國立交通大學

應用數學系

碩 士 論 文

一種新的動差法與其應用

A new variant of the method of moments with
applications

研 究 生：陳哲皓

指導教授：符麥克　教授

中 華 民 國 九 十 八 年 六 月

一種新的動差法與其應用

A new variant of the method of moments with applications

研 究 生：陳哲皓　　　　**Student：Che-Hao Chen**

指導教授：符麥克　　　　**Advisor：Michael Fuchs**

國 立 交 通 大 學

應 用 數 學 系

碩 士 論 文

A Thesis

Submitted to Department of Applied Mathematics

College of Science

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master

in

Applied Mathematics

June 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年六月

一種新的動差法與其應用

研究生：陳哲皓　　　指導老師：符麥克教授

國 立 交 通 大 學

應 用 數 學 系

序

　　演算分析在計算科學裡是一個重要的工作，這可以幫我們了解這些演算法適當的用途。在這篇論文中，我們將會用機率的理論來了解一個演算法所須要運行的時間。為了使這樣機率的方法可以運作，首先我們考慮演算法的輸入為一個適當的隨機模型，然後讓運行的時間成為隨機變數，這樣我們就可以嘗試算出它的期望值和變異數，進而去了解取極限後的行為。在最近幾年，method of moments 已經變成了解決這些問題的標準工具。

　　在論文中，我們對一個叫優先樹的資料結構和它們的分析感到興趣，也會使用 method of moments 來簡化最近一些優先樹結果的證明。然而，在使用這個方法的過程中，我們會遇到一些困難，此時，我們將會介紹一個新的 method of moments 來解決這些問題。

以下是我們的論文概述，在第一章，我們將會介紹 method of moments，並且也會介紹優先樹和關於這資料結構近幾年的結果。第二章，我們會用一個新的 method of moments 重新證明第一章所提到的一個結果。第三章，我們再一次使用這方法來解決一個比較複雜的問題。最後在第四章會給一個結論。

A new variant of the method of moments with applications

Student：Che-Hao Chen　　　**Advisor**：Michael Fuchs

Department of Applied Mathematics

National Chiao Tung University

Preface

Analyzing algorithms in order to understand their usefulness and appropriateness is a key task in computer science. In this thesis, we will use probability theory to analyze the running time of algorithms. In order to perform such a probabilistic analysis, we will first fix a suitable random model for the input. Then, the running time becomes a random variable and properties such as the mean value and the variance are sought. Once the latter properties are understood, one also wants to clarify the limit behavior. In recent years, the method of moments has become a standard tool for this purpose.

In this thesis, we are interested in a data structure, called priority trees, and their analysis. One of the goals of the thesis is to simplify the proofs of some recent results on priority trees by using the method of moments combined with

asymptotic transfers.

However, in doing so, some new problems will arise which we will overcome by introducing a new variant of the method of moments.

We give a short outline of the thesis. In Chapter 1, we are going to introduce the method of moments. Moreover, we will introduce priority trees and explain some recent results concerning this data structure. Then, in Chapter 2, we will introduce our new method and re-prove some of the results mentioned in Chapter 1. In Chapter 3, we will apply our new approach to a more complex example. We will end the thesis with a short summary in Chapter 4.
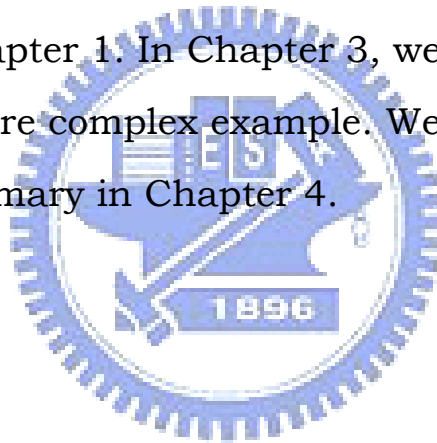
誌　　　　謝

　　首先，感謝我的指導教授符麥克老師在研究所時，不論在課業和未來的方向，都教導我非常多。也感謝所有口試委員，在口試時，給了我很多寶貴的意見。

　　也要謝謝所有在研究所裡的好朋友，不論是１３２的室友和桌球上認識的朋友，和你們在這二年的相處真的很非開心。

　　最後要感謝我的家人，讓我在念書的過程中，能夠沒有後顧之憂專心在學業上。還要感謝所有在研究所裡，有幫助過我的所人好朋友們。

Contents

Chapter 1

Introduction

1.1 Method of moments

In basic probability theory, when proving that a sequence of random variables satisfies a central limit theorem, we usually restrict to sequences that are independent and identically distributed (i.i.d.) because in this situation the proof is simpler. The usual proof then uses characteristic functions and is based on Taylor series expansion and Lévy's continuity theorem. However, for sequences which do not satisfy the i.i.d. assumption, characteristic functions might be intractable. Then, other methods are sought. One such method, the so called "method of moments", proceeds by calculating the limits of the moments. If the resulting sequence uniquely determines a distribution, then weak convergence to this distribution is established. The theoretical basis of this method is Theorem 30.2 in P. Billingsley [1].

Theorem 1. *Suppose that the distribution of X is determined by its moments, that the X_n have moments of all orders, and that $\lim_n \mathbb{E}(X_n^r) = \mathbb{E}(X^r)$ for $r = 1, 2, \ldots$. Then X_n converges in distribution to X.*

One example of a distribution which is uniquely determined by its moments is the standard normal distribution $N(0, 1)$ and the r-th moment is equal to

$$\mathbb{E}\left(N(0,1)^r\right) = \begin{cases} \frac{(2m)!}{2^m m!}, & \text{if } r = 2m, \\ 0, & \text{if } r = 2m + 1. \end{cases}$$

Therefore, by the above theorem, if we want to prove that a sequence of random variables, say $\{X_n\}_{n\geq 0}$, converges in distribution to the standard normal distribution, we only have to show that for all $r \geq 1$,

$$\lim_{n\to\infty} \mathbb{E}\left(X_n^r\right) = \begin{cases} \frac{(2m)!}{2^m m!}, & \text{if } r = 2m, \\ 0, & \text{if } r = 2m + 1. \end{cases} \tag{1.1}$$

Next, we will demonstrate this approach by discussing an example from the analysis of algorithms in which the first order asymptotic of all moments can be obtained by induction.

Example 1: Consider a sequence of random variable $\{X_n\}_{n\geq 0}$ that satisfies the following recurrence

$$X_n \stackrel{d}{=} X_{I_n} + X^*_{n-1-I_n}, \quad \forall n > 1, \tag{1.2}$$

with initial conditions $X_0 = 0$ and $X_1 = 1$, where $I_n = \text{Unif}\{0, n-1\}$, $X_n \stackrel{d}{=} X_n^*$, and $\{X_n\}_{n\geq 0}, \{X_n^*\}_{n\geq 0}$ are independent (for background concerning such sequences of random variables we refer to H. K. Hwang and R. Neininger [7]). We will use the method of moments to prove that

$$\frac{X_n - \mathbb{E}\left(X_n\right)}{\sqrt{\text{Var}\left(X_n\right)}} \stackrel{d}{\longrightarrow} N(0,1).$$

Solution. In the following four steps, we will repeatedly use the following property: If we have

$$a_n = \frac{2}{n} \sum_{j=0}^{n-1} a_j + b_n,$$

where $\{b_n\}_{n\geq 0}$, a_0 and a_1 are given, then for $n \geq 2$,

$$a_n = \frac{n+1}{3} a_1 + 2(n+1) \sum_{j=2}^{n-1} \frac{b_j}{(j+1)(j+2)} + b_n. \tag{1.3}$$

3

1. From the recurrence (1.2),

$$\mathbb{E}\left(X_n\right) = \frac{2}{n}\sum_{k=0}^{n-1}\mathbb{E}\left(X_k\right).$$

Hence by equation (1.3),

$$\mathbb{E}\left(X_n\right) = \begin{cases} 0, & \text{if } n = 0, \\ 1, & \text{if } n = 1, \\ \frac{n+1}{3}, & \text{if } n \geq 2. \end{cases} \tag{1.4}$$

2. Let $x_n = \mathbb{E}\left(X_n\right)$ and $P_n(z) = \mathbb{E}\left(e^{(X_n - x_n)z}\right)$. Moreover, set $P_n^{[r]} = \frac{z}{dz^r}P_n(z)\big|_{z=0}$. Then, again by the recurrence (1.2), we have

$$P_n^{[r]} = \frac{2}{n}\sum_{k=0}^{n-1}P_k^{[r]} + b_n^{[r]},$$

where

$$b_n^{[r]} = \frac{1}{n}\sum_{\substack{i_1+i_2+i_3=r \\ i_1 \neq r, i_2 \neq r}}\binom{r}{i_1, i_2, i_3}\sum_{k=0}^{n-1}P_k^{[i_1]}P_{n-1-k}^{[i_2]}(x_k + x_{n-1-k} - x_n)^{i_3}. \tag{1.5}$$

If $r = 2$, by equation (1.3) and (1.5), we get $\text{Var}\left(X_n\right) = \frac{2}{45}(n+1)$, for all $n \geq 5$.

3. Now, we claim: the r-th centralized and normalized moment of X_n satisfies, as $n \to \infty$,

$$\mathbb{E}\left(\left(\frac{X_n - \frac{1}{3}n}{\sqrt{\frac{2}{45}n}}\right)^r\right) = \begin{cases} \frac{(2m)!}{2^m m!} + o\left(1\right), & \text{if } r = 2m, \\ o\left(1\right), & \text{if } r = 2m + 1. \end{cases}$$

We use induction on r. By Step 1 and Step 2, we know that the claim holds for $r = 1, 2$. Now, we assume that the claim with exponents less than r is true.

Case 1: $r = 2m$

It is easy to see that $i_3 = 0$ gives the main contribution of $b_n^{[2m]}$. Hence

$$b_n^{[2m]} \sim \frac{1}{n}\sum_{i=1}^{2m-1}\binom{2m}{i}\sum_{k=0}^{n-1}P_k^{[i]}P_{n-1-k}^{[2m-i]}$$

$$\sim \frac{1}{n}\sum_{i=1}^{m-1}\binom{2m}{2i}\sum_{k=0}^{n-1}P_k^{[2i]}P_{n-1-k}^{[2(m-i)]}.$$

4

Then, by induction hypothesis,

$$b_n^{[2m]} \sim \frac{1}{n} \sum_{i=1}^{m-1} \binom{2m}{2i} \sum_{k=0}^{n-1} \frac{(2i)!}{2^i i!} \left(\frac{2}{45} k \right)^i \frac{(2(m-i))!}{2^{m-i}(m-i)!} \left(\frac{2}{45}(n-k-1) \right)^{m-i}$$

$$\sim \frac{(m-1)(2m)!}{2^m(m+1)!} \left(\frac{2}{45} \right)^m n^m.$$

By equation (1.3), we have

$$P_n^{[2m]} \sim 2(n+1) \sum_{j=2}^{n-1} \frac{1}{(j+1)(j+2)} \frac{(m-1)(2m)!}{2^m(m+1)!} \left(\frac{2}{45} \right)^m j^m$$

$$+ \frac{(m-1)(2m)!}{2^m(m+1)!} \left(\frac{2}{45} \right)^m n^m$$

$$\sim \frac{2(m-1)(2m)!}{(m-1)2^m(m+1)!} \left(\frac{2}{45} \right)^m n^m + \frac{(m-1)(2m)!}{2^m(m+1)!} \left(\frac{2}{45} \right)^m n^m$$

$$\sim \frac{(2m)!}{2^m m!} \left(\frac{2}{45} \right)^m n^m.$$

Case 2: $r = 2m+1$.

Similarly, we know $i_3 = 0$ gives the main contribution of $b_n^{[2m+1]}$. Hence

$$b_n^{[2m+1]} \sim \frac{1}{n} \sum_{\substack{i_1+i_2=2m+1 \\ i_1 \neq 2m+1, i_2 \neq 2m+1}} \binom{2m+1}{i_1, i_2} \sum_{k=0}^{n-1} P_k^{[i_1]} P_{n-1-k}^{[i_2]}.$$

Then as above,

$$b_n^{[2m+1]} = o\left(n^{m+\frac{1}{2}} \right).$$

Moreover, by equation (1.3), we get

$$P_n^{[2m+1]} = o\left(n^{m+\frac{1}{2}} \right).$$

4. From Step 2 and Step 3, we know

$$\lim_{n \to \infty} \mathbb{E}\left(\left(\frac{X_n^r - x_n}{\sqrt{\mathrm{Var}(X_n)}} \right)^r \right) = \begin{cases} \frac{(2m)!}{2^m m!}, & \text{if } r = 2m, \\ 0, & \text{if } r = 2m+1. \end{cases}$$

Hence our claim follows from Theorem 1.

From the above example, we see that the method of moments is very suitable for sequences of random variables that satisfy a distributional recurrence relation, a situation that is often encountered in the analysis of algorithms. In particular, no advanced theory is needed. However, there are limitation of this method. We will demonstrate this by a second example.

Example 2 : We consider the sequence of random variables $\{X_n\}_{n \geq 0}$ satisfying the following recurrence

$$X_n \overset{d}{=} X_{I_n} + 1, \quad \forall n > 0, \tag{1.6}$$

with initial condition $X_0 = 0$, where $I_n = \text{Unif}\{0, n-1\}$. Again, we claim that

$$\frac{X_n - \mathbb{E}(X_n)}{\sqrt{\text{Var}(X_n)}} \overset{d}{\longrightarrow} N(0, 1).$$

Solution. We will prove later that, if we have the following recurrence

$$a_n = b_n + \frac{1}{n} \sum_{k=1}^{n} a_{k-1},$$

where a_0 and $\{b_n\}_{n \geq 0}$ are given, then, for $n \geq 1$,

$$a_n = b_n + \sum_{k=2}^{n} \frac{1}{k} b_{k-1} + a_0. \tag{1.7}$$

1. From the recurrence (1.6),

$$\mathbb{E}(X_n) = 1 + \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E}(X_k) \quad \forall n \geq 1,$$

Hence by equation (1.7),

$$\mathbb{E}(X_n) = \begin{cases} H_n, & \text{if } n \geq 1, \\ 0, & \text{if } n = 0. \end{cases}$$

2. Let $x_n = \mathbb{E}(X_n)$ and $P_n(z) = \mathbb{E}\left(e^{(X_n - x_n)z}\right)$. Moreover set $P_n^{[r]} = \frac{z^r}{dz^r} P_n(z)\big|_{z=0}$. Then, again by the recurrence (1.6), we have

$$P_n^{[r]} = \frac{1}{n} \sum_{k=0}^{n-1} P_k^{[r]} + b_n^{[r]}, \quad \forall n \geq 1,$$

6

where

$$b_n^{[r]} = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{i=0}^{r-1} \binom{r}{i} P_k^{[r]} (1 + x_k - x_n)^{r-i}. \tag{1.8}$$

If $r = 2$, by equation (1.7) and (1.8), we have $\mathrm{Var}(X_n) \sim \log n$, for all $n \geq 1$.

