

ISOLATING A LEAF IN ROOTED TREES VIA RANDOM CUTTINGS

MARKUS KUBA AND ALOIS PANHOLZER

ABSTRACT. We consider a recursive procedure for destroying rooted trees and isolating a leaf by removing a random edge and keeping the subtree, which does not contain the original root. For two tree families, the simply generated tree families and increasing tree families, we study here the number of random cuts that are necessary to isolate a leaf. We can show limiting distribution results of this parameter for simply generated trees and certain increasing trees.

1. INTRODUCTION

We consider the following edge-removal procedure in a size n rooted tree for isolating a leaf. Pick one of the $n - 1$ edges of the tree at random and remove it. This separates the tree into a pair of rooted trees; the tree containing the root of the original tree retains its root while the tree not containing the root of the original tree is rooted at the vertex adjacent to the edge that was cut. Now we discard the subtree containing the original root and continue this procedure in the other subtree, until we end at a size 1 subtree, which contains a leaf. We are going to study for several tree families under the random tree model a random variable Z_n , which counts here the number of edges that will be removed from a randomly chosen tree of size n by this edge-removal procedure until a leaf is isolated. Since all the analyzed tree families can be considered as weighted trees, this means that for starting the edge-removal procedure we choose a tree of size n with probability proportional to its weight. We can give limiting distribution results of Z_n for general simply generated tree families and some classes of so called *increasing tree families*, which contain recursive trees as a special instance (see Subsection 2.2). Surprisingly the multiple zeta function and its finite counterpart show up in the limit distribution for certain increasing trees.

This edge-removal procedure is the counterpart of another edge-removal procedure, which was already studied in more detail. In the latter procedure the subtree containing the original root of the tree is kept, while the other subtree is discarded (thus it can be seen as the opposite version of the procedure studied in the present paper) and then the procedure is continued recursively on the subtree containing the root until the original root is isolated.

Best to our knowledge the procedure studied here and the results are new, but for the sake of completeness we collect in the following some known results for the opposite procedure, which isolates the root. Meir and Moon [13, 14] considered this edge-removal procedure (= cutting-down procedure) on a rooted tree with n vertices. In papers [12,13] a random variable X_n was studied. This variable counts the number of edges that will be removed from a randomly chosen tree of size n until the root is isolated for the two important tree families *unordered labelled trees* (= Cayley trees) and *recursive trees*. For both tree families they obtained exact and asymptotic formulæ for the expectation $\mathbb{E}(X_n)$ and also asymptotic formulæ resp. bounds for the second moment $\mathbb{E}(X_n^2)$. Limiting distribution results of X_n for some classes of so called *simply generated tree families*, which contain Cayley trees as a special instance (see Subsection 2.1), are given in [16]. The problem for so called non-crossing trees was considered in [17]. The results are extended to general simply generated tree families in [9]: it turns out that for these tree families (after a suitable normalization) X_n is asymptotically Rayleigh distributed. The random tree model is always assumed as the underlying model of randomness.

We want to mention that also the following two-sided variant of the edge-removal procedure was considered in recent papers: after removing the randomly chosen edge one continues the procedure recursively

Date: June 28, 2005.

This work was supported by the Austrian Science Foundation FWF, grant P18009-N12.

on both of the obtained subtrees. Of course, when starting with a tree of size n , this two-sided variant leads to n isolated nodes after $n - 1$ cuts, but one was interested in the total costs when isolating all nodes in the tree, if one assumes that the cost incurred for selecting an edge and splitting the tree is given by a toll function t_n . For toll functions $t_n = n^\alpha$ with $\alpha > 0$, asymptotic results for all moments are obtained in [18] and limiting distribution results for some classes of simply generated tree families are given in [5]. For Cayley trees this procedure is equivalent to a probabilistic model involved in the Union-Find (or equivalence-finding) algorithm, which was analyzed first by Knuth and Schönhage [10].

Basically, to obtain our limiting distribution results for Z_n we treat the recurrences appearing for the probabilities $\mathbb{P}\{Z_n = m\}$ via bivariate generating functions. This leads to exact solvable differential equations. Extracting coefficients of the solutions appearing asymptotically is performed via singularity analysis (see [7]), a complex-analytic technique that relates asymptotics of sequences to the local behaviour of their generating functions in a neighbourhood of the dominant singularities.

Throughout this paper we use the abbreviations $x^{\underline{l}} := x(x-1)\cdots(x-l+1)$ and $x^{\overline{l}} := x(x+1)\cdots(x+l-1)$ for the falling and rising factorials, respectively. Moreover, we use the abbreviations D_x for the differential operator with respect to x , and E_x for the evaluation operator at $x = 1$.

2. PRELIMINARIES

2.1. Simply generated trees. Simply generated trees were introduced in [15] and they include several important tree families as special instances, e. g. binary trees, unordered labelled trees, and ordered trees (= planted plane trees). Moreover, they are strongly related to Galton-Watson branching processes, since it is well known (see [1]), that random simply generated trees are essentially the same as conditioned Galton-Watson trees obtained as the family tree of a Galton-Watson process conditioned on the given total size.

A class \mathcal{T} of simply generated trees can be defined in the following way. A sequence of non-negative real numbers $(\varphi_k)_{k \geq 0}$ with $\varphi_0 > 0$ (φ_k can be seen as the multiplicative weight of a node with out-degree k) is used to define the weight $w(T)$ of any ordered tree T by $w(T) := \prod_v \varphi_{d(v)}$, where v ranges over all vertices of T and $d(v)$ is the out-degree (the number of children) of v . In order to avoid degenerate cases we always assume that there exists a $k \geq 2$ such that $\varphi_k > 0$. The family \mathcal{T} consists then of all trees T with $w(T) \neq 0$ together with their weights $w(T)$. It follows further that for a given degree-weight sequence $(\varphi_k)_{k \geq 0}$ the generating function $T(z) := \sum_{n \geq 1} T_n z^n$ of the quantity total weights $T_n := \sum_{|T|=n} w(T)$, where $|T|$ denotes the size of the tree T , satisfies the functional equation

$$T(z) = z\varphi(T(z)), \quad (1)$$

where the degree-weight generating function $\varphi(t)$ is given by $\varphi(t) = \sum_{k \geq 0} \varphi_k t^k$.

The asymptotic behaviour of $T(z)$ as solution of (1) is discussed in detail in [6] and we collect some of their results concerning $T(z)$ and the growth of its coefficients T_n , where we have to make only few restrictions on $\varphi(t)$. We will suppose that $\varphi(t)$ has a positive radius of convergence $R > 0$ and assume that there exists a minimal positive solution $\tau < R$ of the equation $t\varphi'(t) = \varphi(t)$.

Defining the period $p := \gcd\{k : \varphi_k > 0\}$, it follows that equation (1) has exactly p solutions of smallest modulus given by $\tau_j = \omega^j \tau$ for $0 \leq j \leq p - 1$, where ω is a primitive p -th root of unity. This leads to p dominant singularities of $T(z)$ at $z = \rho_j$ with $\rho_j = \omega^j \rho$ and $\rho = \frac{\tau}{\varphi(\tau)} = \frac{1}{\varphi'(\tau)}$ ($T(z)$ is analytic for $|z| \leq \rho$ except at $z = \rho_j$).

The local expansion around the singularity $z = \rho_j$ is given by the following equation, where κ_j denotes a certain constant:

$$T(z) = \tau_j - \omega^j \sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}} \sqrt{1 - \frac{z}{\rho_j}} + \kappa_j \left(1 - \frac{z}{\rho_j}\right) + \mathcal{O}\left(\left(1 - \frac{z}{\rho_j}\right)^{\frac{3}{2}}\right). \quad (2)$$

By applying singularity analysis one obtains the asymptotic expansion

$$T_n = p \sqrt{\frac{\varphi(\tau)}{2\pi\varphi''(\tau)}} \rho^{-n} n^{-\frac{3}{2}} (1 + \mathcal{O}(n^{-1})), \quad (3)$$

provided that $n \equiv 1 \pmod{p}$. (For $n \not\equiv 1 \pmod{p}$ $T_n = 0$ always holds.)

