Discrete Mathematics and Theoretical Computer Science DMTCS vol. **10:2**, 2008, 1–38

Analysis of some parameters for random nodes in priority trees

Alois Panholzer[1†]

[1]*Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstraße 8–10/104, A-1040 Wien, Austria. e-mail:* `Alois.Panholzer@tuwien.ac.at`

received August 30, 2006, *revised November 12, 2007*, *accepted February 2, 2008*.

Priority trees are a certain data structure used for priority queue administration. Under the model that all permutations of the numbers $1, \dots, n$ are equally likely to construct a priority tree of size n we study the following parameters in size-n trees: depth of a random node, number of right edges to a random node, and number of descendants of a random node. For all parameters studied we give limiting distribution results.

Keywords: priority trees, node depth, descendants, limiting distribution

1 Introduction

The data structure of priority trees (also called p-trees) is used to implement priority queues. Priority queues are important for many scheduling problems (e.g., for jobs, events, etc.). Typical applications are given in operation systems (scheduling the execution of different jobs, resource management, etc.) and in discrete event simulation models (scheduling events according to their time of occurrence). To each element in a priority queue there is associated a fixed key, whose value determines its priority. We will here assume that low key values correspond to high priority. Two basic operations are fundamental for priority queue administration: inserting an element with arbitrary priority (INSERT), and removing the element with the highest priority (DELETE).

Priority trees can be considered as a special implementation of binary trees that make them suitable for usage as a priority queue. Formally they are defined as follows: a priority tree either is empty or consists of a sequence of nodes with non-increasing keys, the so called *left path*, such that to 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. More precisely, if x with key k denotes the left successor of z with key l, then all nodes in the right subtree of z have key values s: $k \leq s < l$. $k = s$ can occur as is seen by considering the insertion algorithm defined in the sequel.

[†]This work was supported by the Austrian Science Foundation FWF, grant S9608-N13, and by the ÖAD funded Austrian-French exchange program "Amadée".

Of course it follows from this definition that not every binary tree of size n, where the size of a tree is given as usual by the number of internal nodes, can appear as a priority tree. The number P_n of different size-n priority trees is given by the number of size-n Motzkin-trees (also called unary-binary trees). Although this is not mentioned explicitly in [10] it is a direct consequence of the formula for P_n presented in that work. Priority trees are generated from the empty tree by successively inserting new elements via the recursive algorithm INSERT. Inserting a new element p into a priority tree T works as follows:

- If T is empty 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.

We give in Figure 1 an example of constructing a priority tree with the algorithm INSERT.

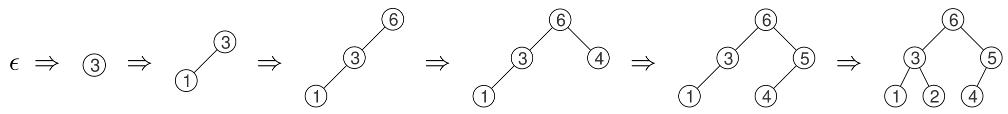

Fig. 1: Construction of a priority tree of size 6 from the keys $3, 1, 6, 4, 5, 2$ starting with the empty tree (denoted by ϵ). The node with key 3 has in the resulting tree exactly 2 ascendants (i.e., there are exactly 2 nodes lying on the path from the root to the node 3; therefore the depth of this node is 2) and 3 descendants (i.e., the size of the subtree with node 3 as root-node is 3). The number of right edges on the path from the root to node 3 is 0.

Priority trees were investigated in the early 1960's as a data structure for maintaining priority queues in the implementation of the programming language SIMULA, which was developed at the Norwegian Computer Center in Oslo by Dahl and Nygaard. A first average-case analysis of the behaviour of priority trees was given by Jonassen and Dahl in [10]. They are also mentioned in handbooks of algorithms and data structure, e.g., in [6], where some of the results obtained in [10] are cited. In the average-case analysis it has been used always the so called *random permutation model* as the underlying probability model. This means that all $n!$ permutations of $\{1, 2, \ldots, n\}$ are assumed to be chosen equally likely as data arrays to construct a priority tree of size n starting with the empty tree using algorithm INSERT repeatedly. As main results in their analysis we cite the expected number of key comparisons I_n and the expected recursion depth R_n when inserting a random element into a random size-n priority tree, and the expected size of the left path L_n in a random size-n priority tree (this gives the depth of the node with the smallest key value, the so called *left leaf*, and thus the depth of the data with the highest priority, where the depth of a node x

in the tree is here always measured by the number of nodes lying on the direct path from the root to x):

$$\mathbb{E}(I_n) = \frac{H_{n+1}^2}{3} + \frac{10H_{n+1}}{9} - \frac{H_{n+1}^{(2)}}{3} - \frac{28}{27}, \quad \text{for } n \geq 2,$$
$$\mathbb{E}(R_n) = \frac{2}{3}H_{n+1} + \frac{1}{9}, \quad \text{for } n \geq 2, \qquad \mathbb{E}(L_n) = 2H_n - 1.$$

Here $H_n := \sum_{k=1}^{n} \frac{1}{k}$ and $H_n^{(2)} := \sum_{k=1}^{n} \frac{1}{k^2}$ denote the first and second order harmonic numbers. Since $\mathbb{E}(I_n) \sim \frac{1}{3}\log^2 n$, it follows that the operation INSERT is for priority trees asymptotically more expensive than for other priority queue structures (like heaps or leftist trees, see, e.g., [6]). We further remark that the dynamic behaviour of priority trees was considered in [9].

The average case analysis of this data structure under the random permutation model has been extended in [13], see also [12]. The focus has been given there in computing exact formulæ for the mean of several tree statistics for nodes in size-n trees with a fixed key j (and not for random nodes). In particular we obtained for size-n priority trees among others exact and asymptotic results for the expected depth of node j ($=$ number of ancestors of j and also called height of j), the expected number of right edges (and left edges) to the node j, the expected number of descendants of node j, and the number of key comparisons when inserting a node with key $j + \frac{1}{2}$. Moreover, we have shown that the length of the left path L_n is asymptotically Gaussian distributed with mean $\mathbb{E}(L_n) \sim 2\log n$ and variance $\mathbb{V}(L_n) \sim 2\log n$. We remark that the latter result can be obtained directly by applying the theory of records (see, e.g., [3]).

In the present work we will reconsider some of these parameters, where we will focus here on establishing limiting distribution results for randomly chosen nodes in size-n priority trees: in particular we will analyse the depth $A_n^{[H]}$ of a random node, the number of right edges $A_n^{[R]}$ to a random node and the number of descendants $A_n^{[D]}$ of a random node. We remark that we use throughout this paper the convention that a node is considered as its own ancestor (and is thus counted for the depth) and descendant. The results obtained for these tree statistics are given in Section 2.

The motivation of the author for a detailed study of this tree family is three-fold.

- Up to now there exist only few results, which describe the average-case behaviour of the operations in priority trees that are of interest when using them as a data structure for priority queue administration. In particular the behaviour of the algorithm INSERT when inserting a random node into a randomly grown priority tree of size n is little analyzed: everything known was obtained more than 30 years ago in [10] leading to the exact and asymptotic formulæ for the expectation of the number of key comparisons I_n and the expectation of the recursion depth R_n given above. Although we will not carry out the distributional analysis of I_n and R_n in the present paper state that these parameters are closely related to $A_n^{[H]}$ and $A_n^{[R]}$ studied here: one obtains that, asymptotically, I_n behaves like $A_n^{[H]}$ and R_n behaves like $A_n^{[R]}$, or to be more precise, Theorem 1 also holds for I_n and Theorem 2 also holds for R_n. We state these results as Corollary 4. We remark that Corollary 4 is not a consequence of Theorem 1-2 (e.g., it does not seem that there is a simple relation between I_n and $A_n^{[H]}$), but a consequence of the proof of these theorems, which is fully analogous. A very raw sketch of the proof of these results appeared in [11] as an extended abstract of a conference.

- This tree model seems to be also of some purely mathematical interest, since the algorithm INSERT gives a quite simple construction rule, which generates a tree family with an unusual behaviour.

Most random tree models are either a member of the class of $\log n$-trees (a randomly chosen node has expected depth $\Theta(\log n)$, like, e. g., the data structures binary search trees, digital search trees, and Quad-trees) or of the class of $\sqrt{n}$-trees (a randomly chosen node has expected depth $\Theta(\sqrt{n})$, like, e. g., binary trees or unordered labelled trees, so called Cayley trees). As a consequence of Theorem 1 it is shown in this paper that priority trees are neither $\log n$-trees nor $\sqrt{n}$-trees, but the depth of a randomly chosen node has expectation $\sim \frac{1}{3}\log^2 n$ and variance $\sim \frac{10}{81}\log^3 n$ and is, after normalization, asymptotically standard normally distributed.

- The proof of the asymptotic normality of the centered and scaled version of the parameter $A_n^{[H]}$ as stated in Theorem 1 could be of some methodological interest, since it was shown by using the method of moments without shifting by the exact or asymptotic mean. This requires of course a very detailed description of the asymptotic behaviour of the ordinary or factorial moments of $A_n^{[H]}$ to take care of the huge cancellations when considering the centered random variable. The motivation comes from the fact that there are several parameters in the study of discrete structures, where one has a simple description of the random variable of interest by means of generating functions leading to functional equations, differential equations, etc., but when applying the method of moments, after shifting by the asymptotic mean, one often (apart from the instance $\mathbb{E}(X_n) \sim cn$) cannot directly continue the computations at the level of generating functions, but one is rather forced to go back and do it at the level of coefficients. In general certainly shifting the mean, i.e., to study directly the recurrences for the centered moments, is advantageous; see, e.g., [1] for an impressive example. However, there might be some intrinsic interest in the question what is necessary to know (and how to obtain this knowledge) about the behaviour of the generating functions of the ordinary (or factorial) moments of a random variable X_n to show asymptotic normality of the normalized version of X_n, in particular for instances, where the so called quasi-power theorem is not applicable.

2 Results and mathematical preliminaries

2.1 Results

As already mentioned in the introduction we use as mathematical model for our analysis of priority trees the random permutation model, i.e., all $n!$ permutations of the numbers $1, \ldots, n$ are equally likely for constructing a priority tree of size n. For this model we give here in Theorems 1–3 limiting distribution results for the following three parameters:

- $A_n^{[H]}$ is the random variable that counts the depth ($=$ height) of a randomly chosen node in a random priority tree of size n, i.e., the number of nodes ($=$ ancestors) lying on the unique path from the root to a randomly chosen node.

- $A_n^{[R]}$ is the random variable that counts the number of right edges on the path from the root to a randomly chosen node in a random priority tree of size n

- $A_n^{[D]}$ is the random variable that counts the number of descendants of a randomly chosen node in a random priority tree of size n. The number of descendants of a node x is here defined as the size of the subtree rooted at node x, i.e., a chosen node is always counted as a descendant of itself.

We formulate now the main results of this paper, where we denote as usual by $\xrightarrow{(d)}$ convergence in distribution and where $\Phi(x)$ denotes the distribution function of the standard normal distribution $\mathcal{N}(0,1)$.

Theorem 1 *The depth $A_n^{[H]}$ of a random node in a random size-n priority tree is asymptotically Gaussian distributed:*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{A_n^{[H]} - \frac{1}{3}\log^2 n}{\sqrt{\frac{10}{81}\log^3 n}} \le x \right\} - \Phi(x) \right| = o(1).$$

The expectation $\mathbb{E}(A_n^{[H]})$ and the variance $\mathbb{V}(A_n^{[H]})$ satisfy

$$\mathbb{E}(A_n^{[H]}) = \frac{1}{3}\log^2 n + \mathcal{O}(\log n), \quad \mathbb{V}(A_n^{[H]}) = \frac{10}{81}\log^3 n + \mathcal{O}(\log^2 n).$$

Theorem 2 *The number $A_n^{[R]}$ of right edges on the path from the root to a random node in a random size-n priority tree is asymptotically Gaussian distributed, where the convergence rate is of order $\mathcal{O}\big(\frac{1}{\sqrt{\log n}}\big)$:*

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

The expectation $\mathbb{E}(A_n^{[R]})$ and the variance $\mathbb{V}(A_n^{[R]})$ satisfy

$$\mathbb{E}(A_n^{[R]}) = \frac{2}{3}\log n + \mathcal{O}(1), \quad \mathbb{V}(A_n^{[R]}) = \frac{10}{27}\log n + \mathcal{O}(1).$$

Theorem 3 *The number $A_n^{[D]}$ of descendants of a random node in a random size-n priority tree converges in distribution to a discrete random variable $A^{[D]}$ (with non-existing r-th moments $\mathbb{E}\big(\big(A^{[D]}\big)^r\big)$, for $r \ge 1$), $A_n^{[D]} \xrightarrow{(d)} A^{[D]}$, which has the distribution*

$$\mathbb{P}\Big\{A^{[D]} = m\Big\} = \frac{2}{3m(m+1)}H_m - \frac{m-1}{3m(m+1)^2}, \quad \textit{for } m \ge 1.$$

The r-th moments of $A_n^{[D]}$ are given asymptotically by

$$\mathbb{E}\left(\big(A_n^{[D]}\big)^r\right) = \begin{cases} \frac{r(r+1)}{(r-1)(r+2)} n^{r-1}\log n + \mathcal{O}\big(n^{r-1}\big), & \textit{for } r \ge 2, \\ \frac{1}{3}\log^2 n + \mathcal{O}\big(\log n\big), \quad \textit{for } r = 1. \end{cases}$$

We further present here limiting distribution results for the following two parameters, which describe the insertion costs when inserting a random element into randomly grown priority trees.

- I_n: 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 $\{\frac{1}{2}, \frac{3}{2}, \dots, n+\frac{1}{2}\}$ can occur with equal probability $\frac{1}{n+1}$ as the key of the inserted element.

- R_n: 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.

By applying the methods used for the proof of Theorem 1-2 we obtained the following results (see [11]).

Corollary 4 *The number of key comparisons I_n when inserting a random element into a random size-n priority tree with the algorithm* INSERT *is asymptotically Gaussian distributed:*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{I_n - \frac{1}{3}\log^2 n}{\sqrt{\frac{10}{81}\log^3 n}} \le x \right\} - \Phi(x) \right| = o(1).$$

The expectation $\mathbb{E}(I_n)$ and the variance $\mathbb{V}(I_n)$ satisfy

$$\mathbb{E}(I_n) = \frac{1}{3}\log^2 n + \mathcal{O}(\log n), \quad \mathbb{V}(I_n) = \frac{10}{81}\log^3 n + \mathcal{O}(\log^2 n).$$

The recursion depth R_n when inserting a random element into a random size-n priority tree with the algorithm INSERT *is asymptotically Gaussian distributed, where the convergence rate is of order $\mathcal{O}(\frac{1}{\sqrt{\log n}})$:*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{R_n - \frac{2}{3}\log n}{\sqrt{\frac{10}{27}\log n}} \le x \right\} - \Phi(x) \right| = \mathcal{O}\Big(\frac{1}{\sqrt{\log n}}\Big).$$

The expectation $\mathbb{E}(R_n)$ and the variance $\mathbb{V}(R_n)$ satisfy

$$\mathbb{E}(R_n) = \frac{2}{3}\log n + \mathcal{O}(1), \quad \mathbb{V}(R_n) = \frac{10}{27}\log n + \mathcal{O}(1).$$

2.2 Mathematical preliminaries and outline of the proofs

In our analysis we use the approach introduced in [13]: we work with three families of combinatorial objects denoted by $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$. The family $\mathcal{A}$ contains ordinary priority trees generated from random permutations (applying algorithm INSERT repeatedly) starting with the empty tree, the auxiliary family $\mathcal{B}$ contains priority trees generated from random permutations starting with the additional element "$+\infty$", the auxiliary family $\mathcal{C}$ contains priority trees generated from random permutations starting with the additional element "$-\infty$". We always assume that these additional elements are not counted for the size of an object, which means that the size of an object is given by the number of nodes without elements $-\infty$ and $+\infty$.

The analysis of the tree statistics introduced above relies heavily on the decomposition of the families $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ according to the first element k of a random permutation of the numbers $1, 2, \ldots, n$. These fundamental combinatorial decompositions are given in Figure 2.