3. Now, we claim: the r-th centralized and normalized moment of X_n satisfies, as $n \to \infty$,

$$\mathbb{E}\left(\left(\frac{X_n - \log n}{\sqrt{\log n}}\right)^r\right) = \begin{cases} \frac{(2m)!}{2^m m!} + o(1), & \text{if } r = 2m, \\ o(1), & \text{if } r = 2m+1. \end{cases}$$

We use induction on r. By Step 1 and Step 2, we know that the claim holds for $r = 1, 2$. Now, we assume that the claim with exponents less than r is true.

Case 1: $r = 2m$

From equation (1.8),

$$b_n^{[2m]} = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{i=0}^{2m-1} \binom{2m}{i} P_k^{[i]} (1 + x_k - x_n)^{2m-i}.$$

In this step, we have some trouble. By induction hypothesis, we have, for the term with $i = 2m - 1$,

$$\frac{1}{n} \sum_{k=0}^{n-1} \binom{2m}{2m-1} P_k^{[2m-1]} (1 + x_k - x_n)^1 = o\left(\log^{m-\frac{1}{2}} n\right).$$

All other terms are smaller. Hence,

$$b_n^{[2m]} = o\left(\log^{m-\frac{1}{2}} n\right).$$

By (1.7), this then yields $P_n^{[2m]} = o\left(\log^{m+\frac{1}{2}} n\right)$. Hence

$$\mathbb{E}\left((X_n - x_n)^{2m}\right) = o\left(\log^{m+\frac{1}{2}} n\right).$$

But this approximation is too large for our purpose.

The latter situation is typical for sequence of random variables that satisfy a "one-side" recurrence of the type (1.6). It is the main task of this thesis to provide a new variant of the method of moments which can be applied to such sequences. Moreover, we will apply our method to some examples from the analysis of priority trees, a data structure which will be introduced next.

1.2 Priority trees and their probabilistic analysis

1.2.1 Priority trees

If a college wants to save the information of thousands of students applying for scholarships, they need a suitable data structure. Assume the college has a formula to calculate the priority of the persons who apply for scholarships. Of course, they may need to insert new information or delete existing information. Now, we will introduce a data structure, called "priority queue", which is suitable for this purpose.

A priority queue is a data structure that supports the following two basic operations:

1. Add an element to the queue with an associated priority.

2. Remove the element from the queue that has the highest priority.

Priority queues are used in operating systems (e.g. job scheduling) and in discrete event simulation models. Every element in a priority queue has a fixed associated key value which determines its priority. Low key values correspond to high priority. Some background on priority queues can be found in T. H. Cormen, C. E. Leiserson, and R. L. Rivest [4]. Next, we will introduce a data structure, called priority trees (or p-trees for short) which is useful for implementing priority queues.

A p-tree is either empty or it is a sorted, non-increasing sequence of nodes, the "left path", such that to each node of the left path except the last one, a p-tree (possibly empty), the "right subtree" is associated. The nodes of the right subtree associated with a node x on the left path, are ranked between x and the left successor of x. The fundamental operations delete and insert for p-trees work as follows.

DELETE

The terminal node on the left path is called the "left leaf". It is the element with the smallest key value and hence the highest priority. Let the node x be the ancestor of the left leaf. If we want to remove an element with the highest priority, we just delete the left leaf, and then take the right subtree attached to x and add it to the left path.

INSERT

Inserting a new element p into a priority tree T works as follows:

INSERT(T, z):

- **if** $T = \emptyset$ or the key associated to the root of T is not larger than p, then let p be the new root and T its left subtree.

- **Otherwise** follow the left path of T and look 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.

In Figure 1.1, we give an example in which we repeatedly use the algorithm INSERT to create a priority tree starting with an empty set.
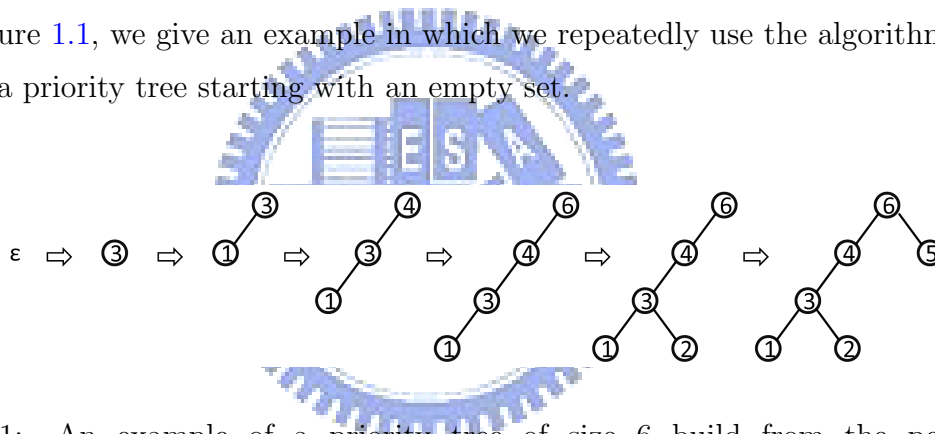

Figure 1.1: An example of a priority tree of size 6 build from the permutation (3,1,4,7,2,5,6) with the length of the left path is 4.

Deleting an element with the highest priority and inserting a new element are the main operations on priority trees. Hence, it is important to thoroughly analyze the running time of these operations. Therefore, we will use a random model on the input. More precisely, we consider the model where all $n!$ permutations of the numbers $1, \ldots, n$ generating all the p-trees of size n are equally likely. Then, characteristic parameters describing the running time become random variables. Apart from the model proposed above (called $\mathcal{A}$ in the sequel) we will need two more models: the model obtained by generating a priority

tree from a random permutation starting with positive infinity "$+\infty$" (this model will be called $\mathcal{B}$) and the model obtained by generating a priority tree from a random permutation starting with negative infinity "$-\infty$" (this model will be called $\mathcal{C}$). Moreover, we make the convention that both $+\infty$ and $-\infty$ do not count towards the size of the p-tree. The reason for considering these three models lies in the natural decomposition obtained by conditioning on the first element in the random permutation (where again $+\infty$ and $-\infty$ do not count); see Figure 1.2. The idea for this decomposition is not new and seems to have appeared first in a paper of A. Panholzer and H. Prodinger [12].
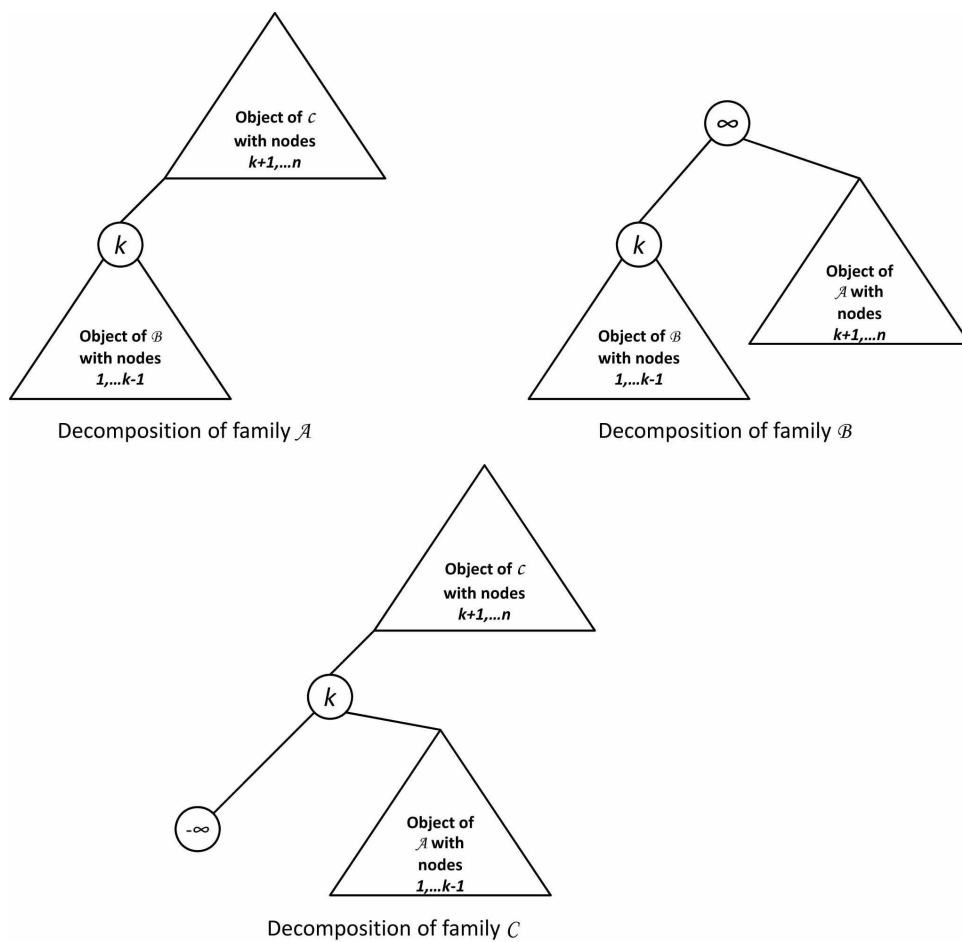
Figure 1.2: The decomposition of family $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$

1.2.2 Previous results on priority trees

Length of the left path

In [12], A. Panholzer and H. Prodinger investigated the length of the left path in a p-tree of size n, where the length is by definition the number of nodes (for the model $\mathcal{C}$ we make the convention that $-\infty$ does not count). This parameter describes the cost of the running time of the algorithm DELETE from the previous section. Under the random model above, the authors obtained mean and variance and proved a central limit theorem. We briefly sketch their method of proof.

1. Let $A_{n,m}$, $B_{n,m}$, and $C_{n,m}$ be the probabilities that a random tree of $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ with size n has a left path of length m. Then, by the decomposition from Figure 1.2, we have the following recurrences:

$$A_{n,m} = \frac{1}{n} \sum_{k=1}^{n} \sum_{i=0}^{m} C_{n-k,i} B_{k-1,m-i} \quad \forall n \geq 1,$$

$$B_{n,m} = \frac{1}{n} \sum_{k=1}^{n} B_{k-1,m-1} \quad \forall n \geq 1, \tag{1.9}$$

$$C_{n,m} = \frac{1}{n} \sum_{k=1}^{n} C_{n-k,m-1} \quad \forall n \geq 1.$$

Consider the bivariate generating functions

$$A(z,v) = \sum_{n \geq 0} \sum_{m \geq 0} A_{n,m} z^n v^m,$$

$$B(z,v) = \sum_{n \geq 0} \sum_{m \geq 0} B_{n,m} z^n v^m,$$

$$C(z,v) = \sum_{n \geq 0} \sum_{m \geq 0} C_{n,m} z^n v^m.$$

From (1.9), we have the following system of differential equations:

$$\frac{\partial}{\partial z} A(z,v) = B(z,v) C(z,v),$$

$$\frac{\partial}{\partial z} B(z,v) = \frac{v}{1-z} B(z,v), \tag{1.10}$$

$$\frac{\partial}{\partial z} C(z,v) = \frac{v}{1-z} C(z,v).$$

This system is easily solved and we obtain

$$B(z, v) = \frac{v}{(1 - z)^v},$$

$$C(z, v) = \frac{1}{(1 - z)^v},$$

and

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

$$A(z, v) = \frac{1}{1 - 2v} \left(1 - v - \frac{v}{(1 - z)^{2v-1}} \right). \tag{1.11}$$

2. Now, denoted by A_n, B_n, and C_n the expectations of the length of the left paths in objects of size n in the families $\mathcal{A}$, $\mathcal{B}$ resp. $\mathcal{C}$. Moreover, define $A(z) := \frac{\partial}{\partial v} A(z, v)\big|_{v=1}$, $B(z) := \frac{\partial}{\partial v} B(z, v)\big|_{v=1}$, and $C(z) := \frac{\partial}{\partial v} C(z, v)\big|_{v=1}$ which are the ordinary generating functions of A_n, B_n, and C_n. By differentiating and reading of the coefficients of $A(z)$, $B(z)$, and $C(z)$ in (1.11), we have

$$A_n = 2H_n - 1 \quad \forall n \geq 1, \quad A_0 = 0,$$
$$B_n = H_n + 1 \quad \forall n \geq 1, \quad B_0 = 0,$$
$$C_n = H_n \quad \forall n \geq 1, \quad C_0 = 0.$$

3. Differentiating equation (1.11) twice with respect to v and evaluating at $v = 1$ yields second factorial moments . For instance the second factorial moment $\check{A}_n$ of the length of the left path in a p-trees of size n from random model $\mathcal{A}$ is given by

$$\check{A}_n = 4H_n^2 - 4H_n - 4H_n^{(n)} + 4 \quad \forall n \geq 1 \quad \check{A}_0 = 0.$$

Consequently, the variance $\hat{A}_n$ is given by

$$\hat{A}_n = \check{A}_n + A_n - A_n^2 = 2 \log n + \mathcal{O}(1).$$

4. Let $A_n(v) = \frac{1}{n}[z^n] \frac{\partial}{\partial z} A(z, v)$. From (1.11) and singularity analysis (see Chapter VI in P. Flajolet and R. Sedgewick [5]), we obtain

$$A_n(v) = \frac{1}{n} \left[z^{n-1}\right] \frac{\partial}{\partial z} A(z, v)$$
$$= \frac{1}{n} \left[z^{n-1}\right] \frac{\partial}{\partial z} \frac{v}{(1-z)^{2v}}$$
$$= \frac{v}{2v-1} \binom{n+2v-2}{n}$$
$$= \frac{vn^{2v-2}}{\Gamma(2v)} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right). \tag{1.12}$$

for fixed v and $n \to \infty$. Next, let

$$\varphi_{\Omega^*}(t) = e^{-i\mu_n t/\sigma_n} A_n\left(e^{it/\sigma_n}\right).$$

By (1.12), we have

$$A_n\left(e^{it/\sigma_n}\right) = \exp\left(2(e^{it/\sigma_n} - 1)\log n\right)(1 + o(1))$$
$$= \exp\left(2\log n \frac{it}{\sigma_n} - \frac{t^2}{2}\right)(1 + o(1)).$$

Then, as $n \to \infty$,

$$\varphi_{\Omega^*}(t) = \exp\left(-2\log n \frac{it}{\sigma_n}\right) \exp\left(2\log n \frac{it}{\sigma_n} - \frac{t^2}{2}\right)(1 + o(1))$$
$$= e^{-t^2/2}(1 + o(1)) \to e^{-t^2/2}.$$

By Lévy's continuity theorem, the central limit theorem follows from this.

