TECHNISCHE UNIVERSITÄT WIEN

VIENNA UNIVERSITY OF TECHNOLOGY

DISSERTATION

Analysis of node isolation procedures and label-based parameters in tree structures

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften
unter der Leitung von

A. o. Univ. Prof. Dipl.-Ing. Dr. techn. Alois Panholzer
Institut für Diskrete Mathematik und Geometrie
E104

eingereicht an der

Technischen Universität Wien
Fakultät für technische Naturwissenschaften und Informatik

von

Dipl.-Ing. Markus Kuba
9925955

Favoritenstraße 14/1/4,
1040 - Wien

Datum

Unterschrift

Danksagung

Ich möchte mich sehr herzlich bei meinem Betreuer Professor Alois Panholzer bedanken, der mich während des gesamten Doktorratsstudiums und auch davor fortwährend unterstützt hat und sich stets viel Zeit für meine Anliegen nahm. Weiters möchte ich mich für die Motivation und Inspiration, ausgehend von einer Vielzahl an Gesprächen, bedanken.

Aufrichtiger Dank gebührt meiner Familie. Ich danke meiner Mutter für die große Unterstützung über all die Jahre hinweg, für die Aufmunterungen und fürs Zuhören, selbst bei der für sie doch recht unverständlichen Mathematik; meinem Vater, durch welchen ich von klein auf mit Mathematik in Berührung kam, und welcher immer ein offenes Ohr für meine kleinen und großen Probleme jeglicher Art hatte. Meinem Bruder danke ich für den Ansporn eine ebenso große Meisterschaft in der Mathematik zu erlangen wie er im Schachsport. Ebenso danke ich Ihm für die Anteilnahme an den manchmal vorhandenen Unbillen des Studiums. Dank an meine Tante, welche mir den Wert von Bildung nahegebracht hat. Weiters danke ich meiner Familie dafür, dass sie mir das Studium ermöglicht haben.

Ein großer Dank gilt Steffi, dafür, dass sie immer für mich da war, und mich auf dem Weg begleitet. Weiters für die vielen schönen Momente, die mir das bewältigen der mathematischen Hürden erleichtert haben.

Dank an alle meine Studienkollegen und Freunde, welche mich durchs Studium begleitet haben und mit mir Freud und Leid des Mathematikstudiums teilten. Ich möchte Bernhard und Christoph für das genaue Korrekturlesen danken

Kurzfassung

Die vorliegende Arbeit beschäftigt sich mit der Analyse von Prozeduren zur Knotenisolation sowie markierungsbasierten Parametern in verschiedenen Baummodellen. Bäume können zur Modellierung der unterschiedlichsten Sachverhalte verwendet werden. Obwohl die meisten unmittelbaren Anwendungen von Bäumen immer noch in der Informatik liegen, hat es sich gezeigt, dass Bäume ebenso gewinnbringend zur Modellierung von Pyramidenspielen, Ausbreitungen von Infektionen und dergleichen verwendet werden können. Da Bäume zu den einfachsten rekursiv definierten Objekten zählen, können in den durch Bäume modellierten Anwendungen oftmals Beweise für ein bestimmtes Verhalten des Modells gegeben werden, welche bei komplexerer Modellierung viel schwerer zu erhalten sind.

Die zur Analyse der Baummodelle bzw. der darin relevanten Parameter angewandte Vorgangsweise ist die folgende. Zunächst werden unter der Vorraussetzung, dass jeder Baum der Größe n einer gewissen Baumfamilie mit gleicher Wahrscheinlichkeit auftritt, Rekursionen für die zu untersuchenden Parameter aufgestellt. Mittels passend definierter erzeugender Funktionen lassen sich diese Rekursionen in Funktional- bzw. Differentialgleichungen übersetzen. Aus den Funktional- bzw. Differentialgleichungen gewinnt man, falls möglich, explizite Lösungen, anhand welcher mittels Koeffizientenablesens exakte Ergebnisse erzielbar sind. Andernfalls läßt sich oft hinreichende Information über das asymptotische Verhalten der untersuchten Parameter aus der Struktur der Funktional- bzw. Differentialgleichungen gewinnen. Um Grenzverteilungen von Zufallsvariablen zu erhalten werden der Stetigkeitssatz von Lévy sowie die sogenannte Methode der Momente (Satz von Fréchet und Shohat) angewandt.

Diese Arbeit ist drei Themenbereichen gewidmet. Die ersten vier Kapiteln beschäftigen sich mit verschiedenen Algorithmen zur Knotenisolierung in Bäumen. Dabei wird in einem zufälligen Baum der Größe n einer gewissen Baumfamilie zufällig eine Kante entfernt (ein Zufallsschnitt). Dadurch entstehen zwei neue Teilbäume. Nun wird je nach Algorithmus in einem oder beiden Teilbäumen diese Prozedur fortgesetzt, bis eine gewisse Auswahl an Knoten isoliert ist. Die hier vorgestellten Algorithmen zur Knotenisolierung verallgemeinern die bisher untersuchten Verfahren. Unter anderen untersuchen wir die benötigte Anzahl an Zufallsschnitten bis der Knoten n eines rekursiven Baumes der Größe n isoliert ist.

Der zweite Schwerpunkt dieser Arbeit liegt auf der Analyse von markierungsbasierten Parametern. Wir untersuchen in der Baumfamilie der Increasing Trees (strikt aufsteigend markierte Bäume) Parameter welche von der Knotenmarkierung abhängen. Im Gegensatz zu den globalen (extremalen) Parametern erhalten wir unterschiedliches Verhalten des Parameters je nach Abhängigkeit der Knotenmarkierung von der Baumgröße. Wir untersuchen unter anderem Knotengrad, Unterbaumgröße, Aststruktur, etc., des Knoten j in einem Baum der Größe n. Für die entsprechend definierten Zufallsvariablen erhalten wir explizite Resultate für die Wahrscheinlichkeitsverteilungen, die (faktoriellen) Momente sowie die Grenzverteilungen. Dabei verwenden wir einen rekursiven Zugang, welchen wir zu einer allgemeinen Methode zur Untersuchung von markierungsbasierten Parametern ausbauen.

Weiters untersuchen wir die Baumfamilie der Scale free trees. Es wird der Zusammenhang mit der Baumfamilie der Increasing Trees gezeigt und es werden wiederum markierungsbasierte Parameter untersucht.

Der dritte Teil beschäftigt sich mit gewichteten Parametern, welche nicht so einfach durch Rekursionen beschreibbar sind, da Ummarkierungsargumente nur schwer verwendet werden können. Diese gewichteten Parameter stellen eine Verallgemeinerung der zuvor untersuchten markierungsbasierten Parameter dar. Mittels der Resultate über markierungsbasierte Parameter sind wir dennoch in der Lage mittels probabilistischer Methoden eine Vielzahl an Ergebnissen zu erlangen.

Diese Dissertation basiert auf mit Prof. Alois Panholzer verfassten Forschungsarbeiten, welche im Rahmen des FWF-Projekt P18009, Analyse von Datenstrukturen und baumartigen Strukturen, sowie des National Research Network S9600 Analytic Combinatorics and Probabilistic Number Theory, Teilprojekt S9608 - Combinatorial Analysis of Data Structures and Tree-Like Structures, erstellt wurden.

Abstract

This thesis is dedicated to the analysis of node isolation procedures and label-based parameters in several tree models. Most of the applications of trees are found in computer science. Nevertheless they are also used in other scientific areas, e.g. for modelling the spread of epidemics, pyramid schemes, growth of networks such as the internet, etc. Trees are some of the easiest recursive structures. Modelling with trees often leads to explicit results concerning the behavior of the model. In contrast when more complex structures are used for modelling it is most likely harder to prove explicit results.

We will use mainly the following approach for obtaining results. Under the assumption that any tree of size n of a certain tree family is picked equally likely, we can set up recurrences for the parameters of interest by using the recursive structure of the examined tree family. By using suitably defined generating functions these recurrences can be translated into either functional equations or differential equations, depending on the recursive structure of the tree family. Most of the times when the arising equations can be explicitly solved, exact results can be obtained by extraction of coefficients. Sometimes we can directly deduce asymptotic results from the functional equations or differential equations. In order to get limiting distribution results for the considered parameters we will rely on two methods. Lévy's continuity theorem and the so-called "Method of moments" will be our main tools.

This thesis is divided into three parts. In the four beginning chapters we will analyze algorithms for node isolation by random cuttings in rooted trees: Pick at random an edge e in a random rooted tree T of size n. Now remove the edge e. This splits T into two rooted subtrees $\hat{T}$ and $\tilde{T}$, where w.l.o.g. $\hat{T}$ is rooted at the original root and $\tilde{T}$ is rooted at the node adjacent to e. Apply this procedure recursively on one or both subtrees $\hat{T}$ and $\tilde{T}$, depending on the algorithm, until a prescribed set of nodes is isolated.

The algorithms analyzed here generalize the known procedures for node isolation. Among others we analyze the number of random cuts necessary to isolate node n in a recursive tree of size n.

The second part is devoted to the analysis of label-based parameters in increasing tree families. Phase transitions occur for label-based parameters depending on the growth of the considered label. We analyze parameters like node degree, subtree size, branching structure, distance, etc.; furthermore we obtain explicit results for both the distribution and the factorial moments of the corresponding random variables. Also limiting distribution results are readily obtained. We use a recursive approach for the description of label-based parameters, which will be turned into a general method for studying arbitrary label-based parameters.

We extend our studies of label-based parameters to the tree family of Scale free trees. After establishing a combinatorial description of this tree family it will turn out that Scale free trees are non-simple increasing trees. Nevertheless we are able to reduce the study of parameters in Scale free trees to the study of a subclass of simple increasing trees.

The third part is devoted to the analysis of weighted parameters in labelled rooted trees. These parameters are generalizations of label-based parameters. It is difficult to get recursive descriptions of the parameters of interest using a combinatorial approach because we cannot use relabelling arguments to obtain recurrences. A probabilistic approach allows us to obtain results for weighted depths and distances and various kinds of weighted node degrees.

This thesis is based on several research papers jointly written with Prof. Alois Panholzer. Most of the work was done within two research projects: FWF-project P18009, "Analyse von Datenstrukturen und baumartigen Strukturen" and National Research Network S9600 "Analytic Combinatorics and Probabilistic Number Theory", project S9608 - "Combinatorial Analysis of Data Structures and Tree-Like Structures".

Contents

Chapter 0

Mathematical Preliminaries

We will give a thorough introduction to the mainly considered tree families, for which different parameters will be analyzed in the next chapters. Some special tree families like the family of non-crossing trees will be presented in the corresponding chapter.

0.1 Tree families

0.1.1 Simply generated trees

Simply generated trees were introduced in [57] and they include several important tree families as special instances, e. g. binary trees, unordered labelled trees (Cayley trees), and ordered trees (= planted plane trees). Moreover, they are strongly related to Galton-Watson branching processes, since it is well known (see [2]), that random simply generated trees are essentially the same as conditioned Galton-Watson trees obtained as the family tree of a Galton-Watson process conditioned on the given total size.

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

$$T(z) = z\varphi\big(T(z)\big), \tag{1}$$

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

The asymptotic behavior of $T(z)$ as solution of (1) is discussed in detail in [24] and we collect some of their results concerning $T(z)$ and the growth of its coefficients T_n, where we have to make only few restrictions on $\varphi(t)$. We will suppose that $\varphi(t)$ has a positive radius of convergence $R > 0$ and assume that there exists a minimal positive solution $\tau < R$ of the equation $t\varphi'(t) = \varphi(t)$.

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

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

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

By applying singularity analysis one obtains the asymptotic expansion

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

provided that $n \equiv 1 \pmod p$. (For $n \not\equiv 1 \pmod p$ $T_n = 0$ always holds.)
We want to mention further that it is often advantageous to describe a simply generated tree family $\mathcal{T}$ by the formal recursive equation

$$\mathcal{T} = \bigcirc \times \Big(\varphi_0 \cdot \{\epsilon\} \,\dot\cup\, \varphi_1 \cdot \mathcal{T} \,\dot\cup\, \varphi_2 \cdot \mathcal{T}\times\mathcal{T} \,\dot\cup\, \varphi_3 \cdot \mathcal{T}\times\mathcal{T}\times\mathcal{T} \,\dot\cup\, \cdots\Big) = \bigcirc \times \varphi(\mathcal{T}), \tag{4}$$

with $\{\epsilon\}$ an empty tree, $\bigcirc$ a node, $\times$ the cartesian product, and $\varphi(\mathcal{T})$ the substituted structure (see e. g. [72]).

0.1.2 Increasing trees

Increasing trees are labelled trees where the nodes of a tree of size n are labelled by distinct integers of the set $\{1,\dots,n\}$ in such a way that each sequence of labels along any branch starting at the root is increasing. As the underlying tree model we use the simply generated trees but, additionally, they are equipped with increasing labellings. We will thus speak about *simple families of increasing trees.* A thorough study of families (= varieties) of increasing trees was conducted in [6].
A class $\mathcal{T}$ of a simple family of increasing trees can thus be defined in analogy to the definition of simply generated tree families in the following way. A sequence of non-negative numbers $(\varphi_k)_{k\ge 0}$ with $\varphi_0 > 0$ is used to define the weight $w(T)$ of any ordered tree T by $w(T) = \prod_v \varphi_{d(v)}$, where v ranges over all vertices of T and $d(v)$ is the out-degree of v (again, we always assume that there exists a $k \ge 2$ with $\varphi_k > 0$). Furthermore, $\mathcal{L}(T)$ denotes the set of different increasing labellings of the tree T with distinct integers $\{1,2,\dots,|T|\}$, where $|T|$ denotes the size of tree T, and $L(T) := \big|\mathcal{L}(T)\big|$ denotes its cardinality. Then the family $\mathcal{T}$ consists of all trees T together with their weights $w(T)$ and the set of increasing labellings $\mathcal{L}(T)$.
For a given degree-weight sequence $(\varphi_k)_{k\ge 0}$ with a degree-weight generating function $\varphi(t) := \sum_{k\ge 0}\varphi_k t^k$, we define now the total weights by $T_n := \sum_{|T|=n} w(T)\cdot L(T)$. It follows then that the exponential generating function $T(z) := \sum_{n\ge 1} T_n \frac{z^n}{n!}$ satisfies the *autonomous* first order differential equation

$$T'(z) = \varphi\big(T(z)\big), \quad T(0) = 0. \tag{5}$$

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

$$\mathcal{T} = ① \times \Big(\varphi_0 \cdot \{\epsilon\} \,\dot\cup\, \varphi_1 \cdot \mathcal{T} \,\dot\cup\, \varphi_2 \cdot \mathcal{T} * \mathcal{T} \,\dot\cup\, \varphi_3 \cdot \mathcal{T} * \mathcal{T} * \mathcal{T} \,\dot\cup\, \cdots\Big) = ① \times \varphi(\mathcal{T}), \tag{6}$$

where additionally $*$ denotes the partition product for labelled objects.
Three specific increasing tree families are of particular interest:

• *Recursive trees* are the family of non-plane increasing trees such that all node degrees are allowed. The degree-weight generating function is $\varphi(t) = \exp(t)$. Solving (5) gives $T(z) = \log\big(\frac{1}{1-z}\big)$ and thus $T_n = (n-1)!$, for $n \ge 1$. Recursive trees have been introduced as simple probability models in several areas. They are used to model the spread of epidemics [55], to aid in the construction of the family trees of preserved copies of ancient manuscripts [61] or to model chain letter and pyramid schemes [28]. Further they are used to model the stochastic growth of networks [11]. See also [52] for a survey of applications and results on random recursive trees.

• *Plane oriented recursive trees* (also called Heap ordered trees) are the family of plane increasing trees such that all node degrees are allowed. The degree-weight generating function is $\varphi(t) = \frac{1}{1-t}$. Equation (5) leads here to $T(z) = 1-\sqrt{1-2z}$ and thus to $T_n = \frac{(n-1)!}{2^{n-1}}\binom{2n-2}{n-1} = (2n-3)!!$, for $n \ge 1$. Plane oriented recursive trees (to be more exact a slight variation of them) are used to model the growth of the internet. See also [52] for a survey on plane oriented recursive trees.

• *Binary increasing trees* (also called tournament trees) have the degree-weight generating function $\varphi(t) = (1+t)^2$. This model is of special importance, since it is isomorphic to the model of *binary search trees* (see [6] and the references therein for binary increasing trees and e. g. [50] for binary search trees). Thus it must follow $T(z) = \frac{z}{1-z}$ and $T_n = n!$, for $n \geq 1$.

Driven from the inspection that all these important increasing tree families satisfy the equation $\frac{T_{n+1}}{T_n} = c_1 n + c_2$, with fixed constants c_1, c_2, for all $n \geq 1$, we will consider such trees in more detail. It turns out from the characterization given below that the defining degree-weight generating functions $\varphi(t)$ are the same as obtained in [63].

We will give now an exact answer to the question, which degree-weight generating functions are actually fulfilling $\frac{T_{n+1}}{T_n} = c_1 n + c_2$.

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

$$\frac{T_{n+1}}{T_n} = c_1 n + c_2, \tag{7}$$

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

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

In applications the subclass of simple families of increasing trees, which can be constructed via an *insertion process* or a *probabilistic growth rule*, is of particular interest. Such tree families $\mathcal{T}$ have the property that for every tree T' of size n with vertices $v_1, \dots, v_n$ there exist probabilities $p_{T'}(v_1), \dots, p_{T'}(v_n)$, such that when starting with a random tree T' of size n, choosing a vertex v_i in T' according to the probabilities $p_{T'}(v_i)$ and attaching node $n+1$ to it, we obtain a *random* increasing tree T of the family $\mathcal{T}$ of size $n+1$. It is well known that the tree families mentioned above, i. e. recursive trees, plane-oriented recursive trees and binary increasing trees, can be constructed via an insertion process. In [68] a full characterization of those simple families of increasing trees, which can be constructed by an insertion process, is given.

Lemma 2 (Panholzer & Prodinger, 2005)**.** *The following three properties of a simple family of increasing trees $\mathcal{T}$ are equivalent:*

1. *The total weights T_n of trees of size n of $\mathcal{T}$ satisfy the equation*

$$\frac{T_{n+1}}{T_n} = c_1 n + c_2, \tag{8}$$

   with fixed constants c_1, c_2, for all $n \in N$.

2. *Starting with a random increasing tree T of size $n \geq j$ of $\mathcal{T}$ and removing all nodes larger than j we obtain a random increasing tree T' of size j of $\mathcal{T}$.*

3. *The family $\mathcal{T}$ can be constructed via an insertion process respectively a probabilistic growth rule.*

Thus the tree families of interest are described by their degree weight generating function as given in Lemma 1. We will call the tree families covered by Lemma 2 throughout this work *grown simple families of increasing trees*.

Solving either the differential equation (5) or using (1) one obtains the following explicit formulæ for the exponential generating function $T(z)$:

$$T(z) = \begin{cases} \frac{\varphi_0}{c_1} \log\left(\frac{1}{1-c_1 z}\right), & \text{Case A,} \\ \frac{\varphi_0}{c_2}\left(\frac{1}{(1-(d-1)c_2 z)^{\frac{1}{d-1}}} - 1\right), & \text{Case B,} \\ \frac{\varphi_0}{c_2}\left(\frac{1}{(1-c_1 z)^{\frac{c_2}{c_1}}} - 1\right), & \text{Case C.} \end{cases} \tag{9}$$

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

$$T_n = \varphi_0 c_1^{n-1}(n-1)!\binom{n-1+\frac{c_2}{c_1}}{n-1}. \tag{10}$$

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 [68]:

- Step 1: The process starts with the root labelled by 1.
- Step $i+1$: At step $i+1$ the node with label $i+1$ is attached to any previous node v (with out-degree $d^+(v)$) of the already grown tree T of size i with probabilities proportional to the weight $\omega(d^+(v))$

  $$\omega(d^+(v)) = \frac{(d^+(v)+1)\varphi_{d^+(v)+1}}{\varphi_{d^+(v)}},$$

  hence the probability of attaching the new node to node v is given be the weight of $d^+(v)$ divided by the total weight of T.

  $$p(v) = \frac{\omega(d^+(v))}{\sum_{u\in T}\omega(d^+(u))}. \tag{11}$$

  I. e.

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

Thus we see that from a probabilistic point of view one could completely reduce the considerations for Case A and Case B to recursive trees ($\varphi_0 = c_1 = 1$) and d-ary trees ($\varphi_0 = c_2 = 1$, $c_1 = d-1$). For Case C we observe that plane-oriented recursive trees are contained due to $\varphi_0 = 1$, $c_1 = 2$, and $c_2 = -1$ (leading to $\alpha = 1$), but we have the possibility of choosing an arbitrary $\alpha > 0$, such that this case can indeed be seen as a generalization of plane-oriented recursive trees.

0.2 Probabilistic tools

We will briefly introduce the main probabilistic tools used in this thesis. The part concerning with the method of moments is based on an article of Hwang and Neininger [34].

0.2.1 Lévy's continuity theorem

Let X be a random variable with characteristic function $\phi(t)$. Assume that X has a continuous distribution function. Further let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables with characteristic functions $\phi_n(t)$, $n \in \mathbb{N}$.

$$X_n \xrightarrow{(d)} X \quad \Longleftrightarrow \quad \lim_{n\to\infty} \phi_n(t) = \phi(t) \quad \text{for all } t \in \mathbb{R}. \tag{12}$$

The corresponding theorem for the moment generating function was proven by Curtiss in [14]. It is sometimes also called Lévy's continuity theorem due to the similarity of the results.

0.2.2 Curtiss' theorem

Let $M_n(t)$ be the moment generating function of a distribution function $F_n(x)$ such that for each n, $M_n(t)$ exists for $|t| < t_1$. Suppose that there is a real function $M(t)$ such that $\lim_{n\to\infty} M_n(t) = M(t)$ for $|t| \le t_2 < t_1$ with $t_2 > 0$. Then there is a distribution function $F(x)$ such that $\lim_{n\to\infty} F_n(x) = F(x)$ at each continuity point of F and the moment generating function of F is $M(t)$ for $|t| < t_2$.

0.2.3 Method of moments

The method of moments is one of the most classical ways of deriving limit distributions. It has been widely applied to problems in diverse fields. It consists in first computing the mean and variance. After properly scaling the random variable, the higher moments of the scaled random variable X are computed by induction. Carleman's criterion provides that the moment sequence $(\mathbb{E}(X^s))_s$ uniquely characterizes a distribution if $\sum_s \mathbb{E}(X^{2s})^{-1/2s} = \infty$. Then one can obtain the convergence in distribution and of all moments (or convergence in L_p for all $p > 0$) by the Fréchet-Shohat moment convergence theorem (see Loève [49]).

0.2.4 Poisson Approximation

In Chapter 6 we are able to use Poisson Approximation technics to obtain limit laws for the considered random variables by following closely the approach of Dobrow and Smythe [18], which is based on results in [4]. The total variation distance d_{TV} of two probability measures P and Q over $\mathbb{Z}_+$ is defined by

$$d_{\mathrm{TV}}(P, Q) = \frac{1}{2} \sum_{k\ge 0} |P(\{k\}) - Q(\{k\})|. \tag{13}$$

We denote with $\mathrm{Po}(\lambda)$ a probability distribution of a Poisson distributed random variable with parameter λ. Further we use the notation $\mathcal{L}(X)$ for the distribution law of the r.v. X. Let X_n be a sequence of random variables for which

$$d_{\mathrm{TV}}(\mathcal{L}(X_n), \mathrm{Po}(\lambda_n)) \to 0 \quad \text{and} \quad \lambda_n \to \infty, \tag{14}$$

then it holds

$$\frac{X_n - \lambda_n}{\sqrt{\lambda_n}} \xrightarrow{(d)} \mathcal{N}(0, 1). \tag{15}$$

Part I

Analysis of node isolation procedures in rooted trees

Chapter 1

Isolating a single node in recursive trees

1.1 Introduction

In [55] Meir and Moon considered the following edge-removal procedure (= cutting-down procedure) for a rooted tree with n vertices. Pick one of the $n-1$ edges of the tree at random and remove it. This separates the tree into a pair of rooted trees; the tree containing the root of the original tree retains its root, while the tree not containing the root of the original tree is rooted at the vertex adjacent to the edge that was cut. Now the subtree that does not contain the original root is discarded and the procedure is continued recursively for the remaining subtree until the original root is isolated. In paper [55] the random variable X_n is studied, which counts the number of edges that will be removed from a randomly chosen recursive tree of size n by above edge-removal procedure until the root, i. e., the node labelled by 1, is isolated. This problem was studied in the context of the spread of contamination in an organism, where it is assumed that the first node is the source of all the contamination. By separating nodes from the source by successively removing edges one eventually isolates the source node. It was shown in [55] the following asymptotic equivalent of the expectation: $\mathbb{E}(X_n) \sim \frac{n}{\log n}$. Thus on average $\sim \frac{n}{\log n}$ random edges have to be removed from a random size-n recursive tree before the root node is isolated.

Recently Javanian and Vahidi-Asl [38] have studied a modification of above edge-removal procedure, motivated by considerations concerning the hierarchy of a workforce of a company: at each stage after removing a random edge, the subtree containing the node with the largest label, i. e., label n, is kept and the other subtree is discarded. Thus finally the node labelled by n will be isolated. Again one is interested in a study of the random variable Y_n, which counts the number of edges that will be removed from a randomly chosen recursive tree of size n until node n, i. e., the last recent entry, is isolated. It was shown in [38] the following asymptotic equivalent of the expectation: $\mathbb{E}(Y_n) \sim \frac{n}{2 \log n}$. Therefore on average $\sim \frac{n}{2 \log n}$ random edges have to be removed from a random size-n recursive tree before node n is isolated.

But isolating node 1 and isolating node n in a tree by removing random edges can be considered as special instances of a natural generalization of the edge-removal procedures described above. In order to isolate via random cuttings the node with a specified label λ, with $1 \le \lambda \le n$, in a tree T with nodes labelled by $1, 2, \dots, n$ we consider the following procedure:

1. Pick one of the $n-1$ edges of the tree at random and remove it. This separates the tree T into two subtrees $\hat{T}$ and $\tilde{T}$. Let us assume that $\lambda \in \hat{T}$.
2. Continue the edge-removal procedure recursively for the subtree $\hat{T}$, which contains the node labelled by λ, until node λ is isolated.

We are going to study this general edge-removal procedure by analyzing the random variable $X_{n,\lambda}$, with $1 \le \lambda \le n$, which counts the number of edge-cuts that are necessary to isolate the node labelled by λ in a random recursive tree of size n. Of course, the margin cases X_n (isolating the root) and Y_n (isolating the largest node) are contained as the special instances $X_{n,1}$ and $X_{n,n}$, respectively. This general edge-removal procedure has a natural interpretation in the model for the spread of a contamination in an organism mentioned above: instead of assuming that the root node is the contamination source we assume that a certain node λ is the contamination source, which one wants to isolate. Figure 1.1 and Figure 1.2 give examples of isolating certain nodes via the edge-removal procedure considered here.

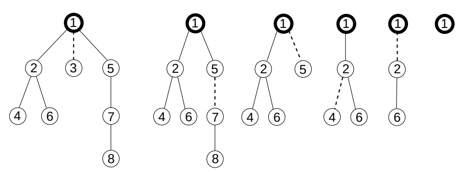

Figure 1.1: Isolating the root in a size-8 recursive tree with 5 cuts.

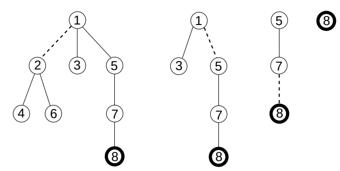

Figure 1.2: Isolating the largest node in a size-8 recursive tree with 3 cuts.

We will analyze the random variable $X_{n,\lambda}$ from "both ends", i. e., we are studying $X_{n,\lambda}$ for small labels: $\lambda = l$, with $l \geq 1$ fixed and $n \to \infty$, and for large labels: $\lambda = n+1-l$, with $l \geq 1$ fixed and $n \to \infty$. By using a recursive approach we are able to give asymptotic expansions of the moments of the random variables $X_{n,l}$ and $X_{n,n+1-l}$, for l fixed and $n \to \infty$. For the instance of large labels we can apply the Theorem of Fréchet and Shohat and characterize the limit law of the normalized random variable $\frac{\log n}{n} X_{n,n+1-l}$ by its moments. It turns out that the random variable $X_{n,n+1-l}$, which counts the number of edge cuts necessary for isolating the l-th largest node, is (after scaling with $\frac{\log n}{n}$) asymptotically, for l fixed and $n \to \infty$, uniformly distributed on $[0,1)$. For the instance of small labels we can show that $\frac{\log n}{n} X_{n,l}$ converges, for l fixed and $n \to \infty$, in probability to 1, but it turns out that a zero-mean and unit-variance normalization of $X_{n,l}$ has (for $s \geq 2$) s-th moments of order $\log^{\frac{s}{2}-1} n$. Thus existence of the limit law (and in the affirmative case a characterization of the limit law) of this normalized random variable cannot be shown by the method of moments. This was already observed for the special instance of isolating the root, i. e., for $X_{n,1}$, in [65].

Remarks:
(i) We want to remark that for the problem of isolating the root node of a tree via random cuttings Janson [37] gave an alternative approach by establishing a very useful connection between the number of cuts to isolate the root and the number of records when assigning random values to the edges of the tree. We want to sketch in the following that one can extend the arguments used in [37] to give also a connection between the number of cuts and the number of records for the problem of isolating a specified label λ. We consider a randomly chosen recursive tree of size n and attach to each edge e a random value γ_e, where we assume that the values γ_e are i. i. d. with an arbitrary continuous distribution. For a given label λ, with $1 \leq \lambda \leq n$, we call a value γ_e a record if it is the largest value in the path from the node labelled by λ to the edge e. Then it holds that the number of records is again distributed as $X_{n,\lambda}$.
(ii) Furthermore we want to remark that the cutting-down procedure for isolating the root node of a random recursive tree has also been used to give an alternative representation of the so called Bothausen-Sznitman coalescent (see [29]).

1.2 Results and outline of the proof

1.2.1 Results

For the s-th moments of the random variables $X_{n,l}$ and $X_{n,n+1-l}$ we get the following asymptotic expansions, for $l \geq 2$ fixed and $n \to \infty$, stated as Theorem 1 and Theorem 2.

Theorem 1. *The s-th moments $\mathbb{E}(X_{n,l}^s)$ of the number of random cuts necessary to isolate node l in a* *random recursive tree of size n are, for $l, s \geq 1$ fixed and $n \to \infty$, asymptotically given by*

$$\mathbb{E}\big(X_{n,l}^s\big) = \frac{n^s}{\log^s n} + \frac{\gamma_{l,s}\, n^s}{\log^{s+1} n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+2} n}\Big), \tag{1.1}$$

where the constants $\gamma_{l,s}$ appearing in above expansion are given by

$$\gamma_{l,s} = (s+1)H_s - sH_{l+s-1} - \frac{1}{2}\big(H_{l-1}^2 - H_{l-1}^{(2)}\big) - H_s H_{l-1} + s\Psi(l+s) + \sum_{k=1}^{l-1} \frac{H_{s+k}}{k} - \sum_{k=1}^{l-1} \frac{1}{k^2\binom{s+k}{k}}.$$

Theorem 2. *The s-th moments $\mathbb{E}\big(X_{n,n+1-l}^s\big)$ of the number of random cuts necessary to isolate node $n+1-l$ in a random recursive tree of size n are, for $l, s \geq 1$ and $n \to \infty$, asymptotically given by*

$$\mathbb{E}\big(X_{n,n+1-l}^s\big) = \frac{n^s}{(s+1)\log^s n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+1} n}\Big). \tag{1.2}$$

From these asymptotic expansions of the s-th moments we obtain the following results for the limiting behavior of $X_{n,l}$ and $X_{n,n+1-l}$, for l fixed and $n \to \infty$, given as Corollary 1 and Theorem 3.

Corollar 1. *The s-th centered moments $\mathbb{E}\Big(\big[X_{n,l} - \mathbb{E}(X_{n,l})\big]^s\Big)$ of the number of random cuts necessary to isolate node l in a random recursive tree of size n are, for $l \geq 1$, $s \geq 2$ fixed and $n \to \infty$, asymptotically given by*

$$\mathbb{E}\Big(\big[X_{n,l} - \mathbb{E}(X_{n,l})\big]^s\Big) = \frac{\delta_{l,s}\, n^s}{\log^{s+1} n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+2} n}\Big), \tag{1.3}$$

where the constants $\delta_{l,s}$ appearing in above expansion are given by

$$\delta_{l,s} = \frac{(-1)^s}{s}\big(H_{l+s-1} - H_s\big) + \frac{(-1)^s}{s^2(s-1)} + \frac{(-1)^s}{s^2\binom{l+s-1}{l-1}}.$$

Thus the scaled random variable $\frac{\log n}{n} X_{n,l}$ converges, for $l \geq 1$ fixed and $n \to \infty$, in probability to 1 with convergence of all moments.

Theorem 3. *The limiting distribution of the normalized random variable $\frac{\log n}{n} X_{n,n+1-l}$ is, for $l \geq 1$ fixed and $n \to \infty$, a standard uniform distribution U_1 with support $[0, 1)$:*

$$\frac{\log n}{n} X_{n,n+1-l} \xrightarrow{(d)} X, \quad X \stackrel{(d)}{=} U_1. \tag{1.4}$$

1.2.2 Outline of the proof

In order to show our results we will basically use a recursive approach, which allows to describe the number of random cuts necessary to isolate label l in a random recursive tree of size n via the corresponding quantities for smaller tree sizes $k < n$ and labels r not larger than l, i. e., $r \leq l$. Such a recursive approach is amenable, since it is well known (see [29; 65]) that random recursive trees satisfy a certain randomness-preservation property, which is stated in Subsection 1.3.1. Using this property we can easily give a distribution recurrence for $X_{n,l}$ (and X_{n+1-l}), where the behavior of the random variables considered are determined by the splitting probabilities $p_{(n,l),(k,r)}$ (and $\mathfrak{p}_{(n,l),(k,r)}$), which give the probability that when starting with a random size-n recursive tree and removing a random edge the subtree containing node l (node $n+1-l$) is of size k and where furthermore node l (node $n+1-l$) is the r-th smallest (the r-th largest) node in this subtree. Using a bijective argument we can give exact formulæ for these splitting probabilities. They are computed in Subsection 1.3.2 and given as Lemma 3.

From the distribution recurrences for $X_{n,l}$ and X_{n+1-l} we easily obtain recurrences for the s-th moments $\mathbb{E}\big(X_{n,l}^s\big)$ and $\mathbb{E}\big(X_{n,n+1-l}^s\big)$. In order to treat these recurrences we use a generating functions approach, which allows to translate these recurrences into linear differential equations for suitably introduced generating functions $M_{l,s}(z)$ and $N_{l,s}(z)$. Since we are able to determine the general solutions of the corresponding homogeneous differential equations it is possible to describe the solutions of these differential equations rather "explicitly". To determine the asymptotic growth behavior of $\mathbb{E}(X_{n,l}^s)$ and $\mathbb{E}(X_{n,n+1-l}^s)$, and thus essentially of the coefficients of the generating functions $M_{l,s}(z)$ and $N_{l,s}(z)$, we use singularity analysis (see [25]), i. e., we study the growth behavior of the functions in a neighborhood of the dominant singularity, together with certain lemmata for singular differentiation and integration; the corresponding Lemma 4 is stated in Subsection 1.4.1. Since, for a given pair (l, s), all generating functions $M_{r,j}(z)$ (and $N_{r,j}(z)$), with $r \leq l$, $j \leq s$ and $(r, j) \neq (l, s)$, appear in the inhomogeneous part of the differential equations determining $M_{l,s}(z)$ (and $N_{l,s}(z)$), we are forced to "pump out" the asymptotic expansions of the generating functions $M_{l,s}(z)$ (and $N_{l,s}(z)$) around the dominant singularity via induction on both parameters l and s. The corresponding computations for small labels (labels $l \geq 1$ fixed) are carried out in Section 1.4, whereas the computations for large labels (labels $n + 1 - l$, with $l \geq 1$ fixed) are given in Section 1.5.

1.3 The recursive approach

1.3.1 Recurrences

As already mentioned in Lemma 2 it has been observed that random recursive trees satisfy the following "randomness-preservation" property, which will allow a recursive approach for the analysis of the parameter considered.

> Choose a random recursive tree of size n and then one of its $n-1$ edges uniformly at random. Cutting this edge produces a pair of trees of size k and $n-k$. Then, after an order preserving relabelling of the subtrees with labels $\{1, \dots, k\}$ and $\{1, \dots, n-k\}$, the subtrees themselves are random recursive trees of size k and $n - k$.

An important step for the recursive description of the probabilities $\mathbb{P}\{X_{n,l} = m\}$ is to introduce the splitting probabilities $p_{(n,l),(k,r)}$: they give the probability that when starting with a random size-n recursive tree and removing a random edge the subtree containing node l is of size k and where furthermore node l is the r-th smallest node in this subtree.

When we treat the analogous problem of isolating the node $n + 1 - l$ it is convenient to introduce the splitting probabilities $\mathfrak{p}_{(n,l),(k,r)}$: they give the probability that when starting with a random size-n recursive tree and removing a random edge the subtree containing node $n + 1 - l$ is of size k and where furthermore node $n + 1 - l$ is the r-th largest node in this subtree. Of course, these quantities are connected via the trivial relation

$$p_{(n,l),(k,r)} = \mathfrak{p}_{(n,n+1-l),(k,k+1-r)}. \tag{1.5}$$

From the recursive nature of the problem together with the randomness-preservation property immediately follows the distribution recurrence for the number of random cuts necessary to isolate the l-th smallest node in a random recursive tree of size n given below.

$$\mathbb{P}\{X_{n,l} = m\} = \sum_{r=1}^{l} \sum_{k=r}^{n-1} p_{(n,l),(k,r)} \mathbb{P}\{X_{k,r} = m - 1\}, \quad n \geq 2, \tag{1.6}$$

with initial value $\mathbb{P}\{X_{1,1} = 0\} = 1$. Furthermore, the distribution recurrence for the number of random cuts necessary to isolate the l-th largest node, i. e., node $n + 1 - l$, in a random recursive tree of size n is given by:

$$\mathbb{P}\{X_{n,n+1-l} = m\} = \sum_{r=1}^{l} \sum_{k=r}^{n-1} \mathfrak{p}_{(n,l),(k,r)} \mathbb{P}\{X_{k,k+1-r} = m - 1\}, \quad n \geq 2, \tag{1.7}$$

with initial value $\mathbb{P}\{X_{1,1} = 0\} = 1$.
The splitting probabilities $p_{(n,l),(k,r)}$ appearing in (1.6) are given by Lemma 3, which also determine the splitting probabilities $\mathfrak{p}_{(n,l),(k,r)}$ appearing in (1.7) due to equation (1.5).

1.3.2 The splitting probabilities

We obtain the following explicit formulæ for the splitting probabilities $p_{(n,l),(k,r)}$ appearing in (1.6).

Lemma 3. *The splitting probabilities $p_{(n,l),(k,r)}$ are, for $1 \le l \le n$, $1 \le r \le k$, $1 \le k \le n-1$ and $n \ge 2$, given as follows:*

$$p_{(n,l),(k,r)} = \begin{cases} \left[(l-1)\binom{n-l}{n-k} + \binom{n-l+1}{n-k+1}\right] \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!}, & r = l, \\ \left[\binom{l-1}{r}\binom{n-l}{k-r} + \binom{l-1}{r-2}\binom{n-l}{k-r}\right] \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!}, & r < l. \end{cases}$$

Proof. If we remove an edge e of a size-n recursive tree we split the tree into two subtrees: we denote with T' the subtree containing the original root, i. e., label 1, and with T'' the other subtree, which is rooted at the vertex adjacent to the edge e that was cut. After an order preserving relabelling with labels $\{1, \dots, |T'|\}$ and $\{1, \dots, |T''|\}$ both subtrees can be considered as recursive trees. Furthermore we denote with B the arising subtree, which contains the node labelled by l in the original tree; we assume that this subtree has size k, with $1 \le k \le n-1$. We distinguish now the cases $r = l$ and $r < l$.
If $r = l$ then it follows that $B = T'$. We want to determine the number of possibilities of removing an edge e of a recursive tree of size n leading (after an order preserving relabelling) to the pair (T', T'') of subtrees. To do this we count the number of different ways of distributing the labels $\{1, \dots, n\}$ order preserving to T' and T'' and adjoining the root of T'' to a node of T' (by inserting edge e), such that the resulting tree is a recursive tree. We consider now the node of T' incident with e: if the node of T' incident with e has label j, with $1 \le j \le k$, then it follows that the labels of T'' must all be larger than j. For $1 \le j \le l$ we can choose $n-k$ of the labels $l+1, l+2, \dots, n$ and distribute them order preserving to T'', whereas the remaining labels are distributed order preserving to T', leading to $\binom{n-l}{n-k}$ possibilities. For $l+1 \le j \le k$ we can choose $n-k$ of the labels $j+1, j+2, \dots, n$ and distribute them order preserving to T'', whereas the remaining labels are distributed order preserving to T', leading to $\binom{n-j}{n-k}$ possibilities. Thus this quantity is independent of the actual choice of T' with $|T'| = k$ and T'' with $|T''| = n-k$. Since there are $T_k = (k-1)!$ and $T_{n-k} = (n-k-1)!$ different recursive trees of size k and $n-k$, this leads together with the fact that there are $n-1$ ways of selecting an edge e for any of the $T_n = (n-1)!$ recursive trees of size n to the following formula:

$$\begin{aligned} p_{(n,l),(k,l)} &= \left[l\binom{n-l}{n-k} + \sum_{j=l+1}^{k}\binom{n-j}{n-k}\right] \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!} \\ &= \left[(l-1)\binom{n-l}{n-k} + \binom{n-l+1}{n-k+1}\right] \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!}, \end{aligned}$$

appealing to a well known identity.
If $r < l$ we have to distinguish further between the two cases $B = T'$ and $B = T''$. If $B = T'$ and we distribute the labels $\{1, \dots, n\}$ order preserving to T' and T'' we have the restriction that exactly $l-r$ nodes of the nodes $2, \dots, l-1$ have to be in T''. If $B = T''$ then we have the restriction that exactly $r-1$ nodes of the nodes $2, \dots, l-1$ have to be in T''. Proceeding the same way as before we obtain eventually the following formula.

$$\begin{aligned} p_{(n,l),(k,r)} &= \left[\binom{n-l}{n-k-(l-r)}\sum_{j=1}^{r-1}\binom{l-1-j}{l-r} + \binom{n-l}{k-r}\sum_{j=1}^{l-r}\binom{l-1-j}{r-1}\right] \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!} \\ &= \left[\binom{l-1}{r-2}\binom{n-l}{k-r} + \binom{l-1}{r}\binom{n-l}{k-r}\right] \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!} \end{aligned}$$

□

1.4 Isolating nodes with small labels

1.4.1 Singular differentiation and integration

In order to treat the recurrences for the s-th moments of $X_{n,l}$ and $X_{n,n+1-l}$ that will be obtained in the sequel we use a generating functions approach, which leads "in principle" to exact formulæ for suitably introduced generating functions. To obtain asymptotic information for the s-th moments we will basically use singularity analysis of generating functions, i. e., the transfer lemmata of Flajolet and Odlyzko [25] to "translate" the asymptotic growth behavior of a generating function in the neighborhood of its dominant singularity into the growth behavior of its coefficients. For the functions studied in here the unique dominant singularity is always located at $z = 1$, thus we will specialize the considerations given below to this case. In order to apply singularity analysis it is necessary that the functions involved are analytic for a domain larger than the circle of convergence, namely the functions have to be analytic for indented discs $\Delta := \Delta(\phi, \eta) = \{z : \ |z| < 1 + \eta, \ |\operatorname{Arg}(z-1)| > \phi\}$, with $\eta > 0$, $0 < \phi < \frac{\pi}{2}$. Such functions are called Δ-regular (see [22]). We want to point out that the functions considered here are always Δ-regular, since they are generated from Δ-regular functions via basic arithmetical functions and the operations differentiation and integration.

We will require the following O-transfer lemma for a Δ-regular function with a certain growth estimate in a neighborhood of $z = 1$:

$$f(z) = \mathcal{O}\Big(\frac{1}{(1-z)^a \log^b\big(\frac{1}{1-z}\big)}\Big), \ \text{ for } \ z \to 1 \quad \Longrightarrow \quad [z^n]f(z) = \mathcal{O}\Big(\frac{n^{a-1}}{\log^b n}\Big). \tag{1.8}$$

together with an asymptotic expansion of the coefficients of the following functions:

$$f(z) = \frac{1}{(1-z)^a \log^b\big(\frac{1}{1-z}\big)} \quad \Longrightarrow \quad [z^n]f(z) = \frac{n^{a-1}}{\Gamma(a)\log^b n}\Big(1 + \frac{b\Psi(a)}{\log n} + \mathcal{O}\big(\frac{1}{\log^2 n}\big)\Big), \tag{1.9}$$

where both formulæ (1.8) and (1.9) hold (at least) for $a > 0$ and $b \geq 0$ (see [25]).

However, for a study of the functions appearing we also require lemmata, which describe the asymptotic behavior of the derivative $f'(z)$ and the antiderivative $\int_0^z f(t)dt$ of a Δ-regular function $f(z)$ in the neighborhood of the dominant singularity $z = 1$, supposing that the asymptotic behavior around $z = 1$ of the function $f(z)$ itself is of a certain kind. Such theorems are known as theorems for singular differentiation and integration and can be found, e. g., in [22]. But in the sequel we will require slightly more general theorems than given there, which are stated in Lemma 4. The proof of this lemma is omitted, since one can essentially "repeat" the arguments used in the proof of the corresponding theorems given in [22].

Lemma 4 (Singular differentiation and integration)**.** *Let $f(z)$ be a Δ-regular function (see [22]), an analytic function in the domain $\Delta := \Delta(\phi, \eta)$,*

$$\Delta(\phi, \eta) = \{z : \ |z| < 1 + \eta, \ |\operatorname{Arg}(z-1)| > \phi\},$$

with $\eta > 0$, $0 < \phi < \frac{\pi}{2}$, satisfying, for $z \to 1$, the expansion

$$f(z) = \mathcal{O}\Big(\frac{1}{(1-z)^a \log^b\big(\frac{1}{1-z}\big)}\Big),$$

for $a > 1$ and $b \geq 1$. Then $\int_0^z f(t)dt$ and $f'(z)$ are also Δ-regular and they admit, for $z \to 1$, the *expansions*

$$\int_0^z f(t)dt = \mathcal{O}\Big(\frac{1}{(1-z)^{a-1}\log^b\big(\frac{1}{1-z}\big)}\Big), \quad \textit{and} \quad f'(z) = \mathcal{O}\Big(\frac{1}{(1-z)^{a+1}\log^b\big(\frac{1}{1-z}\big)}\Big).$$

1.4.2 Expectations

The first step in our proof of Theorem 1 is to show the special case $s = 1$, i. e., asymptotic expansions of the expectations $\mathbb{E}(X_{n,l})$, for l fixed and $n \to \infty$, given as Lemma 5.

Lemma 5. *The expectations $\mathbb{E}(X_{n,l})$ of the number of random cuts necessary to isolate node l in a random recursive tree of size n are, for $l \geq 1$ fixed and $n \to \infty$, asymptotically given by*

$$\mathbb{E}(X_{n,l}) = \frac{n}{\log n} + \big(4 - 2H_l - \frac{1}{l} + \Psi(l+1)\big)\frac{n}{\log^2 n} + \mathcal{O}\Big(\frac{n}{\log^3 n}\Big).$$

The proof of this lemma will be carried out by a generating functions approach using induction on l. The recurrences for the expectations $\mathbb{E}(X_{n,l})$ are obtained easily from the distribution recurrence (1.6) and are given by

$$\mathbb{E}(X_{n,l}) = 1 + \sum_{r=1}^{l}\sum_{k=r}^{n-1} p_{(n,l),(k,r)}\mathbb{E}(X_{k,r}), \tag{1.10}$$

with splitting probabilities given by Lemma 3.
We introduce for $l \geq 1$ the generating functions

$$M_l(z) := \sum_{n\geq l}(n-1)^{\underline{l-1}}\mathbb{E}(X_{n,l})z^{n-l}, \tag{1.11}$$

which allow to translate recurrence (1.10) into the following first order linear differential equation for $M_l(z)$, where the functions $M_r(z)$, with $r < l$, are appearing in the inhomogeneous part $R_l(z)$:

$$(1-z)\log\big(\frac{1}{1-z}\big)\frac{d}{dz}M_l(z) + \Big((l-1) - l\log\big(\frac{1}{1-z}\big)\Big)M_l(z) = R_l(z), \tag{1.12}$$

with inhomogeneous part

$$R_l(z) = \frac{(l-1)!(l-1+z)}{(1-z)^{l+1}} + \sum_{r=1}^{l-1}\left[\binom{l-1}{r} + \binom{l-1}{r-2}\right]\frac{(l-r-1)!}{(1-z)^{l-r}}M_r(z),$$

and initial condition $M_l(0) = (l-1)!\,\mathbb{E}(X_{l,l})$.
The homogeneous differential equation corresponding to (1.12) has the following general solution with C being an arbitrary constant:

$$M_l^{[h]}(z) = \frac{C}{(1-z)^l\log^{l-1}\big(\frac{1}{1-z}\big)}.$$

The method of variation of constants leads then to the following particular solution of (1.12):

$$M_l^{[p]}(z) = \frac{1}{(1-z)^l\log^{l-1}\big(\frac{1}{1-z}\big)}\int_0^z (1-t)^{l-1}\log^{l-2}\big(\frac{1}{1-t}\big)R_l(t)dt, \tag{1.13}$$

and it can be shown that this particular solution matches with the initial condition and is thus the wanted function, so $M_l(z) = M_l^{[p]}(z)$.
It will suffice to show the following asymptotic expansion of $M_l(z)$ around the dominant singularity $z = 1$,

since a direct application of the transfer lemmata (1.8) and (1.9) leads then to Lemma 5:

$$M_l(z) = \frac{l!}{(1-z)^{l+1}\log\left(\frac{1}{1-z}\right)} + \frac{(l-1)!(4l-1-2lH_l)}{(1-z)^{l+1}\log^2\left(\frac{1}{1-z}\right)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+1}\log^3\left(\frac{1}{1-z}\right)}\Big). \tag{1.14}$$

The proof of the expansion (1.14) will be done by induction. The case $l = 1$ gives the following solution, which already appeared in [55]:

$$M_1(z) = \frac{1}{1-z}\int_0^z \frac{t}{(1-t)^2\log\left(\frac{1}{1-t}\right)}dt. \tag{1.15}$$

Integration by parts together with an application of Lemma 4 for singular integration leads then from (1.15) to the expansion (1.14) for the instance $l = 1$.
Now we assume that the functions $M_r(z)$ satisfy for all $r < l$ and a given $l > 1$ the asymptotic expansion (1.14). Plugging these expansions into the formula for the remainder term $R_l(z)$ given above easily leads to the expansion

$$R_l(z) = \frac{l!}{(1-z)^{l+1}} + \frac{(l-1)!(l^2+2l-1-2lH_l)}{(1-z)^{l+1}\log\left(\frac{1}{1-z}\right)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+1}\log^2\left(\frac{1}{1-z}\right)}\Big),$$

and furthermore to

$$\int_0^z (1-t)^{l-1}\log^{l-2}\big(\frac{1}{1-t}\big)R_l(t)dt = \frac{l!\log^{l-2}\left(\frac{1}{1-z}\right)}{1-z} + \frac{(l-1)!\big(4l-1-2lH_l\big)\log^{l-3}\left(\frac{1}{1-z}\right)}{1-z} + \mathcal{O}\Big(\frac{\log^{l-4}\left(\frac{1}{1-z}\right)}{1-z}\Big).$$

Due to formula (1.13) for $M_l(z) = M_l^{[p]}(z)$ expansion (1.14) is also shown for l. Thus (1.14) and as a consequence Lemma 5 is shown for all $l \geq 1$.

1.4.3 Higher moments

In order to show Theorem 1 for the asymptotic behavior of the moments $\mathbb{E}(X_{n,l}^s)$ we will continue our generating functions approach, where we will now use double induction on both parameters: the label l considered and the order s of the moments. To obtain a recurrence for the s-th moments of $X_{n,l}$ we multiply the distribution recurrence (1.6) with $m^s = \sum_{j=0}^{s}\binom{s}{j}(m-1)^j$ and sum up for $m \geq 1$. This leads to the following recurrence valid for $1 \leq l \leq n$ and $n \geq 2$ (with splitting probabilities given by Lemma 3).

$$\mathbb{E}\big(X_{n,l}^s\big) = \sum_{j=0}^{s}\binom{s}{j}\sum_{r=1}^{l}\sum_{k=r}^{n-1} p_{(n,l),(k,r)}\mathbb{E}\big(X_{k,r}^j\big). \tag{1.16}$$

We proceed as before and introduce for $l \geq 1$ and $s \geq 1$ the generating functions

$$M_{l,s}(z) := \sum_{n\geq l}(n-1)^{\underline{l-1}}\,\mathbb{E}\big(X_{n,l}^s\big)z^{n-l}. \tag{1.17}$$

Thus it holds $M_{l,1}(z) = M_l(z)$ for the functions $M_l(z)$ introduced in Subsection 1.4.2. Again we can translate above recurrence (1.16) into the following first order differential equation for $M_{l,s}(z)$, where the functions $M_{r,j}(z)$, with $r \leq l$, $j \leq s$ and $(r,j) \neq (l,s)$, appear in the inhomogeneous part $R_{l,s}(z)$:

$$(1-z)\log\big(\frac{1}{1-z}\big)\frac{d}{dz}M_{l,s}(z) + \Big((l-1) - l\log\big(\frac{1}{1-z}\big)\Big)M_{l,s}(z) = R_{l,s}(z), \tag{1.18}$$

with inhomogeneous part

$$R_{l,s}(z) = \sum_{j=1}^{s-1} \binom{s}{j} \Big[\Big(z - (1-z)\log\big(\frac{1}{1-z}\big)\Big) \frac{d}{dz} M_{l,j}(z) + l \log\big(\frac{1}{1-z}\big) M_{l,j}(z) \Big]$$
$$+ \frac{(l-1)!(l-1+z)}{(1-z)^{l+1}} + \sum_{j=1}^{s} \binom{s}{j} \sum_{r=1}^{l-1} \Big[\binom{l-1}{r} + \binom{l-1}{r-2} \Big] \frac{(l-r-1)!}{(1-z)^{l-r}} M_{r,j}(z),$$

and initial condition $M_{l,s}(0) = (l-1)!\,\mathbb{E}\big(X_{l,l}^s\big)$. Since the homogeneous differential equations corresponding to (1.18) and (1.12) coincide, we already know the shape of the general solution of (1.18):

$$M_{l,s}(z) = \frac{C}{(1-z)^l \log^{l-1}\big(\frac{1}{1-z}\big)} + \frac{1}{(1-z)^l \log^{l-1}\big(\frac{1}{1-z}\big)} \int_0^z (1-t)^{l-1} \log^{l-2}\big(\frac{1}{1-t}\big) R_{l,s}(t) dt.$$

It turns out that the particular solution obtained for $C = 0$ matches with the initial condition and we get thus

$$M_{l,s}(z) = \frac{1}{(1-z)^l \log^{l-1}\big(\frac{1}{1-z}\big)} \int_0^z (1-t)^{l-1} \log^{l-2}\big(\frac{1}{1-t}\big) R_{l,s}(t) dt. \tag{1.19}$$

Again it suffices to show the following asymptotic expansion around the dominant singularity $z = 1$ for the generating functions $M_{l,s}(z)$, since basic singularity analysis immediately leads from this expansion and (1.17) to Theorem 1.

$$M_{l,s}(z) = \frac{(l+s-1)!}{(1-z)^{l+s} \log^{s}\big(\frac{1}{1-z}\big)} + \frac{\alpha_{l,s}}{(1-z)^{l+s} \log^{s+1}\big(\frac{1}{1-z}\big)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s} \log^{s+2}\big(\frac{1}{1-z}\big)}\Big), \tag{1.20}$$

with constants

$$\alpha_{l,s} = (l+s-1)!\Big[(s+1)H_s - sH_{l+s-1} - \frac{1}{2}\big(H_{l-1}^2 - H_{l-1}^{(2)}\big) - H_s H_{l-1} + \sum_{k=1}^{l-1} \frac{H_{s+k}}{k} - \sum_{k=1}^{l-1} \frac{1}{k^2 \binom{s+k}{k}}\Big].$$

To show expansion (1.20) for all $l, s \geq 1$ we use induction on both parameters. The case $s = 1$ with arbitrary $l \geq 1$ was already treated in Subsection 1.4.2, where we computed the following expression, which matches with (1.20):

$$M_{l,1}(z) = M_l(z) = \frac{l!}{(1-z)^{l+1} \log\big(\frac{1}{1-z}\big)} + \frac{(l-1)!(4l-1-2lH_l)}{(1-z)^{l+1} \log^2\big(\frac{1}{1-z}\big)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+1} \log^3\big(\frac{1}{1-z}\big)}\Big).$$

Now we assume that for all pairs $(r, j) < (l, s)$, which means for $r \leq l$, $j \leq s$ and $(r, j) \neq (l, s)$, the functions $M_{r,j}(z)$ have in a neighborhood of the dominant singularity $z = 1$ the asymptotic expansion (1.20). We want to show that (1.20) also holds for the pair (l, s), where we may assume $s > 1$, since the case $s = 1$ is already shown. Plugging the expansions of the functions $M_{r,j}(z)$ into the formula for the inhomogeneous part $R_{l,s}(z)$ we obtain the expansion

$$R_{l,s}(z) = \frac{s(l+s-1)!}{(1-z)^{l+s} \log^{s-1}\big(\frac{1}{1-z}\big)} + \frac{\beta_{l,s}}{(1-z)^{l+s} \log^{s}\big(\frac{1}{1-z}\big)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s} \log^{s+1}\big(\frac{1}{1-z}\big)}\Big),$$

with

$$\beta_{l,s} = s(l+s-1)\alpha_{l,s-1} + (l+s-1)!\big[l - 2 - (s+1)(H_{l+s-1} - H_{s+1})\big]$$
$$+ (l-1)!(s-1)!\Big(\binom{l+s-1}{s} - 1\Big) - s(s-1)(l+s-2)!.$$

This further leads by an application of singular integration to the following expansion around $z = 1$:

$$\int_0^z (1-t)^{l-1} \log^{l-2}\big(\frac{1}{1-t}\big) R_{l,s}(t)dt = \frac{(l+s-1)!}{(1-z)^s \log^{s-l+1}\big(\frac{1}{1-z}\big)} + \frac{(l+s-1)!(s-l+1)+\beta_{l,s}}{s(1-z)^s \log^{s-l+2}\big(\frac{1}{1-z}\big)} + \mathcal{O}\Big(\frac{1}{(1-z)^s \log^{s-l+3}\big(\frac{1}{1-z}\big)}\Big).$$

Using equation (1.19) leads for $M_{l,s}(z)$ to the expansion

$$M_{l,s}(z) = \frac{(l+s-1)!}{(1-z)^{l+s} \log^{s}\big(\frac{1}{1-z}\big)} + \frac{(l+s-1)!(s-l+1)+\beta_{l,s}}{s(1-z)^{l+s} \log^{s+1}\big(\frac{1}{1-z}\big)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s} \log^{s+2}\big(\frac{1}{1-z}\big)}\Big),$$

and thus to the following recurrence for the coefficients $\alpha_{l,s}$:

$$\begin{aligned}\alpha_{l,s} &= \frac{(l+s-1)!(s-l+1)+\beta_{l,s}}{s} \\ &= (l+s-1)\alpha_{l,s-1} + \frac{(l+s-1)!\big[s-1-(s+1)(H_{l+s-1}-H_{s+1})\big]}{s} \\ &\quad + \frac{(l-1)!(s-1)!\big(\binom{l+s-1}{s}-1\big)}{s} - (s-1)(l+s-2)!,\end{aligned}$$

with initial value $\alpha_{l,1} = (l-1)!(4l-1-2lH_l)$. It is not hard to check that the coefficients $\alpha_{l,s}$ defined in equation (1.20) satisfy this recurrence. Thus expansion (1.20) and also Theorem 1 are shown for all $l, s \geq 1$.

1.4.4 The centered moments

It remains to prove Corollary 1 for the centered moments of $X_{n,l}$. To show (1.3) we use the corresponding expansion for the ordinary moments as given by Theorem 1. This leads to

$$\begin{aligned}\mathbb{E}\Big(\big[X_{n,l}-\mathbb{E}(X_{n,l})\big]^s\Big) &= \sum_{j=0}^{s}\binom{s}{j}(-1)^{s-j}\mathbb{E}\big(X_{n,l}^j\big)\big(\mathbb{E}(X_{n,l})\big)^{s-j} \\ &= \Big(\sum_{j=0}^{s}\binom{s}{j}(-1)^{s-j}\Big)\frac{n^s}{\log^s n} + \frac{\delta_{l,s}\, n^s}{\log^{s+1} n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+2} n}\Big),\end{aligned} \tag{1.21}$$

with constants

$$\delta_{l,s} = \sum_{j=0}^{s}\binom{s}{j}(-1)^{s-j} f_{l,s}(j),$$

where the functions $f_{l,s}(j)$ are given as follows:

$$\begin{aligned}f_{l,s}(j) &= (j+1)H_j - jH_{l+j-1} - \frac{1}{2}\big(H_{l-1}^2 - H_{l-1}^{(2)}\big) - H_jH_{l-1} + \sum_{k=1}^{l-1}\frac{H_{j+k}}{k} - \sum_{k=1}^{l-1}\frac{1}{k^2\binom{j+k}{k}} + j\Psi(l+j) \\ &\quad + (s-j)\big[4 - H_l - \frac{1}{l} + \Psi(l+1)\big].\end{aligned}$$

Since it holds that $\sum_{j=0}^{s}\binom{s}{j}(-1)^{s-j} = 0$, for all $s \geq 1$, the first term of (1.21) vanishes. To show Corollary 1 it only remains to simplify the expression for the constants $\delta_{l,s}$. One can do this, for instance, by using the calculus of higher order differences (see, e. g., [30]). Below we give two identities that can be shown by this method, which are required to obtain a closed form expression for $\delta_{l,s}$.

$$\sum_{j=0}^{s}\binom{s}{j}(-1)^{s-j}\sum_{k=1}^{l-1}\frac{H_{j+k}}{k} = (-1)^{s-1}\Big(\frac{1}{s^2} - \frac{1}{s^2\binom{l+s-1}{l-1}}\Big),$$
$$\sum_{j=0}^{s}\binom{s}{j}(-1)^{s-j}\sum_{k=1}^{l-1}\frac{1}{k^2\binom{j+k}{k}} = \frac{(-1)^s}{s}\big(H_{l-1} - H_{l+s-1} + H_s\big).$$

This eventually leads to Corollary 1.

1.5 Isolating nodes with large labels

1.5.1 Isolating node n

Now we are studying the random variable $X_{n,n+1-l}$, for $l \geq 1$ fixed. First we will consider the special case $l = 1$, i. e., the instance of isolating the node with largest label n in a size-n recursive tree. We show the following lemma and prove thus the case $l = 1$ of Theorem 2.

Lemma 6. *The s-th moments $\mathbb{E}\big(X_{n,n}^s\big)$ of the number of cuts necessary to isolate node n in a random recursive tree of size n are, for $s \geq 1$ and $n \to \infty$, asymptotically given by*

$$\mathbb{E}\big(X_{n,n}^s\big) = \frac{n^s}{(s+1)\log^s n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+1} n}\Big).$$

An asymptotic equivalent of the expectation $\mathbb{E}(X_{n,n})$ together with a $\mathcal{O}$-bound for the variance of $X_{n,n}$ was already given in [38].

