A COMBINATORIAL APPROACH FOR ANALYZING THE NUMBER OF DESCENDANTS IN INCREASING TREES AND RELATED PARAMETERS

MARKUS KUBA AND ALOIS PANHOLZER

ABSTRACT. This work is devoted to a study of the number of descendants of node j in random increasing trees, previously treated in [4, 6, 8, 13], and also to a study of a more general quantity called *generalized descendants*, which arises in a certain growth process. Our study is based on a combinatorial approach by establishing a bijection with certain lattice paths. For the parameters considered we derive closed formulæ for the probability distributions, the expectation and the variance, and obtain limiting distribution results also, extending known results in the literature. Furthermore, the bijective approach enables us to study a *weighted version* of the number of descendants of node j in random increasing trees. Moreover, we also discuss the multidimensional case, i.e., the joint distribution of the number of descendants of the nodes j_1 and j_2, and applications.

1. INTRODUCTION

1.1. **Increasing trees.** Increasing trees are labeled trees where the nodes of a tree of size n are labeled 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 [10]) but, additionally, the trees are equipped with increasing labellings. Thus we are considering simple families of increasing trees, which are introduced in [3]. 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 [9]. These models are used in a vast number of applications, e.g., to describe the spread of epidemics, for pyramid schemes, and quite recently as a simplified growth model of the world wide web, since plane-oriented recursive trees and their generalizations are instances of the famous Barabási-Albert [2] network model.

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 $T \in \mathcal{T}$ of size n with vertices $v_1, \dots, v_n$

Date: April 4, 2008.

2000 *Mathematics Subject Classification.* 05C05.

Key words and phrases. Increasing trees, generalized descendants, growth process, limiting distribution.

This research was supported by the Austrian Science Foundation FWF, grant S9608. The second author was also supported by the European FP6-NEST-Adventure Program, contract n° 028875.

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 [12] a full characterization of those simple families of increasing trees, which can be constructed by an insertion process, is given. There this subclass of increasing tree families has been denoted by grown families of increasing trees.

1.2. **Generalized growth process.** In the present paper we consider a *generalized growth process* for so-called label-dependent parameters in increasing tree families, which generalizes the previously considered growth rule [12] at a local level. Such label-dependent parameters are quantifying the local behavior of the tree, by studying, e.g., the number of descendants of node j, the node degree of node j, or the depth of node j, in a size n random increasing tree, for $1 \le j \le n$. The main motivation for a study of such parameters is coming from the necessity of describing the local behaviour of random networks.

In this work we are studying the random variable $D_{n,l,j}$, which counts the number of generalized descendants of node j, under a generalized growth process with parameters j, l and n, where $1 \le j \le n$, $-c_2/c_1 < l < j$ for given n and for certain numbers c_1 and c_2 as specified in Section 3. Note that for $l = 1$ the generalized growth rule coincides with the ordinary growth rule for grown simple families of increasing trees; hence the random variable $D_{n,1,j} = D_{n,j}$ reduces to the number of (ordinary) 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 by the number of nodes) in a random size-n tree.

The random variable $D_{n,j}$ has been treated in [13] 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 [4, 8], where also an explicit formula for the probability $\mathbb{P}\{D_{n,j} = m\}$ is given, which was here obtained by using a description via Pólya-Eggenberger urn models. From this explicit formula limiting distribution results are also derived. It has been shown in [8] that, for $n \to \infty$ and j fixed, the normalized quantity $D_{n,j}/n$ is asymptotically Beta-distributed and in [4] 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. Recently, the probability distribution and limit laws of $D_{n,j}$ were obtained for all grown families of increasing trees in [6], using a generating functions approach.

In this work we extend the results mentioned above by obtaining the probability distribution and limit laws for $D_{n,l,j}$, encompassing the results concerning $D_{n,j} = D_{n,1,j}$ in [4, 6, 8, 13]. In order to obtain our results we do not apply the generating functions approach used in [6] to study $D_{n,l,j}$, but instead we extend the bijective approach of Prodinger [13] to prove our results by studying suitably defined weighted lattice paths.

Furthermore, we study for the case $l = 1$ the random variable $G_{n,j}$, which counts the number of weighted descendants of node j in a random grown simple increasing tree of size n: for this quantity we assume that the trees are growing according to the ordinary growth process as described in [12]. However, we assume further that every node contributes to the number of descendants proportional to its label rather than by one. For example, if node k attaches to the subtree rooted at node j, then the number of weighted descendants increases by k, instead of by one. We also study the joint distribution of the random vector $\mathbf{D}_{n,\mathbf{j}} = (D_{n,1,j_1}, D_{n,1,j_2})$, counting the number of descendants of nodes j_1 and j_2 in a random grown simple increasing tree of size n, with $j_2 > j_1$, assuming the ordinary growth process.

1.3. **Notations.** Throughout this work we interchangeably use the terminology "node j" or just j, which always means the "node labeled j". We denote with $X \overset{(d)}{=} Y$ 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. Furthermore, we denote with $X \oplus Y$ the sum of independent random variables. For the sum of not necessarily independent random variables we write $X + Y$. Let $\{n <_s j\}$ denote the event that node n is contained in the subtree rooted at node j, and with $\{n \not<_s j\}$ the opposite event. We denote with $\mathrm{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ the Beta function, assuming that $\Re(a), \Re(b) > 0$.

1.4. **Plan of the paper.** This work is structured as follows. In the next section we state the definition of simple families of increasing trees and recall the growth process for grown simple families of increasing trees. In Section 3 we introduce the generalized growth process and state some properties of the random variables $D_{n,l,j}$ and $G_{n,j}$. After this we state our main results concerning the distributions of $D_{n,l,j}$, $G_{n,j}$ and $\mathbf{D}_{n,\mathbf{j}}$ in Section 4. Furthermore limiting distribution results are given. The latter sections are devoted to the proofs of the stated results.

2. Preliminaries

2.1. **Formal definition.** 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 $w(T) = \prod_v \varphi_{\deg^+(v)}$, where v ranges over all vertices of T and $\deg^+(v)$ is the out-degree of v (to avoid degenerated cases we always assume that it 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 $\{1, 2, \dots, |T|\}$, 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 $(\varphi_k)_{k\ge 0}$ we define now the total weights by $T_n := \sum_{|T|=n} w(T) \cdot L(T)$.

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

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

where $\textcircled{1}$ denotes the node labeled by 1, $\times$ the cartesian product, $\dot\cup$ the disjoint union, $*$ the partition product for labeled objects, and $\varphi(\mathcal{T})$ the substituted structure (see, e.g., [14]).

For a given degree-weight sequence $(\varphi_k)_{k\ge 0}$, the corresponding degree-weight generating function $\varphi(t)$ is defined by $\varphi(t) := \sum_{k\ge 0} \varphi_k t^k$. It follows from (1) that the exponential generating function $T(z) := \sum_{n\ge 1} T_n \frac{z^n}{n!}$ satisfies the autonomous first order differential equation

$$T'(z) = \varphi\big(T(z)\big), \quad T(0) = 0. \quad (2)$$

2.2. **Examples.** By specializing the degree-weight generating function $\varphi(t)$ in (2) we get 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 (2) gives

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

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 (2) leads here to

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

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

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

2.3. **Growth process.** 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 [12].

