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 [5, 8, 10, 15], and also to a study of the number of descendants of node $j$ in pairs of random trees generated by a certain growth process generalizing the corresponding analysis of various classes of random increasing trees. Our analysis is based on a combinatorial approach, which establishes a bijection with certain lattice paths. For the parameters considered we derive closed formulae 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 $\{1, \ldots, n\}$ 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 [12]) but, additionally, the trees are equipped with increasing labelings. 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 [11]. 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 network model [2].

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 $T$ have the property that for every tree $T \in T$ of size $n$ with vertices $v_1, \ldots, v_n$ there exist probabilities $p_T(v_1), \ldots, p_T(v_n)$, 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 $T$ 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 [14] 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 also consider a generalized growth process, which generates pairs of random trees. This growth rule generalizes at a local level the previously considered growth rule [14] in grown increasing tree families and allows more flexibility in the study of so-called label-dependent parameters. 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 $1 \leq j \leq n$. The main motivation for a study of such parameters is coming from the necessity of describing the local behavior of random growing networks.

In this work we are studying the random variable $D_{n,l,j}$, which counts the number of descendants of node $j$ in a pair of random trees of total size $n$ generated by a generalized growth process with parameter $l$, where we have the restrictions $1 \leq j \leq n$ and $-c_2/c_1 < l < j$ for certain numbers $c_1$ and $c_2$ as specified in Subsection 3.1. Since we will always make the convention that node $j$ is a descendant of itself, we might alternatively say that $D_{n,l,j}$ counts the size of the subtree rooted at $j$ (since, by definition of the growth process, the node labeled $j$ will always be the root of one of the trees, we can also say “size of the tree with root labeled $j$”) in a pair of random trees of total size $n$ generated by a generalized growth process with parameter $l$. A precise description of the generalized growth process considered is given in Subsection 3.1, and a precise definition of the random variable $D_{n,l,j}$ can be found in Subsection 3.2.

For $l = 1$ the generalized growth rule considered in this paper is equivalent to the ordinary growth rule for grown simple families of increasing trees; hence the random variable $D_{n,j} := D_{n,1,j}$ reduces to the number of descendants of a specific node $j$ (with $1 \leq j \leq n$), 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 of a grown increasing tree family.

The random variable $D_{n,j}$ has been treated in [15] for plane-oriented recursive trees and binary increasing trees. For both tree families explicit formulæ for the probabilities $P\{D_{n,j} = m\}$ 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 $E(D_{n,j})$ and the variance $V(D_{n,j})$ are obtained. For recursive trees this parameter has been studied in [5, 10], where also an explicit formula for the probability $P\{D_{n,j} = m\}$ is
ANALYZING THE NUMBER OF DESCENDANTS IN INCREASING TREES

given, which was obtained there by using a description via Pólya-Eggenberger urn models. From this explicit formula limiting distribution results are derived also. It has been shown in [10] that, for \( n \to \infty \) and \( j \) fixed, the normalized quantity \( D_{n,j} / n \) is asymptotically Beta-distributed and in [5] 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 in [8] the probability distribution and limit laws of \( D_{n,j} \) were obtained for all grown families of increasing trees by using a generating functions approach.

In the present work we extend the results mentioned above by obtaining the probability distribution and limit laws for \( D_{n,l,j} \), encompassing the results concerning \( D_{n,j} \) (i.e., \( D_{n,1,j} \)) in [5, 8, 10, 15]. In order to study \( D_{n,l,j} \) we do not apply the generating functions approach used in [8], but instead we prove our results by extending the bijective approach of Prodinger [15] which studies suitably defined weighted lattice paths.

Furthermore, we study for the case \( l = 1 \), i.e., by assuming that the trees are growing according to the ordinary growth process as described in [14], the random variable \( G_{n,j} \), which counts the number of weighted descendants of node \( j \) in a randomly grown simple increasing tree of size \( n \). For this quantity we assume further that every node contributes to the number of descendants proportional to its label rather than by one. For example, if node \( k \) is attached to the subtree rooted at node \( j \), then the number of weighted descendants of node \( j \) increases by \( k \), instead of an increase by one. Again, we make the convention that node \( j \) is a descendant of itself and thus it contributes by \( j \) to the number of weighted descendants of node \( j \).

Moreover, also by assuming that the trees are growing according to the ordinary growth process, we study the joint distribution of the random variables \( D_{n,1,j_1} \) and \( D_{n,1,j_2} \), i.e., we study the distribution of the random vector \( D_{n,j} := (D_{n,1,j_1}, D_{n,1,j_2}) \), which counts simultaneously the number of descendants of nodes \( j_1 \) and \( j_2 \), with \( j_2 > j_1 \), in a random grown simple increasing tree of size \( n \).