After simplifying the expressions for the splitting probabilities $\mathfrak{p}_{(n,1),(k,1)}$ as computed in Section 1.3 we can write the distribution recurrence (1.5) as follows:

$$\begin{aligned}\mathbb{P}\{X_{n,n} = m\} &= \sum_{k=1}^{n-1}\mathfrak{p}_{(n,1),(k,1)}\mathbb{P}\{X_{k,k} = m-1\}\\ &= \frac{1}{n-1}\sum_{k=1}^{n-1}\left(\frac{1}{k} + \frac{k-1}{(n-k)(n+1-k)}\right)\mathbb{P}\{X_{k,k} = m-1\}, \quad n \geq 2,\end{aligned} \tag{1.22}$$

with $\mathbb{P}\{X_{1,1} = 0\} = 1$. For computing the s-th moments of $X_{n,n}$ we multiply (1.22) with $m^s = \sum_{j=0}^{s}\binom{s}{j}(m-1)^s$ and sum up for $m \geq 1$, which leads to the following recurrence:

$$\mathbb{E}\big(X_{n,n}^s\big) = \frac{1}{n-1}\sum_{j=0}^{s}\binom{s}{j}\sum_{k=1}^{n-1}\Big(\frac{1}{k} + \frac{k-1}{(n-k)(n+1-k)}\Big)\mathbb{E}\big(X_{k,k}^j\big). \tag{1.23}$$

We treat (1.23) by introducing the generating functions

$$N_{1,s}(z) := \sum_{n\geq 1}\frac{1}{n}\mathbb{E}\big(X_{n,n}^s\big)z^n. \tag{1.24}$$

In the sequel we will obtain the following asymptotic expansion of the generating function $N_{1,s}(z)$, which leads, after applying basic singularity analysis, to Lemma 6.

$$N_{1,s}(z) = \frac{(s-1)!}{(s+1)(1-z)^s\log^s\big(\frac{1}{1-z}\big)} + \mathcal{O}\Big(\frac{1}{(1-z)^s\log^{s+1}\big(\frac{1}{1-z}\big)}\Big). \tag{1.25}$$

To show (1.25) we will use induction on s. First we have to consider the case $s = 1$. Plugging $s = 1$ into the recurrence (1.23) leads after multiplying with $(n-1)z^{n-1}$ and summing up for $n \geq 2$ to the following second order linear differential equation for the generating function $N_{1,1}(z)$:

$$(1-z)\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}N_{1,1}(z) - \frac{1}{1-z}N_{1,1}(z) = \frac{z}{(1-z)^2}, \tag{1.26}$$

with initial conditions $N_{1,1}(0) = 0$ and $\big(\frac{d}{dz}N_{1,1}(z)\big)\,|_{z=0} = \mathbb{E}(X_{1,1}) = 0$. The solution of the homogeneous differential equation corresponding to (1.26) is given by

$$N_1^{[h]}(z) = C_1 N_1^{[h_1]}(z) + C_2 N_1^{[h_2]}(z) = C_1 \log\big(\frac{1}{1-z}\big) + C_2 \log\big(\frac{1}{1-z}\big)\int_\alpha^z \frac{dt}{\log^2\big(\frac{1}{1-t}\big)}, \tag{1.27}$$

where we may choose an arbitrary $0 < \alpha < 1$. Since the Wronski determinant of the two homogeneous solutions equals one,

$$N_1^{[h_1]}(z)\frac{d}{dz}N_1^{[h_2]}(z) - N_1^{[h_2]}(z)\frac{d}{dz}N_1^{[h_1]}(z) = 1, \tag{1.28}$$

a particular solution of (1.26) is given by

$$\begin{aligned} N_{1,1}^{[p]}(z) &= N_1^{[h_1]}(z)\int_0^z -b_{1,1}(t)N_1^{[h_2]}(t)dt + N_1^{[h_2]}(z)\int_0^z b_{1,1}(t)N_1^{[h_1]}(t)dt \\ &= \log\big(\frac{1}{1-z}\big)\bigg(\int_\alpha^z \frac{dt}{\log^2\big(\frac{1}{1-t}\big)}\int_0^z \frac{t}{(1-t)^3}dt - \int_0^z \frac{t}{(1-t)^3}\int_\alpha^t \frac{1}{\log^2\big(\frac{1}{1-u}\big)}du\,dt\bigg), \end{aligned} \tag{1.29}$$

with $b_{1,1}(z) = \frac{z}{(1-z)^3\log\left(\frac{1}{1-z}\right)}$. Using integration by parts we can simplify the particular solution as follows:

$$N_{1,1}^{[p]}(z) = \log\big(\frac{1}{1-z}\big)\int_0^z \frac{t^2}{2(1-t)^2\log^2\big(\frac{1}{1-t}\big)}dt. \tag{1.30}$$

It turns out that the particular solution (1.30) satisfies also the initial conditions of (1.26) and is thus the required solution, i. e., $N_{1,1}(z) = N_{1,1}^{[p]}(z)$. An application of singular integration, Lemma 4, shows then the instance $s = 1$ of expansion (1.25).

Now we assume that expansion (1.25) is valid for all values $j < s$, with an arbitrary $s > 1$. To show the expansion also for s we again translate recurrence (1.23) by multiplying with $(n-1)z^{n-1}$ and summing up for $n \geq 2$ into the following inhomogeneous second order linear differential equation:

$$(1-z)\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}N_{1,s}(z) - \frac{1}{1-z}N_{1,s}(z) = R_{1,s}(z), \tag{1.31}$$

with inhomogeneous part

$$R_{1,s}(z) = \frac{z}{(1-z)^2} + \sum_{j=1}^{s-1}\binom{s}{j}\bigg[\Big(z - (1-z)\log\big(\frac{1}{1-z}\big)\Big)\frac{d^2}{dz^2}N_{1,j}(z) + \frac{N_{1,j}(z)}{1-z}\bigg],$$

and initial conditions $N_{1,s}(0) = 0$ and $\big(\frac{d}{dz}N_{1,s}(z)\big)\,|_{z=0} = \mathbb{E}\big(X_{1,1}^s\big) = 0$. Since we have already computed the general solution of the corresponding homogeneous differential equation we can immediately state the particular solution of this differential equation:

$$N_{1,s}^{[p]}(z) = N_1^{[h_1]}(z)\int_0^z -b_{1,s}(t)N_1^{[h_2]}(t)dt + N_1^{[h_2]}(z)\int_0^z b_{1,s}(t)N_1^{[h_1]}(t)dt, \tag{1.32}$$

where the homogeneous solutions $N_1^{[h_1]}(z)$ and $N_1^{[h_2]}(z)$ are given by (1.27) and furthermore $b_{1,s}(z) = \frac{R_{1,s}(z)}{(1-z)\log\left(\frac{1}{1-z}\right)}$. Again via integration by parts we can simplify the particular solution and obtain the following formula:

$$N_{1,s}^{[p]}(z) = \log\Big(\frac{1}{1-z}\Big) \int_0^z \Big(\int_0^t \frac{R_{1,s}(u)}{1-u} du\Big) \frac{dt}{\log^2\big(\frac{1}{1-t}\big)}. \quad (1.33)$$

It can be checked easily that (1.33) also satisfies the initial conditions and is thus the solution of (1.31), i. e., $N_{1,s}(z) = N_{1,s}^{[p]}(z)$.

Plugging the expansions (1.25) of $N_{1,j}(z)$, for $j < s$, into the formula for $R_{1,s}(z)$ we obtain the following expansion of the inhomogeneous part in a neighborhood of $z = 1$:

$$R_{1,s}(z) = \frac{s!}{(1-z)^{s+1}\log^{s-1}\big(\frac{1}{1-z}\big)} + \mathcal{O}\Big(\frac{1}{(1-z)^{s+1}\log^{s}\big(\frac{1}{1-z}\big)}\Big).$$

This gives by applying singular integration, Lemma 4,

$$\Big(\int_0^t \frac{R_{1,s}(u)}{1-u} du\Big) \frac{1}{\log^2\big(\frac{1}{1-t}\big)} = \frac{s!}{(s+1)(1-t)^{s+1}\log^{s+1}\big(\frac{1}{1-t}\big)} + \mathcal{O}\Big(\frac{1}{(1-t)^{s+1}\log^{s+2}\big(\frac{1}{1-t}\big)}\Big),$$

and eventually leads, after a further application of Lemma 4, due to (1.33) also for s to expansion (1.25). Thus (1.25) and as a consequence Lemma 6 are shown for all $s \geq 1$.

1.5.2 Expectations for large nodes

Next we show the following asymptotic expansion of the expectation $\mathbb{E}(X_{n,n+1-l})$, for l fixed. Lemma 7 gives thus the special case $s = 1$ of Theorem 2.

Lemma 7. *The expectations $\mathbb{E}(X_{n,n+1-l})$ of the number of random cuts necessary to isolate node $n+1-l$ in a random recursive tree of size n are, for $l \geq 1$ fixed and $n \to \infty$, asymptotically given by*

$$\mathbb{E}(X_{n,n+1-l}) = \frac{n}{2\log n} + \mathcal{O}\Big(\frac{n}{\log^2 n}\Big).$$

To show this lemma we study the following recurrence for $\mathbb{E}(X_{n,n+1-l})$, which is obtained from the distribution recurrence (1.7) after multiplying with $m = (m-1)+1$ and summing up for $m \geq 1$:

$$\mathbb{E}(X_{n,n+1-l}) = 1 + \sum_{r=1}^{l}\sum_{k=r}^{n-1} \mathfrak{p}_{(n,l),(k,r)} \mathbb{E}(X_{k,k+1-r}). \quad (1.34)$$

After simplifying the expressions of the splitting probabilities $\mathfrak{p}_{(n,l),(k,r)}$ as given by (1.5) and Lemma 3 we obtain

$$\begin{aligned}\mathfrak{p}_{(n,l),(k,r)} &= [\![k \leq n+r-l]\!] \frac{\binom{l-1}{r-1}}{(n-1)(n-1)^{\underline{l-1}}} \big[(k-1)^{\underline{r-2}}(n-k-1)^{\underline{l-r}} + (k-1)^{\underline{r}}(n-k-1)^{\underline{l-r-2}}\big] \\ &\quad + [\![k = n+r-l]\!] \binom{l}{r-1} \frac{(l-r-1)!}{(n-1)(n-1)^{\underline{l-r}}}, \quad r \leq l,\end{aligned} \quad (1.35)$$

where we use the convention $(j-1)^{\underline{-p}} := (j^{\overline{p}})^{-1}$, $p \in \mathbb{N}$, see e. g. [30].

To treat recurrence (1.34) we introduce for $l \geq 1$ the following generating functions:

$$N_{l,1}(z) := \sum_{n \geq l} (n-1)^{\underline{l-2}}\, \mathbb{E}(X_{n,n+1-l}) z^{n+1-l}. \quad (1.36)$$

Note that this definition also holds for $l = 1$, where it matches with the definition of $N_{1,1}(z)$ given by (1.24). Again, due to an application of basic singularity analysis, it suffices to show that $N_{l,1}(z)$ admits the following expansion in a neighborhood of the dominant singularity $z = 1$, which proves Lemma 7.

$$N_{l,1}(z) = \frac{(l-1)!}{2(1-z)^l \log\left(\frac{1}{1-z}\right)} + \mathcal{O}\Big(\frac{1}{(1-z)^l \log^2\left(\frac{1}{1-z}\right)}\Big). \tag{1.37}$$

We will show expansion (1.37) by induction on l. The case $l = 1$ was treated already in Subsection 1.5.1, where it turned out that (1.37) holds.

Above recurrence (1.34) can be translated into the following second order linear differential equation for $N_{l,1}(z)$, where the functions $N_{r,1}(z)$, with $r < l$, are all appearing in the inhomogeneous part $R_{l,1}(z)$. $R_{l,1}(z)$ is now a bit "unpleasant", since one had to consider separately the four cases $r = 1$, $1 < r < l-1$, $r = l-1$ and $r = l$. One obtains

$$(1-z)\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}N_{l,1}(z) + (l-1)\frac{d}{dz}N_{l,1}(z) - \frac{1}{1-z}N_{l,1}(z) = R_{l,1}(z), \tag{1.38}$$

with inhomogeneous part

$$\begin{aligned}R_{l,1}(z) &= \frac{(l-1)!(l-1+z)}{(1-z)^{l+1}} + \sum_{r=1}^{l-1}\sum_{n\ge l}\Big[\sum_{k=r}^{n-1}(n-1)(n-1)^{\underline{l-1}}\mathfrak{p}_{(n,l),(k,r)}\mathbb{E}(X_{k,k+1-r})\Big]z^{n-l}\\ &= \frac{(l-1)!(l-1+z)}{(1-z)^{l+1}} + \sum_{r=1}^{l-2}\binom{l-1}{r-1}\Big[\frac{(l-2-r)!}{(1-z)^{l-1-r}}\Big(\frac{d^2}{dz^2}N_{r,1}(z)-1\Big) + N_{r,1}(z)\frac{(l-r)!}{(1-z)^{l+1-r}}\Big]\\ &\quad + (l-1)\Big[\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}N_{l-1,1}(z) + \frac{N_{l-1,1}(z)}{(1-z)^2}\Big] + \sum_{r=1}^{l-1}\binom{l}{r-1}(l-r-1)!\frac{d}{dz}N_{r,1}(z),\end{aligned}$$

and initial conditions $N_{l,1}(0) = 0$, $\big(\frac{d}{dz}N_{l,1}\big)\big|_{z=0} = (l-1)!\,\mathbb{E}(X_{l,1})$. One gets that the homogeneous differential equation corresponding to (1.38) has the general solution

$$\begin{aligned}N_l^{[h]}(z) &= C_1 N_l^{[h_1]}(z) + C_2 N_l^{[h_2]}(z)\\ &= C_1\Big(l-1+\log\big(\frac{1}{1-z}\big)\Big) + C_2\Big(l-1+\log\big(\frac{1}{1-z}\big)\Big)\int_\alpha^z \frac{dt}{\log^{l-1}\big(\frac{1}{1-t}\big)\big[l-1+\log\big(\frac{1}{1-t}\big)\big]^2},\end{aligned} \tag{1.39}$$

where we may choose an arbitrary $0 < \alpha < 1$. This leads to the particular solution

$$N_{l,1}^{[p]}(z) = N_l^{[h_1]}(z)\int_0^z \frac{-b_{l,1}(t)N_l^{[h_2]}(t)}{D_l(t)}dt + N_l^{[h_2]}(z)\int_0^z \frac{b_{l,1}(t)N_l^{[h_1]}(t)}{D_l(t)}dt,$$

with $b_{l,1}(z) = \frac{R_{l,1}(z)}{(1-z)\log\left(\frac{1}{1-z}\right)}$, and where

$$D_l(z) = N_l^{[h_1]}(z)\frac{d}{dz}N_l^{[h_2]}(z) - N_l^{[h_2]}(z)\frac{d}{dz}N_l^{[h_1]}(z) = \frac{1}{\log^{l-1}\left(\frac{1}{1-z}\right)} \tag{1.40}$$

is the Wronski determinant of the two homogeneous solutions $N_l^{[h_1]}(z)$ and $N_l^{[h_2]}(z)$. Again via integration by parts we can write the expression of the particular solution as follows:

$$\begin{aligned}N_{l,1}^{[p]}(z) &= \Big(l-1+\log\big(\frac{1}{1-z}\big)\Big)\times\\ &\times\int_0^z\Big(\int_0^t \frac{\log^{l-2}\left(\frac{1}{1-u}\right)\left(l-1+\log\left(\frac{1}{1-u}\right)\right)R_{l,1}(u)}{1-u}du\Big)\frac{1}{\log^{l-1}\left(\frac{1}{1-t}\right)\left(l-1+\log\left(\frac{1}{1-z}\right)\right)^2}dt.\end{aligned} \tag{1.41}$$

It turns out that (1.41) matches the initial conditions of (1.38) and is thus the required solution, i. e.,

$N_{l,1}(z) = N_{l,1}^{[p]}(z)$.
Using the explicit description (1.41) of the generating function $N_{l,1}(z)$ we can show the required expansion (1.37). We assume that expansion (1.37) of $N_{r,1}(z)$ holds, for all $r < l$ with an arbitrary $l > 1$. Plugging these expansions into the formula of the inhomogeneous part $R_{l,1}(z)$ given above we obtain, after a heavy use of singular differentiation, the following expansion around $z = 1$:

$$R_{l,1}(z) = \frac{(l+1)!}{2(1-z)^{l+1}} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+1}\log\big(\frac{1}{1-z}\big)}\Big). \tag{1.42}$$

Singular integration leads then to the expansion

$$\int_0^t \frac{\log^{l-2}\big(\frac{1}{1-u}\big)\big(l-1+\log\big(\frac{1}{1-u}\big)\big)R_{l,1}(u)}{1-u}du = \frac{l!\log^{l-1}\big(\frac{1}{1-t}\big)}{2(1-t)^{l+1}} + \mathcal{O}\Big(\frac{\log^{l-2}\big(\frac{1}{1-t}\big)}{(1-t)^{l+1}}\Big),$$

and eventually also to (1.37) for $l > 1$. Thus expansion (1.37) and as a consequence Lemma 7 are shown for all $l \geq 1$.

1.5.3 Higher moments for large nodes

The method applied in Subsection 1.5.1-1.5.2 for a study of the s-th moments of $X_{n,n}$ and the expectations of $X_{n,n+1-l}$ can be extended naturally to show Theorem 2, leading to an asymptotic expansion of the s-th moments of the number of random cuts necessary to isolate the l-th largest node in a random size-n recursive tree.
Again we use the distribution recurrence (1.7) to give a recurrence for the s-th moments of $X_{n,n+1-l}$. After multiplying with $m^s = \sum_{j=0}^{s}\binom{s}{j}(m-1)^j$ and summing up for $m \geq 1$ we obtain:

$$\mathbb{E}\big(X_{n,n+1-l}^s\big) = \sum_{j=0}^{s}\binom{s}{j}\sum_{r=1}^{l}\sum_{k=r}^{n-1}\mathfrak{p}_{(n,l),(k,r)}\mathbb{E}\big(X_{k,k+1-r}^j\big). \tag{1.43}$$

It is now appropriate to introduce for $l, s \geq 1$ the generating functions

$$N_{l,s}(z) := \sum_{n\geq l}(n-1)^{\underline{l-2}}\,\mathbb{E}\big(X_{n,n+1-l}^s\big)z^{n+1-l}, \tag{1.44}$$

which are generalizations of $N_{1,s}(z)$ and $N_{l,1}(z)$ as used in Subsection 1.5.1-1.5.2. It is again sufficient to show that $N_{l,s}(z)$ admits the following expansion in a neighborhood of the dominant singularity $z = 1$:

$$N_{l,s}(z) = \frac{(l+s-2)!}{(s+1)(1-z)^{l+s-1}\log^s\big(\frac{1}{1-z}\big)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s-1}\log^{s+1}\big(\frac{1}{1-z}\big)}\Big), \tag{1.45}$$

which will be done by induction on both parameters, l and s. Basic singularity analysis leads then directly to Theorem 2. The margin cases $l = 1$, $s \geq 1$ and $l \geq 1$, $s = 1$ are already shown in Subsection 1.5.1-1.5.2.
We proceed by translating the recurrence (1.43) into the following second order linear differential equation for $N_{l,s}(z)$:

$$(1-z)\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}N_{l,s}(z) + (l-1)\frac{d}{dz}N_{l,s}(z) - \frac{1}{1-z}N_{l,s}(z) = R_{l,s}(z), \tag{1.46}$$

with inhomogeneous part

$$R_{l,s}(z) = \frac{(l-1)!(l-1+z)}{(1-z)^{l+1}} + \sum_{j=1}^{s-1}\binom{s}{j}\sum_{n\geq l}\Big[\sum_{k=l}^{n-1}(n-1)(n-1)^{\underline{l-1}}\mathfrak{p}_{(n,l),(k,l)}\mathbb{E}(X_{k,k+1-l}^j)\Big]z^{n-l}$$

$$
\begin{aligned}
&+\sum_{j=1}^{s}\binom{s}{j}\sum_{r=1}^{l-1}\sum_{n\geq l}\Big[\sum_{k=r}^{n-1}(n-1)(n-1)^{\underline{l-1}}\mathfrak{p}_{(n,l),(k,r)}\mathbb{E}(X_{k,k+1-r}^{j})\Big]z^{n-l}\\
&=\frac{(l-1)!(l-1+z)}{(1-z)^{l+1}}+\sum_{j=1}^{s-1}\binom{s}{j}\left[\frac{N_{l,j}(z)}{1-z}-(1-z)\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}N_{l,j}(z)+\frac{d^2}{dz^2}N_{l,j}(z)\right]\\
&+\sum_{j=1}^{s}\binom{s}{j}\sum_{r=1}^{l-2}\binom{l-1}{r-1}\left[\frac{(l-2-r)!}{(1-z)^{l-1-r}}\Big(\frac{d^2}{dz^2}N_{(r,j)}(z)-1\Big)+N_{r,j}(z)\frac{(l-r)!}{(1-z)^{l+1-r}}\right]\\
&+\sum_{j=1}^{s}\binom{s}{j}(l-1)\left[\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}N_{l-1,j}(z)+N_{l-1,j}(z)\frac{1}{(1-z)^2}\right]\\
&+\sum_{j=1}^{s}\binom{s}{j}\sum_{r=1}^{l-1}\binom{l}{r-1}(l-r-1)!\frac{d}{dz}N_{r,j}(z),
\end{aligned}
$$

and initial conditions $N_{l,s}(0)=0$ and $\big(\frac{d}{dz}N_{l,s}\big)\big|_{z=0}=(l-1)!\,\mathbb{E}(X_{l,1}^{s})$.
Since the homogeneous solution corresponding to (1.45) was already computed in Subsection 1.5.2 we obtain again a particular solution by applying the method of variation of constants. We get after simplifications

$$
\begin{aligned}
N_{l,s}^{[p]}(z)&=\Big(l-1+\log\big(\frac{1}{1-z}\big)\Big)\times \qquad (1.47)\\
&\times\int_0^z\Big(\int_0^t\frac{\log^{l-2}\big(\frac{1}{1-u}\big)\big(l-1+\log\big(\frac{1}{1-u}\big)\big)R_{l,s}(u)}{1-u}du\Big)\frac{1}{\log^{l-1}\big(\frac{1}{1-t}\big)\big(l-1+\log\big(\frac{1}{1-z}\big)\big)^2}dt.
\end{aligned}
$$

Now we assume that for all pairs $(r,j)<(l,s)$, which means for $r\leq l$, $j\leq s$ and $(r,j)\neq(l,s)$, the functions $N_{r,j}(z)$ admit the asymptotic expansion (1.45) in a neighborhood of the dominant singularity $z=1$. We want to show that (1.45) also holds for the pair (l,s), where we may assume $s>1$ and $l>1$, since the margin cases are already treated. Plugging the expansions of the functions $N_{r,j}(z)$ into the formula for the inhomogeneous part $R_{l,s}(z)$ and using singular differentiation we obtain that the main contributions in the expansion of $R_{l,s}(z)$ around $z=1$ are stemming from the terms $\binom{s}{s-1}\frac{d^2}{dz^2}N_{l,s-1}(z)$ and $\binom{s}{s}(l-1)\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}N_{l-1,s}(z)$. This gives

$$
\begin{aligned}
R_{l,s}(z)&=\frac{\binom{s}{s-1}(l+s-1)!}{s(1-z)^{l+s}\log^{s-1}\big(\frac{1}{1-z}\big)}+\frac{(l-1)(l+s-1)!}{(s+1)(1-z)^{l+s}\log^{s-1}\big(\frac{1}{1-z}\big)}+\mathcal{O}\Big(\frac{1}{(1-z)^{l+s}\log^{s}\big(\frac{1}{1-z}\big)}\Big)\\
&=\frac{(l+s)!}{(s+1)(1-z)^{l+s}\log^{s-1}\big(\frac{1}{1-z}\big)}+\mathcal{O}\Big(\frac{1}{(1-z)^{l+s}\log^{s}\big(\frac{1}{1-z}\big)}\Big). \qquad (1.48)
\end{aligned}
$$

One further obtains the expansion

$$
\begin{aligned}
&\int_0^t\frac{\log^{l-2}\big(\frac{1}{1-u}\big)\big(l-1+\log\big(\frac{1}{1-u}\big)\big)R_{l,s}(u)}{1-u}du\\
&\quad=\frac{(l+s-1)!}{(s+1)(1-t)^{l+s}\log^{s-l}\big(\frac{1}{1-t}\big)}+\mathcal{O}\Big(\frac{1}{(1-t)^{l+s}\log^{s-l+1}\big(\frac{1}{1-t}\big)}\Big),
\end{aligned}
$$

which eventually leads to (1.45) also for pairs (l,s), with $l>1$ and $s>1$. Thus (1.45) is shown for all $l\geq 1$ and $s\geq 1$, which completes the proof of Theorem 2.

1.5.4 The limiting distribution

From the asymptotic expansion of the s-th moments of $X_{n,n+1-l}$ as given by Theorem 2 we obtain for the s-th moments of the scaled random variable $\tilde{X}_{n,n+1-l} := \frac{\log n}{n} X_{n,n+1-l}$, for $l \geq 1$, $s \geq 0$ and $n \to \infty$:

$$\mathbb{E}\big(\tilde{X}^s_{n,n+1-l}\big) \to \frac{1}{s+1}.$$

If we consider a random variable X, which has a standard uniform distribution U_1 with support $[0,1)$ then the s-th moments of X are, for $s \geq 0$, given as follows:

$$\mathbb{E}(X^s) = \frac{1}{s+1}.$$

A direct application of the Theorem of Fréchet and Shohat (the second central limit theorem, see, e. g, [49]) proves then Theorem 3.

1.6 Conclusion

Using a generating functions approach in combination with singularity analysis and lemmata for singular differentiation and integration we obtain distribution results for the number of random cuts necessary to isolate large nodes and small nodes in random recursive trees via random cuttings. Although the recurrences obtained in our analysis could be applied to treat the parameter studied for arbitrary labels l and tree sizes n, it turns out that for general growth rates of l compared to n (e. g., when isolating nodes in the central region $l/n \sim \rho$, with $0 < \rho < 1$) it seems considerably more difficult to attack the problem.

Chapter 2

Multiple Isolation of nodes in recursive trees

2.1 Introduction

It turns out that the edge-removal procedure, isolating a node with specified label λ, with $1 \le \lambda \le n$, in a tree T with nodes labelled by $1, 2, \ldots, n$, as considered in Chapter 1, is a special case of an even more general edge-removal procedure. We consider the following edge-removal procedure in a size n labelled rooted tree T for isolating the nodes $(v_{\lambda_1}, v_{\lambda_2}, \ldots, v_{\lambda_i})$ labelled $1 \le \lambda_1 < \lambda_2 < \cdots < \lambda_{i-1} < \lambda_i \le n$: for $1 \le i \le n$.

- Pick one of the $n-1$ edges of the tree at random and remove it. This separates the tree T into two subtrees $\hat{T}$ and $\tilde{T}$. Let us assume that $\lambda_{j_1}, \ldots, \lambda_{j_k} \in \hat{T}$ and $\lambda_{j_{k+1}}, \ldots, \lambda_{j_i} \in \tilde{T}$, with $1 \le k \le i$ and $\{\lambda_{j_1}, \ldots, \lambda_{j_k}\} \cup \{\lambda_{j_{k+1}}, \ldots, \lambda_{j_i}\} = \{\lambda_1, \ldots, \lambda_i\}$.
- Continue the edge-removal procedure recursively in subtree $\hat{T}$, which contains the nodes $\lambda_{j_1}, \ldots, \lambda_{j_k}$, and $\tilde{T}$ containing the nodes $\lambda_{j_{k+1}}, \ldots, \lambda_{j_i}$ until the nodes labelled by $\lambda_1, \ldots, \lambda_i$ are isolated.

We study this procedure for isolating the nodes labelled $1, 2, \ldots, l$, nodes near the root, and $n+1-l, \ldots, n$, the last l inserted nodes. Further we isolate pairs $1, l$, $n+1-l, n$ and $1, n$.

We denote with $X_{n;\lambda_1,\lambda_2,\ldots,\lambda_l}$ the random variable which counts the number of random cuts necessary to isolate the nodes labelled $\lambda_1, \lambda_2, \ldots, \lambda_l$ in a random size n recursive tree. Further we denote with $Y_{n;\lambda_1,\lambda_2,\ldots,\lambda_l}$ the random variable which counts the number of random cuts necessary to isolate the nodes labelled $n+1-\lambda_l, n+1-\lambda_{l-1}, \ldots, n+1-\lambda_1$ in a random size n recursive tree. Thus $X_{n;\lambda_1,\lambda_2,\ldots,\lambda_l} = Y_{n;n+1-\lambda_l,n+1-\lambda_2,\ldots,n+1-\lambda_1}$.

For writing convenience we use the abbreviations $\boldsymbol{l} := 1, \ldots, l$, thus $X_{n;\boldsymbol{l}} = X_{n;1,\ldots,l}$ and $Y_{n;\boldsymbol{l}} = Y_{n;1,\ldots,l} = X_{n;n+1-l,\ldots,n}$.

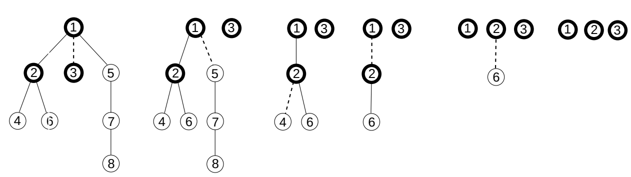

Figure 2.1: Isolating the nodes 1,2 and 3 in a size 8 recursive tree with 5 cuts.

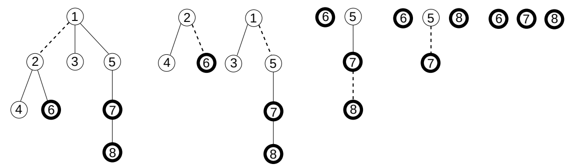

Figure 2.2: Isolating the nodes 6, 7 and 8 in a size 8 recursive tree with 5 cuts.

2.2 Results

For the moments of the random variables $X_{n;\boldsymbol{l}}$, $X_{n;1,l}$, $Y_{n;\boldsymbol{l}}$, $Y_{n;1,l}$ and $X_{n;1,n}$ we obtain the following results.

Theorem 4. *The s-th $\mathbb{E}(X^s_{n;\boldsymbol{l}})$, $s \geq 1$ of the number of cuts necessary to isolate nodes $\boldsymbol{l} = 1, \ldots, l$ in a random recursive tree of size n is for fixed l and $n \to \infty$ asymptotically given by*

$$\mathbb{E}(X^s_{n;\boldsymbol{l}}) = \frac{n^s}{\log^s n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+1} n}\Big).$$

Theorem 5. *The s-th moment $\mathbb{E}(X^s_{n;1,l})$, $s \geq 1$ of the number of cuts necessary to isolate the nodes $(1, l)$ in a random recursive tree of size n is for fixed l and $n \to \infty$ asymptotically given by*

$$\mathbb{E}(X^s_{n;1,l}) = \frac{n^s}{\log^s n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+1} n}\Big).$$

Theorem 6. *The s-th moment $\mathbb{E}(Y^s_{n;\boldsymbol{l}})$, $s \geq 1$ of the number of cuts necessary to isolate the nodes $n+1-l, \ldots, n$ in a random recursive tree of size n is for fixed l and $n \to \infty$ asymptotically given by*

$$\mathbb{E}(Y^s_{n;\boldsymbol{l}}) = \frac{l n^s}{(s+l)\log^s n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+1} n}\Big).$$

Theorem 7. *The s-th moment $\mathbb{E}\big(Y^s_{n;1,l}\big)$ of the number of cuts necessary to isolate nodes $(n+1-l, n)$ in a random recursive tree of size n is for $s \geq 1$ and $n \to \infty$ asymptotically given by*

$$\mathbb{E}\big(Y^s_{n;1,l}\big) = \frac{2n^s}{(s+2)\log^s n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+1} n}\Big),$$

Theorem 8. *The s-th moment $\mathbb{E}\big(X^s_{n;1,n}\big)$ of the number of cuts necessary to isolate nodes $(1, n)$ in a random recursive tree of size n is for $s \geq 1$ and $n \to \infty$ asymptotically given by*

$$\mathbb{E}\big(X^s_{n;1,n}\big) = \frac{n^s}{\log^s n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+1} n}\Big),$$

thus the dominant asymptotic part is the same as for the random variable $X_{n;1}$.

By using the method of moments we obtain from Theorem 6 and 7 immediately the following Theorems.

Theorem 9. *The the random variable $\frac{\log n}{n} Y_{n;\boldsymbol{l}}$ converges in distribution to the continuous random variable Y_l, which has the density $f_{Y_l}(z) = l z^{l-1}$ with support $[0, 1]$ and the distribution function $F_l(z) = z^l$,*

$$\frac{\log n}{n} Y_{n;\boldsymbol{l}} \xrightarrow{(d)} Y_l.$$

Note that if we take the limit $l \to \infty$ of the random variable Y_l, it converges in distribution to the random variable Y_∞, with $\mathbb{P}\{Y_\infty = 1\} = 1$. Below you see the distribution functions of Y_1, Y_2 and Y_5.

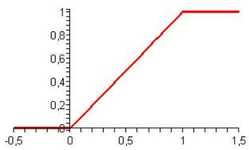

Figure 2.3: $F_1(z)$

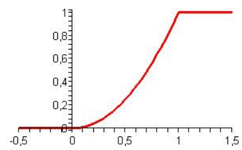

Figure 2.4: $F_2(z)$

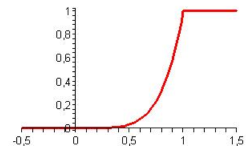

Figure 2.5: $F_5(z)$

Theorem 10. *The normalized random variable* $\frac{\log n}{n} Y_{n;1,l}$ *converges to the continuous random variable* Y_2 *with support* $[0,1]$*, where the density* $f_{Y_l}(z) = 2z$*, for* $0 \le z \le 1$*,*

$$\frac{\log n}{n} Y_{n;1,l} \xrightarrow{(d)} Y_2.$$

2.3 The recursive approach

We will rely on the same approach as in Chapter 1. The most important step for the recursive description of the considered probabilities is again to setup splitting probabilities.

2.3.1 The recurrences

We use the splitting probabilities $p_{(n,l),(k,r)}$, as given by Lemma 3, which give the probability that when starting with a random size-n recursive tree and removing a random edge the subtree containing node l is of size k and where furthermore node l is the r-th smallest node in this subtree. We also use the splitting probabilities $\mathfrak{p}_{(n,l),(k,r)}$, covered by (1.5).

From the recursive description of the problem we immediately obtain the following recurrences for isolating the nodes labelled $\boldsymbol{l} = 1, \dots, l$ and $\boldsymbol{n+1-l} = n+1-l, \dots, n$ respectively. For isolating the nodes labelled $1, \dots, l$ we get the recurrence

$$\mathbb{P}\{X_{n;\boldsymbol{l}} = m\} = \sum_{r=1}^{l} \sum_{k=r}^{n-1} p_{(n,l),(k,r)} \sum_{s=0}^{m-1} \mathbb{P}\{X_{k;\boldsymbol{r}} = s\} \mathbb{P}\{X_{n-k;\boldsymbol{l-r}} = m-1-s\}, \tag{2.1}$$

for $n \ge l$, $m \ge l-1$, with initial value $\mathbb{P}\{X_{1;\boldsymbol{1}} = 0\} = \mathbb{P}\{X_{1;1} = 0\} = 1$ and where the splitting probabilities $p_{(n,l),(k,r)}$ are given in Lemma 3. It is obvious that $\mathbb{P}\{X_{l;\boldsymbol{l}} = l-1\} = 1$. Further we have for isolating the nodes labelled $n+1-l, \dots, n$

$$\begin{aligned} \mathbb{P}\{Y_{n;\boldsymbol{l}} = m\} = &\sum_{r=1}^{l} \sum_{k=r}^{n-1} \mathfrak{p}_{(n,l),(k,r)} \sum_{s=0}^{m-1} \mathbb{P}\{Y_{k;\boldsymbol{r}} = s\} \times \\ &\times \mathbb{P}\{Y_{n-k;\boldsymbol{l-r}} = m-1-s\}, \end{aligned} \tag{2.2}$$

for $n \ge l$, $m \ge l-1$ with initial value $\mathbb{P}\{Y_{1;\boldsymbol{1}} = 0\} = \mathbb{P}\{X_{1;1} = 0\} = 1$ and where the splitting probabilities $\mathfrak{p}_{(n,l),(k,r)}$ are given in Lemma 3 by using (1.5). In (2.1) we use the convention $\mathbb{P}\{X_{n-k,\boldsymbol{0}} = m-1-s\} = [\![s = m-1]\!]$, and in (2.2) $\mathbb{P}\{Y_{n-k,\boldsymbol{0}} = m-1-s\} = [\![s = m-1]\!]$, where we have used the Iverson bracket notation.

Moreover for isolating the pair $1, l$ we need the auxiliary probabilities $\tilde{p}_{(n,l),(k,r)}$ and $\tilde{q}_{(n,k),(r,l)}$. $\tilde{p}_{(n,l),(k,r)}$ denotes the probability that after a random cut in a random recursive tree T_n of size n the tree T containing both node 1 and l is of size k, where l is the r-st smallest node in T. Further we denote with $\tilde{q}_{(n,k),(r,l)}$ the probability that after a random cut the accruing subtree T' containing node l is of size k, where l is the r-st smallest node in T' and the root (node 1) is not in T'. It is obvious that

$\tilde{p}_{(n,k),(r,l)} + \tilde{q}_{(n,l),(k,r)} = p_{(n,k),(r,l)}$, with $p_{(n,l),(k,r)}$ as given in Lemma 3.

$$\begin{aligned}\mathbb{P}\{X_{n;1,l} = m\} = \sum_{r=1}^{l}\sum_{k=r}^{n-1}\Big(&\tilde{p}_{(n,l),(k,r)}\mathbb{P}\{X_{k;1,r} = m-1\}+\\ &+\tilde{q}_{(n,l),(k,r)}\sum_{s=0}^{m-1}\mathbb{P}\{X_{k;r} = s\}\mathbb{P}\{X_{n-k;1} = m-1-s\}\Big),\end{aligned} \tag{2.3}$$

for $n \geq l$, $m \geq 1$, with initial value $\mathbb{P}\{X_{1;1} = 0\} = 1$, and where the splitting probabilities $\tilde{p}_{(n,k),(r,l)}$ and $\tilde{q}_{(n,l),(k,r)}$ are given in Lemma 8.

We denote with $\tilde{\mathfrak{p}}_{(n,n+1-l),(k,r)}$ the probability that after a random cut in a random recursive tree T_n of size n the tree T containing both node n and $n+1-l$ is of size k, where $n+1-l$ is the r-th largest node in T. Further we denote with $\tilde{\mathfrak{q}}_{(n,n+1-l),(k,r)}$ the probability that after a random cut the accruing subtree T' containing node $n+1-l$ is of size k, where $n+1-l$ is the r-th largest node in T' and the last node (node n) is not in T'. It is obvious that $\tilde{\mathfrak{p}}_{(n,n+1-l),(k,r)} + \tilde{\mathfrak{q}}_{(n,n+1-l),(k,r)} = \mathfrak{p}_{(n,n+1-l),(k,r)}$, with $\mathfrak{p}_{(n,n+1-l),(k,r)}$ as given in Lemma 3 and relation (1.5).

$$\begin{aligned}\mathbb{P}\{X_{n;n+1-l,n} = m\} = \sum_{r=1}^{l}\sum_{k=r}^{n-1}\Big(&\tilde{\mathfrak{p}}_{(n,n+1-l),(k,r)}\mathbb{P}\{X_{k;k+1-r,k} = m-1\}+\\ &+\tilde{\mathfrak{q}}_{(n,n+1-l),(k,r)}\sum_{s=0}^{m-1}\mathbb{P}\{X_{k;k+1-r} = s\}\mathbb{P}\{X_{n-k;n-k} = m-1-s\}\Big),\end{aligned} \tag{2.4}$$

for $n \geq l$, $m \geq 1$, with initial value $\mathbb{P}\{X_{1;1} = 0\} = 1$, and where the splitting probabilities $\tilde{\mathfrak{p}}_{(n,n+1-l),(k,r)}$ and $\tilde{\mathfrak{q}}_{(n,n+1-l),(k,r)}$ are given in Lemma 9.

For isolating the nodes labelled 1,n we need the probabilities $p_{n,k}$ and $q_{n,k}$, where $p_{n,k}$ gives the probability that after a random cut the subtree containing both nodes is of size k and $q_{n,k}$ gives the probability that the subtree containing the root 1 is of size k.

$$\begin{aligned}\mathbb{P}\{X_{n;1,n} = m\} &= \sum_{k=2}^{n-1} p_{n,k}\mathbb{P}\{X_{k,;1,k;} = m-1\}\\ &+\sum_{k=1}^{n-1} q_{n,k}\sum_{s=0}^{m-1}\mathbb{P}\{X_{k;1} = s\}\mathbb{P}\{X_{n-k;n-k} = m-1-s\}, \quad n \geq 2,\end{aligned} \tag{2.5}$$

with initial value $\mathbb{P}\{X_{1;1} = 0\} = 1$ and where the splitting probabilities $p_{n,k}$ and $q_{n,k}$ are given in Lemma 10.

We will use a generating functions approach to turn these recurrences into differential equations for suitable defined generating functions. Solving the arising differential equations leads to an explicit form for the generating functions, thus extracting coefficients is possible by using a complex analysis technic called singularity analysis. Further we will use induction to prove our theorems.

2.3.2 The splitting probabilities

Lemma 8. *The splitting probabilities $\tilde{p}_{(n,l),(k,r)}$ and $\tilde{q}_{(n,l),(k,r)}$ are for $1 \le l \le n$, $1 \le r \le k$, $1 \le k \le n-1$ and $n \ge 2$ given as follows:*

$$\tilde{p}_{(n,l),(k,r)} = \begin{cases} p_{(n,l),(k,l)} = \left[(l-1)\binom{n-l}{n-k} + \binom{n-l+1}{n-k+1}\right] \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!}, & r = l, \\ \binom{l-1}{r-2}\binom{n-l}{k-r} \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!}, & 1 < r < l, \\ 0, & r = 1, \end{cases} \tag{2.6}$$

$$\tilde{q}_{(n,l),(k,r)} = \begin{cases} 0, & r = l, \\ \binom{l-1}{r}\binom{n-l}{k-r} \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!}, & 1 \le r < l. \end{cases}$$

The proof is omitted, for details see the proof of Lemma 3.

Lemma 9. *The splitting probabilities $\tilde{\mathfrak{p}}_{(n,n+1-l),(k,r)}$ and $\tilde{\mathfrak{q}}_{(n,n+1-l),(k,r)}$ are for $2 \le l \le n$, $1 \le r \le k$, $1 \le k \le n-1$ and $n \ge 2$ given as follows:*

$$\tilde{\mathfrak{p}}_{(n,n+1-l),(k,r)} = \begin{cases} \mathfrak{p}_{(n,n+1-l),(k,l)} = \left[\binom{n-l}{k-l-1} + \binom{n-l}{k-l+1}\right] \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!}, & r = l, \\ \binom{l-2}{r-2}\left[\binom{n-l}{k-r-1} + \binom{n-l}{k-r+1}\right] + \delta_{k,n+r-l}\binom{l-1}{l-r+1} \frac{(n+r-l-1)!(l-r-1)!}{(n-1)(n-1)!}, & 1 < r < l, \\ 0, & r = 1, \end{cases} \tag{2.7}$$

$$\tilde{\mathfrak{q}}_{(n,n+1-l),(k,r)} = \begin{cases} 0, & r = l, \\ \binom{l-2}{r-1}\left[\binom{n-l}{k-r-1} + \binom{n-l}{k-r+1}\right] + \delta_{k,n+r-l}\binom{l-1}{l-r} \frac{(n+r-l-1)!(l-r-1)!}{(n-1)(n-1)!}, & 1 \le r < l. \end{cases}$$

Proof. The cases $r = l$ and $r = 1$ are obvious. Now lets turn to $1 < r < l$. We cut a random edge e in a random recursive tree of size n. We denote with T the subtree arising after the random cut which contains the root, and with T' the other one. Further let B denote the size k subtree containing the node $n+1-l$ and B' the subtree containing the node n. At first we calculate $\tilde{\mathfrak{p}}_{(n,n+1-l),(k,r)}$. Assume $B = B' = T$. There are $r-2$ nodes out of the set $\{n+2-l, \dots, n-1\}$ in T. We can build the tree T' the following way. For cutting at j, with $1 \le j \le k-r$, we have $\binom{l-2}{l-r}$ ways to select $l-r$ nodes out of the set $\{n+2-l, \dots, n-1\}$ (thus leaving $r-2$ for $B = T$), and $\binom{n-j-l}{n-k-(l-r)}$ ways to choose the remaining $n-k-(l-r)$ nodes of T' out of the set $\{j+1, \dots, n-l\}$. Of course this is only valid if $k < n+r-l$. If $k = n+r-l$ we can also cut at $j = n+1-l, \dots, n+r-1-l$ and choose the remaining $k-j-1$ nodes out of the set $\{j+1, \dots, n-1\}$. This is equivalent to building up the tree T' by choosing $n-k$ nodes out of the set $\{j+1, \dots, n-1\}$. The case $k > n+r-l$ is not possible since then $n+1-l$ could not be the r-th largest node in T. If $B = B' = T'$, thus $j \le n-l$, we choose $r-2$ nodes out of the set $\{n+2-l, \dots, n-1\}$ for T'. Hence we have $\binom{n-j-l}{k-r}$ ways to choose the remaining $k-r$ nodes of T' out of the set $\{j+1, \dots, n-l\}$, which gives us the probabilities $\tilde{\mathfrak{p}}_{(n,n+1-l),(k,r)}$. For $\tilde{\mathfrak{q}}_{(n,n+1-l),(k,r)}$ we investigate the cases $B = T$, $B' = T'$ and $B = T'$, $B' = T$. In the first case we choose $l-1-r$ nodes out of the set $\{n+2-l, \dots, n-1\}$ (thus leaving $r-1$ for $B = T$) Further there are $\binom{n-j-l}{n-k-(l-r)}$ ways to choose the remaining $n-k-(l-r)$ nodes of T' out of the set $\{j+1, \dots, n-l\}$. If $k = n+r-l$ we can also cut at $j = n+1-l, \dots, n+r-l$ and choose the remaining $k-j$ nodes out of the set $\{j+1, \dots, n-1\}$. In the second case we have $\binom{l-2}{r-1}$ to choose $r-1$ nodes to make sure that $n+1-l$ is the r-th largest node in T' and $n \notin T'$. Further we have $\binom{n-j-l}{k-r}$ ways to choose the remaining nodes for T'. This proves the shape of $\tilde{\mathfrak{q}}_{(n,n+1-l),(k,r)}$. □

Lemma 10. *The splitting probabilities $p_{n,k}$ and $q_{n,k}$ for isolating the nodes labelled $1, n$ are for $n \ge 2$*

given as follows:

$$p_{n,k} = \frac{k-1}{(n-1)(n-k+1)(n-k)}, \quad q_{n,k} = \frac{1}{(n-1)(n-k)}.$$

Proof. We remove a random edge e of a random size-n recursive tree T_n, thus splitting T_n into two subtrees. We denote with T the subtree of size k, which contains the 1st node after the removal, and with T' the other one. Further we denote with B the arising subtree containing the n-th node. If the node of T incident with e has label j, then the labels of T' must all be larger than j. Now assume that $B = T$. For $1 \le j \le k-1$ we can select the remaining nodes of T in $\binom{n-j-1}{k-j-1}$ ways out of the set $\{j+1, j+2, \dots, n-2, n-1\}$ of nodes, since the nodes $1, \dots, j$ and n are all in T. The nodes which where not selected for T are then used for building up the tree T'. Next assume $B = T'$. For $1 \le j \le k$ we can select the remaining nodes of T in $\binom{n-j-1}{k-j}$ ways. Having selected the nodes for T and T' there are $T_k = (k-1)!$ and $T_{n-k} = (n-k-1)!$ ways to form the trees T and T' so that when the node labelled j in T is attached by edge e to the smallest labelled node in T' the resulting configuration (after an order preserving relabelling) is a random recursive tree of size n. There are $n-1$ ways to select the edge e, further $T_n = (n-1)!$. Thus we get

$$\begin{aligned} p_{n,k} &= \frac{T_k T_{n-k}}{(n-1)T_n} \sum_{j=1}^{k-1} \binom{n-j-1}{k-j-1} = \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!} \binom{n-1}{n-k+1} \\ &= \frac{k-1}{(n-1)(n-k+1)(n-k)}, \end{aligned}$$

and

$$q_{n,k} = \frac{T_k T_{n-k}}{(n-1)T_n} \sum_{j=1}^{k} \binom{n-j-1}{k-j} = \frac{(k-1)!(n-k-1)!}{(n-1)(n-1)!} \binom{n-1}{n-k} = \frac{1}{(n-1)(n-k)},$$

appealing to a well known identity. □

2.4 Isolating the nodes $\boldsymbol{l} = 1, \dots, l$ for fixed l

By further simplifying the probabilities $p_{(n,k),(r,l)}$ we get for $1 \le l \le n$, $1 \le r \le k$, $1 \le k \le n-1$ and $n \ge 2$ the following.

$$p_{(n,l),(k,r)} = \begin{cases} \left[l-1+\frac{n-l+1}{n-k+1}\right] \frac{(k-1)^{\underline{l-1}}}{(n-1)(n-k)(n-1)^{\underline{l-1}}}, & r = l, \\ \left[\binom{l-1}{r} + \binom{l-1}{r-2}\right] \frac{(k-1)^{\underline{r-1}}(n-k-1)^{\underline{l-r-1}}}{(n-1)(n-1)^{\underline{l-1}}}, & r < l. \end{cases} \tag{2.8}$$

Defining for $l \ge 1$ the generating functions

$$M_l(z,v) := \sum_{n \ge l} \sum_{m \ge l-1} (n-1)^{\underline{l-1}} \mathbb{P}\{X_{n;\boldsymbol{l}} = m\} z^{n-l} v^m, \tag{2.9}$$

the recurrence (2.1) can be translated by multiplication with $(n-1)(n-1)^{\underline{l-1}} z^{n-l} v^m$ and summation over $n \ge l$ and $m \ge l-1$ into the following first order differential equation for $M_l(z,v)$ (where the functions $M_r(z,v)$ with $r < l$ are appearing in the inhomogeneous part):

$$(1-z)v \log\Big(\frac{1}{1-z}\Big) \frac{\partial}{\partial z} M_l(z,v) + \Big((l-1) - lv \log\Big(\frac{1}{1-z}\Big)\Big) M_l(z,v) + z(1-v) \frac{\partial}{\partial z} M_l(z,v) = R(z,v), \tag{2.10}$$

with inhomogeneous part

$$R(z,v) = v \sum_{r=1}^{l-1} \left[\binom{l-1}{r} + \binom{l-1}{r-2} \right] M_r(z,v) M_{l-r}(z,v),$$

and initial condition $M_l(0,v) = v^{l-1}(l-1)!$. Note that it is possible to get an explicit solution of (2.10), which is quite involved. In order to prove Theorem 4 we are interested at first at the expectation of $X_{n,\boldsymbol{l}}$, so we differentiate (2.10) once with respect to v and evaluate at $v = 1$ (we are applying the operator $E_v D_v$ to (2.10)). We will use the abbreviation $M_{l,s}(z) = E_v D_v^s M_l(z,v)$. Further we have $M_{l,0}(z) = M_l(z,1) = \frac{(l-1)!}{(1-z)^l}$ and $\frac{\partial}{\partial z} M_{l,0}(z) = \frac{\partial}{\partial z} M_l(z,1) = \frac{l!}{(1-z)^{l+1}}$. We get the following inhomogeneous differential equation for $M_{l,1}(z)$

$$(1-z)\log\big(\frac{1}{1-z}\big)\frac{d}{dz} M_{l,1}(z) + \big((l-1) - l\log\big(\frac{1}{1-z}\big)\big) M_{l,1}(z) = R_{l,1}(z), \tag{2.11}$$

with inhomogeneous part

$$R_{l,1}(z) = \frac{zl!}{(1-z)^{l+1}} + \sum_{r=1}^{l-1} \left[\binom{l-1}{r} + \binom{l-1}{r-2} \right] \Big(\frac{(r-1)!(l-r-1)!}{(1-z)^l} + 2\frac{M_{r,1}(z)(l-r-1)!}{(1-z)^{l-r}} \Big), \tag{2.12}$$

and initial condition $M_{l,1}(0) = \mathbb{E}(X_{l,\boldsymbol{l}})(l-1)! = (l-1)(l-1)!$ The corresponding homogeneous differential equation has the general solution

$$M_{l,1}^{[h]}(z) = \frac{C}{(1-z)^l \log^{l-1}\big(\frac{1}{1-z}\big)}.$$

Variation of the constants method leads then to the particular solution

$$M_{l,1}^{[p]}(z) = \frac{1}{(1-z)^l \log^{l-1}\big(\frac{1}{1-z}\big)} \int_{t=0}^{z} (1-t)^{l-1} \log^{l-2}\big(\frac{1}{1-t}\big) R_{l,1}(t)\,dt, \tag{2.13}$$

and it can be shown by checking the initial values that this particular solution is already the wanted function, i. e. it matches with the initial value, so $M_{l,1}(z) = M_{l,1}^{[p]}(z)$. It remains to show the following asymptotic expansion around the dominant singularity $z = 1$:

$$M_{l,1}(z) = \frac{l!}{(1-z)^{l+1}\log\big(\frac{1}{1-z}\big)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+1}\log^2\big(\frac{1}{1-z}\big)}\Big),$$

which can be done by induction. The case $l = 1$ gives the solution (already computed in [55]):

$$M_{1,1}(z) = \frac{1}{1-z} \int_{t=0}^{z} \frac{t}{(1-t)^2 \log\big(\frac{1}{1-t}\big)}\,dt,$$

and expanding gives the statement for $l = 1$. Assuming that for all $r < l$ the expectations $M_{r,1}(z)$ have the given asymptotic expansion, then we can obtain easily the expansions

$$R_{l,1}(t) = \frac{l!}{(1-t)^{l+1}} + \mathcal{O}\Big(\frac{1}{(1-t)^{l+1}\log\big(\frac{1}{1-t}\big)}\Big),$$

and thus

$$\int_{t=0}^{z} (1-t)^{l-1}\log^{l-2}\big(\frac{1}{1-t}\big) R_{l,1}(t)\,dt = \frac{l!\log^{l-2}\big(\frac{1}{1-z}\big)}{1-z} + \mathcal{O}\Big(\frac{\log^{l-3}\big(\frac{1}{1-z}\big)}{1-z}\Big). \tag{2.14}$$

Consequently we arrive at the following.

$$M_{l,1}(z) = \frac{\int_{t=0}^{z}(1-t)^{l-1}\log^{l-2}\big(\frac{1}{1-t}\big)R_{l,1}(t)dt}{(1-z)^{l}\log^{l-1}\big(\frac{1}{1-z}\big)} = \frac{l!}{(1-z)^{l+1}\log\big(\frac{1}{1-z}\big)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+1}\log^{2}\big(\frac{1}{1-z}\big)}\Big). \tag{2.15}$$

In (2.14) we used Singular differentiation and Integration as stated in Subsection 1.4.1.
Now an application of singularity analysis [25] finishes the prove of the case $s = 1$ of Theorem 4. Now we turn to case $s \geq 2$. We differentiate (2.10) s times with respect to v and evaluate at $v = 1$. We get the following inhomogeneous differential equation for $M_{l,s}(z)$

$$(1-z)\log\big(\frac{1}{1-z}\big)\frac{d}{dz}M_{l,s}(z) + \big((l-1) - l\log\big(\frac{1}{1-z}\big)\big)M_{l,s}(z) = R_{l,s}(z), \tag{2.16}$$

with inhomogeneous part

$$\begin{aligned} R_{l,s}(z) &= sl\log\big(\frac{1}{1-z}\big)M_{l,s-1}(z) - s(1-z)\log\big(\frac{1}{1-z}\big)\frac{d}{dz}M_{l,s-1}(z) + sz\frac{d}{dz}M_{l,s-1}(z) \\ &\quad + \sum_{r=1}^{l-1}\left[\binom{l-1}{r} + \binom{l-1}{r-2}\right]\Big(s\sum_{i=0}^{s-1}\binom{s-1}{i}M_{r,i}(z)M_{l-r,s-1-i}(z) \\ &\quad + \sum_{i=0}^{s}\binom{s}{i}M_{r,i}(z)M_{l-r,s-i}(z)\Big), \end{aligned} \tag{2.17}$$

with initial condition $M_{l,s}(0) = \mathbb{E}(X_{l,\mathbf{l}}^{s})(l-1)! = (l-1)^{\underline{s}}(l-1)!$. Since we already know the shape of the general solution of (2.16), it remains to show the following asymptotic expansion around the dominant singularity $z = 1$:

$$M_{l,s}(z) = \frac{(l+s-1)!}{(1-z)^{l+s}\log^{s}\Big(\frac{1}{1-z}\Big)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s}\log^{s+1}\Big(\frac{1}{1-z}\Big)}\Big),$$

which can be done by induction. The cases $s = 1$, $1 \leq r \leq l$ were computed before and the cases $s > 1, l = 1$ where already shown in [65]. Assuming that for all pairs $(r, j) < (l, s)$, which means either $r \leq l, j < s$ or $r < l, j \leq s$, the functions $M_{r,j}(z)$ have the given asymptotic expansion, then we can obtain the expansion

$$R_{l,s}(t) = \frac{s(l+s-1)!}{(1-t)^{l+s}\log^{s-1}\Big(\frac{1}{1-t}\Big)} + \mathcal{O}\Big(\frac{1}{(1-t)^{l+s}\log^{s}\Big(\frac{1}{1-t}\Big)}\Big),$$

were only the term $sz\frac{d}{dz}M_{l,s-1}(z)$ of (2.17) contributed to the main asymptotic part. Thus we get

$$\int_{t=0}^{z}(1-t)^{l-1}\log^{l-2}\big(\frac{1}{1-t}\big)R_{l,s}(t)dt = \frac{(l+s-1)!\log^{l-s-1}\big(\frac{1}{1-z}\big)}{(1-z)^{s}} + \mathcal{O}\Big(\frac{\log^{l-s-2}\big(\frac{1}{1-z}\big)}{(1-z)^{s}}\Big).$$

Consequently we arrive at the following.

$$M_{l,s}(z) = \frac{\int_{t=0}^{z}(1-t)^{l-1}\log^{l-2}\big(\frac{1}{1-t}\big)R_{l,s}(t)dt}{(1-z)^{l}\log^{l-1}\big(\frac{1}{1-z}\big)} = \frac{(l+s-1)!}{(1-z)^{l+s}\log^{s}\Big(\frac{1}{1-z}\Big)}\Big(1 + \mathcal{O}\Big(\frac{1}{\log\big(\frac{1}{1-z}\big)}\Big)\Big). \tag{2.18}$$

Since an application of singularity analysis [25] provides

$$\mathbb{E}(X_{n,\boldsymbol{l}}^{\underline{s}}) = \frac{n^s}{\log^s n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+1} n}\Big),$$

we get

$$\mathbb{E}(X_{n,\boldsymbol{l}}^{s}) = \mathbb{E}(X_{n,\boldsymbol{l}}^{\underline{s}}) + \mathcal{O}\Big(\frac{n^{s-1}}{\log^s n}\Big) = \frac{n^s}{\log^s n} + \mathcal{O}\Big(\frac{n^s}{\log^{s+1} n}\Big),$$

which finishes the prove of Theorem 4.

2.5 Isolating the pair $1, l$ for fixed l

Now we will briefly show how to prove Theorem 5. Defining for $l \geq 1$ the generating functions

$$A_l(z, v) := \sum_{n \geq l} \sum_{m \geq 0} (n-1)^{\underline{l-1}} \mathbb{P}\{X_{n;1,l} = m\} z^{n-l} v^m, \tag{2.19}$$

and

$$B_l(z, v) := \sum_{n \geq l} \sum_{m \geq 0} (n-1)^{\underline{l-1}} \mathbb{P}\{X_{n,l} = m\} z^{n-l} v^m, \tag{2.20}$$

the recurrence (2.3) can be translated (see (2.8)) by multiplication with $(n-1)(n-1)^{\underline{l-1}} z^{n-l} v^m$ and summation over $n \geq l$ and $m \geq l-1$ into the following first order differential equation for $A_l(z, v)$ (where the functions $A_r(z, v)$ with $r < l$ are appearing in the inhomogeneous part):

$$(1-z) v \log\big(\frac{1}{1-z}\big) \frac{\partial}{\partial z} A_l(z, v) + \big((l-1) - lv \log\big(\frac{1}{1-z}\big)\big) A_l(z, v) + z(1-v) \frac{\partial}{\partial z} A_l(z, v) = R_l(z, v), \tag{2.21}$$

with inhomogeneous part

$$R_l(z, v) = v \sum_{r=2}^{l-1} \binom{l-1}{r-2} \frac{(l-r-1)!}{(1-z)^{l-r}} A_r(z, v) + v \sum_{r=1}^{l-1} \binom{l-1}{r} B_r(z, v) \frac{\partial^{l-r-1}}{\partial z^{l-r-1}} B_1(z, v),$$

and initial values $A_l(0, v) = (l-1)! \sum_{m=0}^{l-1} \mathbb{P}\{X_{l,(1,l)} = m\} v^m$. Again it is possible to get an explicit solution of (2.21). In order to prove Theorem 5 we calculate the expectation of $X_{n,(1,l)}$, so we apply the operator $E_v D_v$ to (2.21). We will use the abbreviation $A_{l,s}(z) = E_v D_v^s A_l(z, v)$ resp. $B_{l,s}(z) = E_v D_v^s B_l(z, v)$. We have $A_{l,0}(z) = B_{l,0}(z) = M_{l,0}(z) = \frac{(l-1)!}{(1-z)^l}$ and $\frac{d}{dz} A_{l,0}(z) = \frac{d}{dz} M_{l,0}(z) = \frac{l!}{(1-z)^{l+1}}$. We already know from [65] and [43] the expansion

$$B_{l,s}(z) = \frac{(l+s-1)!}{(1-z)^{l+s} \log^s\Big(\frac{1}{1-z}\Big)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s} \log^{s+1}\Big(\frac{1}{1-z}\Big)}\Big).$$

We get the following inhomogeneous differential equation for $A_{l,1}(z)$

$$(1-z) \log\big(\frac{1}{1-z}\big) \frac{d}{dz} A_{l,1}(z) + \big((l-1) - l \log\big(\frac{1}{1-z}\big)\big) A_{l,1}(z) = R_{l,1}(z), \tag{2.22}$$

with initial condition $A_{l,1}(0) = \mathbb{E}(X_{l;1,l})(l-1)!$ and where the inhomogeneous part is given by

$$\begin{aligned} R_{l,1}(z) = {} & \frac{zl!}{(1-z)^{l+1}} + \sum_{r=2}^{l-1} \binom{l-1}{r-2} \Big[\frac{(r-1)!(l-r-1)!}{(1-z)^l} + \frac{(l-r-1)!}{(1-z)^{l-r}} A_{r,1}(z)\Big] \\ & + \sum_{r=1}^{l-1} \binom{l-1}{r} \Big[\frac{(r-1)!(l-r-1)!}{(1-z)^l} + B_{r,1}(z) \frac{(l-r-1)!}{(1-z)^{l-r}} + \frac{(r-1)!}{(1-z)^r} \frac{d^{l-r-1}}{dz^{l-r-1}} B_{1,1}(z)\Big]. \end{aligned} \tag{2.23}$$

Now it is obvious that we can use again induction to prove the case $s = 1$ of Theorem 5, since the case $l = 2$ was carried out in [44] and $R_{l,1}$ allows the expansion

$$R_{l,1}(t) = \frac{l!}{(1-t)^{l+1}} + \mathcal{O}\Big(\frac{1}{(1-t)^{l+1}\log\big(\frac{1}{1-t}\big)}\Big).$$

For $s \geq 2$ it is sufficient to see that

$$R_{l,s}(t) = \frac{s(l+s-1)!}{(1-t)^{l+s}\log^{s-1}\Big(\frac{1}{1-t}\Big)} + \mathcal{O}\Big(\frac{1}{(1-t)^{l+s}\log^{s}\Big(\frac{1}{1-t}\Big)}\Big),$$

where only the term $sz\frac{d}{dz}A_{l,s-1}(z)$ contributes to the dominant asymptotic part. Thus singularity analysis finishes the proof of Theorem 5.

2.6 Isolating the nodes $n+1-l,\dots,n$ for fixed l

We further simplify the binomial coefficients of the splitting probability given by Lemma 3 using (1.5), which leads to

$$\begin{aligned}\mathfrak{p}_{(n,l),(k,r)} &= [\![k \leq n+r-l]\!]\frac{\binom{l-1}{r-1}}{(n-1)(n-1)^{\underline{l-1}}}\big[(k-1)^{\underline{r-2}}(n-k-1)^{\underline{l-r}} + (k-1)^{\underline{r}}(n-k-1)^{\underline{l-r-2}}\big]\\ &\quad + [\![k = n+r-l]\!]\binom{l}{r-1}\frac{(l-r-1)!}{(n-1)(n-1)^{\underline{l-r}}}, \quad r \leq l,\end{aligned} \tag{2.24}$$

where we use the convention $(j-1)^{\underline{-p}} := (j^{\overline{p}})^{-1}$, $p \in \mathbb{N}$, see e. g. [30]. For $l \geq 4$ we always have to distinguish between the four cases $r = 1$, $1 < r < l-1$, $r = l-1$ and $r = l$ when translating (2.2) into a differential equation. Defining for $l \geq 1$ the generating functions

$$N_l(z,v) := \sum_{n\geq l}\sum_{m\geq l-1}(n-1)^{\underline{l-2}}\,\mathbb{P}\{Y_{n,\boldsymbol{l}} = m\}z^{n+1-l}v^m, \tag{2.25}$$

where for $l = 1$ we have $(n-1)^{\underline{-1}} = \frac{1}{n}$, the recurrence (2.2) can be translated by multiplication with $(n-1)(n-1)^{\underline{l-1}}z^{n-l}v^m$ and summation over $n \geq l$ and $m \geq l-1$ into a second order differential equation for $N_l(z,v)$ (where the functions $N_r(z,v)$ with $r < l$ are appearing in the inhomogeneous part). Since $(n-1)^{\underline{l-1}}/(n-1)^{\underline{l-r}} = (n+r-l-1)^{\underline{r-1}}$, we get

$$(l-1)\frac{\partial}{\partial z}N_l(z,v) - v\frac{1}{1-z}N_l(z,v) + v(1-z)\log\Big(\frac{1}{1-z}\Big)\frac{\partial^2}{\partial z^2}N_l(z,v) + (1-v)z\frac{\partial^2}{\partial z^2}N_l(z,v) = R_l(z,v), \tag{2.26}$$

with initial conditions $N_l(0,v) = 0$, $\big(\frac{\partial}{\partial z}N_l(z,v)\big)|_{z=0} = (l-1)!(l-1)v^{l-1}$ and inhomogeneous part

$$\begin{aligned}R_l(z,v) &= v\sum_{r=1}^{l-1}\binom{l-1}{r-1}\Big(N_r(z,v)\frac{\partial^2}{\partial z^2}N_{l-r}(z,v) + N_{l-r}(z,v)\frac{\partial^2}{\partial z^2}N_r(z,v)\Big)\\ &\quad + \sum_{r=1}^{l-1}\binom{l}{r-1}(l-r-1)!v^{l-r}\frac{\partial}{\partial z}N_r(z,v)\\ &= v\sum_{r=1}^{l-1}\binom{l}{r}N_r(z,v)\frac{\partial^2}{\partial z^2}N_{l-r}(z,v) + \sum_{r=1}^{l-1}\binom{l}{r-1}(l-r-1)!v^{l-r}\frac{\partial}{\partial z}N_r(z,v)\end{aligned} \tag{2.27}$$

To prove Theorem 6, we have to show that $N_{l,s}(z)$ admits the following expansion

$$N_{l,s}(z) = \frac{l(l+s-2)!}{(s+l)(1-z)^{l+s-1}\log^{s}\left(\frac{1}{1-z}\right)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s-1}\log^{s+1}\left(\frac{1}{1-z}\right)}\Big), \tag{2.28}$$

where we use the abbreviation $N_{l,s}(z) = E_v D_v^s N_l(z,v)$, which will be done by induction with respect to l and s.