- Step 1: The process starts with the root labeled 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 $\deg^+(v)$, of the already grown tree of size i with probabilities

$$p(v) = \begin{cases} \dfrac{1}{i}, & \text{Case A (recursive trees)}, \\[2mm] \dfrac{d-\deg^+(v)}{(d-1)i+1}, & \text{Case B (d-ary increasing tree)}, \\[2mm] \dfrac{\deg^+(v)+\alpha}{(\alpha+1)i-1}, & \text{Case C (generalized plane-oriented recursive trees).} \end{cases} \quad (3)$$

with $\alpha := -1 - \frac{c_1}{c_2} > 0$ and $0 < -c_2 < c_1$ for generalized plane-oriented recursive trees, and $d = 1 + \frac{c_1}{c_2} \in \mathbb{N} \setminus \{1\}$ for d-ary increasing trees.

The constants c_1, c_2 appearing above are coming from an equivalent characterization of grown simple families of increasing trees obtained in [5]: the total weights T_n of trees of size n of $\mathcal{T}$ satisfy for all $n \in \mathbb{N}$ the equation

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

Furthermore the total weights T_n are given by the following formula, which holds for all grown families of increasing trees (setting $c_2 = 0$ for recursive trees and $d = \frac{c_1}{c_2} + 1$ for d-ary increasing trees):

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

Finally we want to remark that recursive trees are obtained by setting $\varphi_0 = 1$, $c_1 = 1$, binary increasing trees by setting $\varphi_0 = 1$, $c_1 = 1$, $c_2 = 1$ ($\Rightarrow d = 2$), and plane-oriented recursive trees by setting $\varphi_0 = 1$, $c_1 = 2$, $c_2 = -1$.

3. Generalized growth process for a label-dependent parameter

Several important quantities in increasing trees are *label-dependent* (also called label-based), which means that the behavior of a specific node labeled j is inspected when the size of the tree grows. Such label-dependent parameters include, e.g., the quantities number of descendants of node j or the node degree of node j. Thus we are only interested in the local behavior of node j and we can consider the tree as being decomposed into two parts: one part is containing node j together with all nodes contributing to the quantity of interest (e.g., all descendants of node j or all children of node j), and the other part is containing all the remaining nodes. See the example given in Figure 1.

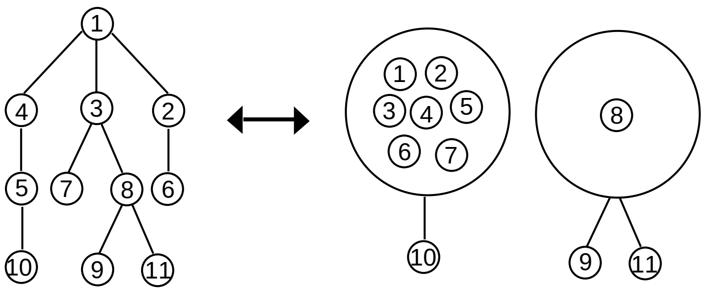

Figure 1. A size-11 increasing tree where node 8 has subtree-size 3, and its equivalent representation.

When considering label-dependent quantities we can generalize the previously stated growth process (3) as follows. We start with two "super nodes " $v_{<j}$ and v_j, where we assume that the outdegrees of the nodes $v_{<j}$ and v_j are zero. The node v_j corresponds to the node labeled j in increasing trees, whereas node $v_{<j}$ corresponds to all nodes with labels smaller than j. Subsequently, nodes with labels larger than j are inserted. Hence, we start at step j with the

nodes $v_{<j}$ and v_j. At step $i+1$, the probability that node $i+1$ is attached to node v_j labeled j is for $i \ge j$ given as follows:

$$p(v_j) = \begin{cases} \frac{l}{i}, & \text{Case A},\\ \frac{l(d-1)+1-\deg^+(v_j)}{(d-1)i+1}, & \text{Case B},\\ \frac{\deg^+(v_j)+l(\alpha+1)-1}{(\alpha+1)i-1}, & \text{Case C}, \end{cases}$$

where $1 \le l \le j$ for integral l (or more general $-c_2/c_1 < l < j$ for arbitrary l), $\alpha := -1 - \frac{c_1}{c_2} > 0$, with $0 < -c_2 < c_1$, for generalized plane-oriented recursive trees, and $d := 1 + \frac{c_1}{c_2} \in \mathbb{N} \setminus \{1\}$ for d-ary increasing trees.

Accordingly, the probability that node $i+1$ attaches to $v_{<j}$ is for $i \ge j$ given as follows.

$$p(v_{<j}) = \begin{cases} \frac{j-l}{i}, & \text{Case A},\\ \frac{(j-l)(d-1)-\deg^+(v_{<j})}{(d-1)i+1}, & \text{Case B},\\ \frac{\deg^+(v_{<j})+(j-l)(\alpha+1)}{(\alpha+1)i-1}, & \text{Case C}, \end{cases}$$

Furthermore we assume that all nodes with labels $k > j$ are attracting a newly inserted node $i+1$ with the probabilities of the ordinary growth process (3), as described in Subsection 2.3. The resulting generalized growth rule leads to a generalization of label-dependent parameters in grown families of increasing trees, such as the subtree-size of node j, the node degree of node j, etc. In this work we will focus on the subtree-size of node j with respect to the generalized growth process.

The generalized growth process presented also makes sense for non-integral, suitably chosen $l \in \mathbb{R}^+$, assuming that $-c_2/c_1 < l < j$, where the fraction c_2/c_1 is given by

$$\frac{c_2}{c_1} = \begin{cases} 0, & \text{Case A},\\ \frac{1}{d-1}, & \text{Case B},\\ -\frac{1}{\alpha+1}, & \text{Case C}. \end{cases} \tag{5}$$

Hence, all our results concerning the generalized growth process also hold for non-integral $l \in \mathbb{R}^+$, assuming that the factorials containing l are replaced by corresponding Gamma functions.

Regarding the parameter subtree-size, we have a natural interpretation of the random variable $D_{n,l,j}$ for integral l as the ordinary subtree-size of, say, node $j-l$, conditioned on the event that the nodes $j+1-l, \dots, j-1, j$ are all attached to $j-l$, and shifted by $l-1$.

Note that for $l = 1$ the cases A, B, C of the generalized growth process coincide with the corresponding cases of the ordinary growth process.

3.1. **Some properties of the random variable $D_{n,l,j}$.** The random variable $D_{n,l,j}$ may be described as a sum of dependent random variables,

$$D_{n,l,j} \stackrel{(d)}{=} \sum_{i=j}^{n} A_{i,l,j}, \quad \text{with } A_{i,l,j} \in \{0,1\}, \quad A_{j,l,j} \stackrel{(d)}{=} 1, \tag{6}$$

and where the conditional probability $\mathbb{P}\{A_{i+1,l,j} = 1 | D_{i,l,j}\}$ is given by

$$\mathbb{P}\{A_{i+1,l,j} = 1 | D_{i,l,j}\} = \begin{cases} \frac{l-1+D_{i,l,j}}{i}, & \text{Case A}, \\ \frac{(d-1)(l-1+D_{i,l,j})+1}{(d-1)i+1}, & \text{Case B}, \\ \frac{(\alpha+1)(l-1+D_{i,l,j})-1}{(\alpha+1)i-1}, & \text{Case C}, \end{cases} \tag{7}$$

with $j \le i \le n-1$. By conditioning one readily obtains the recurrence relation

$$\mathbb{P}\{D_{n,l,j} = m\} = \frac{c_1(m+l-2)+c_2}{c_1(n-1)+c_2}\mathbb{P}\{D_{n-1,l,j} = m-1\} + \frac{c_1(n-m-l)}{c_1(n-1)+c_2}\mathbb{P}\{D_{n-1,l,j} = m\},$$

for $n \ge j+1$ and $m \ge 1$, where c_2/c_1 is given by (5).