They will lead to systems of recurrences for the probabilities $\mathbb{P}\{A_n = m\}$, $\mathbb{P}\{B_n = m\}$ and $\mathbb{P}\{C_n = m\}$ of the random variables A_n, B_n and C_n, where A_n is either $A_n^{[H]}$, $A_n^{[R]}$ or $A_n^{[D]}$, and B_n and C_n are defined as the corresponding random variables for objects in $\mathcal{B}$ and $\mathcal{C}$. To obtain them we introduce

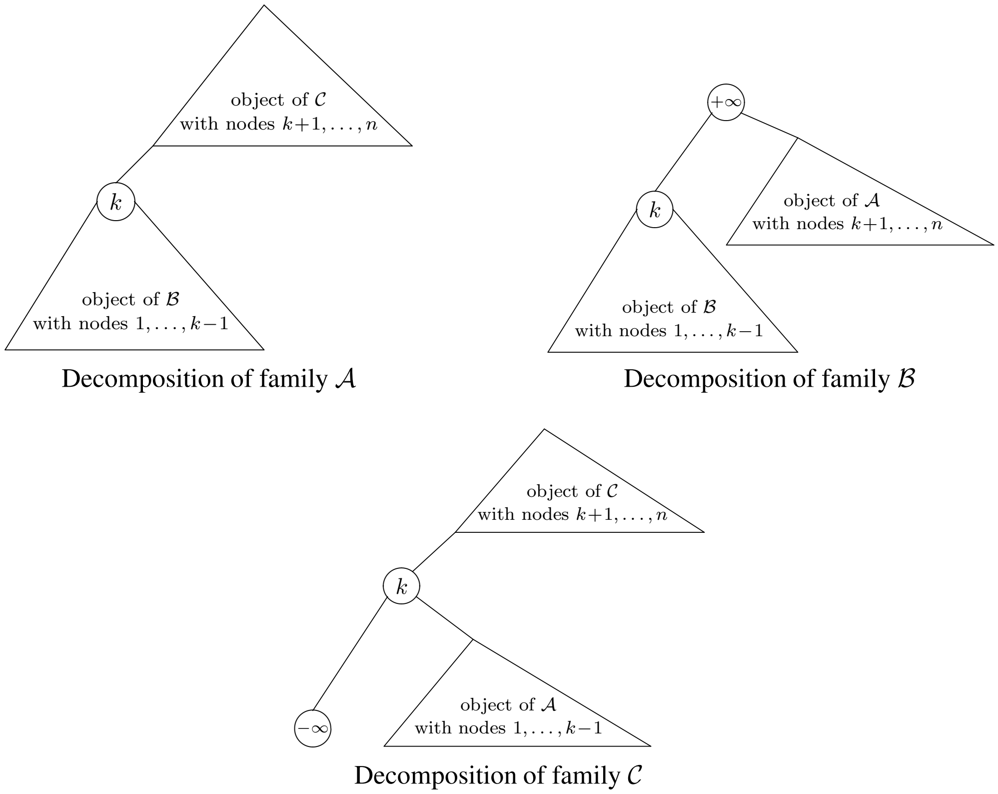

Fig. 2: Decomposition of the families $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ according to the first element.

random variables $A_{n,j}$, $B_{n,j}$ and $C_{n,j}$, which count the corresponding tree parameters in objects of size n in the families $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ for a fixed node j (instead of a random node). Of course we have then

$$A_n = \sum_{j=1}^{n} J_{n,j}\, A_{n,j},$$

where $J_{n,j}$ is the indicator variable of the event $U_n = j$, where U_n is uniformly (and independently of the $A_{n,j}$) distributed on $\{1, 2, \ldots, n\}$. Corresponding equations hold for the B_n and C_n, too. It holds then that $\mathbb{P}\{A_n = m\} = \frac{1}{n}\sum_{j=1}^{n} \mathbb{P}\{A_{n,j} = m\}$, etc.

This approach was used in [13] and one can translate the decompositions of Figure 2 easily into systems of recurrences for the probabilities $\mathbb{P}\{A_{n,j} = m\}$, etc. By introducing trivariate generating functions

$$A(z, u, v) := \sum_{n\geq 1}\sum_{j=1}^{n}\sum_{m\geq 0} \mathbb{P}\{A_{n,j} = m\} z^n u^j v^m, \quad \text{etc.}$$

we obtain then systems of linear differential equations for $A(z,u,v)$, $B(z,u,v)$ and $C(z,u,v)$, which lead after specializing $u=1$ to systems of linear differential equations for

$$A(z,v) := A(z,1,v) = \sum_{n\ge 1}\sum_{m\ge 0} n\,\mathbb{P}\{A_n = m\} z^n v^m$$

and the corresponding functions $B(z,v)$ and $C(z,v)$. All systems of differential equations can be transformed into single higher order linear differential equations for $A(z,v)$.

Before we outline how to proceed from these differential equations we introduce some notation used throughout this paper. We denote by $x^{\underline{m}} := x(x-1)\cdots(x-m+1)$ the falling factorials and by $\lfloor x \rfloor$ the floor function (see, e.g., [7]). In order to avoid clumsy notations we use the abbreviation $\mathfrak{Pol}_x(d)$ for any polynomial in x of degree at most d, which can vary in every occurrence of $\mathfrak{Pol}_x(d)$. If $d<0$ then we use the convention that $\mathfrak{Pol}_x(d) = 0$. With $[x^n]f(x)$ we denote the n-th coefficient f_n of a power series $f(x) = \sum_{k\ge 0} f_k x^k$ or polynomial $f(x) = \sum_{k=0}^{d} f_k x^k$. Furthermore we use the abbreviation $L(z) := \log\left(\frac{1}{1-z}\right)$ and the Iverson-notation $[\![\text{Statement}]\!]$, which gives 1 if Statement is true and 0 otherwise.

In the instance of $A_n^{[H]}$ we will use the method of moments to show the stated central limit theorem. First we can reduce the order of the third order differential equation for $A(z,v)$ to a second order linear differential equation for a function $H(z,v)$. The functions $H_r(z) := \frac{\partial^r}{\partial v^r} H(z,v)\big|_{v=1}$ satisfy for all integers $r \ge 0$ second order Cauchy-Euler differential equations and one can easily pump out the structure of the solutions for $H_r(z)$ as linear combinations of functions $\frac{1}{(1-z)^2}L(z)^{2r-m}$, with $0 \le m \le 2r$ and $(1-z)L(z)^{2r-2-m}$, with $0 \le m \le 2r-2$. This is shown in Lemma 5. This structure will follow also directly from the general study of Cauchy-Euler differential equations given in [2]. But it is essential in our analysis to obtain a full recursive description for the coefficients $\alpha_r^{(m)}$ of the functions $\frac{1}{(1-z)^2}L(z)^m$ in these linear combinations, which is also established in Lemma 5.

Next, these recurrences for the coefficients $\alpha_r^{(m)}$ are studied for fixed m as a function of r. In Lemma 6 we show the representation $\alpha_r^{(m)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \mathfrak{Pol}_r(2m - \lfloor \frac{m+1}{2} \rfloor)$, which holds for fixed m and $r \ge 0$. Moreover, the leading coefficients $\eta^{(m)}$ in the polynomials are given.

Lemma 8 translates these results to the generating functions $A_r(z)$ of the r-th factorial moments of $A_n^{[H]}$, $A_r(z) := \frac{\partial^r}{\partial v^r} A(z,v)\big|_{v=1} = \sum_{n\ge 1} n\,\mathbb{E}\left(\big(A_n^{[H]}\big)^{\underline{r}}\right) z^n$. In particular we obtain the asymptotic expansion $A_r(z) \sim \sum_{m=0}^{2r} a_r^{(m)} \frac{1}{(1-z)^2} L(z)^{2r-m}$ around the dominant singularity $z=1$, where the coefficients $a_r^{(m)}$ are for fixed m and $r \ge 0$ again given by $a_r^{(m)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \mathfrak{Pol}_r(2m - \lfloor \frac{m+1}{2} \rfloor)$, with the same leading coefficients $\eta^{(m)}$ of the polynomials occurring in $a_r^{(m)}$ and $\alpha_r^{(m)}$.

This detailed description of the $a_r^{(m)}$ is now used together with a description of the appearing sums via higher order differences to treat the massive cancellations and to determine the leading terms of the r-th centered moments $\mathbb{E}\left(\big(A_n^{[H]} - \frac{1}{3}\log^2 n\big)^r\right)$. It follows that after normalization these centered moments converge to the moments of the standard normal distribution $\mathcal{N}(0,1)$. This is established in Lemma 9. Together with the Theorem of Fréchet and Shohat (see, e.g., [4]) Theorem 1 is shown.

In the instance of $A_n^{[R]}$ the second order differential equation for $A(z,v)$ is of Cauchy-Euler type, which has an explicit solution. Extracting coefficients $[z^n]A(z,v)$ uniformly in a neighbourhood of $v=1$ only requires singularity analysis of generating functions [5]. Theorem 2 follows from this expansion immediately by an application of the quasi-power theorem due to Hwang [8].

In the instance of $A_n^{[D]}$ the second order differential equation for $A(z, v)$ is again of Cauchy-Euler type. Although the solution is not given in closed form, it is here not hard to extract coefficients $[z^n v^m]A(z, v)$ and thus to give explicit results for the probabilities $\mathbb{P}\{A_n^{[D]} = m\}$. From this explicit formula we obtain immediately the first part of Theorem 3. To obtain the asymptotic formulæ for the r-th moments of $A_n^{[D]}$ as given in the second part of Theorem 3 we expand the solution of $A(z, v)$ in terms of $w := v - 1$ and determine for fixed r the asymptotic expansion of $[w^r]A(z, v)$ around the dominant singularity $z = 1$. Singularity analysis leads then to asymptotic results for $\mathbb{E}\Big(\big(A_n^{[D]}\big)^r\Big)$.

The complete proofs of Theorems 1–3 are given in Sections 3–5.

3 Depth of a random node

3.1 A system of recurrences

Using the combinatorial decomposition given in Figure 2, i.e., the decomposition with respect to the first element k of a random permutation of $\{1, \dots, n\}$, we obtain the following system of recurrences for the probabilities $\mathbb{P}\{A_{n,j}^{[H]} = m\}$, $\mathbb{P}\{B_{n,j}^{[H]} = m\}$ and $\mathbb{P}\{C_{n,j}^{[H]} = m\}$, which give the probabilities that node j has depth m in a random generated object of size n of the families $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ for $1 \le j \le n$ (see [13]). We recall the convention used in this paper that the depth of a node x is always measured by the number of nodes lying on the direct path from the root to x.

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

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

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

The appearing functions $p_{n,m}$ give the probabilities that the size of the left path, i.e., the number of nodes from the root to the left leaf, in a random object of size n of the family $\mathcal{C}$ is m. We will here make the convention to count the element $+\infty$ but not the element $-\infty$ for the depth of a node and the size of the left path. It has been shown already in [13] that these probabilities and their generating function $p(z, v) := \sum_{n\ge0}\sum_{m\ge0} p_{n,m} z^n v^m$ are given by:

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

where $\left[{n \atop m}\right]$ denote the signless Stirling numbers of first kind. This can be seen easily by the observation that when starting with the element $-\infty$ and inserting n distinct elements, an element x will be inserted into the left path if and only if x is larger than all elements already inserted. Thus $p_{n,m}$ is nothing else than the probability that the number of left-to-right maxima of a random permutation of $\{1, \dots, n\}$ is exactly m, and it is well-known that these probabilities are given by the formula stated above.

3.2 The generating functions approach

Introducing the generating function $A(z,u,v) := \sum_{n\ge 1}\sum_{1\le j\le n}\sum_{m\ge 0}\mathbb{P}\{A_{n,j}^{[H]} = m\} z^n u^j v^m$, and the corresponding functions $B(z,u,v)$ and $C(z,u,v)$, this system of recurrences (1) gives the following system of linear differential equations:

$$\begin{aligned}
\frac{\partial}{\partial z}A(z,u,v) &= \frac{u}{1-uz}C(z,u,v) + \frac{uv}{1-uz}p(z,v) + p(z,v)B(z,u,v),\\
\frac{\partial}{\partial z}B(z,u,v) &= \frac{uv}{1-uz}A(z,u,v) + \frac{uv^2}{(1-z)(1-uz)} + \frac{v}{1-z}B(z,u,v),\\
\frac{\partial}{\partial z}C(z,u,v) &= \frac{u}{1-uz}C(z,u,v) + \frac{uv}{1-uz}p(z,v) + vp(z,v)A(z,u,v),
\end{aligned}$$

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

The generating function of interest $A(z,v) := A(z,1,v) = \sum_{n\ge 1}\sum_{m\ge 0} n\,\mathbb{P}\{A_n^{[H]} = m\} z^n v^m$, and the corresponding auxiliary functions $B(z,v)$ and $C(z,v)$, are thus satisfying the following system of differential equations with initial conditions $A(0,v) = B(0,v) = C(0,v) = 0$:

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

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

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

Using the substitutions $\tilde{C}(z,v) := (1-z)C(z,v)$ and $\tilde{B}(z,v) := (1-z)^v B(z,v)$ we can simplify system (2) and get

$$\frac{\partial}{\partial z}A(z,v) = \frac{1}{(1-z)^2}\tilde{C}(z,v) + \frac{1}{(1-z)^{2v}}\tilde{B}(z,v) + \frac{v}{(1-z)^{v+1}}, \tag{3a}$$

$$\frac{\partial}{\partial z}\tilde{B}(z,v) = v(1-z)^{v-1}A(z,v) + v^2(1-z)^{v-2}, \tag{3b}$$

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

Now one can successively eliminate $\tilde{B}(z,v)$ and $\tilde{C}(z,v)$ from (3a) by comparing with (3b) and (3c) and get eventually the following homogeneous third order linear differential equation for $A(z,v)$:

$$\begin{aligned}
\frac{\partial^3}{\partial z^3}A(z,v) - \frac{3+2v}{1-z}\frac{\partial^2}{\partial z^2}A(z,v) + 2v\Big(\frac{2}{(1-z)^2} - \frac{1}{(1-z)^{v+1}}\Big)\frac{\partial}{\partial z}A(z,v)&\\
+ \frac{2v}{(1-z)^{v+2}}A(z,v) = 0,& \quad (4)
\end{aligned}$$

with initial conditions $A(0,v) = 0$, $\frac{\partial}{\partial z}A(z,v)\big|_{z=0} = v$, $\frac{\partial^2}{\partial z^2}A(z,v)\Big|_{z=0} = 2v(v+1)$ resulting from $\mathbb{P}\{A_1^{[H]} = 1\} = 1, \mathbb{P}\{A_2^{[H]} = 1\} = \mathbb{P}\{A_2^{[H]} = 2\} = \frac{1}{2}$.

Since one solution of (4) is given by $\frac{1}{1-z}$, which is checked easily, we can reduce the order of this differential equation when substituting

$$H(z,v) := \frac{\partial}{\partial z}\big((1-z)A(z,v)\big). \tag{5}$$

This leads to the following homogeneous second order linear differential equation, which is the starting point of our studies[(i)]:

$$\frac{\partial^2}{\partial z^2}H(z,v) - \frac{2v}{1-z}\frac{\partial}{\partial z}H(z,v) - \frac{2v}{(1-z)^{v+1}}H(z,v) = 0, \tag{6}$$

with initial values $H(0,v) = v$ and $\frac{\partial}{\partial z}H(z,v)\big|_{z=0} = 2v^2$.

In order to apply the method of moments we define for integers $r \geq 0$ the functions $H_r(z)$ via

$$H_r(z) := \left.\frac{\partial^r}{\partial v^r}H(z,v)\right|_{v=1}.$$

Differentiating equation (6) r times with respect to v and evaluating at $v=1$ leads then to the following differential equations for $H_r(z)$ of Cauchy-Euler type:

$$H_r''(z) - \frac{2}{1-z}H_r'(z) - \frac{2}{(1-z)^2}H_r(z) = S_r(z), \tag{7}$$

where the inhomogeneous parts are given by $S_0(z) = 0$ and for $r \geq 1$:

$$S_r(z) = \frac{2r}{1-z}H_{r-1}'(z) + 2\sum_{k=0}^{r-1}\binom{r}{k}H_k(z)\Big(\frac{1}{(1-z)^2}L(z)^{r-k} + \frac{r-k}{(1-z)^2}L(z)^{r-1-k}\Big). \tag{8}$$

The corresponding homogeneous differential equation has the general solution (with arbitrary constants C_1, C_2):

$$H_r^{[hom]}(z) = C_1\frac{1}{(1-z)^2} + C_2(1-z).$$

The variation of parameters method (see, e.g., [14]) gives then immediately the general solution of the required $H_r(z)$:

$$H_r(z) = \frac{1}{3(1-z)^2}\int_{t=0}^{z}(1-t)^3S_r(t)dt - \frac{1-z}{3}\int_{t=0}^{z}S_r(t)dt + \frac{C_{r,1}}{(1-z)^2} + C_{r,2}(1-z), \tag{9}$$

where the constants $C_{r,1}$ and $C_{r,2}$ are specified by the initial conditions: adapting to the initial values leads to the linear system of equations

$$\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}\begin{pmatrix} C_{r,1} \\ C_{r,2} \end{pmatrix} = \begin{pmatrix} \frac{\partial^r}{\partial v^r}v\big|_{v=1} \\ \frac{\partial^r}{\partial v^r}2v^2\big|_{v=1} \end{pmatrix},$$

(i) It should be remarked that the general solution of this differential equation can be expressed via hypergeometric functions or alternatively via Bessel functions. Maple is capable of finding a required solution basis. However it seems quite involved to show the given central limit theorem directly from the solution obtained after adapting to the initial values.

and thus to the solutions

$$\begin{aligned}
&C_{0,1} = 1,\ C_{0,2} = 0, \quad C_{1,1} = \frac{5}{3},\ C_{1,2} = -\frac{2}{3}, \quad C_{2,1} = \frac{4}{3},\ C_{2,2} = -\frac{4}{3}, \\
&C_{r,1} = 0,\ C_{r,2} = 0, \ \text{ for } \ r \ge 3.
\end{aligned} \tag{10}$$

We get immediately that $H_0(z) = \frac{1}{(1-z)^2}$, which, of course, follows also from the definition.

3.3 The exact solution of $H_r(z)$

We will give now the exact solutions of the functions $H_r(z)$ satisfying (9), where we make throughout this paper the convention that $\sum_{k=a}^{b} f(k) = 0$ if $b < a$. We also recall the abbreviation $L(z) := \log\left(\frac{1}{1-z}\right)$.