First we turn to $s=1$. For the getting the expectation we differentiate once with respect to v and evaluate at $v=1$. Further we have $N_{1,0}(z) = \log\left(\frac{1}{1-z}\right)$, $N_{l,0}(z) = N_l(z,1) = \frac{(l-2)!}{(1-z)^{l-1}} - (l-2)!$ for $l \geq 2$, $\frac{d}{dz}N_{l,0}(z) = M_{l,0}(z) = \frac{(l-1)!}{(1-z)^l}$ and $\frac{d^2}{dz^2}N_{l,0}(z) = \frac{d}{dz}M_{l,0}(z) = \frac{l!}{(1-z)^{l+1}}$. This leads to the following.

$$(l-1)\frac{d}{dz}N_{l,1}(z) - \frac{1}{1-z}N_{l,1}(z) + (1-z)\log\left(\frac{1}{1-z}\right)\frac{d^2}{dz^2}N_{l,1}(z) = R_{l,1}(z), \tag{2.29}$$

with inhomogeneous part

$$\begin{aligned} R_{l,1}(z) &= \frac{1}{1-z}N_{l,0}(z) - (1-z)\log\left(\frac{1}{1-z}\right)\frac{d^2}{dz^2}N_{l,0}(z) + z\frac{d^2}{dz^2}N_{l,0}(z) \\ &+ \sum_{r=1}^{l-1}\binom{l}{r}\Big(N_{r,0}(z)\frac{d^2}{dz^2}N_{l-r,0}(z) + N_{r,1}(z)\frac{d^2}{dz^2}N_{l-r,0}(z) + N_{r,0}(z)\frac{d^2}{dz^2}N_{l-r,1}(z)\Big) \\ &+ \sum_{r=1}^{l-1}\binom{l}{r-1}(l-r-1)!\Big((l-r)\frac{d}{dz}N_{r,0}(z) + \frac{d}{dz}N_{r,1}(z)\Big), \end{aligned} \tag{2.30}$$

and initial conditions $N_{l,1}(0) = 0$, $\big(\frac{d}{dz}N_{l,1}(z)\big)|_{z=0} = (l-1)!\mathbb{E}(Y_{l,\boldsymbol{l}}) = (l-1)(l-1)!$. The corresponding homogeneous differential equation has the general solution

$$\begin{aligned} N_{l,1}^{[h]}(z) &= C_1 N_{l,1}^{[h_1]}(z) + C_2 N_{l,1}^{[h_2]}(z) \\ &= C_1\big(l-1+\log\big(\frac{1}{1-z}\big)\big) + C_2\big(l-1+\log\big(\frac{1}{1-z}\big)\big)\int_\alpha^z \frac{dt}{\log^{l-1}\left(\frac{1}{1-t}\right)\left[l-1+\log\left(\frac{1}{1-t}\right)\right]^2}, \end{aligned} \tag{2.31}$$

where we may choose any real $0 < \alpha < 1$. This leads to the particular solution

$$N_{l,1}^{[p]}(z) = N_{l,1}^{[h_1]}(z)\int_0^z \frac{-b_l(t)N_{l,1}^{[h_2]}(t)}{D_l(t)}dt + N_{l,1}^{[h_2]}(z)\int_0^z \frac{b_l(t)N_{l,1}^{[h_1]}(t)}{D_l(t)}dt, \tag{2.32}$$

where $b_l(z) = \frac{R_{l,1}(z)}{(1-z)\log\left(\frac{1}{1-z}\right)}$ and

$$D_l(z) = N_{l,1}^{[h_1]}(z)\frac{\partial}{\partial z}N_{l,1}^{[h_2]}(z) - N_{l,1}^{[h_2]}(z)\frac{\partial}{\partial z}N_{l,1}^{[h_1]}(z) = \frac{1}{\log^{l-1}\left(\frac{1}{1-z}\right)} \tag{2.33}$$

is the Wronski determinant of the two homogeneous solutions $N_{l,1}^{[h_1]}(z)$ and $N_{l,1}^{[h_2]}(z)$. Using integration by parts we can simplify the particular solution as follows:

$$\begin{aligned} N_{l,1}^{[p]}(z) &= \Big(l-1+\log\big(\frac{1}{1-z}\big)\Big)\times \\ &\times \int_0^z\Big(\int_0^t \frac{\log^{l-2}\left(\frac{1}{1-u}\right)\left(l-1+\log\left(\frac{1}{1-u}\right)\right)R_{l,1}(u)}{1-u}du\Big)\frac{1}{\log^{l-1}\left(\frac{1}{1-t}\right)\left(l-1+\log\left(\frac{1}{1-z}\right)\right)^2}dt. \end{aligned} \tag{2.34}$$

It turns out that the particular solution (2.34) also satisfies the initial conditions of (2.29) and is thus the required solution, i. e., $N_{l,1}(z) = N_{l,1}^{[p]}(z)$. Assuming that the induction hypothesis (2.28) holds for $s = 1$ and $1 \leq k \leq l-1$ the main term of $R_{l,1}(z)$ is given by

$$\begin{aligned} R_{l,1}(z) &= z\frac{d^2}{dz^2}N_{l,0}(z) + N_{1,0}(z)\frac{d^2}{dz^2}N_{l-1,1}(z) \\ &= \frac{l!}{(1-z)^{l+1}} + \binom{l}{1}\frac{(l-1)(l-1)!}{(1-z)^{l+1}} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+1}\log\left(\frac{1}{1-z}\right)}\Big) \\ &= \frac{l\cdot l!}{(1-z)^{l+1}} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+1}\log\left(\frac{1}{1-z}\right)}\Big). \end{aligned} \tag{2.35}$$

By plugging (2.35) into (2.34) and using Singular integration one obtains

$$N_{l,1}(z) = \frac{l\cdot l!}{l(l+1)(1-z)^l\log\left(\frac{1}{1-z}\right)} + \mathcal{O}\Big(\frac{1}{(1-z)^l\log^2\left(\frac{1}{1-z}\right)}\Big). \tag{2.36}$$

Thus the induction step is completed and the part $s = 1$, $l \geq 1$ is proven for Theorem 6. Now we turn to the cases $s > 1, l > 1$, because the case $l = 1$, $s \geq 1$ was already shown in [43]. Applying the operator $E_v D_v^s$ to (2.26) gives

$$(l-1)\frac{d}{dz}N_{l,s}(z) - \frac{1}{1-z}N_{l,s}(z) + (1-z)\log\left(\frac{1}{1-z}\right)\frac{d^2}{dz^2}N_{l,s}(z) = R_{l,s}(z), \tag{2.37}$$

with inhomogeneous part

$$\begin{aligned} R_{l,s}(z) = {} & \frac{s}{1-z}N_{l,s-1}(z) - s(1-z)\log\left(\frac{1}{1-z}\right)\frac{d^2}{dz^2}N_{l,s-1}(z) + sz\frac{d^2}{dz^2}N_{l,s-1}(z) \\ & + \sum_{r=1}^{l-1}\binom{l}{r}\Big(\sum_{j=0}^{s}\binom{s}{j}N_{r,j}(z)\frac{d^2}{dz^2}N_{l-r,s-j}(z) + \sum_{j=0}^{s-1}\binom{s-1}{j}N_{r,j}(z)\frac{d^2}{dz^2}N'_{l-r,s-1-j}(z)\Big) \\ & + \sum_{r=1}^{l-1}\binom{l}{r-1}(l-r-1)!\sum_{j=0}^{s}\binom{s}{j}(l-r)^{\underline{j}}\frac{d}{dz}N_{r,s-j}(z,v), \end{aligned} \tag{2.38}$$

and initial conditions $N_{l,s}(0) = 0$, $\big(\frac{d}{dz}N_{l,s}(z)\big)|_{z=0} = (l-1)!\mathbb{E}(Y_{l,\boldsymbol{l}}^{\underline{s}}) = (l-1)^{\underline{s}}(l-1)!$. Since the homogeneous solution corresponding to (2.37) was already computed before we obtain again a particular solution by applying the method of variation of constants. We get after simplifications

$$\begin{aligned} N_{l,s}^{[p]}(z) = {} & \Big(l-1+\log\big(\frac{1}{1-z}\big)\Big)\times \\ & \times\int_0^z\Big(\int_0^t\frac{\log^{l-2}\left(\frac{1}{1-u}\right)\left(l-1+\log\left(\frac{1}{1-u}\right)\right)R_{l,s}(u)}{1-u}du\Big)\frac{1}{\log^{l-1}\left(\frac{1}{1-t}\right)\left(l-1+\log\left(\frac{1}{1-z}\right)\right)^2}dt. \end{aligned} \tag{2.39}$$

Assuming that for all pairs $(r, j) < (l, s)$, which means either $r \leq l, j < s$ or $r < l, j \leq s$ the functions $N_{r,j}(z)$ have the given asymptotic expansion (2.28), then we can obtain the expansion

$$\begin{aligned} R_{l,s}(z) &= sz\frac{d^2}{dz^2}N_{l,s-1}(z) + \binom{l}{1}\binom{s}{0}N_{1,0}(z)\frac{d^2}{dz^2}N_{l-1,s}(z) + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s}\log^s\left(\frac{1}{1-z}\right)}\Big) \\ &= \frac{l(l+s-1)!}{(1-z)^{l+s}\log^{s-1}\left(\frac{1}{1-z}\right)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s}\log^s\left(\frac{1}{1-z}\right)}\Big). \end{aligned} \tag{2.40}$$

By plugging (2.40) into (2.39) and using Singular integration we get the following result.

$$N_{l,s}(z) = \frac{l(l+1-s)!}{(l+s)(l+s-1)(1-z)^{l+s-1}\log^s\left(\frac{1}{1-z}\right)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s-1}\log^{s+1}\left(\frac{1}{1-z}\right)}\Big). \tag{2.41}$$

This finishes the proof of Theorem 6.

2.7 Isolating the pair $n+1-l, n$ for fixed l

Again we will only outline the proof very briefly since it is basically the same as for Theorem 6. Defining for $l \geq 1$ the generating functions

$$\begin{aligned} F_l(z,v) &:= \sum_{n\geq l}\sum_{m\geq l-1}(n-1)^{\underline{l-2}}\mathbb{P}\{X_{n,(n+1-l,n)} = m\}z^{n-l+1}v^m, \\ G_l(z,v) &:= \sum_{n\geq l}\sum_{m\geq l-1}(n-1)^{\underline{l-2}}\mathbb{P}\{X_{n,n+1-l} = m\}z^{n-l+1}v^m, \end{aligned} \tag{2.42}$$

the recurrence (2.4) can be translated (see (2.24)) by multiplication with $(n-1)(n-1)^{\underline{l-1}}z^{n-l}v^m$ and summation over $n \geq l$ and $m \geq 1$ into the following first order differential equation for $F_l(z,v)$ (where the functions $F_r(z,v)$ with $r < l$ are appearing in the inhomogeneous part):

$$(l-1)\frac{d}{dz}F_l(z,v) - v\frac{1}{1-z}F_l(z,v) + v(1-z)\log\left(\frac{1}{1-z}\right)\frac{d^2}{dz^2}F_l(z,v) + (1-v)z\frac{d^2}{dz^2}F_l(z,v) = R_l(z,v), \tag{2.43}$$

with initial condition $F_l(0,v) = 0$ and where the inhomogeneous part is given by

$$\begin{aligned} R_l(z,v) = {} & v\sum_{r=2}^{l-1}\binom{l-2}{r-2}F_r(z,v)\frac{(l-r)!}{(1-z)^{l-r+1}} + v\sum_{r=2}^{l-2}\binom{l-2}{r-2}\frac{(l-r-2)!}{(1-z)^{l-r-1}}\frac{\partial^2}{\partial z^2}F_r(z,v) \\ & + v\binom{l-2}{l-3}\log\big(\frac{1}{1-z}\big)\frac{\partial^2}{\partial z^2}F_{l-1}(z,v) + v\sum_{r=2}^{l-1}\binom{l-1}{l-r+1}(l-r-1)!\frac{\partial}{\partial z}F_r(z,v) \\ & + v\sum_{r=1}^{l-1}\binom{l-2}{r-1}\Big(G_r(z,v)\frac{d^{l-r+1}}{dz^{l-r+1}}G_1(z,v) + \frac{\partial^2}{\partial z^2}G_r(z,v)\frac{\partial^{l-r-1}}{\partial z^{l-1-r}}G_1(z,v)\Big) \\ & + \sum_{r=1}^{l-1}\binom{l-1}{l-r}(l-r-1)!\sum_{i=0}^{l-r-1}v^{i+1}\frac{\partial}{\partial z}G_r(z,v)\mathbb{P}\{X_{l-r,l-r} = i\} \end{aligned} \tag{2.44}$$

We will use the abbreviation $F_{l,s}(z) = E_vD_v^sF_l(z,v)$. In order to prove Theorem 7 we have to show that $F_{l,s}(z)$ admits the following expansion.

$$F_{l,s}(z) = \frac{2(l+s-2)!}{(s+2)(1-z)^{s+l-1}\log^s\left(\frac{1}{1-z}\right)} + \mathcal{O}\big(\frac{1}{(1-z)^{s+1}\log^{s+1}\left(\frac{1}{1-z}\right)}\big). \tag{2.45}$$

First we turn to the case $s = 1$, $l \geq 2$. For the getting the expectation we differentiate once with respect to v and evaluate at $v = 1$.

$$(l-1)\frac{d}{dz}F_{l,1}(z) - \frac{1}{1-z}F_{l,1}(z) + (1-z)\log\left(\frac{1}{1-z}\right)\frac{d^2}{dz^2}F_{l,1}(z) = R_{l,1}(z), \tag{2.46}$$

where the inhomogeneous part is given by

$$
\begin{aligned}
R_{l,1}(z) &= z\frac{d^2}{dz^2}F_{l,0}(z) + \frac{1}{1-z}F_{l,0}(z) - (1-z)\log\left(\frac{1}{1-z}\right)\frac{d^2}{dz^2}F_{l,0}(z) \\
&+ \sum_{r=2}^{l-1}\binom{l-2}{r-2}F_{r,0}(z)\frac{(l-r)!}{(1-z)^{l-r+1}} + \sum_{r=2}^{l-1}\binom{l-2}{r-2}F_{r,1}(z)\frac{(l-r)!}{(1-z)^{l-r+1}} \\
&+ \sum_{r=2}^{l-2}\binom{l-2}{r-2}\frac{(l-r-2)!}{(1-z)^{l-r-1}}\frac{d^2}{dz^2}F_{r,0}(z) + \sum_{r=2}^{l-2}\binom{l-2}{r-2}\frac{(l-r-2)!}{(1-z)^{l-r-1}}\frac{d^2}{dz^2}F_{r,1}(z) \\
&+ \binom{l-2}{l-3}\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}F_{l-1,1}(z) + \binom{l-2}{l-3}\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}F_{l-1,0}(z) \\
&+ \sum_{r=2}^{l-1}\binom{l-1}{l-r+1}(l-r-1)!\frac{d}{dz}F_{r,0}(z) + \sum_{r=2}^{l-1}\binom{l-1}{l-r+1}(l-r-1)!\frac{d}{dz}F_{r,1}(z) \\
&+ \sum_{r=1}^{l-1}\binom{l-2}{r-1}\Big(G_{r,0}(z)\frac{d^{l-r+1}}{dz^{l-r+1}}G_{1,0}(z) + \frac{d^2}{dz^2}G_{r,0}(z)\cdot\frac{d^{l-r-1}}{dz^{l-1-r}}G_{1,0}(z)\Big) \\
&+ \sum_{r=1}^{l-1}\binom{l-2}{r-1}\Big(G_{r,0}(z)\frac{d^{l-r+1}}{dz^{l-r+1}}G_{1,1}(z) + G_{r,1}(z)\frac{d^{l-r+1}}{dz^{l-r+1}}G_{1,0}(z)\Big) \\
&+ \sum_{r=1}^{l-1}\binom{l-2}{r-1}\Big(\frac{d^2}{dz^2}G_{r,0}(z)\cdot\frac{d^{l-r-1}}{dz^{l-r-1}}G_{1,1}(z) + \frac{d^2}{dz^2}G_{r,1}(z)\cdot\frac{d^{l-r-1}}{dz^{l-r-1}}G_{1,0}(z)\Big) \\
&+ \sum_{r=1}^{l-1}\binom{l-1}{l-r}(l-r-1)!\sum_{i=0}^{l-r-1}\Big((i+1)\frac{d}{dz}G_{r,0}(z)\mathbb{P}\{X_{l-r,l-r}=i\} + \frac{d}{dz}G_{r,1}(z)\mathbb{P}\{X_{l-r,l-r}=i\}\Big).
\end{aligned} \tag{2.47}
$$

We use the following expansion of $G_{r,j}(z)$, $j \geq 1$, around the dominant singularity $z = 1$:

$$
G_{r,j}(z) = \frac{(r+j-2)!}{(j+1)(1-z)^{r+j-1}\log^j\left(\frac{1}{1-z}\right)} + \mathcal{O}\Big(\frac{1}{(1-z)^{r+j-1}\log^{j+1}\left(\frac{1}{1-z}\right)}\Big), \tag{2.48}
$$

as shown in Chapter 1([44]). We have already proven Theorem 7 for $l = 2$, $s \geq 1$ in Theorem 6 before and in Chapter 1([44]). Under the induction hypothesis that (2.45) holds for all $2 \leq r \leq l-1$, with $s = 1$, we get the expansion

$$
\begin{aligned}
R_{l,1}(z) &= z\frac{d^2}{dz^2}F_{l,0}(z) + \binom{l-2}{l-3}\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}F_{l-1,1}(z) + \binom{l-2}{l-2}G_{l-1,1}(z)G_{1,0}(z) \\
&\quad + \binom{l-2}{0}G_{1,0}(z)\frac{d^l}{dz^l}G_{1,1}(z) \\
&= \frac{l!}{(1-z)^{l+1}} + \frac{2l!(l-2)}{3(1-z)^{l+1}} + \frac{2l!}{2(1-z)^{l+1}} + \mathcal{O}\big(\frac{1}{(1-z)^{l+1}\log\left(\frac{1}{1-z}\right)}\big) \\
&= \frac{2l!(l+1)}{3(1-z)^{l+1}} + \mathcal{O}\big(\frac{1}{(1-z)^{l+1}\log\left(\frac{1}{1-z}\right)}\big).
\end{aligned} \tag{2.49}
$$

Thus we get the solution by plugging the expansion above into (2.32) and using (2.35). We obtain the following.

$$
F_{l,1}(z) = \frac{2l!(l+1)}{3l(l+1)(1-z)^l\log\left(\frac{1}{1-z}\right)} + \mathcal{O}\big(\frac{1}{(1-z)^l\log^2\left(\frac{1}{1-z}\right)}\big). \tag{2.50}
$$

Thus an application of singularity analysis proves the part $s = 1$ of Theorem 7. For $s > 1$ we will use induction with respect to s and l.

$$(l-1)\frac{d}{dz}F_{l,s}(z) - \frac{1}{1-z}F_{l,s}(z) + (1-z)\log\left(\frac{1}{1-z}\right)\frac{d^2}{dz^2}F_{l,s}(z) = R_{l,s}(z), \tag{2.51}$$

where the inhomogeneous part is given by

$$\begin{aligned}
R_{l,1}(z) = {} & sz\frac{d^2}{dz^2}F_{l,s-1}(z) + \frac{s}{1-z}F_{l,s-1}(z) - s(1-z)\log\left(\frac{1}{1-z}\right)\frac{d^2}{dz^2}F_{l,s-1}(z) \\
& + s\sum_{r=2}^{l-1}\binom{l-2}{r-2}F_{r,s-1}(z)\frac{(l-r)!}{(1-z)^{l-r+1}} \\
& + \sum_{r=2}^{l-1}\binom{l-2}{r-2}F_{r,s}(z)\frac{(l-r)!}{(1-z)^{l-r+1}} + s\sum_{r=2}^{l-2}\binom{l-2}{r-2}\frac{(l-r-2)!}{(1-z)^{l-r-1}}\frac{d^2}{dz^2}F_{r,s-1}(z) \\
& + \sum_{r=2}^{l-2}\binom{l-2}{r-2}F''_{r,s}(z)\frac{(l-r-2)!}{(1-z)^{l-r-1}} + s\binom{l-2}{l-3}\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}F_{l-1,s-1}(z) \\
& + \binom{l-2}{l-3}\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}F_{l-1,s}(z) + s\sum_{r=2}^{l-1}\binom{l-1}{l-r+1}(l-r-1)!\frac{d}{dz}F_{r,s-1}(z) \\
& + \sum_{r=2}^{l-1}\binom{l-1}{l-r+1}(l-r-1)!\frac{d}{dz}F_{r,s}(z) \\
& + \sum_{r=1}^{l-1}\binom{l-2}{r-1}\sum_{j=0}^{s-1}\binom{s-1}{j}\Big(G_{r,j}(z)\frac{d^{l-r+1}}{dz^{l-r+1}}G_{1,s-1-j}(z) + \frac{d^2}{dz^2}G_{r,j}(z)\frac{d^{l-r-1}}{dz^{l-1-r}}G_{1,s-1-j}(z)\Big) \\
& + \sum_{r=1}^{l-1}\binom{l-2}{r-1}\sum_{j=0}^{s}\binom{s}{j}\Big(G_{r,j}(z)\frac{d^{l-r+1}}{dz^{l-r+1}}G_{1,s-j}(z) + \frac{d^2}{dz^2}G_{r,j}(z)\frac{d^{l-r-1}}{dz^{l-r-1}}G_{1,s-j}(z)\Big) \\
& + \sum_{r=1}^{l-1}\binom{l-1}{l-r}(l-r-1)!\sum_{i=0}^{l-r-1}\sum_{j=0}^{i+1}\binom{s}{j}\ (i+1)^{\underline{j}}\frac{d}{dz}G_{r,s-j}(z)\mathbb{P}\{X_{l-r,l-r} = i\}.
\end{aligned} \tag{2.52}$$

Assuming that for all pairs $(r,j) < (l,s)$, which means either $r \le l, j < s$ or $r < l, j \le s$, the functions $G_{r,j}(z)$ have the given asymptotic expansion, then we can obtain the expansion

$$\begin{aligned}
R_{l,s}(z) &= szF''_{l,s-1}(z) + \binom{l-2}{l-3}\log\big(\frac{1}{1-z}\big)\frac{d^2}{dz^2}F_{l-1,s}(z) + G_{1,0}(z)\frac{\partial^l}{\partial z^l}G_{1,s}(z) + \frac{d^2}{dz^2}G_{l-1,s}(z)G_{1,0}(z) \\
&= \frac{\left(\frac{2s}{s+1} + \frac{2(l-2)}{s+2} + \frac{2}{s+1}\right)(l+s-1)!}{(1-z)^{l+s}\log^{s-1}\left(\frac{1}{1-z}\right)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s}\log^{s}\left(\frac{1}{1-z}\right)}\Big) \\
&= \frac{2(l+s)(l+s-1)!}{(s+2)(1-z)^{l+s}\log^{s-1}\left(\frac{1}{1-z}\right)} + \mathcal{O}\Big(\frac{1}{(1-z)^{l+s}\log^{s}\left(\frac{1}{1-z}\right)}\Big).
\end{aligned} \tag{2.53}$$

Thus we get the solution by plugging the expansion above into (2.32) and using (2.35).

$$F_{l,s}(z) = \frac{2(l+s)(l+s-1)!}{(s+2)(l+s)(l+s-1)(1-z)^{l+s-1}\log^{s}\left(\frac{1}{1-z}\right)} + \mathcal{O}\big(\frac{1}{(1-z)^{s+l-1}\log^{s+1}\left(\frac{1}{1-z}\right)}\big). \tag{2.54}$$

This completes the proof of Theorem 7.

2.8 Isolating the pair $1, n$

We will use a generating functions approach to get the desired result. We define the bivariate generating function

$$P(z,v) := \sum_{n\geq 2}\sum_{m\geq 1} \mathbb{P}\{X_{n;1,n)} = m\} z^{n-1} v^m. \tag{2.55}$$

Further we use the generating functions $M_1(z,v)$ and $N_1(z,v)$ as defined in (2.9) and (2.25). By multiplying (2.5) with $(n-1)z^{n-1}v^m$ and summing up over $n \geq 2$, $m \geq 1$ we get the following functional equation.

$$z\frac{\partial}{\partial z}P(z,v) = vz\log\Big(\frac{1}{1-z}\Big)\frac{\partial}{\partial z}P(z,v) - v\log\Big(\frac{1}{1-z}\Big)\frac{\partial}{\partial z}P(z,v) + vz\frac{\partial}{\partial z}P(z,v) + vM_1(z,v)N_1(z,v). \tag{2.56}$$

This leads to

$$\frac{\partial}{\partial z}P(z,v) = \frac{vM(z,v)N(z,v)}{z(1-v)+v(1-z)\log\big(\frac{1}{1-z}\big)}. \tag{2.57}$$

We already know asymptotic equivalents of $M_{1,s}(z) = E_v D_v^s M_1(z,v)$ and $N_{s,1}(z) = E_v D_v^s N_1(z,v)$. Now we will compute asymptotical results for the factorial moments and deduce then the ordinary moments of $X_{n;1,n}$. The factorial moments $\mathbb{E}(X_{n;1,n}^{\underline{s}})$ can be obtained by extracting coefficients from $E_v D_v^s \frac{\partial}{\partial z}P(z,v)$:

$$\mathbb{E}\big(X_{n;1,n}^{\underline{s}}\big) = \frac{1}{n-1}[z^{n-2}]E_v D_v^s \frac{\partial}{\partial z}P(z,v). \tag{2.58}$$

First we need the following result.

$$\begin{aligned}
E_v D_v^s \frac{\partial}{\partial z}P(z,v) &= E_v D_v^s \frac{vM_1(z,v)N_1(z,v)}{z(1-v)+v(1-z)\log\big(\frac{1}{1-z}\big)} \\
&= \sum_{j=0}^{s}\binom{s}{j} E_v D_v^j\Big(vM_1(z,v)N_1(z,v)\Big) E_v D_v^{s-j}\frac{1}{z(1-v)+v(1-z)\log\big(\frac{1}{1-z}\big)} \\
&= \sum_{j=0}^{s}\binom{s}{j}\Big[\sum_{l=0}^{j}\binom{j}{l} E_v D_v^l\Big(vM_1(z,v)\Big)E_v D_v^{l-j}N_1(z,v)\Big]\times \\
&\qquad \times E_v\frac{(-1)^{s-j}(s-j)!(-z+(1-z)\log\big(\frac{1}{1-z}\big))^{s-j}}{\Big(z(1-v)+v(1-z)\log\big(\frac{1}{1-z}\big)\Big)^{s-j+1}} \\
&= \sum_{j=0}^{s}\binom{s}{j}\Big[\sum_{l=0}^{j}\binom{j}{l}\Big(lM_{1,l-1}(z)+vM_{l,1}(z)\Big)N_{1,l-j}(z)\Big]\times \\
&\qquad \times \frac{(-1)^{s-j}(s-j)!(-z+(1-z)\log\big(\frac{1}{1-z}\big))^{s-j}}{\Big((1-z)\log\big(\frac{1}{1-z}\big)\Big)^{s-j+1}}.
\end{aligned} \tag{2.59}$$

Using the expansions known from [65] and Chapter 1 together with $M_{1,0}(z) = M_1(z,1) = \frac{1}{1-z}$ and $N_{1,0}(z) = N_1(z,1) = \log\big(\frac{1}{1-z}\big)$ we see that the main asymptotic contribution in (2.59) comes from the terms

$$\sum_{j=0}^{s}\binom{k}{j}M_{1,j}(z)N_{1,0}(z)\frac{z^{s-j}(s-j)!}{\Big((1-z)\log\big(\frac{1}{1-z}\big)\Big)^{s-j+1}} = \sum_{j=0}^{s}\binom{s}{j}\frac{M_{1,j}(z)z^{s-j}(s-j)!}{(1-z)^{s-j+1}\Big(\log\big(\frac{1}{1-z}\big)\Big)^{s-j}}. \tag{2.60}$$

Thus an application of singularity analysis (see [25]) leads to Theorem 8.

2.9 Comments

2.9.1 Searching the limit distribution for nodes with small labels

Unfortunately it seems to be difficult to derive a non degenerate limit distribution for the random variables $X_{n;\boldsymbol{l}}$. Note that even in the case $l = 1$, e.g. the random variable $X_{n;1}$, no non-degenerate limit law is known. Although in the case of $X_{n;1}$, one can derive the bivariate generating function explicitly, the limit law seems to out of reach at present.

Lemma 11.

$$M_1(z,v) = e^{\displaystyle\int_{t=0}^{z} \frac{v\log\left(\frac{1}{1-t}\right)}{t - v\Big(t-(1-t)\log\left(\frac{1}{1-t}\right)\Big)}dt}. \tag{2.61}$$

Proof. The splitting probabilities $p_{(n,1),(k,1)}$ for isolating the root were already calculated in [55]:

$$p_{(n,1),(k,1)} = \frac{1}{(n-1)(n-k)(n-k+1)}, \quad 1 \le k \le n, \quad n \ge 2. \tag{2.62}$$

This immediately gives the following recurrence:

$$\mathbb{P}\{X_{n,1} = m\} = \sum_{k=1}^{n-1} \frac{n}{(n-1)(n-k+1)(n-k)}\mathbb{P}\{X_{k,1} = m-1\} \quad \text{for } n \ge 2 \text{ and } m \ge 1 \tag{2.63}$$

with initial value $\mathbb{P}\{X_{1,1} = 0\} = 1$. Using the bivariate generating function $M_1(z,v)$ above recurrence translates by multiplying with $(n-1)z^{n-1}v^m$ and summing up for $n \ge 2$ and $m \ge 1$ into the following first order linear differential equation:

$$\frac{\partial}{\partial z}M_1(z,v) = \frac{v\log\left(\frac{1}{1-z}\right)}{z - v\Big(z-(1-z)\log\left(\frac{1}{1-z}\right)\Big)}M_1(z,v), \tag{2.64}$$

with initial condition $M_1(0,v) = 1$. Solving this differential equation finishes the proof of Lemma 11. □

Note that the function appearing in the integrand of $M_1(z,v)$ is itself a bivariate p.g.f.,

$$\mathcal{M}(z,v) = \frac{v\log\left(\frac{1}{1-z}\right)}{t - v\Big(z-(1-z)\log\left(\frac{1}{1-z}\right)\Big)}, \tag{2.65}$$

which appears if one introduces a slightly modified destruction procedure for recursive trees by counting the number of label comparisons until the root is found.

2.9.2 Isolating nodes by node removal

Instead of removing random edges, randomly selected nodes are removed now. The procedure works as follows. In a size n recursive tree we select at random a node and look if it's label equals one (if we have find the root). If we have picked the root we have done one label comparison and the procedure stops. If not, we remove the selected node together with the subtree rooted at the chosen node and continue the procedure in the subtree containing the original root until the root is found. It turns out that for root isolation the procedure behaves the same way as if for the edge-cutting procedure. For isolating several nodes it should be the same. This node removal procedure was described by Janson [37] where it was shown that indeed the two procedures obey asymptotically the same distribution law.

We denote with $\mathcal{X}_{n,1}$ the random variable counting the number of necessary label comparisons until the root is found in a random size n recursive tree.

The splitting probability $\pi_{n,k}$, which is the probability that after choosing a random node and comparing its label to the root's label and discarding the subtree rooted at the selected node, if the label is not equal to 1, the remaining subtree containing the root is of size k. It is clear, that $\pi_{n,0} = 1/n$ since we pick the root with probability $1/n$, and then there is no tree left containing the root. With probability $(n-1)/n$ we choose another node and remove it. This removal is equivalent to a random edge cut. Since we already know that

$$p_{(n,1),(k,1)} = \frac{n}{(n-1)(n-k)(n-k+1)}, \tag{2.66}$$

we get

Lemma 12. *The splitting probability $\pi_{n,k}$ is given by the following formula.*

$$\pi_{n,k} = [\![1 \le k \le n-1]\!]\frac{n-1}{n}p_{(n,1),(k,1)} + \frac{[\![k=0]\!]}{n} = \frac{[\![1 \le k \le n-1]\!]}{(n-k)(n-k+1)} + \frac{[\![k=0]\!]}{n}. \tag{2.67}$$

Hence it holds the following recurrence.

$$\mathbb{P}\{\mathcal{X}_{n,1} = m\} = \sum_{k=1}^{n-1} \pi_{n,k}\mathbb{P}\{\mathcal{X}_{k,1} = m-1\} = \sum_{k=1}^{n-1} \frac{1}{(n-k)(n-k+1)}\mathbb{P}\{\mathcal{X}_{k,1} = m-1\}, \tag{2.68}$$

for $n \ge 2$, $m \ge 2$ with $\mathbb{P}\{\mathcal{X}_{n,1} = 1\} = \pi_{n,0} = 1/n$ for all $n \ge 1$. We introduce the bivariate generating function

$$\mathcal{M}(z,v) = \sum_{n\ge1}\sum_{m\ge1} \mathbb{P}\{\mathcal{X}_{n,1} = m\}z^{n-1}v^m. \tag{2.69}$$

By multiplying with $z^n v^m$ and summing up over $n \ge 2$ and $m \ge 2$ we get the functional equation

$$z\mathcal{M}(z,v) = \mathcal{M}(z,v)\Big(v(z-1)\log\big(\frac{1}{1-z}\big) + vz\Big) + v\log\big(\frac{1}{1-z}\big), \tag{2.70}$$

which can be simplified to

$$\mathcal{M}(z,v) = \frac{v\log\big(\frac{1}{1-z}\big)}{v(1-z)\log\big(\frac{1}{1-z}\big) + z(1-v)}. \tag{2.71}$$

We rewrite $\mathcal{M}(z,v)$ the following way.

$$\begin{aligned}\mathcal{M}(z,v) &= \frac{v\log\big(\frac{1}{1-z}\big)}{z + v\big((1-z)\log\big(\frac{1}{1-z}\big) - z\big)} = \frac{v\log\big(\frac{1}{1-z}\big)}{z\Big(1 - v\big((z-1)\frac{\log\left(\frac{1}{1-z}\right)}{z} + 1\big)\Big)} \\ &= v\frac{\log\big(\frac{1}{1-z}\big)}{z}\sum_{m\ge0} v^m\Big((z-1)\frac{\log\big(\frac{1}{1-z}\big)}{z} + 1\Big)^m.\end{aligned} \tag{2.72}$$

Thus we can get a closed formula for the probabilities.

Proposition 1.

$$\mathbb{P}\{\mathcal{X}_{n,1} = m\} = \frac{1}{n}[\![m=1]\!] + [\![1 < m \le n]\!]\sum_{k=0}^{m-1}\binom{m-1}{k}\sum_{l=0}^{k}\binom{k}{l}(-1)^{k-l}\frac{{n+k-l \brack k+1}(k+1)!}{(n+k-l)!}. \tag{2.73}$$

For the moments we will use the following identity to obtain the factorial moments.

$$E_v D_v^s \mathcal{M}(z,v) = s![w^s]\mathcal{M}(z,v), \quad \text{where } w := v-1. \tag{2.74}$$

$$\mathcal{M}(z,v) = \frac{1}{(1-z)(1+\frac{z(1-v)}{v(1-z)\log\left(\frac{1}{1-z}\right)})} = \frac{1}{(1-z)(1-\frac{zw}{(w+1)(1-z)\log\left(\frac{1}{1-z}\right)})}$$
$$= \sum_{m\geq 0}\frac{z^m w^m}{(w+1)^m(1-z)^{m+1}\log^m\left(\frac{1}{1-z}\right)} = \sum_{m\geq 0}\frac{z^m w^m}{(1-z)^{m+1}\log^m\left(\frac{1}{1-z}\right)}\sum_{k\geq 0}\binom{k+m-1}{k}(-1)^k w^k. \tag{2.75}$$

Extraction of coefficients gives

$$[w^s]\mathcal{M}(z,v) = \frac{1}{(1-z)^{s+1}\log^s\left(\frac{1}{1-z}\right)} + \mathcal{O}\Big(\frac{1}{(1-z)^s\log^{s-1}\left(\frac{1}{1-z}\right)}\Big). \tag{2.76}$$

Thus we get by using singularity analysis

Proposition 2.

$$\mathbb{E}\Big(X_{n,1}^s\Big) = \frac{n^s}{\log^s n} + \mathcal{O}\Big(\frac{n^{s-1}}{\log^{s-1} n}\Big).$$

2.9.3 Phase changes

When isolating only one node labelled λ, with $1\leq\lambda\leq n$, at least two phase changes occur:

$$\frac{\log n}{n}X_{n;l} \xrightarrow{(d)} 1, \quad \frac{\log n}{n}X_{n;n+1-l} \xrightarrow{(d)} \mathrm{U}_1, \tag{2.77}$$

for fixed $l\in\mathbb{N}$. Unfortunately for the interesting region $\lambda=\rho n$, $0<\rho<1$, e.g. the r.v. $X_{n;\rho n}$, it seems to be difficult to obtain results using the recursive approach.

Now if one looks closer at the behavior of the random variables $X_{n;\boldsymbol{l}}$ and $Y_{n;\boldsymbol{l}}$ it is obvious that analogous phase changes must occur. We get for fixed $l\geq 1$

$$\frac{\log n}{n}X_{n;\boldsymbol{l}} \xrightarrow{(d)} 1, \quad \frac{X_{n;\boldsymbol{n+1-l}}}{n} \xrightarrow{(d)} 1. \tag{2.78}$$

and

$$\frac{\log n}{n}Y_{n;\boldsymbol{l}} \xrightarrow{(d)} Y_l, \quad \frac{Y_{n;\boldsymbol{n+1-l}}}{n} \xrightarrow{(d)} 1, \tag{2.79}$$

where $f_{Y_l}=lz^{l-1}$. As for (2.77) the cases $X_{n;\boldsymbol{\rho n}}$ and $Y_{n;\boldsymbol{\rho n}}$ seem to be out of reach.

Theorem 11. *For fixed $l\geq 1$ it holds $\frac{Y_{n;\boldsymbol{n+1-l}}}{n} \xrightarrow{(d)} 1$ and $\frac{X_{n;\boldsymbol{n+1-l}}}{n} \xrightarrow{(d)} 1$.*

Proof. It is obvious that $X_{n,\boldsymbol{n}} \stackrel{(d)}{=} X_{n,\boldsymbol{n-1}} \stackrel{(d)}{=} Y_{n;\boldsymbol{n}} \stackrel{(d)}{=} Y_{n,\boldsymbol{n-1}} \stackrel{(d)}{=} n-1$. It holds for fixed $l\geq 1$ the following recurrences (2.80) and (2.81).

$$\mathbb{P}\{X_{n;\boldsymbol{n+l-1}}=m\} = \sum_{r=1}^{l}\sum_{k=r}^{n-1}\mathfrak{p}_{(n,l),(k,r)}\sum_{s=0}^{m-1}\mathbb{P}\{X_{k;\boldsymbol{k+1-r}}=s\}\mathbb{P}\{X_{n-k;\boldsymbol{n-k-(l-r)}}=m-1-s\}, \tag{2.80}$$

for $n\geq l$, $m\geq n+1-l$, and where the splitting probabilities $\mathfrak{p}_{(n,l),(k,r)}$ are given in Lemma 3 by using (1.5).

$$\begin{aligned}\mathbb{P}\{Y_{n;\boldsymbol{n+1-l}}=m\} =& \sum_{r=1}^{l}\sum_{k=r}^{n-1}p_{(n,l),(k,r)}\sum_{s=0}^{m-1}\mathbb{P}\{Y_{k;\boldsymbol{k+1-r}}=s\}\times\\ &\times\mathbb{P}\{Y_{n-k;\boldsymbol{n-k-(l-r)}}=m-1-s\},\end{aligned} \tag{2.81}$$

for $n\geq l$, $m\geq l-1$, where the splitting probabilities $p_{(n,l),(k,r)}$ are given in Lemma 3.

Although one could use again generating functions to derive explicit results for the probabilities $\mathbb{P}\{X_{n;\boldsymbol{n+l-1}} = m\}$, $\mathbb{P}\{Y_{n;\boldsymbol{n+1-l}} = m\}$ the distribution can be done by hand. Since for $l > 1$ it takes at least $n+1-l$ edge cuts to isolate the nodes $1, \dots, n+1-l$ or $l, \dots, n$ one proceeds for the moments of $X_{n;\boldsymbol{n+l-1}}$ and $Y_{n;\boldsymbol{n+1-l}}$ as follows.

$$\begin{aligned} \mathbb{E}(X^s_{n;\boldsymbol{n+1-l}}) &= \sum_{m=n+1-l}^{n-1} m^s \mathbb{P}\{X_{n;\boldsymbol{n+1-l}} = m\} = \sum_{m=2}^{l} (n-1-(l-m))^s \mathbb{P}\{X_{n;\boldsymbol{n+1-l}} = m\} \\ &= (n-1)^s + \mathcal{O}((n-1)^{s-1}). \end{aligned} \tag{2.82}$$

The same computation holds for $Y_{n;\boldsymbol{n+1-l}}$. Further the variance is always constant $\mathbb{V}(X_{n;\boldsymbol{n+l-1}}) = c_l \geq 0$ and $\mathbb{V}(Y_{n;\boldsymbol{n+l-1}}) = d_l \geq 0$ for fixed $l \geq 1$. $\square$

Chapter 3

Isolating a leaf in rooted trees via random cuttings

3.1 Introduction

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

This edge-removal procedure can be considered as the counterpart of the cutting down edge-removal procedure considered by Meir and Moon, [55] (the edge removal procedure considered in Chapter 1 with $\lambda = 1$). In the latter procedure the subtree containing the original root of the tree is kept, while the other subtree is discarded (thus it can be seen as the opposite version of the procedure studied here) and then the procedure is continued recursively on the subtree containing the root until the original root is isolated.

We want to mention that also the following two-sided variant of the edge-removal procedure was considered in recent papers: after removing the randomly chosen edge one continues the procedure recursively on both of the obtained subtrees. Of course, when starting with a tree of size n, this two-sided variant leads to n isolated nodes after $n-1$ cuts, but one was interested in the total costs when isolating all nodes in the tree, if one assumes that the cost incurred for selecting an edge and splitting the tree is given by a toll function t_n. For toll functions $t_n = n^\alpha$ with $\alpha > 0$, asymptotic results for all moments are obtained in [65] and limiting distribution results for some classes of simply generated tree families are given in [21]. For Cayley trees this procedure is equivalent to a probabilistic model involved in the Union–Find (or equivalence-finding) algorithm, which was analyzed first by Knuth and Schönhage [39].

Basically, to obtain our limiting distribution results for Z_n we treat the recurrences appearing for the probabilities $\mathbb{P}\{Z_n = m\}$ via bivariate generating functions. This leads to exact solvable differential equations. Extracting coefficients of the solutions appearing asymptotically is performed via singularity analysis (see [25]), a complex-analytic technique that relates asymptotics of sequences to the local behavior of their generating functions in a neighborhood of the dominant singularities.

3.1.1 The recursive approach

We will study the random variable Z_n for the tree families considered by treating the recurrence

$$\mathbb{P}\{Z_n = m\} = \sum_{k=1}^{n-1} q_{n,k}\, \mathbb{P}\{Z_k = m-1\}, \quad \text{for } n \geq 2,\ m \geq 1, \tag{3.1}$$

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

An analogous approach, with transition probabilities $q_{n,n-k}$, was used in [63] for simply generated tree families to study the random variable X_n, i. e. the number of cuts to isolate the root of the tree. There one had to make a strong assumption on the tree family in order to justify this recursive approach: it was necessary that randomness is preserved by cutting off a random edge, which means that starting with a random tree of size n and removing a random edge, the remaining subtree of size k containing the root is actually a *random* tree of size k in this tree family. It turned out that exactly those tree families with $\varphi(t)$ given by Lemma 1 have this property and could be treated with the recursive approach. In [63] such tree families are called very simple tree families (in the context of simply generated trees). As already mentioned before in the preliminaries part we call increasing tree families with $\varphi(t)$ given by Lemma 1 *grown simple families of increasing trees*, due to the fact that they can be described by a probabilistic growth rule.

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

3.2 Results

We state here our findings for simply generated tree families and grown simple families of increasing trees with $\varphi(t)$ satisfying the assumptions made in Subsection 0.1.1 and Subsection 0.1.2, respectively. The proof of these results are given in Section 3.3 and Section 3.4.

Theorem 12. *For simply generated tree families with degree-weight generating function $\varphi(t)$, with period p and τ the minimal positive solution of the equation $t\varphi'(t) = \varphi(t)$, the random variable Z_n, which counts the number of random cuts that are required to isolate a leaf from a randomly chosen tree of size n with the edge-removal procedure considered, converges in distribution, for $n \to \infty$ with $n \equiv 1 \pmod p$, to a shifted Poisson distributed random variable Z, which has the distribution*

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

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

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

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

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

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

Theorem 13. *For grown simple families of increasing trees (with degree-weight generating function $\varphi(t)$ given by Lemma 1 and thus $\frac{T_{n+1}}{T_n} = c_1 n + c_2$, for all $n \geq 1$) let Z_n be the random variable, which counts* *the number of random cuts that are required to isolate a leaf from a randomly chosen tree of size n with the edge-removal procedure considered. Then, for a grown simple increasing tree family, Z_n converges for $n \to \infty$ in distribution to a discrete random variable Z. The probabilities $\mathbb{P}\{Z = m\}$ are for $m \geq 0$ given as the coefficients of the probability generating function $p(v) := \sum_{m \geq 0} \mathbb{P}\{Z = m\} v^m$ as given below.*

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

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

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

From Theorem 13 one gets the following corollaries, which contain results for particular increasing tree families.

Corollar 2. *Using the notation of Theorem 13, we give the following closed formulæ for the probabilities $\mathbb{P}\{Z = m\}$ for recursive tree and binary increasing trees. For recursive trees we get*

$$\mathbb{P}\{Z = m\} = (-1)^m \sum_{k \geq m} (-1)^k \frac{\pi^{2k} \binom{k}{m}}{(2k+1)!}, \tag{3.2}$$

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

$$\mathbb{P}\{Z = m\} = \sum_{k \geq m+1} (-1)^k \frac{\pi^{2k}}{2^{2k}(2k)!} \sum_{j=m+1}^{k} \binom{k}{j} \binom{j-1}{m} (-1)^j 8^j, \tag{3.3}$$

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

Note that the values given for $\mathbb{P}\{Z = 1\}$ are just as expected, since the average number of leaves in recursive trees is $\sim \frac{n}{2}$ for recursive trees and $\sim \frac{n}{3}$ for binary increasing trees (see e. g. [6]).

Corollar 3. *Using the notation of Theorem 13, we give for the instance $c_2 = 0$ the following closed formulæ for the r-th factorial moments of Z_n resp. Z. In the context of the multiple zeta functions*

$$\zeta(a_1, \dots, a_l) := \sum_{1 \leq n_1 < n_2 < \cdots < n_l} \frac{1}{n_1^{a_1} n_2^{a_2} \dots n_l^{a_l}}, \quad \zeta_N(a_1, \dots, a_l) := \sum_{1 \leq n_1 < n_2 < \cdots < n_l \leq N} \frac{1}{n_1^{a_1} n_2^{a_2} \dots n_l^{a_l}}, \tag{3.4}$$

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

$$\mathbb{E}\big(Z_n^{\underline{r}}\big) = r! \zeta_{n-1}(2, \dots, 2). \tag{3.5}$$

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

$$\mathbb{E}\big(Z^{\underline{r}}\big) = r! \zeta(2, \dots, 2) = r! \frac{\pi^{2r}}{(2r+1)!}. \tag{3.6}$$

Further we can decompose the limit distribution of the number of random cuts necessary to isolate a leaf Z as follows.

Corollar 4.

$$Z \stackrel{(d)}{=} \sum_{k=1}^{\infty} B_k, \tag{3.7}$$

where the B_k are Bernoulli distributed random variables $B_k \stackrel{(d)}{=} \mathrm{Be}(\frac{c_1+c_2}{k(c_1k+c_2)})$, for $k \in \mathbb{N}$.

Remark 1. Note that our computations of $\mathbb{E}\big(Z^{\underline{r}}\big)$ give thus a further proof of the identity $\zeta(\underbrace{2,\ldots,2}_{r \text{ times}}) = \frac{\pi^{2r}}{(2r+1)!}$, which was shown first in [32].

In Table 3.1 and 3.2 we collect some results of the limiting distribution of Z_n for a few interesting simply generated tree families and grown simple families of increasing trees.

Table 3.1: Limiting distribution results of Z_n for some important simply generated tree families.

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

Table 3.2: Limiting distribution results of Z_n for some important grown simple families of increasing trees. $H_n := \sum_{k\geq 1} \frac{1}{k}$ resp. $H_n^{(2)} := \sum_{k\geq 1} \frac{1}{k^2}$ denote the first and second order harmonic numbers.

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

3.3 Simply generated tree families

3.3.1 The transition probabilities

The required transition probabilities $q_{n,k}$ as defined in Subsection 3.1.1 were already computed in [63] by a generating functions approach, which is also sketched here. We can define the value $q_{n,k}$ equivalently as the probability that the number of descendants of a node (where the node itself is counted) that was chosen at random from one of the $n-1$ non-root nodes in a random tree of size n is k. We require also

the auxiliary values $\tilde{q}_{n,k}$, which denote the probability that the number of descendants of a randomly chosen node in a random tree of size n is k.
Introducing the generating functions

$$G(z,u) = \sum_{n\ge1}\sum_{k\ge0} nT_n\,\tilde{q}_{n,k}\,z^n u^k, \qquad H(z,u) = \sum_{n\ge1}\sum_{k\ge0}(n-1)T_n\,q_{n,k}\,z^n u^k,$$

we can directly translate the formal equation (6) or the recurrences

$$\begin{aligned}\tilde{q}_{n,k} &= \frac{1}{n}[\![k=n]\!] + \frac{1}{n}\sum_{r\ge1} r\varphi_r \sum_{\substack{k_1+\cdots+k_r=n-1,\\ k_1,\dots,k_r\ge1}} \frac{\tilde{q}_{k_1,k}T_{k_1}\cdots T_{k_m}}{T_n},\\ q_{n,k} &= \frac{1}{n-1}\sum_{r\ge1} r\varphi_r \sum_{\substack{k_1+\cdots+k_r=n-1,\\ k_1,\dots,k_r\ge1}} \frac{q_{k_1,k}T_{k_1}\cdots T_{k_m}}{T_n},\end{aligned} \tag{3.8}$$

into the equations

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

which imply

$$H(z,u) = T(zu)F(z), \quad \text{with} \quad F(z) := \frac{1}{1 - z\varphi'\big(T(z)\big)} - 1. \tag{3.9}$$

Extracting coefficients from (3.9) gives

$$F_n := [z^n]F(z) = \begin{cases} [T^n]\big(\varphi(T)\big)^n, & \text{for } n\ge1,\\ 0, & \text{for } n=0. \end{cases} \tag{3.10}$$

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

$$q_{n,k} = \frac{[z^n u^k]H(z,u)}{(n-1)T_n} = \frac{T_k F_{n-k}}{(n-1)T_n}, \tag{3.11}$$

with F_n defined by equation (3.10).

3.3.2 Solving the recurrence

Using (3.1) we have to study the recurrence

$$\mathbb{P}\{Z_n = m\} = \sum_{k=1}^{n-1} \frac{T_k F_{n-k}}{(n-1)T_n}\mathbb{P}\{Z_k = m-1\}, \tag{3.12}$$

with $\mathbb{P}\{Z_1 = 0\} = 1$ and $\mathbb{P}\{Z_n = 0\} = 0$, for $n\ge2$.
We will perform a generating functions approach using the bivariate generating function

$$M(z,v) := \sum_{n\ge1}\sum_{m\ge0} T_n\mathbb{P}\{Z_n = m\}z^n v^m. \tag{3.13}$$

Multiplying (3.12) with $(n-1)T_n z^n v^m$ and summing up for $n\ge2$ and $m\ge1$ leads to the following first order linear differential equation

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

with initial conditions $M(0,v)=0$ and $\big(\frac{\partial}{\partial z}M(z,v)\big)|_{z=0}=T_1=\varphi_0$, and the function $F(z)$ given by (3.9). Solving this differential equation leads to the solution

$$M(z,v)=\varphi_0 z\exp\Big(v\int_0^z\frac{F(t)}{t}dt\Big). \tag{3.14}$$

Then by using $T'(z)=\frac{\varphi(T(z))}{1-z\varphi'(T(z))}$, which follows from the functional equation (5), and a change of variables, we obtain:

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

Thus we get from (3.14) the following explicit formula for $M(z,v)$:

$$M(z,v)=\varphi_0 z\exp\Big(v\log\frac{\varphi\big(T(z)\big)}{\varphi_0}\Big)=\varphi_0 z\Big(\frac{\varphi\big(T(z)\big)}{\varphi_0}\Big)^v. \tag{3.15}$$

3.3.3 Characterizing the limiting distribution

Extracting coefficients from (3.15) immediately leads to

$$[v^m]M(z,v)=\varphi_0 z\frac{\big(\log\frac{\varphi(T(z))}{\varphi_0}\big)^m}{m!}. \tag{3.16}$$

In our asymptotic study of the coefficients $[z^nv^m]M(z,v)$ (and thus of the probabilities $\mathbb{P}\{Z_n=m\}$) via singularity analysis, which is given below, we will only carry out the instance that the degree-weight generating function $\varphi(t)$ is aperiodic, i. e. $p=1$. But for functions $\varphi(t)$ with period $p>1$ the proof is fully analogous: then we have to consider the contributions of all p dominant singularities, which must be added. This shows Theorem 12 also for $p>1$.

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

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

and further

$$\log\frac{\varphi\big(T(z)\big)}{\varphi_0}=\log\frac{\varphi(\tau)}{\varphi_0}-\frac{1}{\tau}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}+\tilde{\kappa}_2\big(1-\frac{z}{\rho}\big)+\mathcal{O}\Big(\big(1-\frac{z}{\rho}\big)^{\frac{3}{2}}\Big). \tag{3.17}$$

Via (3.17) we obtain thus from (3.16) for $m\geq 1$ the following expansion around $z=\rho$ (again with a certain constant $\tilde{\kappa}$):

$$\begin{aligned}[v^m]M(z,v)&=\sum_{n\geq1}T_n\mathbb{P}\{Z_n=m\}z^n\\&=\frac{\varphi_0\rho}{m!}\Big(\log\frac{\varphi(\tau)}{\varphi_0}\Big)^m-\frac{\varphi_0}{\varphi(\tau)(m-1)!}\Big(\log\frac{\varphi(\tau)}{\varphi_0}\Big)^{m-1}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}+\tilde{\kappa}\big(1-\frac{z}{\rho}\big)+\mathcal{O}\Big(\big(1-\frac{z}{\rho}\big)^{\frac{3}{2}}\Big).\end{aligned} \tag{3.18}$$

Applying singularity analysis to (3.18) gives then

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

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

$$\mathbb{P}\{Z_n = m\} = \frac{\varphi_0}{\varphi(\tau)}\frac{\big(\log\frac{\varphi(\tau)}{\varphi_0}\big)^{m-1}}{(m-1)!}\big(1+\mathcal{O}(n^{-1})\big). \tag{3.19}$$

Thus the probabilities $\mathbb{P}\{Z_n = m\}$ converge for all $m \geq 1$ to the probabilities $\mathbb{P}\{Z = m\}$ of a shifted Poisson distributed random variable Z. This shows the first part of Theorem 12.

3.3.4 Computing the moments

From the generating function $M(z,v)$ as given by (3.15) we can also compute easily the r-th factorial moments $\mathbb{E}\big(Z^{\underline{r}}\big)$.
Evaluating the r-th derivative with respect to v of $M(z,v)$ at $v=1$ gives

$$E_v D_v^r M(z,v) = E_v\Big[\varphi_0 z\Big(\log\frac{\varphi\big(T(z)\big)}{\varphi_0}\Big)^r e^{v\log\frac{\varphi(T(z))}{\varphi_0}}\Big] = z\varphi\big(T(z)\big)\Big(\log\frac{\varphi\big(T(z)\big)}{\varphi_0}\Big)^r \tag{3.20}$$

$$= T(z)\Big(\log\frac{\varphi\big(T(z)\big)}{\varphi_0}\Big)^r. \tag{3.21}$$

We further get by using (2) and (3.17) the asymptotic expansion (with a certain constant $\hat{\kappa}$)

$$\begin{aligned} T(z)\Big(\log\frac{\varphi\big(T(z)\big)}{\varphi_0}\Big)^r &= \tau\Big(\log\frac{\varphi(\tau)}{\varphi_0}\Big)^r - \Big(\log\frac{\varphi(\tau)}{\varphi_0}\Big)^{r-1}\big(r+\log\frac{\varphi(\tau)}{\varphi_0}\big)\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}} \\ &\quad + \hat{\kappa}\big(1-\frac{z}{\rho}\big) + \mathcal{O}\Big(\big(1-\frac{z}{\rho}\big)^{\frac{3}{2}}\Big). \end{aligned} \tag{3.22}$$

Singularity analysis leads then from (3.21) and (3.22) to the asymptotic expansion

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

and by using (3) thus to

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

This completes the proof of Theorem 12.

3.4 Grown simple families of increasing trees

3.4.1 The transition probabilities for general increasing trees

First we show for general increasing tree families an expression for the transition probabilities $q_{n,k}$ as defined in Subsection 3.1.1. We can do this analogous to Subsection 3.3.1 for simply generated tree families: we use the interpretation of the value $q_{n,k}$ as the probability that the number of descendants of a node that was chosen at random from one of the $n-1$ non-root nodes in a random tree of size n is

k, and define the auxiliary value $\tilde{q}_{n,k}$ as the probability that the number of descendants of a randomly chosen node in a random tree of size n is k.
Introducing the generating functions

$$G(z,u) = \sum_{n\ge1}\sum_{k\ge0} nT_n\tilde{q}_{n,k}\frac{z^n}{n!}u^k, \quad H(z,u) = \sum_{n\ge1}\sum_{k\ge0}(n-1)T_n q_{n,k}\frac{z^n}{n!}u^k,$$

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

$$\begin{aligned}
\tilde{q}_{n,k} &= \frac{1}{n}[\![k=n]\!] + \frac{1}{n}\sum_{r\ge1} r\varphi_r \sum_{\substack{k_1+\cdots+k_r=n-1,\\ k_1,\ldots,k_r\ge1}} \frac{\tilde{q}_{k_1,k}T_{k_1}\cdots T_{k_m}}{T_n}\binom{n-1}{k_1,k_2,\ldots,k_r},\\
q_{n,k} &= \frac{1}{n-1}\sum_{r\ge1} r\varphi_r \sum_{\substack{k_1+\cdots+k_r=n-1,\\ k_1,\ldots,k_r\ge1}} \frac{q_{k_1,k}T_{k_1}\cdots T_{k_m}}{T_n}\binom{n-1}{k_1,k_2,\ldots,k_r},
\end{aligned} \tag{3.23}$$

the following differential equations:

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

These differential equations have the solutions

$$G(z,u) = u\varphi\big(T(z)\big)\int_0^z \frac{\varphi\big(T(tu)\big)}{\varphi\big(T(t)\big)}dt, \quad H(z,u) = \varphi\big(T(z)\big)\int_0^z \frac{T(tu)\varphi'\big(T(t)\big)}{\varphi\big(T(t)\big)}dt. \tag{3.24}$$

Equation (3.24) gives immediately

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

and thus

$$q_{n,k} = \frac{n!T_k}{(n-1)T_n k!}[z^n]\varphi\big(T(z)\big)\int_0^z \frac{t^k\varphi'\big(T(t)\big)}{\varphi\big(T(t)\big)}dt. \tag{3.25}$$

For arbitrary degree-weight generating functions $\varphi(t)$ one cannot hope to obtain explicit formulæ for the probabilities $q_{n,k}$, but for the subclass of grown simple families of increasing trees, as given by Lemma 1, we will get an easy expression as is shown in Subsection 3.4.3.

3.4.2 Characterization of grown simple families of increasing trees

Proof of Lemma 1. We will show here Lemma 1, which characterizes increasing tree families that satisfy the equation $\frac{T_{n+1}}{T_n} = c_1 n + c_2$, with arbitrary but fixed constants c_1, c_2, for all $n \ge 1$.
We remark that due to the demand $T_n > 0$ for all $n \ge 1$ we get the a priori restrictions: $c_1 \ge 0$ and $c_2 > -c_1$ (otherwise there would exist $n \ge 1$ such that $\frac{T_{n+1}}{T_n} = c_1 n + c_2 < 0$).
• Now we consider the case $c_1 \ne 0$ and $c_2 \ne 0$ and get for T_n (where we use $T_1 = \varphi_0$):

$$T_n = T_1\prod_{k=1}^{n-1}(c_1k+c_2) = \varphi_0 c_1^{n-1}\prod_{k=1}^{n-1}\big(\frac{c_2}{c_1}+k\big) = \frac{\varphi_0 c_1^n}{c_2}\big(\frac{c_2}{c_1}+n-1\big)^{\underline{n}} = \frac{\varphi_0 c_1^n n!}{c_2}\binom{\frac{c_2}{c_1}+n-1}{n}$$

$$= \frac{\varphi_0(-c_1)^n n!}{c_2}\binom{-\frac{c_2}{c_1}}{n},$$

and further

$$T(z) = \sum_{n\geq 1} T_n \frac{z^n}{n!} = \frac{\varphi_0}{c_2}\sum_{n\geq 1}\binom{-\frac{c_2}{c_1}}{n}(-c_1 z)^n = \frac{\varphi_0}{c_2}\Big(\frac{1}{(1-c_1 z)^{\frac{c_2}{c_1}}} - 1\Big). \tag{3.26}$$

In order to decide which values of c_1, c_2 are indeed possible choices we have to compute the corresponding degree-weight generating functions and check whether they are admissible ($\varphi_k \geq 0$ for all $k \geq 0$). Differentiating (3.26) gives

$$T'(z) = \frac{\varphi_0}{(1-c_1 z)^{\frac{c_2}{c_1}+1}} = \varphi_0\Big(1 + \frac{c_2}{\varphi_0}T(z)\Big)^{\frac{c_1}{c_2}+1}. \tag{3.27}$$

We obtain $[T^n]\varphi(T) = \varphi_n$ and by using (3.24)

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

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

$$\varphi_n = \varphi_0\binom{\frac{c_1}{c_2}+1}{n}\big(\frac{c_2}{\varphi_0}\big)^n, \tag{3.28}$$

resp.

$$\varphi(t) = \sum_{n\geq 0}\varphi_n t^n = \varphi_0\Big(1 + \frac{c_2 t}{\varphi_0}\Big)^{\frac{c_1}{c_2}+1}. \tag{3.29}$$

By considering (3.28) we can now check whether the conditions $\varphi_n \geq 0$, for all $n \geq 0$, with $\varphi_0 > 0$, are satisfied.
(i) We consider first the case $c_2 > 0$: if $1 + \frac{c_1}{c_2} \notin \mathbb{N}$, then it follows that there exists $n \in \mathbb{N}$ such that $\binom{1+\frac{c_1}{c_2}}{n} < 0$ and, since $c_1 > 0$, thus that $\varphi_n < 0$. Therefore we get that this case is not admissible. But if $1 + \frac{c_1}{c_2} =: d \in \mathbb{N}$, then it follows that $\binom{\frac{c_1}{c_2}+1}{n} = 0$, for all $n > d$ and thus that $\varphi_n > 0$, for all $0 \leq n \leq d$ and $\varphi_n = 0$, for all $n > d$. Such degree-weight generating functions are admissible and are covered by Case B in Lemma 1.
(ii) We have to consider also the case $c_2 < 0$: since $c_1 + c_2 > 0$ it follows that $\frac{c_1}{c_2} < -1$ resp. $n - \frac{c_1}{c_2} - 2 > n - 1$ and thus that

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

for all $n \geq 0$. Therefore such degree-weight generating functions are also admissible and are covered by Case C in Lemma 1.
• Next we will consider the case $c_2 = 0$ (and $c_1 > 0$), which gives

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

and

$$T(z) = \sum_{n\geq 1} T_n\frac{z^n}{n!} = \frac{\varphi_0}{c_1}\sum_{n\geq 1}\frac{(c_1 z)^n}{n} = \frac{\varphi_0}{c_1}\log\big(\frac{1}{1-c_1 z}\big). \tag{3.30}$$

Since (3.30) gives

$$T'(z) = \frac{\varphi_0}{1-c_1 z} = \varphi_0 e^{\frac{c_1 T(z)}{\varphi_0}}, \tag{3.31}$$

we obtain

$$\varphi_n = [T^n]\varphi(T) = [T^n]T'(z) = [T^n]\varphi_0 e^{\frac{c_1 T(z)}{\varphi_0}} = \frac{\varphi_0 \left(\frac{c_1}{\varphi_0}\right)^n}{n!}, \tag{3.32}$$

and

$$\varphi(t) = \sum_{n \geq 0} \varphi_n t^n = \varphi_0 e^{\frac{c_1 t}{\varphi_0}}. \tag{3.33}$$

Since $c_1 > 0$, we obtain from (3.33) that $\varphi_n > 0$, for all $n \geq 0$, and thus that all degree-weight generating functions (3.33) are admissible. They are covered by Case A in Lemma 1.