We will provide another proof of the above result by using the method of moments. In doing so, we will encounter similar problems as in Example 2 in Section 1. However, we will overcome these problems by introducing some new ideas (for more details see the next example below). It should be mentioned here that from a technical point of view our proof is more complicated then the above one. So, this example will be mainly used to introduce our new approach.

The number of key comparisons when inserting a random element:

M. Kuba and A. Panholzer [9] considered the number of key comparisons when inserting a random element in a p-tree. This parameter describes the running time of the algorithm

13

INSERT. They authors sketched a proof of a limit law of this quantity (mean and variance were already discussed earlier in [12]). We will repeat their arguments here.

1. Let $A_n^{[I]}$, $B_n^{[I]}$, and $C_n^{[I]}$ be the random variables that count the number of key comparisons when inserting a randomly chosen element from $\{j + \frac{1}{2} : 0 \leq j \leq n\}$ into a random p-tree of the families $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ with size n. Let

$$A(z, v) := \sum_{n \geq 1} \sum_{m > 0} n \mathbb{P}\left(A_n^{[I]} = m\right) z^n v^m,$$

and similar we define $B(z, v)$ and $C(z, v)$. Then, by the decomposition from Section 1.2.1, we have the following system of differential equations:

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

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

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

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

By solving for $A(z, v)$ we get the following homogeneous third order linear differential equation for $A(z, v)$:

$$\frac{\partial}{\partial z^3} A(z, v) - \frac{3 + 2v}{1 - z} \frac{\partial^2}{\partial z^2} A(z, v) + 2v \left(\frac{2}{(1 - z)^2} - \frac{v}{(1 - z)^{v+1}} \right) \frac{\partial}{\partial z} A(z, v)$$

$$+ \frac{2v^2}{(1 - z)^{v+2}} A(z, v) = 0,$$

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

2. Let $H_r(z) := \frac{\partial^r}{\partial v^r} H(z, v)$, where $H(z, v) = \frac{\partial}{\partial z} ((1 - z) A(z, v))$. Then from the above equation we get the following differential equation:

$$H_r''(z) - \frac{2}{1 - z} H_r'(z) - \frac{2}{(1 - z)^2} H_r(z) = S_r(z),$$

where $S_r(z)$ is the inhomogeneous part. Using variation of parameters then yields the following solution

14

$$H_r(z) = \frac{1}{3(1-z)^2} \int_0^z (1-t)^3 S_r(t)dt - \frac{1-z}{3} \int_0^z S_r(t)dt + \frac{C_{r,1}}{(1-z)^2} + C_{r,2}(1-z),$$

$$(1.13)$$

where $C_{r,1}$ and $C_{r,2}$ are suitable constants.

3. By using induction, $H_r(z)$ is shown to exhibit the form

$$H_r(z) = \sum_{m=0}^{2r} \alpha_r^{(m)} \frac{1}{(1-z)^2} L(z)^{2r-m} + \sum_{m=0}^{2r-2} \beta_r^{(m)}(1-z)L(z)^{2r-2-m}, \qquad (1.14)$$

where $\alpha_r^{(m)}$ and $\beta_r^{(m)}$ are some constants which satisfy some recurrences and $L(z) := \log(\frac{1}{1-z})$.

4. Let $A_r(z) := \frac{\partial^r}{\partial v^r} A(z,v)\big|_{v=1} = \sum_{n \geq 1} n\mathbb{E}\left(\left(A_n^{[I]}\right)^{\underline{r}}\right) z^n$, where $\mathbb{E}\left(\left(A_n^{[I]}\right)^{\underline{r}}\right)$ denote the $r-$th factorial moment. By equation (1.14), $A_r(z)$ satisfies the same type of expansion as $H_r(z)$. Therefore by reading of the coefficients of $\frac{1}{(1-z)^2} L(z)^{2r-m}$, we obtain

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

where $c_{r,n}$ are some constants and ϵ is a small constant.

5. Finally, the central moments are obtained from

$$\mathbb{E}\left(\left(A_n^{[I]} - \frac{1}{3}\log^2 n\right)^r\right) = \sum_{k=0}^{r} \binom{r}{k}(-1)^{r-k}\left(\frac{1}{3}\right)^{r-k}(\log^{2r-2k} n)\mathbb{E}\left(\left(A_n^{[I]}\right)^k\right).$$

and

$$\mathbb{E}\left(\left(A_n^{[I]}\right)^k\right) = \sum_{r=0}^{k} \left\{{k \atop r}\right\} \mathbb{E}\left(\left(A_n^{[I]}\right)^{\underline{r}}\right),$$

by using the expansion from Step 4 ($\left\{{k \atop r}\right\}$ denotes the Stirling numbers of first kind). The limits of the central moments are then computed via these exact expressions.

The most complicated step in the proof of A. Panholzer and M. Kuba is the fifths one. This is due to the appearance of many cancellations in the central moments. Hence, the

authors needed a rather precise knowledge of the coefficients in the expansion from Step 4. This part can be simplified by shifting the mean as in the Examples from Section 1.1 and using induction. Then, we however encounter again the same problem as in Example 2 from Section 1.1. Therefore, we will first study the (non-central) moments. In order to do so, we will again use induction. Once the behavior of the non-central moments is clear, we will use this as input for the second induction. This will help us to overcome the problem observed in Section 1.1.

Overall, our method will turn out to be structurally easier than the one used by A. Panholzer and M. Kuba. Carrying out all details is, however, still rather messy. Therefore, as already mentioned before, we will introduce our method by analyzing first the length of the left path.

Chapter 2

Length of the Left Path

2.1 Statement of the result

As already mentioned in Chapter 1, in this chapter we will use the method of moments to give a second proof of the results on the length of the left path from Section 1.2.2. Even though this new proof will be much more complicated, this example is technically easier than the one considered in the next chapter. Hence, we will use this example as a kind of warm up in order to explain details concerning our new approach.

Now, let $A_n^{[L]}, B_n^{[L]}$ resp. $C_n^{[L]}$ be the length of the left path in a random size-n priority tree of the families $\mathcal{A}$, $\mathcal{B}$ resp. $\mathcal{C}$. Then, from the decomposition of Section 1.2.1, we obtain

$$A_n^{[L]} \stackrel{d}{=} B_{I_n-1}^{[L]} + C_{n-I_n}^{[L]} \quad \forall n \geq 1, \tag{2.1}$$

$$B_n^{[L]} \stackrel{d}{=} 1 + B_{I_n-1}^{[L]} \quad \forall n \geq 1, \tag{2.2}$$

$$C_n^{[L]} \stackrel{d}{=} 1 + C_{n-I_n}^{[L]} \quad \forall n \geq 1, \tag{2.3}$$

where $A_0^{[L]} = 0$, $B_0^{[L]} = 1$, $C_0^{[L]} = 0$ and the distribution of I_n is given by

$$\mathbb{P}(I_n = j) = \frac{1}{n}, \quad \text{for all } 1 \leq j \leq n.$$

From this probabilistic description, we will deduce the following result.

17

Theorem 2. *The length of the left path $A_n^{[L]}$ in a random size-n priority tree is asymptotically Gaussian distributed:*

$$\frac{A_n^{[L]} - \mathbb{E}\left(A_n^{[L]}\right)}{\sqrt{\mathrm{Var}\left(A_n^{[L]}\right)}} \xrightarrow{d} N(0,1).$$

Moreover, the mean value and variance satisfy

$$\mathbb{E}\left(A_n^{[L]}\right) \sim 2\ln n, \quad \mathrm{Var}\left(A_n^{[L]}\right) \sim 2\ln n.$$

2.2 Proof of the result - Theorem 2

The steps of this proof are similar to Example 1 in Chapter 1. In order to get all the central moments, first we need to find the mean values of $A_n^{[L]}$, $B_n^{[L]}$, and $C_n^{[L]}$. Let

$$P_n(t) := \mathbb{E}(e^{A_n^{[L]}t}),$$
$$Q_n(t) := \mathbb{E}(e^{B_n^{[L]}t}),$$
$$T_n(t) := \mathbb{E}(e^{C_n^{[L]}t}).$$

By the distributional recurrences form Section 2.1, we get

$$P_n(t) = \frac{1}{n}\sum_{k=1}^{n} Q_{k-1}(t)T_{n-k}(t) \quad \forall n \geq 1, \tag{2.4}$$

$$Q_n(t) = \frac{1}{n}e^t \sum_{k=1}^{n} Q_{k-1}(t) \quad \forall n \geq 1, \tag{2.5}$$

$$T_n(t) = \frac{1}{n}e^t \sum_{k=1}^{n} T_{k-1}(t) \quad \forall n \geq 1, \tag{2.6}$$

with initial conditions $P_0(t) = 1$, $Q_0(t) = e^t$ and $T_0(t) = 1$.

Let $P_n^{[r]} = \frac{d^r}{dt^r}P_n(t)\big|_{t=0}$, $Q_n^{[r]} = \frac{d^r}{dt^r}Q_n(t)\big|_{t=0}$ and $T_n^{[r]} = \frac{d^r}{dt^r}T_n(t)\big|_{t=0}$. Hence,

$$P_n^{[r]} = \frac{1}{n}\sum_{i=0}^{r}\binom{r}{i}\sum_{k=1}^{n} Q_{k-1}^{[i]}T_{n-k}^{[r-i]} \quad \forall n \geq 1, \tag{2.7}$$

$$Q_n^{[r]} = \frac{1}{n}\sum_{k=1}^{n} Q_{k-1}^{[r]} + \frac{1}{n}\sum_{i=1}^{r}\binom{r}{i}\sum_{k=1}^{n} Q_{k-1}^{[r-i]} \quad \forall n \geq 1, \tag{2.8}$$

$$T_n^{[r]} = \frac{1}{n}\sum_{k=1}^{n} T_{k-1}^{[r]} + \frac{1}{n}\sum_{i=1}^{r}\binom{r}{i}\sum_{k=1}^{n} T_{k-1}^{[r-i]} \quad \forall n \geq 1, \tag{2.9}$$

with initial conditions $P_0^{[i]} = 0$, $Q_0^{[i]} = 1$ and $T_0^{[i]} = 0$ for all $i \geq 1$ and $P_0^{[0]} = Q_0^{[0]} = T_0^{[0]} = 1$.

Note that by (2.8) and (2.9), all moments of $B_n^{[L]}$ and $C_n^{[L]}$ satisfy a recurrence of the same type. Later on we will see that the same recurrence is satisfied by the central moments as well. Hence, we first study this recurrence in the next lemma.

Lemma 1. *Consider the recurrence $a_n = b_n + \frac{1}{n}\sum_{k=1}^{n} a_{k-1}$, where a_0, $\{b_n\}_{n\geq 0}$ are given. Then, for $n \geq 1$,*

$$a_n = b_n + \sum_{k=2}^{n}\frac{1}{k}b_{k-1} + a_0.$$

Proof. Because

$$na_n - (n-1)a_{n-1} = nb_n - (n-1)b_{n-1} + a_{n-1}.$$

Therefore

$$a_n = b_n - \frac{n-1}{n}b_{n-1} + a_{n-1}, \quad n \geq 1$$

$$= b_n + \sum_{k=2}^{n}\frac{1}{k}b_{k-1} + a_0. \quad \blacksquare$$

We use Lemma 1 to obtain asymptotic expansions of the mean values.

Lemma 2. *The mean values of $A_n^{[L]}$, $B_n^{[L]}$ and $C_n^{[L]}$ satisfy, as $n \to \infty$,*

$$\mathbb{E}(A_n^{[L]}) = 2\ln n + 2\gamma - 1 + \mathcal{O}\left(\frac{\log n}{n}\right),$$

$$\mathbb{E}(B_n^{[L]}) = \ln n + \gamma + 1 + \mathcal{O}\left(\frac{1}{n}\right),$$

$$\mathbb{E}(C_n^{[L]}) = \ln n + \gamma + \mathcal{O}\left(\frac{1}{n}\right).$$

Proof. Taking $r = 1$ in equation (2.7), (2.8) and (2.9),

$$\mathbb{E}(A_n^{[L]}) = P_n'(0) = \frac{1}{n}\sum_{k=1}^{n}(\mathbb{E}(B_{k-1}^{[L]}) + \mathbb{E}(C_{k-1}^{[L]})) \quad \forall n \geq 1, \tag{2.10}$$

$$\mathbb{E}(B_n^{[L]}) = Q_n'(0) = 1 + \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(B_{k-1}^{[L]}) \quad \forall n \geq 1, \tag{2.11}$$

$$\mathbb{E}(C_n^{[L]}) = T_n'(0) = 1 + \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(C_{k-1}^{[L]}) \quad \forall n \geq 1, \tag{2.12}$$

where $\mathbb{E}(A_0^{[L]}) = 0, \quad \mathbb{E}(B_0^{[L]}) = 1$ and $\mathbb{E}(C_0^{[L]}) = 0$.

Hence, by Lemma 1,

$$\mathbb{E}(B_n^{[L]}) = 1 + \sum_{k=2}^{n}\frac{1}{k} + 1 = H_n + 1 = \ln n + \gamma + 1 + \mathcal{O}\left(\frac{1}{n}\right) \quad \forall n \geq 1,$$

$$\mathbb{E}(C_n^{[L]}) = 1 + \sum_{k=2}^{n}\frac{1}{k} = H_n = \ln n + \gamma + \mathcal{O}\left(\frac{1}{n}\right) \quad \forall n \geq 1.$$

From equation (2.10), we have

$$\mathbb{E}(A_n^{[L]}) = \frac{1}{n}\left(\sum_{k=1}^{n}\mathbb{E}(B_{k-1}^{[L]}) + \mathbb{E}(C_{k-1}^{[L]})\right)$$

$$= \frac{1}{n}\left(1 + \sum_{k=2}^{n}2\ln(k-1) + 2\gamma + 1 + \mathcal{O}\left(\frac{1}{k-1}\right)\right).$$

By Euler-Maclaurin summation formula,

$$\mathbb{E}(A_n^{[L]}) = \frac{2}{n}\left(\int_1^{n-1}\ln x\,dx + \frac{\ln(n-1)}{2} + c + \mathcal{O}\left(\frac{1}{n}\right)\right) + 2\gamma + 1 + \mathcal{O}\left(\frac{\log n}{n}\right),$$

where c is a constant. Then,

$$\mathbb{E}(A_n^{[L]}) = 2\ln n + 2\gamma - 1 + \mathcal{O}\left(\frac{\log n}{n}\right). \quad \blacksquare$$

The next lemma extends the expansions from Lemma 2 to all higher moments. In particular, this lemma will be the crucial new tool for overcoming the problem we have countered in Step 3 of Example 2 of Chapter 1.

Let $\mathfrak{Pol}_n(x)$ represent a polynomial in x with maximum degree $\leq n$ and $\mathfrak{Pol}(x)$ represent a polynomial in x without any restriction to the degree.

