Descendants in increasing trees*

Markus Kuba
Institut für Diskrete Mathematik und Geometrie
Technische Universität Wien
Wiedner Hauptstr. 8-10/104
1040 Wien, Austria
E-mail: 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: Alois.Panholzer@tuwien.ac.at

Submitted: Jul 20, 2005; Accepted: Jan 19, 2006; Published: Jan 25, 2006
Mathematics Subject Classification: 05C05

Abstract

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 $\{1, \ldots, n\}$ 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*. We study the quantity *number of descendants of node j in a random tree of size n* and give closed formulæ for the probability distribution and all factorial moments for those subclass of tree families, which can be constructed via an insertion process. Furthermore limiting distribution results of this parameter are given.

1 Introduction

Increasing trees are labelled trees where the nodes of a tree of size n are labelled by distinct integers of the set $\{1, \ldots, n\}$ in such a way that each sequence of labels along

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

any branch starting at the root is increasing. As the underlying tree model we use the so called *simply generated trees* (see [7]) but, additionally, the trees are equipped with increasing labellings. Thus we are considering *simple families of increasing trees*, which are introduced in [1].

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 by Mahmoud and Smythe in [6]. 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.

In the present paper we are studying for simple families of increasing trees the random variable $D_{n,j}$, which counts the number of descendants of a specific node j (with $1 \le j \le n$), i e. the size of the subtree rooted at j (where size is measured as usual by the number of nodes), in a random size-n tree. Thus the node j is counted as a descendant of itself. We always use as the model of randomness the random tree model, i. e. since all simple families of increasing trees can be considered as weighted trees, we assume that every tree of size n is chosen with probability proportional to its weight. This parameter has been treated in [10] for plane-oriented recursive trees and binary increasing trees. For both tree families explicit formulæ for the probabilities $\mathbb{P}\{D_{n,j} = m\}$ are given, which are obtained by a recursive approach where the sums appearing are brought into closed form via Zeilberger's algorithm. Alternatively a bijective proof of the result for plane-oriented recursive trees is given. Moreover, closed formulæ for the expectation $\mathbb{E}(D_{n,j})$ and the variance $\mathbb{V}(D_{n,j})$ are obtained. For recursive trees this parameter has been studied in [2, 5], where also an explicit formula for the probability $\mathbb{P}\{D_{n,j} = m\}$ is given, obtained from a description via Pólya-Eggenberger urn models. From this explicit formula limiting distribution results are also derived. It has been shown in [5] that, for $n \to \infty$ and j fixed, the normalized quantity $D_{n,j}/n$ is asymptotically Beta-distributed and in [2] it has been proven that, for $n \to \infty$ and $j \to \infty$ such that $j \sim \rho n$ with $0 < \rho < 1$, the random variable $D_{n,j}$ is asymptotically geometrically distributed.

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 [9] 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 has been denoted there by *very simple families of increasing trees* and its characterization via the so called degree-weight generating function is repeated as Lemma 1.

In this work we use a unified 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 $\mathbb{P}\{D_{n,j} = m\}$, which is given in Proposition 1. In the succeeding computations we restrict ourselves to very simple increasing tree families, where we can obtain for all these tree families closed formulæ for the probabilities $\mathbb{P}\{D_{n,j} = m\}$ and the s-th factorial moments $\mathbb{E}\big((D_{n,j})^{\underline{s}}\big) = \sum_{m\ge 0} m^{\underline{s}}\mathbb{P}\{D_{n,j} = m\}$. These explicit results are given in Theorem 1. Furthermore they allow a full characterization of the limiting distribution of $D_{n,j}$, for $n \to \infty$, depending on the growth of j, which is given as Theorem 2. Thus the exact and asymptotic formulæ presented here extend the known results on this subject. We want to mention further that from the closed formula given in Proposition 1 one might derive limiting distribution results for more general families of increasing trees.

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 $\left\{ {n \atop k} \right\}$ the Stirling numbers of the second kind, 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 $(\varphi_k)_{k\ge 0}$ 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 $k \ge 2$ with $\varphi_k > 0$). Furthermore, $\mathcal{L}(T)$ denotes the set of different increasing labellings of the tree T with distinct integers $\{1, 2, \ldots, |T|\}$, where $|T|$ denotes the size of the tree T, and $L(T) := \big|\mathcal{L}(T)\big|$ 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 $(\varphi_k)_{k\ge 0}$ with a degree-weight generating function $\varphi(t) := \sum_{k\ge 0} \varphi_k t^k$, 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 $T(z) := \sum_{n\ge 1} T_n \frac{z^n}{n!}$ satisfies the *autonomous* first order differential equation

$$T'(z) = \varphi\big(T(z)\big), \quad T(0) = 0. \tag{1}$$

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

$$\mathcal{T} = ① \times \Big(\varphi_0 \cdot \{\epsilon\} \,\dot\cup\, \varphi_1 \cdot \mathcal{T} \,\dot\cup\, \varphi_2 \cdot \mathcal{T} * \mathcal{T} \,\dot\cup\, \varphi_3 \cdot \mathcal{T} * \mathcal{T} * \mathcal{T} \,\dot\cup\, \cdots\Big) = ① \times \varphi(\mathcal{T}), \tag{2}$$

where 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., [11]).

