

# Analysis of label-based parameters in increasing trees

Markus Kuba<sup>1†</sup> and Alois Panholzer<sup>1</sup>

<sup>1</sup> Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstr. 8-10/104, A-1040 Wien, Austria. {Markus.Kuba,Alois.Panholzer}@tuwien.ac.at

---

Simple families of increasing trees can be constructed from simply generated tree families, if one considers for every tree of size  $n$  all its increasing labellings, i.e., labellings of the nodes 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. Three such tree families are of particular interest: *recursive trees*, *plane-oriented recursive trees* and *binary increasing trees*.

Here we present a general approach for the analysis of a number of label-based parameters in a random increasing tree of size  $n$  as, e.g., *the degree of node  $j$* , *the subtree-size of node  $j$* , etc. Further we apply the approach to the random variable  $X_{n,j,a}$ , which counts the number of size- $a$  branches (= subtrees) attached to node  $j$  in a random simple increasing tree of size  $n$ . We can give closed formulæ for the probability distribution and the factorial moments for those subclass of tree families, which can be constructed via an insertion process. Furthermore limiting distribution results for  $X_{n,j,a}$  are given dependent on the growth behaviour of  $j = j(n)$  compared to  $n$ .

**Keywords:** increasing trees, label-based parameters, limiting distribution

---

## 1 Introduction

### 1.1 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 so called *simply generated trees* (see Meir and Moon (1978)) but, additionally, the trees are equipped with increasing labellings. Thus we are considering *simple families of increasing trees*, which are introduced in Bergeron et al. (1992).

Several important tree families, in particular *recursive trees*, *plane-oriented recursive trees* (also called heap ordered trees or non-uniform recursive trees) and *binary increasing trees* (also called tournament trees) are special instances of simple families of increasing trees. A survey of applications and results on recursive trees and plane-oriented recursive trees is given in Mahmoud and Smythe (1995). These models are used, e.g., to describe the spread of epidemics, for pyramid schemes, and quite recently as a simplified growth model of the world wide web.

---

<sup>†</sup>This work is supported by the Austrian Science Foundation FWF, grant S9608-N13.

In applications the subclass of simple families of increasing trees, which can be constructed via an *insertion process* or a *probabilistic growth rule*, is of particular interest. Such tree families  $\mathcal{T}$  have the property that for every tree  $T'$  of size  $n$  with vertices  $v_1, \dots, v_n$  there exist probabilities  $p_{T'}(v_1), \dots, p_{T'}(v_n)$ , such that when starting with a random tree  $T'$  of size  $n$ , choosing a vertex  $v_i$  in  $T'$  according to the probabilities  $p_{T'}(v_i)$  and attaching node  $n+1$  to it, we obtain a *random increasing tree*  $T$  of the family  $\mathcal{T}$  of size  $n+1$ . It is well known that the tree families mentioned above, i.e., recursive trees, plane-oriented recursive trees and binary increasing trees, can be constructed via an insertion process. In Panholzer and Prodinger (2005) a full characterization of those simple families of increasing trees, which can be constructed by an insertion process, is given. This subclass of increasing tree families will be denoted by *grown simple families of increasing trees* and its characterization via the so called degree-weight generating function is restated here as Lemma 1.

## 1.2 Problem description

In this work we present a unifying approach for studying a class of label-based parameters for increasing trees. This approach has already been carried out successfully for two special instances, namely for *the subtree-size (= number of descendants) of node  $j$*  and *the degree of node  $j$* , see Kuba and Panholzer (2006a) and Kuba and Panholzer (2006b).

Let us denote by  $X_{n,j}$  a random variable that counts a certain label-based parameter for a specified node  $j$ , with  $1 \leq j \leq n$ , in a random size- $n$  tree, which is assumed to depend only on the subtree rooted at node  $j$ . Important instances of such parameters are the number of descendants of node  $j$  and the degree of node  $j$  as already mentioned. We use a recursive approach, which leads for all simple families of increasing trees (not only those, which can be described via an insertion process) to a closed formula for suitable trivariate generating functions of the probabilities  $\mathbb{P}\{X_{n,j} = m\}$ . We obtain explicit formulæ for the probabilities  $\mathbb{P}\{X_{n,j} = m\}$  and the  $s$ -th factorial moments  $\mathbb{E}((X_{n,j})^{\underline{s}}) = \sum_{m \geq 0} m^{\underline{s}} \mathbb{P}\{X_{n,j} = m\}$ . Such a detailed description turns out to be useful to give a full characterization of the limiting distribution behaviour of  $X_{n,j}$ , dependent on the growth of  $j = j(n)$  compared to  $n$ .

We illustrate this approach for the following parameter analyzing the *branching structure* of increasing trees. Let  $X_{n,j,a}$  denote the random variable, which counts the number of size- $a$  branches (= subtrees) attached to node  $j$  in a random grown simple increasing tree of size  $n$ . This parameter was studied in Su et al. (2006) for the case of the root node  $j = 1$  and for the instance of random recursive trees: they derived the distribution of  $X_{n,1,a}$  and a limit law for it. Further they stated the joint distribution of  $X_{n,1,1}, X_{n,1,2}, \dots, X_{n,1,n-1}$ .