Lemma 3. *The expansions of the r-th moments of $A_n^{[L]}$, $B_n^{[L]}$ and $C_n^{[L]}$ satisfy, as $n \to \infty$*

$$\mathbb{E}\left(A_n^{[L]}\right)^r = \mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right),$$

$$\mathbb{E}\left(B_n^{[L]}\right)^r = \mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right),$$

$$\mathbb{E}\left(C_n^{[L]}\right)^r = \mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right),$$

where ϵ denotes an arbitrarily small constant $\epsilon > 0$.

Proof. First, we will prove the result for $\mathbb{E}\left(B_n^{[L]}\right)^r$. We use induction on r. By Lemma 2, $r = 1$ is true. Now, we assume that the claim with exponents less than r is true. By equation (2.8),

$$Q_n^{[r]} = \frac{1}{n}\sum_{k=1}^{n} Q_{k-1}^{[r]} + \underbrace{\frac{1}{n}\sum_{i=1}^{r}\binom{r}{i}\sum_{k=1}^{n} Q_{k-1}^{[r-i]}}_{:=b_n}.$$

By induction hypothesis, $Q_k^{[r-i]} = \mathfrak{Pol}(\log k) + \mathcal{O}\left(\frac{1}{k^{1-\epsilon}}\right)$ for all $i = 1, \ldots, r$, where ϵ denotes an arbitrarily small constant $\epsilon > 0$. Therefore,

$$b_n = \frac{1}{n}\left(\sum_{i=1}^{r}\binom{n}{i}Q_0^{[r-i]} + \sum_{k=1}^{n-1}\mathfrak{Pol}(\log k) + \mathcal{O}\left(\frac{1}{k^{1-\epsilon}}\right)\right)$$

$$= \frac{1}{n}\sum_{k=1}^{n-1}\mathfrak{Pol}(\log k) + \frac{1}{n}\sum_{k=1}^{n-1}\mathcal{O}\left(\frac{1}{k^{1-\epsilon}}\right) + \mathcal{O}\left(\frac{1}{n}\right)$$

$$= \mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right).$$

Then, we know

$$Q_n^{[r]} = b_n + \sum_{k=2}^{n}\frac{1}{k}b_{k-1} + 1$$

$$= \mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right) + \sum_{k=2}^{n}\frac{1}{k}\left(\mathfrak{Pol}(\log k - 1) + \mathcal{O}\left(\frac{1}{(k-1)^{1-\epsilon}}\right)\right)$$

$$= \mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right) + \sum_{k=2}^{n}\mathcal{O}\left(\frac{1}{k}\frac{1}{(k-1)^{1-\epsilon}}\right)$$

$$= \mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right) + \sum_{k=2}^{n} \mathcal{O}\left(\frac{1}{k}\frac{1}{(k-1)^{1-\epsilon}}\right)$$

$$= \mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right) + \underbrace{\sum_{k=2}^{\infty} \mathcal{O}\left(\frac{1}{k}\frac{1}{(k-1)^{1-\epsilon}}\right)}_{=\text{constant}} - \sum_{k=n+1}^{\infty} \mathcal{O}\left(\frac{1}{k}\frac{1}{(k-1)^{1-\epsilon}}\right)$$

$$= \mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right).$$

The claims for $\mathbb{E}\left(A_n^{[L]}\right)^r$ and $\mathbb{E}\left(C_n^{[L]}\right)^r$ are proved similarly. ∎

Now, we turn to central moments. Therefore, let $\Phi_n(t) := e^{-p_n t}P_n(t)$, $\Psi_n(t) := e^{-q_n t}Q_n(t)$ and $\Omega_n(t) := e^{-t_n t}T_n(t)$ where $p_n = \mathbb{E}(A_n^{[L]})$, $q_n = \mathbb{E}(B_n^{[L]})$ and $t_n = \mathbb{E}(C_n^{[L]})$, Let $\Phi_n^{[r]} = \Phi_n^{(r)}(0), \Psi_n^{[r]} = \Psi_n^{(r)}(0)$ and $\Omega_n^{[r]} = \Omega_n^{(r)}(0)$. Then by equation (2.4), (2.5) and (2.6), we have

$$\Phi_n^{[r]} = \frac{1}{n} \sum_{i_1+i_2+i_3=r} \sum_{k=1}^{n} \binom{r}{i_1, i_2, i_3} \Psi_{k-1}^{[i_1]} \Omega_{n-k}^{[i_2]} (\Delta_{n,k}^{\Phi})^{i_3} \quad \forall n \geq 1, \tag{2.13}$$

$$\Psi_n^{[r]} = \frac{1}{n} \sum_{k=1}^{n} \Psi_{k-1}^{[r]} + \frac{1}{n} \sum_{k=1}^{n} \sum_{i=0}^{r-1} \binom{r}{i} \Psi_{k-1}^{[r]} (\Delta_{n,k}^{\Psi})^{r-i} \quad \forall n \geq 1, \tag{2.14}$$

$$\Omega_n^{[r]} = \frac{1}{n} \sum_{k=1}^{n} \Omega_{k-1}^{[r]} + \frac{1}{n} \sum_{k=1}^{n} \sum_{i=0}^{r-1} \binom{r}{i} \Omega_{k-1}^{[r]} (\Delta_{n,k}^{\Omega})^{r-i} \quad \forall n \geq 1, \tag{2.15}$$

with initial conditions $\Phi_0^{[i]} = \Psi_0^{[i]} = \Omega_0^{[i]} = 0$, for all $i \geq 1$ and $\Phi_0^{[0]} = \Psi_0^{[0]} = \Omega_0^{[0]} = 1$, where $\Delta_{n,k}^{\Phi} = q_{k-1} + t_{n-k} - p_n$, $\Delta_{n,k}^{\Psi} = 1 - q_n + q_{k-1}$ and $\Delta_{n,k}^{\Omega} = 1 - t_n + t_{k-1}$.

First we consider the variance.

Lemma 4. *The variances of $A_n^{[L]}$, $B_n^{[L]}$ and $C_n^{[L]}$ satisfy, as $n \to \infty$,*

$$\mathrm{Var}(A_n^{[L]}) \sim 2\log n, \quad \mathrm{Var}(B_n^{[L]}) \sim \log n, \quad \text{and} \quad \mathrm{Var}(C_n^{[L]}) \sim \log n.$$

Proof. First, we consider $\mathrm{Var}(B_n^{[L]})$. Since $\Psi_n^{[1]} = 0$ and by equation (2.14), we obtain

$$\mathrm{Var}(B_n^{[L]}) = \Psi_n^{[2]} = \frac{1}{n} \sum_{k=1}^{n} \Psi_{k-1}^{[2]} + \underbrace{\frac{1}{n} \sum_{k=1}^{n} (\Delta_{n,k}^{\Psi})^2}_{b_n :=}.$$

22

Next, we have

$$b_n = \frac{1}{n}\sum_{k=1}^{n}(\Delta_{n,k}^{\Psi})^2$$

$$= \frac{1}{n}\sum_{k=1}^{n}(1 - q_n + q_{k-1})^2$$

$$\sim \frac{1}{n}\sum_{k=2}^{n}(1 - \log n + \log(k-1))^2$$

$$\sim \int_0^1 (1 + \log x)^2\,dx$$

$$\sim 1.$$

Hence by Lemma 1,

$$\mathrm{Var}(B_n^{[L]}) \sim 1 + \sum_{k=1}^{n}\frac{1}{k} \sim \log n.$$

Similarly,

$$\mathrm{Var}(C_n^{[L]}) \sim \log n.$$

By the relation between $\Phi_n^{[2]}$, $\Psi_n^{[2]} \sim \log n$, and $\Omega_n^{[2]} \sim \log n$, we have

$$\mathrm{Var}(A_n^{[L]}) = \Phi_n^{[2]}$$

$$= \frac{1}{n}\left(\sum_{k=1}^{n}\Psi_{k-1}^{[2]} + \sum_{k=1}^{n}\Omega_{n-k+1}^{[2]} + \sum_{k=1}^{n}(\Delta_{n,k}^{\Phi})^2\right)$$

$$\sim \frac{1}{n}\left(\sum_{k=2}^{n}2\log(k-1) + \sum_{k=1}^{n}(q_{k-1} + t_{n-k+1} - p_n)^2\right)$$

$$\sim \frac{1}{n}\left(\sum_{k=2}^{n}2\log(k-1) + \sum_{k=2}^{n}(\log(k-1) + \log(n-k+1) - 2\log n)^2\right)$$

$$\sim 2\log n. \quad \blacksquare$$

So far, we have treated means, variances, and all moments. So what is left is to consider r-th central moments with $r \geq 1$. From equation (2.13), the r-th central moment of $A_n^{[L]}$ is a combination of the central moments of $B_n^{[L]}$ and $C_n^{[L]}$. Therefore, we first consider central moments of the latter two random variables.

Lemma 5. *The r-th centralized and normalized moments of B_n and C_n satisfy, as $n \to \infty$,*

$$\mathbb{E}\left(\left(\frac{B_n^{[L]} - \log n}{\sqrt{\log n}}\right)^r\right) = \begin{cases} \frac{(2m)!}{2^m m!} + o(1), & \text{if } r = 2m, \\ o(1), & \text{if } r = 2m+1. \end{cases} \tag{2.16}$$

$$\mathbb{E}\left(\left(\frac{C_n^{[L]} - \log n}{\sqrt{\log n}}\right)^r\right) = \begin{cases} \frac{(2m)!}{2^m m!} + o(1), & \text{if } r = 2m, \\ o(1), & \text{if } r = 2m+1. \end{cases} \tag{2.17}$$

Proof. First, we consider $\Psi_n^{[r]}$. We use induction on r. By Lemma 2 and Lemma 4, we know that the claim holds for $r = 1$ and $r = 2$. Now, we assume that the claim with exponents less than r is true. We consider the two cases r is even and r is odd in equation (2.14).

Case 1: If $r = 2m$, then

$$\Psi_n^{[2m]} = \frac{1}{n}\sum_{k=1}^n \Psi_{k-1}^{[2m]} + \underbrace{\frac{1}{n}\sum_{k=1}^n \sum_{i=0}^{2m-1}\binom{2m}{i}\Psi_{k-1}^{[i]}(\Delta_{n,k}^\Psi)^{2m-i}}_{b_n :=},$$

where $\Delta_{n,k} = 1 - q_n + q_{k-1}$.

We consider two parts according to whether i is even or not,

$$b_n = \underbrace{\frac{1}{n}\sum_{i=0}^{m-1}\sum_{k=1}^n \binom{2m}{2i}\Psi_{k-1}^{[2i]}(\Delta_{n,k}^\Psi)^{2(m-i)}}_{\alpha_n :=}$$

$$+ \underbrace{\frac{1}{n}\sum_{i=1}^m \sum_{k=1}^n \binom{2m}{2i-1}\Psi_{k-1}^{[2i-1]}(\Delta_{n,k}^\Psi)^{2(m-i)+1}}_{\beta_n :=}.$$

First, we consider α_n and use the induction hypothesis to get the following :

$$\alpha_n = \frac{1}{n}\sum_{i=0}^{m-1}\sum_{k=1}^n \binom{2m}{2i}\Psi_{k-1}^{[2i]}(\Delta_{n,k}^\Psi)^{2(m-i)}$$

$$\sim \sum_{i=1}^{m-1}\binom{2m}{2i}\frac{(2i)!}{2^i i!}\underbrace{\frac{1}{n}\sum_{k=2}^n \log^i(k-1)(1 - \log(n) + \log(k-1))^{2(m-i)}}_{:=T_i}.$$

24

Now, we just consider T_i for $1 \leq i \leq m - 1$. By integral approximation,

$$
\begin{aligned}
T_i &= \frac{1}{n} \sum_{k=2}^{n} \left(\log n + \log \left(\frac{k-1}{n} \right) \right)^i (1 - \log(n) + \log(k-1))^{2(m-i)} \\
&\sim \log^i n \int_0^1 (1 + \log x)^{2(m-i)} \, dx \\
&\sim c_i \log^i n,
\end{aligned}
$$

where c_i is a constant. Hence, we know that $i = m - 1$ gives the main contribution of α_n. Consequently,

$$
\begin{aligned}
\alpha_n &\sim \frac{1}{n} \sum_{k=1}^{n-1} \binom{2m}{2m-2} \frac{(2m-2)!}{2^{m-1}(m-1)!} \log^{m-1}(k)(1 - \log(n) + \log(k))^2 \\
&\sim \binom{2m}{2m-2} \frac{(2m-2)!}{2^{m-1}(m-1)!} \log^{m-1} n = \frac{(2m)!}{2^m(m-1)!} \log^{m-1} n.
\end{aligned}
\tag{2.18}
$$

Now, we consider β_n,

$$
\beta_n = \frac{1}{n} \sum_{i=1}^{m} \sum_{k=1}^{n} \binom{2m}{2i-1} \Psi_{k-1}^{[2i-1]} (\Delta_{n,k}^{\Psi})^{2(m-i)+1}.
$$

Similarly as above, for $i \leq m - 1$ we obtain the bound $o \left(\log^{m-\frac{3}{2}} n \right)$. So what is left is to look at the contribution of $i = m$ which is

$$
T := \frac{1}{n} \sum_{k=1}^{n} 2m \Psi_{k-1}^{[2m-1]} \Delta_{n,k}^{\Psi}.
$$

By Lemma 3, $\Psi_{k-1}^{[2m-1]}$ is a polynomial in $\log n$. Therefore by induction hypothesis, we know that $\Psi_{k-1}^{[2m-1]} \sim \sum_{i=0}^{m-1} c_i \log^i(k-1)$, where the c_i's are suitable constants for all i. Consequently,

$$
\begin{aligned}
T &\sim \frac{2m}{n} \sum_{k=2}^{n} \left(\sum_{i=0}^{m-1} c_i \log^i(k-1) \right) \Delta_{n,k}^{\Psi} \\
&= \frac{2m}{n} \sum_{k=2}^{n} \left(\sum_{i=0}^{m-1} c_i \log^i(k-1) \right) (1 - \log(n) + \log(k-1)).
\end{aligned}
$$

Using integral approximation,

$$T \sim c \log^{m-1} n \underbrace{\int_0^1 (1 + \log x) \, dx}_{=0} \quad \Rightarrow \quad T = o\left(\log^{m-1} n\right). \tag{2.19}$$

Overall, $\beta_n = o\left(\log^{m-1} n\right)$. Next, by equation (2.18) and (2.19),

$$b_n = \alpha_n + \beta_n \sim \frac{(2m)!}{2^m(m-1)!} \log^{m-1} n.$$

By Lemma 1,

$$\Psi_n^{[2m]} = b_n + \sum_{k=2}^{n} \frac{1}{k} b_{k-1} \sim \frac{(2m)!}{2^m m!} \log^m n.$$