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 \( X_n \overset{d}{\to} X \) 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 \( B(a,b) := \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \) 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, give a precise definition of the random variable \( D_{n,l,j} \), and also state some properties of the random variables \( D_{n,l,j} \) and \( G_{n,j} \). After this we state in Section 4 our main results
Binary increasing trees have the degree-weight generating function $T$, leads here to $\varphi$ of different increasing labelings of the tree $T$ 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 $w(T) = \prod_v \varphi_{\deg^+(v)}$, 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 $k \geq 2$ with $\varphi_k > 0$). Furthermore, $L(T)$ denotes the set of different increasing labelings of the tree $T$ with distinct integers $\{1, 2, \ldots, |T|\}$, where $|T|$ denotes the size of the tree $T$, and $L(T) := |L(T)|$ its cardinality. Then the family $T$ consists of all trees $T$ together with their weights $w(T)$ and the set of increasing labelings $L(T)$. For a given degree-weight sequence $(\varphi_k)_{k \geq 0}$ 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 $T$ by the formal recursive equation

$$T = \{\varphi_0\} \cup \{\varphi_1 \cdot T \cup \varphi_2 \cdot T \cdot T \cup \varphi_3 \cdot T \cdot T \cdot T \cup \cdots\} = 1 \times \varphi(T), \quad (1)$$

where $\{\}$ denotes the node labeled by 1, $\times$ the cartesian product, $\cup$ the disjoint union, $\cdot$ the partition product for labeled objects, and $\varphi(T)$ the substituted structure (see, e.g., [16]).

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

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

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

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

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 (2) leads here to

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

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

$$T(z) = \frac{z}{1 - z}, \quad \text{and} \quad T_n = n!, \quad \text{for} \ n \geq 1.$$
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 [14].

- 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

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

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.

Hence, when speaking of random grown simple increasing trees of size $n$ we can think of the random tree model, i.e., any tree of size $n$ is chosen with probability proportional to the weight of the tree, or equivalently and more naturally as a tree of size $n$ generated according to the growth process described above.

The constants $c_1, c_2$ appearing above come from an equivalent characterization of grown simple families of increasing trees obtained in [7]: 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):

$$T_n = \varphi_0 c_1^{n-1} (n-1)! \left( \frac{n - 1 + \frac{c_2}{c_1}}{n - 1} \right).$$  

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. A generalized growth process for label-dependent parameters

3.1. Definition of the generalized growth process. 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 $j$”. Since we are thus only interested in the local behavior of node $j$ we can consider the tree as being decomposed into two parts: one part is containing node $j$ together
with all the 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. When analyzing the label-dependent parameter of interest for node \( j \) in a grown increasing tree we can study directly the growth behaviour of these two parts (in particular the part containing node \( j \)). An example of an increasing tree together with a decomposition into two parts is given in Figure 1.

![Figure 1](image.png)

**Figure 1.** A size-11 increasing tree, where node 8 has subtree-size 3 (i.e., node 8 has 3 descendants), and its decomposition into two parts according to a study of the number of descendants of node 8.

In the following we will introduce a certain growth process, which generates a pair of random trees. When considering label-dependent quantities as “the number of descendants of node \( j \)” or “the number of children of node \( j \)” we can consider this growth process, which depends on a parameter \( l \) (restrictions on this parameter \( l \) are given below), as a generalization of the previously stated growth process (3) for grown increasing tree families. We will call this process “generalized growth process with parameter \( l \)”, since, when studying parameters depending on label \( j \), the case \( l = 1 \) will correspond to a study of “ordinary” random increasing trees. A parameter \( l > 1 \) increases the force of node labeled \( j \) and decreases the force of nodes \( 1, 2, \ldots, j - 1 \) to attract further nodes during the growth process, whereas a parameter \( l < 1 \) has the opposite effect.

Now we give the definition of this generalized growth process: we start with two “super nodes” \( v_{<j} \) and \( v_j \), where we assume that at the beginning the out-degrees 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 \( j \), i.e., it “contains” all the labels \( \{1, 2, \ldots, j - 1\} \). 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 \( i \geq j \) and dependent on the parameter \( l \) 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 \( 1 \leq l \leq j \) for integral \( l \) (or more general \(-c_2/c_1 < l < j\) for arbitrary \( l \)), \( \alpha := -1 - \frac{c_2}{c_1} > 0 \), 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 \( i + 1 \) is attached to node \( v_j \) is for \( i \geq j \) and dependent on the parameter \( l \) 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 with parameter \( l \) describes the growth of a pair of trees, namely a tree with root \( v_j \) and a tree with root \( v_{<j} \). As mentioned before a study of certain parameters in these trees leads to generalizations of results for label-dependent parameters in grown families of increasing trees, such as the number of descendants (= subtree-size) of node \( j \) or the number of children (= node-degree) of node \( j \), etc. In this work we will focus on the number of descendants of node \( j \) (i.e., node \( v_j \)) in a pair of random trees generated by the generalized growth process with parameter \( l \) after step \( n \) (i.e., containing nodes \( v_{<j} = \{1, 2, \ldots, j-1\}, v_j = \{j\} \), and the nodes labeled \( j + 1, j + 2, \ldots, n \)). For simplicity, we will always assume that the “super node” \( v_{<j} = \{1, \ldots, j - 1\} \) contributes \( j - 1 \) to the parameter “size of a tree”. This allows us to say that the pair of trees generated by the generalized growth rule after step \( n \) has total size \( n \). An example of a pair of trees generated by this generalized growth process is given in Figure 2.

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

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

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.