Lemma 5 *The solution $H_r(z)$ of the differential equation* (9) *is given by*

$$H_r(z) = \sum_{m=0}^{2r} \alpha_r^{(m)} \frac{1}{(1-z)^2} L(z)^{2r-m} + \sum_{m=0}^{2r-2} \beta_r^{(m)} (1-z) L(z)^{2r-2-m}, \tag{11}$$

with certain constants $\alpha_r^{(m)}$ and $\beta_r^{(m)}$ that are defined recursively via [ii]

$$\alpha_r^{(0)} = \frac{1}{3}\alpha_{r-1}^{(0)}, \quad \text{for } r \ge 1, \tag{12a}$$

$$\begin{aligned}
\alpha_r^{(m)} = {} & \frac{1}{3(2r-m)} \Bigg[4r\alpha_{r-1}^{(m-1)} + 2r(2r-m)\alpha_{r-1}^{(m-2)} + \sum_{k=0}^{m} 2\tbinom{r}{r-1-k} \alpha_{r-1-k}^{(m-k)} \\
& \qquad + \sum_{k=0}^{m-1} 2(k+1)\tbinom{r}{r-1-k} \alpha_{r-1-k}^{(m-1-k)} \Bigg] \\
& + \sum_{l=0}^{m-1} \frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} \Bigg[4r\alpha_{r-1}^{(l-1)} + 2r(2r-l)\alpha_{r-1}^{(l-2)} + \sum_{k=0}^{l} 2\tbinom{r}{r-1-k} \alpha_{r-1-k}^{(l-k)} \\
& \qquad + \sum_{k=0}^{l-1} 2(k+1)\tbinom{r}{r-1-k} \alpha_{r-1-k}^{(l-1-k)} \Bigg], \quad \text{for } 1 \le m \le 2r-1,
\end{aligned} \tag{12b}$$

$$\begin{aligned}
\alpha_r^{(2r)} = {} & \sum_{l=0}^{2r-1} \frac{(-1)^l (2r-1-l)!}{3^{2r-l+1}} \Bigg[4r\alpha_{r-1}^{(l-1)} + 2r(2r-l)\alpha_{r-1}^{(l-2)} + \sum_{k=0}^{l} 2\tbinom{r}{r-1-k} \alpha_{r-1-k}^{(l-k)} \\
& \qquad + \sum_{k=0}^{l-1} 2(k+1)\tbinom{r}{r-1-k} \alpha_{r-1-k}^{(l-1-k)} \Bigg] \\
& + \sum_{l=0}^{2r-3} \frac{(2r-3-l)!}{3^{2r-1-l}} \Bigg[-2r\beta_{r-1}^{(l-1)} + 2r(2r-2-l)\beta_{r-1}^{(l-2)} + \sum_{k=0}^{l} 2\tbinom{r}{r-1-k} \beta_{r-1-k}^{(l-k)} \\
& \qquad + \sum_{k=0}^{l-1} 2(k+1)\tbinom{r}{r-1-k} \beta_{r-1-k}^{(l-1-k)} \Bigg] + C_{r,1},
\end{aligned} \tag{12c}$$

[ii] It must be noticed that in the further proof only the recurrences for $\alpha_r^{(m)}$, for $0 \le m \le 2r-1$, are required. Thus in particular the recurrences for $\beta_r^{(m)}$ will not be used in the sequel, but they are stated here to keep the paper self-contained.

$$\beta_r^{(0)} = -\frac{r}{3(r-1)}\beta_{r-1}^{(0)}, \quad \textit{for } r \geq 2,$$

$$\beta_r^{(m)} = -\frac{1}{3(2r-2-m)}\Big[-2r\beta_{r-1}^{(m-1)} + 2r(2r-2-m)\beta_{r-1}^{(m-2)} + \sum_{k=0}^{m} 2\tbinom{r}{r-1-k}\beta_{r-1-k}^{(m-k)} + \sum_{k=0}^{m-1} 2(k+1)\tbinom{r}{r-1-k}\beta_{r-1-k}^{(m-1-k)}\Big] \tag{12d}$$

$$+ \sum_{l=0}^{m-1} -\frac{(2r-3-l)^{\underline{m-l-1}}}{3^{m-l+1}}\Big[-2r\beta_{r-1}^{(l-1)} + 2r(2r-2-l)\beta_{r-1}^{(l-2)} + \sum_{k=0}^{l} 2\tbinom{r}{r-1-k}\beta_{r-1-k}^{(l-k)} + \sum_{k=0}^{l-1} 2(k+1)\tbinom{r}{r-1-k}\beta_{r-1-k}^{(l-1-k)}\Big], \quad \textit{for } 1 \leq m \leq 2r-3,$$

$$\beta_r^{(2r-2)} = -\alpha_r^{(2r)} + C_{r,1} + C_{r,2}. \tag{12e}$$

The constants $C_{r,1}$ and $C_{r,2}$ are given by (10). *Furthermore we make the convention that $\alpha_r^{(m)} = 0$ if $m < 0$ or $r < 0$ or $m > 2r$, and that $\beta_r^{(m)} = 0$ if $m < 0$ or $r \leq 0$ or $m > 2r-2$.*

Remark: Due to the convention $\sum_{k=a}^{b} f(k) = 0$ if $b < a$ used throughout this paper Lemma 5 gives in particular the values $\alpha_0^{(0)} = C_{0,1} = 1$ and $\beta_0^{(0)} = 0$. **Proof:** To show this result we use induction on r, where we show in parallel the following formulæ for the inhomogeneous parts $S_r(z)$ of (7):

$$S_r(z) = \sum_{l=0}^{2r-1} \gamma_r^{(l)} \frac{1}{(1-z)^4} L(z)^{2r-1-l} + \sum_{l=0}^{2r-3} \delta_r^{(l)} \frac{1}{1-z} L(z)^{2r-3-l}, \tag{13}$$

where the $\gamma_r^{(l)}$ and $\delta_r^{(l)}$ are for $r \geq 0$ given recursively by

$$\gamma_r^{(l)} = 4r\alpha_{r-1}^{(l-1)} + 2r(2r-l)\alpha_{r-1}^{(l-2)} + \sum_{k=0}^{l} 2\tbinom{r}{r-1-k}\alpha_{r-1-k}^{(l-k)} + \sum_{k=0}^{l-1} 2(k+1)\tbinom{r}{r-1-k}\alpha_{r-1-k}^{(l-1-k)}, \tag{14a}$$

$$\delta_r^{(l)} = -2r\beta_{r-1}^{(l-1)} + 2r(2r-2-l)\beta_{r-1}^{(l-2)} + \sum_{k=0}^{l} 2\tbinom{r}{r-1-k}\beta_{r-1-k}^{(l-k)} + \sum_{k=0}^{l-1} 2(k+1)\tbinom{r}{r-1-k}\beta_{r-1-k}^{(l-1-k)}. \tag{14b}$$

For $r = 0$ we obtain from the recurrences $\alpha_0^{(0)} = C_{0,1} = 1$, $\beta_0^{(0)} = 0$, $\gamma_0^{(0)} = 0$ and $\delta_0^{(0)} = 0$, which corresponds with $S_0(z) = 0$ and $H_0(z) = \frac{1}{(1-z)^2}$; thus formulæ (11) and (13), and also the recurrences (12) and (14), hold for $r = 0$.

Now we assume that formulæ (11) and (13) together with the recurrences of the appearing coefficients are proven for all $0 \leq k < r$ with a given $r \geq 1$. To show them also for r we use at first the recursive definition of $S_r(z)$ via (8) and compute by using the induction hypothesis:

$$\frac{1}{1-z}H_{r-1}'(z) = \frac{1}{(1-z)^4}\sum_{l=0}^{2r-2}\alpha_{r-1}^{(l)}\Big(2L(z)^{2r-2-l} + (2r-2-l)L(z)^{2r-3-l}\Big) + \frac{1}{1-z}\sum_{l=0}^{2r-4}\beta_{r-1}^{(l)}\Big(-L(z)^{2r-4-l} + (2r-4-l)L(z)^{2r-5-l}\Big),$$

$$H_k(z)\frac{L(z)^{r-k}}{(1-z)^2} = \sum_{l=0}^{2k} \alpha_k^{(l)} \frac{1}{(1-z)^4} L(z)^{r+k-l} + \sum_{l=0}^{2k-2} \beta_k^{(l)} \frac{1}{1-z} L(z)^{r+k-l-2}, \quad \text{for } 0 \le k \le r-1.$$

Plugging in equation (8) these representations we get after collecting terms:

$$\begin{aligned} S_r(z) = \frac{1}{(1-z)^4} \sum_{l=0}^{2r-1} L(z)^{2r-1-l} \Bigg[& 4r\alpha_{r-1}^{(l-1)} + 2r(2r-l)\alpha_{r-1}^{(l-2)} + \sum_{k=0}^{l} 2\tbinom{r}{r-1-k}\alpha_{r-1-k}^{(l-k)} \\ & + \sum_{k=0}^{l-1} 2(k+1)\tbinom{r}{r-1-k}\alpha_{r-1-k}^{(l-1-k)} \Bigg] \\ + \frac{1}{1-z} \sum_{l=0}^{2r-3} L(z)^{2r-3-l} \Bigg[& -2r\beta_{r-1}^{(l-1)} + 2r(2r-2-l)\beta_{r-1}^{(l-2)} + \sum_{k=0}^{l} 2\tbinom{r}{r-1-k}\beta_{r-1-k}^{(l-k)} \\ & + \sum_{k=0}^{l-1} 2(k+1)\tbinom{r}{r-1-k}\beta_{r-1-k}^{(l-1-k)} \Bigg]. \end{aligned} \tag{15}$$

Extracting in equation (15) the coefficients from the powers of $L(z)$ shows then the validity of the given formulæ (14) for $\gamma_r^{(l)}$ and $\delta_r^{(l)}$ also for $r \ge 1$.

To compute $H_r(z)$ using (9) we require the following antiderivative formulas (for integers $m \ge 0$), which are obtained easily via integration by parts:

$$\int_{t=0}^{z} \frac{1}{1-t} L(t)^m dt = \frac{1}{m+1} L(z)^{m+1},$$

$$\frac{1}{(1-z)^2} \int_{t=0}^{z} (1-t)^2 L(t)^m dt = -(1-z) \sum_{k=0}^{m} \frac{m^{\underline{k}}}{3^{k+1}} L(z)^{m-k} + \frac{1}{(1-z)^2} \frac{m!}{3^{m+1}},$$

$$(1-z) \int_{t=0}^{z} \frac{1}{(1-t)^4} L(t)^m dt = \frac{1}{(1-z)^2} \sum_{k=0}^{m} \frac{(-1)^k m^{\underline{k}}}{3^{k+1}} L(z)^{m-k} + (1-z) \frac{(-1)^{m+1} m!}{3^{m+1}}.$$

Plugging in the representation (15) of $S_r(z)$ into the solution (9) for $H_r(z)$ gives then after a few manipulations and collecting terms:

$$\begin{aligned} H_r(z) = {} & \frac{\gamma_r^{(0)}}{6r} \frac{1}{(1-z)^2} L(z)^{2r} - \frac{\delta_r^{(0)}}{3(2r-2)} (1-z) L(z)^{2r-2} \\ & + \frac{1}{(1-z)^2} \sum_{m=1}^{2r-1} L(z)^{2r-m} \left(\frac{\gamma_r^{(m)}}{3(2r-m)} + \sum_{l=0}^{m-1} \frac{\gamma_r^{(l)} (-1)^{m-l} (2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} \right) \\ & + \frac{1}{(1-z)^2} \left(\sum_{l=0}^{2r-1} \frac{\gamma_r^{(l)} (-1)^l (2r-1-l)!}{3^{2r-l+1}} + \sum_{l=0}^{2r-3} \frac{\delta_r^{(l)} (2r-3-l)!}{3^{2r-1-l}} + C_{r,1} \right) \\ & + (1-z) \sum_{m=1}^{2r-3} L(z)^{2r-2-m} \left(-\frac{\delta_r^{(m)}}{3(2r-2-m)} + \sum_{l=0}^{m-1} -\frac{\delta_r^{(l)} (2r-3-l)^{\underline{m-l-1}}}{3^{m-l+1}} \right) \end{aligned} \tag{16}$$

$$+ (1-z)\Big(- \sum_{l=0}^{2r-1} \frac{\gamma_r^{(l)}(-1)^l (2r-1-l)!}{3^{2r-l+1}} - \sum_{l=0}^{2r-3} \frac{\delta_r^{(l)}(2r-3-l)!}{3^{2r-1-l}} + C_{r,2}\Big).$$

Extracting in equation (16) the coefficients from the powers of $L(z)$ gives then the relations:

$$\alpha_r^{(0)} = \frac{\gamma_r^{(0)}}{6r}, \quad \text{for } r \geq 1,$$

$$\alpha_r^{(m)} = \frac{\gamma_r^{(m)}}{3(2r-m)} + \sum_{l=0}^{m-1} \frac{\gamma_r^{(l)}(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}}, \quad \text{for } 1 \leq m \leq 2r-1,$$

$$\alpha_r^{(2r)} = \sum_{l=0}^{2r-1} \frac{\gamma_r^{(l)}(-1)^l(2r-1-l)!}{3^{2r-l+1}} + \sum_{l=0}^{2r-3} \frac{\delta_r^{(l)}(2r-3-l)!}{3^{2r-1-l}} + C_{r,1},$$

$$\beta_r^{(0)} = -\frac{\delta_r^{(0)}}{6(r-1)}, \quad \text{for } r \geq 2,$$

$$\beta_r^{(m)} = -\frac{\delta_r^{(m)}}{3(2r-2-m)} + \sum_{l=0}^{m-1} \frac{-\delta_r^{(l)}(2r-3-l)^{\underline{m-l-1}}}{3^{m-l+1}}, \quad \text{for } 1 \leq m \leq 2r-3,$$

$$\beta_r^{(2r-2)} = -\sum_{l=0}^{2r-1} \frac{\gamma_r^{(l)}(-1)^l(2r-1-l)!}{3^{2r-l+1}} - \sum_{l=0}^{2r-3} \frac{\delta_r^{(l)}(2r-3-l)!}{3^{2r-1-l}} + C_{r,2}.$$

Plugging in the already proved equations (14) for $\gamma_r^{(l)}$ and $\delta_r^{(l)}$ shows thus the required recurrences (12) for $\alpha_r^{(l)}$ and $\beta_r^{(l)}$ also for $r \geq 1$. □

Lemma 5 (together with equation (5)) allows now to compute at least in principle closed formulæ for the moments of $A_n^{[H]}$ involving higher order harmonic numbers (see [15]). The expectation $\mathbb{E}(A_n^{[H]})$ has been computed already in [13]:

$$\mathbb{E}(A_n^{[H]}) = \frac{n+1}{3n}H_n^2 + \frac{2(2n+5)}{9n}H_n - \frac{n+1}{3n}H_n^{(2)} - \frac{13n+1}{27n}, \quad \text{for } n \geq 2, \quad \mathbb{E}(A_1^{[H]}) = 1.$$

This immediately gives $\mathbb{E}(A_n^{[H]}) \sim \frac{1}{3}\log^2 n$ as asymptotic equivalent of the expectation. A (somewhat lengthy) exact formula for the variance $\mathbb{V}(A_n^{[H]})$ appears in [12, p. 47], which gives the asymptotic equivalent $\mathbb{V}(A_n^{[H]}) \sim \frac{10}{81}\log^3 n$.

3.4 Description of the coefficients $\alpha_r^{(m)}$

Of great importance in our proof of Theorem 1 is a quite detailed description of the structure of the coefficients $\alpha_r^{(m)}$ for m fixed as a function of r. This is done in the following lemma, where we recall the abbreviation $\mathfrak{Pol}_x(d)$ for a certain polynomial in x of degree at most d, which can vary in every occurrence of $\mathfrak{Pol}_x(d)$. If $d < 0$ then we use the convention that $\mathfrak{Pol}_x(d) = 0$.

Lemma 6 *The coefficients $\alpha_r^{(m)}$ are for m fixed and $r \geq 0$ given by*

$$\alpha_r^{(m)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \mathfrak{Pol}_r(2m - \lfloor \tfrac{m+1}{2} \rfloor). \tag{17a}$$

Moreover, the leading coefficients $\eta^{(m)} := [r^{2m}]3^r \alpha_r^{(m)}$ *are given by*

$$\eta^{(m)} = \frac{\left(\frac{5}{9}\right)^m}{m!}. \tag{17b}$$

Remark: Equation (17a) is obviously equivalent to the formulation that for every m there exist constants $c_{m,i}$, $0 \le i \le 2m - \lfloor \frac{m+1}{2} \rfloor$, such that for all $r \ge 0$:

$$\alpha_r^{(m)} = \frac{1}{3^r} r^{\lfloor \frac{m+1}{2} \rfloor} \sum_{i=0}^{2m-\lfloor \frac{m+1}{2} \rfloor} c_{m,i}\, r^{2m-\lfloor \frac{m+1}{2} \rfloor - i}.$$

Thus, for all $i \ge 0$, one can define functions $c_i(m)$, with $m \in \mathbb{N}$, via:

$$c_i(m) := \begin{cases} c_{m,i}, & \text{for } 2m - \lfloor \frac{m+1}{2} \rfloor \ge i, \\ 0, & \text{otherwise,} \end{cases}$$

and obtains the following lemma equivalent to Lemma 6.