By using the approach presented we can extend their results by treating arbitrary grown simple families of increasing trees and considering the node with label  $j$ , not only the root node. We can give closed formulæ for the probability distribution and the factorial moments of  $X_{n,j,a}$ . Furthermore limiting distribution results of  $X_{n,j,a}$  are given, not only for  $j$  fixed, but a full characterization dependent on the growth  $j = j(n)$  compared to  $n$  is presented. Moreover, the joint distribution of  $X_{n,j,1}, X_{n,j,2}, \dots, X_{n,j,n-j}$  is computed for all grown simple families of increasing trees.

## 1.3 Plan of the paper

In Section 2 we will collect some auxiliary results concerning simple families of increasing trees together with the characterization of grown simple families of increasing trees. In Section 3 we state the results for general label-based parameters and for the branching structure of increasing trees. A proof of these results is sketched in Section 4; for a detailed analysis and a lot of further results we refer to Kuba (2006).

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  $x = 1$ . Further we denote with  $\begin{bmatrix} n \\ k \end{bmatrix}$  and  $\begin{Bmatrix} n \\ k \end{Bmatrix}$  the Stirling numbers of the first and second kind, respectively, with  $X \stackrel{(d)}{=} Y$  the equality in distribution of the random variables  $X$  and  $Y$ , and with  $X_n \xrightarrow{(d)} X$  the weak convergence, i.e., the convergence in distribution, of the sequence of random variables  $X_n$  to a random variable  $X$ .

## 2 Preliminaries

Formally, a class  $\mathcal{T}$  of a simple family of increasing trees can be defined 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  $v$  (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 the tree  $T$ , and  $L(T) := |\mathcal{L}(T)|$  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 first order differential equation

$$T'(z) = \varphi(T(z)), \quad T(0) = 0. \quad (1)$$

Often it is advantageous to describe a simple family of increasing trees  $\mathcal{T}$  by the formal equation

$$\mathcal{T} = \textcircled{1} \times \left( \varphi_0 \cdot \{\epsilon\} \dot{\cup} \varphi_1 \cdot \mathcal{T} \dot{\cup} \varphi_2 \cdot \mathcal{T} * \mathcal{T} \dot{\cup} \varphi_3 \cdot \mathcal{T} * \mathcal{T} * \mathcal{T} \dot{\cup} \dots \right) = \textcircled{1} \times \varphi(\mathcal{T}), \quad (2)$$

where  $\textcircled{1}$  denotes the node labelled by 1,  $\times$  the cartesian product,  $*$  the partition product for labelled objects, and  $\varphi(\mathcal{T})$  the substituted structure (see, e.g., Vitter and Flajolet (1990)).

We will now characterize grown simple increasing tree families via the degree-weight generating function  $\varphi(t)$  as obtained in Panholzer and Prodinger (2005).

**Lemma 1 (Panholzer and Prodinger (2005))** *A simple family of increasing trees  $\mathcal{T}$  can be constructed via an insertion process and is thus a grown simple family of increasing trees iff the degree-weight generating function  $\varphi(t) = \sum_{k \geq 0} \varphi_k t^k$  is given by one of the following three formulae (with constants  $c_1, c_2 \in \mathbb{R}$ ).*

**Case A (recursive trees) :**

$$\varphi(t) = \varphi_0 e^{\frac{c_1 t}{\varphi_0}}, \quad \text{for } \varphi_0 > 0, c_1 > 0, \quad (\text{defining } c_2 := 0),$$

**Case B (d-ary trees) :**

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

**Case C (generalized plane-oriented recursive trees) :**

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

Solving the differential equation (1) one obtains the following explicit formulæ for the exponential generating function  $T(z)$ :

$$T(z) = \begin{cases} \frac{\varphi_0}{c_1} \log\left(\frac{1}{1-c_1 z}\right), & \text{Case A,} \\ \frac{\varphi_0}{c_2} \left( \frac{1}{(1-(d-1)c_2 z)^{\frac{1}{d-1}}} - 1 \right), & \text{Case B,} \\ \frac{\varphi_0}{c_2} \left( \frac{1}{(1-c_1 z)^{\frac{c_2}{c_1}}} - 1 \right), & \text{Case C.} \end{cases} \quad (3)$$

Furthermore the coefficients  $T_n$  are given by the following formula, which holds for all three cases of very simple increasing tree families (setting  $c_2 = 0$  in Case A and  $d = \frac{c_1}{c_2} + 1$  in Case B):

$$T_n = \varphi_0 c_1^{n-1} (n-1)! \binom{n-1 + \frac{c_2}{c_1}}{n-1}. \quad (4)$$

Next we are going to describe in more detail the tree evolution process, which generates random trees (of arbitrary size  $n$ ) of grown simple families of increasing trees. This description is a consequence of the considerations made in Panholzer and Prodinger (2005):

- Step 1: The process starts with the root labelled by 1.
- Step  $i+1$ : At step  $i+1$  the node with label  $i+1$  is attached to any previous node  $v$  (with out-degree  $d(v)$ ) of the already grown tree of size  $i$  with probabilities  $p(v)$  proportional to  $\frac{(d(v)+1)\varphi_{d(v)+1}}{\varphi_{d(v)}}$ , i.e.,

$$p(v) = \begin{cases} \frac{1}{i}, & \text{for Case A,} \\ \frac{d-d(v)}{(d-1)i+1}, & \text{for Case B,} \\ \frac{d(v)+\alpha}{(\alpha+1)i-1}, & \text{with } \alpha := -1 - \frac{c_1}{c_2} > 0, \text{ for Case C.} \end{cases}$$