3.2. **Definition of the random variable** \( D_{n,l,j} \). The random variable \( D_{n,l,j} \) counts the number of descendants of the node labeled \( j \) in a pair of random trees of total size \( n \) (i.e., after step \( n \)) generated by the generalized growth process with parameter \( l \) described in Subsection 3.1, where we have the restrictions \( 1 \leq j \leq n \) and \(-c_2/c_1 < l < j\). Alternatively we
**Figure 2.** An example of a pair of trees of total size 15 generated by the generalized growth process with parameter $l$ (with $0 < l < 8$) by starting with the “super nodes” $v_{<8} = \{1, 2, \ldots, 7\}$ and $v_8 = \{8\}$ and inserting the nodes 9 up to 15. The number of descendants of node labeled 8 (i.e., the subtree-size of node labeled 8) is given by 5. When assuming that the generalized growth process grows according to the probabilities given in Case A a further node labeled 16 will be attached to the subtree rooted at 8 with probability $(l + 4)/15$ (with probability $l/15$ it will be attached to 8 and with probability $1/15$ to the nodes 9, 11, 12, 15, respectively) and to the other subtree with probability $(11 - l)/15$ (with probability $(8 - l)/15$ it will be attached to the super node $v_{<8} = \{1, 2, \ldots, 7\}$ and with probability $1/15$ to the nodes 10, 13, 14, respectively).
to the ordinary growth process (3). Consequently, we obtain
\[ \mathbb{P}\{D_{n+1,l,j} = m + 1 | D_{n,l,j} = m\} = \frac{l + m - 1}{n}. \]

In Case B the probability that node \( n + 1 \) is attached to node \( j \) is given by
\[ \frac{l(d-1)+1-\deg^+(v_j)}{(d-1)n+1}, \]
where \( \deg^+(v_j) \) denotes the out-degree of node \( j \). Moreover, due to the ordinary growth process (3), each node \( v \) of the \( m - 1 \) non-root nodes in the tree rooted at node \( j \) attracts the new node \( n + 1 \) with probability \( \frac{d-\deg^+(v)}{(d-1)n+1} \). Since the sum of the node-degrees of the nodes in the tree of size \( m \) rooted at node \( j \) is given by \( m - 1 \), we obtain
\[ \mathbb{P}\{D_{n+1,l,j} = m + 1 | D_{n,l,j} = m\} = \frac{l(d-1)+1-\deg^+(v_j)}{(d-1)n+1} + \sum_{v \in \text{tree with root} j} \frac{d-\deg^+(v)}{(d-1)n+1}. \]

By similar considerations we obtain the corresponding result for Case C also.

By using the notations \( \alpha := -1 - \frac{c_2}{c_1} > 0 \), with \( 0 < -c_2 < c_1 \), and \( d = 1 + \frac{c_2}{c_1} \in \mathbb{N} \setminus \{1\} \), we obtain directly from (6), when conditioning on \( D_{n,l,j} \), the recurrence relation
\[ \mathbb{P}\{D_{n+1,l,j} = m\} = \frac{c_1(m+l-2) + c_2}{c_1n + c_2} \mathbb{P}\{D_{n,l,j} = m - 1\} + \frac{c_1(n+1-m-l)}{c_1n + c_2} \mathbb{P}\{D_{n+l,j} = m\}, \]
for \( n \geq j \) and \( m \geq 1 \).

As mentioned in the introduction \( D_{n,1,j} \) corresponds to the number of descendants \( D_{n,j} \) of a specific node \( j \) (with \( 1 \leq j \leq n \)), i.e., the size of the subtree rooted at \( j \), in a random size-\( n \) tree of a grown increasing tree family, since for \( l = 1 \) the cases \( A, B, C \) of the generalized growth process coincide with the corresponding cases of the ordinary growth process (3).

Concerning the cases of integral \( l \) we have a natural interpretation of the random variable \( D_{n,l,j} \) as the ordinary subtree-size of, say, node \( j-l \), conditioned on the event that the nodes \( j+1-l, \ldots, j-l \) are all attached to \( j-l \), and shifted by \( l \),
\[ \mathbb{P}\{D_{n,l,j} = m\} = \mathbb{P}\{D_{n,1,j-l} = m | D_{j,1,j-l} = l + 1\}. \]

This is readily checked, since by the ordinary growth process (3) the probabilities (6) coincide with the corresponding transition probabilities of \( D_{n,1,j-l} \), assuming that \( D_{j,1,j-l} = l + 1 \). Informally speaking, we assume that the subtree rooted at node \( j-l \) has already \( l+1 \) descendants directly after the insertion of node \( j \); \( D_{j,1,j-l} = l + 1 \); consequently, this subtree attracts new nodes according to (6). However, regarding non-integral \( l \) such interpretations do not seem to make sense.
3.3. Some properties of the random variable $D_{n,l,j}$. Due to the generalized growth process we can decompose the random variable $D_{n,l,j}$ into a sum of dependent random variables,

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

where $A_{i,l,j}$ denotes the indicator of the event that node $i$ is attached to the subtree rooted at node $j$ when considering the generalized growth process with parameter $l$, where the conditional probability $P\{A_{i+1,l,j} = 1 | D_{i,l,j}\}$ is, according to (6), given by

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

with $j \leq i \leq n - 1$.

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

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

and $A_{i,1,j}$ as defined by (8) setting $l = 1$. Hence the random variable $G_{n,j}$ is depending on $D_{n,j} = D_{n,1,j}$.