Lemma 7 *There exist functions* $c_0(m)$, $c_1(m)$, $c_2(m)$, $\ldots$, *such that for all* $m \ge 0$, $r \ge 0$ *the coefficients* $\alpha_r^{(m)}$ *are given as follows:*

$$\alpha_r^{(m)} = \frac{1}{3^r} r^{\lfloor \frac{m+1}{2} \rfloor} \sum_{i=0}^{2m-\lfloor \frac{m+1}{2} \rfloor} c_i(m)\, r^{2m-\lfloor \frac{m+1}{2} \rfloor - i}. \tag{18a}$$

Moreover, the function $\eta^{(m)} := c_0(m) = [r^{2m}]3^r \alpha_r^{(m)}$ *is given by*

$$\eta^{(m)} = \frac{\left(\frac{5}{9}\right)^m}{m!}. \tag{18b}$$

The formulation used in Lemma 7 has the advantage that we have indeed a description of the coefficients $\alpha_r^{(m)}$ valid for all $m \ge 0$, $r \ge 0$. However, the formulation used in Lemma 6 is the direct outcome of the proof by induction carried out in the following.

Proof: We use induction on m.
Step 0: *The initial step.* For $m = 0$ we have due to (12a) the recurrence

$$\alpha_r^{(0)} = \frac{1}{3} \alpha_{r-1}^{(0)}, \quad \text{for } r \ge 1, \qquad \alpha_0^{(0)} = 1,$$

which has the solution $\alpha_r^{(0)} = \frac{1}{3^r}$ for $r \ge 0$. Thus the given representation (17) is true for $m = 0$.
Step 1: *The recurrence.* Now we assume that for $0 \le k < m$ and given $m \ge 1$ the representation (17) holds. We will now study the coefficients $\alpha_r^{(m)}$ for m fixed as a function of r. First we write recurrence (12b), which holds for $m \le 2r - 1$, as

$$\alpha_r^{(m)} = \frac{2r}{3(2r-m)} \alpha_{r-1}^{(m)} + \frac{1}{3(2r-m)} \Bigg[4r\alpha_{r-1}^{(m-1)} + 2r(2r-m)\alpha_{r-1}^{(m-2)} + \sum_{k=1}^{m} 2\binom{r}{r-1-k} \alpha_{r-1-k}^{(m-k)}$$

$$+ \sum_{k=0}^{m-1} 2(k+1)\tbinom{r}{r-1-k}\alpha_{r-1-k}^{(m-1-k)}\Bigg] \tag{19}$$
$$+ \sum_{l=0}^{m-1} \frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}}\Bigg[4r\alpha_{r-1}^{(l-1)} + 2r(2r-l)\alpha_{r-1}^{(l-2)} + \sum_{k=0}^{l} 2\tbinom{r}{r-1-k}\alpha_{r-1-k}^{(l-k)}$$
$$+ \sum_{k=0}^{l-1} 2(k+1)\tbinom{r}{r-1-k}\alpha_{r-1-k}^{(l-1-k)}\Bigg].$$

Step 2: *Simplifying the recurrence.* In the following we will examine every term of the inhomogeneous part of this first order linear recurrence for $\alpha_r^{(m)}$ using the induction hypothesis, where we consider m as fixed as mentioned before. We start with

$$4r\alpha_{r-1}^{(m-1)} = \frac{1}{3^r} r(r-1)^{\underline{\lfloor \frac{m}{2} \rfloor}}\mathfrak{Pol}_r(2m-2-\lfloor \tfrac{m}{2} \rfloor) = \frac{1}{3^r} r^{\underline{1+\lfloor \frac{m}{2} \rfloor}}\mathfrak{Pol}_r(2m-2-\lfloor \tfrac{m}{2} \rfloor), \tag{20}$$

$$\begin{aligned} 2r(2r-m)\alpha_{r-1}^{(m-2)} &= \frac{1}{3^r} r(r-\tfrac{m}{2})(r-1)^{\underline{\lfloor \frac{m-1}{2} \rfloor}}\mathfrak{Pol}_r(2m-4-\lfloor \tfrac{m-1}{2} \rfloor) \\ &= \frac{1}{3^r}(r-\tfrac{m}{2}) r^{\underline{1+\lfloor \frac{m-1}{2} \rfloor}}\mathfrak{Pol}_r(2m-4-\lfloor \tfrac{m-1}{2} \rfloor) = \frac{1}{3^r} r^{\underline{1+\lfloor \frac{m}{2} \rfloor}}\mathfrak{Pol}_r(2m-3-\lfloor \tfrac{m}{2} \rfloor), \end{aligned} \tag{21}$$

where we combined in the last equation the cases m even and m odd.

Next we consider

$$\begin{aligned} 2\binom{r}{r-1-k}\alpha_{r-1-k}^{(m-k)} &= \frac{1}{3^r} r^{\underline{1+k}}(r-1-k)^{\underline{\lfloor \frac{m-k+1}{2} \rfloor}}\mathfrak{Pol}_r(2m-2k-\lfloor \tfrac{m-k+1}{2} \rfloor) \\ &= \frac{1}{3^r} r^{\underline{k+1+\lfloor \frac{m-k+1}{2} \rfloor}}\mathfrak{Pol}_r(2m-2k-\lfloor \tfrac{m-k+1}{2} \rfloor) \\ &= \frac{1}{3^r} r^{\underline{2+\lfloor \frac{m}{2} \rfloor}}\mathfrak{Pol}_r(k+1+\lfloor \tfrac{m-k+1}{2} \rfloor - 2 - \lfloor \tfrac{m}{2} \rfloor + 2m - 2k - \lfloor \tfrac{m-k+1}{2} \rfloor) \\ &= \frac{1}{3^r} r^{\underline{2+\lfloor \frac{m}{2} \rfloor}}\mathfrak{Pol}_r(2m-1-k-\lfloor \tfrac{m}{2} \rfloor), \quad \text{for } 1 \le k \le m. \end{aligned}$$

Thus we get

$$\sum_{k=1}^{m} 2\tbinom{r}{r-1-k}\alpha_{r-1-k}^{(m-k)} = \frac{1}{3^r} r^{\underline{2+\lfloor \frac{m}{2} \rfloor}}\mathfrak{Pol}_r(2m-2-\lfloor \tfrac{m}{2} \rfloor).$$

For computing the leading coefficient $\eta^{(m)}$ it is necessary to determine the leading coefficient of the appearing polynomial. It is seen easily that only the case $k=1$ gives a contribution and we obtain:

$$\begin{aligned} [r^{2m}]3^r \sum_{k=1}^{m} 2\tbinom{r}{r-1-k}\alpha_{r-1-k}^{(m-k)} &= 2[r^{2m}]3^r \tbinom{r}{r-2}\alpha_{r-2}^{(m-1)} \\ &= [r^{2m}]3^r r^{\underline{2}}\frac{1}{3^{r-2}}(r-2)^{\underline{\lfloor \frac{m}{2} \rfloor}}\Big(\eta^{(m-1)} r^{2m-2-\lfloor \frac{m}{2} \rfloor} + \mathfrak{Pol}_r(2m-3-\lfloor \tfrac{m}{2} \rfloor)\Big) \\ &= 9[r^{2m}] r^{\underline{2+\lfloor \frac{m}{2} \rfloor}}\Big(\eta^{(m-1)} r^{2m-2-\lfloor \frac{m}{2} \rfloor} + \mathfrak{Pol}_r(2m-3-\lfloor \tfrac{m}{2} \rfloor)\Big) = 9\eta^{(m-1)}. \end{aligned}$$

Combining these computations we get

$$\sum_{k=1}^{m} 2\binom{r}{r-1-k}\alpha_{r-1-k}^{(m-k)} = \frac{1}{3^r} r^{\underline{2+\lfloor\frac{m}{2}\rfloor}}\Big(9\eta^{(m-1)}r^{2m-2-\lfloor\frac{m}{2}\rfloor} + \mathfrak{Pol}_r(2m-3-\lfloor\tfrac{m}{2}\rfloor)\Big).$$

One shows similarly that

$$2(k+1)\binom{r}{r-1-k}\alpha_{r-1-k}^{(m-1-k)} = \frac{1}{3^r} r^{\underline{1+\lfloor\frac{m}{2}\rfloor}}\mathfrak{Pol}_r(2m-2-k-\lfloor\tfrac{m}{2}\rfloor), \quad \text{for } 0 \le k \le m-1,$$

and thus

$$\sum_{k=0}^{m-1} 2(k+1)\binom{r}{r-1-k}\alpha_{r-1-k}^{(m-1-k)} = \frac{1}{3^r} r^{\underline{1+\lfloor\frac{m}{2}\rfloor}}\mathfrak{Pol}_r(2m-2-\lfloor\tfrac{m}{2}\rfloor). \tag{22}$$

Next we consider for $0 \le l \le m-1$:

$$4r\alpha_{r-1}^{(l-1)} = \frac{1}{3^r} r(r-1)^{\underline{\lfloor\frac{l}{2}\rfloor}}\mathfrak{Pol}_r(2l-2-\lfloor\tfrac{l}{2}\rfloor) = \frac{1}{3^r} r^{\underline{1+\lfloor\frac{l}{2}\rfloor}}\mathfrak{Pol}_r(2l-2-\lfloor\tfrac{l}{2}\rfloor),$$

and thus

$$\frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} 4r\alpha_{r-1}^{(l-1)} = \frac{1}{3^r}(2r-1-l)^{\underline{m-l-1}} r^{\underline{1+\lfloor\frac{l}{2}\rfloor}}\mathfrak{Pol}_r(2l-2-\lfloor\tfrac{l}{2}\rfloor).$$

We distinguish the cases l even and l odd. Fist we consider the case l even, where we use $\lfloor\frac{l}{2}\rfloor + \lfloor\frac{m-l-1}{2}\rfloor = \lfloor\frac{m-1}{2}\rfloor$ (for arbitrary integers m):

$$\begin{aligned}
&\frac{1}{3^r}(2r-1-l)^{\underline{m-l-1}} r^{\underline{1+\lfloor\frac{l}{2}\rfloor}}\mathfrak{Pol}_r(2l-2-\lfloor\tfrac{l}{2}\rfloor) \\
&\quad = \frac{1}{3^r}2^{m-l-1}(r-\tfrac{l+2}{2})^{\underline{\lfloor\frac{m-l-1}{2}\rfloor}}(r-\tfrac{l+1}{2})^{\underline{m-l-1-\lfloor\frac{m-l-1}{2}\rfloor}} r^{\underline{1+\lfloor\frac{l}{2}\rfloor}}\mathfrak{Pol}_r(2l-2-\lfloor\tfrac{l}{2}\rfloor) \\
&\quad = \frac{1}{3^r}(r-1-\lfloor\tfrac{l}{2}\rfloor)^{\underline{\lfloor\frac{m-l-1}{2}\rfloor}} r^{\underline{1+\lfloor\frac{l}{2}\rfloor}}\mathfrak{Pol}_r(m-l-1-\lfloor\tfrac{m-l-1}{2}\rfloor)\mathfrak{Pol}_r(2l-2-\lfloor\tfrac{l}{2}\rfloor) \\
&\quad = \frac{1}{3^r} r^{\underline{1+\lfloor\frac{l}{2}\rfloor+\lfloor\frac{m-l-1}{2}\rfloor}}\mathfrak{Pol}_r(m+l-3-\lfloor\tfrac{l}{2}\rfloor-\lfloor\tfrac{m-l-1}{2}\rfloor) \\
&\quad = \frac{1}{3^r} r^{\underline{\lfloor\frac{m+1}{2}\rfloor}}\mathfrak{Pol}_r(m+l-2-\lfloor\tfrac{m+1}{2}\rfloor).
\end{aligned}$$

Next we consider the case l odd, where we use $\lfloor\frac{l}{2}\rfloor + \lfloor\frac{m-l}{2}\rfloor = \lfloor\frac{m-1}{2}\rfloor$ (for arbitrary integers m):

$$\begin{aligned}
&\frac{1}{3^r}(2r-1-l)^{\underline{m-l-1}} r^{\underline{1+\lfloor\frac{l}{2}\rfloor}}\mathfrak{Pol}_r(2l-2-\lfloor\tfrac{l}{2}\rfloor) \\
&\quad = \frac{1}{3^r}2^{m-l-1}(r-\tfrac{l+1}{2})^{\underline{\lfloor\frac{m-l}{2}\rfloor}}(r-\tfrac{l+2}{2})^{\underline{m-l-1-\lfloor\frac{m-l}{2}\rfloor}} r^{\underline{1+\lfloor\frac{l}{2}\rfloor}}\mathfrak{Pol}_r(2l-2-\lfloor\tfrac{l}{2}\rfloor) \\
&\quad = \frac{1}{3^r}(r-1-\lfloor\tfrac{l}{2}\rfloor)^{\underline{\lfloor\frac{m-l}{2}\rfloor}} r^{\underline{1+\lfloor\frac{l}{2}\rfloor}}\mathfrak{Pol}_r(m-l-1-\lfloor\tfrac{m-l}{2}\rfloor)\mathfrak{Pol}_r(2l-2-\lfloor\tfrac{l}{2}\rfloor)
\end{aligned}$$

$$= \frac{1}{3^r} r^{\underline{1+\lfloor \frac{l}{2} \rfloor + \lfloor \frac{m-l}{2} \rfloor}} \mathfrak{Pol}_r(m+l-3-\lfloor \tfrac{l}{2} \rfloor - \lfloor \tfrac{m-l}{2} \rfloor) = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \mathfrak{Pol}_r(m+l-2-\lfloor \tfrac{m+1}{2} \rfloor).$$

Thus we get for all $0 \le l \le m-1$ the equation

$$\frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} 4r\alpha_{r-1}^{(l-1)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \mathfrak{Pol}_r(m+l-2-\lfloor \tfrac{m+1}{2} \rfloor),$$

which gives

$$\sum_{l=0}^{m-1} \frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} 4r\alpha_{r-1}^{(l-1)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \mathfrak{Pol}_r(2m-3-\lfloor \tfrac{m+1}{2} \rfloor). \tag{23}$$

We further have for $0 \le l \le m-1$:

$$2r(2r-l)\alpha_{r-1}^{(l-2)} = \frac{1}{3^r} r(r-\tfrac{l}{2})(r-1)^{\underline{\lfloor \frac{l-1}{2} \rfloor}} \mathfrak{Pol}_r(2l-4-\lfloor \tfrac{l-1}{2} \rfloor) = \frac{1}{3^r} r^{\underline{1+\lfloor \frac{l}{2} \rfloor}} \mathfrak{Pol}_r(2l-3-\lfloor \tfrac{l}{2} \rfloor),$$

where the last equation is obtained when distinguishing the cases l even and l odd. Similar to the last computation we obtain for $0 \le l \le m-1$ when considering the cases l even and l odd separately:

$$\frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} 2r(2r-l)\alpha_{r-1}^{(l-2)} = \frac{1}{3^r}(2r-1-l)^{\underline{m-l-1}} r^{\underline{1+\lfloor \frac{l}{2} \rfloor}} \mathfrak{Pol}_r(2l-3-\lfloor \tfrac{l}{2} \rfloor)$$
$$= \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \mathfrak{Pol}_r(m+l-3-\lfloor \tfrac{m+1}{2} \rfloor),$$

and further

$$\sum_{l=0}^{m-1} \frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} 2r(2r-l)\alpha_{r-1}^{(l-2)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \mathfrak{Pol}_r(2m-4-\lfloor \tfrac{m+1}{2} \rfloor). \tag{24}$$

Next we get for $0 \le l \le m-1$ due to considerations as above:

$$\sum_{k=0}^{l} 2\tbinom{r}{r-1-k} \alpha_{r-1-k}^{(l-k)} = \frac{1}{3^r} r^{\underline{1+\lfloor \frac{l+1}{2} \rfloor}} \mathfrak{Pol}_r(2l-\lfloor \tfrac{l+1}{2} \rfloor).$$

We are now interested in

$$\frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} \sum_{k=0}^{l} 2\tbinom{r}{r-1-k} \alpha_{r-1-k}^{(l-k)} = \frac{1}{3^r}(2r-1-l)^{\underline{m-l-1}} r^{\underline{1+\lfloor \frac{l+1}{2} \rfloor}} \mathfrak{Pol}_r(2l-\lfloor \tfrac{l+1}{2} \rfloor),$$

for $0 \le l \le m-1$. Again when distinguishing the cases l even and l odd we obtain for $0 \le l \le m-1$:

$$\frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} \sum_{k=0}^{l} 2\tbinom{r}{r-1-k} \alpha_{r-1-k}^{(l-k)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \mathfrak{Pol}_r(m+l-\lfloor \tfrac{m+1}{2} \rfloor),$$

and thus

$$\sum_{l=0}^{m-1} \frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} \sum_{k=0}^{l} 2\binom{r}{r-1-k}\alpha_{r-1-k}^{(l-k)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \mathfrak{Pol}_r(2m-1-\lfloor \tfrac{m+1}{2} \rfloor).$$

Here we are also interested in the leading coefficient of the polynomial, where it is seen easily that only the summand with $k=0$ and $l=m-1$ gives a contribution. We compute

$$\begin{aligned}
&[r^{2m-1}]3^r \sum_{l=0}^{m-1} \frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} \sum_{k=0}^{l} 2\binom{r}{r-1-k}\alpha_{r-1-k}^{(l-k)} = -[r^{2m-1}]3^r \tfrac{2}{9}\binom{r}{r-1}\alpha_{r-1}^{(m-1)} \\
&\quad = -[r^{2m-1}]\tfrac{2}{3} r^{\underline{1+\lfloor \frac{m}{2} \rfloor}} \big(\eta^{(m-1)} r^{2m-2-\lfloor \frac{m}{2} \rfloor} + \mathfrak{Pol}_r(2m-3-\lfloor \tfrac{m}{2} \rfloor)\big) = -\frac{2}{3}\eta^{(m-1)},
\end{aligned}$$

and get thus the more detailed description

$$\begin{aligned}
&\sum_{l=0}^{m-1} \frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} \sum_{k=0}^{l} 2\binom{r}{r-1-k}\alpha_{r-1-k}^{(l-k)} \\
&\qquad = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \Big(-\frac{2}{3}\eta^{(m-1)} r^{2m-1-\lfloor \frac{m+1}{2} \rfloor} + \mathfrak{Pol}_r(2m-2-\lfloor \tfrac{m+1}{2} \rfloor)\Big). \qquad (25)
\end{aligned}$$