In order to state our results concerning arbitrary label-based parameters in Section 3 we introduce two generating functions. For encoding the behaviour of the root node  $j = 1$  we introduce the bivariate generating function

$$M(z, v) = \sum_{n \geq 1} \sum_{m \geq 1} \mathbb{P}\{X_{n,1} = m\} T_n \frac{z^n}{n!} v^m, \quad (5)$$

whereas to describe the behaviour of arbitrary nodes  $j \geq 1$  we introduce the trivariate generating function

$$N(z, u, v) := \sum_{k \geq 0} \sum_{j \geq 1} \sum_{m \geq 0} \mathbb{P}\{X_{k+j,j} = m\} T_{k+j} \frac{z^{j-1}}{(j-1)!} \frac{u^k}{k!} v^m. \quad (6)$$

## 3 Results

### 3.1 General results for label-based parameters of increasing trees

**Proposition 1** *The function  $N(z, u, v)$  as defined in equation (6), which is the trivariate generating function of the probabilities  $\mathbb{P}\{X_{n,j} = m\}$ , which give the probability that a certain label-based parameter*

for the node labelled  $j$ , which depends only on the subtree rooted at node  $j$ , in a randomly chosen size- $n$  tree of a simple family of increasing trees with degree-weight generating function  $\varphi(t)$  equals  $m$ , is given by the following formula, where  $M(z, v)$  is defined by (5):

$$N(z, u, v) = \frac{\varphi(T(z+u)) \frac{\partial}{\partial u} M(u, v)}{\varphi(T(u))}. \quad (7)$$

**Theorem 1** For grown simple families of increasing trees the probability distribution and the factorial moments of  $X_{n,j}$ , which counts a label-based parameter in a random size- $n$  tree that depends only on the subtree rooted at node  $j$ , are given as follows.

$$\mathbb{P}\{X_{n,j} = m\} = \frac{c_1^{j-1} \binom{c_2+j-1}{j-1}}{\binom{n-1}{n-j} \varphi_0 c_1^{n-1} \binom{n-1+c_2}{n-1}} [u^{n-j} v^m] \frac{\frac{\partial}{\partial u} M(u, v)}{(1-c_1 u)^{j-1}}, \quad (8)$$

$$\mathbb{E}(X_{n,j}^s) = \frac{c_1^{j-1} \binom{c_2+j-1}{j-1}}{\binom{n-1}{n-j} \varphi_0 c_1^{n-1} \binom{n-1+c_2}{n-1}} [u^{n-j}] \frac{M_s'(u)}{(1-c_1 u)^{j-1}}, \quad (9)$$

where we use the abbreviation  $M_s(z) := E_v D_v^s M(z, v)$ . For Case A we have to set  $c_2 = 0$  and for Case B:  $d = \frac{c_1}{c_2} + 1$ .

**Remark 1** Slight modifications of the approach presented one can analyze also several label-based parameters in increasing trees, which are dependent on the whole tree, not only on the subtree rooted at  $j$ . In particular, for the depth  $D_{n,j}$  of node  $j$  in a random size- $n$  increasing tree, the trivariate generating function  $N(z, u, v)$  has only a slightly different form, which also allows to gain explicit and limiting distribution results as is figured out in Panholzer and Prodinger (2005).

## 3.2 Results for the branching structure of increasing trees

### Explicit formulæ for the probabilities

**Theorem 2** The probability that there are  $m$  size- $a$  branches attached to node  $j$  is given as follows.

- Case A (recursive trees): we obtain the exact formula

$$\mathbb{P}\{X_{n,j,a} = m\} = \frac{1}{a^m a! \binom{n-1}{j-1}} \sum_{l=0}^{\lfloor \frac{n-j-am}{a} \rfloor} \frac{(-1)^l}{a^l l!} \binom{n-1-a(m+l)}{j-1}.$$

- Case B ( $d$ -ary increasing trees): we obtain the exact formula

$$\begin{aligned} & \mathbb{P}\{X_{n,j,a} = m\} \\ &= \frac{\binom{d}{m} \binom{\frac{d}{d-1}+j-2}{j-1}}{\binom{n-1}{j-1} \binom{n-1+\frac{1}{d-1}}{n-1}} \sum_{i=0}^{\min\{d-m, \lfloor \frac{n-j-am}{m} \rfloor\}} (-1)^i \binom{d-m}{i} \left( \frac{\binom{a-1+\frac{1}{d-1}}{a-1}}{a(d-1)} \right)^{i+m} \binom{n-2-a(i+m) + \frac{d-m-i}{d-1}}{n-j-a(i+m)}. \end{aligned}$$

- Case C (Generalized plane oriented recursive trees): we obtain

$$\mathbb{P}\{X_{n,j,a} = m\} = \frac{\varphi_0(j-1)!(n-j)!c_1^{j-1} \binom{c_2+j-1}{c_1} \binom{c_1+1}{c_2} \left(\frac{c_2 T_a}{\varphi_0 a!}\right)^m}{T_n} \times \\ \times [u^{n-j-am}] \frac{1}{(1-c_1 u)^{j-1} (C(u))^{m-\frac{c_1}{c_2}-1}}, \quad \text{with } C(u) := \frac{1}{(1-c_1 u)^{\frac{c_2}{c_1}}} - \frac{c_2 T_a}{\varphi_0 a!} u^a.$$

### Explicit formulæ for the factorial moments

**Theorem 3** The factorial moments of the random variable  $X_{n,j,a}$ , which counts the number of size- $a$  subtrees attached to node  $j$  in a random size- $n$  grown simple increasing tree are given as follows.