Therefore,

$$\mathbb{E}\left(\left(\frac{B_n^{[L]} - \log n}{\sqrt{\log n}}\right)^{2m}\right) = \frac{(2m)!}{2^m m!} + o(1), \quad \text{as } n \to \infty.$$

Case 2: If $r = 2m + 1$, then

$$\Psi_n^{[2m+1]} = \frac{1}{n} \sum_{k=1}^{n} \Psi_{k-1}^{[2m+1]} + \underbrace{\frac{1}{n} \sum_{k=1}^{n} \sum_{i=0}^{2m} \binom{2m+1}{i} \Psi_{k-1}^{[i]} (\Delta_{n,k}^{\Psi})^{2m+1-i}}_{b_n :=},$$

where $\Delta_{n,k} = 1 - q_n + q_{k-1}$. We consider again two parts according to whether i is even or not:

$$b_n = \underbrace{\frac{1}{n} \sum_{i=0}^{m} \sum_{k=1}^{n} \binom{2m+1}{2i} \Psi_{k-1}^{[2i]} (\Delta_{n,k}^{\Psi})^{2m+1-2i}}_{\alpha_n :=}$$

$$+ \underbrace{\frac{1}{n} \sum_{i=1}^{m} \sum_{k=1}^{n} \binom{2m+1}{2i-1} \Psi_{k-1}^{[2i-1]} (\Delta_{n,k}^{\Psi})^{2(m-i)+2}}_{\beta_n :=}.$$

First, we will treat α_n. Similarly as in Case 1, we only have to consider the term $i = m$. Hence by using the induction hypothesis, we have

$$\alpha_n \sim \frac{1}{n} \sum_{k=2}^{n} \binom{2m+1}{2m} \frac{(2m)!}{2^m m!} \log^m(k-1)\,(1 - \log n + \log(k-1)).$$

Again by integral approximation, we have

$$\alpha_n \sim \binom{2m+1}{2m} \frac{(2m)!}{2^m m!} \log^m n \int_0^1 (1 - \log x) dx = o\left(\log^m n\right).$$

For β_n, we can argue similarly:

$$\beta_n \sim \frac{1}{n} \sum_{k=1}^n \binom{2m+1}{2m-1} \Psi_{k-1}^{[2m-1]} (\Delta_{n,k}^{\Psi})^2.$$

By using induction hypothesis, we have

$$\beta_n = \frac{1}{n} \sum_{k=2}^n \binom{2m+1}{2m-1} o\left(\log^{m-\frac{1}{2}}(k-1)\right) (\Delta_{n,k}^{\Psi})^2 = o\left(\log^{m-\frac{1}{2}} n\right).$$

Therefore, we know

$$b_n = \alpha_n + \beta_n = o\left(\log^{m-\frac{1}{2}} n\right).$$

By Lemma 1,

$$\Psi_n^{[2m+1]} = o\left(\log^{m-\frac{1}{2}} n\right) + \sum_{k=2}^n \frac{1}{k} o\left(\log^{m-\frac{1}{2}}(k-1)\right)$$

$$= o\left(\log^{m+\frac{1}{2}} n\right).$$

Hence, as $n \to \infty$,

$$\mathbb{E}\left(\left(\frac{B_n^{[L]} - \log n}{\sqrt{\log n}}\right)^{2m+1}\right) = o\left(1\right).$$

Similarly, as $n \to \infty$,

$$\mathbb{E}\left(\left(\frac{C_n^{[L]} - \log n}{\sqrt{\log n}}\right)^r\right) = \begin{cases} \frac{(2m)!}{2^m m!} + o\left(1\right), & \text{if } r = 2m, \\ o\left(1\right), & \text{if } r = 2m+1. \end{cases} \quad \blacksquare$$

Lemma 6. *The r-th centralized and normalized moment of $A_n^{[L]}$ satisfies, as $n \to \infty$,*

$$\mathbb{E}\left(\left(\frac{A_n^{[L]} - 2\log n}{\sqrt{2\log n}}\right)^r\right) = \begin{cases} \frac{(2m)!}{2^m m!} + o\left(1\right), & \text{if } r = 2m, \\ o\left(1\right), & \text{if } r = 2m+1. \end{cases} \quad (2.20)$$

Case 1: If $r = 2m$, then

$$\Phi_n^{[2m]} = \frac{1}{n} \sum_{i_1+i_2+i_3=2m} \sum_{k=1}^{n} \binom{2m}{i_1, i_2, i_3} \Psi_{k-1}^{[i_1]} \Omega_{n-k}^{[i_2]} (\Delta_{n,k}^\Phi)^{i_3},$$

where $\Delta_{n,k} = q_{k-1} + t_{n-k} - p_n$.

Then, we consider three parts according to whether $i_1, i_2 \in$ odd, $i_3 = 0$ or $i_1, i_2 \in$ even, $i_3 = 0$ or $i_3 \neq 0$.

$$\Phi_n^{[2m]} = \frac{1}{n} \underbrace{\sum_{i_1+i_2=m} \sum_{k=1}^{n} \binom{2m}{2i_1} \Psi_{k-1}^{[2i_1]} \Omega_{n-k}^{[2i_2]}}_{\alpha_n :=}$$

$$+ \frac{1}{n} \underbrace{\sum_{i_1+i_2=m-1} \sum_{k=1}^{n} \binom{2m}{2i_1+1} \Psi_{k-1}^{[2i_1+1]} \Omega_{n-k}^{[2i_2+1]}}_{\beta_n :=}$$

$$+ \frac{1}{n} \underbrace{\sum_{\substack{i_1+i_2+i_3=2m \\ i_3 \neq 0}} \sum_{k=1}^{n} \binom{2m}{i_1, i_2, i_3} \Psi_{k-1}^{[i_1]} \Omega_{n-k}^{[i_2]} (\Delta_{n,k}^\Phi)^{i_3}}_{\gamma_n :=}.$$

First, we consider α_n

$$\alpha_n = \frac{1}{n} \sum_{i_1+i_2=m} \sum_{k=1}^{n} \binom{2m}{2i_1} \Psi_{k-1}^{[2i_1]} \Omega_{n-k}^{[2i_2]}.$$

By Lemma 5, we have

$$\alpha_n \sim \frac{1}{n} \sum_{i_1+i_2=m} \sum_{k=1}^{n} \binom{2m}{2i_1} \left(\frac{(2i_1)!}{2^{i_1} i_1!} \log^{i_1}(k-1) \right) \left(\frac{(2i_2)!}{2^{i_2} i_2!} \log^{i_2}(n-k) \right).$$

Then, we have

$$\alpha_n \sim \sum_{i_1+i_2=m} \binom{2m}{2i_1} \frac{(2i_1)!}{2_1^i i_1!} \frac{(2i_2)!}{2_2^i i_2!} \log^m n$$

$$= \sum_{i=0}^{m} \frac{(2m)!}{(2m-2i)!(2i)!} \frac{(2i)!}{2^i i!} \frac{(2m-2i)!}{2^{m-i}(m-i)!} \log^m n$$

$$= \frac{(2m)!}{2^m} \log^m n \sum_{i=0}^{m} \frac{1}{i!(m-i)!}$$

28

$$= \frac{(2m)!}{2^m} \log^m n \frac{2^m}{m!}$$

$$= \frac{(2m)!}{2^m m!} 2^m \log^m n.$$

Now, we consider β_n and γ_n,

$$\beta_n = \frac{1}{n} \sum_{i_1+i_2=m-1} \sum_{k=1}^{n} \binom{2m}{2i_1+1} \Psi_{k-1}^{[2i_1+1]} \Omega_{n-k}^{[2i_2+1]}$$

$$= \frac{1}{n} \sum_{i_1+i_2=m-1} \sum_{k=2}^{n-1} \binom{2m}{2i_1+1} o\left(\log^{i_1+1/2}(k-1) \log^{i_2+1/2}(n-k)\right)$$

$$= o\left(\log^m n\right)$$

For γ_n, $i_1 + i_2 = 2m - 1$ gives the main contribution. Consequently,

$$\gamma_n = \frac{1}{n} \sum_{\substack{i_1+i_2+i_3=2m \\ i_3 \neq 0}} \sum_{k=1}^{n} \binom{2m}{i_1, i_2, i_3} \Psi_{k-1}^{[i_1]} \Omega_{n-k}^{[i_2]} (\Delta_{n,k}^{\Phi})^{i_3}$$

$$= \frac{1}{n} \sum_{\substack{i_1+i_2+i_3=2m \\ i_3 \neq 0}} \sum_{k=2}^{n-1} \binom{2m}{i_1, i_2, i_3} \mathcal{O}\left(\log^{i_1/2}(k-1) \log^{i_2/2}(n-k)\right) (\Delta_{n,k}^{\Phi})^{i_3}$$

$$= \mathcal{O}\left(\log^{m-\frac{1}{2}} n\right).$$

Therefore,

$$\Phi_n^{[2m]} = \alpha_n + \beta_n + \gamma_n \sim \frac{(2m)!}{2^m m!} 2^m \log^m n.$$

Hence,

$$\mathbb{E}\left(\left(\frac{A_n^{[L]} - 2\log n}{\sqrt{2\log n}}\right)^{2m}\right) = \frac{(2m)!}{2^m m!} + o(1).$$

Case 2: If $r = 2m + 1$,

$$\Phi_n^{[2m+1]} = \frac{1}{n} \sum_{i_1+i_2+i_3=2m+1} \sum_{k=1}^{n} \binom{2m+1}{i_1, i_2, i_3} \Psi_{k-1}^{[i_1]} \Omega_{n-k}^{[i_2]} (\Delta_{n,k}^{\Phi})^{i_3},$$

where $\Delta_{n,k} = q_{k-1} + t_{n-k} - p_n$. Now, we consider two parts according to whether $i_3 = 0$ or not,

29

$$\Phi_n^{[2m+1]} = \underbrace{\frac{1}{n} \sum_{i_1+i_2=2m+1} \sum_{k=1}^{n} \binom{2m+1}{i_1} \Psi_{k-1}^{[i_1]} \Omega_{n-k}^{[i_2]}}_{\alpha_n :=}$$

$$+ \underbrace{\frac{1}{n} \sum_{\substack{i_1+i_2+i_3=2m+1 \\ i_3 \neq 0}} \sum_{k=1}^{n} \binom{2m+1}{i_1, i_2, i_3} \Psi_{k-1}^{[i_1]} \Omega_{n-k}^{[i_2]} (\Delta_{n,k}^{\Phi})^{i_3}}_{\beta_n :=}.$$

First let us consider α_n. By Lemma 5 we have the following equation:

$$\alpha_n \sim \frac{1}{n} \sum_{i_1+i_2=2m+1} \sum_{k=2}^{n-1} \binom{2m+1}{i_1} o\left(\log^{i_1/2}(k-1)\log^{i_2/2}(n-k)\right)$$

$$= o\left(\log^{m+\frac{1}{2}}(n)\right).$$

For β_n, the main contribution is given by $i_1 + i_2 = 2m$. Consequently,

$$\beta_n = \frac{1}{n} \sum_{\substack{i_1+i_2+i_3=2m+1 \\ i_3 \neq 0}} \sum_{k=1}^{n} \binom{2m+1}{i_1, i_2, i_3} \Psi_{k-1}^{[i_1]} \Omega_{n-k}^{[i_2]} (\Delta_{n,k}^{\Phi})^{i_3}$$

$$= \frac{1}{n} \sum_{\substack{i_1+i_2+i_3=2m+1 \\ i_3 \neq 0}} \sum_{k=1}^{n} \binom{2m+1}{i_1, i_2, i_3} \mathcal{O}\left(\log^{i_1/2}(k-1)\log^{i_2/2}(n-k)\right) (\Delta_{n,k}^{\Phi})^{i_3}$$

$$= \mathcal{O}\left(\log^m n\right).$$

Therefore,

$$\Phi_n^{[2m+1]} = \alpha_n + \beta_n = o\left(\log^{m+\frac{1}{2}}(n)\right).$$

Then,

$$\mathbb{E}\left(\left(\frac{A_n^{[L]} - 2\log n}{\sqrt{2\log n}}\right)^{2m+1}\right) = o(1) \quad \text{as } n \to \infty.$$

By equation (1.1), this concludes the proof. ∎

The proof of Theorem 2 now follows from the last proposition together with Theorem 1 from Chapter 1.

Chapter 3

Number of key comparisons when inserting a random node

3.1 Statement of the result

We will use once more the approach introduced in the last chapter to re-derive the result on the number of key comparisons when inserting a random node from Section 1.2.2. As explained before, our proof will be structurally easier than the previous proof.

Now, let $A_n^{[I]}$, $B_n^{[I]}$ and $C_n^{[I]}$ be the random variables that count the number of key comparisons when inserting a randomly chosen element p in $\{j + \frac{1}{2} : 0 \leq j \leq n\}$ into a random size-n tree of the families $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$. Moreover, let I be the first element in the random permutation (where $+\infty$ and $-\infty$ do not count in model $\mathcal{B}$ and $\mathcal{C}$, respectively) and denote by $\tilde{A}$ (or $\tilde{B}$ or $\tilde{C}$) the tree under consideration. Finally, let $\tilde{B}$ and $\tilde{C}$ (or $\tilde{A}$ and $\tilde{C}$ or $\tilde{A}$ and $\tilde{B}$) be the trees in the decomposition from Section 1.2.1. Then, we have the following distributional recurrences,

$$A_n^{[I]} \overset{d}{=} \begin{cases} C_{n-k}^{[I]}, & \text{if } I = k \text{ and } p \in \tilde{C}, \\ B_{k-1}^{[I]} + L_{n-k}^{(c)}, & \text{if } I = k \text{ and } p \in \tilde{B}. \end{cases}$$

$$B_n^{[I]} \overset{d}{=} \begin{cases} A_{n-k}^{[I]} + 2, & \text{if } I = k \text{ and } p \in \tilde{A}, \\ B_{k-1}^{[I]} + 1, & \text{if } I = k \text{ and } p \in \tilde{B}. \end{cases} \tag{3.1}$$

$$C_n^{[I]} \stackrel{d}{=} \begin{cases} C_{n-k}^{[I]}, & \text{if } I = k \text{ and } p \in \tilde{C}, \\ A_{k-1}^{[I]} + L_{n-k}^{(c)} + 2, & \text{if } I = k \text{ and } p \in \tilde{A}, \end{cases}$$

with initial conditions $A_0^{[I]} = 0$, $B_0^{[I]} = C_0^{[I]} = 1$, where $L_n^{(c)} = $ left path of $\tilde{C}$ and the distribution of I_n is given by

$$\mathbb{P}\left(I = j\right) = \frac{1}{n}, \quad \text{for all } 1 \leq j \leq n.$$

We will again use this to prove the following result.