The last term in (19) can be treated analogous to the previous cases and we get for $0 \le l \le m-1$:

$$\sum_{k=0}^{l-1} 2(k+1)\binom{r}{r-1-k}\alpha_{r-1-k}^{(l-1-k)} = \frac{1}{3^r} r^{\underline{1+\lfloor \frac{l}{2} \rfloor}} \mathfrak{Pol}_r(2l-2-\lfloor \tfrac{l}{2} \rfloor),$$

and further

$$\begin{aligned}
&\sum_{l=0}^{m-1} \frac{(-1)^{m-l}(2r-1-l)^{\underline{m-l-1}}}{3^{m-l+1}} \sum_{k=0}^{l-1} 2(k+1)\binom{r}{r-1-k}\alpha_{r-1-k}^{(l-1-k)} \\
&\qquad = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \mathfrak{Pol}_r(2m-3-\lfloor \tfrac{m+1}{2} \rfloor). \qquad (26)
\end{aligned}$$

Now we collect all terms using (20)–(26) and obtain then from (19) the recurrence:

$$\begin{aligned}
\alpha_r^{(m)} &= \frac{2r}{3(2r-m)}\alpha_{r-1}^{(m)} + \frac{1}{3(2r-m)} \frac{1}{3^r} r^{\underline{1+\lfloor \frac{m}{2} \rfloor}} \Big(9\eta^{(m-1)} r^{2m-1-\lfloor \frac{m}{2} \rfloor} + \mathfrak{Pol}_r(2m-2-\lfloor \tfrac{m}{2} \rfloor)\Big) \\
&\quad + \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \Big(-\frac{2}{3}\eta^{(m-1)} r^{2m-1-\lfloor \frac{m+1}{2} \rfloor} + \mathfrak{Pol}_r(2m-2-\lfloor \tfrac{m+1}{2} \rfloor)\Big). \qquad (27)
\end{aligned}$$

With the substitution $\tilde{\alpha}_r^{(m)} := 3^r \alpha_r^{(m)}$ we obtain after multiplying by 3^r the following recurrence which holds for $1 \le m \le 2r-1$:

$$\tilde{\alpha}_r^{(m)} = \frac{r}{r-\frac{m}{2}}\tilde{\alpha}_{r-1}^{(m)} + \frac{1}{r-\frac{m}{2}} r^{\underline{1+\lfloor \frac{m}{2} \rfloor}} \Big(\frac{3}{2}\eta^{(m-1)} r^{2m-1-\lfloor \frac{m}{2} \rfloor} + \mathfrak{Pol}_r(2m-2-\lfloor \tfrac{m}{2} \rfloor)\Big)$$

$$+ r^{\underline{\lfloor \frac{m+1}{2} \rfloor}}\Big(-\frac{2}{3}\eta^{(m-1)} r^{2m-1-\lfloor \frac{m+1}{2} \rfloor} + \mathfrak{Pol}_r(2m-2-\lfloor \tfrac{m+1}{2} \rfloor)\Big) \qquad (28)$$

We will now study this recurrence, where we have to distinguish the cases m even and m odd.
Step 3: *Solving the recurrence for m even.* We start with m even (and $m > 0$) and set $m = 2q$, where the recurrence (28) holds for all $r > q$. We get

$$\begin{aligned}
\tilde{\alpha}_r^{(2q)} &= \frac{r}{r-q}\tilde{\alpha}_{r-1}^{(2q)} + \frac{1}{r-q} r^{\underline{1+q}}\Big(\frac{3}{2}\eta^{(2q-1)} r^{3q-1} + \mathfrak{Pol}_r(3q-2)\Big) \\
&\qquad + r^{\underline{q}}\Big(-\frac{2}{3}\eta^{(2q-1)} r^{3q-1} + \mathfrak{Pol}_r(3q-2)\Big) \\
&= \frac{r}{r-q}\tilde{\alpha}_{r-1}^{(2q)} + r^{\underline{q}}\Big(\frac{5}{6}\eta^{(2q-1)} r^{3q-1} + \mathfrak{Pol}_r(3q-2)\Big).
\end{aligned} \qquad (29)$$

Dividing equation (29) by $r^{\underline{q}}$ leads then for $r > q$ to

$$\frac{\tilde{\alpha}_r^{(2q)}}{r^{\underline{q}}} = \frac{\tilde{\alpha}_{r-1}^{(2q)}}{(r-1)^{\underline{q}}} + \frac{5}{6}\eta^{(2q-1)} r^{3q-1} + \mathfrak{Pol}_r(3q-2) = \frac{\tilde{\alpha}_{r-1}^{(2q)}}{(r-1)^{\underline{q}}} + \frac{5}{6}\eta^{(2q-1)} r^{\underline{3q-1}} + \mathfrak{Pol}_r(3q-2).$$

Now we write the appearing polynomial of degree $3q-2$ in the basis of the falling factorials as $\sum_{l=0}^{3q-2} c_l r^{\underline{l}}$ with certain constants c_l. After solving this recurrence by summation we get

$$\begin{aligned}
\frac{\tilde{\alpha}_r^{(2q)}}{r^{\underline{q}}} &= \frac{\tilde{\alpha}_q^{(2q)}}{q^{\underline{q}}} + \sum_{k=q+1}^{r} \frac{5}{6}\eta^{(2q-1)} k^{\underline{3q-1}} + \sum_{k=q+1}^{r}\sum_{l=0}^{3q-2} c_l k^{\underline{l}} \\
&= \frac{\tilde{\alpha}_q^{(2q)}}{q^{\underline{q}}} + \frac{5}{6}\eta^{(2q-1)} \frac{(r+1)^{\underline{3q}} - (q+1)^{\underline{3q}}}{3q} + \overbrace{\sum_{l=0}^{3q-2} c_l \frac{(r+1)^{\underline{l+1}} - (q+1)^{\underline{l+1}}}{l+1}}^{\mathfrak{Pol}_r(3q-1)}.
\end{aligned}$$

Since the polynomial in r of degree $3q-1$ is zero at $r = q$ we can write

$$\frac{\tilde{\alpha}_r^{(2q)}}{r^{\underline{q}}} = \frac{\tilde{\alpha}_q^{(2q)}}{q^{\underline{q}}} + \frac{5}{6}\eta^{(2q-1)} \frac{(r+1)^{\underline{3q}} - (q+1)^{\underline{3q}}}{3q} + (r-q)\mathfrak{Pol}_r(3q-2),$$

and further

$$\tilde{\alpha}_r^{(2q)} = r^{\underline{q}}\Big(\frac{\tilde{\alpha}_q^{(2q)}}{q^{\underline{q}}} + \frac{5}{18q}\eta^{(2q-1)}\big((r+1)^{\underline{3q}} - (q+1)^{\underline{3q}}\big) + (r-q)\mathfrak{Pol}_r(3q-2)\Big). \qquad (30)$$

From the validity of the recurrence we get that this formula is valid for $r > q$. But when plugging in $r = q$, the second and third term of (30) vanish and only $\tilde{\alpha}_q^{(2q)}$ remains, thus the formula is also valid for $r = q$. Moreover, when plugging in values $0 \le r < q$, the formula gives $\tilde{\alpha}_r^{(2q)} = 0$ or equivalently $\alpha_r^{(2q)} = 0$, which matches with the definition of the $\alpha_r^{(2q)}$. Thus the formula is actually true for all $r \ge 0$, if $m = 2q$. After back substitution we can write (30) for even $m > 0$ as

$$\alpha_r^{(m)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}}\Big(\frac{5}{9m}\eta^{(m-1)} r^{2m-\lfloor \frac{m+1}{2} \rfloor} + \mathfrak{Pol}_r(2m-1-\lfloor \tfrac{m+1}{2} \rfloor)\Big), \qquad (31a)$$

and get in this case the following recurrence for the leading coefficients $\eta^{(m)}$:

$$\eta^{(m)} = \frac{5}{9m}\eta^{(m-1)}. \tag{31b}$$

Step 4: *Solving the recurrence for m odd.* The case m odd in recurrence (28) is studied next. We assume that $m = 2q+1$, then we have for $r > q$:

$$\begin{aligned}
\tilde{\alpha}_r^{(2q+1)} &= \frac{r}{r-q-\frac{1}{2}}\tilde{\alpha}_{r-1}^{(2q+1)} + \frac{1}{r-q-\frac{1}{2}} r^{\underline{1+q}}\Big(\frac{3}{2}\eta^{(2q)} r^{3q+1} + \mathfrak{Pol}_r(3q)\Big) \\
&\qquad + r^{\underline{1+q}}\Big(-\frac{2}{3}\eta^{(2q)} r^{3q} + \mathfrak{Pol}_r(3q-1)\Big) \\
&= \frac{r}{r-q-\frac{1}{2}}\tilde{\alpha}_{r-1}^{(2q+1)} + \frac{r^{\underline{1+q}}}{r-q-\frac{1}{2}}\Big(\frac{5}{6}\eta^{(2q)} r^{3q+1} + \mathfrak{Pol}_r(3q)\Big).
\end{aligned} \tag{32}$$

Multiplying (32) by the factor $\frac{(r-q-\frac{1}{2})^{\underline{r-q}}}{r^{\underline{r-q}}}$ gives for $r > q$ the recurrence

$$\begin{aligned}
\frac{\tilde{\alpha}_r^{(2q+1)}(r-q-\frac{1}{2})^{\underline{r-q}}}{r^{\underline{r-q}}} &= \frac{\tilde{\alpha}_{r-1}^{(2q+1)}(r-1-q-\frac{1}{2})^{\underline{r-1-q}}}{(r-1)^{\underline{r-1-q}}} \\
&\quad + \frac{(r-1-q-\frac{1}{2})^{\underline{r-1-q}} r^{\underline{q+1}}}{r^{\underline{r-q}}}\Big(\frac{5}{6}\eta^{(2q)} r^{3q+1} + \mathfrak{Pol}_r(3q)\Big).
\end{aligned}$$

We write the polynomial of degree $3q$ as $\sum_{l=0}^{3q} c_l r^l$ with certain constants c_l and obtain after solving this recurrence by summation the following formula which holds for $r > q$:

$$\begin{aligned}
\frac{\tilde{\alpha}_r^{(2q+1)}(r-q-\frac{1}{2})^{\underline{r-q}}}{r^{\underline{r-q}}} &= \frac{\tilde{\alpha}_q^{(2q+1)}(-\frac{1}{2})^{\underline{0}}}{q^{\underline{0}}} + \sum_{k=q+1}^{r}\frac{5}{6}\eta^{(2q)}\frac{(k-1-q-\frac{1}{2})^{\underline{k-1-q}} k^{\underline{q+1}} k^{3q+1}}{k^{\underline{k-q}}} \\
&\quad + \sum_{k=q+1}^{r}\frac{(k-1-q-\frac{1}{2})^{\underline{k-1-q}} k^{\underline{q+1}}}{k^{\underline{k-q}}}\sum_{l=0}^{3q} c_l k^l.
\end{aligned}$$

Since $\tilde{\alpha}_q^{(2q+1)} = 0$ or equivalently $\alpha_q^{(2q+1)} = 0$ per definition, we obtain for $r > q$ the formula

$$\begin{aligned}
\tilde{\alpha}_r^{(2q+1)} &= \frac{r^{\underline{r-q}}}{(r-q-\frac{1}{2})^{\underline{r-q}}}\sum_{k=q+1}^{r}\frac{5}{6}\eta^{(2q)}\frac{(k-1-q-\frac{1}{2})^{\underline{k-1-q}} k^{\underline{q+1}} k^{3q+1}}{k^{\underline{k-q}}} \\
&\quad + \frac{r^{\underline{r-q}}}{(r-q-\frac{1}{2})^{\underline{r-q}}}\sum_{k=q+1}^{r}\frac{(k-1-q-\frac{1}{2})^{\underline{k-1-q}} k^{\underline{q+1}}}{k^{\underline{k-q}}}\sum_{l=0}^{3q} c_l k^l \\
&= \sum_{k=q+1}^{r}\frac{5\eta^{(2q)} r^{\underline{r+q+1-k}} k^{3q+1}}{6(r-q-\frac{1}{2})^{\underline{r+1-k}}} + \sum_{k=q+1}^{r}\frac{r^{\underline{r+q+1-k}}}{(r-q-\frac{1}{2})^{\underline{r+1-k}}}\sum_{l=0}^{3q} c_l k^l,
\end{aligned}$$

where we use here and in the following computations the identity $x^{\underline{m+n}} = x^{\underline{m}}(x-m)^{\underline{n}}$ repeatedly. After shifting k by $-q-1$ and making a base change to the falling factorials we get the following expression with certain constants d_l:

$$\begin{aligned}\tilde{\alpha}_r^{(2q+1)} &= \frac{5}{6}\eta^{(2q)} \sum_{k=0}^{r-q-1} \frac{r^{\underline{r-k}} k^{\underline{3q+1}}}{(r-q-\frac{1}{2})^{\underline{r-q-k}}} + \sum_{l=0}^{3q} d_l \sum_{k=0}^{r-q-1} \frac{r^{\underline{r-k}} k^{\underline{l}}}{(r-q-\frac{1}{2})^{\underline{r-q-k}}} \\ &= \frac{5}{6}\eta^{(2q)} \sum_{k=0}^{r-q-1} \frac{r^{\underline{r+3q+1-k}}}{(r-q-\frac{1}{2})^{\underline{r-q-k}}} + \sum_{l=0}^{3q} d_l \sum_{k=0}^{r-q-1} \frac{r^{\underline{r+l-k}}}{(r-q-\frac{1}{2})^{\underline{r-q-k}}} \\ &= \frac{5\eta^{(2q)} r!}{6(r-q-\frac{1}{2})^{\underline{r-q}}} \sum_{k=0}^{r-q-1} \frac{(k-\frac{1}{2})^{\underline{k}} k^{\underline{3q+1}}}{k!} + \sum_{l=0}^{3q} \frac{d_l r!}{(r-q-\frac{1}{2})^{\underline{r-q}}} \sum_{k=0}^{r-q-1} \frac{(k-\frac{1}{2})^{\underline{k}} k^{\underline{l}}}{k!}. \end{aligned} \tag{33}$$

To evaluate the sums in (33) we use the identity (for s a non-negative integer, where D_z denotes the differential operator with respect to z):

$$\sum_{k=0}^{n} \frac{(k-\frac{1}{2})^{\underline{k}} k^{\underline{s}}}{k!} = [z^n] \frac{z^s}{1-z} D_z^s \frac{1}{\sqrt{1-z}} = [z^n] (s-\tfrac{1}{2})^{\underline{s}} \frac{z^s}{(1-z)^{s+\frac{3}{2}}} = \frac{(s-\frac{1}{2})^{\underline{s}} (n+\frac{1}{2})^{\underline{n-s}} n^{\underline{s}}}{n!}.$$

We get then from equation (33):

$$\begin{aligned}\tilde{\alpha}_r^{(2q+1)} = {} & \frac{5\eta^{(2q)} r! (3q+\frac{1}{2})^{\underline{3q+1}} (r-q-\frac{1}{2})^{\underline{r-4q-2}} (r-q-1)^{\underline{3q+1}}}{6(r-q-\frac{1}{2})^{\underline{r-q}} (r-q-1)!} \\ & + \sum_{l=0}^{3q} \frac{d_l r! (l-\frac{1}{2})^{\underline{l}} (r-q-\frac{1}{2})^{\underline{r-q-1-l}} (r-q-1)^{\underline{l}}}{(r-q-\frac{1}{2})^{\underline{r-q}} (r-q-1)!},\end{aligned}$$

and due to obvious simplifications the following result, which holds for $r > q$:

$$\begin{aligned}\tilde{\alpha}_r^{(2q+1)} &= \frac{5\eta^{(2q)}}{6} \frac{(3q+\frac{1}{2})^{\underline{3q+1}}}{(3q+\frac{3}{2})^{\underline{3q+2}}} r^{\underline{q+1}} (r-q-1)^{\underline{3q+1}} + \sum_{l=0}^{3q} d_l \frac{(l-\frac{1}{2})^{\underline{l}}}{(l+\frac{1}{2})^{\underline{l+1}}} r^{\underline{q+1}} (r-q-1)^{\underline{l}} \\ &= \frac{5\eta^{(2q)}}{6(3q+\frac{3}{2})} r^{\underline{q+1}} (r-q-1)^{\underline{3q+1}} + \sum_{l=0}^{3q} \frac{d_l}{l+\frac{1}{2}} r^{\underline{q+1}} (r-q-1)^{\underline{l}} \\ &= r^{\underline{q+1}} \Big(\frac{5\eta^{(2q)}}{9(2q+1)} (r-q-1)^{\underline{3q+1}} + \mathfrak{Pol}_r(3q) \Big) = r^{\underline{q+1}} \Big(\frac{5\eta^{(2q)}}{9(2q+1)} r^{3q+1} + \mathfrak{Pol}_r(3q) \Big).\end{aligned} \tag{34}$$

But this formula (34) gives $\tilde{\alpha}_r^{(2q+1)} = 0$ or equivalently $\alpha_r^{(2q+1)} = 0$ for all $r \le q$, and this matches with the definition of $\alpha_r^{(2q+1)}$. Thus (34) holds for all $r \ge 0$. Back substitution leads thus for odd m to

$$\alpha_r^{(m)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \Big(\frac{5\eta^{(m-1)}}{9m} r^{2m - \lfloor \frac{m+1}{2} \rfloor} + \mathfrak{Pol}_r(2m-1-\lfloor \tfrac{m+1}{2} \rfloor) \Big), \tag{35a}$$

and we get in this case the following recurrence for the leading coefficient $\eta^{(m)}$:

$$\eta^{(m)} = \frac{5\eta^{(m-1)}}{9m}. \tag{35b}$$

Step 5: *The leading coefficients.* Therefore the only thing remaining in the proof of Lemma 6 is to show formula (17b) for $\eta^{(m)}$, but due to formulæ (31b) and (35b) together with the initial value $\eta^{(0)} = 1$ we obtain finally

$$\eta^{(m)} = \frac{\left(\frac{5}{9}\right)^m}{m!}. \tag{36}$$

□

3.5 Asymptotic expansion of $A_r(z)$

Lemma 6 leads easily to a description of the generating functions of the factorial moments of $A_n^{[H]}$, which is given next.