- Case A (recursive trees):

$$\mathbb{E}(X_{n,j,a}^s) = \frac{1}{a^s} \frac{\binom{n-as-1}{j-1}}{\binom{n-1}{j-1}}.$$

- Case B ( $d$ -ary increasing trees):

$$\mathbb{E}(X_{n,j,a}^s) = \left(\frac{\binom{a-1+\frac{1}{d-1}}{a-1}}{a(d-1)}\right)^s \frac{\binom{j-1+\frac{1}{d-1}}{j-1} \binom{n-as-2+\frac{d-s}{d-1}}{n-j-as}}{\binom{n-1}{j-1} \binom{n-1+\frac{1}{d-1}}{n-1}}.$$

- Case C (Generalized plane oriented recursive trees):

$$\mathbb{E}(X_{n,j,a}^s) = \left(\frac{-c_2 \binom{a-1+\frac{c_2}{c_1}}{a-1}}{c_1 a}\right)^s \frac{\Gamma(s-1-\frac{c_1}{c_2}) \binom{j-1+\frac{c_2}{c_1}}{j-1} \binom{n-as-1-\frac{c_2}{c_1}(s-1)}{n-j-as}}{\Gamma(-1-\frac{c_1}{c_2}) \binom{n-1}{j-1} \binom{n-1+\frac{c_2}{c_1}}{n-1}}.$$

### Limiting distribution results

**Theorem 4** For Case A (recursive trees) we obtain the following results.

- for  $n \rightarrow \infty$ ,  $j = o(n)$  and  $a$  fixed:  $X_{n,j,a}$  is asympt. Poisson distributed with parameter  $1/a$ :

$$X_{n,j,a} \xrightarrow{(d)} X_a, \quad \mathbb{P}\{X_a = m\} = \frac{e^{-\frac{1}{a}}}{a^m m!}.$$

- for  $n \rightarrow \infty$ ,  $j = \rho n$  with  $0 < \rho < 1$  and  $a$  fixed:  $X_{n,j,a}$  is asympt. Poisson distributed with parameter  $(1-\rho)/a$ :

$$X_{n,j,a} \xrightarrow{(d)} X_{\rho,a}, \quad \mathbb{P}\{X_{\rho,a} = m\} = \frac{e^{-\frac{1-\rho}{a}} (1-\rho)^m}{a^m m!}.$$

- for  $n \rightarrow \infty$ ,  $l = n - j = o(n)$  and  $a$  fixed:  $X_{n,j,a}$  is asympt. zero:  $X_{n,j,a} \xrightarrow{(d)} X$ ,  $\mathbb{P}\{X = 0\} = 1$ .

**Theorem 5** For Case B ( $d$ -ary increasing trees) we obtain the following results.

- for  $n \rightarrow \infty$ ,  $j = o(n)$  and  $a$  fixed:  $X_{n,j,a}$  is asympt. zero:  $X_{n,j,a} \xrightarrow{(d)} X$ ,  $\mathbb{P}\{X = 0\} = 1$ .
- for  $n \rightarrow \infty$ ,  $j = \rho n$  with  $0 < \rho < 1$  and  $a$  fixed:  $X_{n,j,a}$  is asympt. binomial distributed  $\text{Bin}(N, p)$  with parameter  $N = d$  and  $p = p(a, d) = \binom{a-1+\frac{1}{d-1}}{a-1} \frac{\rho(1-\rho)^a}{a(d-1)}$ :

$$X_{n,j,a} \xrightarrow{(d)} X_{\rho,a}, \quad \mathbb{P}\{X_{\rho,a} = m\} = \binom{d}{m} p^m (1-p)^{d-m}.$$

- for  $n \rightarrow \infty$ ,  $l = n - j = o(n)$  and  $a$  fixed:  $X_{n,j,a}$  is asympt. zero:  $X_{n,j,a} \xrightarrow{(d)} X$ ,  $\mathbb{P}\{X = 0\} = 1$ .

**Theorem 6** For Case C (Generalized plane oriented recursive trees) we obtain the following results.

- for  $n \rightarrow \infty$ ,  $j$  and  $a$  fixed: the limiting distribution of  $X_{n,j,a}$  is characterized by its moments:

$$n^{\frac{c_2}{c_1}} X_{n,j} \xrightarrow{d} X, \quad \mathbb{E}(X^s) = \frac{\Gamma(j + \frac{c_2}{c_1}) \Gamma(s - 1 - \frac{c_1}{c_2})}{\Gamma(j - \frac{c_2}{c_1}(s-1)) \Gamma(-1 - \frac{c_1}{c_2})} \left( \frac{(-c_2) \binom{a-1+\frac{c_2}{c_1}}{a-1}}{c_1 a} \right)^s.$$