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

$$T(z) = \sum_{n \geq 1} T_n \frac{z^n}{n!} = \frac{\varphi_0}{c_2} \sum_{n \geq 1} \frac{(c_2 z)^n}{n!} = \frac{\varphi_0}{c_2}\left(e^{c_2 z} - 1\right). \tag{3.34}$$

Since (3.34) gives

$$T'(z) = \varphi_0 e^{c_2 z} = \varphi_0 + c_2 T(z), \tag{3.35}$$

this leads to

$$\varphi(t) = \varphi_0 + c_2 t. \tag{3.36}$$

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

3.4.3 The transition probabilities

Now we are going to calculate the probabilities $q_{n,k}$ for grown simple families of increasing trees. Using (5) from (3.25) we get

$$q_{n,k} = \frac{n! T_k}{(n-1) T_n k!} [z^n] \varphi\big(T(z)\big) \int_0^z \frac{t^k \varphi'\big(T(t)\big)}{\varphi\big(T(t)\big)} dt = \frac{n! T_k}{(n-1) T_n k!} [z^n] T'(z) \int_0^z \frac{t^k T''(t)}{(T'(t))^2} dt. \tag{3.37}$$

If $c_2 \neq 0$ then we obtain from (3.37) via integration by parts

$$\begin{aligned}
T'(z) \int_0^z \frac{t^k T''(t)}{(T'(t))^2} dt &= \frac{\varphi_0}{(1-c_1 z)^{\frac{c_2}{c_1}+1}} \int_0^z \frac{t^k \varphi_0 (c_1 + c_2)(1 - c_1 t)^{\frac{2c_2}{c_1}+2}}{\varphi_0^2 (1 - c_1 t)^{\frac{c_2}{c_1}+2}} dt \\
&= \frac{c_1 + c_2}{(1 - c_1 z)^{\frac{c_2}{c_1}+1}} \int_0^z t^k (1 - c_1 t)^{\frac{c_2}{c_1}} dt = \frac{\varphi_0 (c_1 + c_2)}{(1 - c_1 z)^{\frac{c_2}{c_1}+1}} \left[\frac{k!}{T_{k+2}} - \sum_{l=0}^{k} \frac{k^{\underline{l}} z^{k-l} (1 - c_1 z)^{\frac{c_2}{c_1}+1+l}}{T_{l+2}} \right] \\
&= (c_1 + c_2) \left[\frac{k! T'(z)}{T_{k+2}} - \varphi_0 \sum_{l=0}^{k} \frac{k^{\underline{l}} z^{k-l} (1 - c_1 z)^l}{T_{l+2}} \right].
\end{aligned} \tag{3.38}$$

For $n > k$, combining (3.37) and (3.38) leads to

$$\begin{aligned}
q_{n,k} &= \frac{n! T_k}{(n-1) T_n k!} [z^n] T'(z) \int_0^z \frac{t^k T''(t)}{(T'(t))^2} dt = \frac{n! T_k}{(n-1) T_n k!} [z^n] \frac{(c_1 + c_2) k! T'(z)}{T_{k+2}} \\
&= \frac{n! T_k}{(n-1) T_n k!} \frac{(c_1 + c_2) k! T_{n+1}}{T_{k+2} n!} = \frac{(c_1 + c_2)(c_1 n + c_2)}{(n-1)(c_1 (k+1) + c_2)(c_1 k + c_2)}.
\end{aligned} \tag{3.39}$$

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

3.4.4 Solving the recurrence

Using (3.1) we obtain therefore for $n \geq 2$ and $m \geq 1$ the recurrence

$$\mathbb{P}\{Z_n = m\} = \sum_{k=1}^{n-1} \frac{(c_1 + c_2)(c_1 n + c_2)}{(n-1)(c_1(k+1) + c_2)(c_1 k + c_2)} \mathbb{P}\{Z_k = m-1\}, \tag{3.40}$$

with $\mathbb{P}\{Z_1 = 0\} = 1$ and $\mathbb{P}\{Z_n = 0\} = 0$, for $n \geq 2$. We simplify this full history recursion by multiplying with $\frac{n-1}{c_1 n + c_2}$ and taking differences. (3.40) leads then for $n \geq 1$ and $m \geq 1$ to

$$\frac{n}{c_1(n+1) + c_2} \mathbb{P}\{Z_{n+1} = m\} - \frac{n-1}{c_1 n + c_2} \mathbb{P}\{Z_n = m\} = \frac{c_1 + c_2}{(c_1(n+1) + c_2)(c_1 n + c_2)} \mathbb{P}\{Z_n = m-1\}. \tag{3.41}$$

Introducing the generating function

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

recurrence (3.41) leads to the following homogeneous second order linear differential equation:

$$z(1-z) \frac{\partial^2}{\partial z^2} M(z,v) + \frac{3c_1 + c_2}{c_1} (1-z) \frac{\partial}{\partial z} M(z,v) - \frac{(c_1 + c_2)v}{c_1} M(z,v) = 0, \tag{3.42}$$

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

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

$M(z,v)$ satisfies the hypergeometric differential equation with parameters
$a = \frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1}$, $b = \frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1}$, and $c = \frac{3c_1+c_2}{c_1}$.
A solution basis of (3.42) is thus given by the following two functions (see e. g. [3]):

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

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

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

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

$$M(z,v) = C(v)\, {}_2F_1\left(\begin{matrix} \frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1}, \frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1} \\ \frac{3c_1+c_2}{c_1} \end{matrix} \,\middle|\, z \right), \tag{3.43}$$

with a certain function $C(v)$, since the other base function is not analytic at $z = 0$. After adapting (3.43) to the initial conditions we obtain the solution

$$M(z,v) = \frac{1}{(2c_1+c_2)(c_1+c_2)} {}_2F_1\left(\begin{matrix} \frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1}, \frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1} \\ \frac{3c_1+c_2}{c_1} \end{matrix} \,\middle|\, z \right). \tag{3.44}$$

3.4.5 Characterizing the limiting distribution

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

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

and get from equation (3.44) the following local expansion of $M(z, v)$ around the dominant singularity $z = 1$ in a complex neighborhood of $v = 1$ (with certain functions $C_0(v)$, $C_1(v)$, and $C_2(v)$):

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

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

$$
\begin{aligned}
p_n(v) &= (c_1 n + c_2)(c_1(n+1) + c_2)[z^{n-1}]M(z, v) \\
&= \frac{\Gamma(1 + \frac{c_2}{c_1})}{\Gamma(\frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})\Gamma(\frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})} \big(1 + \mathcal{O}(n^{-1})\big). \qquad (3.45)
\end{aligned}
$$

Thus it follows from (3.45) that the moment generating function (= Laplace transform)

$$
\mathbb{E}(e^{Z_n s}) = \sum_{m \ge 0} \mathbb{P}\{Z_n = m\} e^{ms} = p_n(e^s) \qquad (3.46)
$$

of Z_n converges in a neighborhood of $s = 0$ to the moment generating function $\mathbb{E}(e^{Zs}) = p(e^s)$ of a discrete random variable Z with probability generating function $p(v) := \sum_{m\ge 0} \mathbb{P}\{Z = m\} v^s$ given by

$$
p(v) = \frac{\Gamma(1 + \frac{c_2}{c_1})}{\Gamma(\frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})\Gamma(\frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})}. \qquad (3.47)
$$

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

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

one can further simplify the formula of $p(v)$ for binary increasing trees and recursive trees; see Table 3.2. Extracting coefficients leads then for these tree families to expressions for the probabilities $\mathbb{P}\{Z = m\}$ as given in Corollary 2.

Furthermore, we can give a representation of $p(v)$ as an infinite product, where we simply use repeatedly the functional equation $\Gamma(x) = \frac{\Gamma(x+1)}{x}$ for the Gamma function expressions in (3.47). One obtains after

n iteration steps and some simplifications:

$$\begin{aligned} p(v) &= \frac{\Gamma(1)\Gamma(1+\frac{c_2}{c_1})}{\Gamma(\frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})\Gamma(\frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1})} \\ &= f(n)\prod_{k=1}^{n}\Big(1+\frac{(c_1+c_2)(v-1)}{k(c_1k+c_2)}\Big), \end{aligned} \tag{3.48}$$

with

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

Since it holds $f(n) \to 1$ for $n \to \infty$, as can be shown e. g. via Stirling's asymptotic formula for the Gamma function, we obtain from (3.48) the representation

$$p(v) = \prod_{k=1}^{\infty}\Big(1+\frac{(c_1+c_2)(v-1)}{k(c_1k+c_2)}\Big) = \prod_{k=1}^{\infty}\Big(\frac{k(c_1k+c_2)-(c_1+c_2)}{k(c_1k+c_2)} + \frac{v(c_1+c_2)}{k(c_1k+c_2)}\Big). \tag{3.49}$$

By an application of the continuity theorem for the Laplace transform (see e. g. [13]) we obtain from equations (3.47) and (3.49) immediately the first part of Theorem 13. Further we can immediately deduce Corollary 4 from (3.49).

3.4.6 Computing the moments

From the explicit formula (3.44) for the generating function $M(z, v)$ we can also compute exact expressions for the r-th factorial moments $\mathbb{E}\big(Z_n^{\underline{r}}\big)$ of Z_n. To do this we will give first an exact formula for the probability generating function $p_n(v)$. By easy manipulations we obtain

$$\begin{aligned} p_n(v) &= (c_1(n+1)+c_2)(c_1n+c_2)[z^{n-1}]M(z,v) \\ &= \frac{(c_1(n+1)+c_2)(c_1n+c_2)\Big(\frac{2c_1+c_2-\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1}\Big)^{\overline{n-1}}\Big(\frac{2c_1+c_2+\sqrt{(2c_1+c_2)^2-4c_1(c_1+c_2)v}}{2c_1}\Big)^{\overline{n-1}}}{(2c_1+c_2)(c_1+c_2)\Big(\frac{3c_1+c_2}{c_1}\Big)^{\overline{n-1}}(n-1)!} \\ &= \frac{(c_1(n+1)+c_2)(c_1n+c_2)\prod_{k=0}^{n-2}\Big(k^2+(2+\frac{c_2}{c_1})k+(1+\frac{c_2}{c_1})-(1+\frac{c_2}{c_1})(1-v)\Big)}{(2c_1+c_2)(c_1+c_2)\Big(3+\frac{c_2}{c_1}\Big)^{\overline{n-1}}(n-1)!} \\ &= \frac{\prod_{k=1}^{n-1}\Big(k(c_1k+c_2)-(c_1+c_2)(1-v)\Big)}{\prod_{k=1}^{n-1}\big(k(c_1k+c_2)\big)} = \prod_{k=1}^{n-1}\Big(1+\frac{(c_1+c_2)(v-1)}{k(c_1k+c_2)}\Big). \end{aligned} \tag{3.50}$$

Evaluating the r-th derivative of $p_n(v)$ at $v=1$ as given by (3.50) then leads to:

$$\begin{aligned} \mathbb{E}\big(Z_n^{\underline{r}}\big) = E_v D_v^r p_n(v) &= \sum_{1\le k_1<k_2<\cdots<k_r\le n-1}\frac{r!(c_1+c_2)^r}{\prod_{i=1}^{r}\big(k_i(c_1k_i+c_2)\big)} \\ &= r!(c_1+c_2)^r\sum_{k_1=1}^{n-1}\frac{1}{k_1(c_1k_1+c_2)}\sum_{k_2=k_1+1}^{n-1}\frac{1}{k_2(c_1k_2+c_2)}\cdots\sum_{k_r=k_{r-1}+1}^{n-1}\frac{1}{k_r(c_1k_r+c_2)}, \end{aligned} \tag{3.51}$$

which shows also the second part of Theorem 13. We remark that this result for the r-th factorial moments can also be obtained directly from the recurrence (3.41) by using elementary means. We multiply on both sides with $m^{\underline{r}} = (m-1)^{\underline{r}} + r(m-1)^{\underline{r-1}}$ and sum up for $m \ge 1$. Thus we easily get the following Lemma.

Lemma 13. *If $n \geq 1$ we have the following recursion for $\tau_r(n) = \mathbb{E}(Z_n^{\underline{r}})$:*

$$\tau_r(n+1) \quad = \quad \tau_r(n) + r\frac{c_1 + c_2}{n(c_1 n + c_2)}\tau_{r-1}(n). \tag{3.52}$$

Iteration of this result leads directly to

$$\tau_r(n) \quad = \quad r!(c_1+c_2)^r \sum_{k_1=1}^{n-1} \frac{1}{k_1(c_1k_1+c_2)} \sum_{k_2=1}^{k_1-1} \frac{1}{k_2(c_1k_2+c_2)} \cdots \sum_{k_r=1}^{k_{r-1}-1} \frac{1}{k_r(c_1k_r+c_2)}. \tag{3.53}$$

If $c_2 = 0$ we get by using (3.4) the following result

$$\mathbb{E}\big(Z_n^{\underline{r}}\big) = r!\zeta_{n-1}(2,\ldots,2). \tag{3.54}$$

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

$$p(v) = \lim_{n\to\infty} p_n(v) = \frac{\sin(\pi\sqrt{1-v})}{\pi\sqrt{(1-v)}} = \sum_{k\geq 0} \frac{\pi^{2k}(-1)^k(1-v)^k}{(2k+1)!}, \tag{3.55}$$

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

$$\mathbb{E}(Z^{\underline{r}}) = E_v D_v^r p(v) = \frac{r!\pi^{2r}}{(2r+1)!}. \tag{3.56}$$

Since (3.5) yields $\mathbb{E}(Z^{\underline{r}}) = r!\zeta(\underbrace{2,2,\ldots,2}_{r \text{ times}})$, the proof of Corollary 3 is finished.

Chapter 4

Non-crossing-trees: Isolating a leaf and Climbing depth

4.1 Introduction

A non-crossing tree is a tree drawn on the plane having as vertices a set of points on the boundary of a circle, whose edges are straight line segments that do not cross. We consider the vertices labelled clockwise from 1 to n, where the root of the tree is vertex 1.

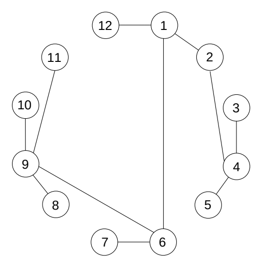

Figure 4.1: A non-crossing tree of size 12.

In this chapter we will consider two procedures to reach an endnode (a leaf) in a non-crossing tree of size n.

4.1.1 Isolating a leaf in non-crossing trees via random cuttings

We consider as before the following version of an edge-removal procedure: after removing a random edge of the tree we discard the subtree containing the original root of the tree and continue the procedure recursively with the other subtree. Thus we will finally isolate a leaf of the original tree and the procedure stops. Under the random tree model we are going to study the random variable Z_n, which counts here the number of edges that will be removed from a randomly chosen non-crossing tree of size n by this edge-removal procedure until a leaf is isolated. This removal procedure was analyzed in Chapter 3 for simply generated trees and a subfamily of increasing trees. We can state the limiting distribution of Z_n.

4.1.2 Climbing depth of non-crossing trees

We are considering here the following procedure to "climb" rooted trees. We start with a non-crossing tree T of size n rooted at node r_0, where the size measures as usual the number of nodes of T. If $n > 1$ then we choose at random one of the edges e in T which are incident with the root r_0 and proceed along

e (say $e = (r_0, r_1)$) to r_1. Then we iterate this procedure with the subtree of T rooted at node r_1. After $m \le n-1$ steps (= number of used edges), we will reach an endnode of T and stop. This procedure was

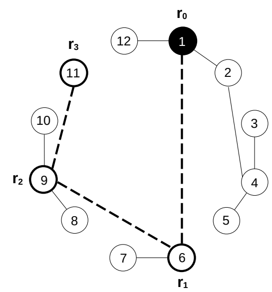

Figure 4.2: Climbing a non-crossing tree of size 12 in 3 steps.

studied in [62] for a wide range of random rooted trees.

4.2 Mathematical Preliminaries

For the analysis of the two considered parameters we will use a generating functions approach. The basic decomposition of non-crossing trees, which is described in [23], can be translated into equations for suitable chosen generating functions for the considered parameters. Following [23] a non-crossing tree consists of a root, which is attached to a (possible empty) sequence of butterflies, where a butterfly is a (ordered) pair of non-crossing trees, that share a common root.

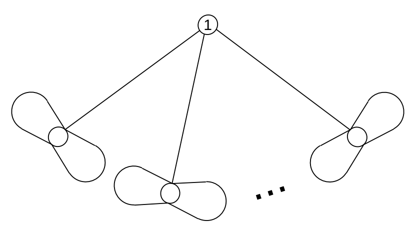

Figure 4.3: The combinatorial description of a non-crossing tree.

The arising formal combinatorial decompositions

$$\begin{aligned} \mathcal{T} &= \bigcirc \times \Big(\{\epsilon\} \,\dot\cup\, \mathcal{B} \,\dot\cup\, \mathcal{B}\times\mathcal{B} \,\dot\cup\, \mathcal{B}\times\mathcal{B}\times\mathcal{B} \,\dot\cup\, \cdots \Big) = \bigcirc \times SEQ(\mathcal{B}), \\ \bigcirc \times \mathcal{B} &= \mathcal{T} \times \mathcal{T}, \end{aligned} \tag{4.1}$$

can be translated immediately into the following system of equations for the generating functions $T(z) = \sum_{n\ge1} T_n z^n$ and $B(z) = \sum_{n\ge1} B_n z^n$ of the number T_n of non-crossing trees of size n resp. the number B_n of butterflies of size n:

$$T(z) = \frac{z}{1 - B(z)}, \qquad B(z) = \frac{T^2(z)}{z}. \tag{4.2}$$

Thus the number T_n of different non-crossing trees of size n can be calculated by an application of the inversion formula of Lagrange - Bürmann:

Lemma 14 (Lagrange - Bürmann inversion formula)**.** *Let $\Phi(z)$ be a formal power series with $\Phi_0 \ne 0$. Further let $T(z)$ denote the only formal power series solution of $\frac{T(z)}{\Phi(T(z))} = z$, thus $T^{[-1]}(z) = \frac{z}{\Phi(z)}$. Then* *the coefficient of* $\Psi(T(z)) = \sum_{n\geq 0} d_n z^n$ *for arbitrary power series* $\Psi(z)$ *is given by*

$$d_n = [z^n]\Psi(T(z)) = \frac{1}{n}[z^{n-1}]\Psi'(z)(\Phi(z))^n, \tag{4.3}$$

where $\Psi'(z) := \frac{d}{dz}\Psi(z)$.

This was done in [23]. The number T_n also follows due to the fact that T_n is equal to the number of ternary trees of size $n-1$, [67]. We arrive at the following.

$$T_n = \frac{1}{2n-1}\binom{3n-3}{n-1}. \tag{4.4}$$

From (4.2) follows also that $B(z) = \frac{z}{(1-B(z))^2}$, and thus that the family of the butterflies (but not the non-crossing trees) are simply generated trees (see Subsection 0.1.1) with the degree generating function $\varphi(t) = \frac{1}{(1-t)^2}$. This means that the family $\mathcal{B}$ resp. the corresponding generating function $B(z)$ fulfills the identity

$$\mathcal{B} = \bigcirc \times \varphi(\mathcal{B}), \qquad B(z) = z\varphi(B(z)). \tag{4.5}$$

Using (4.3) one gets easily that the number of butterflies of size n are given by

$$B_n = \frac{1}{n}\binom{3n-2}{n-1}. \tag{4.6}$$

To obtain our limiting distribution results for Y_n and Z_n we treat the recurrences appearing for the probabilities $\mathbb{P}\{Y_n = m\}$ and $\mathbb{P}\{Z_n = m\}$ via bivariate generating functions. This leads to exact solvable differential equations resp. functional equations and extracting coefficients of the solutions appearing asymptotically is performed via singularity analysis (see [25]), a complex-analytic technique that relates asymptotics of sequences to the local behavior of their generating functions in a neighborhood of the dominant singularities.

4.3 Results

Theorem 14. *For non-crossing trees the random variable* Z_n*, which counts the number of random cuts that are required to isolate a leaf from a randomly chosen non-crossing tree of size n with the edge-removal procedure considered, converges in distribution, for* $n \to \infty$*, to a shifted Poisson distributed random variable* Z*, which has the distribution*

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

with parameter $\lambda := \log\left(\frac{\varphi(\tau)}{\varphi_0}\right)$*, where* τ *is the minimal positive solution of the equation* $t\varphi'(t) = \varphi(t)$*, where* $\varphi(t)$ *is the degree generating function of the butterflies. Moreover, the r-th factorial moments* $\mathbb{E}\big(Z_n^{\underline{r}}\big)$ *have the asymptotic expansion*

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

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

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

Note that as expected the considered parameter behaves the same way as for the butterfly tree family $\mathcal{B}$.

Theorem 15. *For non-crossing trees the random variable* X_n*, which counts the number of steps that are required to climb a random non-crossing tree of size n with the climbing procedure considered, converges* *in distribution, for $n \to \infty$, to a discrete mixed distribution which consist of a geometric distribution with parameter $p = \rho/\tau$ and a negative binomial distribution with the same parameter p, it has the distribution*

$$\mathbb{P}\{X = m\} = \tau \cdot \frac{\rho}{\tau}\Big(1 - \frac{\rho}{\tau}\Big)^{m-1} + (1-\tau) \cdot \Big(1 - \frac{\rho}{\tau}\Big)^{m-2} \frac{(m-1)\rho^2}{\tau^2}, \quad \textit{for } m \geq 1,$$

and $\mathbb{P}\{X = 0\} = 0$. τ is the minimal positive solution of the equation $t\varphi'(t) = \varphi(t)$, where $\varphi(t)$ is the degree generating function of the butterflies. Moreover, the r-th factorial moments $\mathbb{E}\big(Z_n^{\underline{r}}\big)$ have for $r \geq 2$ the asymptotic expansion

$$\mathbb{E}\big(Z_n^{\underline{r}}\big) = r!\tau\frac{\tau^r}{\rho^r}\Big(1 - \frac{\rho}{\tau}\Big)^{r-1} + r!(1-\tau)\frac{\tau^r}{\rho^r}\Big(1 - \frac{\rho}{\tau}\Big)^{r-1}(r + 1 - 2p) + \mathcal{O}(n^{-1}).$$

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

$$\mathbb{E}(Z_n) = \tau\frac{\tau}{\rho} + 2(1-\tau)\frac{\tau}{\rho} + \mathcal{O}(n^{-1}), \quad \textit{and} \quad \mathbb{V}(Z_n) = \frac{\tau^2}{\rho^2}\Big(2 + \rho - 2 - \frac{\rho}{\tau}\tau^2) + \mathcal{O}(n^{-1}).$$

Note that for butterfly tree family only the negative binomial distribution is appearing in the limit distribution for X_n.

4.4 Proof for the isolation of a leaf in non-crossing trees via random cuttings

Since the butterflies are simply generated trees, we already know the behavior of the edge-removal procedure for this kind of trees from [43]. We denote here by Z_n resp. $\tilde{Z}_n$ the random variables that count the number of cuts to isolate a leaf in a random non-crossing tree resp. a random butterfly of size n. Now we turn to the analysis of non-crossing trees. After removing an edge of a non-crossing tree, the two resulting subtrees can be interpreted as a non-crossing tree which contains the root and a butterfly. The transition probabilities $p_{n,k}$, where $p_{n,k}$ denotes the probability that by choosing a random non-crossing tree of size n and removing a random edge, the remaining subtree containing the root is of size k, were already computed in [64]. They are given by

$$p_{n,k} = \frac{(3k-2)T_k B_{n-k}}{(n-1)T_n}, \quad \text{for } 1 \leq k \leq n-1.$$

4.4.1 Solving the recurrence

Now we have to treat the recurrence

$$\mathbb{P}(Z_n = m) = \sum_{k=1}^{n-1} \frac{(3(n-k)-2)T_{n-k}B_k}{(n-1)T_n}\mathbb{P}(\tilde{Z}_k = m-1),$$

with the initial values $\mathbb{P}\{Z_1 = 0\} = 1$ and $\mathbb{P}\{X_n = 0\} = 0$ for $n \geq 2$. We will use the following generating function

$$M(z,v) = \sum_{n\geq 1}\sum_{m\geq 0} T_n\mathbb{P}(Z_n = m)z^n v^m. \tag{4.7}$$

In addition to $M(z,v)$ we have to use the bivariate generating function

$$N(z,v) = \sum_{n\geq 1}\sum_{m\geq 0} B_n\mathbb{P}(\tilde{Z}_n = m)z^n v^m.$$

The recurrence can be translated into the following equation.

$$z\frac{\partial}{\partial z}M(z,v) - M(z,v) = vN(z,v)\big(3\frac{\partial}{\partial z}T(z) - 2T(z)\big), \tag{4.8}$$

with initial conditions $M(0,v) = 0$ and $\frac{\partial}{\partial z}M(z,v)\big|_{z=0} = 1$.
Since butterflies are a simply generated tree family (with degree generating function $\varphi(t) = \frac{1}{(1-t)^2}$) we can use the a result of [43] and get the explicit formula

$$N(z,v) = \varphi_0 z \exp\big(v \log \varphi(B(z))\big) = \frac{z}{\big(1-B(z)\big)^{2v}}.$$

Since the solution of the homogeneous differential equation corresponding to (4.8) is given by Cz, we use the variation of the parameters method and start with $M(z,v) = C(z,v)z$. This gives

$$\frac{\partial}{\partial z}C(z,v) = \frac{v}{(1-B(z))^{2v}} \frac{3zT'(z) - 2T(z)}{z}.$$

Since it holds that

$$\frac{3zT'(z) - 2T(z)}{z} = B'(z),$$

which is checked easily, we obtain

$$\frac{\partial}{\partial z}C(z,v) = \frac{vB'(z)}{(1-B(z))^{2v}}.$$

This leads to the following general solution of (4.8):

$$M(z,v) = vz\int_0^z \frac{B'(t)}{(1-B(t))^{2v}}dt + K(v)z = \frac{vz}{2v-1}\Big(\frac{1}{(1-B(z))^{2v-1}} - 1\Big) + K(v)z,$$

with a function $K(v)$. Adapting to the initial values gives then the required solution

$$M(z,v) = \frac{z}{2v-1}\Big(\frac{v}{(1-B(z))^{2v-1}} + v - 1\Big). \tag{4.9}$$

4.4.2 Characterizing the limiting distribution

Extracting coefficients leads for $n \geq 2$ and $m \geq 1$ to

$$\begin{aligned}[v^m]M(z,v) &= [v^{m-1}]\frac{z}{(2v-1)(1-B(z))^{2v-1}} = [v^{m-1}]\frac{z(1-B(z))}{(2v-1)(1-B(z))^{2v}}\\ &= z(B(z)-1)2^{m-1}\sum_{k=0}^{m-1}\frac{\big(\log\frac{1}{1-B(z)}\big)^k}{k!}.\end{aligned}$$

We require the following local expansion of $B(z)$ around the dominant singularity $\rho = \frac{4}{27}$, with $\tau = \frac{1}{3}$ and $\varphi(t) = \frac{1}{(1-t)^2}$:

$$B(z) = \tau - \sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}} + \mathcal{O}(1-\frac{z}{\rho}). \tag{4.10}$$

This gives the local expansions

$$\frac{1}{1-B(z)} = \frac{1}{1-\tau}\Big(1 - \frac{1}{1-\tau}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}} + \mathcal{O}(1-\frac{z}{\rho})\Big), \tag{4.11}$$

$$\log\frac{1}{1-B(z)} = \log\frac{1}{1-\tau} - \frac{1}{1-\tau}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}} + \mathcal{O}(1-\frac{z}{\rho}), \tag{4.12}$$

and finally

$$\sum_{k=0}^{m-1}\frac{\left(\log\frac{1}{1-B(z)}\right)^k}{k!} = \sum_{k=0}^{m-1}\frac{\left(\log\frac{1}{1-\tau}\right)^k}{k!} - \sum_{k=1}^{m-1}\frac{\left(\log\frac{1}{1-\tau}\right)^{k-1}}{(k-1)!}\frac{1}{1-\tau}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}} + \mathcal{O}(1-\frac{z}{\rho}).$$

This gives due to some cancellations the expansion

$$\begin{aligned}
&z(B(z)-1)2^{m-1}\sum_{k=0}^{m-1}\frac{\left(\log\frac{1}{1-B(z)}\right)^k}{k!}\\
&= 2^{m-1}\rho\Big(\tau-1-\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}+\mathcal{O}(1-\frac{z}{\rho})\Big)\\
&\quad\times\Big(\sum_{k=0}^{m-1}\frac{\left(\log\frac{1}{1-\tau}\right)^k}{k!} - \sum_{k=1}^{m-1}\frac{\left(\log\frac{1}{1-\tau}\right)^{k-1}}{(k-1)!}\frac{1}{1-\tau}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}+\mathcal{O}(1-\frac{z}{\rho})\Big)\\
&= 2^{m-1}\rho(\tau-1)\sum_{k=0}^{m-1}\frac{\left(\log\frac{1}{1-\tau}\right)^k}{k!} - 2^{m-1}\rho\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\frac{\left(\log\frac{1}{1-\tau}\right)^{m-1}}{(m-1)!}\sqrt{1-\frac{z}{\rho}}+\mathcal{O}(1-\frac{z}{\rho}).
\end{aligned}$$

Singularity analysis leads thus to

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

Together with

$$T_n = [z^n]\frac{z}{1-B(z)} = \frac{\rho}{1-\tau} - \frac{\rho}{(1-\tau)^2}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\frac{n^{-\frac{3}{2}}}{\rho^n\Gamma(-\frac{1}{2})}\Big(1+\mathcal{O}(\frac{1}{\sqrt{n}})\Big), \tag{4.13}$$

we obtain

$$\mathbb{P}\{Z_n=m\} = (1-\tau)^2 2^{m-1}\frac{\left(\log\frac{1}{1-\tau}\right)^{m-1}}{(m-1)!}\Big(1+\mathcal{O}(\frac{1}{\sqrt{n}})\Big) = \frac{1}{\varphi(\tau)}\frac{\left(\log\varphi(\tau)\right)^{m-1}}{(m-1)!}\Big(1+\mathcal{O}(\frac{1}{\sqrt{n}})\Big).$$

4.4.3 Computing the moments

From the generating function $M(z,v)$ as given by (4.9) we can also easily compute the r-th factorial moments $E(Z_n^{\underline{r}})$. Evaluating the r-th derivative w. r. t. v of $M(z,v)$ at $v=1$ gives

$$\begin{aligned}
E_vD_v^rM(z,v) &= E_vD_v^r\frac{z}{2v-1}\Big(\frac{v}{(1-B(z))^{2v-1}}+v-1\Big)\\
&= E_vD_v^r\frac{z}{2(2v-1)}\Big(\frac{(2v-1)+1}{(1-B(z))^{2v-1}}+(2v-1)-1\Big)\\
&= \frac{z}{2}E_vD_v^r\Big(\frac{1}{(1-B(z))^{2v-1}}+\frac{1}{(2v-1)(1-B(z))^{2v-1}}+1-\frac{1}{2v-1}\Big)\\
&= \frac{z}{2}\Big(\frac{2^r\log^r\left(\frac{1}{1-B(z)}\right)}{1-B(z)}+\sum_{k=0}^{r}\binom{r}{k}\frac{2^k(-1)^kk!2^{r-k}\log^{r-k}\left(\frac{1}{1-B(z)}\right)}{1-B(z)}-2^r(-1)^rr!\Big).
\end{aligned} \tag{4.14}$$

Using (4.11) and (4.12) leads to the local expansions

$$\log^j\big(\frac{1}{1-B(z)}\big) = \log^j\big(\frac{1}{1-\tau}\big) - \log^{j-1}\big(\frac{1}{1-\tau}\big)\frac{j}{1-\tau}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}} + \mathcal{O}(1-\frac{z}{\rho})$$

and

$$\frac{\log^j\big(\frac{1}{1-B(z)}\big)}{1-B(z)} = \frac{\log^j\big(\frac{1}{1-\tau}\big)}{1-\tau} - \log^{j-1}\big(\frac{1}{1-\tau}\big)\frac{1}{(1-\tau)^2}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}\Big(\log\big(\frac{1}{1-\tau}\big)+j\Big) + \mathcal{O}(1-\frac{z}{\rho}).$$

Thus we get

$$\begin{aligned} E_v D_v^r M(z,v) &= 2^{r-1} z\Bigg(\frac{\log^r\big(\frac{1}{1-\tau}\big)}{1-\tau} - (-1)^r r! + \sum_{k=0}^{r-1}\binom{r}{k}\frac{(-1)^k k!\log^{r-k}\big(\frac{1}{1-\tau}\big)}{(1-\tau)} + \frac{r!(-1)^r}{1-\tau} \\ &- \frac{1}{(1-\tau)^2}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}\Big[r\log^{r-1}\big(\frac{1}{1-\tau}\big) + \log^r\big(\frac{1}{1-\tau}\big) \\ &+ \sum_{k=0}^{r-1}\binom{r}{k}(-1)^k k!\Big((r-k)\log^{r-k-1}\big(\frac{1}{1-\tau}\big) + \log^{r-k}\big(\frac{1}{1-\tau}\big)\Big) + r!(-1)^r\Big] + \mathcal{O}(1-\frac{z}{\rho})\Bigg). \end{aligned} \tag{4.15}$$

Since we can use telescope summation to see

$$\sum_{k=0}^{r-1}\binom{r}{k}(-1)^k k!\Big((r-k)\log^{r-k-1}\big(\frac{1}{1-\tau}\big) + \log^{r-k}\big(\frac{1}{1-\tau}\big)\Big) + r!(-1)^r = \log^r\big(\frac{1}{1-\tau}\big), \tag{4.16}$$

for $r \geq 1$, we get by using (4.13) for $n \to \infty$ the following result

$$\begin{aligned} E(Z_n^r) &= \frac{1}{T_n}[z^n]E_v D_v^r M(z,v) \\ &= \frac{1}{T_n}[z^{n-1}] - \frac{2^{r-1}}{(1-\tau)^2}\log^{r-1}\big(\frac{1}{1-\tau}\big)\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}\Big(r + 2\log\big(\frac{1}{1-\tau}\big)\Big) \\ &= \Big(r2^{r-1}\log^{r-1}\big(\frac{1}{1-\tau}\big) + 2^r\log^r\big(\frac{1}{1-\tau}\big)\Big) + \mathcal{O}(\frac{1}{n}) \\ &= \Big(r\log^{r-1}(\varphi(\tau)) + \log^r(\varphi(\tau))\Big) + \mathcal{O}(\frac{1}{n}), \end{aligned} \tag{4.17}$$

and thus the same asymptotic expansion as obtained for the butterflies. Therefore it holds that non-crossing trees behave for this parameter also like butterflies.

4.5 Proof for the climbing depth of non-crossing trees

Since butterflies are simply generated trees, we already know the climbing depth for this kind of trees from [56]. We denote here by X_n resp. $\tilde{X}_n$ the random variables that counts the number steps until a leaf is reached in a random non-crossing tree resp. a random butterfly of size n. Now we turn to the analysis of non-crossing trees. The probability that a random non-crossing tree is climbed in m steps satisfies

$$\begin{aligned} \mathbb{P}\{X_n = m\} &= \sum_{k=1}^{n-1}\sum_{n_1+\cdots+n_k=n-1}\mathbb{P}\{d(root)=k \wedge |B_1|=n_1 \wedge \cdots \wedge |B_k|=n_k\}\times \\ &\times \frac{\mathbb{P}\{\tilde{X}_{n_1}=m-1\}+\cdots+\mathbb{P}\{\tilde{X}_{n_k}=m-1\}}{k}, \end{aligned} \tag{4.18}$$

for $n \geq 2$ and $m \geq 1$ with initial values $\mathbb{P}\{X_1 = 0\} = 1$ and $\mathbb{P}\{X_n = 0\} = 0$ for $n \geq 2$. The recurrence is obtained by distinguishing the trees, where the degree $d(root)$ of the root of the non-crossing tree is k and the sizes of the subtrees $B_1, \ldots, B_k$ are $n_1, \ldots, n_k$ respectively.

4.5.1 Solving the recurrence

We introduce the bivariate generating functions

$$M(z, v) = \sum_{n\geq 1}\sum_{m\geq 0} T_n \mathbb{P}(X_n = m) z^n v^m, \tag{4.19}$$

and

$$N(z, v) = \sum_{n\geq 1}\sum_{m\geq 0} B_n \mathbb{P}(\tilde{X}_n = m) z^n v^m. \tag{4.20}$$

Using

$$\mathbb{P}\{d(root) = k \wedge |B_1| = n_1 \wedge \cdots \wedge |B_k| = n_k\} = \frac{B_{n_1} \ldots B_{n_k}}{T_n}, \tag{4.21}$$

we get from (4.18) the following equation.

$$M(z, v) - z = \frac{zv}{1 - B(z)} N(z, v). \tag{4.22}$$

We know from [56] that $N_m(z) := \sum_{n\geq 1} B_n \mathbb{P}(\tilde{X}_n = m) z^n$ is given by

$$N_m(z) = z\Big(1 - \frac{z}{B(z)}\Big)^m, \quad \text{for } m \geq 0. \tag{4.23}$$

Multiplication with v^m and summing up over $m \geq 0$ leads directly to the following result.

$$N(z, v) = \frac{z}{1 - v\Big(1 - \frac{z}{B(z)}\Big)}. \tag{4.24}$$

Thus we have the following explicit representation for $M(z, v)$:

$$M(z, v) = \frac{z^2 v}{(1 - B(z))(1 - v\Big(1 - \frac{z}{B(z)}\Big))} + z. \tag{4.25}$$

4.5.2 Characterizing the limiting distribution

We further get for $n \geq 2$

$$\mathbb{P}\{X_n = m\} = \frac{1}{T_n}[z^n v^m] M(z, v) = \frac{1}{T_n}[z^{n-2}] \frac{1}{1 - B(z)} \Big(1 - \frac{z}{B(z)}\Big)^{m-1}. \tag{4.26}$$

We use again the local expansion (4.11) and

$$\frac{1}{B^k(z)} = \frac{1}{\tau^k}\Big(1 + \frac{k}{\tau}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1 - \frac{z}{\rho}} + \mathcal{O}(1 - \frac{z}{\rho})\Big), \tag{4.27}$$

which leads to another expansion

$$\Big(1 - \frac{z}{B(z)}\Big)^{m-1} = \sum_{k=0}^{m-1} \binom{m-1}{k} (-1)^k \frac{z^k}{B^k(z)} = \sum_{k=0}^{m-1} \binom{m-1}{k} (-1)^k \frac{\rho^k}{B^k(z)} + \mathcal{O}(1 - \frac{z}{\rho})$$

$$
\begin{aligned}
&= \Big(1-\frac{\rho}{\tau}\Big)^{m-1} + \sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}\sum_{k=0}^{m-1}\binom{m-1}{k}(-1)^k k \frac{\rho^k}{\tau^{k+1}} \\
&= \Big(1-\frac{\rho}{\tau}\Big)^{m-1} - \frac{(m-1)\rho}{\tau^2}\Big(1-\frac{\rho}{\tau}\Big)^{m-2}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}. \qquad (4.28)
\end{aligned}
$$

Thus we finally get

$$
\begin{aligned}
\frac{1}{1-B(z)}\Big(1-\frac{z}{B(z)}\Big)^{m-1} \quad &= \quad \frac{1}{1-\tau}\Big(1-\frac{\rho}{\tau}\Big)^{m-1} - \Big(1-\frac{\rho}{\tau}\Big)^{m-2}\frac{1}{1-\tau}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}\times \\
&\times\Big(\frac{1-\frac{\rho}{\tau}}{1-\tau} + \frac{(m-1)\rho}{\tau^2}\Big). \qquad (4.29)
\end{aligned}
$$

Together with (4.13) we get the following result.

$$
\mathbb{P}\{X_n = m\} = \Big(\tau\cdot\frac{\rho}{\tau}\Big(1-\frac{\rho}{\tau}\Big)^{m-1} + (1-\tau)\cdot\Big(1-\frac{\rho}{\tau}\Big)^{m-2}\frac{(m-1)\rho^2}{\tau^2}\Big)\big(1+\mathcal{O}(\frac{1}{\sqrt{n}})\big). \qquad (4.30)
$$

4.5.3 Computing the moments

From the generating function $M(z,v)$ as given by (4.25) we can also easily compute the r-th factorial moments $E(X_n^{\underline{r}})$. Evaluating the r-th derivative w. r. t. v of $M(z,v)$ at $v=1$ gives

$$
\begin{aligned}
&E_v D_v^r M(z,v) = E_v D_v^r\Big(\frac{z^2 v}{(1-B(z))(1-v\Big(1-\frac{z}{B(z)}\Big))} + z\Big) \\
&= \frac{z^2}{1-B(z)}E_v\Big(\frac{\big(1-\frac{z}{B(z)}\big)^r r!}{(1-v\big(1-\frac{z}{B(z)}\big))^{r+1}} + r\frac{\big(1-\frac{z}{B(z)}\big)^{r-1}(r-1)!}{(1-v\big(1-\frac{z}{B(z)}\big))^{r}}\Big) = \frac{r!z^2}{1-B(z)}\frac{B^{r+1}(z)}{z^{r+1}}\big(1-\frac{z}{B(z)}\big)^{r-1}. \qquad (4.31)
\end{aligned}
$$

Using (4.10) we get the expansion

$$
B^k(z) = \tau^k - k\tau^{k-1}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}} + \mathcal{O}(1-\frac{z}{\rho}). \qquad (4.32)
$$

By combining (4.29) and (4.32) we get for $r \geq 2$

$$
\begin{aligned}
\frac{B^{r+1}(z)}{1-B(z)}\Big(1-\frac{z}{B(z)}\Big)^{r-1} \quad &= \quad -\frac{\tau^r}{1-\tau}\Big(1-\frac{\rho}{\tau}\Big)^{r-2}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}\times \\
&\times\Big(\tau\Big[\frac{1-\frac{\rho}{\tau}}{1-\tau} + \frac{(r-1)\rho}{\tau^2}\Big] + (r+1)(1-\frac{\rho}{\tau})\Big) + \mathcal{O}(1-\frac{z}{\rho}), \qquad (4.33)
\end{aligned}
$$

where for $r=1$ we simply have

$$
\frac{B^2(z)}{1-B(z)} \quad = \quad -\frac{\tau}{1-\tau}\Big(1-\frac{\rho}{\tau}\Big)^{r-2}\sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}}\sqrt{1-\frac{z}{\rho}}\Big(2-\frac{\tau}{1-\tau}\Big) + \mathcal{O}(1-\frac{z}{\rho}). \qquad (4.34)
$$

Since we known that

$$
E(Z_n^{\underline{r}}) = \frac{1}{T_n}[z^n]E_v D_v^r M(z,v) = \frac{r!}{T_n}[z^{n+r-1}]\frac{B^{r+1}}{1-B(z)}\Big(1-\frac{z}{B(z)}\Big)^{r-1}, \qquad (4.35)
$$

we get for for $n \to \infty$ the following result for $r \geq 2$

$$\mathbb{E}(Z_n^r) = r!\tau\frac{\tau^r}{\rho^r}\Big(1-\frac{\rho}{\tau}\Big)^{r-1} + r!(1-\tau)\frac{\tau^r}{\rho^r}\Big(1-\frac{\rho}{\tau}\Big)^{r-1}(r+1-2p) + \mathcal{O}(\frac{1}{n}), \tag{4.36}$$

and the expectation is given by

$$\mathbb{E}(Z_n) = \tau\frac{\tau}{\rho} + 2(1-\tau)\frac{\tau}{\rho} + \mathcal{O}(\frac{1}{n}). \tag{4.37}$$

Remark 2. Our studies of non-crossing trees were motivated by the following fact. Let $T_{n,k}$ denote the number of non-crossing trees with k butterflies. We use the bivariate generating function $T(z,v) = \sum_{n\geq 1}\sum_{k\geq 0} T_{n,k} z^n v^k$. The combinatorial description of the non-crossing trees can be turned into the following equation.

$$T(z,v) \quad = \quad \frac{z}{1-vB(z)}.$$

Extracting coefficients by an application of (4.3) leads to the result

$$T_{n,k} \quad = \quad \frac{k}{n-1}\binom{3n-k-4}{2n-3}.$$

Let Y_n denote the random variable, that counts the number of non-crossing trees with k butterflies and let $t_{n,k} = \mathbb{P}(Y_n = k)$ denote the probability, that a non-crossing tree has k butterflies.

$$t_{n,k} \quad = \quad \frac{T_{n,k}}{T_n} = \frac{(2n-1)k\binom{3n-k-4}{2n-3}}{(n-1)\binom{3n-3}{n-1}}.$$

Thus we get

$$\mathbb{E}(Y_n) \quad = \quad 2\frac{n-1}{n}.$$

Using Stirling's formula we can calculate the limit distribution of Y_n. We get the following result.

Corollar 5. *The limit distribution of the random variable that counts the number of non-crossing trees with k butterflies is given by*

$$\lim_{n\to\infty} t_{n,k} \quad = \quad \frac{4k}{3^{k+1}}.$$

Part II

Label-based parameters in increasing trees and related tree families

Chapter 5

Label-based parameters in increasing trees

5.1 Introduction

In this chapter we present a unifying approach for studying several label-based parameters for increasing trees. In contrast to global parameters like height, width, etc., which depend on the whole tree, label-based parameters depend on a specific sub-structure of the tree which depends itself on the label j. Let $X_{n,j}$ denote the random variable, which counts a certain label-based parameter of a specified node j in a random size-n tree, where $1 \le j \le n$. We use a recursive approach, which leads for all simple families of increasing trees (not only those which can be described via an insertion process) to a closed formula for suitable trivariate generating functions of the probabilities $\mathbb{P}\{X_{n,j} = m\}$. We obtain formulæ for the probabilities $\mathbb{P}\{X_{n,j} = m\}$ and the s-th factorial moments $\mathbb{E}\big((X_{n,j})^{\underline{s}}\big) = \sum_{m\ge 0} m^{\underline{s}}\mathbb{P}\{X_{n,j} = m\}$ for grown simple families of increasing trees.

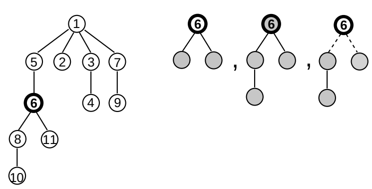

Figure 5.1: A size n=11 increasing tree with $j = 6$: the outdegree of node 6 equals 2, the subtree size of node 6 equals 4 and the branches consists of one size 2 tree and one leaf.

This approach has already been successfully carried out for several parameters, e.g. *level of node* j or *number of descendants of node* j in a grown simple increasing tree of size n, see [46] and [68]. We illustrate this approach here for several parameters like *branching structure*, *subtree size*, *node degree*, *distance between specified nodes*.

In order to state our results concerning arbitrary label-based parameters we introduce the following generating functions. For the root we set up

$$M(z, v) = \sum_{n\ge 1}\sum_{m\ge 1} \mathbb{P}\{X_{n,1} = m\} T_n \frac{z^n}{n!} v^m, \tag{5.1}$$

where for $j \ge 1$ we use a trivariate generating function,

$$N(z, u, v) := \sum_{k\ge 0}\sum_{j\ge 1}\sum_{m\ge 0} \mathbb{P}\{X_{k+j,j} = m\} T_{k+j} \frac{z^{j-1}}{(j-1)!}\frac{u^k}{k!} v^m. \tag{5.2}$$

5.2 Results for label-based parameters

Theorem 16. *The function $N(z,u,v)$ as defined in equation (5.1), which is the trivariate generating function of the probabilities $\mathbb{P}\{X_{n,j}=m\}$, which give the probability that a certain label-based parameter, which depends only on the subtree rooted at node j, in a randomly chosen size-n tree of a simple family of increasing trees with degree-weight generating function $\varphi(t)$, equals m, is given by the following formula:*

$$N(z,u,v)=\frac{\varphi\big(T(z+u)\big)\frac{\partial}{\partial u}M(u,v)}{\varphi\big(T(u)\big)}, \tag{5.3}$$

where $M(z,v)$ is defined by (5.1).

Corollar 6. *For grown simple families of increasing trees the probability distribution and the factorial moments of $X_{n,j}$, where the parameter depends only on the subtree rooted at node j, are given by*

$$\mathbb{P}\{X_{n,j}=m\}=\frac{c_1^{j-1}\binom{\frac{c_2}{c_1}+j-1}{j-1}}{\binom{n-1}{n-j}\varphi_0 c_1^{n-1}\binom{n-1+\frac{c_2}{c_1}}{n-1}}[u^{n-j}v^m]\frac{\frac{\partial}{\partial u}M(u,v)}{(1-c_1u)^{j-1}}, \tag{5.4}$$

and

$$\mathbb{E}\big(X_{n,j}^{\underline{s}}\big)=\frac{c_1^{j-1}\binom{\frac{c_2}{c_1}+j-1}{j-1}}{\binom{n-1}{n-j}\varphi_0 c_1^{n-1}\binom{n-1+\frac{c_2}{c_1}}{n-1}}[u^{n-j}]\frac{M_s'(u)}{(1-c_1u)^{j-1}}, \tag{5.5}$$

were we use the abbreviation with $M_s(z)=E_vD_v^sM(z,v)$. For Case A we set $c_2=0$ and for Case B $d=\frac{c_1}{c_2}+1$.

Remark 3. For the depth $D_{n,j}$ the trivariate generating function has a slightly different form because the depth depends on the ascendants of j. See [68] for results concerning the depth.

Remark 4. For grown simple families of increasing trees one can also obtain the probabilities $\mathbb{P}\{X_{n,j}=m\}$ by conditioning on $Z_{n,j}$, the size of the subtree rooted at node j:

$$\mathbb{P}\{X_{n,j}=m\}=\sum_{k=1}^{n+1-j}\mathbb{P}\{X_{n,j}=m|Z_{n,j}=k\}\mathbb{P}\{Z_{n,j}=k\}=\sum_{k=1}^{n+1-j}\mathbb{P}\{X_{k,1}=m\}\mathbb{P}\{Z_{n,j}=k\}, \tag{5.6}$$

given the probabilities $\mathbb{P}\{X_{n,1}=m\}$ and $\mathbb{P}\{Z_{n,j}=k\}$. See Section 5.4 ([46]) for explicit formulæ for $\mathbb{P}\{Z_{n,j}=k\}$.

5.3 Deriving the generating function for the probabilities

We consider in this section the random variable $X_{n,j}$, which counts a certain label j based parameter in a size n increasing tree. We will present two approaches for deriving the generating functions for the probabilities $\mathbb{P}\{X_{n,j}=m\}$.

5.3.1 A recurrence for the probabilities

We consider in this section the random variable $X_{n,j}$, which counts a certain label-based parameter for the node labelled j in a size n increasing tree. In the following we give a general recurrence for the probability $\mathbb{P}\{X_{n,j}=m\}$. At first we setup a bivariate generating function $M(z,v)$ for the probabilities $\mathbb{P}\{X_{n,1}=m\}$ as follows:

$$M(z,v)=\sum_{n\ge1}\sum_{m\ge1}\mathbb{P}\{X_{n,1}=m\}T_n\frac{z^n}{n!}v^m. \tag{5.7}$$

For increasing trees of size n with root-degree r and subtrees with sizes $k_1, \dots, k_r$, enumerated from left to right, where the node labelled by j lies in the leftmost subtree and is the i-th smallest node in this subtree, we can reduce the computation of the probabilities $\mathbb{P}\{X_{n,j} = m\}$ to the probabilities $\mathbb{P}\{X_{k_1,i} = m\}$, when the label based parameter does only depend on the subtree of node j. Note that for the depth of node j we have the dependence on $\mathbb{P}\{X_{k_1,i} = m-1\}$, since the depth increases by one after attaching the subtree of size k_1 to the root.

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

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

the $i-1$ labels smaller than j are chosen from $2, 3, \dots, j-1$, the k_1-i labels larger than j are chosen from $j+1, \dots, n$, and the remaining $n-1-k_1$ labels are distributed to the second, third, $\dots$, r-th subtree. Again due to symmetry arguments we obtain a factor r, if the node j is the i-th smallest node in the second, third, $\dots$, r-th subtree. Summing up over all choices for the rank i of label j in its subtree, the subtree sizes $k_1, \dots, k_r$, and the degree r of the root node gives the following recurrence (5.8).

$$\begin{aligned}\mathbb{P}\{X_{n,j} = m\} = \sum_{r \ge 1} r\varphi_r \sum_{\substack{k_1 + \cdots + k_r = n-1,\\ k_1, \dots, k_r \ge 1}} & \frac{T_{k_1} \cdots T_{k_r}}{T_n} \times \\ & \times \sum_{i=1}^{\min\{k_1, j-1\}} \mathbb{P}\{X_{k_1,i} = m\} \binom{j-2}{i-1}\binom{n-j}{k_1-i}\binom{n-1-k_1}{k_2, k_3, \dots, k_r},\end{aligned} \tag{5.8}$$

for $n \ge j \ge 2$.

To treat this recurrence (5.8) we set $n := k + j$ with $k \ge 0$ and define the trivariate generating function

$$N(z,u,v) := \sum_{k \ge 0} \sum_{j \ge 1} \sum_{m \ge 0} \mathbb{P}\{X_{k+j,j} = m\} T_{k+j} \frac{z^{j-1}}{(j-1)!} \frac{u^k}{k!} v^m. \tag{5.9}$$

Multiplying (5.8) with $T_{k+j} \frac{z^{j-2}}{(j-2)!} \frac{u^k}{k!} v^m$ and summing up over $k \ge 0$, $j \ge 2$ and $m \ge 0$ gives then $\frac{\partial}{\partial z} N(z,u,v)$ and $\varphi'\big(T(z+u)\big) N(z,u,v)$ for the left and right hand side of (5.8), respectively. Note that for the depth we get $v\varphi'\big(T(z+u)\big)N(z,u,v)$, where the extra factor v is due to the usage of $\mathbb{P}\{X_{k_1,i} = m-1\}$ instead of $\mathbb{P}\{X_{k_1,i} = m\}$. Since these are essentially straightforward, but lengthy computations, they are omitted here; similar considerations are done in [68]. We obtain the following differential equation

$$\frac{\partial}{\partial z} N(z,u,v) = \varphi'\big(T(z+u)\big) N(z,u,v), \tag{5.10}$$

together with the initial condition

$$N(0,u,v) = \sum_{k \ge 0} \sum_{m \ge 0} \mathbb{P}\{X_{k+1,1} = m\} T_{k+1} \frac{u^k}{k!} v^m = M'(u,v). \tag{5.11}$$

The general solution of equation (5.10) is given by

$$N(z,u,v) = C(u,v) \exp\Big(\int_0^z \varphi'\big(T(t+u)\big) dt\Big), \tag{5.12}$$

with some function $C(u,v)$. Adapting to the initial condition (5.11) gives the required solution

$$N(z,u,v) = \frac{\partial}{\partial u} M(u,v) \exp\Big(\int_0^z \varphi'\big(T(t+u)\big) dt\Big). \tag{5.13}$$

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

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

which leads from (5.13) to the following main result. From (5.3) we easily get the probability distribution and the factorial moments of $X_{n,j}$ by extracting coefficients. For grown simple families of increasing trees it holds

$$\frac{\varphi\big(T(z+u)\big)}{\varphi\big(T(u)\big)} = \frac{1}{\Big(1-\frac{c_1 z}{1-c_1 u}\Big)^{\frac{c_2}{c_1}+1}}, \qquad [z^{j-1}]\frac{\varphi\big(T(z+u)\big)}{\varphi\big(T(u)\big)} = \binom{\frac{c_2}{c_1}+j-1}{j-1}\frac{c_1^{j-1}}{(1-c_1u)^{j-1}}, \tag{5.14}$$

where for Case A $c_2 = 0$ and Case B $d = \frac{c_1}{c_2}+1$. Hence by using (10) we obtain Corollary 6.

5.3.2 A combinatorial approach

It is also possible to derive Theorem 16 in a combinatorial way. This was done in [68] Panholzer and Prodinger where they established a description for the depth of node j, which also holds for all label-based parameters of node j. Their approach is summarized as follows.

It is convenient to think of specifically tricolored increasing trees, where the coloring is as follows: exactly one node is colored red, all nodes with a smaller label than the red node are colored black, and all nodes with a larger label than the red node are colored white. We are interested in a parameter, e.g. *depth, node degree, subtree size, etc.*, depending on the red node. Let us consider such a tricolored increasing tree and assume that the out-degree of the root node of T is $r \geq 1$.

We further assume that the red node of T is not the root node. Then the red node is located in one of the r subtrees of the root of T; let us assume that it is in the r-th subtree. Let us now consider these r subtrees. After order preserving relabellings, each subtree $T_1, \dots, T_r$ is an increasing tree by itself. The r-th subtree is again a tricolored increasing tree with one red, j_1 black and k_1 white nodes, whereas the remaining $r-1$ subtrees are only bi-colored in such a way that the nodes with the j_i smallest labels (with $2 \leq i \leq r$ and $0 \leq j_i \leq T_i$) are colored black and the remaining k_i nodes in the subtrees are colored white. Then such a specific r-tuple $T_1, \dots, T_r$ of colored increasing trees appears exactly $\binom{j_1+\cdots+j_r}{j_1,\dots,j_r}\binom{k_1+\cdots+k_r}{k_1,\dots,k_r}$ times, where the labels of the $j_1+\cdots+j_r$ black nodes and the $k_1+\cdots+k_r$ white nodes are distributed over the black and white nodes in $T_1, \dots, T_r$ in an order preserving fashion.

Of course this corresponds to a tricolored increasing tree T of size $|T| = j+k+1$ with $j = j_1+\cdots+j_r$ black nodes and $k = k_1+\cdots+k_r$ white nodes.

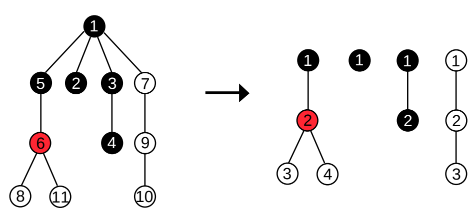

Figure 5.2: Decomposition of a tricolored size 11 tree with root degree 4.

We introduce generating functions which are exponential in both variables z and u, where z marks the black nodes and u marks the white nodes, $f(z,u) = \sum_{j,k\geq 0} f_{j,k}\frac{z^j u^k}{j!k!}$ for sequences $f_{j,k}$ and $f(z,u,v) = \sum_{j,k,m\geq 0} f_{j,k,m}\frac{z^j u^k}{j!k!}v^m$ for sequences $f_{j,k,m}$, where v marks the depth of the red ball.

With this setting, the total weight of all suitably tricolored increasing trees with j black and k white nodes, where the parameter of the red node is exactly m, is given by $\mathbb{P}\{X_{j+k+1,j+1} = m\}T_{j+k+1}$, and thus its generating function is

$$\sum_{k\geq 0}\sum_{j\geq 0}\sum_{m\geq 0}\mathbb{P}\{X_{k+j+1,j+1}=m\}T_{k+j+1}\frac{z^j u^k}{j!k!}v^m = N(z,u,v), \tag{5.15}$$

whereas the total weight of suitably bi-colored increasing trees with j black and k white nodes is T_{k+j} and its generating function is

$$\sum_{k\geq 0}\sum_{j\geq 0}T_{k+j}\frac{z^j u^k}{j!k!} = \sum_{l\geq 0}\frac{T_l}{l!}\sum_{k=0}^{l}\frac{l!}{k!(l-k)!}z^k u^{l-k} = T(z+u). \tag{5.16}$$

The $r-1$ bi-colored trees and the tricolored tree lead to $T(z+u)^{r-1}N(z,u,v)$. Since the red ball can be in the first, second, ..., r-th subtree, we additionally get a factor r. Furthermore, according to (2), the event that the root has out-degree r leads to a factor φ_r. Summing over $r\geq 1$ leads to $\sum_{r\geq 1} r\varphi_r T(z+u)^{r-1}N(z,u,v) = \varphi'(T(z+u))N(z,u,v)$.
Since the root node labelled by 1 is colored black the formal description leads now to

$$\frac{\partial}{\partial z}N(z,u,v) = v\varphi'(T(z+u))N(z,u,v). \tag{5.17}$$

Remark 5. The fact that the *depth* of the red node in the subtree is one more than the depth of the red node in the subtree leads to the additional factor v on the right hand side of (5.17). The equation (5.17) without the additional factor v holds for all parameters depending only on the subtree of the red node.

5.3.3 Combinatorial description for several nodes

Let $\boldsymbol{X}_{n;j_1,\dots,j_r}$ denote the random vector $(X_{n;j_1},\dots,X_{n;j_r})$, which counts a certain parameter depending on the labels $j_1,\dots,j_r$ in a random grown increasing tree of size n.
We have r different colors $c_1,\dots,c_r$. Further we have $r+1$ tones of grey $g_1,\dots,g_{r+1}$, where one can think of g_1 as black and g_{r+1} as white. Now one has to think of specifically $2r+1$ colored increasing trees, where the coloring is as follows: for $1\leq i\leq r$ exactly one node is colored c_i. The smallest labelled node of the r chosen nodes is colored c_1 and in general the i-th smallest node is colored c_i, $1\leq i\leq r$. All nodes with labels smaller than the label of node colored c_1 are colored black, and the all nodes with labels bigger than the label of node colored c_r are colored white. The nodes with a label between the label of the node colored c_i and the label of the node colored c_{i+1} are colored g_{i+1}.

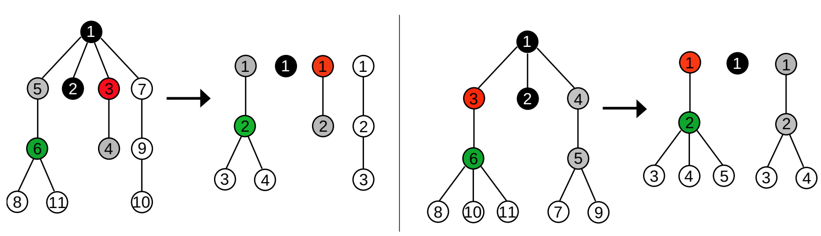

Figure 5.3: Two different decompositions of a 5-colored size 11 tree.

Let us consider such a $2r+1$ increasing tree and assume that the out-degree of the root node of T is $s\geq 1$.
We further assume that root node is colored black. Then the nodes colored $c_1,\dots,c_r$ are located in the s subtrees of the root of T;

- At first let us assume that nodes with colors $c_1,\dots,c_r$ are all in the first subtree of T. After order preserving relabellings, each subtree $T_1,\dots,T_r$ is an increasing tree by itself. The first subtree is again a $(2r+1)$-colored increasing tree with r nodes with colors $c_1, \dots, c_r$ and $j_1^{[g_i]}$ nodes colored g_i, for $1 \le i \le r+1$.

In contrast the remaining $s-1$ subtrees are only $(r+1)$-colored in such a way that $j_l^{[g_i]}$ nodes are colored g_i, with $2 \le l \le s$, $0 \le j_l^{[g_i]} \le T_l$ and $\sum_{i=1}^{r+1} j_l^{[g_i]} = T_l$.

Then such a specific s-tuple $T_1, \dots, T_s$ of colored increasing trees appears exactly

$$\binom{j_1^{[g_1]} + \cdots + j_s^{[g_1]}}{j_1^{[g_1]}, \dots, j_s^{[g_1]}} \binom{j_1^{[g_2]} + \cdots + j_s^{[g_2]}}{j_1^{[g_2]}, \dots, j_s^{[g_2]}} \cdots \binom{j_1^{[g_{r+1}]} + \cdots + j_s^{[g_{r+1}]}}{j_1^{[g_{r+1}]}, \dots, j_s^{[g_{r+1}]}} \tag{5.18}$$

times. The labels of the $j_1^{[g_i]} + \cdots + j_r^{[g_i]}$, $1 \le i \le r+1$, grey nodes in the original $2r+1$ colored tree are distributed over the g_i colored grey nodes in $T_1, \dots, T_s$ in an order preserving fashion.

Let $\mathbf{z}_{r+1}$ denote the random vector $(z_1, \dots, z_{r+1})$ and $\mathbf{j}_{r+1}$ the random vector $(j_1, \dots, j_{r+1})$. Further we use the compact notation $\mathbf{j}_{r+1}! = j_1! \dots j_{r+1}!$. We introduce generating functions which are exponential in all variables z_i, where z_i marks the i-th shade of grey g_i, $1 \le i \le r+1$,

$$f(\mathbf{z}_{r+1}) = \sum_{\mathbf{j}_{r+1} \ge \mathbf{0}} f_{\mathbf{j}_{r+1}} \frac{\mathbf{z}_{r+1}^{\mathbf{j}_{r+1}}}{\mathbf{j}_{r+1}!} \tag{5.19}$$

for sequences $f_{\mathbf{j}_{r+1}} = f_{j_1, \dots, j_{r+1}}$ and

$$f(\mathbf{z}_{r+1}, \mathbf{v}_r) = \sum_{\mathbf{j}_{r+1} \ge 0} \sum_{\mathbf{m}_r \ge 0} f_{\mathbf{j}_{r+1}, \mathbf{m}_r} \frac{\mathbf{z}_{r+1}^{\mathbf{j}_{r+1}}}{\mathbf{j}_{r+1}!} \mathbf{v}_r^{\mathbf{m}_r} \tag{5.20}$$

for sequences $f_{\mathbf{j}_{r+1}, \mathbf{m}_r} = f_{j_1, \dots, j_{r+1}, m_1, \dots, m_r}$, where v_i marks the ball colored c_i.