We want to mention further that it is often advantageous to describe a simply generated tree family \mathcal{T} by the formal recursive equation

$$T = \circ \times \left(\varphi_0 \cdot \{\epsilon\} \dot{\cup} \varphi_1 \cdot T \dot{\cup} \varphi_2 \cdot T \times T \dot{\cup} \varphi_3 \cdot T \times T \times T \dot{\cup} \dots \right) = \circ \times \varphi(T), \quad (4)$$

with \circ a node, \times the cartesian product, and $\varphi(T)$ the substituted structure (see e. g. [19]).

2.2. Increasing trees. Increasing trees are labelled trees where the nodes of a tree of size n are labelled by distinct integers of the set $\{1, \dots, n\}$ in such a way that each sequence of labels along any branch starting at the root is increasing. As the underlying tree model we use the simply generated trees but, additionally, they are equipped with increasing labellings. We will thus speak about *simple families of increasing trees*. A thorough study of families (= varieties) of increasing trees was conducted in [3].

A class \mathcal{T} of a simple family of increasing trees can thus be defined in analogy to the definition of simply generated tree families in the following way. A sequence of non-negative numbers $(\varphi_k)_{k \geq 0}$ with $\varphi_0 > 0$ is used to define the weight $w(T)$ of any ordered tree T by $w(T) = \prod_v \varphi_{d(v)}$, where v ranges over all vertices of T and $d(v)$ is the out-degree of v (again, we always assume that there exists a $k \geq 2$ with $\varphi_k > 0$). Furthermore, $\mathcal{L}(T)$ denotes the set of different increasing labellings of the tree T with distinct integers $\{1, 2, \dots, |T|\}$, where $|T|$ denotes the size of tree T , and $L(T) := |\mathcal{L}(T)|$ denotes its cardinality. Then the family \mathcal{T} consists of all trees T together with their weights $w(T)$ and the set of increasing labellings $\mathcal{L}(T)$.

For a given degree-weight sequence $(\varphi_k)_{k \geq 0}$ with a degree-weight generating function $\varphi(t) := \sum_{k \geq 0} \varphi_k t^k$, we define now the total weights by $T_n := \sum_{|T|=n} w(T) \cdot L(T)$. It follows then that the exponential generating function $T(z) := \sum_{n \geq 1} T_n \frac{z^n}{n!}$ satisfies the *autonomous* first order differential equation

$$T'(z) = \varphi(T(z)), \quad T(0) = 0. \quad (5)$$

Again it is sometimes advantageous to describe an increasing tree family \mathcal{T} by the formal recursive equation

$$T = \textcircled{1} \times \left(\varphi_0 \cdot \{\epsilon\} \dot{\cup} \varphi_1 \cdot T \dot{\cup} \varphi_2 \cdot T * T \dot{\cup} \varphi_3 \cdot T * T * T \dot{\cup} \dots \right) = \textcircled{1} \times \varphi(T), \quad (6)$$

where additionally $*$ denotes the partition product for labelled objects.

Three specific increasing tree families are of particular interest:

- *Recursive trees* are the family of non-plane increasing trees such that all node degrees are allowed. The degree-weight generating function is $\varphi(t) = \exp(t)$. Solving (5) gives $T(z) = \log\left(\frac{1}{1-z}\right)$ and thus $T_n = (n-1)!$, for $n \geq 1$. For a survey of applications and results on random recursive trees see [12].
- *Heap ordered trees* (also called plane recursive trees) are the family of plane increasing trees such that all node degrees are allowed. The degree-weight generating function is $\varphi(t) = \frac{1}{1-t}$. Equation (5) leads here to $T(z) = 1 - \sqrt{1-2z}$ and thus to $T_n = \frac{(n-1)!}{2^{n-1}} \binom{2n-2}{n-1}$, for $n \geq 1$. See also [12] for a survey on heap ordered trees.
- *Binary increasing trees* (also called tournament trees) have the degree-weight generating function $\varphi(t) = (1+t)^2$. This model is of special importance, since it is isomorphic to the model of *binary search trees* (see [3] and the references therein for binary increasing trees and e. g. [11] for binary search trees). Thus it must follow $T(z) = \frac{z}{1-z}$ and $T_n = n!$, for $n \geq 1$.

Driven from the inspection that all these important increasing tree families satisfy the equation $\frac{T_{n+1}}{T_n} = c_1 n + c_2$, with fixed constants c_1, c_2 , for all $n \geq 1$, we will consider such trees in more detail. Throughout this paper we will call increasing tree families satisfying this equation *very simple increasing tree families*, since it turns out from the characterization given below that the defining degree-weight generating functions $\varphi(t)$ are the same as obtained in [16].

We will give now an exact answer to the question, which degree-weight generating functions are actually defining very simple increasing tree families.

Lemma 1. *The total weights T_n of trees of size n in an increasing tree family satisfy for all $n \in \mathbb{N}$ the equation*

$$\frac{T_{n+1}}{T_n} = c_1 n + c_2, \quad (7)$$

if and only if the degree-weight generating function $\varphi(t) = \sum_{k \geq 0} \varphi_k t^k$ is given by one of the following three formulae.

$$\text{Case A: } \varphi(t) = \varphi_0 e^{\frac{c_1 t}{\varphi_0}}, \text{ for } \varphi_0 > 0, c_1 > 0,$$

$$\text{Case B: } \varphi(t) = \varphi_0 \left(1 + \frac{c_2 t}{\varphi_0}\right)^d, \text{ for } \varphi_0 > 0, c_2 > 0, d := \frac{c_1}{c_2} + 1 \in \{2, 3, 4, \dots\},$$

$$\text{Case C: } \varphi(t) = \frac{\varphi_0}{\left(1 + \frac{c_2 t}{\varphi_0}\right)^{-\frac{c_1}{c_2} - 1}}, \text{ for } \varphi_0 > 0, 0 < -c_2 < c_1.$$

2.3. The recursive approach. We will study the random variable Z_n for the tree families considered by treating the recurrence

$$\mathbb{P}\{Z_n = m\} = \sum_{k=1}^{n-1} q_{n,k} \mathbb{P}\{Z_k = m-1\}, \quad \text{for } n \geq 2, m \geq 1, \quad (8)$$

with initial values $\mathbb{P}\{Z_1 = 0\} = 1$ and $\mathbb{P}\{Z_n = 0\} = 0$, for $n \geq 2$. Here the transition probabilities $q_{n,k}$ are given as follows: $q_{n,k}$ denotes the probability that by choosing a random tree of size n from the given tree family and removing a random edge the resulting subtree, which does not contain the original root of the tree, is of size k .

An analogous approach, with transition probabilities $q_{n,n-k}$, was used in [16] to study for simply generated tree families the random variable X_n , i. e. the number of cuts to isolate the root of the tree. There one had to make a strong assumption on the tree family in order to justify this recursive approach: it was necessary that randomness is preserved by cutting off a random edge, which means that starting with a random tree of size n and removing a random edge, the remaining subtree of size k containing the root is actually a *random* tree of size k in this tree family. It turned out that exactly those tree families with $\varphi(t)$ given by Lemma 1 have this property and could be treated with the recursive approach. In [16] such tree families are called very simple tree families.

For the random variable Z_n studied in the present paper things are easier. For the tree families considered it always holds that randomness is preserved by cutting off a random edge: after removing a random edge from a random tree of size n , the subtree which does not contain the original root is always a random tree of this tree family. This follows immediately from the formal recursive equations (4) and (6).

3. RESULTS

We state here our findings for simply generated tree families and very simple increasing tree families with $\varphi(t)$ satisfying the assumptions made in Subsection 2.1 and Subsection 2.2, respectively. The proof of these results are given in Section 4 and Section 5.