The random variable $G_{n,j}$, which counts weighted descendants in grown simple families of increasing trees assuming the ordinary growth process, can be described similar to $D_{n,l,j}$:

$$G_{n,j} \stackrel{(d)}{=} \sum_{i=j}^{n} iA_{i,1,j}, \quad \text{with } A_{j,1,j} \stackrel{(d)}{=} 1,$$

and $A_{i,1,j}$ as defined by (7) setting $l = 1$. Hence $G_{n,j}$ is depending on $D_{n,j} = D_{n,1,j}$.

3.2. **A symmetry relation.** There is a natural symmetry relation between the subtree-sizes of the two nodes v_j and $v_{<j}$, where we allow also non-integral $l \in \mathbb{R}^+$. The distribution of $D_{n,l,j}^{[c]}$, counting the subtree-size of node $v_{<j}$ under the generalized growth process, is given by

$$D_{n,l,j}^{[c]} \stackrel{(d)}{=} D_{n,j-l-\frac{c_2}{c_1},j} \oplus (j-2). \tag{8}$$

Furthermore, we have the obvious relation

$$D_{n,l,j} + D_{n,l,j}^{[c]} \stackrel{(d)}{=} n, \quad D_{n,l,j}^{[c]} \stackrel{(d)}{=} n \oplus (-D_{n,l,j}), \tag{9}$$

since the sum of both of the subtree-sizes must be equal to n. By combining (8) and (9) we obtain then

$$D_{n,l,j} \stackrel{(d)}{=} (n+2-j) \oplus (-D_{n,j-l-\frac{c_2}{c_1},j}). \tag{10}$$

4. RESULTS

4.1. **Results for the probability distributions.**

Theorem 1. *The probability $\mathbb{P}\{D_{n,l,j} = m\}$, which gives the probability that the node with label j has exactly m descendants after step n, for the generalized growth process with parameter l, is given by the following formula:*

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

The expectation and the variance of $D_{n,l,j}$ are given by the following formulæ:

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

Note that for $l = 1$ we regain the results concerning the ordinary descendants.

Theorem 2. *The distribution of the weighted descendants in a grown family of increasing trees is characterized as follows:*

$$\mathbb{P}\{G_{n,j} = m\} = \sum_{k=1}^{n+1-j} \frac{\binom{k-1+\frac{c_2}{c_1}}{k-1}\binom{j-1+\frac{c_2}{c_1}}{j-1}(j-1)}{\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-1}{k}k} a_{m-j,k-1;n,j},$$

for $m \ge j$, where $a_{m,k;n,j}$ denotes the number of partitions of m into k distinct parts, which are restricted to be in $\{j+1, j+2, \dots, n\}$,

$$a_{m,k;n,j} = [u^k v^m] \prod_{l=j+1}^{n} (1 + uv^l).$$

Moreover, the joint distribution of $G_{n,j}$ and $D_{n,j}$ is, for $m \ge j$ and $k \ge 1$, given by

$$\mathbb{P}\{G_{n,j} = m, D_{n,j} = k\} = \frac{\binom{k-1+\frac{c_2}{c_1}}{k-1}\binom{j-1+\frac{c_2}{c_1}}{j-1}(j-1)}{\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-1}{k}k} a_{m-j,k-1;n,j}.$$

Theorem 3. *The distribution of the random vector $\mathbf{D}_{n,\mathbf{j}} = (D_{n,1,j_1}, D_{n,1,j_2})$, counting the number of descendants of the nodes j_1 and j_2 in a random grown simple increasing tree of size n, is given by*

$$\mathbb{P}\{\mathbf{D}_{n,\mathbf{j}} = \mathbf{m}\} = \sum_{k=1}^{j_2-j_1-1} \frac{\binom{m_1-1+\frac{c_2}{c_1}}{m_1-1}\binom{m_2-1+\frac{c_2}{c_1}}{m_2-1}\binom{j_1-1+\frac{c_2}{c_1}}{j_1-1}(j_1-1)\binom{m_1-1}{k-1}\binom{n-m_1-m_2-1}{j_2-k-2}\binom{j_2-1-k}{j_1-1}}{\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-1}{j_2-1}\binom{j_2-2}{j_1-1}(j_2-1)}$$
$$+ \sum_{k=1}^{j_2-j_1-1} \frac{\binom{m_1-m_2-1+\frac{c_2}{c_1}}{m_1-m_2-1}\binom{m_2-1+\frac{c_2}{c_1}}{m_2-1}\binom{j_1-1+\frac{c_2}{c_1}}{j_1-1}\binom{m_1-m_2-1}{k-1}\binom{n-m_1-1}{j_2-k-2}\binom{j_2-2-k}{j_1-2}(k+\frac{c_2}{c_1})}{\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-1}{j_2-1}\binom{j_2-2}{j_1-1}(j_2-1)}.$$

4.2. Results for the limiting distributions.

Theorem 4. *The limiting distribution behavior of the random variable $D_{n,l,j}$, is, for $n \to \infty$, $j = j(n)$ and arbitrary but fixed $l \in (-c_2/c_1, j)$, characterized as follows.*

(1) *The region for j fixed. The normalized random variable $D_{n,l,j}/n$ is asymptotically Beta-distributed, $D_{n,l,j}/n \xrightarrow{(d)} X_{l,j}$, with $X_{l,j} \stackrel{(d)}{=} \beta(l + \frac{c_2}{c_1}, j - l)$. We have the local limit*

$$n\mathbb{P}\{\frac{D_{n,l,j}}{n} = x\} \to f_{X_{l,j}}(x) = \frac{x^{l-1+\frac{c_2}{c_1}}(1-x)^{j-1-l}}{\mathrm{B}(l + \frac{c_2}{c_1}, j - l)}, \quad \text{for } x \in [0, 1].$$

(2) *The region for small j: $j \to \infty$ such that $j = o(n)$. The normalized random variable $jD_{n,l,j}/n$ is asymptotically Gamma-distributed, $jD_{n,l,j}/n \xrightarrow{(d)} X_l$, with $X_l \stackrel{(d)}{=} \gamma(l + \frac{c_2}{c_1}, 1)$. We have the local limit*

$$\frac{n}{j}\mathbb{P}\{\frac{jD_{n,l,j}}{n} = x\} \to f_{X_l}(x) = \frac{x^{l+\frac{c_2}{c_1}-1}}{\Gamma(l + \frac{c_2}{c_1})}e^{-x}, \quad \text{for } x \geq 0.$$