Theorem 3. *The number of key comparison, $A_n^{[I]}$, in a random size-n priority tree is asymptotically Gaussian distributed:*

$$\frac{A_n^{[I]} - \mathbb{E}\left(A_n^{[I]}\right)}{\sqrt{\mathrm{Var}\left(A_n^{[I]}\right)}} \xrightarrow{d} N(0,1).$$

Moreover, the mean value and variance satisfy

$$\mathbb{E}\left(A_n^{[I]}\right) \sim \frac{1}{3}\log^2 n, \quad \mathrm{Var}\left(A_n^{[I]}\right) \sim \frac{10}{81}\log^3 n.$$

3.2   Proof of the result - Theorem 3

By the distributional recurrences from Section 3.1, we get

$$\mathbb{E}\left(e^{A_n^{[I]}v}\right) = \sum_{k=1}^{n} \mathbb{E}\left(e^{C_{n-k}^{[I]}v}\right)\frac{1}{n}\frac{n-k+1}{n+1} + \sum_{k=1}^{n}\mathbb{E}\left(e^{B_{k-1}^{[I]}v}\right)\mathbb{E}\left(e^{L_{n-k}^{(c)}v}\right)\frac{1}{n}\frac{k}{n+1} \quad \forall n \geq 1,$$

(3.2)

$$\mathbb{E}\left(e^{B_n^{[I]}v}\right) = \sum_{k=1}^{n} e^{2v}\mathbb{E}\left(e^{A_{n-k}^{[I]}v}\right)\frac{1}{n}\frac{n-k+1}{n+1} + \sum_{k=1}^{n} e^{v}\mathbb{E}\left(e^{B_{k-1}^{[I]}v}\right)\frac{1}{n}\frac{k}{n+1} \quad \forall n \geq 1, \quad (3.3)$$

$$\mathbb{E}\left(e^{C_n^{[I]}v}\right) = \sum_{k=1}^{n} \mathbb{E}\left(e^{C_{n-k}^{[I]}v}\right)\frac{1}{n}\frac{n-k+1}{n+1} + \sum_{k=1}^{n} e^{2v}\mathbb{E}\left(e^{A_{k-1}^{[I]}v}\right)\mathbb{E}\left(e^{L_{n-k}^{(c)}v}\right)\frac{1}{n}\frac{k}{n+1}$$

$$\forall n \geq 1, \quad (3.4)$$

with initial conditions $\mathbb{E}\left(e^{A_0^{[I]}v}\right) = 1$, $\mathbb{E}\left(e^{B_0^{[I]}v}\right) = \mathbb{E}\left(e^{C_0^{[I]}v}\right) = e^v$ and $\mathbb{E}\left(e^{L_0^{(c)}}\right) = 1$.

Let $A(z,v)$, $B(z,v)$ and $C(z,v)$ be the generating functions of $A_n^{[I]}$, $B_n^{[I]}$ and $C_n^{[I]}$,

$$A(z,v) := \sum_{n \geq 0}(n+1)\mathbb{E}\left(e^{A_n^{[I]}v}\right)z^n,$$

$$B(z,v) := \sum_{n \geq 0}(n+1)\mathbb{E}\left(e^{B_n^{[I]}v}\right)z^n,$$

$$C(z,v) := \sum_{n \geq 0}(n+1)\mathbb{E}\left(e^{C_n^{[I]}v}\right)z^n.$$

By (3.2), (3.3) and (3.4), we get the following system of linear differential equations:

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

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

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

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

This yields the following homogeneous third order linear differential equation for $A(z,v)$,

$$\frac{\partial^3}{\partial z^3}A(z,v) - \frac{3+2e^v}{1-z}\frac{\partial^2}{\partial z^2}A(z,v) + 2e^v\left(\frac{2}{(1-z)^2} - \frac{e^v}{(1-z)^{e^v+1}}\right)\frac{\partial}{\partial z}A(z,v)$$
$$+ \frac{2e^{2v}}{(1-z)^{e^v+2}}A(z,v) = 0,$$

$$(3.5)$$

with initial conditions $A(0,v) = 1$, $\frac{\partial}{\partial z}A(z,v)\big|_{z=0} = 2e^v$, $\frac{\partial^2}{\partial z^2}A(z,v)\big|_{z=0} = 2e^v + 4e^{2v}$.
Our goal is to investigate $\mathbb{E}\left(e^{A_n^{[I]}t}\right)$. Hence let

$$a_n(v) := [z^n]A(z,v) = [z^n]\sum_{n \geq 0}(n+1)\mathbb{E}\left(e^{A_n^{[I]}v}\right)z^n = (n+1)\mathbb{E}\left(e^{A_n^{[I]}v}\right).$$

We can rewrite equation (3.5) to

$$\frac{\partial^3}{\partial z^3}A(z,v) = \frac{3+2e^v}{1-z}\frac{\partial^2}{\partial z^2}A(z,v) - 2e^v\frac{2}{(1-z)^2}\frac{\partial}{\partial z}A(z,v) + \frac{\partial^3}{\partial z^3}g(z,v), \qquad (3.6)$$

33

where $\frac{\partial^3}{\partial z^3} g(z,v) := \frac{2e^{2v}}{(1-z)^{e^v+1}} \frac{\partial}{\partial z} A(z,v) - \frac{2e^{2v}}{(1-z)^{e^v+2}} A(z,v)$.

Because equation (3.6) is a Cauchy Euler differential equations, by [3] we have the recurrence,

$$a_n(v) = \sum_{k=0}^{n-1} \sum_{j=1}^{2} c_j(v) \frac{j!}{3!} \frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} a_k(v) + b_n, \qquad (3.7)$$

where $c_1(v) = -4e^v, c_2(v) = 3 + 2e^v$ and $b_n := [z^n] g(z,v)$. Hence,

$$a_n(v) = \sum_{k=0}^{n-1} \left(\sum_{j=1}^{2} c_j(v) \frac{j!}{3!} \frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} + \frac{2e^{2v}}{3!} \frac{\left(\binom{e^v+n-k-2}{n-k-2}k - \binom{e^v+n-k-2}{n-k-3}\right)}{\binom{n}{3}} \right) a_k(v). \qquad (3.8)$$

Next, we will prove that r-th moments are polynomial in $\log n$. Therefore, we need to differentiate equation (3.8) r times with respect to v. In order to do so, we first proof the following lemma.

Lemma 7. *We have,*

$$\frac{d^m}{dt^m} \binom{e^t+n}{n-1} \bigg|_{t=0} = \frac{(n+1)n}{2} \left(\log^m n + \mathfrak{Pol}_{m-1}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right) \right), \qquad (3.9)$$

and

$$\frac{d^m}{dt^m} \binom{e^t+n}{n} \bigg|_{t=0} = (n+1) \left(\log^m n + \mathfrak{Pol}_{m-1}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right) \right). \qquad (3.10)$$

Hence, we also have

$$\frac{d^m}{dt^m} e^{2t} \binom{e^t+n}{n-1} \bigg|_{t=0} = \frac{(n+1)n}{2} \left(\log^m n + \mathfrak{Pol}_{m-1}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right) \right),$$

and

$$\frac{d^m}{dt^m} e^{2t} \binom{e^t+n}{n} \bigg|_{t=0} = (n+1) \left(\log^m n + \mathfrak{Pol}_{m-1}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right) \right).$$

Proof. First consider equation (3.9). We use induction on m. For, $m = 1$, we have

$$\frac{d}{dt} \binom{e^t+n}{n-1} \bigg|_{t=0} = \frac{d}{dt} \frac{(e^t+n)(e^t+n-1)\cdots(e^t+2)}{(n-1)!} \bigg|_{t=0}$$

$$= \frac{(e^t + n)(e^t + n - 1)\cdots(e^t + 2)}{(n-1)!} \sum_{i=2}^{n} \frac{e^t}{e^t + i}\Bigg|_{t=0}$$

$$= \frac{(n+1)n}{2} \sum_{i=2}^{n} \frac{1}{1+i}$$

$$= \frac{(n+1)n}{2} \left(\log n + \mathfrak{Pol}_0(\log n) + \mathcal{O}\left(\frac{1}{n}\right) \right).$$

Therefore, $m = 1$ is true. Now, we assume that the claim with exponent less than m is true.

$$\frac{d^m}{dt^m}\left(\begin{array}{c} e^t + n \\ n - 1 \end{array}\right)\Bigg|_{t=0} = \frac{d^m}{dt^m} \frac{(e^t + n)(e^t + n - 1)\cdots(e^t + 2)}{(n-1)!}\Bigg|_{t=0}$$

$$= \frac{d^{m-1}}{dt^{m-1}} \frac{(e^t + n)(e^t + n - 1)\cdots(e^t + 2)}{(n-1)!} \sum_{i=2}^{n} \frac{e^t}{e^t + i}\Bigg|_{t=0}$$

$$= \sum_{j=0}^{m-1} \binom{m-1}{j} \left(\frac{(e^t + n)(e^t + n - 1)\cdots(e^t + 2)}{(n-1)!} \right)^{(j)} \left(\sum_{i=2}^{n} \frac{e^t}{e^t + i} \right)^{(m-1-j)}\Bigg|_{t=0}$$

$$= \frac{(n+1)n}{2} \left(\sum_{i=2}^{n} \frac{e^t}{e^t + i} \right)^{m-1}$$

$$+ \underbrace{\sum_{j=1}^{m-1} \binom{m-1}{j} \left(\begin{array}{c} e^t + n \\ n - 1 \end{array} \right)^{(j)} \left(\sum_{i=2}^{n} \frac{e^t}{e^t + i} \right)^{(m-1-j)}}_{:=T_j}\Bigg|_{t=0}.$$

By the induction hypothesis,

$$\left(\begin{array}{c} e^t + n \\ n - 1 \end{array}\right)^{(j)}\Bigg|_{t=0} = \frac{(n+1)n}{2} \left(\log^j n + \mathfrak{Pol}_{j-1}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right) \right),$$

and we can use induction to get the following equation,

$$\left(\sum_{i=2}^{n} \frac{e^t}{e^t + i} \right)^{(m-j)}\Bigg|_{t=0} = \log n + \mathfrak{Pol}_0(\log n) + \mathcal{O}\left(\frac{1}{n}\right) \quad \text{for all } j = 0 \ldots m - 1.$$

Hence, we know T_{m-1} gives the main distribution. Then

$$\frac{d^m}{dt^m}\binom{e^t+n}{n-1}\bigg|_{t=0} = \frac{(n+1)n}{2}\left(\log^m n + \mathfrak{Pol}_{m-1}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right)\right).$$

Similarly, we can prove that

$$\frac{d^m}{dt^m}\binom{e^t+n}{n}\bigg|_{t=0} = (n+1)\left(\log^m n + \mathfrak{Pol}_{m-1}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right)\right).$$

We can use the binomial expansion to get the following equation,

$$\frac{d^m}{dt^m}e^{2t}\binom{e^t+n}{n-1}\bigg|_{t=0} = \sum_{i=0}^m \binom{m}{i}2^{m-i}\binom{e^t+n}{n-1}^{(i)}\bigg|_{t=0}$$
$$= \frac{(n+1)n}{2}\left(\log^m n + \mathfrak{Pol}_{m-1}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right)\right).$$

Similarly, we can prove that

$$\frac{d^m}{dt^m}e^{2t}\binom{e^t+n}{n}\bigg|_{t=0} = (n+1)\left(\log^m n + \mathfrak{Pol}_{m-1}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right)\right). \quad \blacksquare$$

Let $a_n^{[r]} = \frac{1}{n+1}\frac{d^r}{dt^r}a_n(v)\big|_{v=0}$. Differentiating equation (3.8) r times with respect to v and evaluating at $v = 0$ leads then to the following recurrence,

$$(n+1)a_n^{[r]} = \sum_{k=0}^{n-1}\left(\sum_{j=0}^2 \pi_j\frac{j!}{3!}\frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}}\right)(k+1)a_k^{[r]} + b_n, \quad (3.11)$$

where $\pi_0 = -2$, $\pi_1 = -2$, $\pi_2 = 5$, and

$$b_n = \sum_{k=0}^{n-1}\sum_{i=1}^r \binom{r}{i}\left(\sum_{j=1}^2 c_j(v)\frac{j!}{3!}\frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} + \frac{2e^{2v}}{3!}\frac{\left(\binom{e^v+n-k-2}{n-k-2}k - \binom{e^t+n-k-2}{n-k-3}\right)}{\binom{n}{3}}\right)^{(i)}\bigg|_{t=0}$$
$$(k+1)a_k^{[r-i]}.$$

Lemma 8. *The mean value of $A_n^{[I]}$ satisfies, as $n \to \infty$,*

$$\mathbb{E}\left(A_n^{[I]}\right) = \frac{1}{3}\log^2 n + \mathfrak{Pol}_1(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right),$$

where ϵ denotes an arbitrarily small constant $\epsilon > 0$.

36

Proof. Let $r = 1$ in equation (3.11),

$$(n+1)a_n^{[1]} = \sum_{k=0}^{n-1} \left(\sum_{j=0}^{2} \pi_j \frac{j!}{3!} \frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} \right)(k+1)a_k^{[1]} + b_n,$$

where

$$b_n = \sum_{k=0}^{n-1} \left(\sum_{j=1}^{2} c_j(v) \frac{j!}{3!} \frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} + \frac{2e^{2v}}{3!} \frac{\left(\binom{e^v+n-k-2}{n-k-2}k - \binom{e^v+n-k-2}{n-k-3}\right)}{\binom{n}{3}} \right)^{(1)} \Bigg|_{v=0} (k+1).$$

Now, we divide b_n into two parts,

$$b_n = \alpha_n + \beta_n,$$

where

$$\alpha_n = \sum_{k=0}^{n-1} \left(\sum_{j=1}^{2} c_j(v) \frac{j!}{3!} \frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} \right)^{(1)} \Bigg|_{v=0} (k+1),$$

and

$$\beta_n = \sum_{k=0}^{n-1} \left(\frac{2e^{2v}}{3!} \frac{\left(\binom{e^v+n-k-2}{n-k-2}k - \binom{e^v+n-k-2}{n-k-3}\right)}{\binom{n}{3}} \right)^{(1)} \Bigg|_{v=0} (k+1).$$

For α_n, by Euler-Maclaurin summation formula,

$$\alpha_n = n \left(c + \mathcal{O}\left(\frac{1}{n}\right) \right),$$

where c is some constant.

