



## UNIVERSITÀ DI PISA

# Second hands-on: Depth of a node in a Random Search Tree

Algorithm Design (2021/2022)

Gabriele Pappalardo

Email: g.pappalardo4@studenti.unipi.it

Department of Computer Science

March 2022

## 1 Introduction

A **random binary search tree** for a set  $S$  can be defined as follows: if  $S$  is empty, then the null tree is a random search tree; otherwise, choose uniformly at random a key  $k \in S$ : the random search tree is obtained by picking  $k$  as *root*, and the random search trees on  $L = \{x \in S : x < k\}$  and  $R = \{x \in S : x > k\}$  become, respectively, the left and right subtrees of the root  $k$ .

Consider the *Randomized Quick Sort* (RQS) discussed in class and analysed with indicator variables, and observe that the random selection of the pivots follows the above process, thus producing a random search tree of  $n$  nodes.

- Using a variation of the analysis with indicator variables  $X_{ij}$ , prove that the expected depth of a node (i.e. the random variable representing the distance of the node from the root) is nearly  $2 \log n$ .
- Prove that the expected size of its subtree is nearly  $2 \log n$  too, observing that it is a simple variation of the previous analysis.
- Prove that the probability that the depth of a node exceeds  $c 2 \log n$  is small for any given constant  $c > 2$ .

## 2 Solution

We define the elements of a random binary search tree with  $\forall i \in [1, n]. x_i \in S$ . Let  $X_{ij}$  an indicator random variable defined as follows:

$$X_{ij} = \begin{cases} 1 & x_i \text{ is an ancestor of } x_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The depth of a node can be expressed as the sum of all the indicator random variables. Let  $D_i$  be the depth of a node  $x_i$  in a random binary search tree, defined as  $D_i = \sum_{j=1}^n X_{ij}$ . We can express the expected depth as:

$$E[D_i] = E\left[\sum_{j=1}^n X_{ij}\right] = \sum_{j=1}^n E[X_{ij}] = \sum_{j=1}^n \Pr(\{X_{ij} = 1\}) \quad (2)$$

The probability of  $X_{ij} = 1$  can be computed using the following fact: if the node  $x_i$  is an ancestor of  $x_j$  it means that  $x_i$  is inserted before  $x_j$  (and  $i > j$ ), therefore in the set  $\{x_j, x_{j+1}, \dots, x_i\}$  we have  $i - j + 1$  elements to choose  $x_j$ , otherwise, if  $x_i$  is a descendant of  $x_j$  it means that  $x_i$  has been inserted after ( $i < j$ ), therefore in the set  $\{x_i, x_{i+1}, \dots, x_j\}$  we have  $j - i + 1$  elements to choose.

$$\Pr(\{X_{ij} = 1\}) = \begin{cases} \frac{1}{i-j+1} & i > j \\ \frac{1}{j-i+1} & i < j \end{cases} \quad (3)$$

The expected depth of a node can be computed using this fact:

$$\begin{aligned} E[D_i] &= \sum_{j=0}^n \Pr(\{X_{ij} = 1\}) = \\ &= \sum_{j=0}^i \frac{1}{i-j+1} + \sum_{j=i+1}^n \frac{1}{j-i+1} = \\ &= \sum_{k=1}^{i+1} \frac{1}{k} + \sum_{k=2}^{n-i+1} \frac{1}{k} \leq \log n + \log n = 2 \log n \end{aligned}$$