(3) *The central region for j: $j \to \infty$ such that $j \sim \rho n$, with $0 < \rho < 1$. The shifted random variable $D_{n,l,j} - 1$ is asymptotically negative binomial-distributed, $D_{n,l,j} \xrightarrow{(d)} X_{\rho,l}$, with $X_{\rho,l} \stackrel{(d)}{=} \mathrm{NegBin}(l + \frac{c_2}{c_1}, \rho)$,*

$$\mathbb{P}\{X_{\rho,l} = m\} = \binom{m + l - 1 + \frac{c_2}{c_1}}{l - 1}\rho^{l+\frac{c_2}{c_1}}(1-\rho)^m, \quad \text{for } m = 0, 1, \dots$$

(4) *The region for large j: $j \to \infty$ such that $k := n - j = o(n)$. The random variable $D_{n,l,j}$ has asymptotically all its mass concentrated at 1,*

$$\mathbb{P}\{D_{n,l,j} = 1\} = 1 + \mathcal{O}(\frac{k}{n}).$$

Remark 1. The case $l = l(n)$, where l may grow with n, is much more involved. As suggested by the decomposition of $D_{n,j,l}$ given in (6) one often obtains a normal limit law when $l = l(n)$ tends to infinity. One can show that the limiting distribution behavior of the random variable $D_{n,j,l}$, is, for $n \to \infty$, $j = j(n) \to \infty$ and $l = l(n) \to \infty$, assuming that $\mathbb{E}(D_{n,l,j}) \to \infty$, asymptotically normal. For the sake of brevity, we refrain from going into details. In the regions where the expected value of $D_{n,j,l}$ remains bounded, one obtains discrete limiting distributions, such as Binomial distributions or Poisson distributions, which is also not unexpected from the decomposition (6).

5. Generalized descendants

We consider random variables $D_{n,l,j}$, which generalize the previously considered $D_{n,j}$, counting the number of descendants of node j in a random size-n grown increasing tree. We use the idea of Prodinger to count the number of descendants of node j with generalized connectivity. We start at step j with the two nodes $v_{<j}$ and v_j, and continue by attaching

nodes $j+1, j+2, \ldots, n-1, n$ at random, according to the generalized growth process. Note that the larger the subtree of a node, the more likely is the event that a new node will be attached to it.

5.1. **Computations for Case A.** We translate the properties of the growth process into lattice paths. Each state has two entries: the first entry encodes the total number of nodes and the second one gives the size of the subtree rooted at node j $(= v_j)$. Starting at state $j|1$, we can go either to the left or to the right in each step, $k|m \to k+1|m$ or $k|m \to k+1|m+1$. The steps to the left correspond to the events that nodes do not join the subtree rooted at node j, accordingly, steps to the right correspond to the events that nodes join the subtree rooted at node j. The edges are weighted by the probabilities of these events:

$$w(k|m \to k+1|m+1) = \mathbb{P}\{k+1 <_s j | D_{k,l,j} = m\},$$
$$w(k|m \to k+1|m) = \mathbb{P}\{k+1 \not<_s j | D_{k,l,j} = m\}.$$

According to the generalized growth process these probabilities are given by

$$w(k|m \to k+1|m+1) = \frac{m+l-1}{k}, \quad w(k|m \to k+1|m) = \frac{k-m-l+1}{k},$$

for $1 \le m \le k-j+1$ and $k \ge j$. We collect the appropriate weights (transition probabilities) in a diagram, see Figure 2. We are interested in the weight $w(p)$ of paths p starting at $j|1$

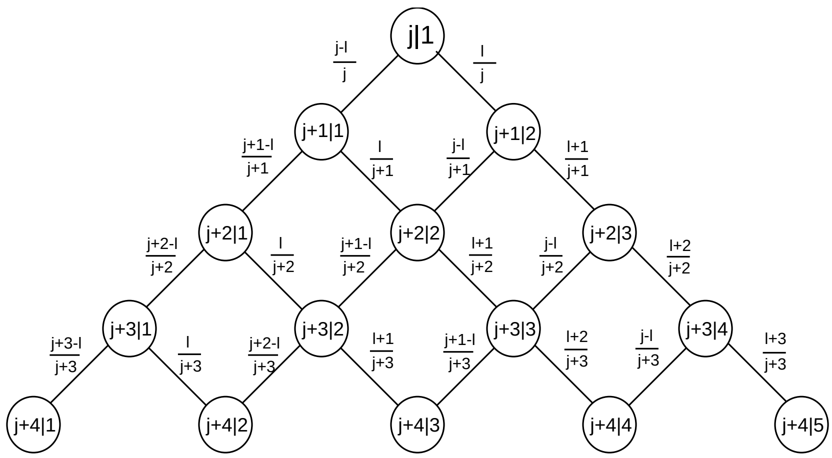

FIGURE 2. The Pascal-like triangle for Case A.

and ending at $n|m$, consisting of the product of the weights ($=$ transition probabilities) of the encountered edges from $j|1$ to $n|m$. The total weight W of all such paths

$$W := \sum_{\text{path } p:\ j|1 \to n|m} w(p),$$

gives then the desired probability that the subtree-size of node j is equal to m after step n:

$$W = \mathbb{P}\{D_{n,l,j} = m\}.$$

The crucial observation for obtaining an exact formula for this quantity is that all paths p, starting at $j|1$ and ending at $n|m$, have the same weight $w(p) = \frac{N}{D}$: regardless of the actual walk we will obtain the denominator D:

$$D = j(j+1)\cdots(n-2)(n-1) = \frac{(n-1)!}{(j-1)!},$$

and the numerator N:

$$N = (\prod_{k=l}^{m+l-2} k)(\prod_{i=j-l}^{n-m-l} i) = \frac{(n-m-l)!(m+l-2)!}{(j-l-1)!(l-1)!}.$$

Hence we have

$$w(p) = \frac{N}{D} = \frac{(n-m-l)!(m+l-2)!(j-1)!}{(j-l-1)!(l-1)!(n-1)!},$$

and further

$$\begin{aligned} W &= \sum_{\text{path } p:\ j|1 \to n|m} w(p) = \frac{(n-m-l)!(m+l-2)!(j-1)!}{(j-l-1)!(l-1)!(n-1)!} \sum_{\text{path } p:\ j|1 \to n|m} 1 \\ &= \frac{(n-m-l)!(m+l-2)!(j-1)!}{(j-l-1)!(l-1)!(n-1)!} \binom{n-j}{m-1} = \frac{\binom{m+l-2}{l-1}\binom{n-m-l}{j-l-1}}{\binom{n-1}{j-1}}, \end{aligned}$$

since the number of paths from $j|1$ to $n|m$ is given by $\binom{n-j}{m-1}$.