3.4. A symmetry relation. There is a natural symmetry relation between the subtree-sizes (i.e., the number of descendants) of the two nodes $v_j$ and $v_{<j}$, where we allow also non-integral $l \in \mathbb{R}^+$. By comparing the probabilities of attracting nodes during the generalized growth process described in Subsection 3.1 one obtains that the distribution of $D_{n,l,j}^{[c]}$, counting the subtree-size of node $v_{<j}$ under the generalized growth process (where we recall that node $v_{<j}$ contributes by $j - 1$ to the size of the tree), is given by

$$D_{n,l,j}^{[c]} = D_{n,j-l-\frac{2}{l+1},j} + (j - 2).$$

Furthermore, we have the obvious relation

$$D_{n,l,j} + D_{n,l,j}^{[c]} \equiv n, \quad D_{n,l,j}^{[c]} \equiv n \oplus (-D_{n,l,j}),$$

since the sum of both of the subtree-sizes must be equal to $n$. By combining (9) and (10) we obtain then

$$D_{n,l,j} \equiv (n + 2 - j) \oplus (-D_{n,j-l-\frac{2}{l+1},j}).$$
4. Results

4.1. Exact results for the probability distributions.

Theorem 1. The probability \( P\{D_{n,l,j} = m\} \), which gives the probability that the node with label \( j \) has exactly \( m \) descendants in a pair of random trees of total size \( n \) (i.e., after step \( n \)) generated by the generalized growth process with parameter \( l \), is given by the following formula:

\[
P\{D_{n,l,j} = m\} = \frac{(j-1+c_2)}{j-1+c_1} \left( \frac{m+l-2+c_2}{m+l-2+c_1} \right) \left( \frac{m+l-2}{m+l-2} \right) \left( \frac{n-m-l}{n-1+c_1} \right)^{l-1+c_2} \left( \frac{n-1}{n-1} \right)^{j-1}, \quad 1 \leq m \leq n-j+1.
\]

The expectation and the variance of \( D_{n,l,j} \) are given as follows:

\[
E(D_{n,l,j}) = \frac{(n-j)(l+c_1)}{j+c_1} + 1, \quad \text{Var}(D_{n,l,j}) = \frac{(n+c_1)(n-j)(l+c_1)(j-l)}{(j+c_1)^2(j+1+c_1)}.
\]

Note that for \( l = 1 \) we regain the results \([5, 8, 10, 15]\) concerning descendants in random grown increasing trees.

Theorem 2. The distribution of the random variable \( G_{n,j} \), which counts the number of weighted descendants (where each descendant contributes by its label rather than by one) of node \( j \) in a random tree of size \( n \) of a grown family of increasing trees is characterized as follows:

\[
P\{G_{n,j} = m\} = \sum_{k=1}^{n+1-j} \frac{(k-1+c_2)}{k-1+c_1} \left( \frac{j-1+c_2}{j-1+c_1} \right) \left( \frac{j-1}{j-1} \right) \left( \frac{n-k}{n-1+k} \right)^{k-1} a_{m-j, k-1; n,j},
\]

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

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

Moreover, the joint distribution of \( G_{n,j} \) and \( D_{n,j} \) is, for \( m \geq j \) and \( k \geq 1 \), given as follows:

\[
P\{G_{n,j} = m, D_{n,j} = k\} = \frac{(k-1+c_2)}{k-1+c_1} \left( \frac{j-1+c_2}{j-1+c_1} \right) \left( \frac{j-1}{j-1} \right) \left( \frac{n-k}{n-1+k} \right)^{k-1} a_{m-j, k-1; n,j}.
\]

Theorem 3. The distribution of the random vector \( D_{n,j} = (D_{n,1,j_1}, D_{n,1,j_2}) \), counting simultaneously 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
\[
\mathbb{P}\{D_{n,j} = m\} = 
\sum_{k=1}^{j_2-j_1-1} \frac{(m_1-1+\frac{c_1}{2})_{k-1} m_2-1+\frac{c_1}{2} j_1-1}{m_1-1}_{j_1-1} (j_1-1)_{j_1-1} (m_1-1-m_2-1)_{j_1-1} (j_2-1)_{j_1-1}
\]
\[
+ \sum_{k=1}^{j_2-j_1-1} \frac{(m_1-m_2-1+\frac{c_1}{2})_{k-1} m_2-1+\frac{c_1}{2} j_1-1}{m_1-1}_{j_1-1} (j_1-1)_{j_1-1} (m_1-1-m_2-1)_{j_1-1} (j_2-1-k)_{j_1-1}
\]
\[
\quad \times \frac{(n-1+\frac{c_1}{2})_{n-j} (n-j-1)_{j_2-j-1}}{(n-j-1)_{j_1-1} (n-j)_{j_1-1} (j_2-1)}.
\]

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 \( X_{l,j} \xrightarrow{d} \beta(l + \frac{c_2}{c_1}, j-l) \). We have the local limit
\[
n\mathbb{P}\{\frac{D_{n,l,j}}{n} = x\} \to f_{X_{l,j}}(x) = \frac{x^{l+\frac{c_2}{c_1}}(1-x)^{j-l-1}}{B(l + \frac{c_2}{c_1}, j-l)}, \quad \text{for } x \in [0, 1].
\]