By specializing the degree-weight generating function $\varphi(t)$ in (1) we get the basic enumerative results for the three most interesting increasing tree families:

• *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 (1) gives

$$T(z) = \log\Big(\frac{1}{1-z}\Big), \quad \text{and} \quad T_n = (n-1)!, \ \text{for } n \geq 1.$$

• *Plane-oriented 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 (1) leads here to

$$T(z) = 1 - \sqrt{1-2z}, \quad \text{and} \quad T_n = \tfrac{(n-1)!}{2^{n-1}}\tbinom{2n-2}{n-1} = 1\cdot 3\cdot 5 \cdots (2n-3) = (2n-3)!!, \ \text{for } n \geq 1.$$

• *Binary increasing trees* have the degree-weight generating function $\varphi(t) = (1+t)^2$. Thus it follows

$$T(z) = \frac{z}{1-z}, \quad \text{and} \quad T_n = n!, \ \text{for } n \geq 1.$$

In the following we describe the characterization of very simple increasing tree families via the degree-weight generating function $\varphi(t)$ as obtained in [9].

Lemma 1 ([9]) *A simple family of increasing trees $\mathcal{T}$ can be constructed via an insertion process and is thus a very 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 formulæ, with constants $c_1, c_2 \in \mathbb{R}$.*

$$\begin{aligned}
&\textbf{\textit{Case A}}: \quad \varphi(t) = \varphi_0 e^{\frac{c_1 t}{\varphi_0}}, \ \textit{for } \varphi_0 > 0, \ c_1 > 0, \quad (\Rightarrow \ c_2 = 0),\\
&\textbf{\textit{Case B}}: \quad \varphi(t) = \varphi_0\Big(1 + \frac{c_2 t}{\varphi_0}\Big)^d, \ \textit{for } \varphi_0 > 0, \ c_2 > 0, \ d := \frac{c_1}{c_2} + 1 \in \{2, 3, 4, \dots\},\\
&\textbf{\textit{Case C}}: \quad \varphi(t) = \frac{\varphi_0}{(1 + \frac{c_2 t}{\varphi_0})^{-\frac{c_1}{c_2} - 1}}, \ \textit{for } \varphi_0 > 0, \ 0 < -c_2 < c_1.
\end{aligned}$$

The constants c_1, c_2 appearing in Lemma 1 are coming from an equivalent characterization of very simple increasing tree families obtained in [3]: The total weights T_n of trees of size n of $\mathcal{T}$ satisfy for all $n \in \mathbb{N}$ the equation

$$\frac{T_{n+1}}{T_n} = c_1 n + c_2. \tag{3}$$

Solving either the differential equation (1) or using (3) one obtains the following explicit formulæ for the exponential generating function $T(z)$:

$$T(z) = \begin{cases} \frac{\varphi_0}{c_1}\log\big(\frac{1}{1-c_1 z}\big), & \text{Case A},\\ \frac{\varphi_0}{c_2}\Big(\frac{1}{(1-(d-1)c_2 z)^{\frac{1}{d-1}}} - 1\Big), & \text{Case B},\\ \frac{\varphi_0}{c_2}\Big(\frac{1}{(1-c_1 z)^{\frac{c_2}{c_1}}} - 1\Big), & \text{Case C}. \end{cases} \tag{4}$$

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):

$$T_n = \varphi_0 c_1^{n-1}(n-1)!\binom{n-1+\frac{c_2}{c_1}}{n-1}. \tag{5}$$

Finally we want to remark that recursive trees are "Case A," for $\varphi_0 = 1$, $c_1 = 1$, binary increasing trees are "Case B," for $\varphi_0 = 1$, $c_1 = 1$, $c_2 = 1$ ($\Rightarrow d = 2$), plane-oriented recursive trees are "Case C," for $\varphi_0 = 1$, $c_1 = 2$, $c_2 = -1$.