For β_n, by Lemma 7,

$$\beta_n = \sum_{k=0}^{n-1} \frac{2}{3!\binom{n}{3}} \left(k(n-k-1) \left(\log(n-k-2) + c_1 + \mathcal{O}\left(\frac{1}{(n-k-2)^{1-\epsilon}}\right) \right) \right.$$
$$\left. - \frac{(n-k-1)(n-k-2)}{2} \left(\log(n-k-2) + c_2 + \mathcal{O}\left(\frac{1}{(n-k-2)^{1-\epsilon}}\right) \right) \right)(k+1).$$

where c_1, c_2 are constants. By Euler-Maclaurin summation formula,

$$\beta_n = n \left(\frac{1}{12} \log^2 n + \mathfrak{Pol}_1(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right) \right).$$

37

Hence,
$$b_n = n\left(\frac{1}{12}\log^2 n + \mathfrak{Pol}_1(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right)\right).$$

By [3], we know
$$(n+1)a_n^{[1]} = n\left(\frac{1}{3}\log^2 n + \mathfrak{Pol}_1(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right)\right).$$

Then,
$$\mathbb{E}\left(A_n^{[I]}\right) = a_n^{[1]} = \frac{1}{3}\log^2 n + \mathfrak{Pol}_1(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right). \quad \blacksquare$$

Lemma 9. *The expansion of the moments satisfy, as $n \to \infty$,*
$$\mathbb{E}\left(\left(A_n^{[I]}\right)^r\right) = \mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right),$$

where ϵ denotes an arbitrarily small constant $\epsilon > 0$.

Proof. We will use induction on r. By Lemma 8, $r = 1$ is true. Now, we assume that the claim with exponents less than r is true. We again divide b_n into two parts
$$b_n = \alpha_n + \beta_n,$$

where
$$\alpha_n = \sum_{k=0}^{n-1}\sum_{i=1}^{r}\binom{r}{i}\left(\sum_{j=1}^{2}c_j(v)\frac{j!}{3!}\frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}}\right)^{(i)}\Bigg|_{v=0}(k+1)a_k^{[r-i]},$$

and
$$\beta_n = \sum_{k=0}^{n-1}\sum_{i=1}^{r}\binom{r}{i}\left(\frac{2e^{2v}}{3!}\frac{\left(\binom{e^v+n-k-2}{n-k-2}k - \binom{e^v+n-k-2}{n-k-3}\right)}{\binom{n}{3}}\right)^{(i)}\Bigg|_{v=0}(k+1)a_k^{[r-i]}.$$

For α_n, by induction hypothesis,
$$\alpha_n = \sum_{k=0}^{n-1}\sum_{i=1}^{r}\binom{r}{i}\left(\sum_{j=1}^{2}c_j(v)\frac{j!}{3!}\frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}}\right)^{(i)}\Bigg|_{v=0}(k+1)\left(\mathfrak{Pol}(\log k) + \mathcal{O}\left(\frac{1}{k^{1-\epsilon}}\right)\right)$$
$$= n\left(\sum_{i=1}^{r}\mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right)\right)$$
$$= n\left(\mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right)\right).$$

For β_n, by induction hypothesis,

$$\beta_n = \sum_{k=0}^{n-1} \sum_{i=1}^{r} \binom{r}{i} \left(\frac{2e^{2v} \left(\binom{v+n-k-2}{n-k-2} k - \binom{v+n-k-2}{n-k-3} \right)}{\binom{n}{3}} \right)^{(i)} \Bigg|_{v=0} (k+1) \left(\mathfrak{Pol}(\log k) + \mathcal{O}\left(\frac{1}{k^{1-\epsilon}} \right) \right).$$

By Lemma 7, we know,

$$\beta_n = \sum_{k=0}^{n-1} \sum_{i=1}^{r} \left(\frac{k(n-k-1) \left(\mathfrak{Pol}((\log n - k - 2)) + \mathcal{O}\left(\frac{1}{(n-k-2)^{1-\epsilon}} \right) \right)}{n(n-1)(n-2)} \right.$$

$$\left. \frac{(n-k-1)(n-k-2) \left(\mathfrak{Pol}(\log(n-k-2)) + \mathcal{O}\left(\frac{1}{(n-k-2)^{1-\epsilon}} \right) \right)}{n(n-1)(n-2)} \right)$$

$$(k+1) \left(\mathfrak{Pol}(\log k) + \mathcal{O}\left(\frac{1}{k^{1-\epsilon}} \right) \right)$$

$$= n \left(\mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}} \right) \right).$$

Hence,

$$b_n = n \left(\mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}} \right) \right).$$

By [3], we know,

$$(n+1)a_n^{[r]} = n \left(\mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}} \right) \right),$$

then,

$$a_n^{[r]} = \mathfrak{Pol}(\log n) + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}} \right). \quad \blacksquare$$

By Lemma 9, we know that the central moments are polynomials in $\log n$ as well. Now, let $\Psi_n(t) := e^{-\mu_n t} \mathbb{E}(e^{A_n^{[I]} t})$, where $\mu_n = \mathbb{E}(A_n^{[I]})$. Then,

$$\Psi_n(t) = \mathbb{E}\left(e^{(A_n^{[I]} - \mu_n)t} \right) = \frac{1}{n+1} a_n(t) e^{-\mu_n t}.$$

Using equation (3.8), we get

$$(n+1)\Psi_n(t) = \sum_{k=0}^{n-1}\left(\sum_{j=1}^{2} c_j(t)\frac{j!}{3!}\frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} + \frac{2e^{2t}}{3!}\frac{\left(\binom{e^t+n-k-2}{n-k-2}k - \binom{e^t+n-k-2}{n-k-3}\right)}{\binom{n}{3}}\right)$$
$$(k+1)\Psi_k(t)e^{(\mu_k-\mu_n)t},$$

$$(3.12)$$

where $c_1(v) = -4e^v$, $c_2(v) = 3+2e^v$.

Let $\Psi_n^{[r]} = \Psi_n^{(r)}(0)$. Then by equation (3.12), we have the following recurrence

$$(n+1)\Psi_n^{[r]} = \sum_{k=0}^{n-1}\left(\sum_{j=0}^{2}\pi_j\frac{j!}{3!}\frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}}\right)(k+1)\Psi_k^{[r]} + b_n, \qquad (3.13)$$

where $\pi_0 = -2$, $\pi_1 = -2$, $\pi_2 = 5$, and

$$b_n = \sum_{k=0}^{n-1}\sum_{\substack{i_1+i_2+i_3=r \\ i_3 \neq r}}\binom{r}{i_1,i_2,i_3}$$

$$\left(\sum_{j=1}^{2} c_j(t)\frac{j!}{3!}\frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} + \frac{2e^{2t}}{3!}\frac{\left(\binom{e^t+n-k-2}{n-k-2}k - \binom{e^t+n-k-2}{n-k-3}\right)}{\binom{n}{3}}\right)^{(i_1)}\Bigg|_{t=0}$$

$$(\mu_k - \mu_n)^{i_2}(k+1)\Psi_k^{[i_3]},$$

where $c_j = c_j(0)$ for $j = 1, 2$.

Lemma 10. *The variance of $A_n^{[I]}$ satisfies, as $n \to \infty$,*

$$\mathrm{Var}(A_n^{[I]}) \sim \frac{10}{81}\log^3 n.$$

Proof. Take $r = 2$ in equation (3.13)

$$b_n = \sum_{k=0}^{n-1}\sum_{\substack{i_1+i_2+i_3=2 \\ i_3 \neq 2}}\binom{r}{i_1,i_2,i_3}$$

$$\left(\sum_{j=1}^{2} c_j(t)\frac{j!}{3!}\frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} + \frac{2e^{2t}}{3!}\frac{\left(\binom{e^t+n-k-2}{n-k-2}k - \binom{e^t+n-k-2}{n-k-3}\right)}{\binom{n}{3}}\right)^{(i_1)}\Bigg|_{t=0}$$

$$(\mu_k - \mu_n)^{i_2}(k+1)\Psi_k^{[i_3]}.$$

40

We divide b_n into two parts,

$$b_n = \alpha_n + \beta_n$$

where

$$\alpha_n = \sum_{\substack{i_1+i_2+i_3=2 \\ i_3 \neq 2}} \binom{2}{i_1, i_2, i_3} \underbrace{\sum_{k=0}^{n-1} \left(\sum_{j=1}^{2} c_j(t) \frac{j!}{3!} \frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} \right)^{(i_1)}\Bigg|_{t=0} (\mu_k - \mu_n)^{i_2}(k+1)\Psi_k^{[i_3]}}_{:=P_{i_1,i_2,i_3}^{\alpha_n}},$$

and

$$\beta_n = \sum_{\substack{i_1+i_2+i_3=2 \\ i_3 \neq 2}} \binom{2}{i_1, i_2, i_3}$$

$$\underbrace{\sum_{k=0}^{n-1} \left(\frac{2e^{2t}}{3!} \frac{\left(\binom{e^t+n-k-2}{n-k-2}k - \binom{e^t+n-k-2}{n-k-3}\right)}{\binom{n}{3}} \right)^{(i_1)}\Bigg|_{t=0} (\mu_k - \mu_n)^{i_2}(k+1)\Psi_k^{[i_3]}}_{:=P_{i_1,i_2,i_3}^{\beta_n}}.$$

For α_n, we first consider $i_2 = 2$ and $i_3 = 0$,

$$P_{0,2,0}^{\alpha_n} \sim \sum_{k=0}^{n-1} \sum_{j=1}^{2} c_j(0) \frac{j!}{3!} \frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} \left(\frac{1}{3}\log^2 k - \frac{1}{3}\log^2 n \right)^2 (k+1)$$

$$\sim -\frac{13}{1944} n \log^2 n.$$

For the other terms in α_n, we can use the same idea to obtain the bound $\mathcal{O}(n \log n)$.
Therefore,

$$\alpha_n \sim -\frac{13}{1944} n \log^2 n.$$

For β_n, we first consider $i_3 = 0$,

$$\sum_{i_1+i_2=2} \binom{2}{i_1, i_2} P_{i_1,i_2,0}^{\beta_n} \sim \sum_{i_1+i_2=2} \binom{2}{i_1, i_2} \sum_{k=0}^{n-1} \left(\frac{2e^{2t}}{3!} \frac{\left(\binom{e^t+n-k-2}{n-k-2}k - \binom{e^t+n-k-2}{n-k-3}\right)}{\binom{n}{3}} \right)^{(i_1)}\Bigg|_{t=0}$$

$$\left(\frac{1}{3}\log^2 k - \frac{1}{3}\log^2 n \right)^{i_2} (k+1)$$

$$\sim -\frac{41}{1944} n \log^2 n - \frac{1}{108} n \log^2 n + \frac{1}{12} n \log^2 n$$

$$= \frac{103}{1944} n \log^2 n.$$

41

For the other terms in β_n, we can also use the same idea to obtain the bound $\mathcal{O}\left(n\log n\right)$. Therefore,

$$\beta_n \sim \frac{103}{1944}n\log^2 n.$$

Hence,

$$b_n \sim -\frac{13}{1944}n\log^2 n + \frac{103}{1944}n\log^2 n = \frac{5}{108}n\log^2 n$$

Then by [3], we know

$$\mathrm{Var}(A_n^{[I]}) \sim \frac{10}{81}\log^3 n. \quad \blacksquare$$

Lemma 11. *The r-th centralized and normalized moments of $A_n^{[I]}$ satisfy, as $n \to \infty$,*

$$\mathbb{E}\left(\left(\frac{A_n^{[I]} - \frac{1}{3}\log^2 n}{\sqrt{\frac{10}{81}\log^3 n}}\right)^r\right) = \begin{cases} \frac{(2m)!}{2^m m!} + o(1), & \text{if } r = 2m, \\ o(1), & \text{if } r = 2m+1. \end{cases} \quad (3.14)$$

Proof. We use induction on r. By Lemma 8 and Lemma 10, we know that the claim holds for $r = 1$ and $r = 2$. Now, we assume that the claim with exponent less than r is true. We divide b_n of equation (3.13) into two parts,

$$b_n = \alpha_n + \beta_n,$$

where,

$$\alpha_n = \sum_{k=0}^{n-1}\sum_{\substack{i_1+i_2+i_3=r \\ i_3 \neq r}}\binom{r}{i_1,i_2,i_3}\left(\sum_{j=1}^{2}c_j(t)\frac{j!\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}}\right)^{(i_1)}\Bigg|_{t=0}(\mu_k - \mu_n)^{i_2}(k+1)\Psi_k^{[i_3]},$$

and

$$\beta_n = \sum_{k=0}^{n-1}\sum_{\substack{i_1+i_2+i_3=r \\ i_3 \neq r}}\binom{r}{i_1,i_2,i_3}\left(\frac{2e^t}{3!}\frac{\left(\binom{e^t+n-k-2}{n-k-2}k - \binom{e^t+n-k-2}{n-k-3}\right)}{\binom{n}{3}}\right)^{(i_1)}\Bigg|_{t=0}$$
$$(\mu_k - \mu_n)^{i_2}(k+1)\Psi_{(}k+1)^{[i_3]}.$$

Then we will consider two cases according to whether r is even or not.

Case 1: $r = 2m$.