- for  $n \rightarrow \infty$ ,  $j = o(n)$  and  $a$  fixed:  $X_{n,j,a}$  is asympt. gamma distributed  $\gamma(k, \lambda)$  with parameters  $k = -\frac{c_1}{c_2} - 1$  and  $\lambda = \frac{ac_1}{(-c_2) \binom{a-1+\frac{c_2}{c_1}}{a-1}}$ :

$$\left(\frac{n}{j}\right)^{\frac{c_2}{c_1}} X_{n,j} \xrightarrow{d} X, \quad X \stackrel{d}{=} \gamma(k, \lambda), \quad \mathbb{E}(X^s) = \left(-\frac{c_1}{c_2} - 1\right)^{\bar{s}} \left(\frac{(-c_2) \binom{a-1+\frac{c_2}{c_1}}{a-1}}{c_1 a}\right)^s.$$

- for  $n \rightarrow \infty$ ,  $j = \rho n$  with  $0 < \rho < 1$  and  $a$  fixed:  $X_{n,j,a}$  is asympt. negative binomial distributed  $\text{NegBin}(r, p)$  with parameters  $r = -1 - \frac{c_1}{c_2}$  and  $p = p(\rho, c_1, c_2, a) = (1 + (1 - \rho)^a \rho^{\frac{c_2}{c_1}} \binom{a-1+\frac{c_2}{c_1}}{a-1} \frac{(-c_2)}{c_1 a})^{-1}$ :

$$X_{n,j,a} \xrightarrow{(d)} X_{\rho,a}, \quad \text{with} \quad \mathbb{P}\{X_{\rho,a} = m\} = \binom{m-2-\frac{c_1}{c_2}}{m} p^{1+\frac{c_2}{c_1}} (1-p)^m.$$

- for  $n \rightarrow \infty$ ,  $l = n - j = o(n)$  and  $a$  fixed:  $X_{n,j,a}$  is asympt. zero:  $X_{n,j,a} \xrightarrow{(d)} X$ ,  $\mathbb{P}\{X = 0\} = 1$ .

### Joint distribution results

**Theorem 7** The joint distribution of  $X_{n,1,1} \dots X_{n,1,n-1}$  is for all Cases A, B and C given as follows.

$$\mathbb{P}\{X_{n,1,1} = m_1, \dots, X_{n,1,n-1} = m_{n-1}\} = \frac{\varphi_{\sum_{i=1}^{n-1} m_i} \varphi_0^{\sum_{i=1}^{n-1} m_i - 1} (\sum_{i=1}^{n-1} m_i)!}{\binom{n-1+\frac{c_2}{c_1}}{n-1}} \prod_{k=1}^{n-1} \frac{\binom{k-1+\frac{c_2}{c_1}}{k-1}}{k^{m_k} m_k!},$$

for all sequences of non-negative integers satisfying  $\sum_{k=1}^{n-1} k m_k = n - 1$ , with  $c_2 = 0$  for Case A and  $\frac{c_2}{c_1} = \frac{1}{d-1}$  for Case B. For Case B we also have the additional constraint  $\sum_{k=1}^{n-1} m_k \leq d$ .

**Corollary 1** *The joint distribution of  $X_{n,j,1} \dots X_{n,j,n-j}$  is for all Cases A, B and C given as follows, where  $Z_{n,j}$  denotes the random variable counting the subtree-size of node  $j$ .*

$$\begin{aligned} & \mathbb{P}\{X_{n,j,1} = m_1, \dots, X_{n,j,n-j} = m_{n-j}\} \\ &= \mathbb{P}\{Z_{n,j} = 1 + \sum_{i=1}^{n-j} im_i\} \mathbb{P}\{X_{1+\sum_{i=1}^{n-j} im_i,1,1} = m_1, \dots, X_{1+\sum_{i=1}^{n-j} im_i,1,n-j} = m_{n-j}\}. \end{aligned}$$

### Results for a randomly chosen node

We denote by  $X_{n,a}^{[R]} = X_{n,U_n,a}$  the random random variable, which counts the number of size- $a$  branches of a randomly chosen node in a random size- $n$  grown simple increasing tree.

**Theorem 8** *The limit law of the random variable  $X_{n,a}^{[R]}$  is given as follows.*

- *Case A (recursive trees): for a fixed  $a$  it holds*

$$X_{n,a}^{[R]} \xrightarrow{(d)} X_a, \quad \mathbb{P}\{X_a = m\} = a \left(1 - e^{-\frac{1}{a}} \sum_{k=0}^m \frac{1}{k! a^k}\right) = \sum_{k \geq m+1} \frac{ae^{-\frac{1}{a}}}{k! a^k}.$$

- *Case B ( $d$ -ary increasing trees): for a fixed  $a$  it holds*

$$X_{n,a}^{[R]} \xrightarrow{(d)} X_a, \quad \mathbb{P}\{X_a = m\} = \sum_{k=0}^{d-m} (-1)^k \frac{\binom{d}{m} \binom{d-m}{k} \left(\frac{a-1+\frac{1}{a-1}}{a-1}\right)^{m+k}}{a^{m+k} (d-1)^{m+k} ((m+k)(d+1)+1) \binom{(d+1)(m+k)}{m+k}}.$$

**Remark 2** *For Case C we were not able to obtain a closed formula for the probability  $\mathbb{P}\{X_a = m\}$ , but of course one has the explicit formula  $\mathbb{P}\{X_a = m\} = \int_0^1 \mathbb{P}\{X_{\rho,a} = m\} d\rho$ .*

## 4 Sketch of the proof of the results presented

### 4.1 General label-based parameters

We consider in this section the random variable  $X_{n,j}$  that counts a certain label-based parameter for a specified node  $j$ , with  $1 \leq j \leq n$ , in a random size- $n$  tree, which is assumed to depend only on the subtree rooted at node  $j$ . In the following we give a general recurrence for the probability  $\mathbb{P}\{X_{n,j} = m\}$ , using the formal description of increasing trees (2).