5.2. **Computations for Case B and Case C.** Using similar considerations as for Case A we obtain exact results for the remaining cases also. For Case C the edges are weighted as follows:

$$w(k|m \to k+1|m+1) = \frac{(m+l-1)(\alpha+1)-1}{k(\alpha+1)-1},$$
$$w(k|m \to k+1|m) = \frac{(k-m-l+1)(\alpha+1)}{k(\alpha+1)-1},$$

for $1 \le m \le k-j+1$ and $k \ge j$, with $\alpha = -1 - \frac{c_1}{c_2} > 0$. See Figure 3 for an example. We derive again the weight $w(p)$ of paths p starting at $j|1$ and ending at $n|m$, consisting of the product of the weights of the encountered edges from $j|1$ to $n|m$, and then the total weight W of all such paths

$$W := \sum_{\text{path } p:\ j|1 \to n|m} w(p) = \mathbb{P}\{D_{n,j} = m\}.$$

We observe again that all paths p, starting at $j|1$ and ending at $n|m$, have the same weight $w(p) = \frac{N}{D}$. Regardless of the actual walk, we will obtain, according to the growth process, the denominator D:

$$D = \prod_{k=j}^{n-1}((\alpha+1)k-1) = (-\frac{c_1}{c_2})^{n-j} \prod_{k=j}^{n-1}(k + \frac{c_2}{c_1}) = (-\frac{c_1}{c_2})^{n-j} \frac{(n-1)!\binom{n-1+\frac{c_2}{c_1}}{n-1}}{(j-1)!\binom{j-1+\frac{c_2}{c_1}}{j-1}},$$

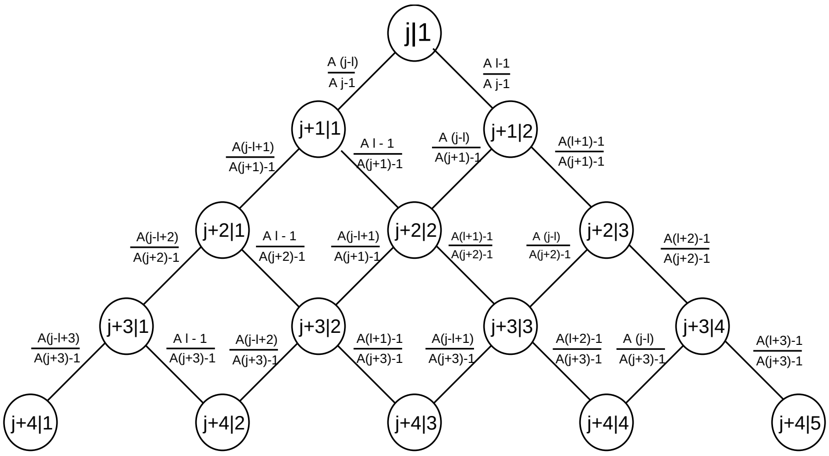

FIGURE 3. The Pascal-like triangle for Case C, with $A := \alpha + 1 = -c_1/c_2 > 1$.

and the numerator N:

$$
\begin{aligned}
B &= \Big(\prod_{k=l}^{m+l-2} (k\alpha + k - 1)\Big)\Big(\prod_{i=j-l}^{n-m-l} (\alpha+1)i\Big) \\
&= (-\frac{c_1}{c_2})^{m-1}\Big(\prod_{k=l}^{m+l-2}(k + \frac{c_2}{c_1})\Big)\frac{(n-m-l)!}{(j-l-1)!}(-\frac{c_1}{c_2})^{n-m-j+1} \\
&= (-\frac{c_1}{c_2})^{n-j}\,\frac{\binom{m+l-2+\frac{c_2}{c_1}}{m+l-2}(m+l-2)!(n-m-l)!}{\binom{l-1+\frac{c_2}{c_1}}{l-1}(l-1)!(j-1-l)!}.
\end{aligned}
$$

Hence we have

$$
w(p) = \frac{\binom{m+l-2+\frac{c_2}{c_1}}{m+l-2}(m+l-2)!\binom{j-1+\frac{c_2}{c_1}}{j-1}(j-1)!(n-m-l)!}{\binom{l-1+\frac{c_2}{c_1}}{l-1}(l-1)!\binom{n-1+\frac{c_2}{c_1}}{n-1}(n-1)!(j-1-l)!},
$$

and consequently

$$
W = w(p)\binom{n-j}{m-1} = \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{m+l-2+\frac{c_2}{c_1}}{m+l-2}\binom{m+l-2}{m-1}\binom{n-m-l}{j-1-l}}{\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{l-1+\frac{c_2}{c_1}}{l-1}\binom{n-1}{j-1}}.
$$

The derivation for Case B is identical. It turns out that this formula is also valid for Case A (by setting $c_2 = 0$).

5.3. **Deriving the expectation and the variance.** To obtain exact formulæ for the first two moments of $D_{n,l,j}$ we use the identity

$$\sum_{m\geq 1}\binom{m+l-2+\frac{c_2}{c_1}}{m+l-2}\binom{m+l-2}{m-1}\binom{n-m-l}{j-1-l} = \frac{\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{l-1+\frac{c_2}{c_1}}{l-1}\binom{n-1}{j-1}}{\binom{j-1+\frac{c_2}{c_1}}{j-1}}. \tag{11}$$

The expectation $\mathbb{E}(D_{n,l,j}) = \sum_{m\geq 1} m\mathbb{P}\{D_{n,l,j} = m\}$ is obtained by using $\binom{n-1}{k-1}\frac{n}{k} = \binom{n}{k}$, the basic decomposition

$$m = -(j-l)\frac{n-m-l+1}{j-l} + n - l + 1,$$

and the identity above (11). We get then the following exact formula for the expectation:

$$\begin{aligned}\mathbb{E}(D_{n,l,j}) &= n-l+1 - \frac{(j-l)\binom{j-1+\frac{c_2}{c_1}}{j-1}}{\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{l-1+\frac{c_2}{c_1}}{l-1}\binom{n-1}{j-1}}\times\\ &\qquad\times\sum_{m\geq 1}\binom{m+l-2+\frac{c_2}{c_1}}{m+l-2}\binom{m+l-2}{m-1}\binom{n+1-m-l}{j-l}\\ &= n-l+1-\frac{(j-l)\frac{n+\frac{c_2}{c_1}}{n}\cdot\frac{n}{j}}{\frac{j+\frac{c_2}{c_1}}{j}} = n-l+1-(j-l)\frac{n+\frac{c_2}{c_1}}{j+\frac{c_2}{c_1}} = \frac{(n-j)(l+\frac{c_2}{c_1})}{j+\frac{c_2}{c_1}}+1.\end{aligned}$$

In order to get the variance $\mathbb{V}(D_{n,l,j}) = \mathbb{E}(D_{n,l,j}^2) - (\mathbb{E}(D_{n,l,j}))^2$ one derives the second moment by using

$$\begin{aligned}m^2 &= (j-l)(j-l+1)\frac{(n-m-l+1)(n-m-l+2)}{(j-l)(j-l+1)}\\ &\quad -(j-l)(2n-2l+3)\frac{n-m-l+1}{j-l}+(n-l+1)^2.\end{aligned}$$

After some simple but lengthy computations, which are omitted here, we finally obtain the following exact formula for the variance:

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

6. WEIGHTED DESCENDANTS

In this section we consider descendants weighted by their label. The ordinary growth process remains unchanged ($l = 1$), but every node contributes to the number of descendants proportional to its label rather than by one. Such weighted generalizations have attained some interest recently, see [1]. Let $G_{n,j}$ denote the weighted subtree-size of node j in a random grown simple increasing tree of size n. We assume that the initial weighted subtree-size is j, $\mathbb{P}\{G_{j,j} = j\} = 1$. We use our previous considerations concerning the ordinary unweighted descendants in order to characterize the distribution of $G_{n,j}$. Every path p_m from $j|1$ to $n|m$ has the same weight $w_m = w(p_m)$:

$$w_m = \frac{\binom{m-1+\frac{c_2}{c_1}}{m-1}(m-1)!\binom{j-1+\frac{c_2}{c_1}}{j-1}(j-1)(n-m-1)!}{\binom{n-1+\frac{c_2}{c_1}}{n-1}(n-1)!}.$$

Going from $j|1$ to $n|m$ means that after node n has been inserted, the subtree rooted at node j is of size m. Equivalently, $m-1$ nodes have been attached to the subtree rooted at node j during the growth from size j to size n. We consider now paths from $j|1$ to $n|k$. Since we are interested in the distribution of $G_{n,j}$, we have to take into account the $k-1$ positions in the path from $j|1$ to $n|k$, where the second coordinate changes - which means that a node is attached to the subtree of node j. We want to compute the number of those paths contributing to $\mathbb{P}\{G_{n,j} = m\}$. As mentioned before we have to look at the $k-1$ transitions, where the second coordinate changes, since $k-1$ nodes are attached to the subtree of node j. At each such transition $b|c \to b+1|c+1$ the weighted number of descendants changes according to the first coordinate b. Hence the searched number of paths from $j|1$ to $n|k$, contributing to $\mathbb{P}\{G_{n,j} = m\}$, is given by $a_{m-j,k-1;n,j}$, the number of partitions of $m-j$ into $k-1$ distinct parts, which are restricted to be in $\{j+1, j+2, \dots, n\}$, $a_{m-j,k-1;n,j} = [u^{k-1}v^{m-j}] \prod_{l=j+1}^{n}(1+uv^l)$. Summing over all possible paths gives

$$\mathbb{P}\{G_{n,j} = m\} = \sum_{k=1}^{n+1-j} w_k a_{m-j,k-1;n,j}.$$

We can also obtain the joint distribution $\mathbb{P}\{G_{n,j} = m, D_{n,j} = k\}$ by considering only paths from $j|1$ to $n|k$:

$$\mathbb{P}\{G_{n,j} = m, D_{n,j} = k\} = \frac{\binom{k-1+\frac{c_2}{c_1}}{k-1}\binom{j-1+\frac{c_2}{c_1}}{j-1}(j-1)}{\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-1}{k}k} a_{m-j,k-1;n,j}.$$

Note that the numbers $a_{m,k;n,j}$ satisfy the relations

$$a_{m,k;n+1,j} = a_{m,k;n,j} + a_{m-n-1,k-1;n,j} = a_{m,k;n+1,j+1} + a_{m-j-1,k-1;n+1,j+1},$$
$$\sum_{m\geq 1} a_{m,k;n,j} = \binom{n-j}{k}.$$

Note that it seems to be a difficult task to study weighted parameters by using the generating functions techniques applied in [6], since the basic tree decomposition (1) cannot be adapted easily to handle weighted parameters, due to the importance of the particular labelings of the subtrees.

7. JOINT DISTRIBUTION OF DESCENDANTS

We determine the distribution of the random vector $\mathbf{D}_{n,\mathbf{j}} = (D_{n,1,j_1}, D_{n,1,j_2})$ by translating again the properties of the growth process into lattice paths. For the sake of simplicity we restrict ourselves to the ordinary growth process. The main difference to the arguments applied before is that we consider a Pascal-like tetrahedra instead of a triangle. Each state has now three entries: the first entry encodes as before the total number of nodes, the second one the size of the subtree rooted at node j_1, and the third one the size of the subtree rooted at node j_2.

The configuration of the initial state is determined by the number of descendants of node j_1 after node j_2 has been inserted. We have to distinguish between the two cases whether node j_2 is contained in the subtree rooted at node j_1 or not: if node j_2 is a descendant of node j_1, an increase of the subtree-size of j_2 simultaneously increases the subtree-size of j_1. Otherwise, this is not the case. We will use the following decomposition:

$$\begin{aligned}\mathbb{P}\{\mathbf{D}_{n,\mathbf{j}} = \mathbf{m}\} = &\sum_{k=1}^{j_2-j_1-1} \mathbb{P}\{\mathbf{D}_{n,\mathbf{j}} = \mathbf{m} | D_{j_2-1,j_1} = k,\, j_2 \not<_s j_1\}\mathbb{P}\{D_{j_2-1,j_1} = k,\, j_2 \not<_s j_1\} \\ &+ \sum_{k=1}^{j_2-j_1-1} \mathbb{P}\{\mathbf{D}_{n,\mathbf{j}} = \mathbf{m} | D_{j_2-1,j_1} = k,\, j_2 <_s j_1\}\mathbb{P}\{D_{j_2-1,j_1} = k,\, j_2 <_s j_1\}.\end{aligned} \tag{12}$$

We already know the quantities $\mathbb{P}\{D_{j_2-1,j_1} = k,\, j_2 \not<_s j_1\}$ and $\mathbb{P}\{D_{j_2-1,j_1} = k,\, j_2 <_s j_1\}$ from our previous considerations and basic properties of the growth process. By lattice path arguments we will determine $\mathbb{P}\{\mathbf{D}_{n,\mathbf{j}} = \mathbf{m} | D_{j_2-1,j_1} = k,\, j_2 \not<_s j_1\}$ and $\mathbb{P}\{\mathbf{D}_{n,\mathbf{j}} = \mathbf{m} | D_{j_2-1,j_1} = k,\, j_2 <_s j_1\}$.

7.1. **Recursive trees.** Starting with a size-j_2 tree, we assume first that node j_2 is not a descendant of node j_1, and further that the subtree-size of node j_1 is k, with $1 \le k \le j_2 - j_1 - 1$. The probability of this event is given by

$$\mathbb{P}\{D_{j_2-1,j_1} = k,\, j_2 \not<_s j_1\} = \mathbb{P}\{D_{j_2-1,j_1} = k\}\mathbb{P}\{j_2 \not<_s j_1 | D_{j_2-1,j_1} = k\}.$$

For recursive trees we have

$$\mathbb{P}\{D_{j_2-1,j_1} = k,\, j_2 \not<_s j_1\} = \mathbb{P}\{D_{j_2-1,j_1} = k\}\frac{j_2 - 1 - k}{j_2 - 1} = \frac{\binom{j_2-2-k}{j_1-2}(j_2 - 1 - k)}{\binom{j_2-2}{j_1-1}(j_2 - 1)}.$$

Starting at state $j_2|k|1$, we can go to the left, to the right, or upwards in each step, $i|k|m \to i+1|k|m$, $i|k|m \to i+1|k+1|m$, or $i|k|m \to i+1|k|m+1$. The steps to the left correspond to the events that nodes do not join the subtrees rooted at node j_1 and j_2, steps to the right correspond to the events that nodes join the subtree rooted at node j_1 and upward steps correspond to the events that nodes join the subtree rooted at node j_2. The edges are weighted by the probabilities of these events:

$$\begin{aligned} w(i|k|m \to i+1|k|m) &= \mathbb{P}\{i+1 \not<_s j_1, j_2 | D_{i,1,j_1} = k, D_{i,1,j_2} = m\}, \\ w(i|k|m \to i+1|k+1|m) &= \mathbb{P}\{i+1 <_s j_1 | D_{i,1,j_1} = k, D_{i,1,j_2} = m\}, \\ w(i|k|m \to i+1|k|m+1) &= \mathbb{P}\{i+1 <_s j_2 | D_{i,1,j_1} = k, D_{i,1,j_2} = m\}. \end{aligned}$$