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 \( X_l \xrightarrow{d} \gamma(l + \frac{c_2}{c_1}, 1) \). We have the local limit
\[
n\mathbb{P}\{\frac{jD_{n,l,j}}{n} = x\} \to f_{X_l}(x) = \frac{x^{l+\frac{c_2}{c_1}}}{\Gamma(l + \frac{c_2}{c_1})} e^{-x}, \quad \text{for } x \geq 0.
\]

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 \( X_{\rho,l} \xrightarrow{d} \operatorname{NegBin}(l + \frac{c_2}{c_1}, \rho) \),
\[
\mathbb{P}\{X_{\rho,l} = m\} = \binom{m + l - 1 + \frac{c_2}{c_1}}{l - 1} \rho^{l+\frac{c_2}{c_1}} (1-\rho)^m, \quad \text{for } m = 0, 1, \ldots
\]

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,
\[
\mathbb{P}\{D_{n,l,j} = 1\} = 1 + O\left(\frac{k}{n}\right).
\]

**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 (7) 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 \( \mathbb{E}(D_{n,j,l}) \to \infty \), 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
ANALYZING THE NUMBER OF DESCENDANTS IN INCREASING TREES

Discrete limiting distributions, such as Binomial distributions or Poisson distributions, which is also not unexpected from the decomposition (7).

5. Proof of the Results for Descendants in the Generalized Growth Process

We study here the random variable \( D_{n,l,j} \), which generalizes the previously in the literature considered random variable \( D_{n,j} \), counting the number of descendants of node \( j \) in a random size-\( n \) grown increasing tree. We use the idea of Prodinger [15] to count the number of descendants of node \( j \) with a generalized force of attraction depending on the parameter \( l \). 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 described in Subsection 3.1. 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:

\[
w(k|m \to k+1|m+1) = \mathbb{P}\{k+1 < s \mid D_{k,l,j} = m\}, \quad w(k|m \to k+1|m) = \mathbb{P}\{k+1 \not< s \mid D_{k,l,j} = m\}.
\]

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 \( 1 \leq m \leq k-j+1 \) and \( k \geq j \). We collect the appropriate weights (transition probabilities) in a diagram, see Figure 3. We are interested in the weight \( w(p) \) of paths \( p \) starting at \( j|1 \) 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 $N$:

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

Hence we have

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

and further

$$W = \sum_{p: \text{ path from } j|1 \text{ to } n|m} w(p) = \frac{(n - m - l)!(m + l - 2)!(j - 1)!}{(j - l - 1)!(l - 1)!(n - l)!} \sum_{p: \text{ path from } j|1 \text{ to } n|m} 1$$

$$= \frac{(n - m - l)!(m + l - 2)!(j - 1)!}{(j - l - 1)!(l - 1)!(n - l)!} \left( \begin{array}{c} n - j \\ m - 1 \end{array} \right) \binom{m+l-2}{j-l-1} \binom{n-m-1}{j-1},$$

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:

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

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

for $1 \leq m \leq k - j + 1$ and $k \geq j$, with $\alpha = -1 - \frac{c_1}{c_2} > 0$. See Figure 4 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

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

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,

\[\begin{array}{c}
\text{FIGURE 4. The Pascal-like triangle for Case C, with } A := \alpha + 1 = -\frac{c_1}{c_2} > 1.
\end{array}\]

the denominator $D$:

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

and the numerator $N$:

$$B = \left( \prod_{k=l}^{m+1-2} (k\alpha + k - 1) \right) \left( \prod_{i=j-1}^{m-l} (\alpha + 1)i \right)$$

$$= \left(-\frac{c_1}{c_2}\right)^{m-l} \left( \prod_{k=l}^{m+1-2} (k + \frac{c_2}{c_1}) \right) \frac{(n - m - l)!}{(j - l - 1)!^2} \left(-\frac{c_1}{c_2}\right)^{m-l+j+1}$$

$$= \left(-\frac{c_1}{c_2}\right)^{n-j} \frac{(m+l-2+c_2)}{(m+l-2)} \frac{(m + l - 2)! (n - m - l)!}{(l-1)!^2 (j - 1 - l)!}.$$
Hence we have
\[
    w(p) = \frac{(m+l-2+\frac{\Theta}{c_i}) (m+l-2)! (j-l+1+\frac{\Theta}{c_i}) (j-1)! (n-m-l)!}{(l-1+\frac{\Theta}{c_i}) (l-1)! (n-l+1)! (j-1-l)!},
\]
and consequently
\[
    W = w(p) \left( \frac{n-j}{m-1} \right) = \frac{(j-l+1+\frac{\Theta}{c_i}) (m+l-2+\frac{\Theta}{c_i}) (m+l-2)! (n-m-l)!}{(n-l+1)! (j-1-l)!}. \tag{12}
\]

