ANALYSIS OF INSERTION COSTS IN PRIORITY TREES

MARKUS KUBA AND ALOIS PANHOLZER

Abstract. Priority trees are a data structure used for priority queue administration. Under the model that all permutations of the numbers $1, \ldots, n$ are equally likely to construct a priority tree of size n we give a detailed average-case analysis of insertion cost measures: we study the *recursion depth* and the *number of key comparisons* when inserting an element into a random size-n priority tree. For inserting a random element we obtain exact and asymptotic results for the expectation and the variance and can further show a central limit law of the parameters studied and for inserting an element with specified priority we can show exact and asymptotic results for the expectation of these quantities.

1. Introduction

Priority trees, also called p-trees for short, are a certain data structure that is used for maintaining priority queues. This data structure was used first in the implementation of the programming language Simula in the early 1960's, which was developed at the Norwegian Computer Center in Oslo by O.-J. Dahl and K. Nygaard.

Priority queues are used for many scheduling problems (e. g., for jobs, events, etc.). Typical applications are given in operating systems (scheduling the execution of different jobs, resource management, etc.) and in discrete event simulation models (scheduling events according to their time of occurrence). In a priority queue every element that is stored is associated with a fixed key, whose value determines its priority. We make here the assumption that a lower key value correspond to a higher priority. For the administration of a priority queue the following two operations are fundamental: "inserting an element with arbitrary priority (Insert)", and "removing the element with highest priority (Delete)".

The data structure of priority trees is a special implementation of binary trees with a particular ordering constraint that makes them suitable for maintaining priority queues: a priority tree is either empty or it consists of a sequence of nodes with non-increasing keys, the so called *left path*, such that with each node on the left path except the last one, there is associated a possibly empty priority tree, the so called *right subtree*. All nodes of the right subtree that are associated with a node z on the left path are ranked between z and the left successor of z: if x with key k denotes the left successor of z with key l, then all nodes on the right subtree of z have key values s: $k \leq s < l$. This is illustrated in Figure 1.

In order to insert a new element p into an already constructed priority tree T one uses the algorithm Insert, which is described in the following:

- If T is empty or the root of T has a key not larger than p, then let p be the new root and T its left subtree.
- Otherwise follow the left path of T for the first node x that has a key not larger than the key of p.
  - If no such node exists then append p to the left path as a new left leaf.
  - Otherwise let us denote by z the predecessor of x; thus the key of p is ranked between the keys of x and z. In this case algorithm Insert will be applied recursively to the right subtree of z to insert node p.

This work was supported by the Austrian Science Foundation FWF, grant S9608-N13.

An example of constructing a priority tree with the algorithm INSERT is given in Figure 2.

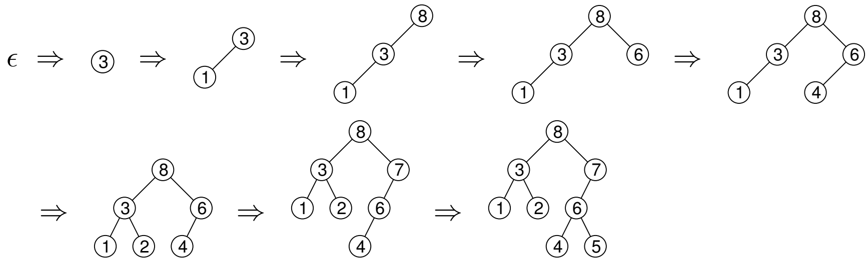


FIGURE 2. Construction of a priority tree of size 8 by successively inserting the keys $3, 1, 8, 6, 4, 2, 7, 5$ starting with an empty tree (denoted by ϵ). Inserting the node with key 5 leads to a recursion depth of 3, where 5 key comparisons are made.

A first average-case analysis of the insertion algorithm INSERT and thus a theoretical study of the efficiency of priority trees was given by Jonassen and Dahl in [6]. For this average-case analysis the so called "random permutation model" was used as the underlying probability model. This means that all $n!$ permutations of $\{1, 2, \ldots, n\}$ are assumed to be chosen equally likely as input data arrays to construct a priority tree of size n starting with the empty tree using the algorithm INSERT repeatedly (in the sequel we will not distinguish between an element and its key).

Under this model the following cost measures are studied, which can be considered as random variables in our probabilistic setting:

- $A_n^{[I]}$: the number of key comparisons that are made when inserting a random element into a random priority tree of size n with the algorithm INSERT. For the model of randomness we assume further that all values of the set $\{\frac{1}{2}, \frac{3}{2}, \ldots, n + \frac{1}{2}\}$ can occur with equal probability $\frac{1}{n+1}$ as the key of the inserted element.
- $A_n^{[R]}$: the recursion depth, i. e., the number of calls of the algorithm INSERT, when inserting a random element into a random priority tree of size n.