Lemma 8 *The generating functions $A_r(z) := \frac{\partial^r}{\partial v^r} A(z,v)\big|_{v=1} = \sum_{n\ge1} n\,\mathbb{E}\left(\left(A_n^{[H]}\right)^{\underline{r}}\right) z^n$ have the following local expansion around the dominant singularity $z = 1$:*

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

The coefficients $a_r^{(m)}$ are for m fixed and $r \ge 0$ given by the following formula, where the same leading coefficients $\eta^{(m)} = \frac{\left(\frac{5}{9}\right)^m}{m!}$ as in (17b) *appear:*

$$a_r^{(m)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \Big(\eta^{(m)} r^{2m-\lfloor \frac{m+1}{2} \rfloor} + \mathfrak{Pol}_r(2m-1-\lfloor \tfrac{m+1}{2} \rfloor)\Big). \tag{37}$$

Remark: Again one could also use the following description of the coefficients $a_r^{(m)}$, which is equivalent to the formulation given above: there exist functions $\tilde{c}_0(m)$, $\tilde{c}_1(m)$, $\tilde{c}_2(m)$, …, such that for all $m \ge 0$, $r \ge 0$ the coefficients $a_r^{(m)}$ are given as follows:

$$a_r^{(m)} = \frac{1}{3^r} r^{\underline{\lfloor \frac{m+1}{2} \rfloor}} \sum_{i=0}^{2m-\lfloor \frac{m+1}{2} \rfloor} \tilde{c}_i(m)\, r^{2m-\lfloor \frac{m+1}{2} \rfloor - i},$$

where the function $\tilde{c}_0(m)$ is given by $\tilde{c}_0(m) = \eta^{(m)} = \frac{\left(\frac{5}{9}\right)^m}{m!}$. **Proof:** Since $A(z,v) = \frac{1}{1-z}\int_{t=0}^{z} H(t,v)dt$ due to (5), we obtain

$$A_r(z) = \frac{1}{1-z}\int_{t=0}^{z} H_r(t)dt,$$

and with Lemma 5 and the coefficients $\alpha_r^{(m)}$ appearing there we obtain further the following local expansion in a complex neighbourhood of $z = 1$:

$$A_r(z) = \sum_{m=0}^{2r} \alpha_r^{(m)} \frac{1}{1-z}\int_{t=0}^{z} \frac{1}{(1-t)^2} L(t)^{2r-m} dt + \mathcal{O}\big(\frac{1}{1-z}\big). \tag{38}$$

Since integration by parts gives

$$\begin{aligned}&\frac{1}{1-z}\int_{t=0}^{z}\frac{1}{(1-t)^2}L(t)^{2r-m}dt\\&\quad=\frac{1}{(1-z)^2}\sum_{k=0}^{2r-m}(-1)^k(2r-m)^{\underline{k}}L(z)^{2r-m-k}-\frac{1}{1-z}(-1)^{2r-m}(2r-m)!,\end{aligned}$$

we get from (38) after easy manipulations

$$\begin{aligned}A_r(z)&=\sum_{m=0}^{2r}\alpha_r^{(m)}\frac{1}{(1-z)^2}\sum_{k=0}^{2r-m}(-1)^k(2r-m)^{\underline{k}}L(z)^{2r-m-k}+\mathcal{O}\big(\frac{1}{1-z}\big)\\&=\frac{1}{(1-z)^2}\sum_{m=0}^{2r}L(z)^{2r-m}\sum_{k=0}^{m}(-1)^{m-k}(2r-k)^{\underline{m-k}}\alpha_r^{(k)}+\mathcal{O}\big(\frac{1}{1-z}\big).\end{aligned}$$

This leads to the stated asymptotic expansion of $A_r(z)$ in Lemma 8 with coefficients

$$a_r^{(m)}=\sum_{k=0}^{m}(-1)^{m-k}(2r-k)^{\underline{m-k}}\alpha_r^{(k)}. \tag{39}$$

Using the representation $\alpha_r^{(k)}=\frac{1}{3^r}r^{\underline{\lfloor\frac{k+1}{2}\rfloor}}\mathfrak{Pol}_r(2k-\lfloor\frac{k+1}{2}\rfloor)$ obtained in Lemma 6 one gets after distinguishing the cases k even and k odd for $0\le k\le m$:

$$(-1)^{m-k}(2r-k)^{\underline{m-k}}\alpha_r^{(k)}=\frac{1}{3^r}r^{\underline{\lfloor\frac{m+1}{2}\rfloor}}\mathfrak{Pol}_r(m+k-\lfloor\tfrac{m+1}{2}\rfloor),$$

and equation (39) leads to

$$a_r^{(m)}=\sum_{k=0}^{m}(-1)^{m-k}(2r-k)^{\underline{m-k}}\alpha_r^{(k)}=\frac{1}{3^r}r^{\underline{\lfloor\frac{m+1}{2}\rfloor}}\mathfrak{Pol}_r(2m-\lfloor\tfrac{m+1}{2}\rfloor). \tag{40}$$

It remains to examine the leading coefficient, where it turns out that only the summand $k=m$ gives a contribution, and by using equation (17b) we get

$$[r^{2m}]3^r a_r^{(m)}=[r^{2m}]\sum_{k=0}^{m}(-1)^{m-k}(2r-k)^{\underline{m-k}}3^r\alpha_r^{(m)}=[r^{2m}]3^r\alpha_r^{(m)}=\eta^{(m)}=\frac{\left(\frac{5}{9}\right)^m}{m!}.$$

This completes the proof of Lemma 8. □

3.6 The r-th centered moments of $A_n^{[H]}$

The last step is to compute the asymptotic growth of the r-th centered moments $\mathbb{E}\Big(\big(A_n^{[H]}-\frac{1}{3}\log^2 n\big)^r\Big)$.

Lemma 9 *The r-th centered and normalized moments of $A_n^{[H]}$ satisfy, for r fixed:*

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

Proof: The basic thing to compute these moments is of course extracting coefficients from the generating functions $A_r(z)$ asymptotically via singularity analysis, which leads to asymptotic expansions for the factorial moments: $\mathbb{E}\left(\left(A_n^{[H]}\right)^{\underline{r}}\right) = \frac{1}{n}[z^n]A_r(z)$. We will require here the full expansion given in [5, p. 227] as Theorem 3A to get (ϵ denotes an arbitrarily small constant > 0):

$$\frac{1}{n}[z^n]\frac{1}{(1-z)^2}L(z)^m = \sum_{k=0}^{m}\binom{m}{k}\kappa_k(\log n)^{m-k} + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right), \tag{41}$$

where the constants κ_k are given as follows (it is here important that they are not depending on m):

$$\kappa_k = (-1)^k \, \frac{d^k}{ds^k}\left(\frac{1}{\Gamma(-s)}\right)\bigg|_{s=-2}.$$

In particular it holds that $\kappa_0 = 1$. This expansion is necessary since every term of $A_k(z)$, $0 \le k \le r$, in the asymptotic expansion in Lemma 8 gives a contribution to the r-th centered moment $\mathbb{E}\Big(\big(A_n^{[H]} - \frac{1}{3}\log^2 n\big)^r\Big)$, and only a very detailed knowledge about the structure of the coefficients of $A_k(z)$ can explain the huge cancellations. Using equation (41), Lemma 8 leads to

$$\begin{aligned} \mathbb{E}\left(\left(A_n^{[H]}\right)^{\underline{l}}\right) &= \sum_{j=0}^{2l} a_l^{(j)}\frac{1}{n}[z^n]\frac{1}{(1-z)^2}L(z)^{2l-j} + \mathcal{O}\left(\tfrac{1}{n}\right) \\ &= \sum_{j=0}^{2l} a_l^{(j)}\sum_{i=0}^{2l-j}\kappa_i\binom{2l-j}{i}(\log n)^{2l-j-i} + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right) \\ &= \sum_{m=0}^{2l}(\log n)^{2l-m}\sum_{j=0}^{m} a_l^{(j)}\binom{2l-j}{m-j}\kappa_{m-j} + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right). \end{aligned} \tag{42}$$

The ordinary moments of $A_n^{[H]}$ are of course related to the corresponding factorial moments via

$$\begin{aligned} \mathbb{E}\left(\left(A_n^{[H]}\right)^k\right) &= \sum_{l=0}^{k}\left\{ k \atop l \right\}\mathbb{E}\left(\left(A_n^{[H]}\right)^{\underline{l}}\right) \\ &= \sum_{l=0}^{k}\left\{ k \atop l \right\}\sum_{m=0}^{2l}(\log n)^{2l-m}\sum_{j=0}^{m} a_l^{(j)}\binom{2l-j}{m-j}\kappa_{m-j} + \mathcal{O}\left(\frac{1}{n^{1-\epsilon}}\right), \end{aligned} \tag{43}$$

where $\left\{ n \atop m \right\}$ denote the Stirling numbers of second kind.

Using (43) gives then the following expansion of the centered moments:

$$\mathbb{E}\Big(\big(A_n^{[H]}-\frac{1}{3}\log^2 n\big)^r\Big)=\sum_{k=0}^{r}\binom{r}{k}(-1)^{r-k}\big(\tfrac{1}{3}\log^2 n\big)^{r-k}\mathbb{E}\left(\big(A_n^{[H]}\big)^k\right) \tag{44}$$
$$=\sum_{k=0}^{r}\binom{r}{k}(-1)^{r-k}\big(\tfrac{1}{3}\log^2 n\big)^{r-k}\sum_{l=0}^{k}\left\{ k \atop l \right\}\sum_{m=0}^{2l}(\log n)^{2l-m}\sum_{j=0}^{m}a_l^{(j)}\binom{2l-j}{m-j}\kappa_{m-j}+\mathcal{O}\big(\frac{1}{n^{1-\epsilon}}\big).$$

The crucial step now is to control the massive cancellations appearing in (44). This is managed with the detailed description of the $a_r^{(m)}$ given in Lemma 8. First we obtain after routine manipulations

$$\mathbb{E}\Big(\big(A_n^{[H]}-\frac{1}{3}\log^2 n\big)^r\Big)$$
$$=\sum_{k=0}^{r}\tbinom{r}{k}(-1)^{r-k}\frac{(\log n)^{2r-2k}}{3^{r-k}}\sum_{l=0}^{k}\left\{ k \atop k-l \right\}\sum_{m=2l}^{2k}(\log n)^{2k-m}\sum_{j=0}^{m-2l}a_{k-l}^{(j)}\tbinom{2k-2l-j}{m-2l-j}\kappa_{m-2l-j}+\mathcal{O}\big(\frac{1}{n^{1-\epsilon}}\big)$$
$$=\sum_{k=0}^{r}\tbinom{r}{k}(-1)^{r-k}\frac{1}{3^r}\sum_{m=0}^{2k}(\log n)^{2r-m}\sum_{l=0}^{\lfloor\frac{m}{2}\rfloor}\left\{ k \atop k-l \right\}\sum_{j=0}^{m-2l}3^k a_{k-l}^{(j)}\tbinom{2k-2l-j}{m-2l-j}\kappa_{m-2l-j}+\mathcal{O}\big(\frac{1}{n^{1-\epsilon}}\big)$$
$$=\sum_{m=0}^{2r}(\log n)^{2r-m}\frac{1}{3^r}\sum_{k=\lfloor\frac{m+1}{2}\rfloor}^{r}\tbinom{r}{k}(-1)^{r-k}\underbrace{\sum_{l=0}^{\lfloor\frac{m}{2}\rfloor}\left\{ k \atop k-l \right\}\sum_{j=0}^{m-2l}3^k a_{k-l}^{(j)}\tbinom{2k-2l-j}{m-2l-j}\kappa_{m-2l-j}}_{=:\,S_m(k)}+\mathcal{O}\big(\frac{1}{n^{1-\epsilon}}\big). \tag{45}$$

Now we have a closer look at the sum $S_m(k)$ in (45) for fixed m as a function of k. First we obtain for m, l, j fixed, with $0\le j\le m-2l$, after distinguishing the cases j even and j odd and using Lemma 8:

$$3^k a_{k-l}^{(j)}\binom{2k-2l-j}{m-2l-j}\kappa_{m-2l-j}=(k-l)^{\underline{\lfloor\frac{j+1}{2}\rfloor}}(2k-2l-j)^{\underline{m-2l-j}}\mathfrak{Pol}_k(2j-\lfloor\tfrac{j+1}{2}\rfloor)$$
$$=(k-l)^{\underline{\lfloor\frac{m+1}{2}\rfloor-l}}\mathfrak{Pol}_k(m+j-l-\lfloor\tfrac{m+1}{2}\rfloor),$$

and thus for m, l fixed:

$$\sum_{j=0}^{m-2l}3^k a_{k-l}^{(j)}\binom{2k-2l-j}{m-2l-j}\kappa_{m-2l-j}=(k-l)^{\underline{\lfloor\frac{m+1}{2}\rfloor-l}}\mathfrak{Pol}_k(2m-3l-\lfloor\tfrac{m+1}{2}\rfloor). \tag{46}$$

Next we require the following property of the Stirling numbers of second kind $\left\{ k \atop k-l \right\}$ for fixed l as a function of k: these functions are for every integer $l\ge 0$ a polynomial in k of degree $2l$ with zeros $0,1,\dots,l-1$ (see [7, p. 271]), thus

$$\left\{ k \atop k-l \right\}=k^{\underline{l}}\,\mathfrak{Pol}_k(l). \tag{47}$$

Combining equations (46) and (47) gives for all m, l fixed, with $0\le l\le\lfloor\frac{m}{2}\rfloor$:

$$\left\{ k \atop k-l \right\}\sum_{j=0}^{m-2l}3^k a_{k-l}^{(j)}\binom{2k-2l-j}{m-2l-j}\kappa_{m-2l-j}=k^{\underline{\lfloor\frac{m+1}{2}\rfloor}}\mathfrak{Pol}_k(2m-2l-\lfloor\tfrac{m+1}{2}\rfloor),$$

and further that, for m fixed, $S_m(k)$ is a polynomial in k of degree (at most) $2m$ with the following factorization:

$$S_m(k) = k^{\lfloor \frac{m+1}{2} \rfloor} \mathfrak{Pol}_k(2m - \lfloor \tfrac{m+1}{2} \rfloor). \tag{48}$$

Of course there is again the equivalent formulation that there exist functions $f_0(m)$, $f_1(m)$, $f_2(m)$, $\dots$, such that we have for all $m \geq 0$, $k \geq 0$ the representation

$$S_m(k) = k^{\lfloor \frac{m+1}{2} \rfloor} \sum_{i=0}^{2m - \lfloor \frac{m+1}{2} \rfloor} f_i(m)\, k^{2m - \lfloor \frac{m+1}{2} \rfloor - i}.$$

Later on we will require also the leading coefficient $[k^{2m}]S_m(k)$ of $S_m(k)$, i.e., the function $f_0(m)$. It is seen easily from the computations above that the only summand that gives a contribution to the main term of order k^{2m} is obtained for the values $l = 0$ and $j = m$ in the sum defining $S_m(k)$. We get thus for all $m \geq 0$:

$$\begin{aligned}
[k^{2m}]S_m(k) &= [k^{2m}] \sum_{l=0}^{\lfloor \frac{m}{2} \rfloor} \begin{Bmatrix} k \\ k-l \end{Bmatrix} \sum_{j=0}^{m-2l} 3^k a_{k-l}^{(j)} \tbinom{2k-2l-j}{m-2l-j} \kappa_{m-2l-j} = [k^{2m}] \begin{Bmatrix} k \\ k \end{Bmatrix} 3^k a_k^{(m)} \tbinom{2k-m}{0} \kappa_0 \\
&= [k^{2m}] 3^k a_k^{(m)} = \eta^{(m)} = \frac{\left(\frac{5}{9}\right)^m}{m!}.
\end{aligned} \tag{49}$$

Now we evaluate the following sum $F(r,m)$ in equation (45) for arbitrary pairs (r,m) with $r \geq 0$, $m \geq 0$:

$$F(r,m) := \sum_{k=\lfloor \frac{m+1}{2} \rfloor}^{r} \binom{r}{k} (-1)^{r-k} S_m(k).$$

First we use that $S_m(k)$ has the factor $k^{\lfloor \frac{m+1}{2} \rfloor}$ and it holds thus $S_m(k) = 0$ for $0 \leq k < \lfloor \frac{m+1}{2} \rfloor$, which means that we can sum over the whole range[(iii)] from 0 to r:

$$F(r,m) = \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} S_m(k). \tag{50}$$

Next we use the calculus of higher order differences (see, e.g., [7]); that means we can interpret this sum as

$$F(r,m) = \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} S_m(k) = \Delta^r S_m(x)\big|_{x=0},$$

where Δ denotes the forward difference operator ($\Delta f(x) := f(x+1) - f(x)$ for functions $f(x)$). If $f(x)$ is a polynomial in x of degree at most r, i.e., $f(x) = \sum_{l=0}^{r} f_l x^l$, it holds then that

$$\sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} f(k) = \Delta^r f(x)\big|_{x=0} = r! f_r. \tag{51}$$

(iii) Now it becomes apparent, why we had to give attention to the factor $r^{\lfloor \frac{m+1}{2} \rfloor}$ of $a_r^{(m)}$ in Lemma 8. It is considerably less effort in Lemma 6 and Lemma 8 to show only the representation $a_r^{(m)} = \frac{1}{3^r} \mathfrak{Pol}_r(2m)$ for $m \leq 2r - 1$, but this is not sufficient for the proof.