For increasing trees of size  $n$  with root-degree  $r$  and subtrees with sizes  $k_1, \dots, k_r$ , enumerated from left to right, where the node labelled by  $j$  lies in the leftmost subtree and is the  $i$ -th smallest node in this subtree, we can reduce the computation of the probabilities  $\mathbb{P}\{X_{n,j} = m\}$  to the probabilities  $\mathbb{P}\{X_{k_1,i} = m\}$ , when the label-based parameter does only depend on the subtree of node  $j$ .

We get as factor the total weight of the  $r$  subtrees and the root node  $\varphi_r T_{k_1} \dots T_{k_r}$ , divided by the total weight  $T_n$  of trees of size  $n$  and multiplied by the number of order preserving relabellings of the  $r$  subtrees, which are given here by

$$\binom{j-2}{i-1} \binom{n-j}{k_1-i} \binom{n-1-k_1}{k_2, k_3, \dots, k_r} :$$



the  $i - 1$  labels smaller than  $j$  are chosen from  $2, 3, \dots, j - 1$ , the  $k_1 - i$  labels larger than  $j$  are chosen from  $j + 1, \dots, n$ , and the remaining  $n - 1 - k_1$  labels are distributed to the second, third,  $\dots$ ,  $r$ -th subtree. Due to symmetry arguments we obtain a factor  $r$ , if the node  $j$  is the  $i$ -th smallest node in the second, third,  $\dots$ ,  $r$ -th subtree. Summing up over all choices for the rank  $i$  of label  $j$  in its subtree, the subtree sizes  $k_1, \dots, k_r$ , and the degree  $r$  of the root node gives for  $n \geq j \geq 2$  the following recurrence.

$$\begin{aligned} \mathbb{P}\{X_{n,j} = m\} &= \sum_{r \geq 1} r \varphi_r \sum_{\substack{k_1 + \dots + k_r = n - 1, \\ k_1, \dots, k_r \geq 1}} \frac{T_{k_1} \dots T_{k_r}}{T_n} \times \\ &\times \sum_{i=1}^{\min\{k_1, j-1\}} \mathbb{P}\{X_{k_1, i} = m\} \binom{j-2}{i-1} \binom{n-j}{k_1-i} \binom{n-1-k_1}{k_2, k_3, \dots, k_r}. \end{aligned} \quad (10)$$

To treat recurrence (10) we set  $n := k + j$ , multiply the equation with  $T_{k+j} \frac{z^{j-2}}{(j-2)!} \frac{u^k}{k!} v^m$  and sum up over  $k \geq 0$ ,  $j \geq 2$  and  $m \geq 0$ . Using the trivariate generating  $N(z, u, v)$  as introduced by (6) we obtain then  $\frac{\partial}{\partial z} N(z, u, v)$  and  $\varphi'(T(z+u))N(z, u, v)$  for the left and right hand side of (10), respectively. These computations are here omitted; similar ones are carried out in Panholzer and Prodinger (2005).

We obtain thus the following differential equation

$$\frac{\partial}{\partial z} N(z, u, v) = \varphi'(T(z+u))N(z, u, v), \quad (11)$$

together with the initial condition

$$N(0, u, v) = \sum_{k \geq 0} \sum_{m \geq 0} \mathbb{P}\{X_{k+1,1} = m\} T_{k+1} \frac{u^k}{k!} v^m = M'(u, v), \quad (12)$$

where we use the bivariate generating function  $M(z, v)$  as introduced by (5). It is an easy task to solve this differential equation, where some simplifications occur due to  $T'(z) = \varphi(T(z))$ . This leads to Proposition 1. Using Lemma 1 and equation (3) we obtain by extracting coefficients Theorem 1.

**Remark 3** We mention that one could also give a more ‘‘combinatorial approach’’ of equation (11) similar to the one, which was presented for the depth of nodes in Panholzer and Prodinger (2005).

## 4.2 The branching structure of increasing trees

### The generating function for the root node

Let  $X_{n,j,a}$  denote the random variable, which counts the number of subtrees of size  $a$  attached to node  $j$  in a random size- $n$  increasing tree. In order to analyze  $X_{n,j,a}$  by the approach presented here and thus to apply Proposition 1 one has to compute at first the generating function  $\frac{\partial}{\partial z} M(z, v)$  as defined in (5).