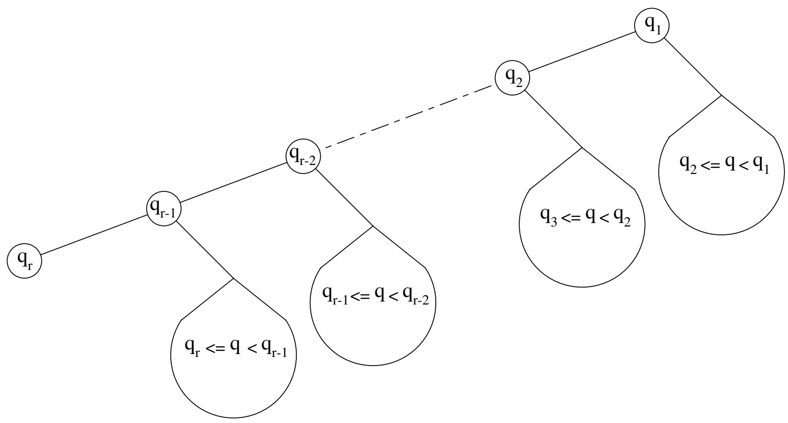


FIGURE 1. The structure of a priority tree. The keys of the elements on the left path satisfy $q_1 \geq q_2 \geq \cdots \geq q_{r-2} \geq q_{r-1} \geq q_r$. If q is the key of an element in the l-th right subtree then it holds that $q_{l+1} \leq q < q_l$.

- $A_n^{[L]}$: the length of the left path of a random priority tree of size n. Since the element with smallest key and thus with highest priority is located at the last node on the left path of the priority tree this quantity is also of interest.

$A_n^{[I]}$ is of prime interest, since it gives the cost of an unsucessful search. The main results of [6] are exact and asymptotic formulæ for the expectation of these quantities, which are given below:

$$\mathbb{E}(A_n^{[I]}) = \frac{H_{n+1}^2}{3} + \frac{10H_{n+1}}{9} - \frac{H_{n+1}^{(2)}}{3} - \frac{28}{27}, \quad \text{for } n \geq 2, \qquad \mathbb{E}(A_n^{[I]}) \sim \frac{1}{3}\log^2 n,$$
$$\mathbb{E}(A_n^{[R]}) = \frac{2}{3}H_{n+1} + \frac{1}{9}, \quad \text{for } n \geq 2, \qquad \mathbb{E}(A_n^{[R]}) \sim \frac{2}{3}\log n,$$
$$\mathbb{E}(A_n^{[L]}) = 2H_n - 1, \quad \text{for } n \geq 1, \qquad \mathbb{E}(A_n^{[L]}) \sim 2\log n,$$

where $H_n := \sum_{k=1}^n \frac{1}{k}$ and $H_n^{(m)} := \sum_{k=1}^n \frac{1}{k^m}$ denote the harmonic numbers and the m-th order harmonic numbers. Since $\mathbb{E}(A_n^{[I]}) \sim \frac{1}{3}\log^2 n$ it follows that the operation Insert is, asymptotically, for priority trees more expensive than for other priority queue structures (like heaps or leftist trees, see, e. g., [3]).

The data structure of priority trees is described in several text books on algorithms and data structures, e. g., in [3], where the results of [6] are cited in parts. Although this data structure has at least some propagation there are very few further theoretical studies of the behaviour of priority trees. One "exception" is the work of Nevalainen and Teuhola [8], where the authors analysed the behaviour of priority trees under so called Hold-operation sequences. A further extension of the work of [6] for random priority trees was given in [9], where the main focus was a study of certain tree parameters for specified nodes (= elements) rather than random nodes. In particular the authors studied the random variable $A_{n,j}^{[I]}$, which counts the number of key comparisons when inserting an element with key $j+\frac{1}{2}$, $0 \leq j \leq n$, i. e., inserting an element with a key between the j-th and $j+1$-th smallest key in the tree, into a random size-n priority tree. They obtained exact and asymptotic formulæ for the expectation of $A_{n,j}^{[I]}$. Moreover, the length of the left path $A_n^{[L]}$ was studied in detail and it was shown that, after normalization, this quantity is asymptotically normal distributed with mean $\mathbb{E}(A_n^{[L]}) \sim 2\log n$ and variance $\mathbb{V}(A_n^{[L]}) \sim 2\log n$.

However, it seems that in the literature there do not exist further results for the parameters $A_n^{[I]}$ and $A_n^{[R]}$, the number of key comparisons and the recursion depth when inserting a random element into a random size-n priority tree. It was the main motivation for the present work to give a thorough theoretical analysis of the behaviour of these quantities leading to exact and asymptotic formulæ for the variance and to a characterization of the limiting distribution. Furthermore we will also study the random variable $A_{n,j}^{[R]}$, which gives the recursion depth when inserting an element with key $j+\frac{1}{2}$, $0 \leq j \leq n$, into a random size-n priority tree. We obtain exact and asymptotic formulæ for the expectation of $A_{n,j}^{[R]}$. The results obtained for these quantities are given in Section 2.

2. Results for the insertion costs

2.1. **Costs for inserting a random element.**

Theorem 1. *The expectation and the variance of the number of key comparisons $A_n^{[I]}$ when inserting a random element into a random size-n priority tree with the algorithm* Insert *are given exact and asymptotically by the following formulæ.*

$$\mathbb{E}(A_n^{[I]}) = \frac{1}{3}H_{n+1}^2 + \frac{10}{9}H_{n+1} - \frac{1}{3}H_{n+1}^{(2)} - \frac{28}{27}, \quad n \geq 2, \quad \mathbb{E}(A_0^{[I]}) = 0, \quad \mathbb{E}(A_1^{[I]}) = 1,$$