Contribution of α_n

$$\alpha_n = \sum_{\substack{i_1+i_2+i_3=2m \\ i_3\neq 2m}} \binom{2m}{i_1,i_2,i_3} \underbrace{\sum_{k=0}^{n-1}\left(\sum_{j=1}^{2} c_j(t)\frac{j!}{3!}\frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}}\right)^{(i_1)}\Bigg|_{t=0}(\mu_k-\mu_n)^{i_2}(k+1)\Psi_k^{[i_3]}}_{:=T_{i_1,i_2,i_3}^{\alpha_n}}$$

We consider two parts according to whether i_3 is even or not,

$$\alpha_n = \underbrace{x_n^{[\alpha]}}_{\text{even part}} + \underbrace{y_n^{[\alpha]}}_{\text{odd part}},$$

where

$$x_n^{[\alpha]} = \sum_{\substack{i_1+i_2+2i_3=2m \\ i_3\neq m}} \binom{2m}{i_1,i_2,2i_3} T_{i_1,i_2,2i_3}^{\alpha_n},$$

and

$$y_n^{[\alpha]} = \sum_{i_1+i_2+2i_3+1=2m} \binom{2m}{i_1,i_2,2i_3+1} T_{i_1,i_2,2i_3+1}^{\alpha_n}.$$

First, we consider $x_n^{[\alpha]}$. By induction hypothesis we know,

$$T_{i_1,i_2,2i_3}^{\alpha_n} \sim \sum_{k=0}^{n-1}\left(\sum_{j=1}^{2} c_j(t)\frac{j!}{3!}\frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}}\right)^{(i_1)}\Bigg|_{t=0}\left(\frac{1}{3}\log^2 k - \frac{1}{3}\log^2 n\right)^{i_2}$$

$$(k+1)\frac{(2i_3)!}{2^{i_3}(i_3)!}\left(\frac{10}{81}\log^3 k\right)^{i_3}$$

$$\sim c_{i_1,i_2,i_3} \log^{i_2+3i_3} n,$$

where c_{i_1,i_2,i_3} are constants. Therefore, $i_1 = 0$, $i_2 = 2$, and $i_3 = m-1$ gives the main contribution. So,
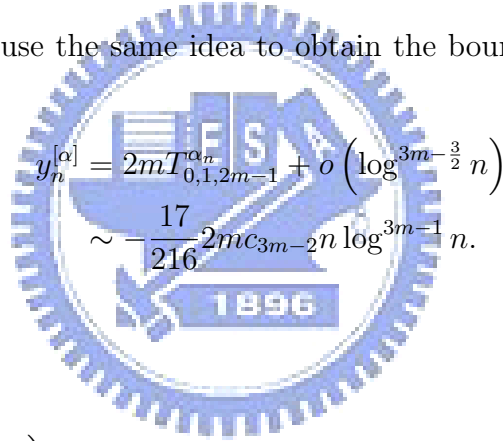
$$x_n^{[\alpha]} \sim \binom{2m}{2, 2m-2} T_{0,2,2m-2}^{\alpha_n}$$

$$\sim \binom{2m}{2, 2m-2} \sum_{k=0}^{n-1} \left(\sum_{j=1}^{2} c_j(t) \frac{j!}{3!} \frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} \right) \left(\frac{1}{3} \log^2 k - \frac{1}{3} \log^2 n \right)^2$$

$$(k+1) \frac{(2i_3)!}{2^{i_3}(i_3)!} \left(\frac{10}{81} \log^3 k \right)^{m-1}$$

$$\sim -\frac{13}{1944} \frac{(2m)!}{2^m(m-1)!} \left(\frac{10}{81} \right)^{m-1} n \log^{3m-1} n.$$

For $y_n^{[\alpha]}$, first, we consider $i_2 = 1$ and $i_3 = m-1$. By Lemma 9, $\Psi_k^{[2m-1]}$ is a polynomial in $\log n$. Therefore by induction hypothesis, we know that $\Psi_k^{[2m-1]} \sim \sum_{i=0}^{3m-2} c_i \log^i(k)$, where the c_i's are suitable constants for all i. Consequently,

$$T_{0,1,2m-1}^{\alpha_n} \sim \sum_{k=0}^{n-1} \left(\sum_{j=1}^{2} c_j(0) \frac{j!}{3!} \frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}} \right) \left(\frac{1}{3} \log^2 k - \frac{1}{3} \log^2 n \right)^1 (k+1) \left(\sum_{i=0}^{3m-2} c_i \log^i(k) \right)$$

$$\sim -\frac{17}{216} c_{3m-2} n \log^{3m-1} n.$$

For $i_3 < m-1$, we can use the same idea to obtain the bound $T_{i_1,i_2,i_3}^{\alpha_n} = o\left(\log^{3m-\frac{3}{2}} n \right)$. Therefore,

$$y_n^{[\alpha]} = 2m T_{0,1,2m-1}^{\alpha_n} + o\left(\log^{3m-\frac{3}{2}} n \right)$$

$$\sim -\frac{17}{216} 2m c_{3m-2} n \log^{3m-1} n.$$

Contribution of β_n

$$\beta_n = \sum_{\substack{i_1+i_2+i_3=2m \\ i_3 \neq 2m}} \binom{2m}{i_1, i_2, i_3}$$

$$\underbrace{\sum_{k=0}^{n-1} \left(\frac{2e^t}{3!} \frac{\left(\binom{e^t+n-k-2}{n-k-2} k - \binom{e^t+n-k-2}{n-k-3} \right)}{\binom{n}{3}} \right)^{(i_1)} \Bigg|_{t=0} (\mu_k - \mu_n)^{i_2} (k+1) \Psi_k^{[i_3]}}_{:=T_{i_1,i_2,i_3}^{\beta_n}}.$$

Now, we also consider two parts according to whether i_3 is even or not,

$$\beta_n = \underbrace{x_n^{[\beta]}}_{\text{even part}} + \underbrace{y_n^{[\beta]}}_{\text{odd part}},$$

44

where

$$x_n^{[\beta]} = \sum_{\substack{i_1+i_2+2i_3=2m \\ i_3 \neq m}} \binom{2m}{i_1,i_2,2i_3} T_{i_1,i_2,2i_3}^{\beta_n},$$

and

$$y_n^{[\beta]} = \sum_{i_1+i_2+2i_3+1=2m} \binom{2m}{i_1,i_2,2i_3} T_{i_1,i_2,2i_3+1}^{\beta_n}.$$

First, we consider $x_n^{[\beta]}$. By induction hypothesis,

$$T_{i_1,i_2,2i_3}^{\beta_n} \sim \sum_{k=0}^{n-1} \left(\frac{2e^t}{3!} \frac{\left(\binom{e^t+n-k-2}{n-k-2} k - \binom{e^t+n-k-2}{n-k-3} \right)}{\binom{n}{3}} \right)^{(i_1)} \Bigg|_{t=0} \left(\frac{1}{3}\log^2 k - \frac{1}{3}\log^2 n \right)^{i_2}$$

$$(k+1)\frac{(2i_3)!}{2^{i_3}(i_3)!} \left(\frac{10}{81}\log^3 k \right)^{i_3}.$$

By Lemma 7 and Euler-Maclaurin summation formula, we know

$$T_{i_1,i_2,2i_3}^{\beta_n} = c_{i_1,i_2,i_3} n \log^{i_1+i_2+3i_3} n,$$

where c_{i_1,i_2,i_3} is a constant. Therefore, $i_3 = m - 1$ gives the main contribution. So,

$$x_n^{[\beta]} \sim \sum_{i_1+i_2=2} \binom{2m}{i_1,i_2} T_{i_1,i_2,2m-2}^{\beta_n}$$

$$\sim 2 \sum_{i_1+i_2=2} \sum_{k=0}^{n-1} \binom{2m}{i_1,i_2} \left(k(n-k-1) - \frac{1}{2}(n-k-1)(n-k-2) \right) \log^{i_1}(n-k-2)$$

$$\left(\frac{1}{3}\log^2 k - \frac{1}{3}\log^2 n \right)^{i_2} (k+1)\frac{(2m-2)!}{2^{m-1}(m-1)!} \left(\frac{10}{81} \right)^{m-1} n \log^{3m-3} n$$

$$\sim \left(\binom{2m}{0,2}\frac{-41}{1944} + \binom{2m}{1,1}\frac{-1}{216} + \binom{2m}{2,0}\frac{1}{12} \right) \frac{(2m-2)!}{2^{m-1}(m-1)!} \left(\frac{10}{81} \right)^{m-1} n \log^{3m-1} n$$

$$= \frac{103}{1944} \frac{(2m)!}{2^m(m-1)!} \left(\frac{10}{81} \right)^{m-1} n \log^{3m-1} n.$$

For $y_n^{[\beta]}$, first we consider $i_3 = m - 1$. By Lemma 9, $\Psi_k^{[2m-1]}$ is polynomial in $\log n$. Therefore by induction hypothesis, we know that $\Psi_k^{[2m-1]} \sim \sum_{i=0}^{3m-2} c_i \log^i(k)$. Consequently,

$$
\sum_{i_1+i_2=1} T_{i_1,i_2,2m-2}^{\beta_n} \sim \sum_{i_1+i_2=1} \sum_{k=0}^{n-1} \binom{2m}{i_1, i_2} \left(k(n-k-1) - \frac{1}{2}(n-k-1)(n-k-2) \right)
$$
$$
\log^{i_1}(n-k-2) \left(\frac{1}{3} \log^2 k - \frac{1}{3} \log^2 n \right)^{i_2} (k+1) n c_{3m-3} \log^{3m-3} n
$$
$$
\sim \frac{17}{216} 2m c_{3m-2} n \log^{3m-1} n.
$$

For $i_3 < m - 1$, we can use the same idea to obtain the bound $T_{i_1,i_2,i_3}^{\beta_n} = o\left(\log^{3m-\frac{5}{2}} n \right)$. Therefore,

$$
y_n^{[\beta]} \sim \frac{17}{216} 2m c_{3m-2} n \log^{3m-1} n.
$$

Hence, we have,

$$
b_n = \alpha_n + \beta_n
$$
$$
\sim -\frac{13}{1944} \frac{(2m)!}{2^m(m-1)!} \left(\frac{10}{81} \right)^{m-1} n \log^{3m-1} n - \frac{17}{216} 2m c_{3m-2} n \log^{3m-1} n
$$
$$
+ \frac{103}{1944} \frac{(2m)!}{2^m(m-1)!} \left(\frac{10}{81} \right)^{m-1} n \log^{3m-1} n + \frac{17}{216} 2m c_{3m-2} n \log^{3m-1} n
$$
$$
= \frac{5}{108} \frac{(2m)!}{2^m(m-1)!} \left(\frac{10}{81} \right)^{m-1} n \log^{3m-1} n.
$$

Case 2: $r = 2m + 1$

Contribution of α_n

Similarly as in Case 1, we only have to consider the term $i_2 = 1$ and $i_3 = 2m$. Hence by induction hypothesis, we have

$$\alpha_n \sim \sum_{k=0}^{n-1}(2m+1)\left(\sum_{j=1}^{2}c_j(t)\frac{j!}{3!}\frac{\binom{k}{j}\binom{n-1-k}{2-j}}{\binom{n}{3}}\right)^{(i_1)}\Bigg|_{t=0}\left(\frac{1}{3}\log^2 k-\frac{1}{3}\log^2 n\right)^{i_2}$$

$$(k+1)\frac{(2m)!}{2^m m!}\left(\frac{10}{81}\log^3 k\right)^m$$

$$\sim -(2m+1)\frac{17}{216}\frac{(2m)!}{2^m m!}n\log^{3m+1}n.$$

Contribution of β_n

We can again argue similarly, $i_3 = m$ gives the main contribution. Consequently,

$$\beta_n \sim \sum_{k=0}^{n-1}\sum_{i_1+i_2=1}\binom{2m+1}{i_1,i_2}\left(\frac{2e^{2t}}{3!}\frac{\left(\binom{e^t+n-k-2}{n-k-2}k-\binom{e^t+n-k-2}{n-k-3}\right)}{\binom{n}{3}}\right)^{(i_1)}\Bigg|_{t=0}$$

$$\left(\frac{1}{3}\log^2 k-\frac{1}{3}\log^2 n\right)^{i_2}(k+1)\frac{(2m)!}{2^m m!}\left(\frac{10}{81}\log^3 k\right)^m$$

$$\sim (2m+1)\frac{17}{216}\frac{(2m)!}{2^m m!}n\log^{3m+1}n.$$

Hence,

$$b_n = \alpha_n + \beta_n$$

$$\sim -(2m+1)\frac{17}{216}\frac{(2m)!}{2^m m!}n\log^{3m+1}n+(2m+1)\frac{17}{216}\frac{(2m)!}{2^m m!}n\log^{3m+1}n$$

$$= o\left(n(\log n)^{3m+1}\right). \tag{3.15}$$

Now, by [3],

$$\Psi_n^{[2m]} = \frac{(2m)!}{2^m m!}\left(\frac{10}{81}\right)^m(\log n)^{3m},$$

and

$$\Psi_n^{[2m+1]} = o\left((\log n)^{3m+2}\right).$$

By Lemma 9, we know that $\Psi_n^{[2m+1]}$ is a polynomial in $\log n$. Hence,

$$\Psi_n^{[2m+1]} = \mathcal{O}\left((\log n)^{3m+1}\right),$$

Therefore, we know, as $n \to \infty$,

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

By equation (1.1), this concludes the proof. ∎

The proof of Theorem 3 follows now from the last result together with Theorem 1 from Chapter 1.

Chapter 4

Conclusion

We conclude this thesis by shortly summarizing the main contributions. As explained in Chapter 1, the method of moments which proceeds by using induction to derive the first order asymptotic of all central moments does not work for some recurrences which are "one-sided". Therefore, we introduced a new variant of the method of moments which uses two induction steps instead of only one. The first induction step studies the non-central moments and is used as input for the second induction for the central moments.

We applied our new approach to re-derive some recent results on the running time of deletion and insertion in random priority trees. In the first case, our method was more complicated than the previous approach. In the second case, our method, however, simplifies the previous proof.

We expect that our method can be used to treat other examples as well. Some examples include cost of unsuccessful search in binary search trees (see H. M. Mahmoud [10]), depths of random nodes in binary search trees (see H. M. Mahmoud [10]), analysis of maximum-finding algorithms for broadcast communication (see W. M. Chen, H. K. Hwang [2]), number of collisions in the $\beta(2, b)$-coalescents (see A.M. Iksanov, A. Marynych, M. Moehle [8]), asymptotic laws for regenerative compositions (see A. Gnedin, J. Pitman, M. Yor [6]), etc.

Bibliography

[1] P. Billingsley, *Probability and Measure*, Third Edition, John Wiley & Sons, New York, 1995.

[2] W. M. Chen, H. K. Hwang, Analysis in distribution of two randomized algorithms for finding the maximum in a broadcast communication model, *J. Algorithms*, **46**, 140-177 (2003).

[3] H. H. Chern, H. K. Hwang, T. H. Tsai, An asymptotic theory for Cauchy-Euler differential equations with applications to the analysis of algorithms, *Journal of Algorithms*, **44:1**, 177-225 (2002).

[4] T. H. Cormen, C. E. Leiserson, and R. L.Rivest, *Introduction to Algorithms*, MIT Press, 2002.

[5] P. Flajolet, R. Sedgewick, *Analytic combinatorics*, Cambridge University Press, 2008.

[6] A. Gnedin, J. Pitman, M. Yor, Asymptotic laws for regenerative compositions: gamma subordinators and the like, *Probab. Theory Relat. Fields*, **135**, 576-602 (2006).

[7] H. K. Hwang, R. Neininger, Phase change of limit laws in the quicksort recurrence under varying toll functions, *SIAM Journal on Computing*, **31:6**, 1687-1722 (2002).

[8] A.M. Iksanov, A. Marynych, M. Moehle, On the number of collisions in beta(2,b)-coalescents, *Bernoulli*, **15**, to appear.

[9] M. Kuba, A. Panholzer. Analysis of insertion costs in priority trees, in: "Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments and the Fourth

Workshop on Analytic Algorithmics and Combinatorics", *SIAM, Philadelphia*, 175-182, (2007).

[10] H. M. Mahmoud, *Evolution of Random Search Trees*, Wiley, New York, 1992

[11] A. Panholzer, Analysis of some parameters for random nodes in priority trees, *Discrete Mathematics and Theoretical Computer Science*, **10(2)**, 1-38 (2008).

[12] A. Panholzer, H. Prodinger, Average-Case Analysis of Priority Trees: A Structure for Priority Queue Administration, *Algorithmica*, **22**, 600-630 (1998).