Theorem 1. *For simply generated tree families with degree-weight generating function $\varphi(t)$, with period p and τ the minimal positive solution of the equation $t\varphi'(t) = \varphi(t)$, the random variable Z_n , which counts the number of random cuts that are required to isolate a leaf from a randomly chosen tree of size n with the edge-removal procedure considered, converges in distribution, for $n \rightarrow \infty$ with $n \equiv 1 \pmod{p}$, to a shifted Poisson distributed random variable Z , which has the distribution*

$$\mathbb{P}\{Z = m\} = \frac{m\lambda^{m-1}}{m!} e^{-\lambda}, \quad \text{for } m \geq 0,$$

with parameter $\lambda := \log\left(\frac{\varphi(\tau)}{\varphi_0}\right)$.

Moreover, the r -th factorial moments $\mathbb{E}(Z_n^r)$ have the asymptotic expansion

$$\mathbb{E}(Z_n^r) = \lambda^{r-1}(\lambda + r) + \mathcal{O}(n^{-1}).$$

In particular, we get for the expectation $\mathbb{E}(Z_n)$ and the variance $\mathbb{V}(Z_n)$:

$$\mathbb{E}(Z_n) = \lambda + 1 + \mathcal{O}(n^{-1}), \quad \text{and} \quad \mathbb{V}(Z_n) = \lambda + \mathcal{O}(n^{-1}).$$

Theorem 2. For a very simple increasing tree family with degree-weight generating function $\varphi(t)$ given by Lemma 1 with $\frac{T_{n+1}}{T_n} = c_1 n + c_2$, for all $n \geq 1$. Let Z_n be random variable, which counts the number of random cuts that are required to isolate a leaf from a randomly chosen tree of size n with the edge-removal procedure considered. Then, for a very simple increasing tree family, Z_n converges for $n \rightarrow \infty$ in distribution to a discrete random variable Z . The probabilities $\mathbb{P}\{Z = m\}$ are for $m \geq 0$ given as the coefficients of the probability generating function $p(v) := \sum_{m \geq 0} \mathbb{P}\{Z = m\} v^m$ as given below.

$$p(v) = \prod_{k=1}^{\infty} \left(1 + \frac{(c_1 + c_2)(v-1)}{k(c_1 k + c_2)} \right) = \frac{\Gamma(1 + \frac{c_2}{c_1})}{\Gamma\left(\frac{2c_1 + c_2 - \sqrt{(2c_1 + c_2)^2 - 4c_1(c_1 + c_2)v}}{2c_1}\right) \Gamma\left(\frac{2c_1 + c_2 + \sqrt{(2c_1 + c_2)^2 - 4c_1(c_1 + c_2)v}}{2c_1}\right)}.$$

Moreover, the r -th factorial moments $\mathbb{E}(Z_n^r)$ are given by the following exact formula.

$$\mathbb{E}(Z_n^r) = r!(c_1 + c_2)^r \sum_{k_1=1}^{n-1} \frac{1}{k_1(c_1 k_1 + c_2)} \sum_{k_2=k_1+1}^{n-1} \frac{1}{k_2(c_1 k_2 + c_2)} \cdots \sum_{k_r=k_{r-1}+1}^{n-1} \frac{1}{k_r(c_1 k_r + c_2)}.$$

From Theorem 2 one gets the following corollaries, which contain results for particular increasing tree families.

Corollary 1. Using the notation of Theorem 2, we give the following closed formulæ for the probabilities $\mathbb{P}\{Z = m\}$ for recursive tree and binary increasing trees. For recursive trees we get

$$\mathbb{P}\{Z = m\} = (-1)^m \sum_{k \geq m} (-1)^k \frac{\pi^{2k} \binom{k}{m}}{(2k+1)!}, \quad (9)$$

which leads to the first few values $\mathbb{P}\{Z = 1\} = 1/2$, $\mathbb{P}\{Z = 2\} = 3/8$, $\mathbb{P}\{Z = 3\} = 5/16 - \pi^2/48$. For binary increasing trees we obtain

$$\mathbb{P}\{Z = m\} = \sum_{k \geq m+1} (-1)^k \frac{\pi^{2k}}{2^{2k} (2k)!} \sum_{j=m+1}^k \binom{k}{j} \binom{j-1}{m} (-1)^j 8^j, \quad (10)$$

which gives in particular $\mathbb{P}\{Z = 1\} = 1/3$.

Note that the values given for $\mathbb{P}\{Z = 1\}$ are just as expected, since the average number of leaves in recursive trees is $\sim \frac{n}{2}$ for recursive trees and $\sim \frac{n}{3}$ for binary increasing trees (see e. g. [3]).

Corollary 2. Using the notation of Theorem 2, we give for the instance $c_2 = 0$ the following closed formulæ for the r -th factorial moments of Z_n resp. Z . In the context of the multiple zeta functions

$$\zeta(a_1, \dots, a_l) := \sum_{1 \leq n_1 < n_2 < \dots < n_l} \frac{1}{n_1^{a_1} n_2^{a_2} \dots n_l^{a_l}}, \quad \zeta_N(a_1, \dots, a_l) := \sum_{1 \leq n_1 < n_2 < \dots < n_l \leq N} \frac{1}{n_1^{a_1} n_2^{a_2} \dots n_l^{a_l}}, \quad (11)$$

the factorial moments $\mathbb{E}(Z_n^r)$ can be expressed for $c_2 = 0$ as follows:

$$\mathbb{E}(Z_n^r) = r! \zeta_{n-1}(2, \dots, 2). \quad (12)$$

Furthermore we obtain for $c_2 = 0$ the following expression for the factorial moments $\mathbb{E}(Z^r)$:

$$\mathbb{E}(Z^r) = r! \zeta(2, \dots, 2) = r! \frac{\pi^{2r}}{(2r+1)!}. \quad (13)$$

We remark that our computations of $\mathbb{E}(Z^r)$ give thus a further proof of the identity $\zeta(\underbrace{2, \dots, 2}_{r \text{ times}}) = \frac{\pi^{2r}}{(2r+1)!}$,

which was shown first in [8].

In Table 1 and 2 we collect some results of the limiting distribution of Z_n for a few interesting simply generated tree families and very simple increasing tree families respectively.

TABLE 1. Limiting distribution results of Z_n for some important simply generated tree families.

Tree family	Degree-weight generating function $\varphi(t)$	$Z_n \rightarrow Z$, shifted Poisson distributed with parameter λ , $\mathbb{E}(Z_n) \sim \lambda + 1$, $\mathbb{V}(Z_n) \sim \lambda$
Cayley trees	$\varphi(t) = e^t$	$\lambda = 1$
d -ary trees	$\varphi(t) = (1+t)^d$, $d \geq 2$	$\lambda = d \log\left(\frac{d}{d-1}\right)$
Ordered trees	$\varphi(t) = \frac{1}{1-t}$	$\lambda = \log(2)$
Motzkin trees	$\varphi(t) = 1+t+t^2$	$\lambda = \log(3)$
Strict binary trees	$\varphi(t) = 1+t^2$	$\lambda = \log(2)$

TABLE 2. Limiting distribution results of Z_n for some important very simple increasing tree families. $H_n := \sum_{k \geq 1} \frac{1}{k}$ resp. $H_n^{(2)} := \sum_{k \geq 1} \frac{1}{k^2}$ denote the first and second order harmonic numbers.