But due to equation (48), for arbitrary $m \geq 0$, $S_m(x)$ is a polynomial in x of degree $2m$, which implies that the r-th power of the difference operator annihilates $S_m(x)$ for all pairs (r, m), which satisfy $r > 2m$, i.e.,

$$\sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} S_m(k) = 0, \tag{52}$$

whenever $r > 2m$ or equivalently $m \leq \lfloor \frac{r-1}{2} \rfloor$. Thus the first non-vanishing term in the expansion of $\mathbb{E}\Big(\big(A_n^{[H]} - \frac{1}{3}\log^2 n\big)^r\Big)$ as given in (45) can occur for $m = \lfloor \frac{r+1}{2} \rfloor$, which gives after distinguishing the cases r even and r odd:

$$\mathbb{E}\Big(\big(A_n^{[H]} - \frac{1}{3}\log^2 n\big)^r\Big) = \begin{cases} \mathcal{O}\big((\log n)^{\frac{3r}{2}}\big), & \text{for } r \text{ even}, \\ \mathcal{O}\big((\log n)^{\frac{3r-1}{2}}\big), & \text{for } r \text{ odd}. \end{cases}$$

To obtain the leading coefficient in this expansion for r even we use again formula (51) together with (49) and get for $r = 2m$:

$$\frac{1}{3^{2m}} \sum_{k=0}^{2m} \binom{2m}{k} (-1)^{2m-k} S_m(k) = \frac{1}{3^{2m}} (2m)! [k^{2m}] S_m(k) = \frac{1}{9^m} (2m)!\, \eta^{(m)} = \frac{(2m)! (\frac{10}{81})^m}{2^m m!}.$$

Expansion (45) gives then:

$$\mathbb{E}\Big(\big(A_n^{[H]} - \frac{1}{3}\log^2 n\big)^r\Big) = \begin{cases} \frac{(2m)!}{2^m m!} \big(\frac{10}{81}\big)^m (\log n)^{\frac{3r}{2}} + \mathcal{O}\big((\log n)^{\frac{3r}{2}-1}\big), & \text{for } r = 2m \text{ even}, \\ \mathcal{O}\big((\log n)^{\frac{3r-1}{2}}\big), & \text{for } r \text{ odd}. \end{cases} \tag{53}$$

Therefore normalization with the main term $\frac{10}{81}\log^3 n$ of the variance of $A_n^{[H]}$ proves this lemma. □

Thus we have shown in Lemma 9 that the r-th moments of $\frac{A_n^{[H]} - \frac{1}{3}\log^2 n}{\sqrt{\frac{10}{81}\log^3 n}}$ converge to the r-th moments of a standard normally distributed random variable X with

$$\mathbb{E}\big(X^r\big) = \begin{cases} \frac{(2m)!}{2^m m!}, & \text{for } r = 2m \text{ even}, \\ 0 & \text{for } r \text{ odd}. \end{cases}$$

Together with the Theorem of Fréchet and Shohat this shows the limiting distribution result for $A_n^{[H]}$ stated as Theorem 1.

4 Number of right edges to a random node

Due to the decomposition of the families $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ we obtain the following system of recurrences for the probabilities $\mathbb{P}\{A_{n,j}^{[R]} = m\}$, $\mathbb{P}\{B_{n,j}^{[R]} = m\}$ and $\mathbb{P}\{C_{n,j}^{[R]} = m\}$, which give the probabilities that the number of right edges on the path from the root to node j in a random generated object of size n are m (for $1 \leq j \leq n$):

$$\mathbb{P}\{A_{n,j}^{[R]} = m\} = \frac{1}{n} \left(\sum_{k=1}^{j-1} \mathbb{P}\{C_{n-k,j-k}^{[R]} = m\} + [\![m = 0]\!] + \sum_{k=j+1}^{n} \mathbb{P}\{B_{k-1,j}^{[R]} = m\} \right),$$

$$\mathbb{P}\{B_{n,j}^{[R]} = m\} = \frac{1}{n}\left(\sum_{k=1}^{j-1} \mathbb{P}\{A_{n-k,j-k}^{[R]} = m-1\} + [\![m=0]\!] + \sum_{k=j+1}^{n} \mathbb{P}\{B_{k-1,j}^{[R]} = m\}\right),$$
$$\mathbb{P}\{C_{n,j}^{[R]} = m\} = \frac{1}{n}\left(\sum_{k=1}^{j-1} \mathbb{P}\{C_{n-k,j-k}^{[R]} = m\} + [\![m=0]\!] + \sum_{k=j+1}^{n} \mathbb{P}\{A_{k-1,j}^{[R]} = m-1\}\right).$$

Introducing the generating function $A(z,u,v) := \sum_{n\ge1}\sum_{1\le j\le n}\sum_{m\ge0} \mathbb{P}\{A_{n,j}^{[R]} = m\} z^n u^j v^m$, and the corresponding functions $B(z,u,v)$ and $C(z,u,v)$, we can convert this system of recurrences into the following system of linear differential equations:

$$\begin{aligned}
\frac{\partial}{\partial z}A(z,u,v) &= \frac{u}{1-uz}C(z,u,v) + \frac{1}{1-z}B(z,u,v) + \frac{u}{(1-z)(1-uz)},\\
\frac{\partial}{\partial z}B(z,u,v) &= \frac{uv}{1-uz}A(z,u,v) + \frac{1}{1-z}B(z,u,v) + \frac{u}{(1-z)(1-uz)},\\
\frac{\partial}{\partial z}C(z,u,v) &= \frac{u}{1-uz}C(z,u,v) + \frac{v}{1-z}A(z,u,v) + \frac{u}{(1-z)(1-uz)},
\end{aligned}$$

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

We obtain thus the following system of differential equations for the generating function $A(z,v) := A(z,1,v) = \sum_{n\ge1}\sum_{m\ge0} n\,\mathbb{P}\{A_n^{[R]} = m\} z^n v^m$, and the corresponding auxiliary functions $B(z,v)$ and $C(z,v)$, with initial conditions $A(0,v) = B(0,v) = C(0,v) = 0$:

$$\frac{\partial}{\partial z}A(z,v) = \frac{1}{1-z}C(z,v) + \frac{1}{1-z}B(z,v) + \frac{1}{(1-z)^2}, \tag{54a}$$
$$\frac{\partial}{\partial z}B(z,v) = \frac{v}{1-z}A(z,v) + \frac{1}{1-z}B(z,v) + \frac{1}{(1-z)^2}, \tag{54b}$$
$$\frac{\partial}{\partial z}C(z,v) = \frac{1}{1-z}C(z,v) + \frac{v}{1-z}A(z,v) + \frac{1}{(1-z)^2}. \tag{54c}$$

With the substitutions $\tilde{C}(z,v) := (1-z)C(z,v)$ and $\tilde{B}(z,v) := (1-z)B(z,v)$ we obtain after adding equations (54b) and (54c):

$$\begin{aligned}
\frac{\partial}{\partial z}A(z,v) &= \frac{1}{(1-z)^2}\Big(\tilde{B}(z,v) + \tilde{C}(z,v)\Big) + \frac{1}{(1-z)^2},\\
\frac{\partial}{\partial z}\Big(\tilde{B}(z,v) + \tilde{C}(z,v)\Big) &= 2vA(z,v) + \frac{2}{1-z}.
\end{aligned}$$

Now a simple elimination leads to the following second order linear differential equation of Cauchy-Euler type for $A(z,v)$:

$$\frac{\partial^2}{\partial z^2}A(z,v) - \frac{2}{1-z}\frac{\partial}{\partial z}A(z,v) - \frac{2v}{(1-z)^2}A(z,v) = \frac{2}{(1-z)^3}, \tag{55}$$

with initial conditions $A(0,v) = 0$, $\frac{\partial}{\partial z}A(z,v)\big|_{z=0} = 1$ coming from $\mathbb{P}\{A_1^{[R]} = 0\} = 1$.

The corresponding homogeneous differential equation has the general solution

$$A^{[hom]}(z,v) = C_1(1-z)^{-\frac{1}{2}-\frac{\sqrt{1+8v}}{2}} + C_2(1-z)^{-\frac{1}{2}+\frac{\sqrt{1+8v}}{2}},$$

with arbitrary constants C_1, C_2. Variation of the parameters method gives after adapting to the initial conditions the solution

$$A(z,v) = \frac{1+2v+\sqrt{1+8v}}{2v\sqrt{1+8v}}(1-z)^{-\frac{1}{2}-\frac{\sqrt{1+8v}}{2}} - \frac{1+2v-\sqrt{1+8v}}{2v\sqrt{1+8v}}(1-z)^{-\frac{1}{2}+\frac{\sqrt{1+8v}}{2}} - \frac{1}{v(1-z)}. \quad (56)$$

Singularity analysis applied to equation (56) gives then immediately the following asymptotic expansion of the coefficients uniformly in a complex neighbourhood of $v = 1$ (ϵ denotes an arbitrary small constant > 0):

$$[z^n]A(z,v) = \frac{1+2v+\sqrt{1+8v}}{2v\sqrt{1+8v}\,\Gamma\big(\frac{1+\sqrt{1+8v}}{2}\big)} n^{\frac{\sqrt{1+8v}-1}{2}} \Big(1+\mathcal{O}\big(\frac{1}{n^{1-\epsilon}}\big)\Big).$$

Thus we get the following asymptotic expansion of the moment generating function $\mathbb{E}\big(e^{A_n^{[R]}s}\big) = \frac{1}{n}[z^n]A(z,e^s)$, which holds uniformly in a neighbourhood of $s = 0$:

$$\mathbb{E}\big(e^{A_n^{[R]}s}\big) = e^{U(s)\log n + V(s)}\Big(1+\mathcal{O}\big(\frac{1}{n^{1-\epsilon}}\big)\Big), \quad (57a)$$

with

$$U(s) = \frac{\sqrt{1+8e^s}-3}{2} \quad \text{and} \quad V(s) = \log\Big(\frac{1+2e^s+\sqrt{1+8e^s}}{2e^s\sqrt{1+8e^s}\,\Gamma\big(\frac{1+\sqrt{1+8e^s}}{2}\big)}\Big). \quad (57b)$$

An application of the quasi-power theorem (see [8]) gives then with $U'(0) = \frac{2}{3}$ and $U''(0) = \frac{10}{27}$ the stated central limit theorem for $A_n^{[R]}$ (Theorem 2).

5 Number of descendants of a random node

5.1 A system of recurrences

Here we want to obtain a system of recurrences for the probabilities $\mathbb{P}\{A_{n,j}^{[D]} = m\}$, $\mathbb{P}\{B_{n,j}^{[D]} = m\}$ and $\mathbb{P}\{C_{n,j}^{[D]} = m\}$, which give the probabilities that node j has exactly m descendants in a random generated object of size n of the families $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ for $1 \le j \le n$. To obtain such equations from the combinatorial decomposition described in Figure 2 we will condition upon the event that the node with key j of object $\mathcal{C}$ is lying on the left path, i.e., node j is lying on the path from the root to the left leaf; we write $j \in$ LP, or not, $j \notin$ LP. Furthermore we denote by $q_{n,j}$ the probability that node j in a random object of size n in family $\mathcal{C}$ is positioned on the left path, i.e., $\mathbb{P}\{j \in \text{LP}(T) : T \text{ a random object of size } n \text{ in family } \mathcal{C}\}$. As already pointed out in [13] this is equivalent to the condition that node j is a left-to-right maximum in a random permutation of size n, thus

$$q_{n,j} = \frac{1}{n-j+1}.$$

We get then directly from Figure 2 the recursive equations

$$\begin{aligned}
\mathbb{P}\{A_{n,j}^{[D]} = m\} &= \frac{1}{n}\Bigg(\sum_{k=1}^{j-1}\Big(q_{n-k,j-k}\mathbb{P}\{C_{n-k,j-k}^{[D]} = m-k \mid j-k \in \mathrm{LP}\} \\
&\quad + (1-q_{n-k,j-k})\mathbb{P}\{C_{n-k,j-k}^{[D]} = m \mid j-k \notin \mathrm{LP}\}\Big) + [\![m=j]\!] + \sum_{k=j+1}^{n}\mathbb{P}\{B_{k-1,j}^{[D]} = m\}\Bigg), \\
\mathbb{P}\{B_{n,j}^{[D]} = m\} &= \frac{1}{n}\left(\sum_{k=1}^{j-1}\mathbb{P}\{A_{n-k,j-k}^{[D]} = m\} + [\![m=j]\!] + \sum_{k=j+1}^{n}\mathbb{P}\{B_{k-1,j}^{[D]} = m\}\right), \\
\mathbb{P}\{C_{n,j}^{[D]} = m\} &= \frac{1}{n}\Bigg(\sum_{k=1}^{j-1}\Big(q_{n-k,j-k}\mathbb{P}\{C_{n-k,j-k}^{[D]} = m-k \mid j-k \in \mathrm{LP}\} \\
&\quad + (1-q_{n-k,j-k})\mathbb{P}\{C_{n-k,j-k}^{[D]} = m \mid j-k \notin \mathrm{LP}\}\Big) + [\![m=j]\!] + \sum_{k=j+1}^{n}\mathbb{P}\{A_{k-1,j}^{[D]} = m\}\Bigg).
\end{aligned}$$

If node j is lying on the left path in an object of $\mathcal{C}$ it has exactly j descendants, namely the nodes $\{1, 2, \ldots, j\}$. We get therefore

$$\mathbb{P}\{C_{n,j}^{[D]} = m \mid j \in \mathrm{LP}\} = [\![m=j]\!].$$

Together with

$$\mathbb{P}\{C_{n,j}^{[D]} = m\} = q_{n,j}\mathbb{P}\{C_{n,j}^{[D]} = m \mid j \in \mathrm{LP}\} + (1-q_{n,j})\mathbb{P}\{C_{n,j}^{[D]} = m \mid j \notin \mathrm{LP}\},$$

we get the following system of recurrences:

$$\begin{aligned}
\mathbb{P}\{A_{n,j}^{[D]} = m\} &= \frac{1}{n}\Bigg(\sum_{k=1}^{j-1}\Big(\mathbb{P}\{C_{n-k,j-k}^{[D]} = m\} + \frac{[\![m=j]\!] - [\![m=j-k]\!]}{n-j+1}\Big) \\
&\qquad + [\![m=j]\!] + \sum_{k=j+1}^{n}\mathbb{P}\{B_{k-1,j}^{[D]} = m\}\Bigg), \qquad (58a) \\
\mathbb{P}\{B_{n,j}^{[D]} = m\} &= \frac{1}{n}\left(\sum_{k=1}^{j-1}\mathbb{P}\{A_{n-k,j-k}^{[D]} = m\} + [\![m=j]\!] + \sum_{k=j+1}^{n}\mathbb{P}\{B_{k-1,j}^{[D]} = m\}\right), \qquad (58b) \\
\mathbb{P}\{C_{n,j}^{[D]} = m\} &= \frac{1}{n}\Bigg(\sum_{k=1}^{j-1}\Big(\mathbb{P}\{C_{n-k,j-k}^{[D]} = m\} + \frac{[\![m=j]\!] - [\![m=j-k]\!]}{n-j+1}\Big) \\
&\qquad + [\![m=j]\!] + \sum_{k=j+1}^{n}\mathbb{P}\{A_{k-1,j}^{[D]} = m\}\Bigg). \qquad (58c)
\end{aligned}$$

5.2 The generating functions approach

Introducing the generating functions $A(z,u,v) := \sum_{n\ge1}\sum_{1\le j\le n}\sum_{m\ge0}\mathbb{P}\{A_{n,j}^{[D]} = m\}z^n u^j v^m$, etc. we get from the system of recurrences (58) the following system of linear differential equations with initial conditions $A(0, u, v) = B(0, u, v) = C(0, u, v) = 0$:

$$\begin{aligned}
\frac{\partial}{\partial z} A(z, u, v) &= \frac{u}{1 - uz} C(z, u, v) + \frac{1}{1 - z} B(z, u, v) + R(z, u, v), \\
\frac{\partial}{\partial z} B(z, u, v) &= \frac{u}{1 - uz} A(z, u, v) + \frac{1}{1 - z} B(z, u, v) + \frac{uv}{(1 - z)(1 - uvz)}, \\
\frac{\partial}{\partial z} C(z, u, v) &= \frac{u}{1 - uz} C(z, u, v) + \frac{1}{1 - z} A(z, u, v) + R(z, u, v),
\end{aligned}$$

with

$$\begin{aligned}
R(z, u, v) = \frac{uv}{(1 - z)(1 - uvz)} &+ \frac{uv}{z(1 - uvz)^2} L(z) + \frac{uv^2}{z(1 - v)(1 - uvz)} L(z) \\
&- \frac{uv}{z(1 - v)(1 - uz)} L(z).
\end{aligned}$$