In the following we will use the notations $J_{r+1} = \sum_{k=1}^{r+1} j_k$, $\mathbf{J}_r = (J_1, \dots, J_r)$ and $\mathbf{r} = (1, \dots, r)$. With this setting, the total weight of all suitably $2r+1$ colored increasing trees with j_i nodes colored g_i, where the parameters of the colored nodes are exactly $m_1, \dots, m_r$, is given by

$$\mathbb{P}\{X_{J_{r+1}+r; J_1+1} = m_1, \dots, X_{J_{r+1}+r; J_r+r} = m_r\} T_{J_{r+1}+r} = \mathbb{P}\{\mathbf{X}_{J_{r+1}+r; \mathbf{J}_r+\mathbf{r}} = \mathbf{m}_r\} T_{J_{r+1}+r}, \tag{5.21}$$

and thus its generating function is

$$N(\mathbf{z}_{r+1}, \mathbf{v}_r) := \sum_{\mathbf{j}_{r+1} \ge 0} \sum_{\mathbf{m}_r \ge 0} \mathbb{P}\{\mathbf{X}_{J_{r+1}+r; \mathbf{J}_r+\mathbf{r}} = \mathbf{m}_r\} T_{J_{r+1}+r} \frac{\mathbf{z}_{r+1}^{\mathbf{j}_{r+1}}}{\mathbf{j}_{r+1}!} \mathbf{v}_r^{\mathbf{m}_r}, \tag{5.22}$$

whereas the total weight of suitably $r+1$ colored increasing trees with colors $g_1, \dots, g_{r+1}$ is $T_{J_{r+1}}$ and its generating function is

$$T(z_1 + \cdots + z_{r+1}) = \sum_{\mathbf{j}_{r+1} \ge 0} T_{J_{r+1}} \frac{\mathbf{z}_{r+1}^{\mathbf{j}_{r+1}}}{\mathbf{j}_{r+1}!}. \tag{5.23}$$

The $s-1$ trees equipped with $r+1$ colors and the $2r+1$ colored tree lead to $T(z_1 + \cdots + z_r)^{r-1} N(\mathbf{z}_{r+1}, \mathbf{v}_r)$. Since the $2r+1$-colored tree can be the first, second, …, s-th subtree, we additionally get a factor s. The event that the root has out-degree s leads to a factor φ_s. Summing over $s \ge 1$ leads to $\sum_{s \ge 1} s \varphi_s T(z_1 + \cdots + z_r)^{s-1} N(\mathbf{z}_{r+1}, \mathbf{v}_r) = \varphi'(T(z_1 + \cdots + z_r)) N(\mathbf{z}_{r+1}, \mathbf{v}_r)$. Since the root is colored by g_1 (black) one obtains the inhomogeneous differential equation

$$\frac{\partial}{\partial z_1} N(\mathbf{z}_{r+1}, \mathbf{v}_r) = \varphi'(T(z_1 + \cdots + z_r)) N(\mathbf{z}_{r+1}, \mathbf{v}_r) + R_r(\mathbf{z}_{r+1}, \mathbf{v}_r), \tag{5.24}$$

where the inhomogeneous part $R_r(\mathbf{z}_{r+1}, \mathbf{v}_r)$ corresponds to the cases where the r nodes colored

$c_1, \dots, c_r$ are not in the same subtree.

- Let us now consider the cases where the nodes colored $c_1, \dots, c_r$ are distributed over at least two subtrees of the root. We have to consider all different partitions of the set $\{j_1, \dots, j_r\}$ or equivalently all partitions of the set $\{1, \dots, r\}$, because every partition corresponds to a different distribution of the nodes $j_1, \dots, j_r$ to the subtrees of the root. We denote with Π the set containing all partitions of $\{1, \dots, r\}$ and with Π_l the set of all partitions of $\{1, \dots, r\}$ into l non-empty subsets. Further for $\mathbf{p} \in \Pi_l$ we get

$$\mathbf{p} = \bigcup_{i=1}^{l} \mathbf{p}_i = \bigcup_{i=1}^{l} \{p_{i,1}, \dots, p_{i,s_i}\}, \tag{5.25}$$

where $|\mathbf{p}_i| = s_i$ and $\sum_{i=1}^{l} s_i = r$. Let $N(\mathbf{z}_{r+1}, \mathbf{v}_{\mathbf{p}_i})$ be defined analogous to (5.22),

$$N(\mathbf{z}_{r+1}, \mathbf{v}_{\mathbf{p}_i}) = N(\mathbf{z}_{r+1}, v_{p_{i,1}}, \dots, v_{p_{i,s_i}}). \tag{5.26}$$

Now we slightly overload the notations $J_{r+1} = \sum_{k=1}^{r+1} j_k$ and $\mathbf{J}_r = (J_1, \dots, J_r)$ by using $J_{p_{i,t}} = \sum_{k=1}^{p_{i,t}} j_k$ and $\mathbf{J}_{\mathbf{p}_i} = (J_{p_{i,1}}, \dots, J_{p_{i,s_i}})$,

$$N(\mathbf{z}_{r+1}, \mathbf{v}_{\mathbf{p}_i}) := \sum_{\mathbf{j}_{r+1} \geq 0} \sum_{\mathbf{m}_{\mathbf{p}_i} \geq 0} \mathbb{P}\{\mathbf{X}_{J_{r+1}+s_i; \mathbf{J}_{\mathbf{p}_i}+\mathbf{s}_i} = \mathbf{m}_r\} T_{J_{r+1}+s_i} \frac{\mathbf{z}_{r+1}^{\mathbf{j}_{r+1}}}{\mathbf{j}_{r+1}!} \mathbf{v}_{\mathbf{p}_i}^{\mathbf{m}_{\mathbf{p}_i}}, \tag{5.27}$$

where $\mathbf{s}_i = (1, 2 \dots, s_i)$ and $\mathbf{m}_{\mathbf{p}_i} = (m_{p_{i,1}}, \dots, m_{p_{i,s_i}})$. The definition (5.27) is a generalization of (5.22) because if all nodes $\{j_1, \dots, j_r\}$ are in the same subtree, then $\mathbf{p} = \{1, 2, \dots, r\}$ and

$$N(\mathbf{z}_{r+1}, \mathbf{v}_{\mathbf{p}}) := N(\mathbf{z}_{r+1}, \mathbf{v}_r). \tag{5.28}$$

Thus we can explicitly specify the right hand side of (5.24):

$$\varphi'(T(z_1+\dots+z_r))N(\mathbf{z}_{r+1}, \mathbf{v}_r) + R_r(\mathbf{z}_{r+1}, \mathbf{v}_r) = \sum_{l=1}^{r} \sum_{\substack{\mathbf{p} \in \Pi_l \\ \mathbf{p} = \cup_{i=1}^{l} \mathbf{p}_i}} l! \varphi^{(l)}\big(T(z_1+\dots+z_r)\big) \prod_{i=1}^{l} N(\mathbf{z}_{r+1}, \mathbf{v}_{\mathbf{p}_i}). \tag{5.29}$$

Thus the inhomogeneous part $R_r(\mathbf{z}_{r+1}, \mathbf{v}_r)$ is given by

$$R_r(\mathbf{z}_{r+1}, \mathbf{v}_r) = \sum_{l=2}^{r} \sum_{\substack{\mathbf{p} \in \Pi_l \\ \mathbf{p} = \cup_{i=1}^{l} \mathbf{p}_i}} l! \varphi^{(l)}\big(T(z_1 + \dots + z_r)\big) \prod_{i=1}^{l} N(\mathbf{z}_{r+1}, \mathbf{v}_{\mathbf{p}_i}). \tag{5.30}$$

Remark 6. Although the differential equation for $N(\mathbf{z}_{r+1}, \mathbf{v}_r)$ is quite involved, it can be used for successive calculations of solutions for $r = 2, 3, \dots$. At least in some cases it should be possible to guess a general solution for $N(\mathbf{z}_{r+1}, \mathbf{v}_r)$, which should be proved by induction with respect to r.

Remark 7. It seems to be more difficult to get the inhomogeneous part $R_r(\mathbf{z}_{r+1}, \mathbf{v}_r)$, (5.30), of the differential equation by setting up a recurrence for $\mathbb{P}\{\mathbf{X}_{J_{r+1}+r; \mathbf{J}_r+\mathbf{r}} = \mathbf{m}_r\}$, because the recurrence turns out to be more and more involved.

5.3.4 Specific description of recursive trees

For some grown simple families of trees like recursive trees there are alternatives to the description by the recurrence (5.8). When considering a random recursive tree of size n for $n \geq 2$ we can decompose the tree into the subtree rooted at the vertex labelled 2 and the rest of the tree, which is then rooted at the root of the original tree.

It is known, see Smythe and Mahmoud [52] and Dobrow and Fill [17], or also [46] that the cardinality J_n of the subtree rooted at the vertex labelled 2 is uniformly distributed on $\{1, \dots, n-1\}$ for $n \geq 2$. Further

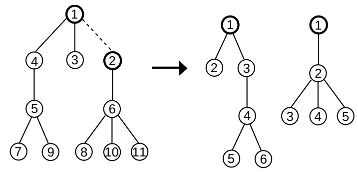

Figure 5.4: Decomposition of a size 11 recursive tree with $J_{11} = 5$.

it holds that the subtree rooted at the vertex labelled 2 is again a random recursive tree of cardinality J_n, which is independent of rest of the tree, which is also a random recursive tree of size $n - J_n$. By conditioning on the value of J_n (the size of the subtree rooted at node 2) we get for $n \geq j \geq 3$ the following recurrence for an arbitrary label-based parameter.

$$T_n \mathbb{P}\{X_{n;j} = m\} = 2 \sum_{k=2}^{n-1} T_k T_{n-k} \sum_{i=2}^{\min\{j-1,k\}} \mathbb{P}\{X_{k;i} = m\} \binom{j-3}{i-2} \binom{n-j}{k-i}. \tag{5.31}$$

5.3.5 Specific description of plane oriented recursive trees

When considering a random plane oriented recursive tree of size n we can decompose the tree into the first subtree of the root with cardinality J_n and the rest of the tree with cardinality $n - J_n$, which is then rooted at the root of the original tree.

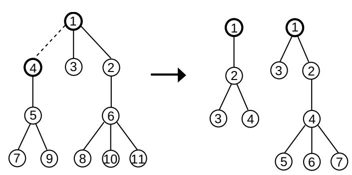

Figure 5.5: Decomposition of a size 11 PORT with $J_{11} = 4$.

$$\mathbb{P}\{J_n = m\} = \frac{(n-1)!}{T_n} [z^{n-1} v^m] \frac{\partial}{\partial z} A_j(z, v) = \frac{T_m T_{n-m} (n-1)!}{T_n m! (n-m-1)!}, \tag{5.32}$$

where T_n is given by (10) with $c_1 = 2$, $c_2 = -1$ and $\varphi_0 = 1$:

$$T_n = 2^{n-1} (n-1)! \binom{n - \frac{3}{2}}{n-1}. \tag{5.33}$$

Remark 8. These alternative recursive descriptions can also be used for other parameters which are not label-based.

Remark 9. Recently there has been some interest in a continuous time increasing tree. The setting in [70] for the continuous time increasing trees is the following. Given a weight function $w : \mathbb{N} \to \mathbb{R}_+$, let $N(t)$ be a Markovian pure birth process with $N(1) = 1$ and birth rates

$$\mathbb{P}\{N_{t+\Delta_t} = n + 1 | N_t = n\} = w(n) \Delta_t + o(\Delta_t), \tag{5.34}$$

Under several mild conditions on $w(k)$ (see [70]) it is assured that the Markov chain does not blow up in finite time.

The discrete random variable $X_{n,j}$ has a continuous analogue, the random variable $\xi_{t,j}$, which counts the size of a certain variety of the node labelled j at the time t, $t \geq 1$, in a continuous time random increasing tree. If we inspect only at the stopping times $(\tau_n)_{n\in\mathbb{N}}$, where $\tau_n = \inf\{t \geq 1 : N_t = n\}$, we have

$$\xi_{\tau_n,j} \overset{(d)}{=} X_{n,j}. \tag{5.35}$$

Thus we have a generalization of the random variables studied before. It is possible to use results of the discrete time model for linear weight functions $w(n)$ ("grown simple families of increasing trees") to obtain results for the continuous time case for linear weight functions by using the basic relation

$$\mathbb{P}\{\xi_{t,j} = m\} = \sum_{n=1}^{\infty} \mathbb{P}\{\xi_{t,j} = m | N_t = n\}\mathbb{P}\{N_t = n\} = \sum_{n=1}^{\infty} \mathbb{P}\{X_{n,j} = m\}\mathbb{P}\{N_t = n\}. \tag{5.36}$$

Furthermore it is obvious that the limit laws must be the same. It is now obvious how we can use explicit formulæ for $\mathbb{P}\{X_{n,j} = m\}$ to derive results for $\xi_{t,j}$ because it is well known that

$$\mathbb{P}\{N_t = n\} = p_{t,n} = \sum_{i=1}^{n} A_n^{(i)} e^{-w(i)t} \quad \text{for } n \in \mathbb{N}, \quad \text{where } A_n^{(i)} = \frac{\prod_{k=1}^{n-1} w(k)}{\prod_{\substack{k=1\\k\neq i}}^{n} \big(w(k) - w(i)\big)}, \tag{5.37}$$

and $w(i) \neq w(j)$ for $i \neq j$.

5.4 Subtree size of node j

Let $X_{n,j}$ be the size of the subtree rooted node j in a random grown simple increasing tree of size n, with $X_{n,n} = 1$. The subtree size of the root in any random grown simple increasing tree of size n equals n, so we get

$$\frac{\partial}{\partial u} M(u,v) = \sum_{k\geq 0}\sum_{m\geq 0} \mathbb{P}\{X_{k+1,1} = m\} T_{k+1}\frac{u^k}{k!} v^m = \sum_{k\geq 0} T_{k+1}\frac{u^k}{k!} v^{k+1} = vT'(uv) = v\varphi\big(T(uv)\big). \quad (5.38)$$

5.4.1 Results for the subtree size of node j

Theorem 17. *The probabilities $\mathbb{P}\{D_{n,j} = m\}$, which give the probability that the node with label j in a randomly chosen size-n tree of a very simple family of increasing trees as given by Lemma 1 (Chapter 1), has exactly m descendants, are, for $m \geq 1$ given by the following formula:*

$$\mathbb{P}\{X_{n,j} = m\} = \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{m-1+\frac{c_2}{c_1}}{m-1}\binom{n-m-1}{j-2}}{\binom{n-1}{j-1}\binom{n-1+\frac{c_2}{c_1}}{n-1}}. \quad (5.39)$$

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

$$\mathbb{E}\big((X_{n,j})^{\underline{s}}\big) = s!\left(\frac{\binom{n-j}{s}\binom{s+\frac{c_2}{c_1}}{s}}{\binom{j-1+\frac{c_2}{c_1}+s}{s}} + \frac{\binom{n-j}{s-1}\binom{s-1+\frac{c_2}{c_1}}{s-1}}{\binom{j-1+\frac{c_2}{c_1}+s-1}{s-1}}\right). \quad (5.40)$$

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

$$\mathbb{E}(X_{n,j}) = \frac{(c_1+c_2)n - c_2(j-1)}{c_1 j + c_2}, \quad (5.41)$$

$$\mathbb{V}(X_{n,j}) = \frac{c_1(c_1+c_2)(c_1 n + c_2)(j-1)(n-j)}{(c_1 j + c_2)^2(c_1 j + c_1 + c_2)}. \quad (5.42)$$

Theorem 18. *The limiting distribution behavior of the random variable $X_{n,j}$, which counts the number of descendants of the node with label j in a randomly chosen size-n tree of a grown simple family of increasing trees as given by Lemma 1, is, for $n \to \infty$ and depending on the growth of j, characterized as follows.*

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

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

- *The region for small j: $j \to \infty$ such that $j = o(n)$. The normalized random variable $\frac{j}{n}X_{n,j}$ is asymptotically Gamma-distributed, $\frac{j}{n}X_{n,j} \xrightarrow{(d)} \gamma(\frac{c_2}{c_1}+1, 1)$, i. e. $\frac{j}{n}X_{n,j} \xrightarrow{(d)} X$, where the s-th moments of X are for $s \geq 0$ given by*

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

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

$X_{n,j} - 1 \xrightarrow{(d)} X$, where the probability mass function of X is given by

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

- *The region for large j: $j \to \infty$ such that $l := n - j = o(n)$. The random variable $X_{n,j}$ converges to a random variable, which has all its mass concentrated at 1, i. e. $X_{n,j} \xrightarrow{(d)} X$, with*

$$\mathbb{P}\{X = 1\} = 1.$$

From Theorem 5.3 and Corollary 6 we can easily compute explicit formulæ for the probabilities $\mathbb{P}\{X_{n,j} = m\}$ for grown simple increasing tree families, i. e. increasing tree families, which can be constructed via an insertion process. We will figure out only the Case C and omit the analogous computations for Case A and Case B.
Using Lemma 1 and equation (9) we get

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

and thus from equation (5.3):

$$N(z,u,v) = \frac{v\varphi_0(1-c_1 u)^{\frac{c_2}{c_1}+1}}{(1-c_1 uv)^{\frac{c_2}{c_1}+1}\big(1-c_1(z+u)\big)^{\frac{c_2}{c_1}+1}} = \frac{v\varphi_0}{(1-c_1 uv)^{\frac{c_2}{c_1}+1}\big(1-\frac{c_1 z}{1-c_1 u}\big)^{\frac{c_2}{c_1}+1}}. \tag{5.43}$$

Extracting coefficients from (5.43) gives then by using (5.3) and (10):

$$\begin{aligned}
\mathbb{P}\{X_{k+j,j} = m\} &= \frac{(j-1)!k!}{T_{k+j}}[z^{j-1}u^k v^m]N(z,u,v) \\
&= \frac{(j-1)!k!\varphi_0}{(k+j-1)!\varphi_0 c_1^{k+j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}}[z^{j-1}u^k v^{m-1}]\frac{1}{(1-c_1 uv)^{\frac{c_2}{c_1}+1}\big(1-\frac{c_1 z}{1-c_1 u}\big)^{\frac{c_2}{c_1}+1}} \\
&= \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}}{c_1^k\binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}}[u^k v^{m-1}]\frac{1}{(1-c_1 uv)^{\frac{c_2}{c_1}+1}(1-c_1 u)^{j-1}} \\
&= \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{m-1+\frac{c_2}{c_1}}{m-1}}{c_1^{k-m+1}\binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}}[u^k]\frac{u^{m-1}}{(1-c_1 u)^{j-1}} \\
&= \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{m-1+\frac{c_2}{c_1}}{m-1}\binom{k-m+j-1}{j-2}}{\binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}}.
\end{aligned} \tag{5.44}$$

It turns out that this formula (5.44) is indeed valid for all three cases of very simple families of increasing trees. Thus we obtain the first part of Theorem 17 after the substitution $n := k + j$.

5.4.2 An exact formula for the factorial moments

To obtain the s-th factorial moments of $D_{n,j}$ we use (6), we differentiate $N(z,u,v)$ s times w. r. t. v and evaluate it at $v = 1$. For Case C this gives

$$E_v D_v^s N(z,u,v) = \frac{\varphi_0 c_1^s u^s \big(\frac{c_2}{c_1}+1\big)^{\overline{s}}}{(1-c_1 u)^{\frac{c_2}{c_1}+s+1}\big(1-\frac{c_1 z}{1-c_1 u}\big)^{\frac{c_2}{c_1}+1}} + \frac{s\varphi_0 c_1^{s-1} u^{s-1}\big(\frac{c_2}{c_1}+1\big)^{\overline{s-1}}}{(1-c_1 u)^{\frac{c_2}{c_1}+s}\big(1-\frac{c_1 z}{1-c_1 u}\big)^{\frac{c_2}{c_1}+1}}. \tag{5.45}$$

Extracting coefficients of (5.45) leads then by using (10) to

$$\begin{aligned}
\mathbb{E}\big((X_{k+j,j})^{\underline{s}}\big) &= \sum_{m\geq 0} m^{\underline{s}} \mathbb{P}\{D_{k+j,j}=m\} = \frac{(j-1)!k!}{T_{k+j}}[z^{j-1}u^k]E_v D_v^s N(z,u,v) \\
&= \frac{1}{\varphi_0 c_1^{k+j-1}\binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}} \Bigg[\varphi_0 c_1^{s+j-1}\big(\frac{c_2}{c_1}+1\big)^{\overline{s}}\binom{j-1+\frac{c_2}{c_1}}{j-1}[u^k]\frac{u^s}{(1-c_1u)^{\frac{c_2}{c_1}+s+j}} \\
&\quad + s\varphi_0 c_1^{s+j-2}\big(\frac{c_2}{c_1}+1\big)^{\overline{s-1}}\binom{j-1+\frac{c_2}{c_1}}{j-1}[u^k]\frac{u^{s-1}}{(1-c_1u)^{\frac{c_2}{c_1}+s+j-1}}\Bigg] \\
&= \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}}{\binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}}\Bigg[\big(\frac{c_2}{c_1}+1\big)^{\overline{s}}\binom{k+j+\frac{c_2}{c_1}-1}{k-s} + s\big(\frac{c_2}{c_1}+1\big)^{\overline{s-1}}\binom{k+j+\frac{c_2}{c_1}-1}{k-s+1}\Bigg] \\
&= \frac{s!\binom{j-1+\frac{c_2}{c_1}}{j-1}}{\binom{k+j-1}{j-1}\binom{k+j-1+\frac{c_2}{c_1}}{k+j-1}}\Bigg[\binom{s+\frac{c_2}{c_1}}{s}\binom{k+j-1+\frac{c_2}{c_1}}{k-s} + \binom{s-1+\frac{c_2}{c_1}}{s-1}\binom{k+j-1+\frac{c_2}{c_1}}{k-s+1}\Bigg],
\end{aligned}$$

which can be slightly simplified and we get

$$\mathbb{E}\big((X_{k+j,j})^{\underline{s}}\big) = s!\left(\frac{\binom{k}{s}\binom{s+\frac{c_2}{c_1}}{s}}{\binom{j-1+\frac{c_2}{c_1}+s}{s}} + \frac{\binom{k}{s-1}\binom{s-1+\frac{c_2}{c_1}}{s-1}}{\binom{j-1+\frac{c_2}{c_1}+s-1}{s-1}}\right). \tag{5.46}$$

Since formula (5.46) is valid also for Case A and Case B, the second part of Theorem 17 follows after substituting $n := k+j$.

5.4.3 The case j fixed

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

$$\mathbb{E}(X^s) = \prod_{k=0}^{s-1}\frac{a+k}{a+b+k} = \frac{a^{\overline{s}}}{(a+b)^{\overline{s}}}. \tag{5.47}$$

Using Stirling's formula for the Gamma function

$$\Gamma(z) = \Big(\frac{z}{e}\Big)^z\frac{\sqrt{2\pi}}{\sqrt{z}}\Big(1+\frac{1}{12z}+\frac{1}{288z^2}+\mathcal{O}(\frac{1}{z^3})\Big), \tag{5.48}$$

we obtain for j and s fixed:

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

Thus we get from equation (5.40) the following asymptotic expansion of the s-th factorial moment of $X_{n,j}$:

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

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

$$\begin{aligned} \mathbb{E}\big((X_{n,j})^s\big) &= \mathbb{E}\big((X_{n,j})^{\underline{s}}\big) + \sum_{k=1}^{s-1} \left\{ {s \atop k} \right\} \mathbb{E}\big((X_{n,j})^{\underline{k}}\big) \\ &= \frac{\binom{s+\frac{c_2}{c_1}}{s}}{\binom{j-1+\frac{c_2}{c_1}+s}{s}} n^s\big(1+\mathcal{O}(n^{-1})\big) + \mathcal{O}(n^{s-1}) = \frac{\binom{s+\frac{c_2}{c_1}}{s}}{\binom{j-1+\frac{c_2}{c_1}+s}{s}} n^s\big(1+\mathcal{O}(n^{-1})\big). \end{aligned} \tag{5.49}$$

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

$$\mathbb{E}\Big(\big(\frac{X_{n,j}}{n}\big)^s\Big) \to \frac{\binom{s+\frac{c_2}{c_1}}{s}}{\binom{j-1+\frac{c_2}{c_1}+s}{s}} = \frac{\big(\frac{c_2}{c_1}+1\big)^{\overline{s}}}{\big(\frac{c_2}{c_1}+j\big)^{\overline{s}}}, \tag{5.50}$$

which shows together with the Theorem of Fréchet and Shohat (see e. g. [49]) the first part of Theorem 18.

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

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

$$\mathbb{E}(X^s) = \frac{1}{\lambda^s}\prod_{k=0}^{s-1}(a+k) = \frac{a^{\overline{s}}}{\lambda^s}. \tag{5.51}$$

Again by using Stirling's formula (5.48) for the Gamma function we obtain for s fixed:

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

and thus from equation (5.40) the following expansion of the s-th factorial moments of $X_{n,j}$:

$$\mathbb{E}\big((X_{n,j})^{\underline{s}}\big) = s!\binom{s+\frac{c_2}{c_1}}{s}\big(\frac{n}{j}\big)^s\Big(1+\mathcal{O}\big(\frac{1}{j}\big)+\mathcal{O}\big(\frac{j}{n}\big)\Big). \tag{5.52}$$

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

$$\mathbb{E}\big((X_{n,j})^{s}\big) = s!\binom{s+\frac{c_2}{c_1}}{s}\big(\frac{n}{j}\big)^s\Big(1+\mathcal{O}\big(\frac{1}{j}\big)+\mathcal{O}\big(\frac{j}{n}\big)\Big). \tag{5.53}$$

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

$$\mathbb{E}\Big(\big(\frac{j}{n}X_{n,j}\big)^s\Big) \to s!\binom{s+\frac{c_2}{c_1}}{s} = \big(\frac{c_2}{c_1}+1\big)^{\overline{s}}. \tag{5.54}$$

This proves the second part of Theorem 18.

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

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

$$\mathbb{P}\{X = m\} = \binom{m+r-1}{m} p^r (1-p)^m, \quad \text{for } m \geq 0. \tag{5.55}$$

We start with the following form of $\mathbb{P}\{X_{n,j} = m\}$ equivalent to (5.39):

$$\mathbb{P}\{X_{n,j} = m\} = \frac{(j-1)\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{n-j}{m-1}\binom{m-1+\frac{c_2}{c_1}}{m-1}}{m\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-1}{m}}, \tag{5.56}$$

and apply Stirling's formula (5.48). This leads to

$$\mathbb{P}\{X_{n,j} = m\} = \binom{m-1+\frac{c_2}{c_1}}{m-1}\Big(\frac{j}{n}\Big)^{\frac{c_2}{c_1}+1}\Big(1-\frac{j}{n}\Big)^{m-1}\Big(1+\mathcal{O}\big(\frac{1}{n}\big)+\mathcal{O}\big(\frac{1}{j}\big)+\mathcal{O}\big(\frac{1}{n-j}\big)\Big). \tag{5.57}$$

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

$$\mathbb{P}\{X_{n,j} - 1 = m\} \to \binom{m+\frac{c_2}{c_1}}{m}\rho^{\frac{c_2}{c_1}+1}(1-\rho)^m. \tag{5.58}$$

Thus the third part ot Theorem 18 follows.

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

Substituting $l := n - j$, the probabilities $\mathbb{P}\{X_{n,j} = m\}$ given by (5.56) can be written as follows:

$$\mathbb{P}\{X_{n,j} = m\} = \frac{\binom{n-l-1+\frac{c_2}{c_1}}{n-l-1}}{\binom{n-1+\frac{c_2}{c_1}}{n-1}} \frac{\binom{l+1}{m}}{\binom{n-1}{m}} \frac{n-l-1}{l+1}\binom{m-1+\frac{c_2}{c_1}}{m-1}. \tag{5.59}$$

In the sequel we want to obtain a suitable bound for the probabilities $\mathbb{P}\{X_{n,j} = m\}$, which holds uniformly for all $m \geq 2$. Since we are only interested in the case $l := n - j = o(n)$ we make in the following computations the assumptions $l \leq \frac{n}{3}$ and $n \geq 3$.
First we consider for $2 \leq m \leq l+1$:

$$\frac{\binom{l+1}{m}}{\binom{n-1}{m}} = \frac{(l+1)l}{(n-1)(n-2)}\prod_{k=1}^{m-2}\frac{l-k}{n-2-k}. \tag{5.60}$$

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

$$\frac{(l+1)l}{(n-1)(n-2)} \leq \frac{9\,l^2}{n^2}, \quad \text{and} \quad \frac{l-k}{n-2-k} \leq \frac{l}{n}, \text{ for } 1 \leq k \leq l. \tag{5.61}$$

Combining (5.60) and (5.61) leads to the estimate

$$\frac{\binom{l+1}{m}}{\binom{n-1}{m}} \leq 9\Big(\frac{l}{n}\Big)^m, \tag{5.62}$$

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

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

$$\begin{aligned} \frac{\binom{n-l-1+\frac{c_2}{c_1}}{n-l-1}}{\binom{n-1+\frac{c_2}{c_1}}{n-1}} &= \frac{(n-1)(n-2)\cdots(n-l)}{(n-1+\frac{c_2}{c_1})(n-2+\frac{c_2}{c_1})\cdots(n-l+\frac{c_2}{c_1})} \le \frac{(n-1)(n-2)\cdots(n-l)}{(n-2)(n-3)\cdots(n-l-1)} \\ &= \frac{n-1}{n-l-1} \le 2. \end{aligned} \tag{5.63}$$

Analogously we compute:

$$\binom{m-1+\frac{c_2}{c_1}}{m-1} = \big(m-1+\frac{c_2}{c_1}\big)\frac{(m-2+\frac{c_2}{c_1})(m-3+\frac{c_2}{c_1})\cdots(1+\frac{c_2}{c_1})}{(m-1)(m-2)\cdots 2} \le m-1+\frac{c_2}{c_1} \le m. \tag{5.64}$$

Together with the trivial bound

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

we finally get from (5.59) by using (5.62), (5.63) and (5.64) the following estimate, which holds uniformly for all $m \ge 2$:

$$\mathbb{P}\{X_{n,j} = m\} \le 18m\big(\frac{l}{n}\big)^{m-1}. \tag{5.65}$$

Equation (5.65) leads then (for $l \le \frac{n}{3}$) to the bound

$$\sum_{m\ge 2} \mathbb{P}\{X_{n,j} = m\} \le 18 \sum_{m\ge 2} m\big(\frac{l}{n}\big)^{m-1} = \frac{18l}{n}\frac{2-\frac{l}{n}}{(1-\frac{l}{n})^2} \le \frac{36\frac{l}{n}}{(1-\frac{l}{n})^2} \le \frac{81\,l}{n}. \tag{5.66}$$

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

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

Thus also the last part of Theorem 18 is shown.

5.4.7 Auxiliary results concerning subtree sizes

Let $\pi_{n;m_1,\dots,m_j}$ denote the probability that in a random plane oriented recursive tree the first subtree has size m_1, the second m_2, and so on. Let $M_j = \sum_{l=1}^{j} m_l$. Hence by definition it holds for generalized plane oriented recursive trees

$$\begin{aligned} T_n \pi_{n;m_1,\dots,m_j} = \sum_{r\ge j+1} \varphi_r \sum_{\substack{k_{j+1}+\cdots+k_r = n-M_j-1,\\ k_{j+1},\dots,k_r \ge 1}} T_{m_1}\dots T_{m_j} T_{k_{j+1}}\cdots T_{k_r} \binom{n-1}{m_1,\dots,m_j,k_{j+1},\dots,k_r} \\ + [\![n = m_1+\cdots+m_j+1]\!] \varphi_j T_{m_1}\dots T_{m_j}\binom{n-1}{m_1,\dots,m_j}. \end{aligned} \tag{5.67}$$

By defining

$$A_j(z, v_1, \dots, v_j) = \sum_{n\ge j+1} \sum_{m_1,\dots,m_j\ge 1} T_n \pi_{n;m_1,\dots,m_j} \frac{z^n}{n!} v_1^{m_1}\dots v_j^{m_j}, \tag{5.68}$$

we get

$$\frac{\partial}{\partial z} A_j(z, v_1, \dots, v_j) = \frac{\varphi\big(T(z)\big) - \sum_{k=0}^{j-1}\varphi_j T(z)^k}{T(z)^j} \prod_{l=1}^{j} T(zv_l). \tag{5.69}$$

For PORTs one has $\varphi(t) = 1/(1-t)$, $\varphi(T(z)) = 1/(1-T(z)) = T'(z)$ and consequently

$$\frac{\partial}{\partial z} A_j(z, v_1, \ldots, v_j) = \frac{1}{\sqrt{1-2z}} \prod_{l=1}^{j} (1 - \sqrt{1-2zv_l}). \tag{5.70}$$

Thus for $j = 1$ one obtains $\pi_{n,m} = \mathbb{P}\{J_n = m\}$ be extracting coefficients from $\frac{\partial}{\partial z} A_j(z, v)$.

$$\mathbb{P}\{J_n = m\} = \frac{(n-1)!}{T_n} [z^{n-1} v^m] \frac{\partial}{\partial z} A_j(z, v) = \frac{T_m T_{n-m} (n-1)!}{T_n m! (n-m-1)!}, \tag{5.71}$$

where $T_n = 2^{n-1}(n-1)!\binom{n-\frac{3}{2}}{n-1} = (2n-3)!!$.

Remark 10. A simple application of Stirling's formula shows the probabilities $\mathbb{P}\{J_n = m\}$ converge without convergence of any integral moment. This was already pointed out by Hwang. A related phenomena occurs for the probabilities $\mathbb{P}\{X_n^{[R]} = m\}$ for all grown simple families of increasing trees. Here $X_n^{[R]}$ denotes the random variable counting the subtree size of a randomly chosen node in a grown simple increasing tree of size n.

Theorem 19. *The probabilities $\mathbb{P}\{X_n^{[R]} = m\}$ are for all grown simple families of increasing trees given by the following explicit formula.*

$$\mathbb{P}\{X_n^{[R]} = m\} = \frac{[\![m = n]\!]}{n} + [\![m < n]\!] \frac{\left(\frac{c_2}{c_1} + 1\right)\left(\frac{c_2}{c_1} + n\right)}{n(m + 1 + \frac{c_2}{c_1})(m + \frac{c_2}{c_1})}. \tag{5.72}$$

The limit distribution of $X_\infty^{[R]} = X^{[R]}$ is given by

$$\mathbb{P}\{X^{[R]} = m\} = \frac{\frac{c_2}{c_1} + 1}{(m + 1 + \frac{c_2}{c_1})(m + \frac{c_2}{c_1})}, \tag{5.73}$$

where $X_n^{[R]}$ converges without convergence of any integral moment.

By using the explicit formula for the probabilities $\mathbb{P}\{X_{n,j} = m\}$ as given in [46]

$$\mathbb{P}\{X_{n,j} = m\} = \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{m-1+\frac{c_2}{c_1}}{m-1}\binom{n-m-1}{j-2}}{\binom{n-1}{j-1}\binom{n-1+\frac{c_2}{c_1}}{n-1}}, \tag{5.74}$$

we obtain the probabilities $\mathbb{P}\{X_n^{[R]} = m\}$ by summation.

$$\mathbb{P}\{X_n^{[R]} = m\} = \frac{[\![m = n]\!]}{n} + \frac{1}{n} \sum_{j=2}^{n} \mathbb{P}\{X_{n,j} = m\} = \frac{[\![m = n]\!]}{n} + \frac{1}{n} \sum_{j=2}^{n-m+1} \mathbb{P}\{X_{n,j} = m\}. \tag{5.75}$$

We get further

$$\begin{aligned}
\sum_{j=2}^{n-m+1} \mathbb{P}\{X_{n,j} = m\} &= \frac{\binom{m-1+\frac{c_2}{c_1}}{m-1}}{\binom{n-1+\frac{c_2}{c_1}}{n-1}} \sum_{j=2}^{n-m+1} \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{n-m-1}{j-2}}{\binom{n-1}{j-1}} = \frac{\binom{m-1+\frac{c_2}{c_1}}{m-1}}{\binom{n-1+\frac{c_2}{c_1}}{n-1}} \sum_{j=0}^{n-m-1} \frac{\binom{j+1+\frac{c_2}{c_1}}{j+1}\binom{n-m-1}{j}}{\binom{n-1}{j+1}} \\
&= \frac{\binom{m-1+\frac{c_2}{c_1}}{m-1}}{\binom{n-1+\frac{c_2}{c_1}}{n-1}} \sum_{j=0}^{n-m-1} \frac{(j+1)}{(n-1)} \frac{\binom{j+1+\frac{c_2}{c_1}}{j+1}\binom{n-2-j}{m-1}}{\binom{n-2}{m-1}} \\
&= \frac{\binom{m-1+\frac{c_2}{c_1}}{m-1}}{(n-1)\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-2}{m-1}} \sum_{j=0}^{n-m-1} (j+1) \binom{j+1+\frac{c_2}{c_1}}{j+1} \binom{n-2-j}{m-1}.
\end{aligned} \tag{5.76}$$

The arising sum can be simplified as follows

$$\begin{aligned}\sum_{j=0}^{n-m-1}(j+1)\binom{n-2-j}{m-1}\binom{j+1+\frac{c_2}{c_1}}{j+1} &= \big(\frac{c_2}{c_1}+1\big)\sum_{j=0}^{n-m-1}(-1)^j\binom{n-2-j}{m-1}\binom{-\frac{c_2}{c_1}-2}{j}\\ = \big(\frac{c_2}{c_1}+1\big)(-1)^{m+m-3}\binom{-\frac{c_2}{c_1}-2-m}{n-m-1} &= \big(\frac{c_2}{c_1}+1\big)(-1)^{m+m-3}(-1)^{n-m-1}\binom{n+\frac{c_2}{c_1}}{n-m-1}\\ &= \big(\frac{c_2}{c_1}+1\big)\binom{n+\frac{c_2}{c_1}}{n-m-1}. \end{aligned} \tag{5.77}$$

Thus we finally get

$$\mathbb{P}\{X_n^{[R]}=m\} = \frac{[\![m=n]\!]}{n} + \frac{\binom{m-1+\frac{c_2}{c_1}}{m-1}\big(\frac{c_2}{c_1}+1\big)\binom{n+\frac{c_2}{c_1}}{n-m-1}}{n(n-1)\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-2}{m-1}} = \frac{[\![m=n]\!]}{n} + [\![m<n]\!]\frac{\big(\frac{c_2}{c_1}+1\big)\big(\frac{c_2}{c_1}+n\big)}{n(m+1+\frac{c_2}{c_1})(m+\frac{c_2}{c_1})}. \tag{5.78}$$

5.5 Degree of node j

Let $X_{n,j}$ count the degree of node j in a random increasing tree of size n. From the work of Bergeron, Flajolet and Salvy [6] we get

$$M(z,v) = \int_0^z \varphi\big(vT(t)\big)dt, \tag{5.79}$$

and thus

$$\frac{\partial}{\partial u} M(u,v) = \varphi\big(vT(u)\big). \tag{5.80}$$

Hence Theorem 16 gives

$$N(z,u,v) = \frac{\varphi\big(vT(u)\big)\varphi\big(T(z+u)\big)}{\varphi\big(T(u)\big)}. \tag{5.81}$$

Using the characterization of the degree-weight generating function in Lemma 1 we easily get

$$\varphi(vT(u)) = \begin{cases} \dfrac{\varphi_0}{(1-c_1u)^v} = \varphi_0 \sum_{k\ge 0} \dfrac{v^k}{k!}\log^k\big(\dfrac{1}{1-c_1u}\big), & \text{Case A},\\ \varphi_0\Big(1+v\big(\dfrac{1}{(1-(d-1)c_2u)^{\frac{1}{d-1}}}-1\big)\Big)^d, & \text{Case B},\\ \dfrac{\varphi_0}{\Big(1+v\big(\frac{1}{(1-c_1u)^{\frac{c_2}{c_1}}}-1\big)^{-\frac{c_1}{c_2}-1}\Big)}, & \text{Case C}. \end{cases} \tag{5.82}$$

5.5.1 Results for the degree of node j

Theorem 20. *The probabilities $\mathbb{P}\{X_{n,j}=m\}$, which give the probability that the node with label j in a randomly chosen size-n tree of a grown simple family of increasing trees as given by Lemma 1, has outdegree m, are, for $m \ge 1$, given by the following formula:*

$$\mathbb{P}\{X_{n,j}=m\} = \begin{cases} \dfrac{1}{\binom{n-1}{j-1}} \sum_{k=m}^{n-j} \dbinom{n-k-2}{j-2} \dfrac{\left[{k \atop m}\right]}{k!}, & \textit{Case A},\\ \dbinom{d}{m} \sum_{k=0}^{m} \dbinom{m}{k} (-1)^{m-k} \dfrac{\Gamma(n-1+k\frac{1}{d-1})\Gamma(j+\frac{1}{d-1})}{\Gamma(j-1+k\frac{1}{d-1})\Gamma(n+\frac{1}{d-1})}, & \textit{Case B},\\ \dbinom{m-2-\frac{c_1}{c_2}}{m} \sum_{k=0}^{m} \dbinom{m}{k} (-1)^k \dfrac{\Gamma(n-1+k\frac{c_2}{c_1})\Gamma(j+\frac{c_2}{c_1})}{\Gamma(j-1+k\frac{c_2}{c_1})\Gamma(n+\frac{c_2}{c_1})}, & \textit{Case C}. \end{cases} \tag{5.83}$$

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

$$\mathbb{E}(X_{n,j}^{\underline{s}}) = \begin{cases} \dfrac{s!}{\binom{n-1}{j-1}} \sum_{l=0}^{n-j} \dbinom{n-l-1}{j-1} \dfrac{\left[{l \atop s}\right]}{l!}, & \textit{Case A},\\ \dfrac{d^{\underline{s}}\binom{\frac{1}{d-1}+j-1}{j-1}(n+\frac{s}{d-1}-2)^{\underline{n-j}}}{\binom{n-1}{j-1}\binom{n-1+\frac{1}{d-1}}{n-1}(n-j)!}, & \textit{Case B},\\ \dfrac{\Gamma(s-1-\frac{c_1}{c_2})}{\Gamma(-1-\frac{c_1}{c_2})} \sum_{k=0}^{s} \dbinom{s}{k}(-1)^k \dfrac{\Gamma(n-\frac{c_2}{c_1}(s-1-k))\Gamma(j+\frac{c_2}{c_1})}{\Gamma(j-\frac{c_2}{c_1}(s-1-k))\Gamma(n+\frac{c_2}{c_1})}, & \textit{Case C}. \end{cases} \tag{5.84}$$

Theorem 21. *The limiting distribution behavior of the random variable $X_{n,j}$, which counts the outdegree of the node with label j in a randomly chosen size-n tree of a grown simple family of increasing trees of Case A as given by Lemma 1, is, for $n \to \infty$ and depending on the growth of j:*

- *The region $j = o(n)$. The centralized and normalized random variable $X_{n,j}^*$ is asymptotically gaussian distributed,*
$$X_{n,j}^* = \frac{X_{n,j} - (\log n - \log j)}{\sqrt{\log n - \log j}} \to \mathcal{N}(0,1). \tag{5.85}$$

- *The region j: $j \to \infty$ such that $j = \rho n$. The random variable $X_{n,j}$ is asymptotically poisson distributed with parameter $\lambda = -\log \rho$.*
$$X_{n,j} \xrightarrow{(d)} X_j, \quad \mathbb{P}\{X_j = m\} = \frac{\rho(-\log \rho)^m}{m!}. \tag{5.86}$$

- *The region j: $l := n - j = o(n)$. $\mathbb{P}\{X_{n,j} = 0\} \to 1$.*

Theorem 22. *The limiting distribution behavior of the random variable $X_{n,j}$ in a randomly chosen size-n tree of a grown simple family of increasing trees of Case B as given by Lemma 1, is, for $n \to \infty$ and depending on the growth of j*

- *The regions for j fixed and small j with $j \to \infty$ such that $j = o(n)$: $\mathbb{P}\{X_{n,j} = d\} \to 1$.*
- *The region j: $j \to \infty$ such that $j = \rho n$.*
$$X_{n,j} \xrightarrow{(d)} X_j, \quad \mathbb{P}\{X_j = m\} = \binom{d}{m} \rho^{1+\frac{1}{d-1}} \left(\rho^{-\frac{1}{d-1}} - 1\right)^m. \tag{5.87}$$

- *The region j with $l := n - j = o(n)$: $\mathbb{P}\{X_{n,j} = 0\} \to 1$.*

Theorem 23. *The limiting distribution behavior of the random variable $X_{n,j}$ in a randomly chosen size-n tree of a grown simple family of increasing trees of Case C as given by Lemma 1, is, for $n \to \infty$ and depending on the growth of j*

- *The region for j fixed. For general c_1 and c_2 we have*
$$n^{\frac{c_2}{c_1}} X_{n,j} \xrightarrow{(d)} X_j, \quad \mathbb{E}(X_j^s) = \frac{\Gamma(s - 1 - \frac{c_1}{c_2})\Gamma(j + \frac{c_2}{c_1})}{\Gamma(-1 - \frac{c_1}{c_2})\Gamma(j - \frac{(s-1)c_2}{c1})}. \tag{5.88}$$

  The density $f_j(x)$ of X_j is given by
$$f_j(x) = \frac{\Gamma(j + \frac{c_2}{c_1})}{\Gamma(1 + \frac{c_2}{c_1})\Gamma(j-1)} \int_0^1 t^{2\frac{c_2}{c_1}} (1-t)^{j-2} f_1\big(x t^{\frac{c_2}{c_1}}\big) \, dt \, dx, \tag{5.89}$$

  where
$$f_1(x) = \frac{-\Gamma(\frac{c_2}{c_1}) x^{-2-\frac{c_1}{c_2}}}{\Gamma(-1 - \frac{c_1}{c_2})\pi} \int_0^\infty e^{-(r^{-\frac{c_1}{c_2}})} r^{-\frac{c_1}{c_2}-1} e^{-xr\cos(\frac{c_2}{c_1}\pi)} \sin\big(xr\sin(-\frac{c_2}{c_1}\pi)\big) \, dr. \tag{5.90}$$

  For plane oriented recursive trees ($c_1 = 2$, $c_2 = -1$) we get by specialization the following result. For $j = 1$, $X_{n,1}$ is asymptotically Rayleigh distributed with parameter $\sigma = \sqrt{2}$.
$$n^{\frac{c_2}{c_1}} X_{n,1} \xrightarrow{(d)} X_1, \quad f_{X_1}(x) = \frac{x}{2} e^{-\frac{x^2}{2}}. \tag{5.91}$$

  For $j > 1$ we get the following
$$n^{-\frac{1}{2}} X_{n,j} \xrightarrow{(d)} X_j, \quad f_{X_j}(x) = \frac{2j-3}{2^{2j-3}(j-2)!} \int_x^\infty (t-x)^{2j-4} e^{-\frac{t^2}{4}} \, dt. \tag{5.92}$$

For $j \geq 1$ the moments of X_j are given by

$$\mathbb{E}(X_j^s) = \frac{(2j-2)!2^s s!\Gamma(j+\frac{s}{2})}{(j-1)!(s+2j-2)!}. \tag{5.93}$$

- *The region small j: $j \to \infty$ such that $j = o(n)$. $n^{\frac{c_2}{c_1}} Xn, j$ is asymptotically Gamma-distributed $X_j \xrightarrow{(d)} \gamma(a, \lambda)$, with parameter $a = -1 - \frac{c_1}{c_2}$ and $\lambda = 1$, where the moments of X_j are given by*

$$\mathbb{E}(X_j^s) = \frac{a^{\overline{s}}}{\lambda^s} = (-\frac{c_1}{c_2} - 1)^{\overline{s}} = \frac{\Gamma(s-1-\frac{c_1}{c_2})}{\Gamma(-1-\frac{c_1}{c_2})}. \tag{5.94}$$

- *The region j: $j \to \infty$ such that $j = \rho n$. We have a negative binomial distribution* $\mathrm{NegBin}(r, p)$ *with parameters $r = -1 - \frac{c_1}{c_2}$ and $p = \rho^{-\frac{c_2}{c_1}}$:*

$$X_{n,j} \xrightarrow{(d)} X_j, \quad \mathbb{P}\{X_j = m\} = \binom{m - 2 - \frac{c_1}{c_2}}{m} \rho^{1+\frac{c_2}{c_1}} \big(1 - \rho^{-\frac{c_2}{c_1}}\big)^m. \tag{5.95}$$

- *The region j: $l := n - j = o(n)$. $\mathbb{P}\{X_{n,j} = 0\} \to 1$.*

5.5.2 An exact formula for the probabilities

From Proposition 5.81 we can easily compute explicit formulæ for the probabilities $\mathbb{P}\{X_{n,j} = m\}$ for grown simple families of increasing trees, i. e. increasing tree families, which can be constructed via an insertion process. We will figure out Case A and Case B, Case C follows from an analogous computation. We observe that

$$N(z,u,v) = \begin{cases} \dfrac{\varphi_0}{(1-c_1u)^v(1-\frac{c_1 z}{1-c_1 u})}, & \text{Case A,} \\[2ex] \dfrac{\varphi_0\Big(1 + v\big(\frac{1}{(1-(d-1)c_2u)^{\frac{1}{d-1}}} - 1\big)\Big)^d}{\Big(1 - \frac{(d-1)c_2 z}{1-(d-1)c_2 u}\Big)^{\frac{d}{d-1}}}, & \text{Case B,} \\[2ex] \dfrac{\varphi_0}{\Big(1 + v\big(\frac{1}{(1-c_1u)^{\frac{c_2}{c_1}}} - 1\big)\Big)^{-\frac{c_1}{c_2}-1}\Big(1 - \frac{c_1 z}{1-c_1 u}\Big)^{\frac{c_2}{c_1}+1}}, & \text{Case C.} \end{cases} \tag{5.96}$$

For Case A we obtain for the probability

$$\begin{aligned} \mathbb{P}\{X_{n,j} = m\} &= \frac{1}{c_1^{n-1}\binom{n-1}{j-1}} [u^{n-j}v^m] \frac{c_1^{j-1}}{(1-c_1u)^{v+j-1}} \\ &= \frac{1}{\binom{n-1}{j-1}} \sum_{k=0}^{n-j} [u^{n-j-k}] \frac{1}{(1-c_1u)^{j-1}} [u^k v^m] \frac{c_1^{j-1}}{(1-c_1u)^{v+j-1}}, \end{aligned} \tag{5.97}$$

which leads with the generating function identity for the first order Stirling numbers

$$\sum_{n \geq 0} \sum_{m=0}^{m} \begin{bmatrix} n \\ m \end{bmatrix} \frac{z^n}{n!} v^m = \frac{1}{(1-z)^v} \tag{5.98}$$

to the desired result. For Case B we proceed as follows.

$$
\begin{aligned}
\mathbb{P}\{X_{n,j}=m\} &= \frac{\binom{\frac{1}{d-1}+j-1}{j-1}}{c_1^{n-j}\binom{n-1}{j-1}\binom{n-1+\frac{1}{d-1}}{n-1}}[u^{n-j}v^m]\frac{\Big(1+v\big(\frac{1}{(1-(d-1)c_2u)^{\frac{1}{d-1}}}-1\big)\Big)^d}{(1-(d-1)c_2u)^{j-1}} \\
&= \frac{\binom{\frac{1}{d-1}+j-1}{j-1}\binom{d}{m}}{c_1^{n-j}\binom{n-1}{j-1}\binom{n-1+\frac{1}{d-1}}{n-1}}[u^{n-j}]\sum_{k=0}^{m}\binom{m}{k}\frac{(-1)^{m-k}}{(1-(d-1)c_2u)^{j-1+\frac{k}{d-1}}} \\
&= \frac{\binom{\frac{1}{d-1}+j-1}{j-1}\binom{d}{m}}{\binom{n-1}{j-1}\binom{n-1+\frac{1}{d-1}}{n-1}}\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}\binom{n-2+\frac{k}{d-1}}{n-j} \\
&= \binom{d}{m}\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}\frac{\Gamma(n-1+k\frac{1}{d-1})\Gamma(j+\frac{1}{d-1})}{\Gamma(j-1+k\frac{1}{d-1})\Gamma(n+\frac{1}{d-1})}.
\end{aligned}
\tag{5.99}
$$

An analogous computation to Case B leads to Case C.

5.5.3 An exact formula for the factorial moments

We only present the calculations for Case C, the other cases are completely analogous. To obtain the s-th factorial moments of $X_{n,j}$ we use (5.81), but differentiate s times w. r. t. v and evaluate it at $v=1$. For Case C this gives

$$
E_vD_v^sN(z,u,v)=\frac{\varphi_0(-\frac{c_1}{c_2}-1)^{\overline{s}}(1-(1-c_1u)^{-\frac{c_2}{c_1}})^s}{(1-c_1u)^{1-\frac{c_2}{c_1}(s-1)}\Big(1-\frac{c_1z}{1-c_1u}\Big)^{\frac{c_2}{c_1}+1}}.
\tag{5.100}
$$

Extracting coefficients of (5.100) leads then by using (10) to

$$
\begin{aligned}
\mathbb{E}\big((X_{n,j})^{\underline{s}}\big) &= \frac{(j-1)!(n-j)!}{T_n}[z^{j-1}u^{n-j}]E_vD_v^sN(z,u,v) \\
&= \frac{(-\frac{c_1}{c_2}-1)^{\overline{s}}\varphi_0c_1^{j-1}\binom{j-1+\frac{c_2}{c_1}}{j-1}}{\varphi_0c_1^{n-1}\binom{n-1}{j-1}\binom{n-1+\frac{c_2}{c_1}}{n-1}}[u^{n-j}]\frac{(1-(1-c_1u)^{-\frac{c_2}{c_1}})^s}{(1-c_1u)^{j-\frac{c_2}{c_1}(s-1)}} \\
&= \frac{\Gamma(s-1-\frac{c_1}{c_2})\binom{j-1+\frac{c_2}{c_1}}{j-1}}{\Gamma(-1-\frac{c_1}{c_2})c_1^{n-j}\binom{n-1}{j-1}\binom{n-1+\frac{c_2}{c_1}}{n-1}}\sum_{k=0}^{s}\binom{s}{k}(-1)^k\frac{1}{(1-c_1u)^{j-\frac{c_2}{c_1}(s-k-1)}} \\
&= \frac{\Gamma(s-1-\frac{c_1}{c_2})}{\Gamma(-1-\frac{c_1}{c_2})}\sum_{k=0}^{s}\binom{s}{k}(-1)^k\frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{n-1-\frac{c_2}{c_1}(s-k-1)}{n-j}}{\binom{n-1}{j-1}\binom{n-1+\frac{c_2}{c_1}}{n-1}} \\
&= \frac{\Gamma(s-1-\frac{c_1}{c_2})}{\Gamma(-1-\frac{c_1}{c_2})}\sum_{k=0}^{s}\binom{s}{k}(-1)^k\frac{\Gamma(n-\frac{c_2}{c_1}(s-1-k))\Gamma(j+\frac{c_2}{c_1})}{\Gamma(j-\frac{c_2}{c_1}(s-1-k))\Gamma(n+\frac{c_2}{c_1})}.
\end{aligned}
$$

5.5.4 Proofs of the limit distribution results

5.5.5 Case A

First we turn our attention to the region $j = o(n)$. For the expectation and the second factorial moment we get by using

$$[z^n]\frac{\log\left(\frac{1}{1-z}\right)}{(1-z)^{j+1}} = \binom{n+j}{j}(H_{n+j} - H_j), \quad [z^n]\frac{\log^2\left(\frac{1}{1-z}\right)}{(1-z)^{j+1}} = \binom{n+j}{j}\big((H_{n+j} - H_j)^2 - (H^{(2)}_{n+j} - H^{(2)}_j)\big), \tag{5.101}$$

the following result

$$\begin{aligned}
\mathbb{E}(X_{n,j}) &= \frac{1}{\varphi_0 c_1^{n-j}\binom{n-1}{j-1}}[u^{n-j}]\frac{M_1'(u)}{(1-c_1u)^{j-1}} = \frac{[u^{n-j}]}{c_1^{n-j}\binom{n-1}{j-1}}\frac{\log\left(\frac{1}{1-c_1u}\right)}{(1-c_1u)^j} \\
&= H_{n-1} - H_{j-1} = \log n - \log j + \mathcal{O}(1), \\
\mathbb{E}(X^2_{n,j}) &= \frac{1}{\varphi_0 c_1^{n-j}\binom{n-1}{j-1}}[u^{n-j}]\frac{M_2'(u)}{(1-c_1u)^{j-1}} = \frac{[u^{n-j}]}{c_1^{n-j}\binom{n-1}{j-1}}\frac{\log^2\left(\frac{1}{1-c_1u}\right)}{(1-c_1u)^j} \\
&= (H_{n-1} - H_{j-1})^2 - (H^{(2)}_{n-1} - H^{(2)}_{j-1}),
\end{aligned} \tag{5.102}$$

and thus for the variance

$$\mathbb{V}(X_{n,j}) = \mathbb{E}(X^2_{n,j}) + \mathbb{E}(X_{n,j}) - \mathbb{E}(X_{n,j})^2 = H_{n-1} - H_{j-1} - (H^{(2)}_{n-1} - H^{(2)}_{j-1}) = \log n - \log j + \mathcal{O}(1). \tag{5.103}$$

We will use the abbreviations $\mu_{n,j} = \sigma^2_{n,j} = \log n - \log j$. Since we want to apply Lévys continuity theorem to the moment generating function of $X^*_{n,j}$ we calculate first the probability generating function $p_{n,j}(v)$ of $X_{n,j}$. We get

$$\begin{aligned}
p_{n,j}(v) = \mathbb{E}(v^{X_{n,j}}) &= \sum_{m\ge 0}\mathbb{P}\{X_{n,j} = m\}v^m = \frac{1}{c_1^{n-j}\binom{n-1}{j-1}}[u^{n-j}]\frac{1}{(1-c_1u)^{v+j-1}} = \frac{\binom{n+v-2}{v+j-2}}{\binom{n-1}{j-1}} = \frac{\binom{n+v-2}{n-1}}{\binom{j+v-2}{j-1}} \\
&= \frac{\Gamma(n+v-1)\Gamma(j)}{\Gamma(n)\Gamma(j+v-1)}.
\end{aligned} \tag{5.104}$$

The moment generating function $\mathcal{M}_{n,j}(t)$ of $X^*_{n,j} = (X_{n,j} - \mu_{n,j})/\sigma_{n,j}$ is then given by

$$\mathcal{M}_{n,j}(t) = \mathbb{E}(e^{tX^*_{n,j}}) = e^{-\frac{\mu_{n,j}}{\sigma_{n,j}}t}\mathbb{E}(e^{\frac{X_{n,j}}{\sigma_{n,j}}t}) = e^{-\sigma_{n,j}t}p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}). \tag{5.105}$$

Using Stirling's formula for the Gamma function

$$\Gamma(z) = \left(\frac{z}{e}\right)^z\frac{\sqrt{2\pi}}{\sqrt{z}}\left(1 + \frac{1}{12z} + \frac{1}{288z^2} + \mathcal{O}(\frac{1}{z^3})\right) \tag{5.106}$$

we get

$$\frac{\Gamma(n+v-1)}{\Gamma(n)} = n^{v-1}\big(1 + \mathcal{O}(\frac{1}{n})\big) = e^{(v-1)\log n}\big(1 + \mathcal{O}(\frac{1}{n})\big). \tag{5.107}$$

We begin considering the region $j = o(n)$, such that $j \le \log n$. It holds for fixed t

$$\begin{aligned}
e^{-\sigma t} &= e^{-\sqrt{\log n}t\left(1+\mathcal{O}(\frac{\log j}{\log n})\right)} = e^{-\sqrt{\log n}t\left(1+\mathcal{O}(\frac{\log\log n}{\log n})\right)} = e^{-\sqrt{\log n}t}\Big(1+\mathcal{O}(\frac{\log\log n}{\sqrt{\log n}})\Big),\\
\frac{1}{\sigma_{n,j}} &= \frac{1}{\sqrt{\log n}\sqrt{1-\frac{\log j}{\log n}}} = \frac{1}{\sqrt{\log n}}\Big(1+\mathcal{O}(\frac{\log j}{\log n})\Big),\\
e^{\frac{t}{\sigma_{n,j}}} &= e^{\frac{t}{\sqrt{\log n}}}\Big(1+\mathcal{O}(\frac{\log j}{\log^{\frac{3}{2}} n})\Big) = e^{\frac{t}{\sqrt{\log n}}}\Big(1+\mathcal{O}(\frac{\log\log n}{\log^{\frac{3}{2}} n})\Big),
\end{aligned} \tag{5.108}$$

and consequently

$$e^{\frac{t}{\sigma_{n,j}}}-1 = \big(1+\frac{t}{\sqrt{\log n}}+\frac{t^2}{2\log n}+\mathcal{O}(\frac{1}{\log^{\frac{3}{2}} n})\big)\Big(1+\mathcal{O}(\frac{\log\log n}{\log^{\frac{3}{2}} n})\Big)-1 = \frac{t}{\sqrt{\log n}}+\frac{t^2}{2\log n}+\mathcal{O}(\frac{\log\log n}{\log^{\frac{3}{2}} n}). \tag{5.109}$$

We get

$$\frac{\Gamma(j+e^{\frac{t}{\sigma_{n,j}}}-1)}{\Gamma(j)} = 1+\mathcal{O}(\frac{\log\log n}{\sqrt{\log n}}), \tag{5.110}$$

by applying the trivial estimate

$$\frac{\Gamma^{(k)}(j)}{\Gamma(j)} \le k!(\log\log n)^k, \quad k\ge 2 \tag{5.111}$$

to

$$\frac{\Gamma(j+e^{\frac{t}{\sigma_{n,j}}}-1)}{\Gamma(j)} = \sum_{k=0}^{\infty}\frac{\Gamma^{(k)}(j)}{k!\Gamma(j)}(e^{\frac{t}{\sigma_{n,j}}}-1)^k. \tag{5.112}$$

By combining the previous results we obtain

$$\begin{aligned}
p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}) &= n^{e^{\frac{t}{\sigma_{n,j}}}-1}\Big(1+\mathcal{O}(\frac{\log\log n}{\sqrt{\log n}}\Big) = e^{\left(e^{\frac{t}{\sigma_{n,j}}}-1\right)\log n}\Big(1+\mathcal{O}(\frac{\log\log n}{\sqrt{\log n}})\Big)\\
&= e^{\left(t\sqrt{\log n}+\frac{t^2}{2}+\mathcal{O}(\frac{\log\log n}{\sqrt{\log n}})\right)}\Big(1+\mathcal{O}(\frac{\log\log(n)}{\sqrt{\log n}}\Big).
\end{aligned} \tag{5.113}$$

This leads to the required expansion

$$\begin{aligned}
\mathcal{M}_{n,j}(t) &= e^{-\sigma_{n,j}t}p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}) = e^{-\sqrt{\log n}t}e^{\left(t\sqrt{\log n}+\frac{t^2}{2}+\mathcal{O}(\frac{\log\log n}{\sqrt{\log n}})\right)}\Big(1+\mathcal{O}(\frac{\log\log n}{\sqrt{\log n}})\Big)\\
&= e^{\frac{t^2}{2}}\Big(1+\mathcal{O}(\frac{\log\log n}{\sqrt{\log n}})\Big).
\end{aligned} \tag{5.114}$$

For j: $j\to\infty$ such that $j=o(n)$ we simply observe

$$\frac{\Gamma(j+v-1)}{\Gamma(j)} = j^{v-1}\big(1+\mathcal{O}(\frac{1}{j})\big) = e^{(v-1)\log j}\big(1+\mathcal{O}(\frac{1}{j})\big), \quad p_{n,j}(v) = e^{(v-1)\mu_{n,j}}(1+\mathcal{O}(\frac{1}{j})). \tag{5.115}$$

Further

$$\begin{aligned}\mathcal{M}_{n,j}(t) &= e^{-\sigma_{n,j}t}p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}) = e^{-\sigma_{n,j}t}e^{(e^{\frac{t}{\sigma_{n,j}}}-1)\mu_{n,j}}(1+\mathcal{O}(\frac{1}{j}))\\ &= e^{-\sigma_{n,j}t}e^{(\frac{t}{\sigma_{n,j}}+\frac{t^2}{2\mu_{n,j}}+\mathcal{O}(\frac{1}{\sigma^3_{n,j}}))\mu_{n,j}}(1+\mathcal{O}(\frac{1}{j})) = e^{\frac{t^2}{2}+\mathcal{O}(\frac{1}{\sigma_{n,j}})}(1+\mathcal{O}(\frac{1}{j}))\\ &= e^{\frac{t^2}{2}}(1+\mathcal{O}(\frac{1}{j}))(1+\mathcal{O}(\frac{1}{\sigma_{n,j}})) = e^{\frac{t^2}{2}}(1+\mathcal{O}(\frac{1}{j})+\mathcal{O}(\frac{1}{\sigma_{n,j}})).\end{aligned} \tag{5.116}$$

For the region $j = \rho n$ we use (5.115) to obtain the desired result

$$p_{n,j}(v) = e^{-(v-1)\log\rho}(1+\mathcal{O}(\frac{1}{n})) \xrightarrow[n\to\infty]{} p_j(v) = e^{-(v-1)\log\rho}. \tag{5.117}$$

For the region j: $l := n - j = o(n)$:

$$\mathbb{P}\{X_{n,j}=0\} = \frac{\binom{n-2}{j-2}}{\binom{n-1}{j-1}} = \frac{j-1}{n-1} = \frac{n-1-l}{n-1} = 1 - \frac{l}{n-1} = 1-o(1). \tag{5.118}$$

5.5.6 Case B

For Case B we proceed as follows. For fixed j we get

$$\begin{aligned}\mathbb{P}\{X_{n,j}=d\} &= 1+\sum_{k=0}^{d-1}\binom{d}{k}(-1)^{d-k}\frac{\Gamma(n-1+k\frac{1}{d-1})\Gamma(j+\frac{1}{d-1})}{\Gamma(j-1+k\frac{1}{d-1})\Gamma(n+\frac{1}{d-1})}\\ &= 1+\sum_{k=0}^{d-1}\binom{d}{k}(-1)^{d-k}\frac{\Gamma(j+\frac{1}{d-1})}{\Gamma(j-1+k\frac{1}{d-1})}\mathcal{O}(\frac{1}{n^{1-\frac{k-1}{d-1}}}) = 1+\mathcal{O}(\frac{1}{n^{\frac{1}{d-1}}}),\end{aligned} \tag{5.119}$$

and for j: $j \to \infty$ such that $j = o(n)$

$$\begin{aligned}\mathbb{P}\{X_{n,j}=d\} &= 1+\sum_{k=0}^{d-1}\binom{d}{k}(-1)^{d-k}\frac{\Gamma(n-1+k\frac{1}{d-1})\Gamma(j+\frac{1}{d-1})}{\Gamma(j-1+k\frac{1}{d-1})\Gamma(n+\frac{1}{d-1})}\\ &= 1+\sum_{k=0}^{d-1}\binom{d}{k}(-1)^{d-k}\mathcal{O}(\frac{j^{1-\frac{k-1}{d-1}}}{n^{1-\frac{k-1}{d-1}}}) = 1+\mathcal{O}\Big((\frac{j}{n})^{\frac{1}{d-1}}\Big).\end{aligned} \tag{5.120}$$

The same approach also works for j: $j \to \infty$ such that $j = \rho n$

$$\begin{aligned}\mathbb{P}\{X_{n,j}=m\} &= \binom{d}{m}\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}\frac{\Gamma(n-1+k\frac{1}{d-1})\Gamma(\rho n+\frac{1}{d-1})}{\Gamma(\rho n-1+k\frac{1}{d-1})\Gamma(n+\frac{1}{d-1})}\\ &= \binom{d}{m}\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}\rho^{1-\frac{k-1}{d-1}}(1+\mathcal{O}(\frac{1}{n}))\\ &= \binom{d}{m}\rho^{1+\frac{1}{d-1}}\sum_{k=0}^{m}\binom{m}{k}(-1)^{m-k}\rho^{-\frac{k}{d-1}}+\mathcal{O}(\frac{1}{n}) = \binom{d}{m}\rho^{1+\frac{1}{d-1}}(\rho^{-\frac{1}{d-1}}-1)^m+\mathcal{O}(\frac{1}{n}).\end{aligned} \tag{5.121}$$

For the region j: $l := n - j = o(n)$:

$$\mathbb{P}\{X_{n,j}=0\} = \frac{\Gamma(n-1)\Gamma(j+\frac{1}{d-1})}{\Gamma(j-1)\Gamma(n+\frac{1}{d-1})} = \frac{(n-2)^{\underline{l}}}{(n-1+\frac{1}{d-1})^{\underline{l}}} \geq \frac{(n-2)^{\underline{l}}}{(n-1)^{\underline{l}}} = \frac{j-1}{n-1} = 1-\frac{l}{n-1}. \tag{5.122}$$

5.5.7 Case C

Since we already know the factorial moments from Subsection 5.5.3, we can express the ordinary moments by the using the Stirling numbers of the second kind $\left\{ {n \atop k} \right\}$.

$$\mathbb{E}\big(X_{n,j}^{s}\big) = \mathbb{E}\big(X_{n,j}^{\underline{s}}\big) + \sum_{k=1}^{s-1} \left\{ {s \atop k} \right\} \mathbb{E}\big(X_{n,j}^{\underline{k}}\big). \tag{5.123}$$

Using again Stirling's formula for the Gamma function (5.106) we get from the explicit result for the factorial moments in Theorem 20 the asymptotic expansion

$$\mathbb{E}(X_{n,j}^{\underline{s}}) = \frac{\Gamma(s-1-\frac{c_1}{c_2})\Gamma(j+\frac{c_2}{c_1})}{\Gamma(-1-\frac{c_1}{c_2})\Gamma(j-\frac{(s-1)c_2}{c1})} n^{-\frac{c_2}{c_1}s} + \mathcal{O}(n^{-\frac{c_2}{c_1}(s-1)}), \tag{5.124}$$

which leads by (5.123) to an asymptotic expansion of the ordinary s-th moment. Thus the moments of $n^{\frac{c_2}{c_1}} X_{n,j}$ converge to a random variable X_j with s-th moment

$$\mathbb{E}(X_j^s) = \frac{\Gamma(s-1-\frac{c_1}{c_2})\Gamma(j+\frac{c_2}{c_1})}{\Gamma(-1-\frac{c_1}{c_2})\Gamma(j-\frac{(s-1)c_2}{c1})}. \tag{5.125}$$

For the region $j \to \infty$ such that $j = o(n)$ we proceed exactly as before arriving at $\frac{n^{\frac{c_2}{c_1}}}{j^{\frac{c_2}{c_1}}} X_{n,j} \xrightarrow{(d)} X_j \stackrel{(d)}{=} \gamma(a, \lambda)$, where X_j is Gamma distributed with parameters $a = -\frac{c_1}{c_2} - 1$ and $\lambda = 1$.

$$\mathbb{E}(X_j) = \frac{\Gamma(s-1-\frac{c_1}{c_2})}{\Gamma(-1-\frac{c_1}{c_2})}. \tag{5.126}$$

For the region $j \to \infty$ such that $j = \rho n$ we get after applying (5.106) to $\mathbb{P}\{X_{n,j} = m)$ the asymptotic expansion

$$\mathbb{P}\{X_{n,j} = m\} = \binom{m-2-\frac{c_1}{c_2}}{m} \rho^{1+\frac{c_2}{c_1}} \big(1-\rho^{-\frac{c_2}{c_1}}\big)^m + \mathcal{O}(\frac{1}{n}), \tag{5.127}$$

which proves the result. The result corresponding to the region $l = n - j = o(n)$ follows by a similar observation.