Tree family	$\varphi(t)$	$\frac{T_{n+1}}{T_n}$	$Z_n \rightarrow Z$, with $p(v) = \sum_{m \geq 0} \mathbb{P}\{Z = m\} v^m$	$\mathbb{E}(Z_n)$
Recursive trees	e^t	n	$p(v) = \frac{\sin(\pi\sqrt{1-v})}{\pi\sqrt{1-v}}$	$H_{n-1}^{(2)} \sim \frac{\pi^2}{6} \approx 1.6449$
Binary increasing trees	$(1+t)^2$	$n+1$	$p(v) = \frac{\cos(\frac{\pi}{2}\sqrt{9-8v})}{2\pi(v-1)}$	$2 - \frac{2}{n} \sim 2$
Heap ordered trees	$\frac{1}{1-t}$	$2n-1$	$p(v) = \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{3-\sqrt{9-8v}}{4})\Gamma(\frac{3+\sqrt{9-8v}}{4})}$	$2(H_{2n-2} - H_{n-1}) \sim 2 \log 2 \approx 1.3863$

4. SIMPLY GENERATED TREE FAMILIES

4.1. The transition probabilities. The required transition probabilities $q_{n,k}$ as defined in Subsection 2.3 were already computed in [16] by a generating functions approach, which is also here sketched. We can define the value $q_{n,k}$ equivalently as the probability that the number of descendants of a node (where the node itself is counted) that was chosen at random from one of the $n-1$ non-root nodes in a random tree of size n is k . We require also the auxiliary value $\tilde{q}_{n,k}$, which denote the probability that the number of descendants of a randomly chosen node in a random tree of size n is k .

Introducing the generating functions

$$G(z, u) = \sum_{n \geq 1} \sum_{k \geq 0} n T_n \tilde{q}_{n,k} z^n u^k, \quad H(z, u) = \sum_{n \geq 1} \sum_{k \geq 0} (n-1) T_n q_{n,k} z^n u^k,$$

we can translate the formal equation (4) into the equations

$$G(z, u) = T(zu) + z\varphi'(T(z))G(z, u), \quad H(z, u) = z\varphi'(T(z))G(z, u),$$

which imply

$$H(z, u) = T(zu)F(z), \quad \text{with} \quad F(z) := \frac{1}{1 - z\varphi'(T(z))} - 1. \quad (14)$$

Extracting coefficients from (14) gives

$$F_n := [z^n]F(z) = \begin{cases} [T^n](\varphi(T))^n, & \text{for } n \geq 1, \\ 0, & \text{for } n = 0. \end{cases} \quad (15)$$

Thus the required transition probabilities $q_{n,k}$ for $1 \leq k \leq n-1$ are given as follows:

$$q_{n,k} = \frac{[z^n u^k]H(z, u)}{(n-1)T_n} = \frac{T_k F_{n-k}}{(n-1)T_n}, \quad (16)$$

with F_n defined by equation (15).

4.2. Solving the recurrence. Using (8) we have to study the recurrence

$$\mathbb{P}\{Z_n = m\} = \sum_{k=1}^{n-1} \frac{T_k F_{n-k}}{(n-1)T_n} \mathbb{P}\{Z_k = m-1\}, \quad (17)$$

with $\mathbb{P}\{Z_1 = 0\} = 1$ and $\mathbb{P}\{Z_n = 0\} = 0$, for $n \geq 2$.

We will perform a generating functions approach using the bivariate generating function

$$M(z, v) := \sum_{n \geq 1} \sum_{m \geq 0} T_n \mathbb{P}\{Z_n = m\} z^n v^m. \quad (18)$$

Multiplying (17) with $(n-1)T_n z^n v^m$ and summing up for $n \geq 2$ and $m \geq 1$ leads to the following first order linear differential equation

$$z \frac{\partial}{\partial z} M(z, v) - M(z, v) = v F(z) M(z, v),$$

with initial conditions $M(0, v) = 0$ and $(\frac{\partial}{\partial z} M(z, v))|_{z=0} = T_1 = \varphi_0$, and the function $F(z)$ given by (14). Solving this differential equation leads to the solution

$$M(z, v) = \varphi_0 z \exp\left(v \int_0^z \frac{F(t)}{t} dt\right). \quad (19)$$

Then by using $T'(z) = \frac{\varphi(T(z))}{1-z\varphi'(T(z))}$, which follows from the functional equation (1), and a change of variables, we obtain:

$$\begin{aligned} \int_0^z \frac{F(t)}{t} dt &= \int_0^{T(z)} \frac{\frac{1}{1-t\varphi'(T(t))} - 1}{t} \frac{dT}{T'(t)} = \int_0^{T(z)} \frac{\varphi'(T(t))}{T'(t)(1-t\varphi'(T(t)))} dT = \int_0^{T(z)} \frac{\varphi'(T)}{\varphi(T)} dT \\ &= \log \varphi(T(z)) - \log \varphi(0) = \log \frac{\varphi(T(z))}{\varphi_0}. \end{aligned}$$

Thus we get from (19) the following explicit formula for $M(z, v)$:

$$M(z, v) = \varphi_0 z \exp\left(v \log \frac{\varphi(T(z))}{\varphi_0}\right) = \varphi_0 z \left(\frac{\varphi(T(z))}{\varphi_0}\right)^v. \quad (20)$$

4.3. Characterizing the limiting distribution. Extracting coefficients from (20) immediately leads to

$$[v^m]M(z, v) = \varphi_0 z \frac{\left(\log \frac{\varphi(T(z))}{\varphi_0}\right)^m}{m!}. \quad (21)$$

In our asymptotic study of the coefficients $[z^n v^m]M(z, v)$ (and thus of the probabilities $\mathbb{P}\{Z_n = m\}$) via singularity analysis, which is given below, we will only carry out the instance that the degree-weight generating function $\varphi(t)$ is aperiodic, i. e. $p = 1$. But for functions $\varphi(t)$ with period $p > 1$ the proof is fully analogous: then we have to consider the contributions of all p dominant singularities, which must be added. This shows Theorem 1 also for $p > 1$.

Using the singular expansion (2) of $T(z)$ we obtain the following local expansion around the dominant singularity $z = \rho$, with certain constants $\tilde{\kappa}_1, \tilde{\kappa}_2$:

$$\frac{\varphi(T(z))}{\varphi_0} = \frac{T(z)}{\varphi_0 z} = \frac{\tau}{\rho \varphi_0} - \frac{1}{\varphi_0 \rho} \sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}} \sqrt{1 - \frac{z}{\rho}} + \tilde{\kappa}_1 \left(1 - \frac{z}{\rho}\right) + \mathcal{O}\left(\left(1 - \frac{z}{\rho}\right)^{\frac{3}{2}}\right),$$

and further

$$\log \frac{\varphi(T(z))}{\varphi_0} = \log \frac{\varphi(\tau)}{\varphi_0} - \frac{1}{\tau} \sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}} \sqrt{1 - \frac{z}{\rho}} + \tilde{\kappa}_2 \left(1 - \frac{z}{\rho}\right) + \mathcal{O}\left(\left(1 - \frac{z}{\rho}\right)^{\frac{3}{2}}\right). \quad (22)$$

Via (22) we obtain thus from (21) for $m \geq 1$ the following expansion around $z = \rho$ (again with a certain constant $\hat{\kappa}$):

$$\begin{aligned} [v^m]M(z, v) &= \sum_{n \geq 1} T_n \mathbb{P}\{Z_n = m\} z^n \\ &= \frac{\varphi_0 \rho}{m!} \left(\log \frac{\varphi(\tau)}{\varphi_0} \right)^m - \frac{\varphi_0}{\varphi(\tau)(m-1)!} \left(\log \frac{\varphi(\tau)}{\varphi_0} \right)^{m-1} \sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}} \sqrt{1 - \frac{z}{\rho}} + \hat{\kappa} \left(1 - \frac{z}{\rho}\right) + \mathcal{O}\left(\left(1 - \frac{z}{\rho}\right)^{\frac{3}{2}}\right). \end{aligned} \quad (23)$$

Applying singularity analysis to (23) gives then

$$[z^n v^m]M(z, v) = T_n \mathbb{P}\{Z_n = m\} = \frac{\varphi_0}{\varphi(\tau)(m-1)!} \left(\log \frac{\varphi(\tau)}{\varphi_0} \right)^{m-1} \sqrt{\frac{\varphi(\tau)}{2\pi\varphi''(\tau)}} \rho^{-n} n^{-\frac{3}{2}} (1 + \mathcal{O}(n^{-1})),$$

and, together with (3), for $m \geq 1$ the asymptotic expansion

$$\mathbb{P}\{Z_n = m\} = \frac{\varphi_0}{\varphi(\tau)} \frac{\left(\log \frac{\varphi(\tau)}{\varphi_0} \right)^{m-1}}{(m-1)!} (1 + \mathcal{O}(n^{-1})). \quad (24)$$

Thus the probabilities $\mathbb{P}\{Z_n = m\}$ converge for all $m \geq 1$ to the probabilities $\mathbb{P}\{Z = m\}$ of a shifted Poisson distributed random variable Z . This shows the first part of Theorem 1.