This leads to the following system of differential equations for the generating functions $A(z, v) := A(z, 1, v) = \sum_{n \geq 1} \sum_{m \geq 0} n \, \mathbb{P}\{A_n^{[D]} = m\} z^n v^m$, etc. with initial conditions $A(0, v) = B(0, v) = C(0, v) = 0$:

$$\frac{\partial}{\partial z} A(z, v) = \frac{1}{1 - z} C(z, v) + \frac{1}{1 - z} B(z, v) + R(z, v), \tag{59a}$$

$$\frac{\partial}{\partial z} B(z, v) = \frac{1}{1 - z} A(z, v) + \frac{1}{1 - z} B(z, v) + \frac{v}{(1 - z)(1 - vz)}, \tag{59b}$$

$$\frac{\partial}{\partial z} C(z, v) = \frac{1}{1 - z} C(z, v) + \frac{1}{1 - z} A(z, v) + R(z, v), \tag{59c}$$

where $R(z, v)$ is given as

$$R(z, v) = \frac{v}{(1 - z)(1 - vz)} + \frac{v}{z(1 - vz)^2} L(z) + \frac{v^2}{z(1 - v)(1 - vz)} L(z) - \frac{v}{z(1 - v)(1 - z)} L(z).$$

With the substitutions $\tilde{C}(z, v) = (1 - z) C(z, v)$ and $\tilde{B}(z, v) = (1 - z) B(z, v)$ we obtain after adding equations (59b) and (59c):

$$\begin{aligned}
\frac{\partial}{\partial z} A(z, v) &= \frac{1}{(1 - z)^2} \Big(\tilde{B}(z, v) + \tilde{C}(z, v) \Big) + R(z, v), \\
\frac{\partial}{\partial z} \Big(\tilde{B}(z, v) + \tilde{C}(z, v) \Big) &= 2A(z, v) + (1 - z) \Big(\frac{v}{(1 - z)(1 - vz)} + R(z, v) \Big),
\end{aligned}$$

and a simple elimination leads to the following second order linear differential equation of Cauchy-Euler type for $A(z, v)$:

$$\frac{\partial^2}{\partial z^2} A(z, v) - \frac{2}{1 - z} \frac{\partial}{\partial z} A(z, v) - \frac{2}{(1 - z)^2} A(z, v) = \frac{2v^2}{(1 - z)(1 - vz)^2} - \frac{2v^2(1 - v)}{(1 - z)(1 - vz)^3} L(z), \tag{60}$$

with initial conditions $A(0, v) = 0$ and $\frac{\partial}{\partial z} A(z, v) \big|_{z=0} = v$ resulting from $\mathbb{P}\{A_1^{[D]} = 1\} = 1$.

The general solution of the corresponding homogeneous differential equation is given by

$$A^{[h]}(z,v) = C_1(v)\frac{1}{(1-z)^2} + C_2(v)(1-z),$$

with arbitrary functions $C_1(v)$ and $C_2(v)$. Variation of the parameters method gives then after adapting to the initial conditions the solution of differential equation (60):

$$\begin{aligned} A(z,v) = {} & \frac{1}{3(1-z)^2}\int_{t=0}^{z}\bigg(\frac{2v^2(1-t)^2}{(1-vt)^2} - \frac{2v^2(1-v)(1-t)^2}{(1-vt)^3}L(t)\bigg)\,dt \\ & - \frac{1-z}{3}\int_{t=0}^{z}\bigg(\frac{2v^2}{(1-t)(1-vt)^2} - \frac{2v^2(1-v)}{(1-t)(1-vt)^3}L(t)\bigg)\,dt + \frac{v}{3(1-z)^2} - \frac{v(1-z)}{3}. \end{aligned} \tag{61}$$

5.3 The limiting distribution

To characterize the limiting distribution of $A_n^{[D]}$ we will first compute an explicit formula for the probabilities $\mathbb{P}\{A_n^{[D]} = m\}$. To do this we will extract coefficients from (61), where we require the following formulæ, which are obtained by routine (but somehow lengthy) computations. It holds then for $m \ge 2$:

$$\begin{aligned}
&[z^n v^m]\frac{1}{(1-z)^2}\int_{t=0}^{z}\frac{2v^2(1-t)^2}{(1-vt)^2}dt = \frac{4n}{m(m+1)} + \frac{4}{m(m+1)}, \\
&[z^n v^m](1-z)\int_{t=0}^{z}\frac{2v^2}{(1-t)(1-vt)^2}dt = -\frac{2(m-1)}{n(n-1)}, \\
&[z^n v^m]\frac{1}{(1-z)^2}\int_{t=0}^{z}\frac{2v^2(1-v)(1-t)^2}{(1-vt)^3}L(t)dt \\
&\quad = 2\binom{m}{2}\bigg(\Big[\frac{n-m+2}{m-1}H_{n+1-m} - \frac{n+1}{m-1}\big(H_n - H_{m-1}\big)\Big] \\
&\quad - 2\Big[\frac{n-m+1}{m}H_{n-m} - \frac{n+1}{m}\big(H_n - H_m\big)\Big] + \Big[\frac{n-m}{m+1}H_{n-1-m} - \frac{n+1}{m+1}\big(H_n - H_{m+1}\big)\Big]\bigg) \\
&\quad - 2\binom{m-1}{2}\bigg(\Big[\frac{n-m+3}{m-2}H_{n+2-m} - \frac{n+1}{m-2}\big(H_n - H_{m-2}\big)\Big] \\
&\quad - 2\Big[\frac{n-m+2}{m-1}H_{n+1-m} - \frac{n+1}{m-1}\big(H_n - H_{m-1}\big)\Big] + \Big[\frac{n-m+1}{m}H_{n-m} - \frac{n+1}{m}\big(H_n - H_m\big)\Big]\bigg), \\
&[z^n v^m](1-z)\int_{t=0}^{z}\frac{2v^2(1-v)}{(1-t)(1-vt)^3}L(t)dt \\
&\quad = 2\binom{m}{2}\left(\frac{H_{n-m+1}}{n} - \frac{H_{n-m}}{n-1}\right) - 2\binom{m-1}{2}\left(\frac{H_{n-m+2}}{n} - \frac{H_{n-m+1}}{n-1}\right).
\end{aligned}$$

Using these equations we obtain from (61) the following explicit formula for the probabilities $\mathbb{P}\{A_n^{[D]} = m\} = \frac{1}{n}[z^n v^m]A(z,v)$:

$$\mathbb{P}\{A_n^{[D]} = m\} = \frac{2(n-m)(n^2+mn+m^2-1)}{3n^2(n-1)m(m+1)}H_{n-m} - \frac{2(n+1)}{3nm(m+1)}\big(H_n - H_m\big) \tag{62}$$

$$+\frac{n+1}{3nm}-\frac{n-1}{3n(m+1)}-\frac{2(n+1)}{3n(m+1)^2}-\frac{2n-m-1}{3n^2(n-1)}+\frac{1}{n(n-m+1)}, \quad \text{for } m \geq 1.$$

With the asymptotic expansion $H_n = \log n + \gamma + \mathcal{O}\big(\frac{1}{n}\big)$ of the harmonic numbers we get from (62) the discrete limiting distribution of $A_n^{[D]}$ given in Theorem 3.

5.4 The r-th moments

To compute the r-th moments of $A_n^{[D]}$ we define $w := v-1$ and use

$$[z^n w^r]A(z,v) = \frac{n}{r!}\mathbb{E}\left(\left(A_n^{[D]}\right)^{\underline{r}}\right).$$

Thus we want to extract the coefficients $[z^n w^r]A(z,v)$ asymptotically for fixed r and $n \to \infty$. To do this we use the following formulæ:

$$\begin{aligned}
&[w^r]\frac{1}{3(1-z)^2}\int_{t=0}^{z}\frac{2v^2(1-t)^2}{(1-vt)^2}dt = [w^r]\frac{2}{3(1-z)^2}\int_{t=0}^{z}\frac{(w+1)^2(1-t)^2}{(1-t)^2\left(1-\frac{t}{1-t}w\right)^2}dt \\
&= \frac{2}{3}\frac{1}{(1-z)^2}\int_{t=0}^{z}\left([w^{r-2}]\frac{1}{(1-\frac{t}{1-t}w)^2}+2[w^{r-1}]\frac{1}{(1-\frac{t}{1-t}w)^2}+[w^r]\frac{1}{(1-\frac{t}{1-t}w)^2}\right)dt \\
&= \frac{2}{3}\frac{1}{(1-z)^2}\int_{t=0}^{z}\left[(r-1)\Big(\frac{t}{1-t}\Big)^{r-2}+2r\Big(\frac{t}{1-t}\Big)^{r-1}+(r+1)\Big(\frac{t}{1-t}\Big)^{r}\right]dt \\
&= \begin{cases} \frac{2}{3}\frac{1}{(1-z)^2}\left(\frac{r+1}{r-1}\frac{1}{(1-z)^{r-1}}+\mathcal{O}\big(\frac{1}{(1-z)^{r-2}}\big)\right), & \text{for } r \geq 3, \\ \frac{2}{3}\frac{1}{(1-z)^2}\left(\frac{3}{1-z}+\mathcal{O}\big(L(z)\big)\right), & \text{for } r = 2, \\ \frac{2}{3}\frac{1}{(1-z)^2}\left(2L(z)+\mathcal{O}(1)\right), & \text{for } r = 1, \end{cases} \\
&= \begin{cases} \mathcal{O}\big(\frac{1}{(1-z)^{r+1}}\big), & \text{for } r \geq 2, \\ \mathcal{O}\big(\frac{1}{(1-z)^2}L(z)\big), & \text{for } r = 1. \end{cases}
\end{aligned} \tag{63a}$$

$$\begin{aligned}
&[w^r]\frac{1}{3(1-z)^2}\int_{t=0}^{z}\frac{2v^2(1-v)(1-t)^2}{(1-vt)^3}L(t)dt \\
&= -[w^r]\frac{1}{3(1-z)^2}\int_{t=0}^{z}\frac{2(w+1)^2w(1-t)^2}{(1-t)^3(1-\frac{t}{1-t}w)^3}L(t)dt \\
&= -\frac{2}{3(1-z)^2}\int_{t=0}^{z}\frac{1}{1-t}L(t)\Big([w^{r-3}]+2[w^{r-2}]+[w^{r-1}]\Big)\frac{1}{\left(1-\frac{t}{1-t}w\right)^3}dt \\
&= -\frac{2}{3(1-z)^2}\int_{t=0}^{z}\frac{1}{1-t}L(t)\Big(\tbinom{r-1}{2}\big(\tfrac{t}{1-t}\big)^{r-3}+2\tbinom{r}{2}\big(\tfrac{t}{1-t}\big)^{r-2}+\tbinom{r+1}{2}\big(\tfrac{t}{1-t}\big)^{r-1}\Big)dt \\
&= \begin{cases} -\frac{2}{3(1-z)^2}\int_{t=0}^{z}\Big(\tbinom{r+1}{2}\frac{1}{(1-t)^r}L(t)+\mathcal{O}\big(\frac{1}{(1-t)^{r-1}}L(t)\big)\Big)dt, & \text{for } r \geq 2, \\ -\frac{2}{3(1-z)^2}\int_{t=0}^{z}\frac{1}{1-t}L(t)dt, & \text{for } r = 1, \end{cases} \\
&= \begin{cases} -\frac{r(r+1)}{3(r-1)}\frac{1}{(1-z)^{r+1}}L(z)+\mathcal{O}\big(\frac{1}{(1-z)^{r+1}}\big), & \text{for } r \geq 2, \\ -\frac{1}{3(1-z)^2}L^2(z), & \text{for } r = 1. \end{cases}
\end{aligned} \tag{63b}$$

$$[w^r]\frac{1-z}{3}\int_{t=0}^{z}\frac{2v^2}{(1-t)(1-vt)^2}dt = [w^r]\frac{2(1-z)}{3}\int_{t=0}^{z}\frac{(w+1)^2}{(1-t)^3\big(1-\frac{t}{1-t}w\big)^2}dt$$
$$= \frac{2(1-z)}{3}\int_{t=0}^{z}\frac{1}{(1-t)^3}\Big([w^{r-2}]+2[w^{r-1}]+[w^r]\Big)\frac{1}{\big(1-\frac{t}{1-t}w\big)^2}dt$$
$$= \frac{2(1-z)}{3}\int_{t=0}^{z}\Big((r+1)\frac{1}{(1-t)^{r+3}}+\mathcal{O}\big(\frac{1}{(1-t)^{r+2}}\big)\Big)dt$$
$$= \mathcal{O}\Big(\frac{1}{(1-z)^{r+1}}\Big), \quad \text{for } r \ge 1, \tag{63c}$$
$$[w^r]\frac{1-z}{3}\int_{t=0}^{z}\frac{2v^2(1-v)}{(1-t)(1-vt)^3}L(t)dt = -[w^r]\frac{1-z}{3}\int_{t=0}^{z}\frac{2(w+1)^2 w}{(1-t)^4\big(1-\frac{t}{1-t}w\big)^3}L(t)dt$$
$$= -\frac{2(1-z)}{3}\int_{t=0}^{z}\frac{1}{(1-t)^4}L(t)\Big([w^{r-3}]+2[w^{r-2}]+[w^{r-1}]\Big)\frac{1}{\big(1-\frac{t}{1-t}w\big)^3}dt$$
$$= \begin{cases} -\frac{2(1-z)}{3}\int_{t=0}^{z}\Big(\frac{(r+1)r}{2}\frac{1}{(1-t)^{r+3}}L(t)+\mathcal{O}\big(\frac{1}{(1-t)^{r+2}}L(t)\big)\Big)\,dt, & \text{for } r \ge 2,\\ -\frac{2(1-z)}{3}\int_{t=0}^{z}\frac{1}{(1-t)^4}L(t)dt, \quad \text{for } r = 1, \end{cases}$$
$$= \begin{cases} -\frac{(r+1)r}{3(r+2)}\frac{1}{(1-z)^{r+1}}L(z)+\mathcal{O}\big(\frac{1}{(1-z)^{r+1}}\big), & \text{for } r \ge 2,\\ \mathcal{O}\big(\frac{1}{(1-z)^2}L(z)\big), \quad \text{for } r = 1. \end{cases} \tag{63d}$$

Using the asymptotic expansions (63) we get from (61):

$$[w^r]A(z,v) = \begin{cases} \frac{r(r+1)}{(r-1)(r+2)}\frac{1}{(1-z)^{r+1}}L(z)+\mathcal{O}\big(\frac{1}{(1-z)^{r+1}}\big), & \text{for } r \ge 2,\\ \frac{1}{3}\frac{1}{(1-z)^2}L^2(z)+\mathcal{O}\big(\frac{1}{(1-z)^2}L(z)\big), & \text{for } r = 1. \end{cases} \tag{64}$$

Singularity analysis gives then

$$\mathbb{E}\left(\big(A_n^{[D]}\big)^{\underline{r}}\right) = \frac{r!}{n}[z^n w^r]A(z,v) = \begin{cases} \frac{r(r+1)}{(r-1)(r+2)}n^{r-1}\log n+\mathcal{O}\big(n^{r-1}\big), & \text{for } r \ge 2,\\ \frac{1}{3}\log^2 n+\mathcal{O}\big(\log n\big), \quad \text{for } r = 1. \end{cases} \tag{65}$$

Since $\mathbb{E}\left(\big(A_n^{[D]}\big)^r\right) = \mathbb{E}\left(\big(A_n^{[D]}\big)^{\underline{r}}\right) + \mathcal{O}\Big(\mathbb{E}\left(\big(A_n^{[D]}\big)^{\underline{r-1}}\right)\Big)$, the second part of Theorem 3 is also proven.

6 Conclusion

Priority trees are an easy-to-describe binary tree model, where the behaviour of basic random node-parameters is fundamentally different from the binary search tree model and the Catalan tree model.

Acknowledgement

The author thanks the anonymous referees for their comments made on an earlier version of that manuscript helping to improve the presentation of the contents.

References

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

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

[3] L. Devroye, Applications of the theory of records in the study of random trees, *Acta Informatica* 26, 123–130, 1988.

[4] M. Fisz, *Probability theory and mathematical statistics*, John Wiley, New York, 1963.

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

[6] G. Gonnet and R. Baeza-Yates, *Handbook of algorithms and data structures*, second edition, Addison-Wesley, Wokingham, 1991.

[7] R. Graham, D. Knuth and O. Patashnik, *Concrete mathematics*, second edition, Addison-Wesley, Reading, 1994.

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

[9] A. Jonassen, The stationary p-tree forest, Technical Report, Stanford University, Computer Science Department, 1976.

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

[11] M. Kuba and 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*, 175–182, SIAM, Philadelphia, 2007.

[12] A. Panholzer, *Untersuchungen zur durchschnittlichen Gestalt gewisser Baumfamilien. Mit besonderer Berücksichtigung von Anwendungen in der Informatik*, Dissertationen der Technischen Universität Wien, Band 84, Österreichischer Kunst und Kulturverlag, Wien, 1999.

[13] A. Panholzer and H. Prodinger, Average-case analysis of priority trees: a structure for priority queue administration, *Algorithmica* 22(4), 600–630, 1998.

[14] L. S. Pontryagin, *Ordinary differential equations*, Addison-Wesley, Reading, 1962.

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