5.5.8 Density appearing in Case C

At first we show that $f_1(x)$ is a density with the required moments

$$\frac{\Gamma(s-1-\frac{c_1}{c_2})\Gamma(1+\frac{c_2}{c_1})}{\Gamma(-1-\frac{c_1}{c_2})\Gamma(1-\frac{(s-1)c_2}{c1})}. \tag{5.128}$$

The density $f_1(x)$ can be derived by using Hankel contour technics similar to [10; 66]. Then we sketch how to obtain the density $f_j(x)$ by combining a previous result ([46]) about the random variable $D_{n,j}$, which counts the size of the subtree attached to node j in a random increasing tree of size n, and the shape of $f_1(x)$. First we turn to the root $j = 1$:

$$f_1(x) = \frac{-\Gamma(\frac{c_2}{c_1})x^{-2-\frac{c_1}{c_2}}}{\Gamma(-1-\frac{c_1}{c_2})\pi} \int_0^\infty e^{-(r^{-\frac{c_1}{c_2}})} r^{-\frac{c_1}{c_2}-1} e^{-xr\cos(\frac{c_2}{c_1}\pi)} \sin\big(xr\sin(-\frac{c_2}{c_1}\pi)\big) dr. \tag{5.129}$$

We proceed as follows.

$$\begin{aligned}\int_0^\infty x^s f(x)dx &= \frac{-\Gamma(\frac{c_2}{c_1})}{\Gamma(-1-\frac{c_1}{c_2})\pi}\int_0^\infty e^{-(r^{-\frac{c_1}{c_2}})}r^{-\frac{c_1}{c_2}-1}\int_0^\infty x^{s-2-\frac{c_1}{c_2}}e^{-xr\cos(\frac{c_2}{c_1}\pi)}\sin\big(xr\sin(-\frac{c_2}{c_1}\pi)\big)dx\,dr\\ &= \Im\Big(\frac{-\Gamma(\frac{c_2}{c_1})}{\Gamma(-1-\frac{c_1}{c_2})\pi}\int_0^\infty e^{-(r^{-\frac{c_1}{c_2}})}r^{-\frac{c_1}{c_2}-1}\int_0^\infty x^{s-2-\frac{c_1}{c_2}}e^{-xr\cos(\frac{c_2}{c_1}\pi)}e^{ixr\sin(-\frac{c_2}{c_1}\pi)}dx\,dr\Big)\\ &= \Im\Big(\frac{-\Gamma(\frac{c_2}{c_1})}{\Gamma(-1-\frac{c_1}{c_2})\pi}\int_0^\infty e^{-(r^{-\frac{c_1}{c_2}})}r^{-\frac{c_1}{c_2}-1}\int_0^\infty x^{s-2-\frac{c_1}{c_2}}e^{-xr(e^{i\frac{c_2}{c_1}\pi})}dx\,dr\Big).\end{aligned} \tag{5.130}$$

Using the substitution $u = xre^{i\frac{c_2}{c_1}\pi}$, $\frac{du}{dx} = re^{i\frac{c_2}{c_1}\pi}$ we get further

$$\begin{aligned}\int_0^\infty x^s f_1(x)dx &= \Im\Big(\frac{-\Gamma(\frac{c_2}{c_1})}{\Gamma(-1-\frac{c_1}{c_2})\pi}\int_0^\infty e^{-(r^{-\frac{c_1}{c_2}})}r^{-\frac{c_1}{c_2}-1}\int_0^\infty (\frac{ue^{-i\frac{c_2}{c_1}\pi}}{r})^{s-2-\frac{c_1}{c_2}}e^{-u}\frac{e^{-i\frac{c_2}{c_1}\pi}}{r}du\,dr\Big)\\ &= \Im\Big(\frac{-\Gamma(\frac{c_2}{c_1})e^{-\frac{c_2}{c_1}\pi i(s-1-\frac{c_1}{c_2})}}{\Gamma(-1-\frac{c_1}{c_2})\pi}\int_0^\infty e^{-(r^{-\frac{c_1}{c_2}})}r^{-s}\int_0^\infty u^{s-2-\frac{c_1}{c_2}}e^{-u}du\,dr\Big)\\ &= \frac{-\Gamma(\frac{c_2}{c_1})\sin(1-\frac{c_2}{c_1}(s-1))\Gamma(s-1-\frac{c_1}{c_2})}{\Gamma(-1-\frac{c_1}{c_2})\pi}\int_0^\infty e^{-(r^{-\frac{c_1}{c_2}})}r^{-s}dr\\ &= \frac{-\Gamma(\frac{c_2}{c_1})\sin(1-\frac{c_2}{c_1}(s-1))\Gamma(s-1-\frac{c_1}{c_2})(-\frac{c_2}{c_1})\Gamma(\frac{c_2}{c_1}(s-1))}{\Gamma(-1-\frac{c_1}{c_2})\pi}\\ &= \frac{\Gamma(s-1-\frac{c_1}{c_2})\Gamma(1+\frac{c_2}{c_1})}{\Gamma(-1-\frac{c_1}{c_2})\Gamma(1-\frac{(s-1)c_2}{c1})},\end{aligned} \tag{5.131}$$

where we have used the identity $\frac{\pi}{\sin(\pi x)} = \Gamma(x)\Gamma(1-x)$. For the density $f_j(x)$ we can proceed as follows. We know (see [46] for details) that for fixed j it holds $\frac{D_{n,j}}{n} \xrightarrow{(d)} \beta(\frac{c_2}{c_1}+1, j-1)$, where $\beta(a,b)$ denotes a Beta distribution with parameter a and b, and thus

$$\lim_{n\to\infty} n\mathbb{P}\{\frac{D_{n,j}}{n} = t\} = \frac{t^{-\frac{c_2}{c_1}}(1-t)^{j-2}}{\mathrm{B}(\frac{c_2}{c_1}+1, j-1)}, \tag{5.132}$$

where $\mathrm{B}(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ denotes the Beta function of p and q. By conditioning on the size of the subtree rooted at node j the r.v. $Z_{n,j}$ we get

$$\mathbb{P}\{X_{n,j} = m\} = \sum_{k\ge 0}\mathbb{P}\{X_{k,1} = m | Z_{n,j} = k\}\mathbb{P}\{Z_{n,j} = k\} = \sum_{k\ge 0}\mathbb{P}\{X_{k,1} = m\}\mathbb{P}\{Z_{n,j} = k\}. \tag{5.133}$$

A standard argument leads then to (5.134).

$$f_j(x) = \frac{\Gamma(j+\frac{c_2}{c_1})}{\Gamma(1+\frac{c_2}{c_1})\Gamma(j-1)}\int_0^1 t^{2\frac{c_2}{c_1}}(1-t)^{j-2}f_1\big(xt^{\frac{c_2}{c_1}}\big)dt\,dx, \tag{5.134}$$

where $f_1(x)$ is given by (5.129). Using the substitution $u = xt^{\frac{c_2}{c_1}}$ it can easily be seen that the density $f_j(x)$ has the required moments.

5.6 The branching structure

Let $X_{n,j,a}$ denote the random variable, which counts the number of size a branches (subtrees) attached to node j in a random grown simple increasing tree of size n. This was studied by Hu, Feng and Su in [33] for random recursive trees: they derived the distribution of $X_{n,1,a}$ and a limit law for it. Further they stated the joint distribution of $X_{n,1,1}X_{n,1,2}\dots X_{n,1,n}$.

By using our approach we can extend their results to arbitrary grown simple families of increasing trees. We can give closed formulæ for the probability distribution and factorial moments. The joint distribution is computed for all grown simple families of increasing trees. Furthermore limiting distribution results are given for the full region of j, $1 \le j \le n$.

5.6.1 Results for the probabilities

Theorem 24. *The probability that there are m size a branches in the subtree rooted at node j, for every $j \ge 1$,*

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

$$\mathbb{P}\{X_{n,j,a} = m\} = \frac{1}{a^m a! \binom{n-1}{j-1}} \sum_{l=0}^{\lfloor \frac{n-j-am}{a} \rfloor} \frac{(-1)^l}{a^l l!} \binom{n-1-a(m+l)}{j-1}. \tag{5.135}$$

- *Case B (d-ary increasing trees): we obtain the following exact formula.*

$$\begin{aligned}\mathbb{P}\{X_{n,j,a} = m\} = \frac{\binom{d}{m}\binom{\frac{d}{d-1}+j-2}{j-1}}{\binom{n-1}{j-1}\binom{n-1+\frac{1}{d-1}}{n-1}} &\sum_{i=0}^{min\{d-m,\lfloor \frac{n-j-am}{m} \rfloor\}} \binom{d-m}{i}(-1)^i \left(\frac{\binom{a-1+\frac{1}{d-1}}{a-1}}{a(d-1)}\right)^{i+m} \times \\ &\times \binom{n-2-a(i+m)+\frac{d-m-i}{d-1}}{n-j-a(i+m)}.\end{aligned} \tag{5.136}$$

- *Case C (Generalized plane oriented recursive trees): For the root $j=1$ it holds*

$$\mathbb{P}\{X_{n,1,a} = m) = \binom{m-2-\frac{c_1}{c_2}}{m} \Big(\frac{\varphi_0 c_1 a}{(-c_2)\binom{a-1+\frac{c_2}{c_1}}{a-1}}\Big)^{-\frac{c_2}{c_1}} \frac{(1-\frac{c_2}{c_1}(m-1))}{n} + \mathcal{O}(\frac{1}{n^{1-\frac{c_2}{c_1}}}), \tag{5.137}$$

and for $j > 1$

$$\mathbb{P}\{X_{n,j,a} = m) = \binom{\frac{c_2}{c_1}+j-1}{j-1}\binom{m-2-\frac{c_2}{c_1}}{m}\Big(\frac{\varphi_0 c_1 a}{(-c_2)\binom{a-1+\frac{c_2}{c_1}}{a-1}}\Big)^{-\frac{c_2}{c_1}-1} \frac{(j-1)\Gamma(1+\frac{c_2}{c_1})}{n^{1-\frac{c_2}{c_1}}} + \mathcal{O}(\frac{1}{n^{1-2\frac{c_2}{c_1}}}). \tag{5.138}$$

5.6.2 Explicit formulæ for the factorial moments

Theorem 25. *The factorial moments of the random variable $X_{n,j,a}$ counting the number of size a subtrees attached to node j in a size n random grown simple increasing tree is given as follows.*

- *Case A (recursive trees):*

$$\mathbb{E}(X^{\underline{s}}_{n,j,a}) = \frac{1}{a^s} \frac{\binom{n-as-1}{j-1}}{\binom{n-1}{j-1}}. \tag{5.139}$$

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

$$\mathbb{E}(X_{n,j,a}^{\underline{s}}) = \left(\frac{\binom{a-1+\frac{1}{d-1}}{a-1}}{a(d-1)} \right)^s \frac{\binom{j-1+\frac{1}{d-1}}{j-1}\binom{n-as-2+\frac{d-s}{d-1}}{n-j-as}}{\binom{n-1}{j-1}\binom{n-1+\frac{1}{d-1}}{n-1}}. \tag{5.140}$$

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

$$\mathbb{E}(X_{n,j,a}^{\underline{s}}) = \left(\frac{-c_2\binom{a-1+\frac{c_2}{c_1}}{a-1}}{c_1 a} \right)^s \frac{\Gamma(s-1-\frac{c_1}{c_2})}{\Gamma(-1-\frac{c_1}{c_2})} \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\binom{n-as-1-\frac{c_2}{c_1}(s-1)}{n-j-as}}{\binom{n-1}{j-1}\binom{n-1+\frac{c_2}{c_1}}{n-1}}. \tag{5.141}$$

5.6.3 Joint distributions

Theorem 26. *The joint distribution of the random variables $X_{n,1,1} \dots X_{n,1,n}$ is for all Cases A,B and C given as follows.*

$$\mathbb{P}\{X_{n,1,1} = m_1, \dots, X_{n,1,n-1} = m_{n-1}\} = \frac{\varphi_{\sum_{i=1}^{n-1} m_i} \varphi_0^{\sum_{i=1}^{n-1} m_i - 1} (\sum_{i=1}^{n-1} m_i)!}{\binom{n-1+\frac{c_2}{c_1}}{n-1}} \prod_{k=1}^{n-1} \frac{\binom{k-1+\frac{c_2}{c_1}}{k-1}}{k^{m_k} m_k!}, \tag{5.142}$$

for all sequences of non-negative integers satisfying $\sum_{k=1}^{n-1} k m_k = n-1$, where one has $c_2 = 0$ for Case A and $\frac{c_2}{c_1} = \frac{1}{d-1}$ for Case B. Further for Case B (d-ary increasing trees) we have the additional constraint $\sum_{k=1}^{n-1} m_k \leq d$.

Corollar 7. *The joint distribution of the random variables $X_{n,j,1} \dots X_{n,j,n}$ is for all Cases A,B and C given as follows.*

$$\begin{aligned} \mathbb{P}\{X_{n,j,1} = m_1, \dots, X_{n,j,n-j} = m_{n-j}\} &= \mathbb{P}\{Z_{n,j} = 1 + \sum_{i=1}^{n-j} i m_i\} \times \\ &\quad \times \mathbb{P}\{X_{1+\sum_{i=1}^{n-j} i m_i,1,1} = m_1, \dots, X_{1+\sum_{i=1}^{n-j} i m_i,1,n-j} = m_{n-j}\}, \end{aligned} \tag{5.143}$$

where $Z_{n,j}$ denotes the r.v. counting the subtree size of node j.

5.6.4 Limit distribution results

Theorem 27. *For Case A (recursive trees) we get*

- *for $n \to \infty$, $j = o(n)$ and a fixed: the random variable $X_{n,j,a}$ is asymptotically Poisson distributed with parameter $1/a$:*

$$X_{n,j,a} \xrightarrow{(d)} X_a, \qquad \mathbb{P}\{X_a = m\} = \frac{e^{-\frac{1}{a}}}{a^m m!}. \tag{5.144}$$

- *for $n \to \infty$, $j = \rho n$ with $0 < \rho < 1$ and a fixed: the random variable $X_{n,j,a}$ is asymptotically Poisson distributed with parameter $(1-\rho)/a$:*

$$X_{n,j,a} \xrightarrow{(d)} X_{\rho,a}, \qquad \mathbb{P}\{X_{\rho,a} = m\} = \frac{e^{-\frac{1-\rho}{a}}(1-\rho)^m}{a^m m!}. \tag{5.145}$$

- *for $n \to \infty$ and all other cases of j and a it holds*

$$X_{n,j,a} \xrightarrow{(d)} X, \qquad \mathbb{P}\{X = 0\} = 1. \tag{5.146}$$

Theorem 28. *In the Case B (d-ary increasing trees) we obtain the following characterization.*

- *for $n \to \infty$, $j = o(n)$ and $1 \le a \le n-1$ fixed the random variable $X_{n,j,a}$ is asymptotically zero.*

$$X_{n,j,a} \xrightarrow{(d)} X, \qquad \mathbb{P}\{X = 0\} = 1. \tag{5.147}$$

- *for $n \to \infty$, $j = \rho n$ with $0 < \rho < 1$ and a fixed the random variable $X_{n,j,a}$ is asymptotically binomial distributed* $\text{Bin}(N, p)$ *with parameter $N = d$ and $p = p(a,d) = \binom{a-1+\frac{1}{d-1}}{a-1}\frac{\rho(1-\rho)^a}{a(d-1)}$:*

$$X_{n,j,a} \xrightarrow{(d)} X_{\rho,a}, \quad \mathbb{P}\{X_{\rho,a} = m\} = \binom{d}{m} p^m (1-p)^{d-m} \tag{5.148}$$

- *for $n \to \infty$, $l = n - j = o(n)$ and $1 \le a \le n-1$ fixed the random variable $X_{n,j,a}$ is asymptotically zero.*

$$X_{n,j,a} \xrightarrow{(d)} X, \qquad \mathbb{P}\{X = 0\} = 1. \tag{5.149}$$

Theorem 29. *For Case C (Generalized plane oriented recursive trees) we obtain the following characterization.*

- *For fixed $a \ge 1$, fixed $j \ge 1$ and $n \to \infty$ we get*

$$n^{\frac{c_2}{c_1}} X_{n,j} \xrightarrow{d} X, \quad \mathbb{E}(X^s) = \frac{\Gamma(j + \frac{c_2}{c_1})\Gamma(s - 1 - \frac{c_1}{c_2})}{\Gamma(j - \frac{c_2}{c_1}(s-1))\Gamma(-1 - \frac{c_1}{c_2})} \Big(\frac{(-c_2)\binom{a-1+\frac{c_2}{c_1}}{a-1}}{c_1 a}\Big)^s. \tag{5.150}$$

- *The region j: $j \to \infty$ such that $j = o(n)$ and fixed $a \ge 1$ we get a gamma distribution $\gamma(k, \lambda)$ with parameters $k = -\frac{c_1}{c_2} - 1$ and $\lambda = \frac{ac_1}{(-c_2)\binom{a-1+\frac{c_2}{c_1}}{a-1}}$*

$$\Big(\frac{n}{j}\Big)^{\frac{c_2}{c_1}} X_{n,j} \xrightarrow{d} X, \quad X \stackrel{d}{=} \gamma(k,\lambda), \quad \mathbb{E}(X^s) = (-\frac{c_1}{c_2} - 1)^{\overline{s}} \Big(\frac{(-c_2)\binom{a-1+\frac{c_2}{c_1}}{a-1}}{c_1 a}\Big)^s. \tag{5.151}$$

- *The region j: $j \to \infty$ such that $j = \rho n$ and fixed $a \ge 1$: The r. v. $X_{n,j,a}$ is asymptotically negative binomial distributed* $\text{NegBin}(r, p)$ *with parameters $r = -1 - \frac{c_1}{c_2}$ and $p = p(\rho, c_1, c_2, a) = (1 + (1-\rho)^a \rho^{\frac{c_2}{c_1}} \binom{a-1+\frac{c_2}{c_1}}{a-1} \frac{(-c_2)}{c_1 a})^{-1}$.*

$$X_{n,j,a} \xrightarrow{(d)} X_{\rho,a}, \quad \textit{with} \quad \mathbb{P}\{X_\rho = m\} = \binom{m - 2 - \frac{c_1}{c_2}}{m} p^{1+\frac{c_2}{c_1}} (1-p)^m. \tag{5.152}$$

- *The region j: $n - j = l = o(n)$*

$$X_{n,j,a} \xrightarrow{(d)} X, \qquad \mathbb{P}\{X = 0\} = 1. \tag{5.153}$$

Remark 11. As mentioned earlier the result concerning recursive trees (basically Case A) and $j = 1$ in the Theorems 24, 26 and 27 already appeared in Hu, Feng & Su; 2005, which motivated this research.

Remark 12. For Cases B and C the limit laws of $X_{n,j,a}$ resemble the limit laws of the out-degree $Y_{n,j}$ of node j (compare with [47]). This is no coincidence since

$$Y_{n,j} \stackrel{(d)}{=} X_{n,j,1} + \cdots + X_{n,j,n-j}. \tag{5.154}$$

For Case A the limit law is different from the behavior of the node degree. Note that the random variable $X_{n,a}$, which counts the number of size a trees on the fringe of a size n random grown simple increasing tree is related to $X_{n,j,a}$ by

$$X_{n,a} \stackrel{(d)}{=} X_{n,1,a} + X_{n,2,a} + \cdots + X_{n,n+1-a,a}. \tag{5.155}$$

5.6.5 Results for a randomly chosen node

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

Theorem 30. *The limit law of the r. v. $X_{n,a}^{[R]}$ counting the number of size a branches in a size n grown simple increasing tree is given as follows.*

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

$$X_{n,a}^{[R]} \xrightarrow{(d)} X_a, \qquad \mathbb{P}\{X_a = m\} = a\Big(1 - e^{-\frac{1}{a}} \sum_{k=0}^{m} \frac{1}{k!a^k}\Big) = \sum_{k \ge m+1} \frac{ae^{-\frac{1}{a}}}{k!a^k}. \tag{5.156}$$

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

$$X_{n,a}^{[R]} \xrightarrow{(d)} X_a, \qquad \mathbb{P}\{X_a = m\} = \sum_{k=0}^{d-m} (-1)^k \frac{\binom{d}{m}\binom{d-m}{k}\binom{a-1+\frac{1}{d-1}}{a-1}^{m+k}}{a^{m+k}(d-1)^{m+k}((m+k)(d+1)+1)\binom{(d+1)(m+k)}{m+k}}. \tag{5.157}$$

Remark 13. For Case C we were not able to obtain an explicit formula for the probability $\mathbb{P}\{X_a = m\}$, although we have the closed formula $\mathbb{P}\{X_a = m\} = \int_0^1 \mathbb{P}\{X_{\rho,a} = m\} d\rho$.

5.6.6 Deriving the generation function for the root

In order to use the approach (5.3) developed here to characterize $X_{n,j,a}$ one has to calculate at first $\frac{\partial}{\partial z} M(z,v)$. By definition one gets the following explicit characterization for the probabilities $\mathbb{P}\{X_{n,1,a} = m\}$:

$$\mathbb{P}\{X_{n,1,a} = m\} = \sum_{r \ge m} \varphi_r \binom{r}{m} \sum_{\substack{k_1 + \cdots + k_r = n-1 \\ k_i = a \text{ for } 1 \le i \le m \\ k_j \ne a \text{ for } m+1 \le j \le r}} \frac{T_{k_1} \dots T_{k_r}}{T_n} \binom{n-1}{k_1, \dots, k_r} \tag{5.158}$$

for $n \ge 1$ and $m \ge 0$. By multiplying with $T_n z^{n-1} v^m/(n-1)!$ and summing up over $n \ge 1$, $m \ge 0$ this can be turned into the an explicit formula for $\frac{\partial}{\partial z} M(z,v)$.

$$\begin{aligned} \frac{\partial}{\partial z} M(z,v) &= \sum_{n \ge 1} \sum_{m \ge 0} \varphi_n \binom{n}{m} (T_a z^a v)^m (T(z) - T_a z^a)^{n-m} = \sum_{n \ge 1} \varphi_n (T_a z^a v + T(z) - T_a z^a)^n \\ &= \varphi(T(z) + T_a z^a (v-1)). \end{aligned} \tag{5.159}$$

This leads for grown simple families of increasing trees to

$$\frac{\partial}{\partial z} M(z,v) = \begin{cases} \dfrac{\varphi_0}{1 - c_1 z} \exp\Big(\dfrac{c_1 T_a}{\varphi_0 a!} z^a (v-1)\Big), & \text{Case A,} \\ \varphi_0 \Big(\dfrac{c_2 T_a}{\varphi_0 a!} z^a (v-1) + \dfrac{1}{(1-(d-1)c_2 z)^{\frac{1}{d-1}}}\Big)^d, & \text{Case B,} \\ \dfrac{\varphi_0}{\Big(\frac{1}{(1-c_1 z)^{\frac{c_2}{c_1}}} + \frac{c_2 T_a}{\varphi_0 a!} z^a (v-1)\Big)^{-\frac{c_1}{c_2}-1}} & \text{Case C,} \end{cases} \tag{5.160}$$

where $T_a = \varphi_0 c_1^{a-1}(a-1)!\binom{a-1+\frac{c_2}{c_1}}{a-1}$.

5.6.7 Deriving the probabilities

We will restrict ourselves to Case A and Case C, Case B is fully analogous. By simply applying (5.4) we get for Case A

$$\begin{aligned}
\mathbb{P}\{X_{n,j,a}=m\} &= \frac{(j-1)!(n-j)!}{T_n}[u^{n-j}v^m]\frac{c_1^{j-1}}{(1-c_1u)^{j-1}}\frac{\partial}{\partial u}M(u,v)\\
&= \frac{1}{\binom{n-1}{j-1}c_1^{n-j}}[u^{n-j}v^m]\frac{\exp\Big(-\frac{c_1T_a}{\varphi_0 a!}u^a\Big)}{(1-c_1u)^j}\exp\Big(v\frac{c_1T_a}{\varphi_0 a!}u^a\Big)\\
&= \frac{c_1^{am}}{a^m m!\binom{n-1}{j-1}c_1^{n-j}}[u^{n-j-ma}]\frac{\exp\Big(-\frac{c_1^a}{a}u^a\Big)}{(1-c_1u)^j}\\
&= \frac{1}{a^m m!\binom{n-1}{j-1}}\sum_{l=0}^{\lfloor\frac{n-j-am}{a}\rfloor}\frac{(-1)^l}{a^l l!}\binom{n-1-a(m+l)}{j-1}.
\end{aligned} \tag{5.161}$$

For Case B we get

$$\begin{aligned}
\mathbb{P}\{X_{n,j}=m\} &= \frac{(c_2(d-1))^{j-1}\binom{j-1+\frac{1}{d-1}}{j-1}}{\binom{n-1}{n-j}\varphi_0(c_2(d-1))^{n-1}\binom{n-1+\frac{1}{d-1}}{n-1}}[u^{n-j}v^m]\frac{\varphi_0\Big(\frac{c_2T_a}{\varphi_0 a!}z^a(v-1)+\frac{1}{(1-(d-1)c_2z)^{\frac{1}{d-1}}}\Big)^d}{(1-c_2(d-1)u)^{j-1}}\\
&= \frac{\binom{j-1+\frac{1}{d-1}}{j-1}\binom{d}{m}\Big(\frac{c_2T_a}{\varphi_0 a!}\Big)^m}{\binom{n-1}{j-1}(c_2(d-1))^{n-j}\binom{n-1+\frac{1}{d-1}}{n-1}}[u^{n-j-am}]\sum_{i=0}^{d-m}\binom{d-m}{i}\frac{(-1)^i\Big(\frac{c_2T_a}{\varphi_0 a!}\Big)^i u^{ai}}{(1-(d-1)c_2u)^{j-1+\frac{d-m-i}{d-1}}}\\
&= \frac{\binom{d}{m}\binom{\frac{d}{d-1}+j-2}{j-1}}{\binom{n-1}{j-1}\binom{n-1+\frac{1}{d-1}}{n-1}}\sum_{i=0}^{min\{d-m,\lfloor\frac{n-j-am}{m}\rfloor\}}\binom{d-m}{i}(-1)^i\left(\frac{\binom{a-1+\frac{1}{d-1}}{a-1}}{a(d-1)}\right)^{i+m}\times\\
&\quad\times\binom{n-2-a(i+m)+\frac{d-m-i}{d-1}}{n-j-a(i+m)}.
\end{aligned} \tag{5.162}$$

For Case C we observe the following

$$\begin{aligned}
\mathbb{P}\{X_{n,j,a}=m\} &= \frac{(j-1)!(n-j)!}{T_n}[u^{n-j}v^m]\frac{\binom{\frac{c_2}{c_1}+j-1}{j-1}c_1^{j-1}}{(1-c_1u)^{j-1}}M'(u,v)\\
&= \frac{(j-1)!(n-j)!}{T_n}[u^{n-j}v^m]\frac{\varphi_0\binom{\frac{c_2}{c_1}+j-1}{j-1}c_1^{j-1}}{(1-c_1u)^{j-1}\Big(\frac{1}{(1-c_1u)^{\frac{c_2}{c_1}}}+\frac{c_2T_a}{\varphi_0 a!}u^a(v-1)\Big)^{-\frac{c_1}{c_2}-1}}\\
&= \frac{(j-1)!(n-j)!}{T_n}[u^{n-j}v^m]\frac{\varphi_0\binom{\frac{c_2}{c_1}+j-1}{j-1}c_1^{j-1}}{(1-c_1u)^{j-1}\big(C(u)\big)^{-\frac{c_1}{c_2}-1}\Big(1+\frac{\frac{c_2T_a}{\varphi_0 a!}u^a v}{C(u)}\Big)^{-\frac{c_1}{c_2}-1}},
\end{aligned} \tag{5.163}$$

where $C(u)$ is defined as

$$C(u) := \frac{1}{(1-c_1u)^{\frac{c_2}{c_1}}} - \frac{c_2T_a}{\varphi_0 a!}u^a. \tag{5.164}$$

Thus we get further

$$\mathbb{P}\{X_{n,j,a}=m\}=\frac{(j-1)!(n-j)!}{T_n}[u^{n-j}]\frac{\varphi_0\binom{\frac{c_2}{c_1}+j-1}{j-1}c_1^{j-1}}{(1-c_1u)^{j-1}\big(C(u)\big)^{-\frac{c_1}{c_2}-1}}\binom{\frac{c_1}{c_2}+1}{m}\Big(\frac{c_2T_a}{\varphi_0a!}\Big)^m\frac{u^{am}}{(C(u))^m}. \quad (5.165)$$

Now we expand $\frac{1}{(C(u))^{m-\frac{c_1}{c_2}-1}}$ around $u=1/c_1$.

$$\begin{aligned}\frac{1}{(C(u))^{-\frac{c_1}{c_2}-1+m}}&=\frac{1}{\Big((1-c_1u)^{-\frac{c_2}{c_1}}-\frac{c_2T_a}{\varphi_0a!}u^a\Big)^{-\frac{c_1}{c_2}-1+m}}\\&=\frac{1}{\Big(\frac{-c_2T_a}{\varphi_0a!}u^a\Big)^{-\frac{c_1}{c_2}-1+m}}\frac{1}{\Big(1-\frac{\varphi_0a!}{u^ac_2T_a}(1-c_1u)^{-\frac{c_2}{c_1}}\Big)^{-\frac{c_1}{c_2}-1+m}}\\&=\frac{1}{\Big(\frac{-c_2T_a}{\varphi_0a!c_1^a}\Big)^{-\frac{c_1}{c_2}-1+m}}\Bigg(1+\frac{\frac{c_1}{c_2}+1-m}{\frac{-c_2T_a}{\varphi_0a!c_1^a}}(1-c_1u)^{-\frac{c_2}{c_1}}+\mathcal{O}((1-c_1u)^{-\frac{2c_2}{c_1}})\Bigg).\end{aligned} \quad (5.166)$$

For the root,$j=1$, we proceed from (5.165) as follows.

$$\begin{aligned}\mathbb{P}\{X_{n,1,a}=m\}&=\frac{(n-1)!}{T_n}[u^{n-1}]\varphi_0\binom{\frac{c_1}{c_2}+1}{m}\Big(\frac{c_2T_a}{\varphi_0a!}\Big)^m\frac{u^{am}}{(C(u))^{-\frac{c_1}{c_2}-1+m}}\\&=\frac{(n-1)!}{T_n}[u^{n-1-am}]\varphi_0\binom{\frac{c_1}{c_2}+1}{m}\Big(\frac{c_2T_a}{\varphi_0a!}\Big)^m\frac{(\frac{c_1}{c_2}+1-m)(1-c_1u)^{-\frac{c_2}{c_1}}}{\Big(\frac{-c_2T_a}{\varphi_0a!c_1^a}\Big)^{-\frac{c_1}{c_2}+m}}\\&=\frac{(n-1)!}{T_n}[u^{n-1-am}]\varphi_0\binom{\frac{c_1}{c_2}+1}{m}(-1)^mc_1^{am}\Big(\frac{-c_2T_a}{\varphi_0a!c_1^a}\Big)^{\frac{c_1}{c_2}}(\frac{c_1}{c_2}+1-m)(1-c_1u)^{-\frac{c_2}{c_1}},\end{aligned} \quad (5.167)$$

while for $j>1$ we have

$$\begin{aligned}\mathbb{P}\{X_{n,j,a}=m\}&=\frac{(j-1)!(n-j)!}{T_n}[u^{n-j-am}]\frac{\varphi_0\binom{\frac{c_2}{c_1}+j-1}{j-1}c_1^{j-1}}{(1-c_1u)^{j-1}}\binom{\frac{c_1}{c_2}+1}{m}\frac{\Big(\frac{c_2T_a}{\varphi_0a!}\Big)^m}{\Big(\frac{-c_2T_a}{\varphi_0a!c_1^a}\Big)^{-\frac{c_1}{c_2}-1+m}}\\&=\frac{(j-1)!(n-j)!}{T_n}[u^{n-j-am}]\frac{\varphi_0\binom{\frac{c_2}{c_1}+j-1}{j-1}c_1^{j-1}}{(1-c_1u)^{j-1}}\binom{\frac{c_1}{c_2}+1}{m}(-1)^mc_1^{am}\Big(\frac{-c_2T_a}{\varphi_0a!c_1^a}\Big)^{\frac{c_1}{c_2}+1}.\end{aligned} \quad (5.168)$$

Now we use singularity analysis and Stirling's formula for the Gamma function

$$\Gamma(z)=\Big(\frac{z}{e}\Big)^z\frac{\sqrt{2\pi}}{\sqrt{z}}\Big(1+\frac{1}{12z}+\frac{1}{288z^2}+\mathcal{O}(\frac{1}{z^3})\Big), \quad (5.169)$$

to get an asymptotic expression for $n\to\infty$. This leads directly to the stated results.

5.6.8 Deriving the factorial moments

We evaluate (5.5) using (5.160) which leads for Cases A, B and C to the desired results. We derive exemplary $M'_s(u)$ for Case A and Case C. We get for Case A

$$M'_s(u)=E_vD_v^s\frac{\partial}{\partial u}M(u,v)=E_vD_v^s\frac{\varphi_0}{1-c_1u}\exp\Big(\frac{c_1T_a}{\varphi_0a!}u^a(v-1)\Big)=\frac{\varphi_0\big(\frac{u^ac_1T_a}{\varphi_0a!}\big)^s}{1-c_1u}=\frac{\varphi_0u^{as}\big(\frac{c_1^a}{a}\big)^s}{1-c_1u}, \quad (5.170)$$

and Case C

$$M_s'(u) = \frac{\varphi_0(\frac{c_1}{c_2}+1)^{\underline{s}}\Big(\frac{c_2T_a}{\varphi_0 a!}\Big)^s u^{as}}{(1-c_1u)^{1-\frac{c_2}{c_1}(s-1)}} = \frac{\varphi_0\Gamma(s-1-\frac{c_1}{c_2})\Big(\frac{-c_2c_1^{a-1}\binom{a-1+\frac{c_2}{c_1}}{a-1}}{a}\Big)^s u^{as}}{\Gamma(-1-\frac{c_1}{c_2})(1-c_1u)^{1-\frac{c_2}{c_1}(s-1)}}. \tag{5.171}$$

5.6.9 Joint distributions

We observe

$$\begin{aligned} T_n\mathbb{P}\{X_{n,1,1}=m_1,\dots,X_{n,1,n-1}=m_{n-1}\} &= \varphi_{\sum_{i=1}^{n-1}m_i}\binom{n-1}{\underbrace{1,\dots,1}_{m_1},\underbrace{2,\dots,2}_{m_2},\dots,m_{n-1}}\times \\ &\times\binom{\sum_{i=1}^{n-1}m_i}{m_1,\dots,m_{n-1}}\prod_{k=1}^{n-1}T_k^{m_k}. \end{aligned} \tag{5.172}$$

The factor $\varphi_{\sum_{i=1}^{n-1}m_i}$ corresponds to the root degree, the factor $\binom{n-1}{\underbrace{1,\dots,1}_{m_1},\underbrace{2,\dots,2}_{m_2},\dots,m_{n-1}}$ to the choices for the labels and the factor $\binom{\sum_{i=1}^{n-1}m_i}{m_1,\dots,m_{n-1}}$ to the different positions of the subtrees. By using (10) we get the desired result.

5.6.10 Limit distribution results

The results for Case A simply follows by an application of Stirling's formula:

$$\mathbb{P}\{X_{n,j,a}=m\} = \frac{1}{a^m a!\binom{n-1}{j-1}}\sum_{l=0}^{\lfloor\frac{n-j-am}{a}\rfloor}\frac{(-1)^l}{a^l l!}\binom{n-1-a(m+l)}{j-1} \sim \frac{1}{a^m a!}\sum_{l=0}^{\lfloor\frac{n-j-am}{a}\rfloor}\frac{(-1)^l}{a^l l!}\Big(1+\mathcal{O}(\frac{1}{n}\Big). \tag{5.173}$$

For $j=\rho n$ we get an additional factor $(1-\rho)^{m+l}$.
The Case B is proved by application of Stirling's formula to either the probabilities $\mathbb{P}\{X_{n,j,a}=m\}$ or the factorial moments $\mathbb{E}(X_{n,j,a}^{\underline{s}})$. An asymptotic expansion of the factorial moments leads to an asymptotic expansion of the ordinary moments by using the Stirling numbers of the second kind $\left\{ {n \atop k} \right\}$.

$$\mathbb{E}\big(X_{n,j}^s\big) = \mathbb{E}\big(X_{n,j}^{\underline{s}}\big) + \sum_{k=1}^{s-1}\left\{ {s \atop k} \right\}\mathbb{E}\big(X_{n,j}^{\underline{k}}\big). \tag{5.174}$$

For the degenerate cases we use the Method of moments. For Case C we simplify the factorial moments.

$$\mathbb{E}\big(X_{n,j}^{\underline{s}}\big) = \frac{(-c_2)^s\binom{a-1+\frac{c_2}{c_1}}{a-1}^s}{c_1^s a^s}\frac{\Gamma(j+\frac{c_2}{c_1})(n-j)!\Gamma(n-as-\frac{c_2}{c_1}(s-1))}{\Gamma(j-\frac{c_2}{c_1}(s-1))(n-j-as)!\Gamma(n+\frac{c_2}{c_1})}. \tag{5.175}$$

For fixed a and fixed j application of Stirling's formula to (5.175) leads by using (5.174) to

$$\mathbb{E}\big(X_{n,j}^s\big) = \frac{\Gamma(j+\frac{c_2}{c_1})\Gamma(s-1-\frac{c_1}{c_2})}{\Gamma(j-\frac{c_2}{c_1}(s-1))\Gamma(-1-\frac{c_1}{c_2})}\Big(\frac{(-c_2)\binom{a-1+\frac{c_2}{c_1}}{a-1}}{c_1 a}\Big)^s\big(n^{-\frac{c_2}{c_1}}\big)^s(1+\mathcal{O}(\frac{1}{n^{-\frac{c_2}{c_1}}})). \tag{5.176}$$

For $j: j\to\infty$, $j=o(n)$ one gets similarly

$$\mathbb{E}\big(X_{n,j}^s\big) = \frac{\Gamma(s-1-\frac{c_1}{c_2})}{\Gamma(-1-\frac{c_1}{c_2})}\Big(\frac{(-c_2)\binom{a-1+\frac{c_2}{c_1}}{a-1}}{c_1 a}\Big)^s\big(\frac{n}{j}\big)^{-s\frac{c_2}{c_1}}(1+\mathcal{O}((\frac{j}{n})^{-\frac{c_2}{c_1}})). \tag{5.177}$$

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

$$\mathbb{E}\big(X_{n,j}^{\underline{s}}\big) = \frac{\Gamma(s-1-\frac{c_1}{c_2})}{\Gamma(-1-\frac{c_1}{c_2})}\Big(\frac{(-c_2)\binom{a-1+\frac{c_2}{c_1}}{a-1}}{c_1 a}\Big)^s \rho^{s\frac{c_2}{c_1}}(1-\rho)^{as}(1+\mathcal{O}(\frac{1}{n})), \tag{5.178}$$

which are asymptotically the factorial moments of a negative binomial distribution $X_\rho \stackrel{(d)}{=} \text{NegBin}(r,p)$,

$$\mathbb{E}(X_\rho^{\underline{s}}) = \frac{\Gamma(r+s)}{\Gamma(r)}(\frac{1}{p}-1)^s, \tag{5.179}$$

with parameters $r = -1 - \frac{c_1}{c_2}$ and $p = (1 + (1-\rho)^a \rho^{\frac{c_2}{c_1}} \binom{a-1+\frac{c_2}{c_1}}{a-1}\frac{(-c_2)}{c_1 a})^{-1}$.

5.6.11 Results for a randomly chosen node

We use the limiting distribution results for the central region of $X_{n,j,a}$, i. e. $j = \rho n$, with $0 < \rho < 1$, to derive result for the number of size a branches $X_{n,a}^{[R]}$ of a randomly chosen node in a grown simple increasing tree. Since $X_{n,j,a} \xrightarrow{(d)} X_{\rho,a}$ we obtain $X_{n,a}^{[R]} \xrightarrow{(d)} X_a$, where the probabilities $\mathbb{P}\{X_a = m\}$ of the discrete r. v. X_a can be obtained via

$$\mathbb{P}\{X_a = m\} = \int_0^1 \mathbb{P}\{X_{\rho,a} = m\}\, d\rho. \tag{5.180}$$

We obtain for Case A and Case B closed formulæ for these integrals. We present the computations for Case A:

Case A (recursive trees):

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

5.7 Number of leaves of node j

5.7.1 Short introduction

We denote with $X_{n,j}$ the number of leaves in the subtree rooted at node j in a size n random grown simple increasing tree. We will use the following Lemma provided in [6].

Lemma 15. *$M(z,v)$ is implicitly characterized by the following equation.*

$$\int_0^M \frac{dt}{(u-1)\varphi_0 + \varphi(t)} = z, \tag{5.181}$$

or explicitly by the following differential equation.

$$M'(z,v) = \varphi\big(M(z,v)\big) - \varphi_0(1-v), \tag{5.182}$$

with $M(0,v) = 0$. This leads to explicit formulæ for $M(z,v)$.
Binary increasing trees

$$M(z,v) = \xi \frac{\xi \tan(z\xi) + 1}{\xi - \tan(z\xi)} - 1, \quad \xi := (v-1)^{1/2}, \tag{5.183}$$

Recursive trees

$$M(z,v) = \log \frac{1-v}{1 - ve^{z(1-v)}}, \tag{5.184}$$

Plane recursive trees

$$M(z,v) = \frac{C(ve^{-v}e^{z(v-1)^2}) - C(ve^{-v})}{v-1}, \quad \textit{where} \quad C(z) = \sum_{n\geq 1} n^{n-1}\frac{z^n}{n!} \tag{5.185}$$

is the Cayley function $C(z)\exp\big(C(z)\big) = z$.

5.7.2 Results

Theorem 31. *For recursive trees, binary increasing trees and plane oriented recursive trees the limiting distribution behavior of the random variable $X_{n,j}$, which counts the number of leafs in the subtree root at the node with label j in a randomly chosen size-n tree, is, for $n \to \infty$ and depending on the growth of j, characterized as follows.*

- *The region for j fixed. The normalized random variable $c \cdot X_{n,j}/n$ converges for fixed $j \geq 2$ in distribution to a random variable X, which is Beta distributed with parameter 1 and $j-1$ for recursive tree and Beta distributed with parameters 2 and $j-1$ for binary increasing trees. For plane oriented recursive trees we obtain a beta distribution with parameters $1/2$ and $j-1$,*

$$\frac{cX_{n,j}}{n} \xrightarrow{(d)} X_j, \quad X_j \overset{(d)}{=} \begin{cases} \beta(1, j-1), & \textit{recursive trees,} \\ \beta(2, j-1), & \textit{binary increasing trees,} \\ \beta(\frac{1}{2}, j-1), & \textit{plane oriented recursive trees} \end{cases} \tag{5.186}$$

  where $c = 2$ for recursive trees, $c = 3$ for binary increasing trees and $c = 3/2$ for plane oriented recursive trees.

- *The region for small j: $j \to \infty$ such that $j = o(n)$. The normalized random variable $\frac{j}{n}Xn, j$ is asymptotically Gamma-distributed $\gamma(a, \lambda)$, with parameter $a = 1$ and $\lambda = 2$ for recursive tree, parameters $a = 2$ and $\lambda = 3$ for binary increasing trees and $a = \frac{1}{2}$ and $\lambda = \frac{3}{2}$ for plane oriented* *recursive trees.*

$$\frac{jX_{n,j}}{n} \xrightarrow{(d)} X, \quad X \stackrel{(d)}{=} \begin{cases} \gamma(1,2), & \textit{recursive trees}, \\ \gamma(2,3), & \textit{binary increasing trees}, \\ \gamma(\frac{1}{2},\frac{2}{3}), & \textit{plane oriented recursive trees}. \end{cases} \tag{5.187}$$

5.7.3 Proofs for the region j fixed

First we turn to the proof of the case j fixed. We will use the Method of moments the prove these results. If X is a Beta-distributed random variable with parameters a and b, $X \stackrel{(d)}{=} \beta(a,b)$, then the s-th moment of X is given by

$$\mathbb{E}(X^s) = \prod_{k=0}^{s-1} \frac{a+k}{a+b+k} = \frac{a^{\overline{s}}}{(a+b)^{\overline{s}}}. \tag{5.188}$$

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

$$\mathbb{E}\big((X_{n,j})^s\big) = \mathbb{E}\big((X_{n,j})^{\underline{s}}\big) + \sum_{k=1}^{s-1} \left\{ {s \atop k} \right\} \mathbb{E}\big((X_{n,j})^{\underline{k}}\big). \tag{5.189}$$

In both cases we will use the following identity to obtain the factorial moments.

$$E_v D_v^s M(z,v) = s![w^s]M(z,v), \quad \text{where } w := v-1. \tag{5.190}$$

For recursive trees we have the following calculation.

$$\begin{aligned}
\mathbb{E}(X_{k+j,j}^{\underline{s}}) &= \frac{(j-1)!k!}{(k+j-1)!}[z^{j-1}u^k]E_v D_v^s N(z,u,v) = \frac{(j-1)!k!}{(k+j-1)!}[u^k]\frac{1}{(1-u)^{j-1}}E_v D_v^s M'(u,v) \\
&= \frac{(j-1)!k!s!}{(k+j-1)!}[u^k w^s]\frac{1}{(1-u)^{j-1}}M'(u,v) = \frac{(j-1)!k!s!}{(k+j-1)!}[u^k w^s]\frac{1}{(1-u)^{j-1}}\frac{v(1-v)e^{u(1-v)}}{1-ve^{u(1-v)}} \\
&= \frac{(j-1)!k!s!}{(k+j-1)!}[u^k w^s]\frac{1}{(1-u)^{j-1}}\frac{v(v-1)}{v-e^{u(v-1)}} = \frac{(j-1)!k!s!}{(k+j-1)!}[u^k w^s]\frac{1}{(1-u)^{j-1}}\frac{(w+1)w}{w+1-e^{uw}} \\
&= \frac{(j-1)!k!s!}{(k+j-1)!}[u^k w^s]\frac{1}{(1-u)^{j-1}}\frac{(w+1)w}{w(1-u)-\sum_{l\geq 2}\frac{(uw)^l}{l!}} \\
&= \frac{(j-1)!k!s!}{(k+j-1)!}[u^k w^s]\frac{1}{(1-u)^{j-1}}\frac{w+1}{(1-u)\big(1-\sum_{l\geq 2}\frac{u^l w^{l-1}}{l!(1-u)}\big)} \\
&= \frac{(j-1)!k!s!}{(k+j-1)!}[u^k]\frac{1}{(1-u)^j}\Big(\frac{1}{2^s(1-u)^s}+\mathcal{O}(\frac{1}{(1-u)^{s-1}})\Big),
\end{aligned} \tag{5.191}$$

by setting $k = n-j$ this further simplifies to

$$\begin{aligned}
\mathbb{E}(X_{n,j}^{\underline{s}}) &= \frac{(j-1)!(n-j)!s!}{2^s(n-1)!}[u^{n-j}]\Big(\frac{1}{2^s(1-u)^{s+j}}+\mathcal{O}(\frac{1}{(1-u)^{s+j-1}})\Big) = \frac{n^s(j-1)!s!}{2^s(j+s-1)!}+\mathcal{O}(n^{s-1}) \\
&= \Big(\frac{n}{2}\Big)^s\frac{1^{\overline{s}}}{j^{\overline{s}}}+\mathcal{O}(n^{s-1}),
\end{aligned} \tag{5.192}$$

which proves the first part concerning the recursive trees. For binary increasing trees we proceed as follows.

$$
\begin{aligned}
\mathbb{E}(X_{k+j,j}^{\underline{s}}) &= \frac{(j-1)!k!s!}{(k+j)!}[u^k w^s]\frac{j}{(1-u)^{j-1}}M'(u,v) = \frac{j!k!s!}{(k+j)!}[u^k w^s]\frac{w(w+1)(1+\tan^2(u\sqrt{w}))}{(1-u)^{j-1}(\sqrt{w}-\tan(u\sqrt{w}))^2}\\
&= \frac{j!k!s!}{(k+j)!}[u^k w^s]\frac{1}{(1-u)^{j-1}}\frac{(w+1)(1+\tan^2(u\sqrt{w}))}{(1-\frac{\tan(u\sqrt{w})}{\sqrt{w}})^2}\\
&= \frac{j!k!s!}{(k+j)!}[u^k w^s]\frac{1}{(1-u)^{j-1}}\frac{(w+1)(1+\tan^2(u\sqrt{w}))}{(1-u)^2(1-\sum_{n\geq 2}\frac{t_n u^{2n-1}w^{n-1}}{1-u})^2}\\
&= \frac{j!k!s!}{(k+j)!}[u^k]\frac{1}{(1-u)^{j+1}}\Big(\frac{(s+1)t_2^s}{(1-u)^s}+\mathcal{O}(\frac{1}{(1-u)^{s-1}})\Big),
\end{aligned}
\tag{5.193}
$$

with

$$
\tan z = \sum_{n\geq 1}(-1)^{n-1}\frac{2^{2n}(2^{2n}-1)B_{2n}}{(2n)!}z^{2n-1} = \sum_{n\geq 1}t_n z^{2n-1}, \tag{5.194}
$$

where B_{2n} denotes the $2n$-th Bernoulli number. Since $t_2 = 1/3$ this finishes the proof of Theorem. For the plane oriented recursive trees we proceed differently by using induction. We start with the differential equation as given in Lemma 15.

$$
M'(z,v) = \frac{1}{1-M(z,v)} + v - 1, \tag{5.195}
$$

with $M(0,v)=0$. Applying the operator $E_v D_v^s$ to this differential equation leads to

$$
\begin{aligned}
M_s'(z) &= E_v D_v^s \frac{1}{1-M(z,v)} + \delta_{s,1}\\
&= \frac{M_s(z)}{1-2z} + \sum_{k=0}^{s-1}\binom{s-1}{k}M_{k+1}E_v D_v^{s-1-k}\frac{1}{(1-M(z,v))^2} + \delta_{s,1}\\
&= \frac{M_s(z)}{1-2z} + R_s(z),
\end{aligned}
\tag{5.196}
$$

with $M_s(0)=0$. It can easily be seen that the dominant term in the singular expansions of $M_s(z)$ and $E_v D_v^s \frac{1}{(1-M(z,v))^2}$ around $z=1/2$ are given as follows

$$
M_s(z) = \frac{m_s}{(1-2z)^{\frac{2s-1}{2}}} + \mathcal{O}\Big(\frac{1}{(1-2z)^{s-2}}\Big), \quad E_v D_v^s \frac{1}{(1-M(z,v))^2} = \frac{f_{s,2}}{(1-2z)^{s+1}} + \mathcal{O}\Big(\frac{1}{(1-2z)^{s}}\Big), \tag{5.197}
$$

We will show that $m_s = \frac{2^{s-1}}{3^s}\big(\frac{1}{2}\big)^{\overline{s-1}}$, which is equivalent to

$$
\mathbb{E}(X_{n,j}^{\underline{s}}) = \Big(\frac{2n}{3}\Big)^s \frac{\big(\frac{1}{2}\big)^{\overline{s}}}{(j-\frac{1}{2})^{\overline{s}}} + \mathcal{O}(n^{s-1}). \tag{5.198}
$$

It holds that $f_{s,1}$, which is the leading coefficient of the singular expansion of $E_v D_v^s 1/(1-M(z,v))$, is given by $(2s-1)m_s$ due to (5.196) and (5.197). The equation

$$
E_v D_v^s \frac{1}{(1-M(z,v))^2} = \sum_{k=0}^{s}\binom{s}{k}E_v D_v^k \frac{1}{1-M(z,v)}E_v D_v^{s-k}\frac{1}{1-M(z,v)} \tag{5.199}
$$

translates into the following equation concerning $f_{s,1}$ and $f_{s,2}$

$$f_{s,2} = \sum_{k=0}^{s} \binom{s}{k} f_{k,1} f_{s-k,1} = \sum_{k=0}^{s} \binom{s}{k} (2k-1) m_{k,1} (2s-2k-1) m_{s-k,1}. \tag{5.200}$$

Since the solution of (5.196) is given by

$$M_s(z) = \frac{\int_0^z R_s(t)\sqrt{1-2t}dt}{\sqrt{1-2z}}, \tag{5.201}$$

we obtain the recurrence

$$m_s = \frac{1}{2(s-1)} \sum_{l=0}^{s-2} \binom{s-1}{l} m_{l+1} \sum_{k=0}^{s-l} \binom{s-l}{k} (2k-1) m_{k,1} (2s-2k-2l-1) m_{s-l-k,1}, \tag{5.202}$$

with initial value $m_1 = \frac{1}{3}$. Assuming m_s has the suggested shape for all $k < s$, we easily get that $f_{s,2} = s!\frac{2^s}{3^s}$ and consequently by (5.202) the induction step.

5.7.4 Proofs for the region $j \to \infty$ such that $j = o(n)$

For the region j: $j \to \infty$ such that $j = o(n)$ we consider the normalized random variable $jX_{n,j}/n$ and will show via the method of moments that $jX_{n,j}/n \xrightarrow{(d)} \gamma(1,c)$, where $\gamma(a,\lambda)$ denotes the Gamma-distribution with shape parameter a and scale parameter λ, and c is the same constant as for in the case j fixed. If X is a Gamma-distributed random variable, $X \stackrel{(d)}{=} \gamma(a,\lambda)$, then the s-th moment of X is given by

$$\mathbb{E}(X^s) = \frac{1}{\lambda^s} \prod_{k=0}^{s-1} (a+k) = \frac{a^{\overline{s}}}{\lambda^s}. \tag{5.203}$$

From (5.191) we obtain

$$\mathbb{E}(X_{n,j}^{\underline{s}}) = \frac{(j-1)!(n-j)!s!}{2^s(n-1)!} [u^{n-j}] \Big(\frac{1}{2^s(1-u)^{s+j}} + \sum_{l=0}^{s+j-1} \frac{K_l}{(1-u)^l} \Big), \tag{5.204}$$

where K_l are certain constants which do not depend explicitly on n. Thus we get further

$$\begin{aligned} \mathbb{E}(X_{n,j}^{\underline{s}}) &= \frac{(j-1)!(n-j)!s!}{2^s(n-1)!} \Big(\frac{\binom{n+s-1}{n-j}}{2^s} + \sum_{l=0}^{s+j-1} K_l \binom{n-j+l-1}{n-j} \Big) \\ &= \frac{s!(n+s-1)^{\underline{s}}}{2^s(j+s-1)^{\underline{s}}} \Big(1 + \mathcal{O}(\frac{1}{j}) + \mathcal{O}(\frac{j}{n}) \Big) = \frac{s!n^s}{2^s j^s} \Big(1 + \mathcal{O}(\frac{1}{j}) + \mathcal{O}(\frac{j}{n}) \Big) \end{aligned} \tag{5.205}$$

which finishes the proof for recursive trees. For binary increasing trees we proceed as before which leads to

$$\mathbb{E}(X_{n,j}^{\underline{s}}) = \frac{(s+1)!n^s}{3^s j^s} \Big(1 + \mathcal{O}(\frac{1}{j}) + \mathcal{O}(\frac{j}{n}) \Big). \tag{5.206}$$

For plane oriented recursive trees we proceed analogously.

5.8 Subtrees on the fringe of plane oriented recursive trees

5.8.1 Short introduction

We denote with $X_{n,a}$ the number of size a trees on the fringe of a plane oriented increasing tree. Note that for $a = 1$ the r. v. $X_{n,a}$ just counts the number of leaves. This random variable was studied in [20] by an analytic approach. For grown simple families of increasing trees we can set up the following recurrence.

$$\begin{aligned}\mathbb{P}\{X_{n,a} = m\} = \sum_{r\ge1}\varphi_r \sum_{\substack{n_1+\dots+n_r=n-1,\\ n_1,\dots,n_r\ge1}} \frac{T_{n_1}\cdots T_{n_r}}{T_n}\binom{n-1}{n_1,n_2,\dots,n_r}\times \\ \times \sum_{\substack{m_1+\dots+m_r=m,\\ m_1,\dots,m_r\ge1}} \mathbb{P}\{X_{n_1,a}=m_1\}\cdots\mathbb{P}\{X_{n_r,a}=m_r\}, \quad \text{for } n > a,\end{aligned} \tag{5.207}$$

with initial values $\mathbb{P}\{X_{a,a} = 1\} = 1$ and $\mathbb{P}\{X_{n,a} = 0\} = 1$ for $n < a$. Introducing the bivariate generating function

$$M_a(z,v) = \sum_{n\ge1} T_n \mathbb{E}(v^{X_{n,a}})\frac{z^n}{n!} = \sum_{n\ge1}\sum_{m\ge1} T_n\mathbb{P}\{X_{n,a}=m\}\frac{z^n}{n!}v^m, \tag{5.208}$$

recurrence (5.207) can be translated into

$$\frac{\partial}{\partial z}M_a(z,v) = \varphi\big(M_a(z,v)\big) + (v-1)\frac{T_k}{(k-1)!}z^{a-1}, \tag{5.209}$$

with initial condition $M_a(0,v) = 0$.

5.8.2 Results

Theorem 32. *In a random plane oriented recursive tree of size n the number $X_{n,a}$ of subtrees of size a for fixed a satisfies*

$$\begin{aligned}\mathbb{E}(X_{n,a}) &= \frac{2n-1}{(2a-1)(2a+1)}, \\ \mathbb{V}(X_{n,a}) &= n\Big(\frac{8a^2-4a-8}{(2a-1)^2(2a+1)^2} - \frac{\big((2a-3)!!\big)^2}{\big((a-1)!\big)^2 2^{2a-2}a(2a+1)}\Big) - \frac{4a^2-2k-2}{(2a-1)^2(2a+1)^2} \\ &\quad + \frac{\big((2a-3)!!\big)^2}{\big((a-1)!\big)^2 2^{2a-1}a(2a+1)} + \mathcal{O}\Big(\frac{1}{n^{\frac52}}\Big).\end{aligned} \tag{5.210}$$

Theorem 33. *The random variable $X_{n,a}$ satisfies for fixed a and $n \to \infty$ the limit law*

$$\frac{X_{n,a}-\mu n}{\sqrt{\sigma^2 n}} = \frac{X^*_{n,a}}{\sqrt{\sigma^2 n}} \xrightarrow{(d)} \mathcal{N}(0,1), \tag{5.211}$$

where $\mu n = \frac{2n}{(2a-1)(2a+1)} \sim \mathbb{E}(X_{n,a})$ *and* $\sigma^2 n = n\Big(\frac{8a^2-4k-8}{(2a-1)^2(2a+1)^2} - \frac{\big((2a-3)!!\big)^2}{\big((a-1)!\big)^2 2^{2a-2}a(2a+1)}\Big) \sim \mathbb{V}(X_{n,a})$.