4.4. Computing the moments. From the generating function $M(z, v)$ as given by (20) we can also compute easily the r -th factorial moments $\mathbb{E}(Z^r)$.

Evaluating the r -th derivative with respect to v of $M(z, v)$ at $v = 1$ gives

$$E_v D_v^r M(z, v) = E_v \left[\varphi_0 z \left(\log \frac{\varphi(T(z))}{\varphi_0} \right)^r e^{v \log \frac{\varphi(T(z))}{\varphi_0}} \right] = z \varphi(T(z)) \left(\log \frac{\varphi(T(z))}{\varphi_0} \right)^r \quad (25)$$

$$= T(z) \left(\log \frac{\varphi(T(z))}{\varphi_0} \right)^r. \quad (26)$$

We further get by using (2) and (22) the asymptotic expansion (with a certain constant $\hat{\kappa}$)

$$\begin{aligned} T(z) \left(\log \frac{\varphi(T(z))}{\varphi_0} \right)^r &= \tau \left(\log \frac{\varphi(\tau)}{\varphi_0} \right)^r - \left(\log \frac{\varphi(\tau)}{\varphi_0} \right)^{r-1} \left(r + \log \frac{\varphi(\tau)}{\varphi_0} \right) \sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}} \sqrt{1 - \frac{z}{\rho}} \\ &\quad + \hat{\kappa} \left(1 - \frac{z}{\rho}\right) + \mathcal{O}\left(\left(1 - \frac{z}{\rho}\right)^{\frac{3}{2}}\right). \end{aligned} \quad (27)$$

Singularity analysis leads then from (26) and (27) to the asymptotic expansion

$$[z^n] E_v D_v^r M(z, v) = \left(\log \frac{\varphi(\tau)}{\varphi_0} \right)^{r-1} \left(r + \log \frac{\varphi(\tau)}{\varphi_0} \right) \sqrt{\frac{\varphi(\tau)}{2\pi\varphi''(\tau)}} \rho^{-n} n^{-\frac{3}{2}} (1 + \mathcal{O}(n^{-1})),$$

and by using (3) thus to

$$\mathbb{E}(Z_n^r) = \frac{[z^n] E_v D_v^r M(z, v)}{T_n} = \left(\log \frac{\varphi(\tau)}{\varphi_0} \right)^{r-1} \left(r + \log \frac{\varphi(\tau)}{\varphi_0} \right) (1 + \mathcal{O}(n^{-1})).$$

This completes the proof of Theorem 1.

5. VERY SIMPLE INCREASING TREE FAMILIES

5.1. The transition probabilities for general increasing trees. First we show for general increasing tree families an expression for the transition probabilities $q_{n,k}$ as defined in Subsection 2.3. We can do this analogous to Subsection 4.1 for simply generated tree families: we use the interpretation of the value $q_{n,k}$ as the probability that the number of descendants of a node that was chosen at random from one of the $n-1$ non-root nodes in a random tree of size n is k , and define the auxiliary value $\tilde{q}_{n,k}$ as the probability that the number of descendants of a randomly chosen node in a random tree of size n is k .

Introducing the generating functions

$$G(z, u) = \sum_{n \geq 1} \sum_{k \geq 0} n T_n \tilde{q}_{n,k} \frac{z^n}{n!} u^k, \quad H(z, u) = \sum_{n \geq 1} \sum_{k \geq 0} (n-1) T_n q_{n,k} \frac{z^n}{n!} u^k,$$

we obtain from the formal equation (6) (or by setting up recurrences for $\tilde{q}_{n,m}$ and $q_{n,k}$) the following differential equations:

$$\begin{aligned} \frac{\partial}{\partial z} G(z, u) &= u \varphi(T(zu)) + \varphi'(T(z)) G(z, u), \quad G(0, u) = 0, \\ \frac{\partial}{\partial z} H(z, u) &= \varphi'(T(z)) G(z, u), \quad H(0, u) = 0. \end{aligned}$$

These differential equations have the solutions

$$G(z, u) = u \varphi(T(z)) \int_0^z \frac{\varphi(T(tu))}{\varphi(T(t))} dt, \quad H(z, u) = \varphi(T(z)) \int_0^z \frac{T(tu) \varphi'(T(t))}{\varphi(T(t))} dt. \quad (28)$$

Equation (28) gives immediately

$$[z^n u^k] H(z, u) = \frac{(n-1) T_n q_{n,k}}{n!} = [z^n] \frac{T_k}{k!} \varphi(T(z)) \int_0^z \frac{t^k \varphi'(T(t))}{\varphi(T(t))} dt,$$

and thus

$$q_{n,k} = \frac{n! T_k}{(n-1) T_n k!} [z^n] \varphi(T(z)) \int_0^z \frac{t^k \varphi'(T(t))}{\varphi(T(t))} dt. \quad (29)$$

For arbitrary degree-weight generating functions $\varphi(t)$ one cannot hope to obtain explicit formulæ for the probabilities $q_{n,k}$, but for the subclass of very simple increasing tree families as given by Lemma 1 we will get an easy expression as is shown in Subsection 5.3.

5.2. Characterization of very simple increasing tree families.

Proof of Lemma 1. We will show here Lemma 1, which characterizes increasing tree families that satisfy the equation $\frac{T_{n+1}}{T_n} = c_1 n + c_2$, with arbitrary but fixed constants c_1, c_2 , for all $n \geq 1$.

We remark that due to the demand $T_n > 0$ for all $n \geq 1$ we get the a priori restrictions: $c_1 \geq 0$ and $c_2 > -c_1$ (otherwise there would exist $n \geq 1$ such that $\frac{T_{n+1}}{T_n} = c_1 n + c_2 < 0$).

• Now we consider the case $c_1 \neq 0$ and $c_2 \neq 0$ and get for T_n (where we use $T_1 = \varphi_0$):

$$\begin{aligned} T_n &= T_1 \prod_{k=1}^{n-1} (c_1 k + c_2) = \varphi_0 c_1^{n-1} \prod_{k=1}^{n-1} \left(\frac{c_2}{c_1} + k \right) = \frac{\varphi_0 c_1^n}{c_2} \left(\frac{c_2}{c_1} + n - 1 \right)^n = \frac{\varphi_0 c_1^n n!}{c_2} \binom{\frac{c_2}{c_1} + n - 1}{n} \\ &= \frac{\varphi_0 (-c_1)^n n!}{c_2} \binom{-\frac{c_2}{c_1}}{n}, \end{aligned}$$

and further

$$T(z) = \sum_{n \geq 1} T_n \frac{z^n}{n!} = \frac{\varphi_0}{c_2} \sum_{n \geq 1} \binom{-\frac{c_2}{c_1}}{n} (-c_1 z)^n = \frac{\varphi_0}{c_2} \left(\frac{1}{(1 - c_1 z)^{\frac{c_2}{c_1}}} - 1 \right). \quad (30)$$

In order to decide which values of c_1, c_2 are indeed possible choices we have to compute the corresponding degree-weight generating functions and check, whether they are admissible ($\varphi_k \geq 0$ for all $k \geq 0$). Differentiating (30) gives

$$T'(z) = \frac{\varphi_0}{(1 - c_1 z)^{\frac{c_2}{c_1} + 1}} = \varphi_0 \left(1 + \frac{c_2}{\varphi_0} T(z) \right)^{\frac{c_1}{c_2} + 1}. \quad (31)$$

We obtain $[T^n] \varphi(T) = \varphi_n$ and by using (28)

$$[T^n] T'(z) = [T^n] \varphi_0 \left(1 + \frac{c_2}{\varphi_0} T \right)^{\frac{c_1}{c_2} + 1} = \varphi_0 \binom{\frac{c_1}{c_2} + 1}{n} \left(\frac{c_2}{\varphi_0} \right)^n.$$

Since $T'(z) = \varphi(T(z))$ this gives

$$\varphi_n = \varphi_0 \binom{\frac{c_1}{c_2} + 1}{n} \left(\frac{c_2}{\varphi_0}\right)^n, \quad (32)$$

resp.

$$\varphi(t) = \sum_{n \geq 0} \varphi_n t^n = \varphi_0 \left(1 + \frac{c_2 t}{\varphi_0}\right)^{\frac{c_1}{c_2} + 1}. \quad (33)$$

By considering (32) we can now check, whether the conditions $\varphi_n \geq 0$, for all $n \geq 0$, with $\varphi_0 > 0$, are satisfied.