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 apply the Pfaff-Saalschütz identity (see, e.g., [6], p. 214) and get
\[
    \sum_{m \geq 1} \binom{m+l-2+\frac{\Theta}{c_i}}{m+l-2} \binom{m+l-2}{m-1} \binom{n-m-l}{j-1-l} = \frac{(n-l+1)! (j-1-l)!}{(n-l+1)! (j-1-l)!} \tag{12}
\]
The expectation \( \mathbb{E}(D_{n,l,j}) = \sum_{m \geq 1} m P\{D_{n,l,j} = m\} \) is obtained by using \( \binom{n-1}{k-1} \frac{n}{k} = \binom{\binom{n}{k}}{k} \), the basic decomposition
\[
    m = -(j-l) \frac{n-m-l+1}{j-l} + n - l + 1,
\]
and the identity (12) stated above. We get then the following exact formula for the expectation:
\[
    \mathbb{E}(D_{n,l,j}) = n - l + 1 - \frac{(j-l) (j-l+1+\frac{\Theta}{c_i})}{(n-l+1)! (j-1-l)!} \times 
    \sum_{m \geq 1} \binom{m+l-2+\frac{\Theta}{c_i}}{m+l-2} \binom{m+l-2}{m-1} \binom{n+1-m-l}{j-l} 
    = n - l + 1 - \frac{(j-l) (j-l+1+\frac{\Theta}{c_i}) n}{j+\frac{\Theta}{c_i}} = n - l + 1 - \frac{(j-l) (l+\frac{\Theta}{c_i})}{j+\frac{\Theta}{c_i}} + 1.
\]
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
\[
    m^2 = (j-l) (j-l+1) \frac{(n-m-l+1)(n-m-l+2)}{(j-l)(j-l+1)} 
    - (j-l)(2n-2l+3) \frac{n-m-l+1}{j-l} + (n-l+1)^2.
\]
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. PROOF OF THE RESULTS FOR 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 $j$, $P\{G_{j,j} = j\} = 1$. 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 $w_m = w(p_m)$:

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

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 $P\{G_{n,j} = m\}$. 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 $b|c \rightarrow b + 1|c + 1$ 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 $P\{G_{n,j} = m\}$, 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 $\{j + 1, j + 2, \ldots, n\}$, $a_{m-j,k-1,n,j} = \left\lfloor \frac{(n - 1 + \frac{c_2}{c_1})}{k} \right\rfloor \prod_{l=1}^{n} (1 + uv^l)$. Summing over all possible paths gives

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

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

$$P\{G_{n,j} = m, D_{n,j} = k\} = \frac{\left(\frac{n - 1 + \frac{c_2}{c_1}}{k - 1}\right) \left(\frac{j - 1 + \frac{c_2}{c_1}}{j - 1}\right)(j - 1)}{\left(\frac{n - 1 + \frac{c_2}{c_1}}{n - 1}\right) \left(\frac{k}{k}\right)} a_{m-j,k-1,n,j}.$$
Note that the numbers $a_{m,k;n,j}$ satisfy the relations
\[
a_{m,k;n+1,j} = a_{m,k;n,j} + a_{m-n-1,k-1;n,j} = a_{m,k;n+1,j+1} + a_{m-j-1,k-1;n+1,j+1},
\]
\[
\sum_{m \geq 1} a_{m,k;n,j} = \binom{n-j}{k}.
\]

It seems to be a difficult task to study weighted parameters by using the generating functions techniques applied in [8], 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. Proof of the Results Concerning the Joint Distribution of Descendants

We determine the distribution of the random vector $D_{n,j} = (D_{n,1,j_1}, D_{n,1,j_2})$ 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 entry the properties of the growth process into lattice paths. For the sake of simplicity we will determine the distribution of the random vector $D_{n,j}$ from our previous considerations and basic properties of the growth process. By lattice path arguments we will determine $\mathbb{P}\{D_{n,j} = m | D_{n,1,j_1} = k, j_2 \not< s j_1\}$ and $\mathbb{P}\{D_{n,j} = m | D_{n,1,j_1} = k, j_1 \not< s j_2\}$.

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 $\mathbb{P}\{D_{n,j} = m | D_{n,1,j_1} = k, j_2 \not< s j_1\}$ and $\mathbb{P}\{D_{n,j} = m | D_{n,1,j_1} = k, j_2 < s j_1\}$.

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 $1 \leq k \leq j_2 - j_1 - 1$. 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
\[ P\{D_{j_2-1,j_1} = k, j_2 \not\leq s, j_1\} = P\{D_{j_2-1,j_1} = k\} \frac{j_2 - 1 - k}{j_2 - 1} = \frac{(j_2-2-k)(j_2-1-k)}{(j_2-1)(j_2-1)}. \]

Starting at state \( j_2|k|1 \), we can go to the left, to the right, or upwards in each step, \( i|k|m \rightarrow i + 1|k|m, i|k|m \rightarrow i + 1|k+1|m, \) or \( i|k|m \rightarrow i + 1|k|m + 1 \). 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:
\[ w(i|k|m \rightarrow i + 1|k|m) = P\{i + 1 \not\leq s, j_1, j_2|D_{i,1,j_1} = k, D_{i,1,j_2} = m\}, \]
\[ w(i|k|m \rightarrow i + 1|k+1|m) = P\{i + 1 < s, j_1|D_{i,1,j_1} = k, D_{i,1,j_2} = m\}, \]
\[ w(i|k|m \rightarrow i + 1|k|m + 1) = P\{i + 1 < s, j_2|D_{i,1,j_1} = k, D_{i,1,j_2} = m\}. \]