3 Results for very simple families of increasing trees

Theorem 1 *The probabilities $\mathbb{P}\{D_{n,j} = m\}$, which give the probability that the node with label j in a randomly chosen size-n tree of a very simple family of increasing trees as given by Lemma 1, has exactly m descendants, are, for $m \geq 1$ given by the following formula:*

$$\mathbb{P}\{D_{n,j} = m\} = \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{m-1+\frac{c_2}{c_1}}{m-1}\binom{n-m-1}{j-2}}{\binom{n-1}{j-1}\binom{n-1+\frac{c_2}{c_1}}{n-1}}. \tag{6}$$

The s-th factorial moments $\mathbb{E}\big((D_{n,j})^{\underline{s}}\big) = \sum_{m\geq 0} m^{\underline{s}}\mathbb{P}\{D_{n,j} = m\}$ are for $s \geq 1$ given by the following formula:

$$\mathbb{E}\big((D_{n,j})^{\underline{s}}\big) = s!\left(\frac{\binom{n-j}{s}\binom{s+\frac{c_2}{c_1}}{s}}{\binom{j-1+\frac{c_2}{c_1}+s}{s}} + \frac{\binom{n-j}{s-1}\binom{s-1+\frac{c_2}{c_1}}{s-1}}{\binom{j-1+\frac{c_2}{c_1}+s-1}{s-1}}\right). \tag{7}$$

In particular we obtain the following results for the expectation $\mathbb{E}(D_{n,j})$ and the variance $\mathbb{V}(D_{n,j})$:

$$\mathbb{E}(D_{n,j}) = \frac{(c_1+c_2)n - c_2(j-1)}{c_1 j + c_2}, \tag{8}$$

$$\mathbb{V}(D_{n,j}) = \frac{c_1(c_1+c_2)(c_1 n + c_2)(j-1)(n-j)}{(c_1 j + c_2)^2(c_1 j + c_1 + c_2)}. \tag{9}$$

Theorem 2 *The limiting distribution behaviour of the random variable $D_{n,j}$, which counts the number of descendants of the node with label j in a randomly chosen size-n tree of a very simple family of increasing trees as given by Lemma 1, is, for $n \to \infty$ and depending on the growth of j, characterized as follows.*

- *The region for j fixed. The normalized random variable $\frac{D_{n,j}}{n}$ is asymptotically Beta-distributed, $\frac{D_{n,j}}{n} \xrightarrow{(d)} \beta(\frac{c_2}{c_1}+1, j-1)$, i. e. $\frac{D_{n,j}}{n} \xrightarrow{(d)} X$, where the s-th moments of X are for $s \geq 0$ given by*

$$\mathbb{E}(X^s) = \frac{\left(\frac{c_2}{c_1}+1\right)^{\overline{s}}}{\left(\frac{c_2}{c_1}+j\right)^{\overline{s}}}.$$

- *The region for small j: $j \to \infty$ such that $j = o(n)$. The normalized random variable $\frac{j}{n} D_{n,j}$ is asymptotically Gamma-distributed, $\frac{j}{n} D_{n,j} \xrightarrow{(d)} \gamma(\frac{c_2}{c_1}+1, 1)$, i. e. $\frac{j}{n} D_{n,j} \xrightarrow{(d)} X$, where the s-th moments of X are for $s \ge 0$ given by*
$$\mathbb{E}(X^s) = \big(\frac{c_2}{c_1}+1\big)^{\overline{s}}.$$
- *The central region for j: $j \to \infty$ such that $j \sim \rho n$, with $0 < \rho < 1$. The shifted random variable $D_{n,j}-1$ is asymptotically negative binomial-distributed, $D_{n,j}-1 \xrightarrow{(d)}$ $\mathrm{NegBin}(\frac{c_2}{c_1}+1, \rho)$, i. e. $D_{n,j}-1 \xrightarrow{(d)} X$, where the probability mass function of X is given by*
$$\mathbb{P}\{X = m\} = \binom{m+\frac{c_2}{c_1}}{m} \rho^{\frac{c_2}{c_1}+1}(1-\rho)^m, \quad \text{for } m \ge 0.$$
- *The region for large j: $j \to \infty$ such that $l := n - j = o(n)$. The random variable $D_{n,j}$ converges to a random variable, which has all its mass concentrated at 1, i. e. $D_{n,j} \xrightarrow{(d)} X$, with*
$$\mathbb{P}\{X = 1\} = 1.$$

In Section 4 we treat a recurrence for the probabilities $\mathbb{P}\{D_{n,j} = m\}$ via generating functions. This leads for all simple families of increasing trees to a closed formula for this generating function, which is given in Proposition 1. In Section 5 we prove the explicit results for very simple families of increasing trees which are given by Theorem 1, and the corresponding limiting distribution results of Theorem 2 are shown in Section 6.

4 A recurrence for the probabilities

We consider in this section the random variable $D_{n,j}$, which counts the number of descendants of node j in a random increasing tree of size n, for general simple families of increasing trees with degree-weight generating function $\varphi(t)$. In the following we give a recurrence for the probabilities $\mathbb{P}\{D_{n,j} = m\}$, which is obtained from the formal recursive description (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 $\mathbb{P}\{D_{n,j} = m\}$ to the probabilities $\mathbb{P}\{D_{k_1,i} = m\}$. We get as factor the total weight of the r subtrees and the root node $\varphi_r T_{k_1} \cdots T_{k_r}$, 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 that 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, ..., r-th subtree. Again due to symmetry arguments we obtain a factor r, if the node j is the i-th smallest node in the second, third, ..., 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 the following recurrence (10).