5.8.3 Deriving the expectation and the variance

We start with the differential equation obtained from (5.209) using $\varphi(t) = 1/(1-t)$

$$\frac{\partial}{\partial z}M_a(z,v) = \frac{1}{1-M_a(z,v)} + (v-1)z^{a-1}\frac{T_a}{(a-1)!}, \tag{5.212}$$

with $M_a(0,v)=0$. We will use the abbreviation $M_a^{[s]}(z) = E_v D_v^s M_a(z,v)$. Applying the operator $E_v D_v^s$ to this differential equation leads to

$$\begin{aligned}\frac{d}{dz}M_a^{[s]}(z) &= E_v D_v^s \frac{1}{1-M_a(z,v)} + \delta_{s,1} z^{a-1}\frac{T_a}{(a-1)!} = E_v D_v^{s-1}\frac{\frac{\partial}{\partial v}M_a(z,v)}{(1-M_a(z,v))^2} + \delta_{s,1} z^{a-1}\frac{T_a}{(a-1)!}\\ &= \frac{M_a^{[s]}(z)}{1-2z} + \sum_{l=0}^{s-2}\binom{s-1}{l} M_a^{[l+1]}(z) E_v D_v^{s-1-l}\frac{1}{(1-M_a(z,v))^2} + \delta_{s,1} z^{k-1}\frac{T_k}{(a-1)!}\\ &= \frac{M_a^{[s]}(z)}{1-2z} + R_s(z),\end{aligned} \tag{5.213}$$

with $M_a^{[s]}(0)=0$. The solution of (5.213) is given by

$$M_a^{[s]}(z) = \frac{\int_0^z R_s(t)\sqrt{1-2t}dt}{\sqrt{1-2z}}. \tag{5.214}$$

At first we will calculate the expectation and the variance. For the expectation we get

$$\frac{d}{dz}M_a^{[1]}(z) = \frac{M_a^{[1]}(z)}{1-2z} + z^{a-1}\frac{T_a}{(a-1)!}, \tag{5.215}$$

which leads to the solution

$$M_a^{[1]}(z) = \frac{T_a\int_0^z \sqrt{1-2t}\, t^{a-1}dt}{(a-1)!\sqrt{1-2z}}. \tag{5.216}$$

Since we obtain by partial integration

$$\begin{aligned}\int_0^z \sqrt{1-2t}\, t^N dt &= \frac{N!(N+2)!2^{N+2}}{(2N+4)!} - \sum_{l=0}^{N}\frac{(N+2-l)!2^{N+2-l}}{(2N+4-2l)!} N^{\underline{N-l}} z^l (1-2z)^{\frac{2(N-l)+3}{2}}\\ &= \frac{N!}{(2N+3)!!} - \sum_{l=0}^{N}\frac{N^{\underline{N-l}}}{(2N-2l+3)!!} z^l (1-2z)^{\frac{2(N-l)+3}{2}},\end{aligned} \tag{5.217}$$

or by expanding around $t=1/2$,

$$\begin{aligned}\int_0^z \sqrt{1-2t}\, t^N dt &= \frac{1}{2^N}\sum_{l=0}^{N}\binom{N}{l}(-1)^l \int_0^z (1-2t)^{l+\frac12}dt\\ &= \frac{1}{2^N}\sum_{l=0}^{N}\binom{N}{l}\frac{(-1)^l}{2l+3} - \frac{1}{2^N}\sum_{l=0}^{N}\binom{N}{l}(-1)^l\frac{(1-2z)^{l+\frac32}}{2l+3},\end{aligned} \tag{5.218}$$

where

$$\begin{aligned}\frac{1}{2^N}\sum_{l=0}^{N}\binom{N}{l}\frac{(-1)^l}{2l+3} &= \frac{(-1)^N}{2^{N+1}}\sum_{l=0}^{N}\binom{N}{l}\frac{(-1)^{N-l}}{l+\frac32} = \frac{(-1)^N}{2^{N+1}}\Big(\sum_{l=0}^{N}\binom{N}{l}(-1)^{N-l}\mathrm{E}^l\frac{1}{x}\Big)\Big|_{x=\frac32}\\ &= \frac{(-1)^N}{2^{N+1}}\Big(\Delta^N\frac{1}{x}\Big)\Big|_{x=\frac32} = \frac{N!}{2^{N+1}\left(\frac32\right)^{\overline{N+1}}} = \frac{N!}{(2N+3)!!},\end{aligned} \tag{5.219}$$

and Δ denotes the ordinary difference operator, we can refine our previous result by setting $N = a - 1$:

$$M_a^{[1]}(z) = \frac{T_a}{(2a+1)!!\sqrt{1-2z}} - \frac{1}{2^{a-1}} \sum_{l=0}^{a-1} \binom{a-1}{l} (-1)^l \frac{(1-2z)^{l+1}}{2l+3}. \tag{5.220}$$

This leads to

$$\begin{aligned} \mathbb{E}(X_{n,a}) &= \frac{n!}{T_n}[z^n] \frac{1}{(2a-1)(2a+1)\sqrt{1-2z}} = \frac{n!}{(2n-3)!!(2a-1)(2a+1)} \binom{-\frac{1}{2}}{n} (-2)^n \\ &= \frac{2n-1}{(2a-1)(2a+1)}. \end{aligned} \tag{5.221}$$

For the variance we get

$$\frac{d}{dz} M_a^{[2]}(z) = \frac{M_a^{[2]}(z)}{1-2z} + \frac{2\big(M_a^{[1]}(z)\big)^2}{(1-2z)^{\frac{3}{2}}}, \tag{5.222}$$

which leads to the solution

$$M_a^{[2]}(z) = \frac{2\int_0^z \frac{\left(M_a^{[1]}(t)\right)^2}{1-2t}\, dt}{\sqrt{1-2z}}. \tag{5.223}$$

We will use the expression

$$\big(M_a^{[1]}(t)\big)^2 = \frac{T_a^2}{\big((a-1)!\big)^2 (1-2t)} \Big(\int_0^t \sqrt{1-2x}\, x^{a-1} dt \Big)^2, \tag{5.224}$$

with

$$\begin{aligned} \Big(\int_0^t \sqrt{1-2x}\, x^{a-1} dt \Big)^2 &= \big(\frac{(a-1)!}{(2a+1)!!}\big)^2 - \frac{2(a-1)!}{2^{a-1}(2a+1)!!} \sum_{l=0}^{a-1} \binom{a-1}{l} (-1)^l \frac{(1-2t)^{l+\frac{3}{2}}}{2l+3} \\ &\quad + \frac{1}{2^{2a-2}} \sum_{l=0}^{a-1} \binom{a-1}{l} (-1)^l \sum_{i=0}^{a-1} \binom{a-1}{i} (-1)^i \frac{(1-2t)^{i+l+3}}{(2i+3)(2l+3)}, \end{aligned} \tag{5.225}$$

which leads to

$$\begin{aligned} \frac{\big(M_a^{[1]}(t)\big)^2}{1-2t} &= \frac{T_a^2}{\big((a-1)!\big)^2} \Big(\big(\frac{(a-1)!}{(2a+1)!!}\big)^2 \frac{1}{(1-2t)^2} - \frac{2(a-1)!}{2^{a-1}(2a+1)!!} \sum_{l=0}^{a-1} \binom{a-1}{l} (-1)^l \frac{(1-2t)^{l-\frac{1}{2}}}{2l+3} \\ &\quad + \frac{1}{2^{2a-2}} \sum_{l=0}^{a-1} \binom{a-1}{l} (-1)^l \frac{1}{2l+3} \sum_{i=0}^{a-1} \binom{a-1}{i} (-1)^i \frac{(1-2t)^{i+l+1}}{2i+3}, \Big). \end{aligned} \tag{5.226}$$

This leads to an expansion of $M_a^{[2]}(z)$ around $z = 1/2$:

$$\begin{aligned}
M_a^{[2]}(z) &= 2\frac{T_a^2}{\big((a-1)!\big)^2}\Bigg[\Big(\frac{(a-1)!}{(2a+1)!!}\Big)^2\frac{1}{2(1-2z)^{\frac{3}{2}}} + \frac{2(a-1)!}{2^{a-1}(2a+1)!!}\sum_{l=0}^{a-1}\binom{a-1}{l}(-1)^l\frac{(1-2z)^l}{(2l+1)(2l+3)}\\
&\quad - \frac{1}{2^{2a-2}}\sum_{l=0}^{a-1}\binom{a-1}{l}(-1)^l\sum_{i=0}^{a-1}\binom{a-1}{i}(-1)^i\frac{(1-2z)^{i+l+\frac{3}{2}}}{(2i+3)(2l+3)(2i+2l+4)}\Bigg]\\
&\quad + 2\frac{T_a^2}{\big((a-1)!\big)^2\sqrt{1-2z}}\Bigg[-\Big(\frac{(a-1)!}{(2a+1)!!}\Big)^2\frac{1}{2} - \frac{2(a-1)!}{2^{a-1}(2a+1)!!}\sum_{l=0}^{a-1}\binom{a-1}{l}\frac{(-1)^l}{(2l+1)(2l+3)}\\
&\quad + \frac{1}{2^{2a-2}}\sum_{l=0}^{a-1}\binom{a-1}{l}(-1)^l\sum_{i=0}^{a-1}\binom{a-1}{i}\frac{(-1)^i}{(2i+3)(2l+3)(2i+2l+4)}\Bigg].
\end{aligned} \tag{5.227}$$

Now we use partial fraction decomposition to obtain a simplified form of the coefficient of $1/\sqrt{1-2z}$:

$$\begin{aligned}
&-\frac{1}{(2a+1)(2a-1)^2} - \frac{T_a^2}{\big((a-1)!\big)^2 2^{2a-2}}\sum_{l=0}^{a-1}\binom{a-1}{l}\frac{(-1)^l}{(2l+1)(2l+3)(l+2)\binom{l+2+a}{a}}\\
&\qquad = -\frac{1}{(2a+1)(2a-1)^2} - \frac{\big((2a-3)!!\big)^2}{\big((a-1)!\big)^2 2^{2a-1}k(2a+1)}.
\end{aligned} \tag{5.228}$$

Now we get by extracting coefficients the result.

$$\mathbb{E}(X_{n,a}^2) = \frac{(2n+1)(2n-1)}{(2a-1)^2(2a+1)^2} - \frac{2n-1}{(2a+1)(2a-1)^2} - \frac{(2n-1)\big((2a-3)!!\big)^2}{\big((a-1)!\big)^2 2^{2a-1}k(2a+1)} + \mathcal{O}\Big(\frac{1}{n^{\frac{5}{2}}}\Big). \tag{5.229}$$

The variance can now be obtained from

$$\begin{aligned}
\mathbb{V}(X_{n,a}) &= \mathbb{E}(X_{n,a}^2) + \mathbb{E}(X_{n,a}) - \mathbb{E}(X_{n,a})^2 = \mathbb{E}(X_{n,a}^2) + \frac{2n-1}{(2a-1)(2a+1)} - \frac{4n^2-4n+1}{(2a-1)^2(2a+1)^2}\\
&= n\Big(\frac{8a^2-4ak-8}{(2a-1)^2(2a+1)^2} - \frac{\big((2a-3)!!\big)^2}{\big((a-1)!\big)^2 2^{2a-2}k(2a+1)}\Big) - \frac{4a^2-2a-2}{(2a-1)^2(2a+1)^2}\\
&\quad + \frac{\big((2a-3)!!\big)^2}{\big((a-1)!\big)^2 2^{2a-1}a(2a+1)} + \mathcal{O}\Big(\frac{1}{n^{\frac{5}{2}}}\Big).
\end{aligned} \tag{5.230}$$

Combining (5.221) and (5.230) proves Theorem 32.

5.9 Higher moments

It can easily be seen that the dominant term in the singular expansions of $M_a^{[s]}(z)$ and $E_v D_v^s \frac{1}{(1-M_a(z,v))^2}$ around $z = 1/2$ are given as follows

$$M_a^{[s]}(z) = \frac{m_s}{(1-2z)^{\frac{2s-1}{2}}} + \mathcal{O}\Big(\frac{1}{(1-2z)^{s-2}}\Big), \quad E_v D_v^s \frac{1}{(1-M_a(z,v))^2} = \frac{f_s}{(1-2z)^{s+1}} + \mathcal{O}\Big(\frac{1}{(1-2z)^s}\Big), \tag{5.231}$$

Proposition 3. *The coefficient of the leading term in the asymptotic expansion of $M_a^{[s]}(z)$ around $z = 1/2$*

is given as follows.

$$m_s = \frac{2^{s-1}}{(2a-1)^s(2a+1)^s}\left(\frac{1}{2}\right)^{\overline{s-1}}. \tag{5.232}$$

Proof. It holds that the leading coefficient e_s of the singular expansion of $E_v D_v^s 1/(1-M_a(z,v))$ is given by $e_s = (2s-1)m_s$ due to (5.213) and (5.231). The equation

$$E_v D_v^s \frac{1}{(1-M_a(z,v))^2} = \sum_{l=0}^{s} \binom{s}{l} E_v D_v^l \frac{1}{1-M_a(z,v)} E_v D_v^{s-l} \frac{1}{1-M_a(z,v)} \tag{5.233}$$

translates into the following equation concerning f_s

$$f_s = \sum_{l=0}^{s} \binom{s}{l} e_l e_{s-l} = \sum_{l=0}^{s} \binom{s}{l} (2l-1)m_l(2s-2l-1)m_{s-l}. \tag{5.234}$$

Since the solution of (5.213) is given by

$$M_a^{[s]}(z) = \frac{\int_0^z R_s(t)\sqrt{1-2t}dt}{\sqrt{1-2z}}, \tag{5.235}$$

we obtain the recurrence

$$m_s = \frac{1}{2(s-1)} \sum_{l=0}^{s-2} \binom{s-1}{l} m_{l+1} \sum_{i=0}^{s-l-1} \binom{s-l-1}{i} (2i-1)m_i(2(s-l-1-i)-1)m_{s-l-1-i}, \tag{5.236}$$

with initial value $m_1 = \frac{1}{(2a-1)(2a+1)}$. Assuming m_s has the suggested shape for all $i < s$, we easily get

$$\begin{aligned} f_s &= \frac{2^s}{(2a-1)^s(2a+1)^s} \sum_{l=0}^{s} \binom{s}{l} \left(\frac{1}{2}\right)^{\overline{l}} \left(\frac{1}{2}\right)^{\overline{s-l}} = \frac{2^s}{(2a-1)^s(2a+1)^s} s! \sum_{k=0}^{s} [z^k]\frac{1}{\sqrt{1-z}}[z^{s-l}]\frac{1}{\sqrt{1-z}} \\ &= \frac{2^s}{(2a-1)^s(2a+1)^s} s![z^s]\frac{1}{1-z} = s!\frac{2^s}{(2a-1)^s(2a+1)^s}, \end{aligned} \tag{5.237}$$

and consequently by (5.236)

$$\begin{aligned} m_s &= \frac{2^{s-2}}{(2k-1)^s(2k+1)^s(s-1)} \sum_{l=0}^{s-2} \binom{s-1}{l} \left(\frac{1}{2}\right)^{\overline{l}} (s-1-l)! = \frac{2^{s-2}}{(2k-1)^s(2k+1)^s}(s-2)! \sum_{l=0}^{s-2} \frac{(\frac{1}{2})^{\overline{l}}}{l!} \\ &= \frac{2^{s-2}}{(2k-1)^s(2k+1)^s}(s-2)![z^{s-2}]\frac{1}{(1-z)^{\frac{3}{2}}} = \frac{2^{s-2}}{(2k-1)^s(2k+1)^s}\left(\frac{3}{2}\right)^{\overline{s-2}} \\ &= \frac{2^{s-1}}{(2k-1)^s(2k+1)^s}\left(\frac{1}{2}\right)^{\overline{s-1}}, \end{aligned} \tag{5.238}$$

the induction step. $\square$

Introducing the random variable $X_{n,j,a}$, which counts the number of fixed size a trees in the subtree rooted at the node with label j in a randomly chosen size-n tree, we can deduce immediately that

$$\mathbb{E}(X_{n,j,a}^{\underline{s}}) = \left(\frac{2n}{(2a-1)(2a+1)}\right)^s \frac{(\frac{1}{2})^{\overline{s}}}{(j-\frac{1}{2})^{\overline{s}}} + \mathcal{O}(n^{s-1}), \tag{5.239}$$

by applying the approach for label-based parameters. The expansion (5.239) leads to the following corollary.

Corollar 8. *For plane oriented recursive trees the limiting distribution behavior of the random variable $X_{n,j,a}$, which counts the number of fixed size k trees in the subtree rooted at the node with label j in a randomly chosen size-n tree, is, for $n \to \infty$ and depending on the growth of j, characterized as follows*

- *The region j fixed:*

$$\frac{(2a-1)(2a+1)X_{n,j,a}}{2n} \xrightarrow{(d)} X_{j,a} \quad X_{j,a} \overset{(d)}{=} \beta(\frac{1}{2}, j-1), \tag{5.240}$$

- *The region for small j: $j \to \infty$ such that $j = o(n)$. The normalized random variable $\frac{j}{n}X_{n,j,a}$ is asymptotically Gamma-distributed $\gamma(a, \lambda)$*

$$\frac{jX_{n,j,a}}{n} \xrightarrow{(d)} X_a, \quad X_a \overset{(d)}{=} \gamma(\frac{1}{2}, \frac{2}{(2a-1)(2a+1)}) \tag{5.241}$$

5.9.1 Shifting the mean

To get the limit law of the normalized and centralized random variable $X_{n,a}$, we shift the random variable $X_{n,a}$ by its mean $X^*_{n,a} = X_{n,a} - \mathbb{E}(X_{n,a}) = X_{n,a} - \mu n$, where $\mu = \frac{2}{(2a-1)(2a+1)}$. Introducing the bivariate generating function $N_a(z,v) = \sum_{n\geq 1} T_n \mathbb{E}(v^{X^*_{n,a}})\frac{z^n}{n!} = M_k(zv^{-\mu}, v)$ leads to the differential equation

$$\frac{\partial}{\partial z}N_a(z,v) = \frac{v^{-\mu}}{1-N_a(z,v)} + \frac{T_a}{(a-1)!}z^{a-1}v^{-a\mu}, \tag{5.242}$$

with $N_a(0,v) = 0$. Now we extract higher moments where we will use the abbreviation $N_a^{[s]}(z) = E_v D_v^s N_a(z,v)$. Differentiating (5.242) s times with respect to v and evaluation at $v = 1$ leads to

$$\begin{aligned} \frac{d}{dz}N_a^{[s]}(z) &= \frac{N_a^{[s]}(z)}{1-2z} + \sum_{i=0}^{s-1}\binom{s}{i}(-\mu)^{\underline{s-i}}E_v D_v^i \frac{1}{1-N_a(z,v)} \\ &\quad + \sum_{l=0}^{s-2}\binom{s-1}{l}N_a^{[l+1]}(z)E_v D_v^{s-1-l}\frac{1}{(1-N_a(z,v))^2} + s(a\mu)^{\underline{s-1}}\frac{T_a}{a!}z^{a-1} = \frac{N_a^{[s]}(z)}{1-2z} + R_s(z), \end{aligned} \tag{5.243}$$

with $N_a^{[s]}(0) = 0$. The solution of (5.243) is given by

$$N_a^{[s]}(z) = \frac{\int_0^z R_s(t)\sqrt{1-2t}dt}{\sqrt{1-2z}}. \tag{5.244}$$

In order to prove the gaussian limit law with the method of moments we have to show the expansions

$$\begin{aligned} N_a^{[2s]}(z) &= \frac{(2s)!\sigma^{2s}\Gamma(\frac{2s-1}{2})}{2^{s+1}s!\Gamma(\frac{1}{2})(1-2z)^{\frac{2s-1}{2}}} + \mathcal{O}\big(\frac{1}{(1-2z)^{s-1}}\big), \\ N_a^{[2s+1]}(z) &= \mathcal{O}\big(\frac{1}{(1-2z)^{\frac{2s-1}{2}}}\big), \end{aligned} \tag{5.245}$$

which are equivalent to

$$\mathbb{E}\big((X^*_{n,a})^{2s}\big) = \frac{(2s)!n^s\sigma^{2s}}{2^s s!} + \mathcal{O}(n^{s-\frac{1}{2}}), \quad \mathbb{E}\big((X^*_{n,a})^{2s+1}\big) = \mathcal{O}(n^s). \tag{5.246}$$

We will use induction to prove the expansions stated in (5.245). We need the auxiliary expansions

$$E_v D_v^{2s} \frac{1}{1-N_a(z,v)} = \frac{(2s)!\sigma^{2s}\Gamma(\frac{2s+1}{2})}{2^s s!\Gamma(\frac{1}{2})(1-2z)^{\frac{2s+1}{2}}} + \mathcal{O}\Big(\frac{1}{(1-2z)^s}\Big), \quad E_v D_v^{2s+1} \frac{1}{1-N_a(z,v)} = \mathcal{O}\Big(\frac{1}{(1-2z)^{\frac{2s+1}{2}}}\Big). \tag{5.247}$$

The expansions (5.247) lead to

$$E_v D_v^{2s} \frac{1}{(1-N_a(z,v))^2} = \sum_{i=0}^{2s} \binom{2s}{i} E_v D_v^{i} \frac{1}{1-N_a(z,v)} E_v D_v^{2s-i} \frac{1}{1-N_a(z,v)}, \tag{5.248}$$

where we split the sum into two parts.

$$\begin{aligned} E_v D_v^{2s} \frac{1}{(1-N_a(z,v))^2} &= \sum_{i=0}^{s} \binom{2s}{2i} E_v D_v^{2i} \frac{1}{1-N_a(z,v)} E_v D_v^{2(s-i)} \frac{1}{1-N_a(z,v)} \\ &\quad + \sum_{i=0}^{s-1} \binom{2s}{2i+1} E_v D_v^{2i+1} \frac{1}{1-N_a(z,v)} E_v D_v^{2s-2i+1} \frac{1}{1-N_a(z,v)} \end{aligned} \tag{5.249}$$

By proceeding as in (5.237) we get further

$$\begin{aligned} E_v D_v^{2s} \frac{1}{(1-N_a(z,v))^2} &= \frac{1}{(1-2z)^{s+1}} \sum_{i=0}^{s} \binom{2s}{2i} \frac{(2i)!\sigma^{2i}\Gamma(\frac{2i+1}{2})}{2^i i!\Gamma(\frac{1}{2})} \frac{(2s-2i)!\sigma^{2s-2i}\Gamma(\frac{2s-2i+1}{2})}{2^{s-i}(s-i)!\Gamma(\frac{1}{2})} \\ &\quad + \mathcal{O}\Big(\frac{1}{(1-2z)^{\frac{2s+1}{2}}}\Big) = \frac{(2s)!}{2^s(1-2z)^{s+1}} + \mathcal{O}\Big(\frac{1}{(1-2z)^{\frac{2s+1}{2}}}\Big). \end{aligned} \tag{5.250}$$

It also holds

$$E_v D_v^{2s+1} \frac{1}{(1-N_a(z,v))^2} = \mathcal{O}\Big(\frac{1}{(1-2z)^{\frac{2s+1}{2}}}\Big). \tag{5.251}$$

We already know that (5.245) holds for $N_a^{[1]}(z)$ and $N_a^{[2]}(z)$. Assuming that for all $1 \le a \le 2s-1$ the expansions (5.245) hold, it also holds (5.247) and (5.248) in the required range, and we get according to (5.243) the result

$$\begin{aligned} \frac{d}{dz} N_a^{[2s]}(z) &= \frac{N_a^{[2s]}(z)}{1-2z} + \sum_{l=0}^{2s-2} \binom{2s-1}{l} N_a^{[l+1]}(z) E_v D_v^{2s-1-l} \frac{1}{(1-N_a(z,v))^2} + \mathcal{O}\Big(\frac{1}{(1-2z)^{\frac{2s-1}{2}}}\Big) \\ &= \frac{N_a^{[2s]}(z)}{1-2z} + \sum_{l=0}^{s-2} \binom{2s-1}{2l+1} \frac{(2l+2)!\sigma^{2l+2}\Gamma(\frac{2l+1}{2})(2s-2l-2)!\sigma^{2s-2l-2}}{(1-2z)^{s+\frac{1}{2}} 2^{l+2}(l+1)!\Gamma(\frac{1}{2})2^{s-l-1}} + \mathcal{O}\Big(\frac{1}{(1-2z)^{\frac{2s-1}{2}}}\Big) \\ &= \frac{N_a^{[2s]}(z)}{1-2z} + \frac{(2s)!\sigma^{2s}(s-1)\Gamma(\frac{2s-1}{2})}{s!2^s(1-2z)^{s+\frac{1}{2}}\Gamma(\frac{1}{2})} + \mathcal{O}\Big(\frac{1}{(1-2z)^{\frac{2s-1}{2}}}\Big), \end{aligned} \tag{5.252}$$

which proves after using the solution (5.244) the result. The other case $1 \le a \le 2s$ to $2s+1$ can easily be seen to be true

$$N_a^{[2s+1]}(z) = \mathcal{O}\Big(\frac{1}{(1-2z)^{\frac{2s-1}{2}}}\Big). \tag{5.253}$$

Thus it holds that

$$\frac{X_{n,a} - \mu n}{\sqrt{\sigma^2 n}} \xrightarrow{(d)} \mathcal{N}(0,1), \tag{5.254}$$

which proves Theorem 33.

Chapter 6

The distribution of distances in increasing trees

6.1 Introduction

Recently there have been a lot of studies devoted to a distributional analysis of distances between random nodes in a lot of tree families of interest. We mention here Mahmoud and Neininger [51] for binary search trees, Christophi and Mahmoud [12] for the digital data structure called Tries, and Panholzer [65] for simply generated trees (= Galton Watson trees).

Considerably less studies are made to analyze the distribution of distances between *specified* nodes in *labelled* tree structures. "Exceptions" are the work of Dobrow [16] and Dobrow and Smythe [18], who have shown a central limit theorem for the distance between the nodes labelled by j and n (= the largest node), respectively, in a random recursive tree of size n for all sequences $(n, j(n))_{n\in\mathbb{N}}$, with $1 \leq j = j(n) < n$, and the work of Devroye and Neininger [15], who have shown a central limit theorem for the distance between the nodes labelled by j_1 and j_2 in a random binary search tree of size n for all sequences $(n, j_1(n), j_2(n))_{n\in\mathbb{N}}$ with $1 \leq j_1 = j_1(n) < j_2 = j_2(n) \leq n$, provided that $j_2 - j_1 \to \infty$.

In this chapter we "continue" the work of [16; 18] by extending the results from recursive trees to a larger class of tree families: we give a distributional analysis (by showing a central limit theorem) of the r. v. $\Delta_{n,j}$, which counts the distance, measured by the number of edges lying on the connecting path between node j and node n in a random grown simple increasing tree of size n.

6.2 Results for grown simple families of increasing trees

6.2.1 Exact formulæ

Here we give the exact formulæ for the distribution, the expectation and the variance of the random variable $\Delta_{n,j}$. In the following formula for the probabilities $\mathbb{P}\{\Delta_{n,j} = m\}$ we have to distinguish between the case of plane-oriented recursive trees ($c_1 = -2c_2$) and the other instances of grown simple families of increasing trees ($c_1 \neq -2c_2$).

Theorem 34. *The probabilities $\mathbb{P}\{\Delta_{n,j} = m\}$, which give the probability that the distance between the node with label j and the node with label n in a randomly chosen size-n tree of a grown simple family of increasing trees as given by Lemma 1, is m, are, for $m \geq 1$ and $1 \leq j < n$ given by the following formula.*

- *Case C with instance $c_1 = -2c_2$ (Plane-oriented recursive trees): it holds*

$$\mathbb{P}\{\Delta_{n,j} = m\} = \frac{2^{2n-3}}{(n-1)\binom{2n-2}{n-1}\binom{n-2}{j-1}} \sum_{k=0}^{m-1} \frac{1}{2^{m-1-k}} \Big(\sum_{l=0}^{j-1} \begin{bmatrix} l \\ k \end{bmatrix} \frac{1}{l!} \binom{j-l-\frac{3}{2}}{j-l-1} \Big) \times$$

$$\times \Big(\sum_{l=0}^{n-j-1} \left[{l \atop m-1-k} \right] \frac{1}{l!} \binom{n-2-l}{j-1} \Big). \tag{6.1}$$

- *Case A, Case B, and Case C with instance $c_1 \neq -2c_2$: it holds*

$$\begin{aligned}
\mathbb{P}\{\Delta_{n,j} = m\} &= \frac{\binom{j-1+\frac{c_2}{c_1}}{j-1}\big(1+\frac{c_2}{c_1}\big)}{(n-1)\binom{n-2}{j-1}\binom{n-1-\frac{c_2}{c_1}}{n-1}} \Bigg(\sum_{l=0}^{n-j-1} \binom{n-l-2}{j-1} \big(1+\frac{c_2}{c_1}\big)^{m-1} \frac{1}{l!} \left[{l \atop m-1} \right] \\
&\quad + \frac{c_1}{c_1+2c_2} \sum_{k=0}^{n-j-1} \binom{n-k-2}{j-1} \sum_{l=0}^{m-2} \frac{2^l (c_1+c_2)^l}{(c_1+2c_2)^l} \big(1+\frac{c_2}{c_1}\big)^{m-2-l} \frac{1}{k!} \left[{k \atop m-2-l} \right] \Bigg) \\
&- \frac{1}{(n-1)\binom{n-2}{j-1}\binom{n-1-\frac{c_2}{c_1}}{n-1}} \sum_{l=0}^{m-2} \frac{2^{m-2-l}(c_1+c_2)^{m-1-l}}{(c_1+2c_2)^{m-1-l}} \sum_{k=0}^{l} \Bigg(\Big(\sum_{i=0}^{j-1} \binom{j-2-i-\frac{c_2}{c_1}}{j-1-i} 2^k \big(1+\frac{c_2}{c_1}\big)^k \frac{1}{k!} \left[{i \atop k} \right] \Big) \times \\
&\qquad \times \Big(\sum_{i=0}^{n-j-1} \binom{n-2-i}{j-1} \big(1+\frac{c_2}{c_1}\big)^{l-k} \frac{1}{(l-k)!} \left[{i \atop l-k} \right] \Big) \Bigg).
\end{aligned} \tag{6.2}$$

Theorem 35. *The expectation and the variance of the random variable $\Delta_{n,j}$, which counts the distance between the node with label j and the node with label n in a randomly chosen tree of size n, are for all grown simple families of increasing trees as given by Lemma 1 (and $1 \leq j < n$) given by*

$$\begin{aligned}
\mathbb{E}(\Delta_{n,j}) &= \big(1+\frac{c_2}{c_1}\big)\Bigg(H_{n+\frac{c_2}{c_1}-1} + H_{j+\frac{c_2}{c_1}} - 2H_{1+\frac{c_2}{c_1}} + \frac{1+\frac{2c_2}{c_1}}{j+\frac{c_2}{c_1}} \Bigg) - 2\frac{c_2}{c_1}. \\
\mathbb{V}(\Delta_{n,j}) &= \big(1+\frac{c_2}{c_1}\big) H_{n+\frac{c_2}{c_1}-1} + \Bigg(\big(1+\frac{c_2}{c_1}\big) - 4\frac{\big(1+\frac{2c_2}{c_1}\big)\big(1+\frac{c_2}{c_1}\big)^2}{j+\frac{c_2}{c_1}} \Bigg) H_{j+\frac{c_2}{c_1}} \\
&\quad - 2\Bigg(\big(1+\frac{c_2}{c_1}\big) - 2\frac{\big(1+\frac{2c_2}{c_1}\big)\big(1+\frac{c_2}{c_1}\big)^2}{j+\frac{c_2}{c_1}} \Bigg) H_{1+\frac{c_2}{c_1}} - \big(1+\frac{c_2}{c_1}\big)^2 \Big(H_{n+\frac{c_2}{c_1}+1}2 + 3H_{j+\frac{c_2}{c_1}}2 - 4H_{\frac{c_2}{c_1}+1}2 \Big) \\
&\quad + 2\big(1+\frac{c_2}{c_1}\big)\big(1+\frac{2c_2}{c_1}\big) - \frac{\big(1+\frac{2c_2}{c_1}\big)\big(1+\frac{c_2}{c_1}\big)}{j+\frac{c_2}{c_1}} - \frac{\big(1+\frac{2c_2}{c_1}\big)^2\big(1+\frac{c_2}{c_1}\big)^2}{(j+\frac{c_2}{c_1})^2}.
\end{aligned} \tag{6.3}$$

We explicitly give the formulæ for the three most prominent members of grown simple tree families. The result for recursive trees already appears in [58].

Corollar 9. *The expectation and the variance of the random variable $\Delta_{n,j}$ (for $1 \leq j < n$) are for plane-oriented recursive trees ($\varphi_0 = 1, c_1 = 2, c_2 = -1$) given by*

$$\begin{aligned}
\mathbb{E}(\Delta_{n,j}) &= H_{2n-2} - \frac{1}{2}H_{n-1} + H_{2j} - \frac{1}{2}H_j - 1, \\
\mathbb{V}(\Delta_{n,j}) &= H_{2n-2} - \frac{1}{2}H_{n-1} + H_{2j} - \frac{1}{2}H_j - H_{2n-2}2 + \frac{1}{4}H_{n-1}2 - 3H_{2j}2 + \frac{3}{4}H_j2 + 2.
\end{aligned} \tag{6.4}$$

The expectation and the variance of the random variable $\Delta_{n,j}$ (for $1 \leq j < n$) are for recursive trees ($\varphi_0 = 1, c_1 = 1, c_2 = 0$) given by

$$\begin{aligned}
\mathbb{E}(\Delta_{n,j}) &= H_{n-1} + H_j + \frac{1}{j} - 2, \\
\mathbb{V}(\Delta_{n,j}) &= H_{n-1} + H_j - H_{n-1}2 - 3H_j2 - \frac{4}{j}H_j + 4 + \frac{3}{j} - \frac{1}{j^2}.
\end{aligned} \tag{6.5}$$

The expectation and the variance of the random variable $\Delta_{n,j}$ (for $1 \leq j < n$) are for binary increasing *trees ($\varphi_0 = 1, c_1 = c_2 = 1$, and thus $d = 1 + \frac{c_2}{c_1} = 2$) given by*

$$\begin{aligned}\mathbb{E}(\Delta_{n,j}) &= 2H_n + 2H_{j+1} + \frac{6}{j+1} - 8,\\ \mathbb{V}(\Delta_{n,j}) &= 2H_n + 2H_{j+1} - 4H_n2 - 12H_{j+1}2 - \frac{48}{j+1}H_{j+1} + 26 + \frac{66}{j+1} - \frac{36}{(j+1)^2}.\end{aligned} \tag{6.6}$$

6.2.2 Distribution laws

We can easily reprove a result of [18].

Corollar 10 ([18]). *The random variable $\Delta_{n,j}$ satisfies the following distribution law.*

$$\Delta_{n,j} \overset{(d)}{=} \Delta_{j+1,j} \oplus \bigoplus_{k=j+1}^{n-1} B_k, \quad \text{for } j < n \tag{6.7}$$

where the B_k's are Bernoulli distributed random variables

$$B_k \overset{(d)}{=} \mathrm{Be}(p_k), \quad p_k = \frac{1 + \frac{c_2}{c_1}}{k + \frac{c_2}{c_1}}, \quad \text{for } j + 1 \le k \le n - 1. \tag{6.8}$$

Note that in [18] an interpretation for the probabilities p_k appearing in Corollary 10 where given: p_k is the probability that node $k + 1$ is attached (a child) of node k.
Now we are going to calculate the distribution law of $\Delta_{j+1,j}$. In [18] the law $\Delta_{j+1,j}$ for recursive trees was given. We extend this result to arbitrary grown increasing tree families.

Theorem 36. *The distribution law of $\Delta_{j+1,j}$ is given as follows. For $c_1 \neq -2c_2$:*

$$\Delta_{j+1,j} \overset{(d)}{=} \sum_{k=1}^{\eta_j} \tilde{B}_k, \tag{6.9}$$

where $\tilde{B}_k \overset{(d)}{=} \mathrm{Be}(\tilde{p_k})$, $\tilde{p}_0 = \tilde{p}_1 = 1$ and $\tilde{p}_k = \frac{1}{k-1+\frac{c_2}{c_1}}$ for $3 \le k \le j$. Further $\mathbb{P}\{\eta_j = 1\} = \frac{1+\frac{c_2}{c_1}}{j+\frac{c_2}{c_1}}$ and $\mathbb{P}\{\eta_j = m\} = \frac{1}{j+\frac{c_2}{c_1}}$ for $2 \le m \le j$.
For $c_1 = -2c_2$ we find an even simpler decomposition:

$$\Delta_{j+1,j} \overset{(d)}{=} \bigoplus_{k=1}^{j} \tilde{B}_k = 1 \oplus \bigoplus_{k=1}^{j-1} \mathbb{1}(A_k), \tag{6.10}$$

where $\tilde{B}_1 \overset{(d)}{=} 1$ and $\tilde{B}_k \overset{(d)}{=} \mathrm{Be}(\frac{2}{2k-1})$ for $2 \le k \le j$. Note that $\tilde{B}_{k+1} = \mathbb{1}(A_k)$, $1 \le k \le j-1$ where A_k denotes the event that node k is on the path from $j + 1$ to j.

6.2.3 Limiting distribution results

In the following we give the main theorem this chapter, i. e. the central limit theorems for the r. v. $\Delta_{n,j}$ and $\Delta_{n;j_1,j_2}$, respectively.

Theorem 37. *The centralized and normalized random variable $\Delta^*_{n,j}$, where $\Delta_{n,j}$ counts the distance between the nodes with the label j and the label n in a randomly chosen size-n tree of a grown simple family of increasing trees as given by Lemma 1, is, for arbitrary sequences $(n, j(n))_{n\in\mathbb{N}}$, with $1 \le j = j(n) < n$, asymptotically for $n \to \infty$ Gaussian distributed,*

$$\Delta^*_{n,j} := \frac{\Delta_{n,j} - \mu_{n,j}}{\sigma_{n,j}} \xrightarrow{(d)} \mathcal{N}(0, 1), \tag{6.11}$$

where $\mu_{n,j} := \big(1+\frac{c_2}{c_1}\big)(\log n + \log j)$ *and* $\sigma_{n,j}^2 := \big(1+\frac{c_2}{c_1}\big)(\log n + \log j)$.

Corollar 11. *The centralized and normalized random variable* $\Delta^*_{n;j_1,j_2}$*, where* $\Delta_{n;j_1,j_2}$ *counts the distance between the nodes with the label* j_1 *and the label* j_2 *in a randomly chosen size-*n *tree of a grown simple family of increasing trees as given by Lemma 1, is, for arbitrary sequences* $(n, j_1(n), j_2(n))_{n\in\mathbb{N}}$*, with* $1 \le j_1 = j_1(n), j_2 = j_2(n) < n$*, provided that* $\max(j_1, j_2) \to \infty$*, asymptotically for* $n \to \infty$ *Gaussian distributed,*

$$\Delta^*_{n;j_1,j_2} := \frac{\Delta_{n,;j_1,j_2} - \mu_{n;j_1,j_2}}{\sigma_{n;j_1,j_2}} \xrightarrow{(d)} \mathcal{N}(0,1), \tag{6.12}$$

where $\mu_{n;j_1,j_2} := \big(1+\frac{c_2}{c_1}\big)(\log j_1 + \log j_2)$*,* $\sigma^2_{n;j_1,j_2} := \big(1+\frac{c_2}{c_1}\big)(\log j_1 + \log j_2)$.

6.3 A recurrence for the probabilities

By using the combinatorial description of increasing trees as stated in the Preliminaries (6) we will obtain a recursive description of $\Delta_{n,j}$ and thus of the probabilities $\mathbb{P}\{\Delta_{n,j} = m\}$ for simple families of increasing trees. For these considerations we also have to introduce the r. v. $D_{n,j}$, which counts the depth (= the number of edges lying on the path connecting the root, i. e. the node with label 1, with the node considered) of node j in a random size-n tree of a simple family of increasing trees. One may thus also define $D_{n,j} := \Delta_{n;j,1}$. We will use an approach similar to Chapter 5 Section 5.3 to obtain a recurrence for the probabilities.

For increasing trees of size n with root-degree r and subtrees with sizes $k_1, \dots, k_r$, enumerated from left to right, we will distinguish between two cases that cover all possible cases by symmetry arguments . For the first case we assume that node j and node n are both lying in the leftmost subtree of the root, where the node labelled by j is the i-th smallest node in this subtree. We can then reduce the computation of the probabilities $\mathbb{P}\{\Delta_{n,j} = m\}$ to the probabilities $\mathbb{P}\{\Delta_{k_1,i} = m\}$. For the second case we assume that node j is lying in the leftmost subtree and is the i-th smallest node in this subtree, whereas node n is lying in the second subtree (from left to right). We can thus reduce the computation of the probabilities $\mathbb{P}\{\Delta_{n,j} = m\}$ to the probabilities of the depths $\mathbb{P}\{D_{k_1,i} = t\}$ and $\mathbb{P}\{D_{k_2,k_2} = m - 2 - t\}$.

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

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

the $i-1$ labels smaller than j are chosen from $2, 3, \dots, j-1$, the $k_1 - 1 - i$ labels larger than j but different from n are chosen from $j+1, \dots, n-1$, and the remaining $n-1-k_1$ labels are distributed to the second, third, …, r-th subtree. Due to symmetry arguments we obtain a factor r, if the node j is the i-th smallest node in the second, third, …, r-th subtree.

Analogously, in the second case we get the factor $\varphi_r T_{k_1} \cdots T_{k_r}$ divided by the total weight T_n of trees of size n and multiplied by the number of order preserving relabellings of the r subtrees, which are given here by

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

the $i-1$ labels smaller than j are chosen from $2, 3, \dots, j-1$, the $k_1 - i$ labels larger than j are chosen from $j+1, \dots, n-1$ (since node n must be in the second subtree), and the remaining $n-2-k_1$ labels are distributed to the second, third, …, r-th subtree. Again due to symmetry arguments we obtain a factor $r(r-1)$.

Summing up over all choices for the rank i of label j in its subtree, the subtree sizes $k_1, \dots, k_r$, and the degree r of the root node gives the following recurrence (6.13).

$$\begin{aligned}
\mathbb{P}\{\Delta_{n,j}=m\} &= \sum_{r\ge 1} r\varphi_r \sum_{\substack{k_1+\cdots+k_r=n-1,\\ k_1,\dots,k_r\ge 1}} \frac{T_{k_1}\cdots T_{k_r}}{T_n}\times \\
&\quad \times \sum_{i=1}^{\min\{k_1,j-1\}} \mathbb{P}\{\Delta_{k_1,i}=m\}\binom{j-2}{i-1}\binom{n-1-j}{k_1-1-i}\binom{n-1-k_1}{k_2,k_3,\dots,k_r} \\
&\quad + \sum_{r\ge 1} r(r-1)\varphi_r \sum_{\substack{k_1+\cdots+k_r=n-1,\\ k_1,\dots,k_r\ge 1}} \frac{T_{k_1}\cdots T_{k_r}}{T_n}\times \\
&\times \sum_{i=1}^{\min\{k_1,j-1\}} \sum_{t=0}^{m-2} \mathbb{P}\{D_{k_1,i}=t\}\mathbb{P}\{D_{k_2,k_2}=m-2-t\}\binom{j-2}{i-1}\binom{n-1-j}{k_1-i}\binom{n-2-k_1}{k_2-1,k_3,\dots,k_r},
\end{aligned} \tag{6.13}$$

for $2 \le j \le n-1$, with $\mathbb{P}\{\Delta_{n,1}=m\} = \mathbb{P}\{D_{n,n}=1\}$ and $\mathbb{P}\{\Delta_{n,n}=m\} = \delta_{m,0}$.
To treat this recurrence (6.13) we set $n := k+j$ with $k \ge 0$ and define the trivariate generating functions

$$\begin{aligned}
M(z,u,v) &:= \sum_{k\ge 1}\sum_{j\ge 1}\sum_{m\ge 0} \mathbb{P}\{\Delta_{k+j,j}=m\} T_{k+j} \frac{z^{j-1}}{(j-1)!}\frac{u^{k-1}}{(k-1)!}v^m, \\
N(z,u,v) &:= \sum_{k\ge 0}\sum_{j\ge 1}\sum_{m\ge 0} \mathbb{P}\{D_{k+j,j}=m\} T_{k+j} \frac{z^{j-1}}{(j-1)!}\frac{u^{k}}{k!}v^m.
\end{aligned} \tag{6.14}$$

Multiplying (6.13) with $T_{k+j}\frac{z^{j-2}}{(j-2)!}\frac{u^{k-1}}{(k-1)!}v^m$ and summing up over $k \ge 1$, $j \ge 2$ and $m \ge 0$ gives then $\frac{\partial}{\partial z}M(z,u,v)$ for the left hand side and $\varphi'\big(T(z+u)\big)M(z,u,v)$ as well as $v^2 N(z,u,v)N(z+u,0,v)\varphi''\big(T(z+u)\big)$ for the right hand side of (6.13). Since these are essentially straightforward, but quite lengthy computations, they are omitted here; similar considerations are done in [68] for a study of the r. v. $D_{n,j}$, where the (somewhat simpler) recurrences appearing there are treated analogously. In any case we obtain the following differential equation:

$$\frac{\partial}{\partial z}M(z,u,v) = \varphi'\big(T(z+u)\big)M(z,u,v) + v^2 N(z,u,v)N(z+u,0,v)\varphi''\big(T(z+u)\big), \tag{6.15}$$

together with the initial condition

$$\begin{aligned}
M(0,u,v) &= \sum_{k\ge 1}\sum_{m\ge 0}\mathbb{P}\{\Delta_{k+1,1}=m\}T_{k+1}\frac{u^{k-1}}{(k-1)!}v^m = \sum_{k\ge 1}\sum_{m\ge 0}\mathbb{P}\{D_{k+1,k+1}=m\}T_{k+1}\frac{u^{k-1}}{(k-1)!}v^m \\
&= \frac{\partial}{\partial u}N(u,0,v).
\end{aligned} \tag{6.16}$$

As mentioned before, the random variable $D_{n,j}$ was already analyzed in [68], where the following result was obtained:

$$N(z,u,v) = \varphi\big(T(u)\big)\Big(\frac{\varphi\big(T(z+u)\big)}{\varphi\big(T(u)\big)}\Big)^v = T'(u)\Big(\frac{T'(z+u)}{T'(u)}\Big)^v. \tag{6.17}$$

Consequently we get

$$N(z,0,v) = \varphi_0\Big(\frac{\varphi\big(T(z)\big)}{\varphi_0}\Big)^v = \varphi_0\Big(\frac{T'(z)}{\varphi_0}\Big)^v, \tag{6.18}$$

and further

$$M(0,u,v) = \frac{\partial}{\partial u}N(u,0,v) = \frac{\partial}{\partial u}\Big(\varphi_0\big(\frac{T'(u)}{\varphi_0}\big)^v\Big) = \varphi_0\Big(\frac{T'(u)}{\varphi_0}\Big)^{v-1}\frac{T''(u)}{\varphi_0} = vT''(u)\Big(\frac{T'(u)}{\varphi_0}\Big)^{v-1}. \tag{6.19}$$

Thus the differential equation (6.15) can be rewritten into

$$\frac{\partial}{\partial z} M(z,u,v) = \varphi'\big(T(z+u)\big) M(z,u,v) + \frac{v^2 \varphi''\big(T(z+u)\big)\big(T'(z+u)\big)^{2v}}{\big(T'(u)\big)^{v-1} \varphi_0^{v-1}}, \tag{6.20}$$

with initial condition $M(0,u,v) = vT''(u)\Big(\frac{T'(u)}{\varphi_0}\Big)^{v-1}$. The corresponding homogeneous differential equation has the solution

$$M^{[h]}(z,u,v) = C(u,v) \exp\Big(\int_0^z \varphi'\big(T(t+u)\big) dt\Big) = C(u,v) \frac{\varphi\big(T(z+u)\big)}{\varphi\big(T(u)\big)} = C(u,v) \frac{T'(z+u)}{T'(u)}, \tag{6.21}$$

with some function $C(u,v)$. Variation of the constants method leads to the particular solution

$$M^{[p]}(z,u,v) = \frac{v^2 T'(z+u)}{\varphi_0^{v-1}\big(T'(u)\big)^{v-1}} \int_0^z \varphi''\big(T(t+u)\big)\big(T'(t+u)\big)^{2v-1} dt. \tag{6.22}$$

Adapting to the initial condition leads to $C(u,v) = M(0,u,v) = vT''(u)\Big(\frac{T'(u)}{\varphi_0}\Big)^{v-1}$, and therefore to the following proposition.

Proposition 4. *The function $M(z,u,v)$ as defined in equation (6.14), which is the trivariate generating function of the probabilities $\mathbb{P}\{\Delta_{n,j} = m\}$, which give the probability that the distance (measured by the number of edges on the connecting path) between the node with label j and the node with label n in a randomly chosen size-n tree of a simple family of increasing trees with degree-weight generating function $\varphi(t)$, is m, is given by the following formula:*

$$M(z,u,v) = vT''(u)\Big(\frac{T'(u)}{\varphi_0}\Big)^{v-1} \frac{T'(z+u)}{T'(u)} + \frac{v^2 T'(z+u)}{\varphi_0^{v-1}\big(T'(u)\big)^{v-1}} \int_0^z \varphi''\big(T(t+u)\big)\big(T'(t+u)\big)^{2v-1} dt. \tag{6.23}$$

This immediately has the following consequence.

Corollar 12. *The trivariate generating function $M(z,v,u)$ is for all grown simple families of increasing trees as given by Lemma 1 given by the following formula:*

$$\begin{aligned} M(z,u,v) = {} & \frac{\varphi_0(c_1+c_2)v\big(1 - \frac{vc_1}{(c_1+c_2)(2v-1)-c_2}\big)}{(1-c_1u)^{(\frac{c_2}{c_1}+1)(v-1)+1}\big(1-c_1(z+u)\big)^{\frac{c_2}{c_1}+1}} \\ & + \frac{\varphi_0 c_1(c_1+c_2)v^2(1-c_1u)^{(\frac{c_2}{c_1}+1)(v-1)}}{\big((c_1+c_2)(2v-1)-c_2\big)\big(1-c_1(z+u)\big)^{(\frac{c_2}{c_1}+1)(2v-1)+1}}, \end{aligned} \tag{6.24}$$

where we have to set $c_2 = 0$ for Case A and $d = \frac{c_1}{c_2} + 1$ for Case B.

Next we will distinguish between two cases, namely $c_1 = -2c_2$, and $c_1 \neq -2c_2$. In the former case it holds

$$\frac{vc_1}{(c_1+c_2)(2v-1)-c_2} = \frac{vc_1}{\frac{c_1}{2}(2v-1)+\frac{c_1}{2}} = 1, \tag{6.25}$$

and thus we are able to refine Corollary 12 for this instance.

Corollar 13. *For grown simple families of increasing trees of Case C as given by Lemma 1 that are satisfying $c_1 = -2c_2$, and therefore in particular for plane-oriented recursive trees ($\varphi_0 = 1, c_1 = 2, c_2 = -1$), the generating function $M(z,u,v)$ simplifies to*

$$M(z,u,v) = \frac{\varphi_0 c_1 v(1-c_1u)^{\frac{v-1}{2}}}{2(1-c_1(z+u))^{v+\frac{1}{2}}}. \tag{6.26}$$

Remark 14. As in Chapter 5 Section 5.3 we can also use a combinatorial approach to derive the differential equation (6.15) for $M(z,u,v)$. Think of a specifically 4-colored increasing tree T. Exactly one node is colored red, all nodes with a smaller label than the red node are colored black, and nodes with larger label than the red node are colored white except the node with the largest label in T, which is colored green. Assume that the red node of T is not the root node.

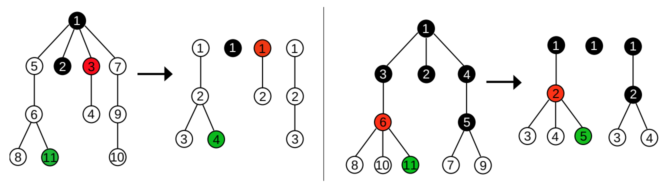

Figure 6.1: Two decompositions of size n=11 increasing trees.

Then the red node is located in one of the r subtrees of the root of T; let us assume that it is in the r-th subtree. Furthermore assume that the green node is also in the r-th subtree. Then the r-th subtree is possibly itself a 4-colored increasing tree (after order preserving relabelling), where the remaining $r-1$ subtrees of the root are only bicolored trees colored black and white. The total weight of the suitably 4-colored increasing trees with j black and k white nodes, where the parameter of the red node is exactly m, is given by $\mathbb{P}\{\Delta_{j+k+2,j+1}=m\}T_{j+k+2}$. Setting up the generating function

$$M(z,u,v) := \sum_{k\geq 0}\sum_{j\geq 0}\sum_{m\geq 0}\mathbb{P}\{\Delta_{k+j+2,j+1}=m\}T_{k+j+2}\frac{z^j}{j!}\frac{u^k}{k!}v^m, \tag{6.27}$$

we obtain by the same considerations as in Chapter 5 Section 5.3 the differential equation

$$\frac{\partial}{\partial z}M(z,u,v) = \varphi'\big(T(z+u)\big)M(z,u,v) + R(z,u,v), \tag{6.28}$$

where the inhomogeneous part $R(z,u,v)$ is due to the case where the red and the green node are not in the same subtree. Let's consider this case in more detail. Assume that the red node is in the r-th subtree whereas the green node is in the $r-1$-th subtree. The r-th subtree is possibly a tricolored increasing tree with colors black, red and white. The $r-1$-th subtree is also a tricolored increasing tree with colors black, white and green. By introducing the generating functions for the total weights

$$N(z,u,v) := \sum_{k\geq 0}\sum_{j\geq 0}\sum_{m\geq 0}\mathbb{P}\{D_{k+j+1,j+1}=m\}T_{k+j+1}\frac{z^j}{j!}\frac{u^k}{k!}v^m, \tag{6.29}$$

the generating functions of the total weights of the $r-1$-th subtree is given by

$$\begin{aligned}\sum_{k\geq 0}\sum_{j\geq 0}\sum_{m\geq 0}\mathbb{P}\{D_{k+j+1,k+j+1}=m\}T_{k+j+1}\frac{z^j}{j!}\frac{u^k}{k!}v^m &= \sum_{l\geq 0}\sum_{i=0}^{l}\sum_{m\geq 0}\mathbb{P}\{D_{l+1,l+1}=m\}T_{l+1}\frac{z^i}{i!}\frac{u^{l-i}}{(l-i)!}v^m \\ &= \sum_{l\geq 0}\sum_{m\geq 0}\mathbb{P}\{D_{l+1,l+1}=m\}T_{l+1}\frac{(z+u)^l}{l!}v^m = N(z+u,0,v).\end{aligned}$$

Finally we obtain the inhomogeneous part :

$$R(z,u,v) = v^2\varphi''\big(T(z+u)\big)N(z,u,v)N(z+u,0,v), \tag{6.30}$$

where the additional factor v^2 is due to the fact that the depth of the red and the green node increase by one in the original tree T.

6.4 Closed formulæ for the probabilities

For extracting coefficients from the trivariate g. f. $M(z,u,v)$ as given by Corollary 12 it is convenient to split $M(z,u,v)$ into two parts $M(z,u,v) = M_1(z,u,v) + M_2(z,u,v)$, where the first part, i. e. $M_1(z,u,v)$, disappears for trees characterized by Corollary 13. Furthermore we use the well known relation for the Stirling numbers of the first kind

$$\sum_{n\ge 0}\sum_{m=0}^{m} \left[{n \atop m}\right] \frac{z^n}{n!} v^m = \frac{1}{(1-z)^v}. \tag{6.31}$$

We only present the calculations for the trees characterized by Corollary 13. The general case, $c_1 \neq -2c_2$, can be treated by the same method; the calculations become a bit lengthier, but are only slightly more complicated. We extract coefficients according to (6.14).

$$\begin{aligned}
\mathbb{P}\{\Delta_{n,j} = m\} &= \frac{(j-1)!(n-j-1)!}{T_n}[z^{j-1}u^{n-j-1}v^m]M(z,u,v) \\
&= \frac{(j-1)!(n-j-1)!}{T_n}[z^{j-1}u^{n-j-1}v^m]\frac{\varphi_0 c_1 v(1-c_1 u)^{\frac{v-1}{2}}}{2(1-c_1(z+u))^{v+\frac{1}{2}}} \\
&= \frac{(j-1)!(n-j-1)!\varphi_0 c_1^{n-1}}{2T_n}[z^{j-1}u^{n-j-1}v^{m-1}]\frac{(1-u)^{\frac{v-1}{2}}}{(1-z-u)^{v+\frac{1}{2}}} \\
&= \frac{(j-1)!(n-j-1)!}{2(n-1)!\binom{n-\frac{3}{2}}{n-1}}[z^{j-1}u^{n-j-1}v^{m-1}]\frac{1}{(1-u)^{\frac{v}{2}+1}(1-\frac{z}{1-u})^{v+\frac{1}{2}}},
\end{aligned} \tag{6.32}$$

where we have used $[z^n]f(c_1 z) = c_1^n [z^n]f(z)$. We get further

$$\begin{aligned}
\mathbb{P}\{\Delta_{n,j} = m\} &= \frac{2^{2n-3}(j-1)!(n-j-1)!}{(n-1)!\binom{2n-2}{n-1}}[u^{n-j-1}v^{m-1}]\frac{\binom{v+j-\frac{3}{2}}{j-1}}{(1-u)^{\frac{v}{2}+j}} \\
&= \frac{2^{2n-3}}{(n-1)\binom{n-2}{j-1}\binom{2n-2}{n-1}}[v^{m-1}]\binom{v+j-\frac{3}{2}}{j-1}\binom{\frac{v}{2}+n-2}{n-j-1}.
\end{aligned} \tag{6.33}$$

The remaining part of the proof follows by using (6.31) and

$$\sum_{l\ge 0}\binom{v+K+l-1}{l}z^l = \frac{1}{(1-z)^{v+K}}, \quad [v^m]\binom{v+j-\frac{3}{2}}{j-1} = [z^{j-1}v^{m-1}]\frac{1}{(1-z)^{v+\frac{1}{2}}}. \tag{6.34}$$

6.5 Closed formulæ for expectation and variance

To avoid lengthy computations we restrict ourselves again to the case covered by Corollary 13, i. e. $c_1 = -2c_2$. We basically use

$$\begin{aligned}
\mathbb{E}(\Delta_{n,j}) &= \frac{(j-1)!(n-j-1)!}{T_n}[z^{j-1}u^{n-j-1}]E_v D_v M(z,u,v), \\
\mathbb{E}(\Delta_{n,j}^2) &= \frac{(j-1)!(n-j-1)!}{T_n}[z^{j-1}u^{n-j-1}]E_v D_v^2 M(z,u,v).
\end{aligned} \tag{6.35}$$

For the calculation of the expectation we begin with

$$E_v D_v M(z,u,v) = \frac{\varphi_0 c_1}{2(1-c_1(z+u))^{\frac{3}{2}}}\Big(1 - \frac{1}{2}\log\big(\frac{1}{1-c_1 u}\big) + \log\big(\frac{1}{1-c_1(z+u)}\big)\Big). \tag{6.36}$$

Now we use the relations

$$\frac{1}{(1-c_1(z+u))^\alpha} = \frac{1}{(1-c_1u)^\alpha(1-\frac{c_1z}{1-c_1u})^\alpha}, \log^\alpha\Big(\frac{1}{1-c_1(z+u)}\Big) = \Big(\log\Big(\frac{1}{1-c_1u}\Big) + \log\Big(\frac{1}{1-\frac{c_1z}{1-c_1u}}\Big)\Big)^\alpha \tag{6.37}$$

and

$$[z^n]\frac{\log\big(\frac{1}{1-z}\big)}{(1-z)^{\alpha+1}} = \binom{n+\alpha}{n}(H_{n+\alpha} - H_\alpha), \tag{6.38}$$

to obtain

$$[z^{j-1}u^{n-j-1}]E_vD_vM(z,u,v) = \frac{\varphi_0 c_1^{n-1}}{2}\binom{n-\frac{3}{2}}{n-j-1}\binom{j-\frac{1}{2}}{j-1}\Big(\frac{1}{2}H_{n-\frac{3}{2}} + \frac{1}{2}H_{j-\frac{1}{2}} - H_{\frac{1}{2}} + 1\Big). \tag{6.39}$$

Together with

$$\frac{(j-1)!(n-j-1)!}{T_n} = \frac{(j-1)!(n-j-1)!}{\varphi_0 c_1^{n-1}(n-1)!\binom{n-\frac{3}{2}}{n-1}} = \frac{2}{\varphi_0 c_1^{n-1}\binom{n-\frac{3}{2}}{n-j-1}\binom{j-\frac{1}{2}}{j-1}}, \tag{6.40}$$

and by converting into “integer” harmonic numbers we get the desired result for $\mathbb{E}(\Delta_{n,j})$.
For the second factorial moment we get

$$\begin{aligned} E_vD_v^2M(z,u,v) = \frac{\varphi_0 c_1}{8(1-c_1(z+u))^{\frac{3}{2}}}\Big(&-4\log\Big(\frac{1}{1-c_1u}\Big) + 8\log\Big(\frac{1}{1-c_1(z+u)}\Big) + \log^2\Big(\frac{1}{1-c_1u}\Big) \\ &- 4\log\Big(\frac{1}{1-c_1u}\Big)\log\Big(\frac{1}{1-c_1(z+u)}\Big) + 4\log^2\Big(\frac{1}{1-c_1(z+u)}\Big)\Big). \end{aligned} \tag{6.41}$$

For extracting coefficients we use again (6.37) and (6.38) together with

$$[z^n]\frac{\log^2\big(\frac{1}{1-z}\big)}{(1-z)^{\alpha+1}} = \binom{n+\alpha}{n}\big((H_{n+\alpha} - H_\alpha)^2 - (H^{(2)}_{n+\alpha} - H^{(2)}_\alpha)\big), \tag{6.42}$$

and obtain

$$\begin{aligned} \mathbb{E}(\Delta^2_{n,j}) = &\frac{1}{4}(H_{n-\frac{3}{2}} - H_{j-\frac{1}{2}})^2 + (H_{n-\frac{3}{2}} - H_{\frac{1}{2}})(H_{j-\frac{1}{2}} - H_{\frac{1}{2}} + 1) + 2(H_{j-\frac{1}{2}} - H_{\frac{1}{2}}) + (H_{j-\frac{1}{2}} - H_{\frac{1}{2}})^2 \\ &- \frac{1}{4}(H^{(2)}_{n-\frac{3}{2}} - H^{(2)}_{j-\frac{1}{2}}) - (H^{(2)}_{j-\frac{1}{2}} - H^{(2)}_{\frac{1}{2}}). \end{aligned} \tag{6.43}$$

The variance follows then via

$$\mathbb{V}(\Delta_{n,j}) = \mathbb{E}(\Delta^2_{n,j}) + \mathbb{E}(\Delta_{n,j}) - \big(\mathbb{E}(\Delta_{n,j})\big)^2. \tag{6.44}$$

For the general case the usage of a computer algebra system becomes handy for carrying out the simplifications.

6.5.1 Proving the distribution laws

First we turn our attention to the proof of Corollary 10. From the explicit representation of the generating function $p_n(v) = \sum_{m\geq 1} \mathbb{P}\{\Delta_{n,j} = m\}$

$$\begin{aligned} p_{n,j}(v) = \big(1+\frac{c_2}{c_1}\big)&\frac{v\binom{n-2+v\left(1+\frac{c_2}{c_1}\right)}{n-j-1}}{(n-1)\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-2}{j-1}}\Bigg(\big(1-\frac{vc_1}{(c_1+c_2)(2v-1)-c_2}\big)\binom{j-1+\frac{c_2}{c_1}}{j-1} \\ &+\frac{vc_1}{(c_1+c_2)(2v-1)-c_2}\binom{j-1+(\frac{c_2}{c_1}+1)(2v-1)}{j-1}\Bigg), \end{aligned} \tag{6.45}$$

we immediately get

$$\begin{aligned} p_{n,j}(v) &= j\binom{j+\frac{c_2}{c_1}}{j}p_{j+1,j}(v)\frac{\binom{n-2+v\left(1+\frac{c_2}{c_1}\right)}{n-j-1}}{(n-1)\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-2}{j-1}} \\ &= j\binom{j+\frac{c_2}{c_1}}{j}p_{j+1,j}(v)\frac{(n-2+v(1+\frac{c_2}{c_1}))^{\underline{n-j-1}}(j-1)!}{(n-1+\frac{c_2}{c_1})^{\underline{n-j-1}}(j+\frac{c_2}{c_1})^{\underline{j}}} \\ &= p_{j+1,j}(v)\prod_{k=j+1}^{n-1}\big(\frac{k-1}{k+\frac{c_2}{c_1}}+\frac{v(1+\frac{c_2}{c_1})}{k+\frac{c_2}{c_1}}\big), \end{aligned} \tag{6.46}$$

which proves Corollary 10. We split the proof of Theorem 36 into two parts. First we turn to the case $c_1 = -2c_2$.

$$\begin{aligned} p_{j+1,j}(v) &= \frac{v2^{2j-1}}{j\binom{2j}{j}}\binom{v+j-\frac{1}{2}}{j-1} = \frac{v2^{2j-1}j!(v+j-\frac{1}{2})^{\underline{j-1}}}{(2j)!} = \frac{v2^{j-1}(v+j-\frac{3}{2})^{\underline{j-1}}}{(2j-1)!!} \\ &= \frac{v\prod_{k=1}^{j-1}(2v+2k-1)}{(2j-1)!!} = v\prod_{k=1}^{j-1}\big(\frac{2k-1}{2k+1}+\frac{2v}{2k+1}\big). \end{aligned} \tag{6.47}$$

We still have to prove that $\tilde{B}_{k+1} = \mathbb{1}(A_k)$. Due to the independence of the random variables we only have to show $\tilde{B}_j = \mathbb{1}(A_{j-1})$ for $\Delta_{j+1,j}$, which holds then for arbitrary $j \geq 1$.

$$\begin{aligned} \mathbb{P}\{A_{j-1}\} &= \mathbb{P}\{A_{j-1}|\Delta_{j,j-1}=1\}\mathbb{P}\{\Delta_{j,j-1}=1\} + \mathbb{P}\{A_{j-1}|\Delta_{j,j-1}>1\}\mathbb{P}\{\Delta_{j,j-1}>1\} \\ &= \frac{\mathbb{P}\{A_{j-1}|\Delta_{j,j-1}=1\}}{2j-3} + \frac{\mathbb{P}\{A_{j-1}|\Delta_{j,j-1}>1\}(2j-4)}{2j-3} \\ &= \frac{2j-2}{(2j-3)(2j-1)} + \frac{2j-4}{(2j-3)(2j-1)} = \frac{2}{2j-1}. \end{aligned} \tag{6.48}$$

Now one can proceed as follows

$$\mathbb{P}\{\Delta_{j+1,j} = m\} = \mathbb{P}\{\Delta_{j+1,j} = m|A_{j-1}\}\mathbb{P}\{A_{j-1}\} + \mathbb{P}\{\Delta_{j+1,j} = m|A_{j-1}^c\}\mathbb{P}\{A_{j-1}^c\}. \tag{6.49}$$

Now we use the decomposition of $\Delta_{j+1,j} = \mathbb{1}(A_{j-1}) \oplus \Delta_{j,j-1}$ and the independence of $\Delta_{j,j-1}$ and $\mathbb{1}(A_{j-1})$ to obtain

$$\mathbb{P}\{\Delta_{j+1,j} = m\} = \mathbb{P}\{\Delta_{j,j-1} = m-1\}\mathbb{P}\{A_{j-1}\} + \mathbb{P}\{\Delta_{j,j-1} = m\}\mathbb{P}\{A_{j-1}^c\}. \tag{6.50}$$

For the other case $c_1 \neq -2c_2$ we have

$$\begin{aligned} p_{j+1,j}(v) = \big(1+\frac{c_2}{c_1}\big)\frac{v}{j\binom{j+\frac{c_2}{c_1}}{j}}\Bigg(\big(1-\frac{vc_1}{(c_1+c_2)(2v-1)-c_2}\big)\binom{j-1+\frac{c_2}{c_1}}{j-1} \\ + \frac{vc_1}{(c_1+c_2)(2v-1)-c_2}\binom{j-1+(\frac{c_2}{c_1}+1)(2v-1)}{j-1}\Bigg), \end{aligned} \tag{6.51}$$

which can be written as

$$p_{j+1,j}(v) = \frac{v(1+\frac{c_2}{c_1})}{(j+\frac{c_2}{c_1})}\Big(1-\frac{v}{(1+\frac{c_2}{c_1})(2v-1)-\frac{c_2}{c_1}}\Big) + \frac{v^2(1+\frac{c_2}{c_1})\big(j-1+(\frac{c_2}{c_1}+1)(2v-1)\big)^{\underline{j-1}}}{(j+\frac{c_2}{c_1})\big((1+\frac{c_2}{c_1})(2v-1)-\frac{c_2}{c_1}\big)(j-1+\frac{c_2}{c_1})^{\underline{j-1}}}\Bigg). \tag{6.52}$$

We get further

$$\begin{aligned} \frac{\big(j-1+(\frac{c_2}{c_1}+1)(2v-1)\big)^{\underline{j-1}}}{\big((1+\frac{c_2}{c_1})(2v-1)-\frac{c_2}{c_1}\big)(j-1+\frac{c_2}{c_1})^{\underline{j-1}}} &= \frac{(\frac{c_2}{c_1}+1)(2v-1)-\frac{c_2}{c_1}+j-1+\frac{c_2}{c_1}}{j-1+\frac{c_2}{c_1}}\times \\ &\times \frac{\big(j-2+(\frac{c_2}{c_1}+1)(2v-1)\big)^{\underline{j-2}}}{\big((1+\frac{c_2}{c_1})(2v-1)-\frac{c_2}{c_1}\big)(j-2+\frac{c_2}{c_1})^{\underline{j-2}}}, \end{aligned} \tag{6.53}$$

which can be written as

$$a_{j-1} = \frac{\big((\frac{c_2}{c_1}+1)(2v-1)-\frac{c_2}{c_1}\big)}{j-1+\frac{c_2}{c_1}}a_{j-2} + a_{j-2}, \quad a_j := \frac{\big(j+(\frac{c_2}{c_1}+1)(2v-1)\big)^{\underline{j}}}{\big((1+\frac{c_2}{c_1})(2v-1)-\frac{c_2}{c_1}\big)(j+\frac{c_2}{c_1})^{\underline{j}}}. \tag{6.54}$$

Iterating this argument leads to

$$a_{j-1} = \sum_{k=0}^{j-2}\frac{\big((\frac{c_2}{c_1}+1)(2v-1)-\frac{c_2}{c_1}\big)}{k+1-\frac{c_2}{c_1}}a_k + a_0. \tag{6.55}$$

By using (6.55) we get

$$p_{j+1,j}(v) = \frac{v(1+\frac{c_2}{c_1})}{(j+\frac{c_2}{c_1})} + \frac{v^2(1+\frac{c_2}{c_1})}{(j+\frac{c_2}{c_1})}\sum_{k=0}^{j-2}\frac{a_k}{k+1+\frac{c_2}{c_1}} = \frac{v(1+\frac{c_2}{c_1})}{(j+\frac{c_2}{c_1})} + \frac{v^2}{(j+\frac{c_2}{c_1})}\sum_{k=0}^{j-2}\frac{\big(k+(\frac{c_2}{c_1}+1)(2v-1)\big)^{\underline{k}}}{(k+1+\frac{c_2}{c_1})^{\underline{k}}}. \tag{6.56}$$

Remark 15. Note that a decomposition of the form

$$\Delta_{j+1,j} \stackrel{(d)}{=} 1 \oplus \sum_{k=1}^{j-1}\mathbb{1}(A_k) \tag{6.57}$$

is possible for arbitrary grown simple families of increasing trees, but only in the case $c_1 = -2c_2$ the indicators are mutually independent. E.g. for recursive trees we get

$$\begin{aligned} \mathbb{P}\{A_{j-1}\} &= \mathbb{P}\{A_{j-1}|\Delta_{j,j-1}=1\}\mathbb{P}\{\Delta_{j,j-1}=1\} + \mathbb{P}\{A_{j-1}|\Delta_{j,j-1}>1\}\mathbb{P}\{\Delta_{j,j-1}>1\} \\ &= \frac{j-1}{(j-1)j} + \frac{j-2}{(j-1)j} = \frac{2j-3}{j(j-1)}. \end{aligned} \tag{6.58}$$

Assuming that the A_k's are mutually independent we get further $\mathbb{P}\{A_k\} = \frac{2k-1}{k(k+1)}$. But it can easily be seen that $\mathbb{P}\{A_{j-1}\}\mathbb{P}\{A_{j-2}\} \neq \mathbb{P}\{A_{j-1}A_{j-2}\}$, which leads to a contradiction.