(i) We consider first the case $c_2 > 0$: if $1 + \frac{c_1}{c_2} \notin \mathbb{N}$, then it follows that there exists $n \in \mathbb{N}$ such that $\binom{1 + \frac{c_1}{c_2}}{n} < 0$ and, since $c_1 > 0$, thus that $\varphi_n < 0$. Therefore we get that this case is not admissible. But if $1 + \frac{c_1}{c_2} =: d \in \mathbb{N}$, then it follows that $\binom{c_1}{c_2 n} = 0$, for all $n > d$ and thus that $\varphi_n > 0$, for all $0 \leq n \leq d$ and $\varphi_n = 0$, for all $n > d$. Such degree-weight generating functions are admissible and are covered by Case B in Lemma 1.

(ii) We have to consider also the case $c_2 < 0$: since $c_1 + c_2 > 0$ it follows that $\frac{c_1}{c_2} < -1$ resp. $n - \frac{c_1}{c_2} - 2 > n - 1$ and thus that

$$\varphi_n = \varphi_0 \binom{\frac{c_1}{c_2} + 1}{n} (-1)^n \left(-\frac{c_2}{\varphi_0}\right)^n = \varphi_0 \binom{n - \frac{c_1}{c_2} - 2}{n} \left(-\frac{c_2}{\varphi_0}\right)^n > 0,$$

for all $n \geq 0$. Therefore such degree-weight generating functions are also admissible and are covered by Case C in Lemma 1.

• Next we will consider the case $c_2 = 0$ (and $c_1 > 0$), which gives

$$T_n = T_1 \prod_{k=1}^{n-1} (c_1 k) = \varphi_0 c_1^{n-1} (n-1)!,$$

and

$$T(z) = \sum_{n \geq 1} T_n \frac{z^n}{n!} = \frac{\varphi_0}{c_1} \sum_{n \geq 1} \frac{(c_1 z)^n}{n} = \frac{\varphi_0}{c_1} \log\left(\frac{1}{1 - c_1 z}\right). \quad (34)$$

Since (34) gives

$$T'(z) = \frac{\varphi_0}{1 - c_1 z} = \varphi_0 e^{\frac{c_1 T(z)}{\varphi_0}}, \quad (35)$$

we obtain

$$\varphi_n = [T^n] \varphi(T) = [T^n] T'(z) = [T^n] \varphi_0 e^{\frac{c_1 T(z)}{\varphi_0}} = \frac{\varphi_0 \left(\frac{c_1}{\varphi_0}\right)^n}{n!}, \quad (36)$$

and

$$\varphi(t) = \sum_{n \geq 0} \varphi_n t^n = \varphi_0 e^{\frac{c_1 t}{\varphi_0}}. \quad (37)$$

Since $c_1 > 0$, we obtain from (37) that $\varphi_n > 0$, for all $n \geq 0$, and thus that all degree-weight generating functions (37) are admissible. They are covered by Case A in Lemma 1.

• The remaining case is $c_1 = 0$ (and thus $c_2 > 0$), which leads to $T_n = \varphi_0 c_2^{n-1}$ and to

$$T(z) = \sum_{n \geq 1} T_n \frac{z^n}{n!} = \frac{\varphi_0}{c_2} \sum_{n \geq 1} \frac{(c_2 z)^n}{n!} = \frac{\varphi_0}{c_2} (e^{c_2 z} - 1). \quad (38)$$

Since (38) gives

$$T'(z) = \varphi_0 e^{c_2 z} = \varphi_0 + c_2 T(z), \quad (39)$$

this leads to

$$\varphi(t) = \varphi_0 + c_2 t. \quad (40)$$

This degenerate case (all trees are “chains”) is excluded from our further considerations due to the demand that there exists a $k \geq 2$ with $\varphi_k > 0$. \square

5.3. The transition probabilities. Now we are going to calculate the probabilities $q_{n,k}$ for very simple increasing tree families. Using (5) from (29) we get

$$q_{n,k} = \frac{n!T_k}{(n-1)T_n k!} [z^n] \varphi(T(z)) \int_0^z \frac{t^k \varphi'(T(t))}{\varphi(T(t))} dt = \frac{n!T_k}{(n-1)T_n k!} [z^n] T'(z) \int_0^z \frac{t^k T''(t)}{(T'(t))^2} dt. \quad (41)$$

If $c_2 \neq 0$ then we obtain from (41) via integration by parts

$$\begin{aligned} T'(z) \int_0^z \frac{t^k T''(t)}{(T'(t))^2} dt &= \frac{\varphi_0}{(1-c_1 z)^{\frac{c_2}{c_1}+1}} \int_0^z \frac{t^k \varphi_0 (c_1 + c_2) (1-c_1 t)^{\frac{2c_2}{c_1}+2}}{\varphi_0^2 (1-c_1 t)^{\frac{c_2}{c_1}+2}} dt \\ &= \frac{c_1 + c_2}{(1-c_1 z)^{\frac{c_2}{c_1}+1}} \int_0^z t^k (1-c_1 t)^{\frac{c_2}{c_1}} dt = \frac{\varphi_0 (c_1 + c_2)}{(1-c_1 z)^{\frac{c_2}{c_1}+1}} \left[\frac{k!}{T_{k+2}} - \sum_{l=0}^k \frac{k^l z^{k-l} (1-c_1 z)^{\frac{c_2}{c_1}+1+l}}{T_{l+2}} \right] \\ &= (c_1 + c_2) \left[\frac{k! T'(z)}{T_{k+2}} - \varphi_0 \sum_{l=0}^k \frac{k^l z^{k-l} (1-c_1 z)^l}{T_{l+2}} \right]. \end{aligned} \quad (42)$$

For $n > k$, combining (41) and (42) leads to

$$\begin{aligned} q_{n,k} &= \frac{n!T_k}{(n-1)T_n k!} [z^n] T'(z) \int_0^z \frac{t^k T''(t)}{(T'(t))^2} dt = \frac{n!T_k}{(n-1)T_n k!} [z^n] \frac{(c_1 + c_2) k! T'(z)}{T_{k+2}} \\ &= \frac{n!T_k}{(n-1)T_n k!} \frac{(c_1 + c_2) k! T_{n+1}}{T_{k+2} n!} = \frac{(c_1 + c_2)(c_1 n + c_2)}{(n-1)(c_1(k+1) + c_2)(c_1 k + c_2)}. \end{aligned} \quad (43)$$

It turns out that this formula also holds for $c_2 = 0$, thus covering all cases of very simple increasing tree families.