$$\begin{aligned}\mathbb{P}\{D_{n,j} = m\} = \sum_{r\ge1} r\varphi_r \sum_{\substack{k_1+\cdots+k_r=n-1,\\ k_1,\dots,k_r\ge1}} \frac{T_{k_1}\cdots T_{k_r}}{T_n} \times \\ \times \sum_{i=1}^{\min\{k_1,j-1\}} \mathbb{P}\{D_{k_1,i} = m\} \binom{j-2}{i-1}\binom{n-j}{k_1-i}\binom{n-1-k_1}{k_2,k_3,\dots,k_r}, \end{aligned} \tag{10}$$

for $n \ge j \ge 2$. For $j = 1$ we obtain $\mathbb{P}\{D_{n,1} = m\} = \delta_{m,n}$.

To treat this recurrence (10) we set $n := k + j$ with $k \ge 0$ and define the trivariate generating function

$$N(z,u,v) := \sum_{k\ge0}\sum_{j\ge1}\sum_{m\ge0} \mathbb{P}\{D_{k+j,j} = m\} T_{k+j} \frac{z^{j-1}}{(j-1)!}\frac{u^k}{k!} v^m. \tag{11}$$

Multiplying (10) with $T_{k+j}\frac{z^{j-2}}{(j-2)!}\frac{u^k}{k!}v^m$ and summing up over $k \ge 0$, $j \ge 2$ and $m \ge 0$ gives then $\frac{\partial}{\partial z}N(z,u,v)$ and $\varphi'\big(T(z+u)\big)N(z,u,v)$ for the left and right hand side of (10), respectively. Since these are essentially straightforward, but lengthy computations, they are omitted here; similar considerations are done in [9], where the recurrences appearing there are treated analogously. In any case we obtain the following differential equation

$$\frac{\partial}{\partial z}N(z,u,v) = \varphi'\big(T(z+u)\big)N(z,u,v), \tag{12}$$

together with the initial condition

$$\begin{aligned} N(0,u,v) &= \sum_{k\ge0}\sum_{m\ge0} \mathbb{P}\{D_{k+1,1} = m\} T_{k+1}\frac{u^k}{k!}v^m = \sum_{k\ge0} T_{k+1}\frac{u^k v^{k+1}}{k!} = vT'(uv) \\ &= v\varphi\big(T(uv)\big). \end{aligned} \tag{13}$$

The general solution of equation (12) is given by

$$N(z,u,v) = C(u,v)\exp\Big(\int_0^z \varphi'\big(T(t+u)\big)dt\Big), \tag{14}$$

with some function $C(u,v)$. Adapting to the initial condition (13) gives the required solution

$$N(z,u,v) = v\varphi\big(T(uv)\big)\exp\Big(\int_0^z \varphi'\big(T(t+u)\big)dt\Big). \tag{15}$$

Due to the equation $T'(z) = \varphi(T(z))$ we further get the simplifications

$$\int_0^z \varphi'\big(T(t+u)\big)dt = \int_0^z \frac{\varphi'\big(T(t+u)\big)T'(t+u)}{\varphi\big(T(t+u)\big)}dt = \int_{T(u)}^{T(z+u)} \big(\log\varphi(w)\big)' dw$$
$$= \log\Big(\frac{\varphi\big(T(z+u)\big)}{\varphi\big(T(u)\big)}\Big),$$

which leads from (15) to the following result.

Proposition 1 *The function $N(z,u,v)$ as defined in equation (11), which is the trivariate generating function of the probabilities $\mathbb{P}\{D_{n,j} = m\}$, which give the probability that the node with label j in a randomly chosen size-n tree of a simple family of increasing trees with degree-weight generating function $\varphi(t)$ has exactly m descendants, is given by the following formula:*

$$N(z,u,v) = \frac{v\varphi\big(T(uv)\big)\varphi\big(T(z+u)\big)}{\varphi\big(T(u)\big)}. \tag{16}$$

5 Computing the probabilities and moments

5.1 An exact formula for the probabilities

From Proposition 1 we can easily compute explicit formulæ for the probabilities $\mathbb{P}\{D_{n,j} = m\}$ for very simple increasing tree families, i. e. increasing tree families, which can be constructed via an insertion process. We will figure out only the Case C and omit the analogous computations for Case A and Case B.