$$\mathbb{V}(A_n^{[I]}) = \frac{10}{81}H_{n+1}^3 + \frac{85}{81}H_{n+1}^2 - \frac{4}{9}H_{n+1}^2 H_{n+1}^{(2)} + \frac{602}{243}H_{n+1} - \frac{50}{27}H_{n+1}H_{n+1}^{(2)} + \frac{8}{9}H_{n+1}H_{n+1}^{(3)}$$
$$- \frac{8}{81(n+1)n(n-1)}H_{n+1} + \frac{2}{9}\big(H_{n+1}^{(2)}\big)^2 - \frac{185}{81}H_{n+1}^{(2)} + \frac{140}{81}H_{n+1}^{(3)} - \frac{2}{3}H_{n+1}^{(4)}$$
$$- \frac{752}{243} + \frac{476}{243(n+1)} + \frac{20}{243n} - \frac{10}{243(n-1)} + \frac{4}{81(n+1)^2} - \frac{8}{81n^2}$$
$$+ \frac{4}{81(n-1)^2}, \quad \text{for } n \geq 2, \quad \mathbb{V}(A_0^{[I]}) = 0, \quad \mathbb{V}(A_1^{[I]}) = 0,$$
$$\mathbb{E}(A_n^{[I]}) = \frac{1}{3}\log^2 n + \big(\frac{10}{9} + \frac{2\gamma}{3}\big)\log n - \frac{28}{27} + \frac{\gamma^2}{3} + \frac{10\gamma}{9} - \frac{\pi^2}{18} + \mathcal{O}\big(\frac{\log n}{n}\big),$$
$$\mathbb{V}(A_n^{[I]}) = \frac{10}{81}\log^3 n + \big(\frac{85}{81} + \frac{10\gamma}{27} - \frac{2\pi^2}{27}\big)\log^2 n + \mathcal{O}(\log n).$$

Moreover, after centering and scaling, $A_n^{[I]}$ converges in distribution to a standard normal distributed random variable:

$$\mathbb{P}\left\{\frac{A_n^{[I]} - \frac{1}{3}\log^2 n}{\sqrt{\frac{10}{81}\log^3 n}} \leq x\right\} \to \Phi(x), \quad \text{for } x \in \mathbb{R}.$$

Here $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$ denotes the distribution function of the standard normal distribution $\mathcal{N}(0,1)$.

Theorem 2. *The expectation and the variance of the recursion depth $A_n^{[R]}$ when inserting a random element into a random size-n priority tree with the algorithm* Insert *are given exact and asymptotically by the following formulæ.*

$$\mathbb{E}(A_n^{[R]}) = \frac{2}{3}H_{n+1} + \frac{1}{9}, \quad \text{for } n \geq 2, \quad \mathbb{E}(A_0^{[R]}) = 1, \quad \mathbb{E}(A_1^{[R]}) = 1,$$
$$\mathbb{V}(A_n^{[R]}) = \frac{10}{27}H_{n+1} - \frac{4}{9}H_{n+1}^{(2)} + \frac{8}{81} + \frac{8}{27(n+1)n(n-1)}, \quad \text{for } n \geq 2,$$
$$\mathbb{V}(A_0^{[R]}) = 0, \quad \mathbb{V}(A_1^{[R]}) = 0,$$
$$\mathbb{E}(A_n^{[R]}) = \frac{2}{3}\log n + \frac{1}{9} + \frac{2\gamma}{3} + \mathcal{O}(n^{-1}), \quad \mathbb{V}(A_n^{[R]}) = \frac{10}{27}\log n + \frac{8}{81} + \frac{10\gamma}{27} - \frac{2\pi^2}{27} + \mathcal{O}(n^{-1}).$$

Moreover, after centering and scaling, $A_n^{[R]}$ converges in distribution to a standard normal distributed random variable, where the rate of convergence is of order $\mathcal{O}\big(\frac{1}{\sqrt{\log n}}\big)$:

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{A_n^{[R]} - \frac{2}{3}\log n}{\sqrt{\frac{10}{27}\log n}} \leq x\right\} - \Phi(x)\right| = \mathcal{O}\Big(\frac{1}{\sqrt{\log n}}\Big), \quad \text{for } x \in \mathbb{R}.$$

2.2. **Costs for inserting a specified element.**

Theorem 3. [9] *The expectation of the number of key comparisons $A_{n,j}^{[I]}$ when inserting the element $j + \frac{1}{2}$, with $0 \leq j \leq n$, into a random size-n priority tree with the algorithm* Insert *are given exact and asymptotically by the following formulæ.*

$$\mathbb{E}(A_{n,j}^{[I]}) = \frac{1}{3}H_{n-j}^2 - \big(\frac{1}{3} + \frac{2}{3j}\big)H_n + \big(\frac{1}{3} + \frac{2}{3j}\big)H_j + \big(\frac{7}{3} + \frac{2}{3j} + \frac{2j^2}{3n} - \frac{2j(j-1)}{3(n-1)}\big)H_{n-j}$$
$$- \frac{1}{3}H_{n-j}^{(2)} - \frac{j(2j+5)}{3n} + \frac{2(j^2+j-1)}{3(n-1)} + \frac{1}{3(n+1-j)} + \frac{1}{3(n-j)} - \frac{1}{3(j+1)}$$

$$- \frac{1}{3j} - \frac{2}{3} \sum_{k=j+1}^{n} \frac{H_{k-j}}{k}, \quad \text{for } 1 \le j \le n-1,$$

