Rightarrow \mathbb{E}[Y] = t\mu p = t\frac{k}{p}p = tk \Rightarrow$

$\Rightarrow \Pr[E_2] = \Pr[Y < \mathbb{E}[Y]/t] \Rightarrow \Pr[Y < (1-\varepsilon)\mathbb{E}[Y]], \; 1/t = 1-\varepsilon$

EXERCISE: Given a sorted array $A[1..n]$, assume that it can be read only via method

$$A.\text{read}(i) = \begin{cases} A[i] & p = 1/2 \\ \text{error} & p = 1/2 \end{cases}$$

(models faulty storage).

Write a randomized binary search routine SEARCH(A, 1, n, K) returning 1 iff $\exists i : A[i] = K$ executing $O(\log n)$ reads w.h.p.

HINT: Same structure of binary search but each value is read multiple times until A.read(·) ≠ error. Use the geometric bound in the analysis.

EXERCISE 4 Consider n i.i.d. variables $X_i \in \{-1, 1\}$, with

$$Pr(X_i = 1) = Pr(X_i = -1) = \frac{1}{2}$$

Let $X = \sum_{i=1}^{n} X_i$.

1. Show how to adapt Chernoff Bound 2 to get an upper bound
$$Pr(X > \delta), \quad \delta > 0$$

2. Determine a value $\delta_n : Pr(X > \delta_n) < \frac{1}{n}$

EXERCISE 5: Consider the unit square $S:[0,1]\times[0,1] \in \mathbb{R}^2$ and its K^2 subsquares of size $1/K \times 1/K$. How many calls to RANDOM(S) to cover all subsquares with $p \geq 1 - \frac{1}{K^2}$?