Using Lemma 1 and equation (4) we get

$$\varphi\big(T(z)\big) = \frac{\varphi_0}{(1-c_1z)^{\frac{c_2}{c_1}+1}},$$

and thus from equation (16):

$$N(z,u,v) = \frac{v\varphi_0(1-c_1u)^{\frac{c_2}{c_1}+1}}{(1-c_1uv)^{\frac{c_2}{c_1}+1}(1-c_1(z+u))^{\frac{c_2}{c_1}+1}} = \frac{v\varphi_0}{(1-c_1uv)^{\frac{c_2}{c_1}+1}\big(1-\frac{c_1z}{1-c_1u}\big)^{\frac{c_2}{c_1}+1}}. \tag{17}$$

Extracting coefficients from (17) gives then by using (11) and (5):

$$\mathbb{P}\{D_{k+j,j} = m\} = \frac{(j-1)!k!}{T_{k+j}}[z^{j-1}u^kv^m]N(z,u,v)$$
$$= \frac{(j-1)!k!\varphi_0}{(k+j-1)!\varphi_0c_1^{k+j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}}[z^{j-1}u^kv^{m-1}]\frac{1}{(1-c_1uv)^{\frac{c_2}{c_1}+1}\big(1-\frac{c_1z}{1-c_1u}\big)^{\frac{c_2}{c_1}+1}}$$

$$= \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}}{c_1^k \binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}} [u^k v^{m-1}] \frac{1}{(1-c_1uv)^{\frac{c_2}{c_1}+1}(1-c_1u)^{j-1}}$$

$$= \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{m-1+\frac{c_2}{c_1}}{m-1}}{c_1^{k-m+1} \binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}} [u^k] \frac{u^{m-1}}{(1-c_1u)^{j-1}}$$

$$= \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{m-1+\frac{c_2}{c_1}}{m-1}\binom{k-m+j-1}{j-2}}{\binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}}. \tag{18}$$

It turns out that this formula (18) is indeed valid for all three cases of very simple families of increasing trees. Thus we obtain the first part of Theorem 1 after the substitution $n := k + j$.

5.2 An exact formula for the factorial moments

To obtain the s-th factorial moments of $D_{n,j}$ we use again Proposition 1, but differentiate equation (16) s times w. r. t. v and evaluate it at $v = 1$. For Case C this gives

$$E_v D_v^s N(z,u,v) = \frac{\varphi_0 c_1^s u^s \big(\frac{c_2}{c_1}+1\big)^{\overline{s}}}{(1-c_1u)^{\frac{c_2}{c_1}+s+1}\big(1-\frac{c_1z}{1-c_1u}\big)^{\frac{c_2}{c_1}+1}} + \frac{s\varphi_0 c_1^{s-1} u^{s-1} \big(\frac{c_2}{c_1}+1\big)^{\overline{s-1}}}{(1-c_1u)^{\frac{c_2}{c_1}+s}\big(1-\frac{c_1z}{1-c_1u}\big)^{\frac{c_2}{c_1}+1}}. \tag{19}$$

Extracting coefficients of (19) leads then by using (5) to

$$\mathbb{E}\big((D_{k+j,j})^{\underline{s}}\big) = \sum_{m\ge 0} m^{\underline{s}} \mathbb{P}\{D_{k+j,j}=m\} = \frac{(j-1)!k!}{T_{k+j}} [z^{j-1}u^k] E_v D_v^s N(z,u,v)$$

$$= \frac{1}{\varphi_0 c_1^{k+j-1}\binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}} \Bigg[\varphi_0 c_1^{s+j-1}\Big(\frac{c_2}{c_1}+1\Big)^{\overline{s}}\binom{j-1+\frac{c_2}{c_1}}{j-1}[u^k]\frac{u^s}{(1-c_1u)^{\frac{c_2}{c_1}+s+j}}$$

$$+ s\varphi_0 c_1^{s+j-2}\Big(\frac{c_2}{c_1}+1\Big)^{\overline{s-1}}\binom{j-1+\frac{c_2}{c_1}}{j-1}[u^k]\frac{u^{s-1}}{(1-c_1u)^{\frac{c_2}{c_1}+s+j-1}}\Bigg]$$

$$= \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}}{\binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}} \Bigg[\Big(\frac{c_2}{c_1}+1\Big)^{\overline{s}}\binom{k+j+\frac{c_2}{c_1}-1}{k-s} + s\Big(\frac{c_2}{c_1}+1\Big)^{\overline{s-1}}\binom{k+j+\frac{c_2}{c_1}-1}{k-s+1}\Bigg]$$

$$= \frac{s!\binom{j-1+\frac{c_2}{c_1}}{j-1}}{\binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}} \Bigg[\binom{s+\frac{c_2}{c_1}}{s}\binom{k+j-1+\frac{c_2}{c_1}}{k-s} + \binom{s-1+\frac{c_2}{c_1}}{s-1}\binom{k+j-1+\frac{c_2}{c_1}}{k-s+1}\Bigg],$$

which can be slightly simplified and we get

$$\mathbb{E}\big((D_{k+j,j})^{\underline{s}}\big) = s!\left(\frac{\binom{k}{s}\binom{s+\frac{c_2}{c_1}}{s}}{\binom{j-1+\frac{c_2}{c_1}+s}{s}} + \frac{\binom{k}{s-1}\binom{s-1+\frac{c_2}{c_1}}{s-1}}{\binom{j-1+\frac{c_2}{c_1}+s-1}{s-1}}\right). \tag{20}$$

Since formula (20) is valid also for Case A and Case B, the second part of Theorem 1 follows after substituting $n := k + j$.