According to the ordinary growth process these probabilities are given by

$$w(i|k|m \to i+1|k|m) = \frac{i-k-m}{i}, \quad w(i|k|m \to i+1|k+1|m) = \frac{k}{i},$$
$$w(i|k|m \to i+1|k|m+1) = \frac{m}{i}.$$

We have to determine again the weight $w(p)$ of paths p starting at $j_2|k|1$ and ending at $n|m_1|m_2$, consisting of the product of weights (= transition probabilities) of the encountered edges from $j_2|k|1$ to $n|m_1|m_2$. The total weight W of all such paths

$$W := \sum_{\text{path } p\text{:} \quad j_2|k|1 \to n|m_1|m_2} w(p),$$

gives then the desired probability:

$$W = \mathbb{P}\{\mathbf{D}_{n,\mathbf{j}} = \mathbf{m} | D_{j_2-1,j_1} = k,\, j_2 \not<_s j_1\}.$$

We obtain as in the computations of Section 5 that the weight $w(p)$ of all paths p is identical and it is computed easily that they are given by

$$w(p) = \frac{(m_1-1)!(m_2-1)!(j_2-1)!(n-m_1-m_2-1)!}{(k-1)!(n-1)!(j_2-k-2)!}.$$

Hence the total weight W is given by the number of paths from $j_2|k|1$ to $n|m_1|m_2$ times the weight $w(p)$:

$$W = \binom{n-j_2}{m_1-k, m_2-1, n-j_2-m_1-m_2+k+1} w(p) = \frac{\binom{m_1-1}{k-1}\binom{n-m_1-m_2-1}{j_2-k-2}}{\binom{n-1}{j_2-1}}.$$

We assume now that node j_2 is a descendant of node j_1, and further that the subtree-size of node j_1, where we do not count the node j_2, is k, with $1 \le k \le j_2 - j_1 - 1$. The probability of this event is given by

$$\mathbb{P}\{D_{j_2-1,j_1} = k,\, j_2 <_s j_1\} = \mathbb{P}\{D_{j_2-1,j_1} = k\}\mathbb{P}\{j_2 <_s j_1 | D_{j_2-1,j_1} = k\}.$$

For recursive trees we have then

$$\mathbb{P}\{D_{j_2-1,j_1} = k,\, j_2 <_s j_1\} = \mathbb{P}\{D_{j_2-1,j_1} = k\}\frac{k}{j_2-1} = \frac{\binom{j_2-2-k}{j_1-2}k}{\binom{j_2-2}{j_1-1}(j_2-1)}.$$

Instead of considering paths from $j_2|k|1$ to $n|m_1|m_2$, we consider now paths from $j_2|k|1$ to $n|(m_1-m_2)|m_2$, since the subtree-size of of j_2 contributes to the subtree-size of node j_1. The total weight of all paths from $j_2|k|1$ to $n|(m_1-m_2)|m_2$, assuming that $\{j_2 <_s j_1\}$, contributes then to the desired number of descendants as follows:

$$W = \mathbb{P}\{\mathbf{D}_{n,\mathbf{j}} = \mathbf{m} | D_{j_2-1,j_1} = k,\, j_2 \not<_s j_1\}.$$

Since the weight $w(p)$ of all such paths p is given by

$$w(p) = \frac{(m_1-1)!(m_1-m_2)!(j_2-1)!(n-m_1-1)!}{(k-1)!(n-1)!(j_2-k-2)!},$$

we finally obtain

$$W = \binom{n-j_2}{m_1-m_2-k, m_2-1, n-j_2-m_1+1+k} w(p) = \frac{\binom{m_1-m_2-1}{k-1}\binom{n-m_1-1}{j_2-k-2}}{\binom{n-1}{j_2-1}}.$$

Collecting all the results and using (12) gives then the desired expression for $\mathbb{P}\{\mathbf{D}_{n,\mathbf{j}} = \mathbf{m}\}$.

7.2. **Generalized plane-oriented recursive tree and d-ary increasing trees.** For generalized plane-oriented recursive trees (and d-ary increasing trees) we use exactly the same consideration as for recursive trees.

Starting with a size-j_2 tree, we assume first that node j_2 is not a descendant of node j_1, and further that the subtree-size of node j_1 is k, with $1 \leq k \leq j_2 - j_1 - 1$. The probability of this event is given by

$$\begin{aligned}\mathbb{P}\{D_{j_2-1,j_1} = k,\, j_2 \not<_s j_1\} &= \mathbb{P}\{D_{j_2-1,j_1} = k\}\frac{(\alpha+1)(j_2-k)}{(\alpha+1)j_2-1}\\ &= \frac{\binom{j_1-1+\frac{c_2}{c_1}}{j_1-1}\binom{k-1+\frac{c_2}{c_1}}{k-1}\binom{j_2-1-k}{j_1-1}(j_1-1)}{\binom{j_2-1+\frac{c_2}{c_1}}{j_2-1}\binom{j_2-2}{j_1-1}(j_2-1)},\end{aligned}$$

with $\alpha = -1 - \frac{c_1}{c_2} > 0$.

The edges of the corresponding lattice are weighted as follows:

$$\begin{aligned}w(i|h|m \to i+1|h|m) &= \frac{(i-h-m)(\alpha+1)+1}{i(\alpha+1)-1},\\ w(i|h|m \to i+1|h+1|m) &= \frac{h(\alpha+1)-1}{i(\alpha+1)-1},\\ w(i|h|m \to i+1|h|m+1) &= \frac{m(\alpha+1)-1}{i(\alpha+1)-1}.\end{aligned}$$

Regardless of the actual walk from $j_2|k|1$ to $n|m_1|m_2$, we will obtain for $w(p) = \frac{N}{D}$, according to the growth process, the denominator D,

$$D = \prod_{i=j_2}^{n-1}((\alpha+1)i-1) = (-\frac{c_1}{c_2})^{n-j_2}\prod_{i=j_2}^{n-1}(i+\frac{c_2}{c_1}) = (-\frac{c_1}{c_2})^{n-j_2}\frac{(n-1)!\binom{n-1+\frac{c_2}{c_1}}{n-1}}{(j_2-1)!\binom{j_2-1+\frac{c_2}{c_1}}{j_2-1}},$$

and the numerator N:

$$\begin{aligned}N &= \Big(\prod_{i=k}^{m_1-1}(i\alpha+i-1)\Big)\Big(\prod_{i=1}^{m_2-1}(i\alpha+i-1)\Big)\Big(\prod_{i=j_2-k-1}^{n-m_1-m_2-1}(\alpha+1)i\Big)\\ &= (-\frac{c_1}{c_2})^{m_1-k}\Big(\prod_{i=k}^{m_1-1}(i+\frac{c_2}{c_1})\Big)(-\frac{c_1}{c_2})^{m_2}\Big(\prod_{i=1}^{m_2-1}(i+\frac{c_2}{c_1})\Big)\times\\ &\quad\times\frac{(n-m_1-m_2-1)!}{(j_2-k-2)!}(-\frac{c_1}{c_2})^{n-m_1-m_2-j_2+k}\end{aligned}$$

$$= (-\frac{c_1}{c_2})^{n-j_2} \frac{\binom{m_1-1+\frac{c_2}{c_1}}{m_1-1}(m_1-1)!\binom{m_2-1+\frac{c_2}{c_1}}{m_2-1}(m_2-1)!(n-m_1-m_2-1)!}{\binom{k-1+\frac{c_2}{c_1}}{k-1}(k-1)!(j_2-2-k)!}.$$