$$\mathbb{E}(A_{n,n}^{[I]}) = 1, \quad \text{for } n \ge 1, \quad \mathbb{E}(A_{0,0}^{[I]}) = 0, \quad \mathbb{E}(A_{n,0}^{[I]}) = 2H_n - 1, \quad \text{for } n \ge 1.$$

$$\mathbb{E}(A_{n,j}^{[I]}) = 2 \log n + \mathcal{O}(1), \quad \text{for } j \ge 1 \text{ fixed},$$

$$\mathbb{E}(A_{n,j}^{[I]}) = \frac{1}{3} H_l^2 + \frac{7}{3} H_l - \frac{1}{3} H_l^{(2)} - \frac{1}{3} + \frac{1}{3l} + \frac{1}{3(l+1)} + \mathcal{O}(n^{-1}), \quad \text{for } l := n-j \ge 1 \text{ fixed},$$

$$\mathbb{E}(A_{n,j}^{[I]}) = \frac{1}{3} \log^2 n + \big(\frac{2}{3} \log \rho + \frac{2}{3} \log(1-\rho) + \frac{2\gamma}{3} + \frac{7}{3} + \frac{2}{3} \rho(1-\rho)\big) \log n + \mathcal{O}(1), \quad \text{for } j = \rho n, \ 0 < \rho < 1.$$

Theorem 4. *The expectation of the recursion depth $A_{n,j}^{[R]}$ when inserting the element $j + \frac{1}{2}$, with $0 \le j \le n$, into a random size-n priority tree with the algorithm* Insert *are given exact and asymptotically by the following formulæ.*

$$\mathbb{E}(A_{n,j}^{[R]}) = \frac{2}{3} H_j + \frac{2}{3} H_{n-j} - \frac{2}{3} H_n + \frac{4}{3} + \frac{2j}{3n} - \frac{2j(j-1)}{3n(n-1)}, \quad \text{for } 1 \le j \le n-1,$$

$$\mathbb{E}(A_{n,n}^{[R]}) = 1, \quad \text{for } n \ge 0, \quad \mathbb{E}(A_{n,0}^{[I]}) = 1, \quad \text{for } n \ge 1.$$

$$\mathbb{E}(A_{n,j}^{[R]}) = \frac{2}{3} H_j + \frac{4}{3} + \mathcal{O}(n^{-1}), \quad \text{for } j \ge 1 \text{ fixed},$$

$$\mathbb{E}(A_{n,j}^{[R]}) = \frac{2}{3} H_l + \frac{4}{3} + \mathcal{O}(n^{-1}), \quad \text{for } l := n-j \ge 1 \text{ fixed},$$

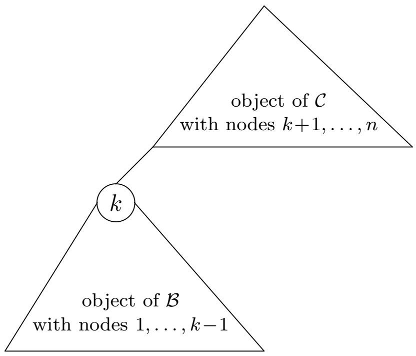
$$\mathbb{E}(A_{n,j}^{[R]}) = \frac{2}{3} \log n + \frac{2}{3} \log \rho + \frac{2}{3} \log(1-\rho) + \frac{2\gamma}{3} + \frac{4}{3} + \frac{2}{3} \rho(1-\rho) + O(n^{-1}), \quad \text{for } j = \rho n, \ 0 < \rho < 1.$$

3. Mathematical preliminaries

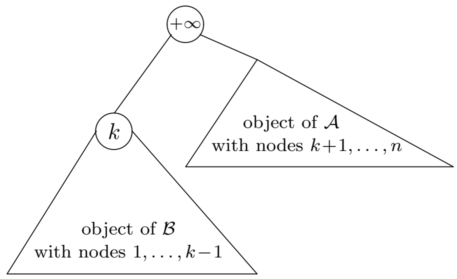
The mathematical analysis of the insertion costs presented here is based on the following description of priority trees, which was introduced in [9]. We use three families of combinatorial objects (= tree families) denoted by $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$. The family $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots$ is the family of ordinary priority trees: the trees in $\mathcal{A}_n$ are generated by successively inserting the elements of a random permutation of $\{1, \dots, n\}$ with the algorithm Insert starting with an empty tree. The auxiliary families $\mathcal{B} = \mathcal{B}_0 \cup \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots$ and $\mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots$ are defined as follows: the trees in $\mathcal{B}_n$ resp. $\mathcal{C}_n$ are generated by successively inserting the elements of a random permutation of $\{1, \dots, n\}$ with the algorithm Insert starting with the additional element "$+\infty$" (i. e., an element that has a larger key than every other element) resp. starting with the additional element "$-\infty$" (i. e., an element that has a smaller key than every other element). We always assume that these additional elements are not counted for the size of an object, which means that the size of a tree is given by the number of nodes without elements $-\infty$ or $+\infty$.

In our description of the parameters studied we will heavily use the decomposition of the families $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ according to the first element k of a random permutation of $\{1, 2, \dots, n\}$. These fundamental combinatorial decompositions are given in Figure 3.