According to the ordinary growth process these probabilities are given by
\[ w(i|k|m \rightarrow i + 1|k|m) = \frac{i - k - m}{i}, \quad w(i|k|m \rightarrow i + 1|k+1|m) = \frac{k}{i}, \quad w(i|k|m \rightarrow i + 1|k|m + 1) = \frac{m}{i}. \]

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
\[ W := \sum_{\text{path } p: \ j_2|k|1 \rightarrow n|m_1|m_2} w(p), \]
gives then the desired probability:
\[ W = P\{D_{n,1} = m|D_{j_2-1,j_1} = k, j_2 \not\leq s, j_1\}. \]

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 it is 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 \( w(p) \):
\[ W = \binom{n-j_2}{m_1-k, m_2-1, n-j_2-m_1-m_2+k+1} w(p) = \frac{(m_1-1)!(n-m_1-m_2-1)}{(j_2-k-2)!}. \]

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 \( 1 \leq k \leq j_2 - j_1 - 1 \). The probability of this event is given by
\[ P\{D_{j_2-1,j_1} = k, j_2 \not\leq s, j_1\} = P\{D_{j_2-1,j_1} = k\} P\{j_2 < s, j_1|D_{j_2-1,j_1} = k\}. \]
For recursive trees we have then

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

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 node $j_2$ 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 $\{j_2 < j_1\}$, contributes then to the desired number of descendants as follows:

$$W = \mathbb{P}\{D_{n,j} = m|D_{j_2-1,j_1} = k, j_2 \neq j_1\}.$$  

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

$$W = \left(\frac{n - j_2}{m_1 - m_2 - k, m_2 - 1, n - j_2 - m_1 + 1 + k}\right) w(p) = \frac{(m_1 - m_2 - 1)}{(n - 1)} \frac{(n - m_1 - 1)}{1} \frac{(j_2 - k - 2)}{(j_{j_1 - 1} - 1)}.$$

Collecting all the results and using (13) gives then the desired expression for $\mathbb{P}\{D_{n,j} = m\}$.

7.2. Generalized plane-oriented recursive tree and $d$-ary increasing trees. For generalized plane-oriented recursive trees (and $d$-ary increasing trees) we proceed exactly 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 $1 \leq k \leq j_2 - 1 - 1$. The probability of this event is given by

$$\mathbb{P}\{D_{j_2-1,j_1} = k, j_2 \neq j_1\} = \mathbb{P}\{D_{j_2-1,j_1} = k\} \frac{(\alpha + 1)(j_2 - k)}{(\alpha + 1)j_2 - 1} = \frac{(j_2 - k + 1)}{j_{j_1 - 1} - 1} \frac{(j_2 - k)}{j_{j_1 - 1} - 1},$$

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

The edges of the corresponding lattice are weighted as follows:

$$w(i|h|m \rightarrow i + 1|h|m) = \frac{(i - h - m)(\alpha + 1) + 1}{i(\alpha + 1) - 1},$$

$$w(i|h|m \rightarrow i + 1|h + 1|m) = \frac{h(\alpha + 1) - 1}{i(\alpha + 1) - 1},$$

$$w(i|h|m \rightarrow i + 1|h|m + 1) = \frac{m(\alpha + 1) - 1}{i(\alpha + 1) - 1}.$$
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 \( D \),

\[
D = \prod_{i=j_2}^{n-1} ((\alpha + 1)i - 1) = (-\frac{c_1}{c_2})^{n-j_2} \prod_{i=j_2}^{n-1} (i + \frac{c_2}{c_1}) = (-\frac{c_1}{c_2})^{n-j_2} (n-1)!^{\left(\frac{n+1+c_2}{n-1+c_1}\right)}{(j_2 - 1)!^{\left(\frac{j_2+1+c_2}{j_2-1+c_1}\right)}},
\]

and the numerator \( N \):

\[
N = \left( \prod_{i=k}^{m_1-1} (i\alpha + i - 1) \right) \left( \prod_{i=1}^{m_2-1} (i\alpha + i - 1) \right) \left( \prod_{i=j_2-k-1}^{m_1} (\alpha + 1)i \right)
\]

\[
= (-\frac{c_1}{c_2})^{m_1-k} \left( \prod_{i=k}^{m_1-1} (i + \frac{c_2}{c_1}) \right) (-\frac{c_1}{c_2})^{m_2-k} \left( \prod_{i=1}^{m_2-1} (i + \frac{c_2}{c_1}) \right) \times
\]

\[
\frac{(n - m_1 - m_2 - 1)!}{(j_2 - 2 - k)!} \left( -\frac{c_1}{c_2} \right)^{(n-m_1-m_2-j_2+k)}
\]

\[
= (-\frac{c_1}{c_2})^{n-j_2} \left( \frac{m_1+\frac{c_2}{c_1}}{m_1-1} \right) (m_1 - 1)! \left( \frac{m_2+\frac{c_2}{c_1}}{m_2-1} \right) (m_2 - 1)! (n - m_1 - m_2 - 1)!
\]

\[
\frac{(n-1+c_2)}{(n-1+c_1)} (k-1)! (j_2 - 2 - k)! (n - 1)!^{\left(\frac{m_1+\frac{c_2}{c_1}}{m_2+\frac{c_2}{c_1}}\right)}{(j_2+1+c_2)}
\]