5.4. Solving the recurrence. Using (8) we obtain therefore for $n \geq 2$ and $m \geq 1$ the recurrence

$$\mathbb{P}\{Z_n = m\} = \sum_{k=1}^{n-1} \frac{(c_1 + c_2)(c_1 n + c_2)}{(n-1)(c_1(k+1) + c_2)(c_1 k + c_2)} \mathbb{P}\{Z_k = m-1\}, \quad (44)$$

with $\mathbb{P}\{Z_1 = 0\} = 1$ and $\mathbb{P}\{Z_n = 0\} = 0$, for $n \geq 2$. We simplify this full history recursion by multiplying with $\frac{n-1}{c_1 n + c_2}$ and taking differences. (44) leads then for $n \geq 1$ and $m \geq 1$ to

$$\frac{n}{c_1(n+1) + c_2} \mathbb{P}\{Z_{n+1} = m\} - \frac{n-1}{c_1 n + c_2} \mathbb{P}\{Z_n = m\} = \frac{c_1 + c_2}{(c_1(n+1) + c_2)(c_1 n + c_2)} \mathbb{P}\{Z_n = m-1\}. \quad (45)$$

Introducing the generating function

$$M(z, v) := \sum_{n \geq 1} \sum_{m \geq 0} \mathbb{P}\{Z_n = m\} \frac{z^{n-1}}{(c_1(n+1) + c_2)(c_1 n + c_2)} v^m,$$

recurrence (45) leads to the following homogeneous second order linear differential equation:

$$z(1-z) \frac{\partial^2}{\partial z^2} M(z, v) + \frac{3c_1 + c_2}{c_1} (1-z) \frac{\partial}{\partial z} M(z, v) - \frac{(c_1 + c_2)v}{c_1} M(z, v) = 0, \quad (46)$$

with initial conditions $M(0, v) = \frac{1}{(2c_1 + c_2)(c_1 + c_2)}$ and $\frac{\partial}{\partial z} M(z, v) \Big|_{z=0} = \frac{v}{(3c_1 + c_2)(2c_1 + c_2)}$. Since the hypergeometric differential equation with parameters a, b, c is given by

$$z(1-z)F''(z) + (c - (a+b+1)z)F'(z) - abF(z) = 0,$$

$M(z, v)$ satisfies the hypergeometric differential equation with parameters

$$a = \frac{2c_1 + c_2 - \sqrt{(2c_1 + c_2)^2 - 4c_1(c_1 + c_2)v}}{2c_1}, \quad b = \frac{2c_1 + c_2 + \sqrt{(2c_1 + c_2)^2 - 4c_1(c_1 + c_2)v}}{2c_1}, \quad \text{and } c = \frac{3c_1 + c_2}{c_1}.$$

A solution basis of (46) is thus given by the following two functions (see e. g. [2]):

$$\begin{aligned} {}_2F_1 \left(\begin{matrix} a, b \\ c \end{matrix} \middle| z \right) &= {}_2F_1 \left(\begin{matrix} \frac{2c_1 + c_2 - \sqrt{(2c_1 + c_2)^2 - 4c_1(c_1 + c_2)v}}{2c_1}, \frac{2c_1 + c_2 + \sqrt{(2c_1 + c_2)^2 - 4c_1(c_1 + c_2)v}}{2c_1} \\ \frac{3c_1 + c_2}{c_1} \end{matrix} \middle| z \right), \\ z^{1-c} {}_2F_1 \left(\begin{matrix} a+1-c, b+1-c \\ 2-c \end{matrix} \middle| z \right) \end{aligned}$$

$$= z^{-\frac{2c_1+c_2}{c_1}} {}_2F_1\left(\frac{-(2c_1+c_2)-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1}, \frac{-(2c_1+c_2)+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1} \middle| z\right),$$

where ${}_2F_1\left(\begin{smallmatrix} a, b \\ c \end{smallmatrix} \middle| z\right) := \sum_{n \geq 0} \frac{a^{\bar{n}} b^{\bar{n}} z^n}{c^{\bar{n}} n!}$ denotes the Gauss hypergeometric series.

Since $M(z, v)$ has a power series expansion around $z = 0$ (and $v = 0$) it must follow that

$$M(z, v) = C(v) {}_2F_1\left(\frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1}, \frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1} \middle| z\right), \quad (47)$$

with a certain function $C(v)$, since the other base function is not analytic at $z = 0$. After adapting (47) to the initial conditions we obtain the solution

$$M(z, v) = \frac{1}{(2c_1+c_2)(c_1+c_2)} {}_2F_1\left(\frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1}, \frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1} \middle| z\right). \quad (48)$$

5.5. Characterizing the limiting distribution. To obtain a limiting distribution result we will apply the following instance of the z to $1-z$ transformation (see e. g. [2]) with $m \in \{1, 2, 3, \dots\}$:

$$\begin{aligned} {}_2F_1\left(\begin{smallmatrix} a, b \\ a+b+m \end{smallmatrix} \middle| z\right) &= \frac{\Gamma(m)\Gamma(a+b+m)}{\Gamma(a+m)\Gamma(b+m)} \sum_{n=0}^{m-1} \frac{a^{\bar{n}} b^{\bar{n}}}{n!(1-m)^{\bar{n}}} (1-z)^n \\ &\quad - \frac{\Gamma(a+b+m)}{\Gamma(a)\Gamma(b)} (z-1)^m \sum_{n=0}^{\infty} \frac{(a+m)^{\bar{n}} (b+m)^{\bar{n}}}{n!(n+m)!} (1-z)^n \\ &\quad \times \left(\log(1-z) - \Psi(n+1) - \Psi(n+m+1) + \Psi(a+n+m) + \Psi(b+n+m) \right), \end{aligned}$$

and get from equation (48) the following local expansion of $M(z, v)$ around the dominant singularity $z = 1$ in a complex neighbourhood of $v = 1$ (with certain functions $C_0(v)$, $C_1(v)$, and $C_2(v)$):

$$\begin{aligned} M(z, v) &= \frac{1}{(2c_1+c_2)(c_1+c_2)} \frac{\Gamma(3+\frac{c_2}{c_1})}{\Gamma(\frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})\Gamma(\frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})} \\ &\quad \times (z-1) \log \frac{1}{1-z} + C_0(v) + C_1(v)(1-z) + C_2(v)(1-z)^2 + \mathcal{O}((1-z)^2 \log \frac{1}{1-z}). \end{aligned}$$

Singularity analysis gives then the following expansion of the probability generating function $p_n(v) := \sum_{m \geq 0} \mathbb{P}\{Z_n = m\} v^m$ of Z_n .

$$\begin{aligned} p_n(v) &= (c_1 n + c_2)(c_1(n+1) + c_2)[z^{n-1}]M(z, v) \\ &= \frac{\Gamma(1+\frac{c_2}{c_1})}{\Gamma(\frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})\Gamma(\frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})} (1 + \mathcal{O}(n^{-1})). \quad (49) \end{aligned}$$

Thus it follows from (49) that the moment generating function (= Laplace transform) $\mathbb{E}(e^{Z_n s}) = \sum_{m \geq 0} \mathbb{P}\{Z_n = m\} e^{ms} = p_n(e^s)$ of Z_n converges in a neighbourhood of $s = 0$ to the moment generating function $\mathbb{E}(e^{Z s}) = p(e^s)$ of a discrete random variable Z with probability generating function $p(v) := \sum_{m \geq 0} \mathbb{P}\{Z = m\} v^m$ given by

$$p(v) = \frac{\Gamma(1+\frac{c_2}{c_1})}{\Gamma(\frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})\Gamma(\frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})}. \quad (50)$$

We want to remark that by using the reflection law of the Gamma function

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)},$$

one can further simplify the formula of $p(v)$ for binary increasing trees and recursive trees; see Table 2. Extracting coefficients leads then for these tree families to expressions for the probabilities $\mathbb{P}\{Z = m\}$ as given in Corollary 1.