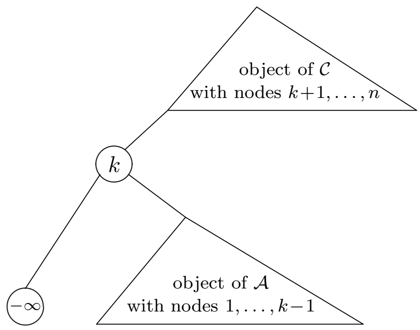
The combinatorial decompositions can be translated directly into distribution recurrences for the random variables $A_{n,j}$, $B_{n,j}$, and $C_{n,j}$, where $A_{n,j}$ is either $A_{n,j}^{[R]}$ or $A_{n,j}^{[I]}$, and $B_{n,j}$ and $C_{n,j}$ are the corresponding random variables for the objects of $\mathcal{B}$ and $\mathcal{C}$. In order to obtain results for random elements, i. e., for the random variables A_n, which either denote



Decomposition of the family $\mathcal{A}$



Decomposition of the family $\mathcal{B}$



Decomposition of the family $\mathcal{C}$

FIGURE 3. Decomposition of the families $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ according to the first element.

$A_n^{[R]}$ or $A_n^{[I]}$, we use the following distribution equation ($\stackrel{(d)}{=}$ denotes equality in distribution of two random variables):

$$A_n \stackrel{(d)}{=} \sum_{j=1}^{n} I_{n,j}\, A_{n,j},$$

where $I_{n,j}$ is the indicator variable of the event $U_n = j$, where U_n is uniformly distributed on the set $\{0, 1, \ldots, n\}$. Analogous equations hold for B_n and C_n, too. This leads in particular to equations $\mathbb{P}\{A_n = m\} = \frac{1}{n+1} \sum_{j=0}^{n} \mathbb{P}\{A_{n,j} = m\}$, etc.

To treat the distribution recurrences for the random variables studied we will use a generating functions approach. A sketch of the proof of the results for the recursion depth when inserting an element into a random priority tree is given in Section 4, whereas the proof of the corresponding results for the number of key comparisons is sketched in Section 5.

For our recursive description of the parameter $A_{n,j}^{[I]}$, and thus also for $A_n^{[I]}$, we require results for $C_n^{[L]}$ obtained in [9], i. e., results for the length of the left path (= the number of nodes lying on the path from the root to node $-\infty$, where we use the convention that the node $-\infty$ itself is not counted) in a random size-n tree of the family $\mathcal{C}$. We obtain that $C_n^{[L]}$ is distributed exactly like the number of cycles, or equivalently the number of left-to-right maxima, in a random permutation of length n. In the following we use the abbreviation $p_{n,m} := \mathbb{P}\{C_n^{[L]} = m\}$ and $p(z,v) := \sum_{n\ge 0} \sum_{m \ge 0} p_{n,m} z^n v^m$. Then equation (1) holds, where $\left[{n \atop m}\right]$ denotes the signless Stirling numbers of the first kind (see, e. g., [4]).

$$p_{n,m} = \frac{\left[{n \atop m}\right]}{n!}, \quad \text{and} \quad p(z,v) = \frac{1}{(1-z)^v}. \tag{1}$$

4. The recursion depth

4.1. **Inserting a specified element.** We consider the quantities $A_{n,j}^{[R]}$, which count the recursion depth (= the number of calls of the algorithm Insert) when inserting the element $j+\frac{1}{2}$, with $0 \le j \le n$, into a random size-n priority tree (i. e., a random size-n tree of the family $\mathcal{A}$). Analogously we define the quantities $B_{n,j}^{[R]}$ and $C_{n,j}^{[R]}$ as the recursion depth when inserting the element $j+\frac{1}{2}$, with $0 \le j \le n$, into a random size-n tree of the families $\mathcal{B}$ and $\mathcal{C}$.

A direct translation of the combinatorial decomposition of the objects $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ as given in Section 3 leads to the following distribution recurrences, which hold for $n \ge 1$ and $0 \le j \le n$, and initial values $\mathbb{P}\{A_{0,0}^{[R]} = 1\} = \mathbb{P}\{B_{0,0}^{[R]} = 1\} = \mathbb{P}\{C_{0,0}^{[R]} = 1\} = 1$:

$$\mathbb{P}\{A_{n,j}^{[R]} = m\} = \frac{1}{n}\Big(\sum_{k=1}^{j} \mathbb{P}\{C_{n-k,j-k}^{[R]} = m\} + \sum_{k=j+1}^{n} \mathbb{P}\{B_{k-1,j}^{[R]} = m\}\Big), \tag{2a}$$

$$\mathbb{P}\{B_{n,j}^{[R]} = m\} = \frac{1}{n}\Big(\sum_{k=1}^{j} \mathbb{P}\{A_{n-k,j-k}^{[R]} = m-1\} + \sum_{k=j+1}^{n} \mathbb{P}\{B_{k-1,j}^{[R]} = m\}\Big), \tag{2b}$$

$$\mathbb{P}\{C_{n,j}^{[R]} = m\} = \frac{1}{n}\Big(\sum_{k=1}^{j} \mathbb{P}\{C_{n-k,j-k}^{[R]} = m\} + \sum_{k=j+1}^{n} \mathbb{P}\{A_{k-1,j}^{[R]} = m-1\}\Big). \tag{2c}$$