By definition one gets the following explicit result for the probabilities  $\mathbb{P}\{X_{n,1,a} = m\}$ :

$$\mathbb{P}\{X_{n,1,a} = m\} = \sum_{r \geq m} \varphi_r \binom{r}{m} \sum_{\substack{k_1 + \dots + k_r = n - 1 \\ k_i = a \text{ for } 1 \leq i \leq m \\ k_j \neq a \text{ for } m + 1 \leq j \leq r}} \frac{T_{k_1} \dots T_{k_r}}{T_n} \binom{n-1}{k_1, \dots, k_r}, \quad (13)$$

for  $n \geq 1$  and  $m \geq 0$ . By multiplying with  $T_n z^{n-1} v^m / (n-1)!$  and summing up over  $n \geq 1, m \geq 0$  this yields to an explicit formula for  $\frac{\partial}{\partial z} M(z, v)$ , which is given below.

$$\frac{\partial}{\partial z} M(z, v) = \varphi(T(z) + T_a z^a (v-1)). \quad (14)$$

This leads for grown simple families of increasing trees to the following formulæ.

$$\frac{\partial}{\partial z} M(z, v) = \begin{cases} \frac{\varphi_0}{1-c_1 z} \exp\left(\frac{c_1 T_a}{\varphi_0 a!} z^a (v-1)\right), & \text{Case A,} \\ \varphi_0 \left( \frac{c_2 T_a}{\varphi_0 a!} z^a (v-1) + \frac{1}{(1-(d-1)c_2 z)^{\frac{1}{d-1}}} \right)^d, & \text{Case B,} \\ \frac{\varphi_0}{\left( \frac{1}{(1-c_1 z)^{\frac{c_2}{c_1}}} + \frac{c_2 T_a}{\varphi_0 a!} z^a (v-1) \right)^{-\frac{c_1}{c_2}-1}}, & \text{Case C} \end{cases} \quad (15)$$

### Explicit formulæ for the probabilities

Applying Theorem 1 and using (15) leads for all three cases to the results stated. We will here only present exemplarily the computations for Case A, where we get

$$\begin{aligned} \mathbb{P}\{X_{n,j,a} = m\} &= \frac{(j-1)!(n-j)!}{T_n} [u^{n-j} v^m] \frac{c_1^{j-1}}{(1-c_1 u)^{j-1}} \frac{\partial}{\partial u} M(u, v) \\ &= \frac{1}{\binom{n-1}{j-1} c_1^{n-j}} [u^{n-j} v^m] \frac{\exp\left(-\frac{c_1 T_a}{\varphi_0 a!} u^a\right)}{(1-c_1 u)^j} \exp\left(v \frac{c_1 T_a}{\varphi_0 a!} u^a\right) \\ &= \frac{c_1^{am}}{a^m m! \binom{n-1}{j-1} c_1^{n-j}} [u^{n-j-ma}] \frac{\exp\left(-\frac{c_1}{a} u^a\right)}{(1-c_1 u)^j} = \frac{1}{a^m m! \binom{n-1}{j-1}} \sum_{l=0}^{\lfloor \frac{n-j-am}{a} \rfloor} \frac{(-1)^l}{a^l l!} \binom{n-1-a(m+l)}{j-1}. \end{aligned} \quad (16)$$

### Explicit formulæ for the factorial moments

Again by applying the corresponding part of Theorem 1 one obtains by using (15) for all three cases explicit formulæ for the factorial moments. We derive exemplarily  $M'_s(u)$  for Case A, where we get

$$M'_s(u) = E_v D_v^s \frac{\partial}{\partial u} M(u, v) = E_v D_v^s \frac{\varphi_0}{1-c_1 u} \exp\left(\frac{c_1 T_a}{\varphi_0 a!} u^a (v-1)\right) = \frac{\varphi_0 \left(\frac{u^a c_1 T_a}{\varphi_0 a!}\right)^s}{1-c_1 u} = \frac{\varphi_0 u^{as} \left(\frac{c_1}{a}\right)^s}{1-c_1 u}. \quad (17)$$

### Limiting distribution results

From the explicit results obtained for the probabilities and the factorial moments as stated in Theorem 2 and Theorem 3 one can for all three cases of grown simple families of increasing trees characterize the limiting distribution of  $X_{n,j,a}$  dependent on the growth of  $j = j(n)$  compared to  $n$  (and  $a$  fixed). We restrict ourselves here to present the computations for Case C. To obtain limiting distribution results for that case we consider the factorial moments of  $X_{n,j,a}$ , which are after simplifications given as follows.

$$\mathbb{E}(X_{n,j,a}^s) = \frac{(-c_2)^s \left(\frac{a-1+\frac{c_2}{c_1}}{a-1}\right)^s}{c_1^s a^s} \frac{\Gamma\left(j + \frac{c_2}{c_1}\right) (n-j)! \Gamma\left(n - as - \frac{c_2}{c_1}(s-1)\right)}{\Gamma\left(j - \frac{c_2}{c_1}(s-1)\right) (n-j-as)! \Gamma\left(n + \frac{c_2}{c_1}\right)}. \quad (18)$$

For  $j$  (and  $a$ ) fixed an application of Stirling's formula to (18) leads to

$$\mathbb{E}(X_{n,j,a}^s) = \frac{\Gamma(j + \frac{c_2}{c_1})\Gamma(s - 1 - \frac{c_1}{c_2})}{\Gamma(j - \frac{c_2}{c_1}(s - 1))\Gamma(-1 - \frac{c_1}{c_2})} \left( \frac{(-c_2)^{\binom{a-1+\frac{c_2}{c_1}}{a-1}}}{c_1 a} \right)^s \left( n^{-\frac{c_2}{c_1}} \right)^s \left( 1 + \mathcal{O}\left(\frac{1}{n^{-\frac{c_2}{c_1}}}\right) \right). \quad (19)$$