6 Limiting distribution results

6.1 The case j fixed

We will show via the method of moments that $D_{n,j}/n \xrightarrow{(d)} \beta(\frac{c_2}{c_1}+1, j-1)$, where $\beta(a,b)$ denotes the Beta-distribution with parameters a and b. If X is a Beta-distributed random variable, $X \stackrel{(d)}{=} \beta(a,b)$, then the s-th moment of X is given by

$$\mathbb{E}(X^s) = \prod_{k=0}^{s-1} \frac{a+k}{a+b+k} = \frac{a^{\overline{s}}}{(a+b)^{\overline{s}}}. \tag{21}$$

Using Stirling's formula for the Gamma function

$$\Gamma(z) = \Big(\frac{z}{e}\Big)^z \frac{\sqrt{2\pi}}{\sqrt{z}} \Big(1 + \frac{1}{12z} + \frac{1}{288z^2} + \mathcal{O}(\frac{1}{z^3})\Big), \tag{22}$$

we obtain for j and s fixed:

$$\binom{n-j}{s} = \frac{n^s}{s!}\big(1+\mathcal{O}(n^{-1})\big).$$

Thus we get from equation (7) the following asymptotic expansion of the s-th factorial moment of $D_{n,j}$:

$$\mathbb{E}\big((D_{n,j})^{\underline{s}}\big) = \frac{\binom{s+\frac{c_2}{c_1}}{s}}{\binom{j-1+\frac{c_2}{c_1}+s}{s}} n^s \big(1+\mathcal{O}(n^{-1})\big).$$

The ordinary moments of $D_{n,j}$ can be expressed by the factorial moments of $D_{n,j}$, where the Stirling numbers of the second kind $\left\{ {n \atop k} \right\}$ are appearing. We obtain then

$$\begin{aligned} \mathbb{E}\big((D_{n,j})^s\big) &= \mathbb{E}\big((D_{n,j})^{\underline{s}}\big) + \sum_{k=1}^{s-1} \left\{ {s \atop k} \right\} \mathbb{E}\big((D_{n,j})^{\underline{k}}\big) \\ &= \frac{\binom{s+\frac{c_2}{c_1}}{s}}{\binom{j-1+\frac{c_2}{c_1}+s}{s}} n^s \big(1+\mathcal{O}(n^{-1})\big) + \mathcal{O}(n^{s-1}) = \frac{\binom{s+\frac{c_2}{c_1}}{s}}{\binom{j-1+\frac{c_2}{c_1}+s}{s}} n^s \big(1+\mathcal{O}(n^{-1})\big). \end{aligned} \tag{23}$$

Thus, for $n \to \infty$ and j fixed, the s-th moments of the normalized random variable $D_{n,j}/n$ converge for all integers $s \geq 1$ to the s-th moments of a Beta-distributed random variable:

$$\mathbb{E}\Big(\big(\frac{D_{n,j}}{n}\big)^s\Big) \to \frac{\binom{s+\frac{c_2}{c_1}}{s}}{\binom{j-1+\frac{c_2}{c_1}+s}{s}} = \frac{\big(\frac{c_2}{c_1}+1\big)^{\overline{s}}}{\big(\frac{c_2}{c_1}+j\big)^{\overline{s}}}, \tag{24}$$

which shows together with the Theorem of Fréchet and Shohat (see e. g. [4]) the first part of Theorem 2.

6.2 The case $j \to \infty$ such that $j = o(n)$

For this region of j we consider the normalized random variable $jD_{n,j}/n$ and will show via the method of moments that $jD_{n,j}/n \xrightarrow{(d)} \gamma(\frac{c_2}{c_1}+1,1)$, where $\gamma(a,\lambda)$ denotes the Gamma-distribution with shape parameter a and scale parameter λ. If X is a Gamma-distributed random variable, $X \stackrel{(d)}{=} \gamma(a,\lambda)$, then the s-th moment of X is given by

$$\mathbb{E}(X^s) = \frac{1}{\lambda^s}\prod_{k=0}^{s-1}(a+k) = \frac{a^{\overline{s}}}{\lambda^s}. \tag{25}$$

Again by using Stirling's formula (22) for the Gamma function we obtain for s fixed:

$$\binom{n-j}{s} = \frac{n^s}{s!}\Big(1+\mathcal{O}\big(\frac{j}{n}\big)\Big), \quad \text{and} \quad \binom{j-1+\frac{c_2}{c_1}+s}{s} = \frac{j^s}{s!}\Big(1+\mathcal{O}\big(\frac{1}{j}\big)\Big),$$

and thus from equation (7) the following expansion of the s-th factorial moments of $D_{n,j}$:

$$\mathbb{E}\big((D_{n,j})^{\underline{s}}\big) = s!\binom{s+\frac{c_2}{c_1}}{s}\big(\frac{n}{j}\big)^s\Big(1+\mathcal{O}\big(\frac{1}{j}\big)+\mathcal{O}\big(\frac{j}{n}\big)\Big). \tag{26}$$