Introducing the generating functions $A(z,u,v) := \sum_{n\ge 0}\sum_{0\le j\le n}\sum_{m\ge 0}\mathbb{P}\{A_{n,j}^{[R]} = m\}z^n u^j v^m$, etc., and multiplying the system of recurrences (2) with $nz^{n-1}u^j v^m$ and summing up for $n \ge 1$, $0 \le j \le n$, and $m \ge 0$, we can translate the recurrences into the following system of linear differential equations:

$$\frac{\partial}{\partial z}A(z,u,v) = \frac{u}{1-uz}C(z,u,v) + \frac{1}{1-z}B(z,u,v), \tag{3a}$$

$$\frac{\partial}{\partial z}B(z,u,v) = \frac{uv}{1-uz}A(z,u,v) + \frac{1}{1-z}B(z,u,v), \tag{3b}$$

$$\frac{\partial}{\partial z}C(z,u,v) = \frac{u}{1-uz}C(z,u,v) + \frac{v}{1-z}A(z,u,v), \tag{3c}$$

with the initial conditions $A(0,u,v) = B(0,u,v) = C(0,u,v) = v$.

One can successively eliminate the auxiliary functions $B(z,u,v)$ and $C(z,u,v)$ from the system (3) and obtains a third order homogeneous linear differential equation for the function $A(z,u,v)$ of interest:

$$\frac{\partial^3}{\partial z^3}A(z,u,v) + \frac{3(2uz-u-1)}{(1-z)(1-uz)}\frac{\partial^2}{\partial z^2}A(z,u,v) + \frac{2u(3-v)}{(1-z)(1-uz)}\frac{\partial}{\partial z}A(z,u,v) + \frac{2uv(u+1-2uz)}{(1-z)^2(1-uz)^2}A(z,u,v) = 0. \tag{4}$$

The initial conditions $A(0,u,v) = v$, $\frac{\partial}{\partial z}A(z,u,v)\big|_{z=0} = v+uv$, $\frac{\partial^2}{\partial z^2}A(z,u,v)\Big|_{z=0} = 2(v+uv^2+u^2v)$ can be obtained easily when considering trees with sizes 0, 1, and 2. For a study of the expectation as given in Theorem 4 we introduce the generating function $E(z,u) := \frac{\partial}{\partial v}A(z,u,v)\big|_{v=1}$, differentiate (4) w. r. t. v and evaluate at $v=1$. This leads to a homogeneous third order linear differential equation for $E(z,u)$, which can be solved explicitly (e.g., by reducing to a hypergeometric differential equation). Extracting coefficients by using well-known formulæ leads then to Theorem 4.

4.2. **Inserting a random element.** We study now the random variables $A_n^{[R]}$, which count the recursion depth when inserting a random element into a random size-n priority tree. Analogously we define the random variables $B_n^{[R]}$ and $C_n^{[R]}$ for trees of the families $\mathcal{B}$ and $\mathcal{C}$. Again we introduce generating functions $A(z,v) := A(z,1,v) = \sum_{n\ge 0}\sum_{m\ge 0}(n+1)\mathbb{P}\{A_n^{(R)} = m\}z^n v^m$, etc. From the system of differential equations (3) we obtain, after evaluating at $v=1$ and eliminating the functions $B(z,v)$ and $C(z,v)$, for $A(z,v)$ a second order linear differential equation of Cauchy-Euler type, which has the solution

$$A(z,v) = \frac{v(3+\sqrt{1+8v})}{2\sqrt{1+8v}}(1-z)^{-\frac{1}{2}-\frac{\sqrt{1+8v}}{2}} + \frac{v(-3+\sqrt{1+8v})}{2\sqrt{1+8v}}(1-z)^{-\frac{1}{2}+\frac{\sqrt{1+8v}}{2}}. \tag{5}$$

From this explicit solution of $A(z,v)$ one directly obtains closed formulæ for the moments of $A_n^{[R]}$. Thus the first part of Theorem 2 follows after differentiating w. r. t. v, evaluating at $v=1$ and extracting coefficients.

To show the central limit theorem for $A_n^{[R]}$ we will apply basic singularity analysis [1] to the explicit formula (5), which directly leads to the following expansion of the moment generating function of $A_n^{[R]}$ that holds uniformly in a complex neighbourhood of $s=0$:

$$\mathbb{E}(e^{A_n^{[R]}s}) = \frac{1}{n+1}[z^n]A(z,e^s) = e^{U(s)\log n + V(s)}\big(1+\mathcal{O}(n^{-1+\epsilon})\big), \tag{6}$$

with an arbitrary small $\epsilon > 0$, $U(s) = \frac{\sqrt{1+8e^s}-3}{2}$ and a certain analytic function $V(s)$. The central limit theorem stated in Theorem 2 follows now from (6) by an immediate application of the quasi power theorem due to Hwang (see [5]).

5. Number of key comparisons

5.1. **Inserting a specified element.** We consider the quantities $A_{n,j}^{[I]}$, which count the number of key comparisons when inserting the element $j+\frac{1}{2}$, with $0\le j\le n$, into a random size-n priority tree (i. e., a random size-n tree of the family $\mathcal{A}$). Analogously we define the quantities $B_{n,j}^{[I]}$ and $C_{n,j}^{[I]}$ as the number of key comparisons when inserting the element $j+\frac{1}{2}$, with $0\le j\le n$, into a random size-n tree of the families $\mathcal{B}$ and $\mathcal{C}$.