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

\[
w(p) = \frac{\left( \frac{m_1+\frac{c_2}{c_1}}{m_1-1} \right) (m_1 - 1)! \left( \frac{m_2+\frac{c_2}{c_1}}{m_2-1} \right) (m_2 - 1)! (n - m_1 - m_2 - 1)! (j_2 - 1)!^{\left(\frac{j_2+1+c_2}{j_2-1+c_1}\right)}}{(k-1)! (j_2 - 2 - k)! (n - 1)!^{\left(\frac{m_1+\frac{c_2}{c_1}}{m_1-1}+\frac{c_2}{c_1}\right)}{(j_2+1+c_2)}}
\]

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

\[
W = \frac{\left( \frac{m_1+\frac{c_2}{c_1}}{m_1-1} \right) \left( \frac{m_2+\frac{c_2}{c_1}}{m_2-1} \right) \left( \frac{j_2+\frac{c_2}{c_1}}{j_2-1} \right) \left( \frac{m_1+\frac{c_2}{c_1}}{m_2+\frac{c_2}{c_1}} \right) \left( \frac{m_1-m_2-1}{n-1} \right) \left( \frac{m_1-m_2-1}{j_2-1} \right)}{(k-1)! \left( \frac{n+\frac{c_2}{c_1}}{n-1} \right) \left( \frac{n+\frac{c_2}{c_1}}{n-1} \right) \left( \frac{n+\frac{c_2}{c_1}}{n-1} \right)}
\]

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 \( 1 \leq k \leq j_2 - j_1 - 1 \). The probability of this event is given by

\[
\mathbb{P}\{D_{j_2-1,j_1} = k; j_2 < s | j_1\} = \mathbb{P}\{D_{j_2-1,j_1} = k\} \frac{k(\alpha + 1) - 1}{j_2 - 1} = \frac{(j_1+1+c_2)(k+1+c_2)(j_2-2-k)}{(j_2+1+c_2)(j_1-1)(j_2-1)}.
\]

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

\[
w = \frac{\left( \frac{m_1-m_2-1+c_2}{m_1-m_2-1+c_1} \right) (m_1 - m_2 - 1)! \left( \frac{m_2-1+c_2}{m_2-1+c_1} \right) (m_2 - 1)! (n - m_1 - 1)! (j_2 - 1)!^{\left(\frac{j_2+1+c_2}{j_2-1+c_1}\right)}}{(k-1)! (j_2 - 2 - k)! (n - 1)!^{\left(\frac{m_1+\frac{c_2}{c_1}}{m_1-1+c_1}\right)}{(j_2+1+c_2)}}
\]
Hence the total weight $W$ of all such paths is given by
\[
W = \frac{(m_1 - m_2 + \frac{c_2}{c_1})(m_2 - 1 + \frac{c_2}{c_1})(j_2 - 1 + \frac{c_2}{c_1})(m_1 - m_2 - 1)(j_2 - k - 2)}{(k_1 - 1 + \frac{c_2}{c_1})(n_1 - 1 + \frac{c_2}{c_1})(n - 1)(j_2 - 1)}.
\]

8. Deriving the limiting distribution results

The main tool to obtain limiting distribution results is Stirling’s formula for the Gamma function:
\[
\Gamma(z) = \left(\frac{z}{e}\right)^z \sqrt{\frac{2\pi}{z}} \left(1 + \frac{1}{12z} + \frac{1}{288z^2} + O\left(\frac{1}{z^3}\right)\right),
\]
which is applied to the simplified probabilities:
\[
\mathbb{P}\{D_{n,l,j} = m\} = \frac{\Gamma(j + \frac{c_2}{c_1})\Gamma(m + l - 1 + \frac{c_2}{c_1})(n - m - l)!(n - j)!}{\Gamma(n + \frac{c_2}{c_1})\Gamma(l + \frac{c_2}{c_1})(m - 1)!(j - 1)!((n - m - j + 1))!}.
\]
For $l$ fixed, we basically proceed in the spirit of [8] 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 [8].

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 $n$. Equivalently, one can check, e.g., by induction, that 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 following distributional equations:
\[
P_{n,l} \overset{(d)}{=} \begin{cases} 
P_1 \oplus D_{n,l-1,j}, & \text{Case A}, \\
P_1 + D_{n,(d-2)(l-1) + X_{l,1,j}}, & \text{Case B}, \\
P_1 + D_{n,(\alpha+1)(l-1) - X_{l,1,j}}, & \text{Case C}, 
\end{cases}
\]
where $X_{l,1}$ denotes the outdegree of node 1 in a random size-$l$ tree. Hence, it seems to be possible to combine the results in the literature concerning $P_n$ (see, e.g., [4] for results in random recursive trees), and the results for $D_{n,l,j}$ presented in this work to give a precise analysis of $P_{n,l}$. The authors plan to investigate this matter in a future work.

10. Conclusion and Acknowledgement

In this work we considered several generalizations of the original problem of analyzing the number of descendants of a specific node in random grown increasing trees by applying and extending the original approach of Prodinger [15]. We remark that this approach could be extended further, e.g., to obtain even the joint distribution of nodes $j_1, \ldots, j_p$ for an arbitrary but fixed integer $p$. However, the expressions appearing get more and more involved and we refrain here from giving such an analysis.
The authors thank the referees for many valuable remarks for improving the presentation of this work.

REFERENCES


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