Again, by expressing the ordinary moments of $D_{n,j}$ by its factorial moments, we obtain

$$\mathbb{E}\big((D_{n,j})^{s}\big) = s!\binom{s+\frac{c_2}{c_1}}{s}\big(\frac{n}{j}\big)^s\Big(1+\mathcal{O}\big(\frac{1}{j}\big)+\mathcal{O}\big(\frac{j}{n}\big)\Big). \tag{27}$$

Thus, for $n \to \infty$ and $j \to \infty$ such that $j = o(n)$, the s-th moments of the normalized random variable $jD_{n,j}/n$ converge for all integers $s \geq 1$ to the s-th moments of a Gamma-distributed random variable:

$$\mathbb{E}\Big(\big(\frac{j}{n}D_{n,j}\big)^s\Big) \to s!\binom{s+\frac{c_2}{c_1}}{s} = \big(\frac{c_2}{c_1}+1\big)^{\overline{s}}. \tag{28}$$

This proves the second part of Theorem 2.

6.3 The case $j \to \infty$ such that $j \sim \rho n$

For the central region of j we compute an asymptotic equivalent of the probabilities $\mathbb{P}\{D_{n,j} = m\}$ under the assumption that $j \sim \rho n$ with $0 < \rho < 1$ and show by convergence of the probability mass function that $D_{n,j} - 1 \xrightarrow{(d)} \mathrm{NegBin}(\frac{c_2}{c_1}+1,\rho)$, where $\mathrm{NegBin}(r,p)$ denotes the negative binomial distribution with parameters r and p. If X is a negative binomial-distributed random variable, $X \stackrel{(d)}{=} \mathrm{NegBin}(r,p)$, then the probability mass function of X is given by

$$\mathbb{P}\{X = m\} = \binom{m+r-1}{m}p^r(1-p)^m, \quad \text{for } m \geq 0. \tag{29}$$

We start with the following form of $\mathbb{P}\{D_{n,j}=m\}$ equivalent to (6):

$$\mathbb{P}\{D_{n,j}=m\}=\frac{(j-1)\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{n-j}{m-1}\binom{m-1+\frac{c_2}{c_1}}{m-1}}{m\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-1}{m}}, \tag{30}$$

and apply Stirling's formula (22). This leads to

$$\mathbb{P}\{D_{n,j}=m\}=\binom{m-1+\frac{c_2}{c_1}}{m-1}\big(\frac{j}{n}\big)^{\frac{c_2}{c_1}+1}\big(1-\frac{j}{n}\big)^{m-1}\Big(1+\mathcal{O}\big(\frac{1}{n}\big)+\mathcal{O}\big(\frac{1}{j}\big)+\mathcal{O}\big(\frac{1}{n-j}\big)\Big). \tag{31}$$

Thus, for $n\to\infty$ and $j\to\infty$ such that $j\sim\rho n$ with $0<\rho<1$, the probabilities $\mathbb{P}\{D_{n,j}-1=m\}=\mathbb{P}\{D_{n,j}=m+1\}$ of the shifted random variable $D_{n,j}-1$ converge for all $m\geq 0$ to the probabilities of a negative binomial-distribution:

$$\mathbb{P}\{D_{n,j}-1=m\}\to\binom{m+\frac{c_2}{c_1}}{m}\rho^{\frac{c_2}{c_1}+1}(1-\rho)^m. \tag{32}$$

Thus the third part of Theorem 2 follows.

6.4 The case $l:=n-j=o(n)$

Substituting $l:=n-j$, the probabilities $\mathbb{P}\{D_{n,j}=m\}$ given by (30) can be written as follows:

$$\mathbb{P}\{D_{n,j}=m\}=\frac{\binom{n-l-1+\frac{c_2}{c_1}}{n-l-1}}{\binom{n-1+\frac{c_2}{c_1}}{n-1}}\frac{\binom{l+1}{m}}{\binom{n-1}{m}}\frac{n-l-1}{l+1}\binom{m-1+\frac{c_2}{c_1}}{m-1}. \tag{33}$$

In the sequel we want to obtain a suitable bound for the probabilities $\mathbb{P}\{D_{n,j}=m\}$, which holds uniformly for all $m\geq 2$. Since we are only interested in the case $l:=n-j=o(n)$ we make in the following computations the assumptions $l\leq\frac{n}{3}$ and $n\geq 3$.