A direct translation of the combinatorial decomposition of the objects $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ as given in Section 3 leads to the following distribution recurrences, which hold for $n\ge 1$ and $0\le j\le n$, with the initial values $\mathbb{P}\{A_{0,0}^{[I]} = 0\} = 1$, $\mathbb{P}\{B_{0,0}^{[I]} = 1\} = \mathbb{P}\{C_{0,0}^{[I]} = 1\} = 1$, and $p_{n,k}$ defined by (1):

$$\mathbb{P}\{A_{n,j}^{[I]} = m\} = \frac{1}{n}\Big(\sum_{k=1}^{j}\mathbb{P}\{C_{n-k,j-k}^{[I]} = m\} + \sum_{k=j+1}^{n}\sum_{i=0}^{m}p_{n-k,i}\mathbb{P}\{B_{k-1,j}^{[I]} = m-i\}\Big), \tag{7a}$$

$$\mathbb{P}\{B_{n,j}^{[I]} = m\} = \frac{1}{n}\Big(\sum_{k=1}^{j}\mathbb{P}\{A_{n-k,j-k}^{[I]} = m-2\} + \sum_{k=j+1}^{n}\mathbb{P}\{B_{k-1,j}^{[I]} = m-1\}\Big), \tag{7b}$$

$$\mathbb{P}\{C_{n,j}^{[I]} = m\} = \frac{1}{n}\Big(\sum_{k=1}^{j}\mathbb{P}\{C_{n-k,j-k}^{[I]} = m\} + \sum_{k=j+1}^{n}\sum_{i=0}^{m-2}p_{n-k,i}\mathbb{P}\{A_{k-1,j}^{[I]} = m-i-2\}\Big). \tag{7c}$$

Introducing the generating functions $A(z,u,v) := \sum_{n\ge 0}\sum_{0\le j\le n}\sum_{m\ge 0}\mathbb{P}\{A_{n,j}^{[I]} = m\}z^n u^j v^m$, etc., and multiplying the system of recurrences (7) with $nz^{n-1}u^j v^m$ and summing up for $n\ge 1$, $0\le j\le n$, and $m\ge 0$, we can translate the recurrences into the following system of linear differential equations:

$$\frac{\partial}{\partial z}A(z,u,v) = \frac{u}{1-uz}C(z,u,v) + \frac{1}{(1-z)^v}B(z,u,v), \tag{8a}$$

$$\frac{\partial}{\partial z}B(z,u,v) = \frac{uv^2}{1-uz}A(z,u,v) + \frac{v}{1-z}B(z,u,v), \tag{8b}$$

$$\frac{\partial}{\partial z}C(z,u,v) = \frac{u}{1-uz}C(z,u,v) + \frac{v^2}{(1-z)^v}A(z,u,v), \tag{8c}$$

with the initial conditions $A(0,u,v)=1$ and $B(0,u,v)=C(0,u,v)=v$.

In order to study the expectation of the quantity $A_{n,j}^{[I]}$ one can introduce the generating functions $E_A(z,u) := \frac{\partial}{\partial v}A(z,u,v)\big|_{v=1}$, etc., and obtains after differentiating (8) w. r. t. v and evaluating at $v=1$ a system of linear differential equations, which can be solved explicitly. This eventually leads to the results of Theorem 3.

5.2. **Inserting a random element.** Now we study the random variables $A_n^{[I]}$, which count the number of key comparisons when inserting a random element into a random size-n priority tree. Analogously we define the random variables $B_n^{[I]}$ and $C_n^{[I]}$ for trees of the families $\mathcal{B}$ and $\mathcal{C}$. We introduce the generating functions $A(z,v) := A(z,1,v) = \sum_{n\geq 0}\sum_{m\geq 0}(n+1)\mathbb{P}\{A_n^{[I]} = m\}z^n v^m$, etc., and obtain after evaluating (8) at $v=1$ the following system of differential equations:

$$\frac{\partial}{\partial z}A(z,v) = \frac{1}{1-z}C(z,v) + \frac{1}{(1-z)^v}B(z,v), \tag{9}$$

$$\frac{\partial}{\partial z}B(z,v) = \frac{v^2}{1-z}A(z,v) + \frac{v}{1-z}B(z,v), \quad \frac{\partial}{\partial z}C(z,v) = \frac{1}{1-z}C(z,v) + \frac{v^2}{(1-z)^v}A(z,v),$$

with initial conditions $A(0,v)=1$, $B(0,v)=C(0,v)=v$. Eliminating the functions $B(z,v)$ and $C(z,v)$ of (9) leads to the following third order homogeneous linear differential equation for the function $A(z,v)$:

$$\frac{\partial^3}{\partial z^3}A(z,v) - \frac{2v+3}{1-z}\frac{\partial^2}{\partial z^2}A(z,v) + 2v\Big(\frac{2}{(1-z)^2} - \frac{v}{(1-z)^{v+1}}\Big)\frac{\partial}{\partial z}A(z,v) + \frac{2v^2}{(1-z)^{v+2}}A(z,v) = 0, \tag{10}$$

with initial conditions $A(0,v)=1$, $\frac{\partial}{\partial z}A(z,v)\big|_{z=0} = 2v$, and $\frac{\partial^2}{\partial z^2}A(z,v)\Big|_{z=0} = 2v+4v^2$.