Similarly one gets for  $j \rightarrow \infty$ , such that  $j = o(n)$  (and  $a$  fixed) that

$$\mathbb{E}(X_{n,j,a}^s) = \frac{\Gamma(s - 1 - \frac{c_1}{c_2})}{\Gamma(-1 - \frac{c_1}{c_2})} \left( \frac{(-c_2)^{\binom{a-1+\frac{c_2}{c_1}}{a-1}}}{c_1 a} \right)^s \left( \frac{n}{j} \right)^{-s \frac{c_2}{c_1}} \left( 1 + \mathcal{O}\left(\left(\frac{j}{n}\right)^{-\frac{c_2}{c_1}}\right) \right). \quad (20)$$

For  $j = \rho n$ ,  $0 < \rho < 1$ , one obtains the following expansion of the factorial moments

$$\mathbb{E}(X_{n,j,a}^s) = \frac{\Gamma(s - 1 - \frac{c_1}{c_2})}{\Gamma(-1 - \frac{c_1}{c_2})} \left( \frac{(-c_2)^{\binom{a-1+\frac{c_2}{c_1}}{a-1}}}{c_1 a} \right)^s \rho^{s \frac{c_2}{c_1}} (1 - \rho)^{as} \left( 1 + \mathcal{O}\left(\frac{1}{n}\right) \right), \quad (21)$$

which are asymptotically the factorial moments of a negative binomial distribution  $X_\rho \stackrel{(d)}{=} \text{NegBin}(r, p)$ , with parameters  $r = -1 - \frac{c_1}{c_2}$  and  $p = (1 + (1 - \rho)^a \rho^{\frac{c_2}{c_1}} \binom{a-1+\frac{c_2}{c_1}}{a-1} \frac{(-c_2)}{c_1 a})^{-1}$ :

$$\mathbb{E}(X_\rho^s) = \frac{\Gamma(r + s)}{\Gamma(r)} \left( \frac{1}{p} - 1 \right)^s. \quad (22)$$

### Joint distribution results

One observes that

$$\begin{aligned} T_n \mathbb{P}\{X_{n,1,1} = m_1, \dots, X_{n,1,n-1} = m_{n-1}\} \\ = \varphi_{\sum_{i=1}^{n-1} m_i} \left( \underbrace{1, \dots, 1}_{m_1}, \underbrace{2, \dots, 2}_{m_2}, \dots, m_{n-1} \right) \binom{n-1}{m_1, \dots, m_{n-1}} \prod_{k=1}^{n-1} T_k^{m_k}. \end{aligned} \quad (23)$$

The factor  $\varphi_{\sum_{i=1}^{n-1} m_i}$  corresponds to the root degree, the factor  $\binom{n-1}{m_1, \dots, m_{n-1}}$  to the choices for the labels and the factor  $\left( \underbrace{1, \dots, 1}_{m_1}, \underbrace{2, \dots, 2}_{m_2}, \dots, m_{n-1} \right)$  to the different positions of the subtrees.

### Results for a randomly chosen node

We use the limiting distribution results for the central region of  $X_{n,j,a}$ , i.e.,  $j = \rho n$ , with  $0 < \rho < 1$ , to derive results for the number of size- $a$  branches  $X_{n,a}^{[R]}$  attached to a randomly chosen node in a grown simple increasing tree. Since  $X_{n,j,a} \stackrel{(d)}{\rightarrow} X_{\rho,a}$  we obtain  $X_{n,a}^{[R]} \stackrel{(d)}{\rightarrow} X_a$ , where the probabilities  $\mathbb{P}\{X_a = m\}$  of the discrete random variable  $X_a$  can be obtained via  $\mathbb{P}\{X_a = m\} = \int_0^1 \mathbb{P}\{X_{\rho,a} = m\} d\rho$ . For Case A and Case B we obtain closed formulæ for these integrals. We present only the computations for Case A.

$$\mathbb{P}\{X = m\} = \int_0^1 \frac{e^{-\frac{1-\rho}{a}} (1-\rho)^m}{m! a^m} d\rho = \frac{a}{m!} \int_0^{\frac{1}{a}} e^{-u} u^m du = a - \sum_{k=0}^m \frac{a e^{-\frac{1}{a}}}{a^{m-k} (m-k)!}, \quad \text{for } m \geq 0.$$

## References

- F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. *Lecture Notes in Computer Science*, 581:24–48, 1992.
- M. Kuba. *Analysis of node isolation procedures and label-based parameters in tree structures*. PhD thesis, TU Wien, Wien, 2006.
- M. Kuba and A. Panholzer. Descendants in increasing trees. *Electronic Journal of Combinatorics*, 13(1), 2006a. paper R8.
- M. Kuba and A. Panholzer. On the degree distribution of the nodes in increasing trees. manuscript, 2006b.
- M. Mahmoud and R. Smythe. A survey of recursive trees. *Theoretical Probability and Mathematical Statistics*, 51:1–37, 1995.
- A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, 30:997–1015, 1978.
- A. Panholzer and H. Prodinger. the level of nodes in increasing trees revisited. *Random Structures and Algorithms*, 2005. to appear.
- C. Su, Q. Feng, and Z. Hu. Uniform recursive trees: Branching structure and simple random downward walk. *Journal of Mathematical Analysis and Applications*, 315:225–243, 2006.
- J. Vitter and P. Flajolet. Average case analysis of algorithms and data structures. In *Handbook of Theoretical Computer Science*, pages 431–524. Elsevier, Amsterdam, 1990.