First we consider for $2\leq m\leq l+1$:

$$\frac{\binom{l+1}{m}}{\binom{n-1}{m}}=\frac{(l+1)l}{(n-1)(n-2)}\prod_{k=1}^{m-2}\frac{l-k}{n-2-k}. \tag{34}$$

Using the assumptions $l\leq\frac{n}{3}$ and $n\geq 3$ we further get the bounds

$$\frac{(l+1)l}{(n-1)(n-2)}\leq\frac{9\,l^2}{n^2},\quad\text{and}\quad\frac{l-k}{n-2-k}\leq\frac{l}{n},\ \text{for } 1\leq k\leq l. \tag{35}$$

Combining (34) and (35) leads to the estimate

$$\frac{\binom{l+1}{m}}{\binom{n-1}{m}}\leq 9\big(\frac{l}{n}\big)^m, \tag{36}$$

which holds for all $m\geq 2$, since $\binom{l+1}{m}=0$ for $m>l+1$.

For the following estimates we use the bound $|\frac{c_2}{c_1}| \le 1$, which follows from the characterization of very simple families of increasing trees as given by Lemma 1. Together with $l \le \frac{n}{3}$ we get

$$\begin{aligned}\frac{\binom{n-l-1+\frac{c_2}{c_1}}{n-l-1}}{\binom{n-1+\frac{c_2}{c_1}}{n-1}} &= \frac{(n-1)(n-2)\cdots(n-l)}{(n-1+\frac{c_2}{c_1})(n-2+\frac{c_2}{c_1})\cdots(n-l+\frac{c_2}{c_1})} \le \frac{(n-1)(n-2)\cdots(n-l)}{(n-2)(n-3)\cdots(n-l-1)} \\ &= \frac{n-1}{n-l-1} \le 2. \end{aligned} \tag{37}$$

Analogously we compute:

$$\binom{m-1+\frac{c_2}{c_1}}{m-1} = \big(m-1+\frac{c_2}{c_1}\big)\frac{(m-2+\frac{c_2}{c_1})(m-3+\frac{c_2}{c_1})\cdots(1+\frac{c_2}{c_1})}{(m-1)(m-2)\cdots 2} \le m-1+\frac{c_2}{c_1} \le m. \tag{38}$$

Together with the trivial bound

$$\frac{n-l-1}{l+1} \le \frac{n}{l},$$

we finally get from (33) by using (36), (37) and (38) the following estimate, which holds uniformly for all $m \ge 2$:

$$\mathbb{P}\{D_{n,j} = m\} \le 18m\big(\frac{l}{n}\big)^{m-1}. \tag{39}$$

Equation (39) leads then (for $l \le \frac{n}{3}$) to the bound

$$\sum_{m\ge 2} \mathbb{P}\{D_{n,j} = m\} \le 18\sum_{m\ge 2} m\big(\frac{l}{n}\big)^{m-1} = \frac{18l}{n}\frac{2-\frac{l}{n}}{(1-\frac{l}{n})^2} \le \frac{36\frac{l}{n}}{(1-\frac{l}{n})^2} \le \frac{81\,l}{n}. \tag{40}$$

Thus, for $n \to \infty$ and $j \to \infty$ such that $l := n - j = o(n)$, we have

$$\sum_{m\ge 2} \mathbb{P}\{D_{n,j} = m\} \to 0, \quad \text{which implies} \quad \mathbb{P}\{D_{n,j} = 1\} \to 1.$$

Thus also the last part of Theorem 2 is shown.

References

[1] F. Bergeron, P. Flajolet and B. Salvy, Varieties of Increasing Trees, *Lecture Notes in Computer Science* 581, 24–48, 1992.

[2] J. L. Gastwirth and P. K. Bhattacharya, Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable, *Operations Research* 32, 527–536, 1984.

[3] M. Kuba and A. Panholzer, Isolating a leaf in rooted trees via random cuttings, submitted, 2005.
available at `http://info.tuwien.ac.at/panholzer/cuttingleaves8.pdf`

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

[5] H. Mahmoud and R. Smythe, On the distribution of leaves in rooted subtrees of recursive trees, *Annals of Applied Probability* 1, 406–418, 1991.

[6] H. Mahmoud and R. Smythe, A Survey of Recursive Trees, *Theoretical Probability and Mathematical Statistics* 51, 1–37, 1995.

[7] A. Meir and J. W. Moon, On the altitude of nodes in random trees, *Canadian Journal of Mathematics* 30, 997–1015, 1978.

[8] D. Najock and C. C. Heyde, On the number of terminal vertices in certain random trees with an application to stemma construction in philology, *Journal of Applied Probability* 19, 675–680, 1982.

[9] A. Panholzer and H. Prodinger, The level of nodes in increasing trees revisited, *Random Structures & Algorithms*, accepted for publication, 2005. available at `http://info.tuwien.ac.at/panholzer/levelincrev.pdf`

[10] H. Prodinger, Descendants in heap ordered trees, or, A triumph of computer algebra, *Electronic Journal of Combinatorics* 3, paper #29, 1996.

[11] J. Vitter and P. Flajolet, *Average Case Analysis of Algorithms and Data Structures*, in *Handbook of Theoretical Computer Science*, 431–524, Elsevier, Amsterdam, 1990.