Hence, the weight $w(p) = \frac{N}{D}$ of all paths p from $j_2|k|1$ to $n|m_1|m_2$ is given by

$$w(p) = \frac{\binom{m_1-1+\frac{c_2}{c_1}}{m_1-1}(m_1-1)!\binom{m_2-1+\frac{c_2}{c_1}}{m_2-1}(m_2-1)!(n-m_1-m_2-1)!(j_2-1)!\binom{j_2-1+\frac{c_2}{c_1}}{j_2-1}}{\binom{k-1+\frac{c_2}{c_1}}{k-1}(k-1)!(j_2-2-k)!(n-1)!\binom{n-1+\frac{c_2}{c_1}}{n-1}}.$$

Thus the total weight W of all such paths is given by

$$W = \frac{\binom{m_1-1+\frac{c_2}{c_1}}{m_1-1}\binom{m_2-1+\frac{c_2}{c_1}}{m_2-1}\binom{j_2-1+\frac{c_2}{c_1}}{j_2-1}\binom{m_1-1}{k-1}\binom{n-m_1-m_2-1}{j_2-k-2}}{\binom{k-1+\frac{c_2}{c_1}}{k-1}\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-1}{j_2-1}}.$$

We assume now that node j_2 is a descendant of node j_1, and further that the subtree-size of node j_1, where we do not count the node j_2, is k, with $1 \le k \le j_2 - j_1 - 1$. The probability of this event is given by

$$\mathbb{P}\{D_{j_2-1,j_1} = k,\, j_2 <_s j_1\} = \mathbb{P}\{D_{j_2-1,j_1} = k\}\frac{k(\alpha+1)-1}{j_2-1} = \frac{\binom{j_1-1+\frac{c_2}{c_1}}{j_1-1}\binom{k+\frac{c_2}{c_1}}{k}\binom{j_2-2-k}{j_1-2}k}{\binom{j_2-1+\frac{c_2}{c_1}}{j_2-1}\binom{j_2-2}{j_1-1}(j_2-1)}.$$

The weight $w(p)$ of all paths p from $j_2|k|1$ to $n|(m_1-m_2)|m_2$ is identical to w, which is given by

$$w = \frac{\binom{m_1-m_2-1+\frac{c_2}{c_1}}{m_1-m_2-1}(m_1-m_2-1)!\binom{m_2-1+\frac{c_2}{c_1}}{m_2-1}(m_2-1)!(n-m_1-1)!(j_2-1)!\binom{j_2-1+\frac{c_2}{c_1}}{j_2-1}}{\binom{k-1+\frac{c_2}{c_1}}{k-1}(k-1)!(j_2-2-k)!(n-1)!\binom{n-1+\frac{c_2}{c_1}}{n-1}}.$$

Hence the total weight W of all such paths is given by

$$W = \frac{\binom{m_1-m_2-1+\frac{c_2}{c_1}}{m_1-m_2-1}\binom{m_2-1+\frac{c_2}{c_1}}{m_2-1}\binom{j_2-1+\frac{c_2}{c_1}}{j_2-1}\binom{m_1-m_2-1}{k-1}\binom{n-m_1-1}{j_2-k-2}}{\binom{k-1+\frac{c_2}{c_1}}{k-1}\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-1}{j_2-1}}.$$

8. Deriving the limiting distribution

The main tool to obtain limiting distribution results is Stirling's formula for the Gamma function:

$$\Gamma(z) = \left(\frac{z}{e}\right)^z \frac{\sqrt{2\pi}}{\sqrt{z}}\Big(1 + \frac{1}{12z} + \frac{1}{288z^2} + \mathcal{O}(\frac{1}{z^3})\Big), \qquad (13)$$

which is applied to the simplified probabilities:

$$\mathbb{P}\{D_{n,l,j} = m\} = \frac{\Gamma(j+\frac{c_2}{c_1})\Gamma(m+l-1+\frac{c_2}{c_1})(n-m-l)!(n-j)!}{\Gamma(n+\frac{c_2}{c_1})\Gamma(l+\frac{c_2}{c_1})(m-1)!(j-1-l)!(n-m-j+1)!}.$$

For l fixed, we basically proceed in the spirit of [6] by deriving (local) limit laws for the growth regions of interest; for the sake of brevity these cases are skipped here, since the proofs can be carried out in the same fashion as [6].

9. An application

There is an interesting application of the random variable $D_{n,l,j}$ to the pathlengths of increasing trees. The pathlength of a size-n tree is defined as the sum of the depths of the nodes 2 up to n. Moreover, the pathlength is also given as the sum of the descendants $P_n = \sum_{k=2}^{n} D_{n,1,k}$. For the generalized quantity $P_{n,l} = \sum_{k=2}^{l} D_{n,1,k}$, with $P_{n,n} = P_n$, we obtain the distributional equations

$$P_{n,l} \overset{(d)}{=} \begin{cases} P_l \oplus D_{n,l-1,l}, & \text{Case A}, \\ P_l + D_{n,(d-2)(l-1)+X_{l,1},l}, & \text{Case B}, \\ P_l + D_{n,(\alpha+1)(l-1)-X_{l,1},l}, & \text{Case C}, \end{cases}$$

where $X_{l,1}$ denotes the outdegree of node 1 in a random size-l tree. Hence, one may combine the results in the literature concerning P_n and the results for $D_{n,l,j}$ to give a precise analysis of $P_{n,l}$.

10. Conclusion

In principle, the approach applied in this paper can be extended to obtain the joint distribution of nodes $j_1, \dots, j_p$ for arbitrary but fixed integer p. However, the expressions get more and more involved.

References

[1] R. Aguech, N. Lasmar, Nabil, and H. Mahmoud, Extremal weighted path lengths in random binary search trees, *Probability in the Engineering and Informational Sciences* 21, 133–141, 2007.

[2] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* 286, 509512, 1999.

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

[4] 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.

[5] M. Kuba and A. Panholzer, Isolating a leaf in rooted trees via random cuttings, accepted for publication in *Annals of Combinatorics*.

[6] M. Kuba and A. Panholzer, Descendants in increasing trees, *Electronic Journal of Combinatorics* 13, research paper #8, 2006.

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

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

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

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

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

[12] A. Panholzer and H. Prodinger, The level of nodes in increasing trees revisited, accepted for publication in *Random Structures and Algorithms*.

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

[14] 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+e104@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