Furthermore, we can give a representation of $p(v)$ as an infinite product, where we simply use repeatedly the functional equation $\Gamma(x) = \frac{\Gamma(x+1)}{x}$ for the Gamma function expressions in (50). One obtains after n iteration steps and some simplifications:

$$\begin{aligned} p(v) &= \frac{\Gamma(1)\Gamma(1 + \frac{c_2}{c_1})}{\Gamma(\frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})\Gamma(\frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})} \\ &= f(n) \prod_{k=1}^n \left(1 + \frac{(c_1+c_2)(v-1)}{k(c_1k+c_2)}\right), \end{aligned} \quad (51)$$

with

$$f(n) = \frac{\Gamma(n+1)\Gamma(n+1 + \frac{c_1}{c_2})}{\Gamma(n + \frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})\Gamma(n + \frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})}.$$

Since it holds $f(n) \rightarrow 1$ for $n \rightarrow \infty$, as can be shown e. g. via Stirling's asymptotic formula for the Gamma function, we obtain from (51) the representation

$$p(v) = \prod_{k=1}^{\infty} \left(1 + \frac{(c_1+c_2)(v-1)}{k(c_1k+c_2)}\right). \quad (52)$$

By an application of the continuity theorem for the Laplace transform (see e. g. [4]) we obtain from equations (50) and (52) immediately the first part of Theorem 2.

5.6. Computing the moments. From the explicit formula (48) for the generating function $M(z, v)$ we can also compute exact expressions for the r -th factorial moments $\mathbb{E}(Z_n^r)$ of Z_n . To do this we will give first an exact formula for the probability generating function $p_n(v)$. By easy manipulations we obtain

$$\begin{aligned} p_n(v) &= (c_1(n+1) + c_2)(c_1n + c_2)[z^{n-1}]M(z, v) \\ &= \frac{(c_1(n+1) + c_2)(c_1n + c_2) \left(\frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1}\right)^{n-1} \left(\frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1}\right)^{n-1}}{(2c_1+c_2)(c_1+c_2) \left(\frac{3c_1+c_2}{c_1}\right)^{n-1} (n-1)!} \\ &= \frac{(c_1(n+1) + c_2)(c_1n + c_2) \prod_{k=0}^{n-2} \left(k^2 + (2 + \frac{c_2}{c_1})k + (1 + \frac{c_2}{c_1}) - (1 + \frac{c_2}{c_1})(1-v)\right)}{(2c_1+c_2)(c_1+c_2) \left(3 + \frac{c_2}{c_1}\right)^{n-1} (n-1)!} \\ &= \frac{\prod_{k=1}^{n-1} \left(k(c_1k+c_2) - (c_1+c_2)(1-v)\right)}{\prod_{k=1}^{n-1} (k(c_1k+c_2))} = \prod_{k=1}^{n-1} \left(1 + \frac{(c_1+c_2)(v-1)}{k(c_1k+c_2)}\right). \end{aligned} \quad (53)$$

Evaluating the r -th derivative of $p_n(v)$ at $v = 1$ as given by (53) leads then to:

$$\begin{aligned} \mathbb{E}(Z_n^r) &= E_v D_v^r p_n(v) = \sum_{1 \leq k_1 < k_2 < \dots < k_r \leq n-1} \frac{r!(c_1+c_2)^r}{\prod_{i=1}^r (k_i(c_1k_i+c_2))} \\ &= r!(c_1+c_2)^r \sum_{k_1=1}^{n-1} \frac{1}{k_1(c_1k_1+c_2)} \sum_{k_2=k_1+1}^{n-1} \frac{1}{k_2(c_1k_2+c_2)} \dots \sum_{k_r=k_{r-1}+1}^{n-1} \frac{1}{k_r(c_1k_r+c_2)}, \end{aligned} \quad (54)$$

which shows also the second part of Theorem 2. We remark that this result for the r -th factorial moments can also be obtained directly from the recurrence (45) by using elementary means.

If $c_2 = 0$ we get by using (11) the following result

$$\mathbb{E}(Z_n^r) = r! \zeta_{n-1}(2, \dots, 2). \quad (55)$$

Using the probability generating function $p(v)$ for $c_1 = 1$, $c_2 = 0$

$$p(v) = \lim_{n \rightarrow \infty} p_n(v) = \frac{\sin(\pi\sqrt{1-v})}{\pi\sqrt{(1-v)}} = \sum_{k \geq 0} \frac{\pi^{2k}(-1)^k(1-v)^k}{(2k+1)!}, \quad (56)$$

we obtain the r -th factorial moment of Z by differentiating r times with respect to v and evaluating at $v = 1$:

$$\mathbb{E}(Z^{\underline{r}}) = E_v D_v^r p(v) = \frac{r! \pi^{2r}}{(2r+1)!}. \quad (57)$$

Since (12) yields $\mathbb{E}(Z^{\underline{r}}) = r! \zeta(\underbrace{2, 2, \dots, 2}_{r \text{ times}})$, the proof of Corollary 2 is finished.

REFERENCES

- [1] D. Aldous, The Continuum Random Tree II: An Overview, *London Mathematical Society Lecture Note Series* 167, 23–70, 1991.
- [2] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, ninth edition, Dover, New York, 1970.
- [3] F. Bergeron, P. Flajolet and B. Salvy, Varieties of Increasing Trees, *Lecture Notes in Computer Science* 581, 24–48, 1992.
- [4] K. L. Chung, *A course in probability theory*, Academic Press, San Diego, 1974.
- [5] J. A. Fill, N. Kapur and A. Panholzer, Destruction of very simple trees, *Algorithmica*, to appear, 2004.
- [6] P. Flajolet and A. Odlyzko, The Average Height of Binary Trees and Other Simple Trees, *Journal of Computer and System Sciences* 25, 171–213, 1982.
- [7] P. Flajolet and A. Odlyzko, Singularity Analysis of Generating Functions, *SIAM Journal on Discrete Mathematics* 3, 216–240, 1990.
- [8] M. Hoffman, Multiple harmonic series, *Pacific Journal of Mathematics* 152, 275–290, 1992.
- [9] S. Janson, Random cutting and records in deterministic and random trees, *Random Structures and Algorithms*, to appear, 2004.
- [10] D. Knuth and A. Schönhage, The expected linearity of a simple equivalence algorithm, *Theoretical Computer Science* 6, 281–315, 1978.
- [11] H. Mahmoud, *Evolution of random search trees*, Wiley, New York, 1992.
- [12] H. Mahmoud and R. Smythe, A Survey of Recursive Trees, *Theoretical Probability and Mathematical Statistics* 51, 1–37, 1995.
- [13] A. Meir and J. W. Moon, Cutting down random trees, *Journal of the Australian Mathematical Society* 11, 313–324, 1970.
- [14] A. Meir and J. W. Moon, Cutting down recursive trees, *Mathematical Biosciences* 21, 173–181, 1974.
- [15] A. Meir and J. W. Moon, On the altitude of nodes in random trees, *Canadian Journal of Mathematics* 30, 997–1015, 1978.
- [16] A. Panholzer, Cutting down very simple trees, submitted, 2003.
- [17] A. Panholzer, Non-crossing trees revisited: cutting down and spanning subtrees, in *Discrete Random Walks*, C. Banderier and C. Krattenthaler (eds.), *Discrete Mathematics and Theoretical Computer Science*, Proceedings AC, 265–276, 2003.
- [18] A. Panholzer, Destruction of Recursive Trees, *Proceedings of the "Third Colloquium on Mathematics and Computer Science"*, M. Drmota et al (eds.), Birkhäuser, Basel, 267–280, 2004.
- [19] J. Vitter and P. Flajolet, *Average Case Analysis of Algorithms and Data Structures*, in *Handbook of Theoretical Computer Science*, 431–524, Elsevier, Amsterdam, 1990.

MARKUS KUBA, INSTITUT FÜR DISKRETE MATHEMATIK UND GEOMETRIE, TECHNISCHE UNIVERSITÄT WIEN, WIEDNER HAUPTSTR. 8-10/104, 1040 WIEN, AUSTRIA

E-mail address: Markus.Kuba@tuwien.ac.at

ALOIS PANHOLZER, INSTITUT FÜR DISKRETE MATHEMATIK UND GEOMETRIE, TECHNISCHE UNIVERSITÄT WIEN, WIEDNER HAUPTSTR. 8-10/104, 1040 WIEN, AUSTRIA

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