In order to study the moments of $A_n^{[I]}$ we introduce the generating functions $A_r(z) := \frac{\partial^r}{\partial v^r}A(z,v)\big|_{v=1} = \sum_{n\geq 0}(n+1)\mathbb{E}\Big(\big(A_n^{[I]}\big)^{\underline{r}}\Big)z^n$. Differentiating equation (10) once and twice w. r. t. v and evaluating at $v=1$ leads to third order linear differential equations of Cauchy-Euler type for $A_1(z)$ and $A_2(z)$, which can be solved explicitly. This gives after extracting coefficients the explicit and asymptotic results for the expectation and the variance of $A_n^{[I]}$ stated in Theorem 1.

In order to show the central limit theorem for the quantity $A_n^{[I]}$ we can use the method described in [10], which has been applied there to a study of the depth of a random node in a random priority tree; a generating functions approach for the latter parameter leads to similar (but not the same) differential equations as appearing here. Basically we use the method of moments, but in order to control the huge cancellations when considering the centered moments, we are forced to give a very detailed description of the coefficients appearing in the asymptotic expansion of the r-th factorial moments of $A_n^{[I]}$. We will only state the "key steps" leading to the result.

- The generating functions $A_r(z)$ have the following local expansion around the dominant singularity $z=1$:

$$A_r(z) = \sum_{m=0}^{2r} a_r^{(m)} \frac{1}{(1-z)^2}\log^{2r-m}\big(\frac{1}{1-z}\big) + \mathcal{O}\big(\frac{1}{1-z}\big),$$

where the coefficients $a_r^{(m)}$ are, for all $m \geq 0$ and $r \geq 0$, given as follows:

$$a_r^{(m)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \Big(\frac{\left(\frac{5}{9}\right)^m}{m!} r^{2m-\lfloor \frac{m+1}{2} \rfloor} + \sum_{l=1}^{2m-\lfloor \frac{m+1}{2} \rfloor} c_l(m)\, r^{2m-\lfloor \frac{m+1}{2} \rfloor - l} \Big),$$

with certain functions $c_1(m)$, $c_2(m)$, ... Here $x^{\underline{m}} := x(x-1)\cdots(x-m+1)$ denote the falling factorials and $\lfloor x \rfloor$ the floor function (see, e. g., [4]).

- The r-th factorial moments of $A_n^{[I]}$ have the following asymptotic expansion:

$$\mathbb{E}\left((A_n^{[I]})^{\underline{r}} \right) = \sum_{m=0}^{2r} (\log n)^{2r-m} \sum_{j=0}^{m} a_r^{(j)} \binom{2r-j}{m-j} \kappa_{m-j} + \mathcal{O}\big(\frac{1}{n^{1-\epsilon}}\big),$$

with constants $\kappa_k := (-1)^k \, \frac{d^k}{ds^k} \left(\frac{1}{\Gamma(-s)} \right) \Big|_{s=-2}$, and $a_r^{(j)}$ as described above.

- The r-th centralized and normalized moments of $A_n^{[I]}$ have the following asymptotic expansions:

$$\mathbb{E}\left(\left(\frac{A_n^{[I]} - \frac{1}{3}\log^2 n}{\sqrt{\frac{10}{81}\log^3 n}} \right)^r \right) = \begin{cases} \frac{(2m)!}{2^m m!} + \mathcal{O}\big(\frac{1}{\log n}\big), & \text{if } r = 2m \text{ even}, \\ \mathcal{O}\big(\frac{1}{\sqrt{\log n}}\big), & \text{if } r \text{ odd}. \end{cases}$$

- The Theorem of Fréchet and Shohat (see, e. g., [7]) shows then the central limit theorem stated in Theorem 1.

References

[1] P. Flajolet and A. Odlyzko, Singularity analysis of generating functions, *SIAM Journal on Discrete Mathematics* 3, 216–240, 1990.

[2] D. H. Greene and D. E. Knuth, *Mathematics for the Analysis of Algorithms*, second edition, Birkhäuser, Boston, 1982.

[3] G. Gonnet and R. Baeza-Yates, *Handbook of algorithms and data structures*, Second Edition, Addison-Wesley, Wokingham, 1991.

[4] R. Graham, D. Knuth and O. Patashnik, *Concrete Mathematics*, Second Edition, Addison-Wesley, Reading, 1994.

[5] H.-K. Hwang, On convergence rates in the central limit theorems for combinatorial structures, *European Journal of Combinatorics* 19, 329–343, 1998.

[6] A. Jonassen and O.-J. Dahl, Analysis of an algorithm for priority queue administration. *BIT* 15, 409–422, 1975.

[7] M. Loève, *Probability Theory I*, 4th Edition, Springer-Verlag, New York, 1977.

[8] O. Nevalainen and J. Teuhola, The efficiency of two indexed priority queue algorithms, *BIT* 18, 320–333, 1978.

[9] A. Panholzer and H. Prodinger, Average-case analysis of priority trees: a structure for priority queue administration, *Algorithmica* 22, 600–630, 1998.

[10] A. Panholzer, Analysis of some parameters for random nodes in priority trees, manuscript, 2005. available at `http://info.tuwien.ac.at/panholzer/ptreerev.pdf`

[11] D. A. Zave, A series expansion involving the harmonic numbers, *Information Processing Letters* 5, 75–77, 1976.

Markus Kuba, Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstr. 8-10/104, 1040 Wien, Austria

E-mail address: `markus.kuba@tuwien.ac.at`

Alois Panholzer, Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstr. 8-10/104, 1040 Wien, Austria

E-mail address: `Alois.Panholzer@tuwien.ac.at`