6.6 Proving the central limit theorem

As in the previous sections we consider mainly the case covered by Corollary 13, i. e. $c_1 = -2c_2$. At the end of this section we sketch the analogous calculations for the general case. We start with an expression for the probability generating function $p_{n,j}(v) = \sum_{m\geq 0} \mathbb{P}\{\Delta_{n,j} = m\} v^m$ obtained from (6.33):

$$\begin{aligned} p_{n,j}(v) &= \frac{(j-1)!(n-j-1)!}{T_n}[z^{j-1}u^{n-j-1}]M(z,u,v) = \frac{2^{2n-3}}{(n-1)\binom{n-2}{j-1}\binom{2n-2}{n-1}}\binom{v+j-\frac{3}{2}}{j-1}\binom{\frac{v}{2}+n-2}{n-j-1} \\ &= \frac{2^{2n-3}(n-1)!\Gamma(v+j-\frac{1}{2})\Gamma(n-1+\frac{v}{2})}{(2n-2)!\Gamma(v+\frac{1}{2})\Gamma(\frac{v}{2}+j)}. \end{aligned} \tag{6.59}$$

The moment generating function $\mathcal{M}_{n,j}(t)$ of $\Delta^*_{n,j} := (\Delta_{n,j} - \mu_{n,j})/\sigma_{n,j}$ is then given by

$$\mathcal{M}_{n,j}(t) := \mathbb{E}(e^{t\Delta^*_{n,j}}) = e^{-\frac{\mu_{n,j}}{\sigma_{n,j}}t}\mathbb{E}(e^{\frac{\Delta_{n,j}}{\sigma_{n,j}}t}) = e^{-\frac{\mu_{n,j}}{\sigma_{n,j}}t} p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}). \tag{6.60}$$

For our further computations we split the region $1 \leq j < n$ into two cases, namely j big, such that $j \geq \log n$, and j small, such that $j \leq \log n$. In both cases we set $\mu_{n,j} := (\log n + \log j)/2$ and $\sigma^2_{n,j} := (\log n + \log j)/2$. In the former case $j \geq \log n$ we get by using Stirling's formula for the Gamma function

$$\Gamma(z) = \Big(\frac{z}{e}\Big)^z \frac{\sqrt{2\pi}}{\sqrt{z}}\Big(1 + \frac{1}{12z} + \frac{1}{288z^2} + \mathcal{O}(\frac{1}{z^3})\Big), \tag{6.61}$$

the expansion

$$p_{n,j}(v) = \frac{\sqrt{\pi}}{2\Gamma(v+\frac{1}{2})} n^{\frac{v-1}{2}} j^{\frac{v-1}{2}}\Big(1 + \mathcal{O}(\frac{1}{n}) + \mathcal{O}(\frac{1}{j})\Big) = \frac{\sqrt{\pi}}{2\Gamma(v+\frac{1}{2})} e^{(v-1)\mu_{n,j}}\Big(1 + \mathcal{O}(\frac{1}{n}) + \mathcal{O}(\frac{1}{j})\Big). \tag{6.62}$$

We get further

$$\begin{aligned} p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}) &= \frac{\sqrt{\pi}}{2\Gamma(e^{\frac{t}{\sigma_{n,j}}} + \frac{1}{2})} e^{(\frac{t}{\sigma_{n,j}} + \frac{t^2}{2!\mu_{n,j}} + \mathcal{O}(\frac{1}{\sigma^3_{n,j}}))\mu_{n,j}}\Big(1 + \mathcal{O}(\frac{1}{n}) + \mathcal{O}(\frac{1}{j})\Big) \\ &= e^{t\sigma_{n,j} + \frac{t^2}{2!}}\Big(1 + \mathcal{O}(\frac{1}{\sqrt{\log n}}) + \mathcal{O}(\frac{1}{j})\Big), \end{aligned} \tag{6.63}$$

where we have used

$$\frac{\sqrt{\pi}}{2\Gamma(e^{\frac{t}{\sigma_{n,j}}} + \frac{1}{2})} = \frac{\sqrt{\pi}}{2\Gamma(\frac{3}{2})}\Big(1 + \mathcal{O}(\frac{1}{\sqrt{\log n}})\Big) = 1 + \mathcal{O}(\frac{1}{\sqrt{\log n}}). \tag{6.64}$$

This leads to

$$\mathcal{M}_{n,j}(t) = e^{-\sigma_{n,j}t} p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}) = e^{\frac{t^2}{2}}\Big(1 + \mathcal{O}(\frac{1}{\sqrt{\log n}}) + \mathcal{O}(\frac{1}{j})\Big). \tag{6.65}$$

In the latter case $j \leq \log n$ we get for $p_{n,j}(v)$ the asymptotic expansion

$$p_{n,j}(v) = \frac{\sqrt{\pi}}{2\Gamma(v+\frac{1}{2})} n^{\frac{v-1}{2}} \frac{\Gamma(v+j-\frac{1}{2})}{\Gamma(\frac{v}{2}+j)}\Big(1 + \mathcal{O}(\frac{1}{n})\Big). \tag{6.66}$$

This leads to

$$\begin{aligned} p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}) &= \frac{\sqrt{\pi}\,\Gamma(e^{\frac{t}{\sigma_{n,j}}}+j-\frac{1}{2})}{2\Gamma(e^{\frac{t}{\sigma_{n,j}}}+\frac{1}{2})\Gamma(\frac{e^{\frac{t}{\sigma_{n,j}}}}{2}+j)} e^{(\frac{t}{\sigma_{n,j}}+\frac{t^2}{2!\mu_{n,j}}+\mathcal{O}(\frac{1}{\sigma_{n,j}^3}))(\mu_{n,j}-\frac{1}{2}\log j)}\Big(1+\mathcal{O}(\frac{1}{n})\Big) \\ &= e^{t\sigma_{n,j}+\frac{t^2}{2!}}\Big(1+\mathcal{O}(\frac{\log\log n}{\sqrt{\log n}})\Big), \end{aligned} \tag{6.67}$$

where we have used (6.64) and

$$\frac{\Gamma(e^{\frac{t}{\sigma_{n,j}}}+j-\frac{1}{2})}{\Gamma(\frac{e^{\frac{t}{\sigma_{n,j}}}}{2}+j)} = 1+\mathcal{O}(\frac{\log\log n}{\sqrt{\log n}}). \tag{6.68}$$

This leads to

$$\mathcal{M}_{n,j}(t) = e^{-\sigma_{n,j}t}p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}) = e^{\frac{t^2}{2}}\Big(1+\mathcal{O}(\frac{\log\log n}{\sqrt{\log n}}\Big). \tag{6.69}$$

Thus for $1 \le j < n$ the moment generating function $\mathcal{M}_{n,j}(t)$ of $\Delta^*_{n,j}$ converges in a real neighborhood of $t=0$ to the moment generating function $e^{\frac{t^2}{2}}$ of the standard normal distribution. The continuity theorem of Lévy shows thus convergence in distribution of $\Delta^*_{n,j}$ to a Gaussian distributed random variable.

Now we will sketch the proof of the general case $c_1 \neq -2c_2$. We set, as stated in Theorem 35, $\mu_{n,j} := \big(1+\frac{c_2}{c_1}\big)(\log n+\log j)$ and $\sigma^2_{n,j} := \big(1+\frac{c_2}{c_1}\big)(\log n + \log j)$. The probability generating function is given by

$$\begin{aligned} p_{n,j}(v) = \big(1+\frac{c_2}{c_1}\big)&\frac{v\binom{n-2+v\left(1+\frac{c_2}{c_1}\right)}{n-j-1}}{(n-1)\binom{n-1+\frac{c_2}{c_1}}{n-1}\binom{n-2}{j-1}}\Bigg(\big(1-\frac{vc_1}{(c_1+c_2)(2v-1)-c_2}\big)\binom{j-1+\frac{c_2}{c_1}}{j-1} \\ &+\frac{vc_1}{(c_1+c_2)(2v-1)-c_2}\binom{j-1+(\frac{c_2}{c_1}+1)(2v-1)}{j-1}\Bigg), \end{aligned} \tag{6.70}$$

and the moment generating function $\mathcal{M}_{n,j}(t)$ of $\Delta^*_{n,j} := (\Delta_{n,j}-\mu_{n,j})/\sigma_{n,j}$ is given by

$$\mathcal{M}_{n,j}(t) := \mathbb{E}(e^{t\Delta^*_{n,j}}) = e^{-\frac{\mu_{n,j}}{\sigma_{n,j}}t}\mathbb{E}(e^{\frac{\Delta_{n,j}}{\sigma_{n,j}}t}) = e^{-\frac{\mu_{n,j}}{\sigma_{n,j}}t}p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}). \tag{6.71}$$

We split the region $1 \le j < n$ again into two cases j big, such that $j \ge \log n$, and j small, such that $j \le \log n$. Since it holds that

$$\frac{e^{\frac{t}{\sigma_{n,j}}}c_1}{(c_1+c_2)(2e^{\frac{t}{\sigma_{n,j}}}-1)-c_2} = 1+\mathcal{O}(\frac{1}{\sqrt{\log n}}), \tag{6.72}$$

it can be seen that only the second summand of $p_{n,j}(e^{\frac{t}{\sigma_{n,j}}})$, as given by (6.70), gives a main contribution to the asymptotic behavior. Writing $p_{n,j}(e^{\frac{t}{\sigma_{n,j}}})$ as

$$\begin{aligned} p_{n,j}(v) = \frac{\Gamma(n-1+v(\frac{c_2}{c_1}+1))\Gamma(2+\frac{c_2}{c_1})}{\Gamma(v+j(\frac{c_2}{c_1}+1))\Gamma(n+\frac{c_2}{c_1})}&\Bigg(\big(1-\frac{e^{\frac{t}{\sigma_{n,j}}}c_1}{(c_1+c_2)(2e^{\frac{t}{\sigma_{n,j}}}-1)-c_2}\big)\frac{\Gamma(j+\frac{c_2}{c_1})}{\Gamma(1+\frac{c_2}{c_1})} \\ &+\frac{e^{\frac{t}{\sigma_{n,j}}}c_1}{(c_1+c_2)(2e^{\frac{t}{\sigma_{n,j}}}-1)-c_2}\frac{\Gamma(j+(2v-1)(\frac{c_2}{c_1}+1))}{\Gamma(1+(2v-1)(\frac{c_2}{c_1}+1))}\Bigg), \end{aligned} \tag{6.73}$$

using Stirling's formula for the Gamma function (6.61) and proceeding as in the proof of the special case $c_1 = -2c_2$ covered by Corollary 13, leads then to Theorem 37.

Remark 16. Note that an alternative proof of the limit law of $\Delta_{n,j}$ is available by following very closely the approach of Dobrow and Smythe [18]. They have already outlined the benefits of using passion approximation there, we will sketch how their proof for recursive trees can be extended to the general case.

6.6.1 Poisson approximation of $\Delta_{n,j}$

As already stated in the preliminary part of this thesis, the total variation distance d_{TV} of two probability measures P and Q over $\mathbb{Z}_+$ is defined by

$$d_{\mathrm{TV}}(P,Q) = \frac{1}{2}\sum_{k\geq 0} |P(\{k\}) - Q(\{k\})|. \tag{6.74}$$

We denote with $\mathrm{Po}(\lambda)$ a probability distribution of a Poisson distributed random variable with parameter λ. Further we use the notation $\mathcal{L}(X)$ for the distribution law of the r. v. X. Let X_n be a sequence of random variables for which

$$d_{\mathrm{TV}}(\mathcal{L}(X_n), \mathrm{Po}(\lambda_n)) \to 0 \quad \text{and} \quad \lambda_n \to \infty, \tag{6.75}$$

then it holds

$$\frac{X_n - \lambda_n}{\sqrt{\lambda_n}} \xrightarrow{(d)} \mathcal{N}(0,1). \tag{6.76}$$

We will use a general setup based on [4]. For a finite or countable index set I and every $k \in I$ let X_k denote a Bernoulli random variable with $p_k = \mathbb{P}\{X_k = 1\} > 0$. Further we set $Y = \sum_{k\in I} X_k$ and $\lambda = \mathbb{E}(Y)$. For each $k \in I$ we choose $U_k \subset I$ such that $k \in U_k$ and if $i \notin U_k$ then X_k and X_i are almost independent. Define $b_1 = \sum_{k\in I}\sum_{i\in U_k} p_k p_i$, $b_2 = \sum_{k\in I}\sum_{i\in U_k, i\neq k} p_{ki}$ and $b_3 = \sum_{k\in I} \mathbb{E}\big(|\mathbb{E}(X_k - p_k | X_i : i \notin U_k|\big)$.

Theorem 38 (Arratia, et al. [4])**.**

$$d_{\mathrm{TV}}(\mathcal{L}(Y), \mathrm{Po}(\lambda)) \leq 2\Big((b_1 + b_2)\frac{1-e^{-\lambda}}{\lambda} + b_3 \min\{1, \frac{1.4}{\sqrt{\lambda}}\}\Big)$$

When the X_k are independent , we may drop the factor 2 and $b_2 = b_3 = 0$. Further $U_k = \{k\}$ and thus $b_1 = \sum_{k\in I} p_k^2$. For grown simple trees admitting $c_1 = -2c_2$ we use Theorem 38. For the other grown simple trees we need the following result.

Theorem 39. *Let Y have the mixed Poisson distribution* $\mathrm{Po}(\Lambda)$*, where Λ denotes a random variable. For $\lambda > 0$ it holds:*

$$d_{\mathrm{TV}}(\mathcal{L}(Y), \mathrm{Po}(\lambda)) \leq \frac{1-e^{-\lambda}}{\lambda}\mathbb{V}(\Lambda).$$

We also need the following lemma.

Lemma 16. *Let X, Y, Z, W be random variables such that W, X are independent and Y, Z are independent. Then*

$$d_{\mathrm{TV}}(\mathcal{L}(W+X), \mathcal{L}(Y+Z)) \leq d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(Y)) + d_{\mathrm{TV}}(\mathcal{L}(X), \mathcal{L}(Z)). \tag{6.77}$$

Let's consider first the case $c_1 = -2c_2$ (plane oriented recursive trees). From Corollary 10 and Theorem 36 we get

$$\Delta_{n,j} \stackrel{(d)}{=} \sum_{k=1}^{j} \tilde{B}_k + \sum_{k=j+1}^{n-1} B_k, \quad \text{for } j < n \tag{6.78}$$

$\tilde{B}_1 \stackrel{(d)}{=} 1$, $\tilde{B}_k \stackrel{(d)}{=} \mathrm{Be}(\frac{1}{k-\frac{1}{2}})$ for $2 \leq k \leq j$ and $B_k \stackrel{(d)}{=} \mathrm{Be}(\frac{\frac{1}{2}}{k-\frac{1}{2}})$ for $j+1 \leq k \leq n-1$. If we set $\lambda_n = \mathbb{E}(\Delta_{n,j}) = H_{2n-2} - \frac{1}{2}H_{n-1} + H_{2j} - \frac{1}{2}H_j - 1$ we get by using Theorem 38 the result

Theorem 40. *For plane oriented recursive trees and grown simple trees admitting $c_1 = -2c_2$ it holds for all $1 \leq j \leq n-1$:*

$$d_{\mathrm{TV}}(\mathcal{L}(\Delta_{n,j}), \mathrm{Po}(\lambda_n)) \leq \frac{1}{\lambda_n}\Big(\frac{\pi^2}{6}+1\Big) \leq \frac{2}{\log n + \log j}\Big(\frac{\pi^2}{6}+1\Big). \tag{6.79}$$

In the other case $c_1 \neq -2c_2$ one would need two estimates to handle the full range of j. They can be obtained as in [18]. This also leads then by using Lemma 16 to the desired result.

Chapter 7

Scale free trees

There has been much interest in using random graphs to model complex real-world networks such as the world-wide web. Barabási and Albert noticed that in many real-world examples the degree sequence has a scale-free power law distribution: the fraction $P(d)$ of vertices with degree d is proportional over a large range to d^γ, where γ is a constant independent of the size of the network. To explain this phenomenon, Barabási and Albert [5] suggested the following random graph process as a model:

> Starting with a small number m_0 of vertices, at every time step we add a new vertex with $m \leq m_0$ edges that link the new vertex to m different vertices already present in the system. To incorporate preferential attachment, we assume that the probability p that a new vertex will be connected to a vertex i depends on the connectivity k_v of that vertex, so that $p(k_i) = k_i / \sum_j k_j$. After t steps the model leads to a random network with $t + m_0$ vertices and mt edges.

The random graph model of Barabási and Albert can be applied to model the growth of the world-wide web, see [5]. As already known for $m = 1$ this process is very similar to the plane oriented recursive tree introduced in [71], which is a well studied object in the "combinatorial world". As pointed out by Bollobás and Riordan in [7], the standard definition of the Barabási and Albert model for $m = 1$ treats the root differently, where the branches are plane oriented recursive trees. A further generalization for the case $m = 1$ was introduced by Móri in [59] and can be defined by the following model. Starting with a non-decreasing weight sequence $(\tilde{\omega}(k))_{k\in\mathbb{N}}$ we attach node $i + 1$ to node $v \in T$ of degree $d(v)$, where T is a size i tree, with a probability $\tilde{p}(v)$ proportional to $\tilde{\omega}(d(v))$:

$$\tilde{p}(v) = \frac{\tilde{\omega}(d(v))}{\sum_{u\in T} \tilde{\omega}(d(u))} \tag{7.1}$$

Again this model (in the strict sense) treats the root differently compared to generalized plane oriented recursive trees. We will show how to reduce the trees grown according to (7.1) to generalized plane oriented recursive trees. To do so we will show how the trees defined by Móri are related to the class of non-simple increasing trees. We will use the abbreviations PORTs for plane oriented recursive trees and GPORTs for generalized plane oriented recursive trees.

7.1 Combinatorial description of scale free tree families

Formally, the class $\tilde{\mathcal{T}}$ of scale free trees falls into a broader class of non-simple increasing trees, which can be defined in the following way. Two sequences of non-negative numbers $(\tilde{\varphi}_k)_{k\geq 0}$ and $(\varphi_k)_{k\geq 0}$ with $\tilde{\varphi}_0 > 0$ and $\varphi_0 > 0$ (we further assume that there exists a $k \geq 2$ with $\varphi_k > 0$) are used to define the weight $w(T)$ of any ordered tree T rooted at node r by

$$\tilde{w}(T) := \tilde{\varphi}_{d(r)} \prod_{\substack{v \in T \\ v \neq r}} \varphi_{d(v)-1} = \tilde{\varphi}_{d^+(r)} \prod_{\substack{v \in T \\ v \neq r}} \varphi_{d^+(v)}, \tag{7.2}$$

where v ranges over all vertices of T, $d(v)$ denotes the degree and $d^+(v)$ the out-degree of v. 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 $\tilde{\mathcal{T}}$ consists of all trees T together with their weights $\tilde{w}(T)$ and the set of increasing labellings $\mathcal{L}(T)$.

For given degree-weight sequences $(\tilde{\varphi}_k)_{k\geq 0}$ and $(\varphi_k)_{k\geq 0}$ with the degree-weight generating functions $\tilde{\varphi}(t) := \sum_{k\geq 0} \tilde{\varphi}_k t^k$ and $\varphi(t) := \sum_{k\geq 0} \varphi_k t^k$, we define now the total weights by $\tilde{T}_n := \sum_{|T|=n} \tilde{w}(T) \cdot L(T)$. It follows then that the exponential generating function $\tilde{T}(z) := \sum_{n\geq 1} \tilde{T}_n \frac{z^n}{n!}$ satisfies

$$\tilde{T}'(z) = \tilde{\varphi}\big(T(z)\big), \quad \tilde{T}(0) = 0, \tag{7.3}$$

where

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

and $T(z)$ is the exponential generating function of a simple increasing tree family. We can also describe a scale free increasing trees $\tilde{\mathcal{T}}$ by the formal recursive equation, where $\mathcal{T}$ denotes the corresponding GPORT family

$$\tilde{\mathcal{T}} = ① \times \Big(\tilde{\varphi}_0 \cdot \{\epsilon\} \,\dot{\cup}\, \tilde{\varphi}_1 \cdot \mathcal{T} \,\dot{\cup}\, \tilde{\varphi}_2 \cdot \mathcal{T} * \mathcal{T} \,\dot{\cup}\, \tilde{\varphi}_3 \cdot \mathcal{T} * \mathcal{T} * \mathcal{T} \,\dot{\cup}\, \cdots\Big) = ① \times \tilde{\varphi}(\mathcal{T}), \tag{7.4}$$

where again ① denotes the node labelled by 1, $\times$ the cartesian product, $*$ the partition product for labelled objects, and $\tilde{\varphi}(\mathcal{T})$ the substituted structure (see, e. g., [72]).

7.1.1 Degree weight generating functions

Now we show which degree generating functions are corresponding to the scale free tree model as introduced by Móri in [59]. Starting with a non-decreasing weight sequence $(\tilde{\omega}(k))_{k\in\mathbb{N}}$ we attach node $i+1$ to node $v \in T$ of degree $d(v)$, where T is a size i tree, with a probability $\tilde{p}(v)$ proportional to $\tilde{\omega}(d(v))$:

$$\tilde{\omega}(d(v)) = \frac{(d(v)+1)\tilde{\varphi}_{d(v)+1}}{\tilde{\varphi}_{d(v)}}, \quad \tilde{p}(v) = \frac{\tilde{\omega}(d(v))}{\sum_{u\in T} \tilde{\omega}(d(u))}$$

We will investigate the weight function $\tilde{\omega}(k) = k + \beta$, with $\beta > -1$. Since in a size i tree T the sum $\sum_{u\in T} d(u) = 2(i-1)$ we get for $\tilde{\omega}(k) = k + \beta$

$$\tilde{p}(v) = \frac{d(v) + \beta}{2(i-1) + i\beta}. \tag{7.5}$$

The difference between the scale free trees defined by this model and the GPORTs is that for the scale free trees the probability $\tilde{p}(v)$ depends on the degree $d(v)$, where for grown increasing trees $p(v)$ depends on the out-degree $d^+(v)$. Since for all nodes in $v \in T$ except the root it holds $d^+(v) + 1 = d(v)$ we can naturally reduce the study of this model to the study of generalized plane oriented recursive trees, since only the root of such trees behaves differently to GPORTs. The parameter $\alpha > 0$ of the probabilistic description of GPORTs corresponds to the parameter β in the obvious relation

$$\alpha = \beta + 1, \tag{7.6}$$

e.g. for $\beta = 0$ we attach ordinary PORTs, $\alpha = 1$, to the root. The weight of the GPORTs $\omega(k)$ fulfills

$$\omega(k) = \tilde{\omega}(k+1), \quad \text{or equivalently } \frac{\varphi_k k}{\varphi_{k-1}} = \frac{(k+1)\tilde{\varphi}_{k+1}}{\tilde{\varphi}_k} \text{ for } k \geq 1, \tag{7.7}$$

with $\tilde{\varphi}_1 = \tilde{\varphi}_0 = 1$. For the GPORTs with degree weight generating function $\varphi(t) = (1-t)^{-\alpha}$, (where $\alpha = -\frac{c_1}{c_2} - 1 > 0$), we have $\varphi_k = \binom{k+\alpha-1}{k}$. This leads to the recurrence

$$\tilde{\varphi}_{k+1} = \frac{k\binom{k+\alpha-1}{k}}{(k+1)\binom{k+\alpha-2}{k-1}}\tilde{\varphi}_k = \frac{k+\alpha-1}{k+1}\tilde{\varphi}_k \text{ for } k \geq 1. \tag{7.8}$$

Iteration leads to

$$\tilde{\varphi}_k = \frac{1}{k}\binom{k+\alpha-2}{k-1}, \quad \text{for } k \geq 2. \tag{7.9}$$

We finally get the degree weight generating functions $\tilde{\varphi}(t)$.

$$\tilde{\varphi}(t) = 1 + \int_0^t \varphi(x)dx = 1 + \int_0^t \frac{1}{(1-t)^\alpha} = \begin{cases} 1 + \log\left(\frac{1}{1-t}\right), & \alpha = 1, \\ 1 + \frac{1}{\alpha-1}\left(\frac{1}{(1-z)^{\alpha-1}} - 1\right), & \alpha \neq 1. \end{cases} \tag{7.10}$$

Proposition 5. *The class of scale-free trees can be described as a non-simple increasing tree family, where the degree weight generating function $\tilde{\varphi}(t)$ is given by*

$$\tilde{\varphi}(t) = 1 + \int_0^t \varphi(x)dx, \tag{7.11}$$

with $\varphi(t) = (1-t)^{-\alpha}$.

By using (7.3) and $T(z)$ covered by Case C of the grown simple families of increasing trees, where $\alpha = -\frac{c_1}{c_2} - 1$, we obtain for the generating functions of the total weights the following result.

$$\tilde{T}(z) = \int_0^z \left(1 - \frac{1}{2}\log(1-2t)\right)dt = \frac{1}{4}(2z-1)\log\left(\frac{1}{1-2z}\right) + \frac{3}{2}z, \quad \alpha = 1, \tag{7.12}$$

$$\tilde{T}(z) = \int_0^z \left(1 - \frac{1}{\alpha-1} + \frac{1}{(\alpha-1)(1-(\alpha+1)t)^{\frac{\alpha-1}{\alpha+1}}}\right)dt = \frac{\alpha-2}{\alpha-1}z - \frac{(1-(\alpha+1)z)^{\frac{2}{(\alpha+1)}}}{2(\alpha-1)}, \quad \alpha \neq 1. \tag{7.13}$$

This leads to the following Corollary.

Corollar 14. *The generating function of the Scale free trees with weight function $\tilde{\omega}(k) = k + \alpha$, $\alpha > -1$ is given by*

$$\tilde{T}(z) = \begin{cases} \frac{1}{4}(2z-1)\log\left(\frac{1}{1-2z}\right) + \frac{3}{2}z, & \alpha = 1, \\ \frac{\alpha-2}{\alpha-1}z - \frac{(1-(\alpha+1)z)^{\frac{2}{(\alpha+1)}}}{2(\alpha-1)}, & \alpha \neq 1. \end{cases} \tag{7.14}$$

Further the total weights $\tilde{T}_n$ are given as follows:

$$\tilde{T}_n = \begin{cases} 1, & n = 1, \\ \prod_{k=2}^{n-1}(k(\alpha+1) - 2), & n > 1. \end{cases} \tag{7.15}$$

7.1.2 Characterization of the admissibly tree families

Lemma 17. *The following two properties of scale free families of increasing trees $\tilde{\mathcal{T}}$ are equivalent:*

1. *Starting with a random scale free increasing tree T of size $n \geq j$ of $\tilde{\mathcal{T}}$ and removing all nodes larger than j we obtain a random scale free increasing tree T' of size j of $\tilde{\mathcal{T}}$.*

2. *The family $\tilde{\mathcal{T}}$ can be constructed via an insertion process resp. a probabilistic growth rule.*

7.1.3 Randomness preserving property

In order to find such scale free families of increasing trees we consider families of non-simple increasing trees $\mathcal{T}$ with the property that when starting with a random tree of $\mathcal{T}$ of size n and removing all nodes larger than j, we obtain a random tree of $\mathcal{T}$ of size j. By iterating the argument, it is sufficient to show that after removing node n in a random size-n tree we get a random tree of $\tilde{\mathcal{T}}$ of size $n-1$. This randomness preserving property can be described easily via the equation

$$\frac{\tilde{w}(T')}{\tilde{w}(T'')} = \frac{\sum_{T:|T|=n, T \xrightarrow{\text{n}} T'} \tilde{w}(T)}{\sum_{T:|T|=n, T \xrightarrow{\text{n}} T''} \tilde{w}(T)}, \tag{7.16}$$

which must hold for all ordered trees T', T'' of size $|T'| = |T''| = n-1$ and ordered trees T of size $|T| = n$. Here, $T \xrightarrow{\text{n}} T'$ describes the fact that by cutting off node n from the tree T we get the tree T'. We assume now that T' is obtained from T by cutting off node n, which was originally attached at node v. Let r' denote the root of the tree T'. If the root r' has degree $d(r')$, there are $d(r')+1$ different trees T that lead to the same tree T' when cutting off node n. In contrast if $d(v)$ is the degree of node v, $v \neq r'$, in the tree T', hence the out degree of v is $d^+(v) = d(v) - 1$, there are $d(v)$ different trees T that lead to the same tree T' when cutting off node n. We obtain then the equivalent characterization

$$\frac{\tilde{w}(T')}{\tilde{w}(T'')} = \frac{\tilde{\varphi}_{d(r')} \prod_{\substack{v\in T'\\ v\neq r'}} \varphi_{d(v)}}{\tilde{\varphi}_{d(r'')} \prod_{\substack{v\in T''\\ v\neq r''}} \varphi_{d(v)}} = \frac{\tilde{\varphi}_{d(r')} \prod_{\substack{v\in T'\\ v\neq r'}} \varphi_{d(v)} \left(\frac{(d(r')+1)\tilde{\varphi}_{d(r')+1}}{\tilde{\varphi}_{d(r')}} + \sum_{\substack{v\in T'\\ v\neq r'}} \frac{d(v)\varphi_{d(v)}}{\varphi_{d(v)-1}} \right)}{\tilde{\varphi}_{d(r'')} \prod_{\substack{v\in T''\\ v\neq r''}} \varphi_{d(v)} \left(\frac{(d(r'')+1)\tilde{\varphi}_{d(r'')+1}}{\tilde{\varphi}_{d(r'')}} + \sum_{\substack{v\in T''\\ v\neq r''}} \frac{d(v)\varphi_{d(v)}}{\varphi_{d(v)-1}} \right)}, \tag{7.17}$$

and further

$$\frac{(d(r')+1)\tilde{\varphi}_{d(r')+1}}{\tilde{\varphi}_{d(r')}} + \sum_{\substack{v\in T'\\ v\neq r'}} \frac{d(v)\varphi_{d(v)}}{\varphi_{d(v)-1}} = \frac{(d(r'')+1)\tilde{\varphi}_{d(r'')+1}}{\tilde{\varphi}_{d(r'')}} + \sum_{\substack{v\in T''\\ v\neq r''}} \frac{d(v)\varphi_{d(v)}}{\varphi_{d(v)-1}} \tag{7.18}$$

7.1.4 Probabilistic growth rule

We study non-simple increasing trees which can be constructed by an *insertion process* or a *probabilistic growth rule*. Such trees have the property that for every tree T of size $n-1$ with vertices $v_1, \dots, v_{n-1}$ there exist probabilities $p_T(v_1), \dots, p_T(v_{n-1})$, such that when starting with a random tree T of size $n-1$, choosing a vertex v_i in T according to the probabilities $p_T(v_i)$, i. e., $\sum_{i=1}^{n-1} p_T(v_i) = 1$, and attaching node n to it at one of the $d^+(v_i)+1 = d(v_i)$ possible positions (which must be all equally likely due to the "symmetric" recursive description of increasing tree families), we obtain a random non-simple increasing tree T of the family $\mathcal{T}$ of size n.

We start with two given trees T' and T'' of size $|T'| = |T''| = n-1$ with weights $\tilde{w}(T')$ and $\tilde{w}(T'')$, nodes $v_1', \dots, v_{n-1}'$ and $v_1'', \dots, v_{n-1}''$, respectively, with probabilities $p_{T'}(v_1'), \dots, p_{T'}(v_{n-1}')$ and $p_{T''}(v_1''), \dots, p_{T''}(v_{n-1}'')$ fulfilling $\sum_{i=1}^{n-1} p_{T'}(v_i') = \sum_{i=1}^{n-1} p_{T''}(v_i'') = 1$. After attaching node n to a vertex $v_k' \in T'$ and $v_l'' \in T''$ one has to distinguish if any of the nodes v_k' or v_l'' is the root. If e.g. $v_k' = r'$ node n has $d(r)+1$ positions to join the tree whereas if $v_k' \neq r'$ there are only $d(v_k')$ positions. One obtains in any case trees T_k' and T_l'' which have weights

$$\begin{aligned} \tilde{w}(T_k') &= \varphi_0 \frac{\varphi_{d(v_k')+1}}{\varphi_{d(v_k')}} \tilde{\varphi}_{d(r')} \prod_{\substack{v\in T'\\ v\neq r'}} \varphi_{d(v)} \quad \text{if } v_k' \neq r', \quad \text{else} \quad \tilde{w}(T_{r'}') = \varphi_0 \frac{\tilde{\varphi}_{d(r')+1}}{\tilde{\varphi}_{d(r')}} \tilde{\varphi}_{d(r')} \prod_{\substack{v\in T'\\ v\neq r'}} \varphi_{d(v)}; \\ \tilde{w}(T_l'') &= \varphi_0 \frac{\varphi_{d(v_l')+1}}{\varphi_{d(v_l')}} \tilde{\varphi}_{d(r'')} \prod_{\substack{v\in T''\\ v\neq r''}} \varphi_{d(v)} \quad \text{if } v_l'' \neq r'', \quad \text{else} \quad \tilde{w}(T_{r''}'') = \varphi_0 \frac{\tilde{\varphi}_{d(r'')+1}}{\tilde{\varphi}_{d(r'')}} \tilde{\varphi}_{d(r'')} \prod_{\substack{v\in T''\\ v\neq r''}} \varphi_{d(v)} \end{aligned} \tag{7.19}$$

When starting with random non-simple increasing trees T' and T'' of size $|T'| = |T''| = n-1$, which are chosen with probability proportional to their weights, we obtain the following probabilities that the trees T'_k resp. T''_l are obtained by the insertion process:

$$\mathfrak{w}(T'_k) = \frac{p_{T'}(v'_k)}{d(v'_k)} \frac{\tilde{\varphi}_{d(r')} \prod_{\substack{v\in T' \\ v\neq r'}} \varphi_{d(v)}}{\tilde{T}_{n-1}} \quad \text{if } v'_k \neq r', \quad \text{else} \quad \mathfrak{w}(T'_{r'}) = \frac{p_{T'}(r')}{d(r')+1} \frac{\tilde{\varphi}_{d(r')} \prod_{\substack{v\in T' \\ v\neq r'}} \varphi_{d(v)}}{\tilde{T}_{n-1}}$$
$$\mathfrak{w}(T''_l) = \frac{p_{T''}(v''_l)}{d(v''_l)} \frac{\tilde{\varphi}_{d(r'')} \prod_{\substack{v\in T''' \\ v\neq r'}} \varphi_{d(v)}}{\tilde{T}_{n-1}} \quad \text{if } v''_l \neq r'', \quad \text{else} \quad \mathfrak{w}(T''_{r''}) = \frac{p_{T''}(r'')}{d(r'')+1} \frac{\tilde{\varphi}_{d(r'')} \prod_{\substack{v\in T'' \\ v\neq r''}} \varphi_{d(v)}}{\tilde{T}_{n-1}} \tag{7.20}$$

Since the resulting trees T'_k and T''_l must be random increasing trees of $\mathcal{T}$ of size n it must hold that

$$\frac{\tilde{w}(T'_k)}{\tilde{w}(T''_l)} = \frac{\mathfrak{w}(T'_k)}{\mathfrak{w}(T''_l)}, \quad \frac{\tilde{w}(T'_{r'})}{\tilde{w}(T''_l)} = \frac{\mathfrak{w}(T'_{r'})}{\mathfrak{w}(T''_l)}, \quad \frac{\tilde{w}(T'_k)}{\tilde{w}(T''_{r''})} = \frac{\mathfrak{w}(T'_k)}{\mathfrak{w}(T''_{r''})}, \quad \frac{\tilde{w}(T'_{r'})}{\tilde{w}(T''_{r''})} = \frac{\mathfrak{w}(T'_{r'})}{\mathfrak{w}(T''_{r''})}. \tag{7.21}$$

Consequently we obtain

$$\begin{aligned} \frac{1}{p_{T'}(v'_k)} \frac{d(v'_k)\varphi_{d(v'_k)}}{\varphi_{d(v'_k)-1}} &= \frac{1}{p_{T''}(v''_l)} \frac{d(v''_l)\varphi_{d(v''_l)}}{\varphi_{d(v''_l)-1}} = \frac{1}{p_{T'}(r')} \frac{(d(r')+1)\varphi_{d(r')+1}}{\varphi_{d(r')}} \\ &= \frac{1}{p_{T''}(r'')} \frac{(d(r'')+1)\varphi_{d(r'')+1}}{\varphi_{d(r'')}} := c(n-1), \end{aligned} \tag{7.22}$$

for all trees T', T'' with size $|T'| = |T''| = n-1$, and all vertices $v'_k \in T'$ resp. $v''_l \in T''$ where $c(n)$ is a function depending only on n. Further one obtains

$$\frac{(d(r')+1)\varphi_{d(r')+1}}{\varphi_{d(r')}} = c(n-1)p_{T'}(r') = \frac{d(v'_k)\varphi_{d(v'_k)}}{\varphi_{d(v'_k)-1}} = c(n-1)p_{T'}(v'_k), \tag{7.23}$$

and by summing up over all nodes in T' and T'' and dividing by $c(n-1)$

$$\frac{(d(r')+1)\tilde{\varphi}_{d(r')+1}}{\tilde{\varphi}_{d(r')}} + \sum_{\substack{v\in T' \\ v\neq r'}} \frac{d(v)\varphi_{d(v)}}{\varphi_{d(v)-1}} = \frac{(d(r'')+1)\tilde{\varphi}_{d(r'')+1}}{\tilde{\varphi}_{d(r'')}} + \sum_{\substack{v\in T'' \\ v\neq r''}} \frac{d(v)\varphi_{d(v)}}{\varphi_{d(v)-1}}. \tag{7.24}$$

For all tree families satisfying this property we can simply define the probabilities $p_{T'}(v')$ as $\frac{1}{c(n-1)} \frac{d(v')\varphi_{d(v')}}{\varphi_{d(v')-1}}$ and $p_{T'}(r')$ as $\frac{1}{c(n-1)} \frac{(d(r')+1)\varphi_{d(r')+1}}{\varphi_{d(r')}}$. E.g. by choosing $\tilde{\varphi}(t) = 1 + \log\left(\frac{1}{1-t}\right)$ and $\varphi(t) = 1/(1-t)$ we get $p_{T'}(r') = \frac{d(r')}{c(n-1)}$ and $p_{T'}(v') = \frac{d(v')}{c(n-1)}$. Hence by $\sum_{v\in T'} d(v') = 2|T'| - 2 = 2(n-1) - 2$ we end at $c(n) = 2(n-1)$. This just corresponds to the weight function $\tilde{\omega}(d(v)) = d(v)$, i.e. $\beta = 0$ (and $\alpha = 1$).

7.2 Results

7.2.1 Results for the depth of node j

Theorem 41. *The depth of node j in a size $n \geq j$ random scale free tree admits the following distribution law.*

$$\tilde{D}_{n,j} \stackrel{(d)}{=} \bigoplus_{k=1}^{j-1} B_k, \tag{7.25}$$

where $B_k \overset{(d)}{=} \mathrm{Be}(p_k)$ *with* $p_1 = 1$ *and* $p_k = \frac{\alpha}{k(\alpha+1)-2}$ *for* $2 \le k \le j-1$. *Further* $\tilde{D}_{n,j} \overset{(d)}{=} \tilde{D}_{j,j}$ *since the distribution does not depend on* n. *The expectation and the variance of* $\tilde{D}_{n,j}$ *are given as follows.*

$$\begin{aligned}\mathbb{E}(\tilde{D}_{n,j}) &= 1 + \sum_{k=2}^{j-1} p_k = 1 + \sum_{k=2}^{j-1} \frac{\alpha}{k(\alpha+1)-2} = 1 + \frac{\alpha}{\alpha+1}\big(H_{j-1-\frac{2}{\alpha+1}} - H_{1-\frac{2}{\alpha+1}}\big),\\ \mathbb{V}(\tilde{D}_{n,j}) &= \sum_{k=1}^{j-1} p_k(1-p_k) = \sum_{k=2}^{j-1} \frac{\alpha}{k(\alpha+1)-2} - \sum_{k=2}^{j-1} \frac{\alpha^2}{\big(k(\alpha+1)-2\big)^2}\\ &= \frac{\alpha}{\alpha+1}\big(H_{j-1-\frac{2}{\alpha+1}} - H_{1-\frac{2}{\alpha+1}}\big) - \frac{\alpha^2}{(\alpha+1)^2}\big(H^{(2)}_{j-1-\frac{2}{\alpha+1}} - H^{(2)}_{1-\frac{2}{\alpha+1}}\big).\end{aligned} \tag{7.26}$$

For $\alpha = 1$ *this can be simplified to*

$$\begin{aligned}\mathbb{E}(\tilde{D}_{n,j}) &= 1 + \frac{1}{2}H_{j-2},\\ \mathbb{V}(\tilde{D}_{n,j}) &= \frac{1}{2}H_{j-2} - \frac{1}{2}H^{(2)}_{j-2}.\end{aligned} \tag{7.27}$$

Theorem 42. *The probabilities* $\mathbb{P}\{\tilde{D}_{n,j} = m\}$ *are for* $m \ge 1$ *given by*

$$\mathbb{P}\{\tilde{D}_{n,j} = m\} = \frac{(j-2)!(\alpha+1)^{j-2}}{\prod_{l=2}^{j-1}(l(\alpha+1)-2)}\big(\frac{\alpha}{\alpha+1}\big)^{m-1} \sum_{l=0}^{j-2} \binom{\frac{\alpha}{\alpha+1}+j-3-k}{j-2-k} \frac{\left[{k \atop m-1}\right]}{k!}, \tag{7.28}$$

where $\mathbb{P}\{\tilde{D}_{n,j} = 0\} = \delta_{j,1}$.

Theorem 43. *The centralized and normalized random variable* $\tilde{D}^\star_{n,j} = \frac{\tilde{D}_{n,j}-\mu_j}{\sigma_j}$ *is for* $n \ge j$ *and* $j \to \infty$ *asymptotically gaussian distributed,*

$$\tilde{D}^*_{n,j} := \frac{\tilde{D}_{n,j} - \mu_{n,j}}{\sigma_{n,j}} \xrightarrow{(d)} \mathcal{N}(0,1), \tag{7.29}$$

where $\mu_j = \frac{\alpha}{\alpha+1}\log j \sim \mathbb{E}(\tilde{D}_{n,j})$ *and* $\sigma_j^2 = \mu_j \sim \mathbb{V}(\tilde{D}_{n,j})$.

7.2.2 Results for the distances

Theorem 44. *The random variable* $\tilde{\Delta}_{n,j}$ *satisfies for* $1 \le j \le n-1$ *the following distribution law.*

$$\tilde{\Delta}_{n,j} \overset{(d)}{=} \tilde{\Delta}_{j+1,j} \oplus \bigoplus_{k=j+1}^{n-1} B_k, \tag{7.30}$$

where $B_k \overset{(d)}{=} \mathrm{Be}(p_k)$, $p_k = \alpha/(k(\alpha+1)-2)$ *for* $j+1 \le k \le n-1$

Theorem 45. *The random variable* $\tilde{\Delta}_{j+1,j}$ *satisfies for* $\alpha \ne 1$ *and* $j \ge 1$ *the following distribution law*

$$\tilde{\Delta}_{j+1,j} \overset{(d)}{=} \sum_{k=1}^{\tilde{\eta}_j} B_k, \tag{7.31}$$

$B_k \overset{(d)}{=} \mathrm{Be}(p_k)$ *with* $p_1 = p_2 = 1$ *and* $p_k = 2\alpha/(k(\alpha+1)-2)$ *for* $3 \le k \le j$; *the r. v.* η_j *is distributed as follows:* $\mathbb{P}\{\tilde{\eta}_j = 1\} = \alpha/(j(\alpha+1)-2)$, $\mathbb{P}\{\tilde{\eta}_j = m\} = (\alpha+1)/(j(\alpha+1)-2)$, $2 \le m \le j-1$ *and*

$\mathbb{P}\{\tilde{\eta}_j = j\} = 1/2$. *For* $\alpha = 1$ *it holds a more simply decomposition.*

$$\tilde{\Delta}_{j+1,j} \overset{(d)}{=} \bigoplus_{k=1}^{j} B_k, \tag{7.32}$$

where $B_k \overset{(d)}{=} \mathrm{Be}(p_k)$ *with* $p_1 = 1$, $p_2 = 1/2$ *and* $p_k = 1/(k-1)$ *for* $3 \le k \le j$.

Theorem 46. *The probability distribution of* $\tilde{\Delta}_{n,j}$ *is given by the following exact formulæ.*

$$\begin{aligned}\mathbb{P}(\tilde{\Delta}_{n,j} = m) = {}& \frac{(j-2)!(n-j)!\alpha(\alpha+1)^{n-3}}{\prod_{k=2}^{n-1}(k(\alpha+1)-2)} \Bigg[\frac{\alpha^{m-1}}{(\alpha+1)^{m-1}} \sum_{k=0}^{m-1} \sum_{l=0}^{n-j-1} \frac{\left[l \atop k \right]}{l!} \binom{\frac{\alpha}{\alpha+1} + n - 3 - l}{n-j-1-l} \times \\ & \times \Bigg(2^{m-1-k} \frac{\left[j-2 \atop m-2-k \right]}{(j-2)!} + \frac{\alpha}{2(\alpha+1)} \sum_{l=0}^{j-2} \frac{\left[l \atop m-2-k \right]}{l!} \Bigg) \Bigg] \\ & + \frac{(j-2)!(n-j)!\alpha(\alpha+1)^{n-3}}{\prod_{k=2}^{n-1}(k(\alpha+1)-2)} \Bigg[\sum_{k=0}^{m-2} \frac{\alpha^k}{(\alpha+1)^k} \Bigg(\sum_{l=0}^{n-j-1} \frac{\left[l \atop k \right]}{l!} \binom{\frac{\alpha}{\alpha+1} + n - 3 - l}{n-j-1-l} \Bigg) \times \\ & \times \sum_{r=0}^{m-2-k} \frac{\alpha+1}{1-\alpha} \Big(\frac{2\alpha}{\alpha-1} \Big)^r \Big(\frac{2\alpha}{\alpha+1} \Big)^{m-2-k-r} \Bigg(\sum_{l=0}^{j-2} \frac{\left[l \atop m-2-k-2 \right]}{l!} - \frac{\left[j-2 \atop m-2-k-r \right]}{l!} \Bigg) \Bigg].\end{aligned} \tag{7.33}$$

Theorem 47. *The expectation and the variance of* $\tilde{\Delta}_{n,j}$ *are given by the following exact formulæ.*

$$\begin{aligned}\mathbb{E}(\tilde{\Delta}_{n,j}) &= \frac{\alpha}{\alpha+1} \big(H_{n-3+\frac{2\alpha}{\alpha+1}} + H_{j-2+\frac{2\alpha}{\alpha+1}} - 2H_{\frac{2\alpha}{\alpha+1}} \big) + \frac{\alpha+2}{\alpha+1} + \frac{\alpha(\alpha-1)}{(\alpha+1)(j(\alpha+1)-2)}, \\ \mathbb{V}(\tilde{\Delta}_{n,j}) &= \frac{\alpha}{\alpha+1} H_{n-3+\frac{2\alpha}{\alpha+1}} + \big(\frac{\alpha}{\alpha+1} - \frac{4\alpha^2(\alpha-1)}{(\alpha+1)^2(j(\alpha+1)-2)} \big) H_{j-2+\frac{2\alpha}{\alpha+1}} \\ &\quad + \big(-\frac{2\alpha}{\alpha+1} + \frac{4\alpha^2(\alpha-1)}{(\alpha+1)^2(j(\alpha+1)-2)} \big) H_{\frac{2\alpha}{\alpha+1}} - \frac{\alpha^2}{(\alpha+1)^2} \Big(H^{(2)}_{n-3+\frac{2\alpha}{\alpha+1}} + 3H^{(2)}_{j-2+\frac{2\alpha}{\alpha+1}} - 4H^{(2)}_{\frac{2\alpha}{\alpha+1}} \Big) \\ &\quad + \frac{(2\alpha-1)\alpha}{(\alpha+1)^2} - \frac{(3\alpha+1)(\alpha-1)\alpha}{(j(\alpha+1)-2)(\alpha+1)^2} - \frac{(\alpha-1)^2\alpha^2}{(\alpha+1)^2(j(\alpha+1)-2)^2}.\end{aligned} \tag{7.34}$$

In the following we give the main theorem of this chapter, i. e. the central limit theorems for the r. v. $\tilde{\Delta}_{n,j}$ and $\tilde{\Delta}_{n;j_1,j_2}$, respectively.

Theorem 48. *The centralized and normalized random variable* $\tilde{\Delta}^*_{n,j}$, *where* $\tilde{\Delta}_{n,j}$ *counts the distance between the nodes with the label* j *and the label* n *in a randomly chosen size-*n *tree is, for arbitrary sequences* $(n, j(n))_{n\in\mathbb{N}}$, *with* $1 \le j = j(n) < n$, *asymptotically for* $n \to \infty$ *Gaussian distributed,*

$$\tilde{\Delta}^*_{n,j} := \frac{\Delta_{n,j} - \mu_{n,j}}{\sigma_{n,j}} \xrightarrow{(d)} \mathcal{N}(0,1), \tag{7.35}$$

where $\mu_{n,j} := \frac{\alpha}{\alpha+1}(\log n + \log j)$ *and* $\sigma^2_{n,j} := \frac{\alpha}{\alpha+1}(\log n + \log j)$.

Corollar 15. *The centralized and normalized random variable* $\tilde{\Delta}^*_{n;j_1,j_2}$, *where* $\Delta_{n;j_1,j_2}$ *counts the distance between the nodes with the label* j_1 *and the label* j_2 *in a randomly chosen size-*n *tree is, for arbitrary sequences* $(n, j_1(n), j_2(n))_{n\in\mathbb{N}}$, *with* $1 \le j_1 = j_1(n), j_2 = j_2(n) < n$, *provided that* $\max(j_1, j_2) \to \infty$, *asymptotically for* $n \to \infty$ *Gaussian distributed,*

$$\tilde{\Delta}^*_{n;j_1,j_2} := \frac{\Delta_{n,;j_1,j_2} - \mu_{n;j_1,j_2}}{\sigma_{n;j_1,j_2}} \xrightarrow{(d)} \mathcal{N}(0,1), \tag{7.36}$$

where $\mu_{n;j_1,j_2} := \frac{\alpha}{\alpha+1}(\log j_1 + \log j_2)$, $\sigma^2_{n;j_1,j_2} := \frac{\alpha}{\alpha+1}(\log j_1 + \log j_2)$.

7.3 Labelbased varieties

The combinatorial description allows us to provide a general scheme for labelbased varieties such as *subtree size of node* j, *degree of node* j or *depth of node* j. Let $\tilde{X}_{n,j}$ denote the random variable, which counts a certain variety depending on the label j in a random scale free tree of size n. Further let $X_{n,j}$ denote the r. v. counting the same variety in the corresponding GPORT. At first we introduce a bivariate generating function $\tilde{M}(z,v)$ for root $j=1$, which is defined as follows.

$$\tilde{M}(z,v) := \sum_{n\geq 1}\sum_{m\geq 0}\mathbb{P}\{\tilde{X}_{n,1}=m\}\tilde{T}_n\frac{z^n}{n!}v^m. \tag{7.37}$$

Further we set $n := k+j$ with $k\geq 0$ and define the trivariate generating function

$$\tilde{N}(z,u,v) := \sum_{k\geq 0}\sum_{j\geq 1}\sum_{m\geq 0}\mathbb{P}\{\tilde{X}_{k+j,j}=m\}\tilde{T}_{k+j}\frac{z^{j-1}}{(j-1)!}\frac{u^k}{k!}v^m. \tag{7.38}$$

Due to considerations done in [68] and [47] it holds the recurrence

$$\begin{aligned}\mathbb{P}\{\tilde{X}_{n,j}=m\} = \sum_{r\geq 1} r\tilde{\varphi}_r \sum_{\substack{k_1+\cdots+k_r=n-1,\\ k_1,\dots,k_r\geq 1}} \frac{T_{k_1}\cdots T_{k_r}}{\tilde{T}_n}\times \\ \times \sum_{i=1}^{\min\{k_1,j-1\}} \mathbb{P}\{X_{k_1,i}=m\}\binom{j-2}{i-1}\binom{n-j}{k_1-i}\binom{n-1-k_1}{k_2,k_3,\dots,k_r},\end{aligned} \tag{7.39}$$

for $n\geq j\geq 2$. We get

$$\frac{\partial}{\partial z}\tilde{N}(z,u,v) = \tilde{\varphi}'\big(T(z+u)\big)N(z,u,v) = \varphi\big(T(z+u)\big)N(z,u,v), \tag{7.40}$$

where we already know the shape of $N(z,u,v)$.

Lemma 18. *The generating function of a variety depending only on the subtree rooted at j for GPORTs is given by*

$$N(z,u,v) = \frac{\varphi\big(T(z+u)\big)\frac{\partial}{\partial u}M(u,v)}{\varphi\big(T(u)\big)}, \tag{7.41}$$

where

$$M(z,v) := \sum_{n\geq 1}\sum_{m\geq 0}\mathbb{P}\{X_{n,1}=m\}T_n\frac{z^n}{n!}v^m. \tag{7.42}$$

Note that (7.41) holds for varieties such as *subtree size of node* j, *degree of node* j, *subtrees of various sizes*, etc., where the variety counts a property depending only on the subtree rooted at node j. Integration gives

$$\begin{aligned}\tilde{N}(z,u,v) &= \int_0^z \tilde{\varphi}'\big(T(t+u)\big)N(t,u,v)dt + \frac{\partial}{\partial u}\tilde{M}(u,v)\\ &= \frac{\tilde{\varphi}\big(T(z+u)\big)\frac{\partial}{\partial u}M(u,v)}{\varphi\big(T(u)\big)} - \frac{\tilde{\varphi}\big(T(u)\big)\frac{\partial}{\partial u}M(u,v)}{\varphi\big(T(u)\big)} + \frac{\partial}{\partial u}\tilde{M}(u,v).\end{aligned} \tag{7.43}$$

Proposition 6. *The function $\tilde{N}(z,u,v)$ as defined in equation* (7.38)*, which is the trivariate generating function of the probabilities $\mathbb{P}\{\tilde{X}_{n,j}=m\}$ that give the probability that a certain variety of label j in a randomly chosen size-n scale free tree with degree-weight generating function $\tilde{\varphi}(t)$ is of size m, is given* *by the following formula:*

$$\tilde{N}(z,u,v) = \frac{\tilde{\varphi}\big(T(z+u)\big)\frac{\partial}{\partial u}M(u,v)}{\varphi\big(T(u)\big)} - \frac{\tilde{\varphi}\big(T(u)\big)\frac{\partial}{\partial u}M(u,v)}{\varphi\big(T(u)\big)} + \frac{\partial}{\partial u}\tilde{M}(u,v). \tag{7.44}$$

Note that this approach can be used to determine the behavior of $\tilde{X}_{n,j}$ over the full range of j. We will sketch the results concerning the node degree in the following subsection.

7.3.1 Node degree

Móri analyzed in [60] the maximal degree in a random size n scale free tree. Further he gives closed formulæ for the r. v. $\tilde{X}_{n,j}$ counting the degree of the node j in a size n scale free tree using a martingale approach. He also obtained the limit laws for $\tilde{X}_{n,j_1}\tilde{X}_{n,j_2}\dots\tilde{X}_{n,j_k}$, with $1 \le j_1 < j_2 < \dots < j_k$ fixed. The approach introduced in Section 7.3 enables to state the following auxiliary result (compare with [47]).

Theorem 49. *The probability distribution of $\tilde{X}_{n,j}$ is given as follows. For the root $j=1$ one gets*

$$\mathbb{P}\{\tilde{X}_{n,1}=m\} = \begin{cases} \dfrac{2(n-1)}{m}\displaystyle\sum_{k=0}^{m}\binom{m}{k}(-1)^k\binom{n-2-\frac{k}{2}}{n-1}, & \text{for } \alpha = 1,\\[2ex] \dfrac{(\alpha+1)\binom{m-2+\alpha}{m}}{\alpha-1}\displaystyle\sum_{k=0}^{m}\binom{m}{k}(-1)^k\frac{\Gamma(n-1+\frac{k}{\alpha+1})\Gamma(2-\frac{2}{\alpha+1})}{\Gamma(-\frac{k}{\alpha+1})\Gamma(n-\frac{2}{\alpha+1})}, & \text{for } \alpha \neq 1. \end{cases} \tag{7.45}$$

Further the s-th factorial moments $\mathbb{E}(\tilde{X}_{n,1}^{\underline{s}})$ are given as follows.

$$\mathbb{E}(\tilde{X}_{n,1}^{\underline{s}}) = \begin{cases} 2(n-1)\displaystyle\sum_{k=0}^{s}\binom{s}{k}(-1)^k\frac{\Gamma(s)\Gamma(n-1-\frac{k}{2})}{\Gamma(n)\Gamma(\frac{k}{2})}, & \text{for } \alpha = 1,\\[2ex] \dfrac{\Gamma(\alpha+s-1)}{\Gamma(\alpha-1)}\displaystyle\sum_{k=0}^{s}\binom{s}{k}(-1)^k\frac{\Gamma(n-\frac{s-k-2}{\alpha+1})\Gamma(1-\frac{2}{\alpha+1})}{\Gamma(1-\frac{s-k-2}{\alpha+1})\Gamma(n-\frac{2}{\alpha+1})}, & \text{for } \alpha \neq 1. \end{cases} \tag{7.46}$$

7.4 Depth of node j

Let $\tilde{D}_{n,j}$ denote the random variable which counts the depth of node j in a size n scale-free tree. Further let $D_{n,j}$ denote the corresponding random variable in GPORTs. For the *depth of node* j, Lemma 18 does not hold. Instead it holds the following result of [68]:

Lemma 19 (Panholzer & Prodinger, 2005+)**.** *For the family of GPORTs ($\varphi(t) = (1-t)^{-\alpha}$, $\varphi_0 = -c_2 = 1$, $\alpha = -1-\frac{c_1}{c_2}$), the generating function of the probabilities $\mathbb{P}\{D_{n,j}=m\}$ is given as follows.*

$$N(z,u,v) = \varphi\big(T(u)\big)\Big(\frac{\varphi\big(T(z+u)\big)}{\varphi\big(T(u)\big)}\Big)^v = \frac{1}{(1-2u)^{\frac{\alpha}{\alpha+1}}}\Big(\frac{1-2u}{1-2(z+u)}\Big)^{\frac{v\alpha}{\alpha+1}}. \tag{7.47}$$

The generating function $\tilde{N}(z,u,v)$ is given by

$$\frac{\partial}{\partial z}\tilde{N}(z,u,v) = v\tilde{\varphi}'(T(z+u))N(z,u,v). \tag{7.48}$$

Since

$$\tilde{N}(0,u,v) = \sum_{k\ge0}\sum_{m\ge0}\mathbb{P}\{\tilde{D}_{k+1,1}=m\}\tilde{T}_{k+1}\frac{u^k}{k!}v^m = \frac{\partial}{\partial u}\tilde{M}(u,v), \tag{7.49}$$

we get by using Lemma 19 the result

$$\tilde{N}(z,u,v) = \int_0^z v\varphi(T(t+u))N(t,u,v)dt + \tilde{\varphi}(T(u))$$
$$= \frac{v\big((v+1)\frac{\alpha}{\alpha+1}-1\big)(1-(\alpha+1)u)^{(v-1)\frac{\alpha}{\alpha+1}}}{(\alpha+1)\big(1-(\alpha+1)(z+u)\big)^{\frac{v\alpha-1}{\alpha+1}}} + \begin{cases} 1-\frac{1}{2}\log(1-2u), & \alpha = 1, \\ 1-\frac{1}{\alpha-1}+\frac{1}{(\alpha-1)(1-(\alpha+1)u)^{\frac{\alpha-1}{\alpha+1}}}, & \alpha \neq 1. \end{cases} \tag{7.50}$$

Due to computational reasons extracting coefficients from (7.48) is more easy. First we will show the following distribution law.

7.4.1 Proofs

Proof of Theorem 41. The probability generating function $\tilde{p}_{n,j}(v)$ can be obtained as follows.

$$\begin{aligned} \tilde{p}_{n,j}(v) &= \frac{(j-2)!(n-j)!}{\tilde{T}_n}[z^{j-2}u^{n-j}]\frac{\partial}{\partial z}\tilde{N}(z,u,v) \\ &= \frac{(j-2)!(n-j)!}{\tilde{T}_n}[z^{j-2}u^{n-j}]\frac{v}{(1-(\alpha+1)u)^{\frac{2\alpha}{\alpha+1}}\big(1-\frac{(\alpha+1)z}{1-(\alpha+1)u}\big)^{\frac{(v+1)\alpha}{\alpha+1}}} \\ &= \frac{(j-2)!(n-j)!(\alpha+1)^{n-2}}{\prod_{k=2}^{n-1}(k(\alpha+1)-2)}v\binom{\frac{(v+1)\alpha}{\alpha+1}+j-3}{j-2}\binom{\frac{2\alpha}{\alpha+1}+n-3}{n-j}. \end{aligned} \tag{7.51}$$

After simplification of the binomial coefficients one gets

$$\tilde{p}_{n,j}(v) = v\prod_{k=2}^{j-1}\big(\frac{k(\alpha+1)-2-\alpha}{k(\alpha+1)-2} + v\frac{\alpha}{k(\alpha+1)-2}\big), \tag{7.52}$$

which shows the distribution law. The expectation and the variance follow easily since the B_k's are independent. □

Proof of Theorem 42. We extract coefficients of $\tilde{p}_{n,j}(v)$.

$$[v^m]\tilde{p}_{n,j}(v) = \frac{(j-2)!(\alpha+1)^{j-2}}{\prod_{l=2}^{j-1}(l(\alpha+1)-2)}[v^{m-1}]\binom{\frac{(v+1)\alpha}{\alpha+1}+j-3}{j-2} = \frac{(j-2)!(\alpha+1)^{j-2}}{\prod_{l=2}^{j-1}(l(\alpha+1)-2)}[z^{j-2}v^{m-1}]\frac{1}{(1-z)^{\frac{(v+1)\alpha}{\alpha+1}}}. \tag{7.53}$$

Now the well known relation for the Stirling numbers of the first kind $\left[{n \atop m}\right]$

$$\sum_{n\geq 0}\sum_{m=0}^{n}\left[{n \atop m}\right]\frac{z^n}{n!}v^m = \frac{1}{(1-z)^v} \tag{7.54}$$

shows the shape of the probabilities $\mathbb{P}\{D_{n,j}=m\}$. □

The limit law can either be seen by applying Lévy's continuity theorem for the moment generating function of the normalized and centralized random variable $D^\star_{n,j}$ or by poisson approximation.

Proof of Theorem 43 - Poisson approximation. Using Theorem 38, see the end of Chapter 6, we proceed as follows. Let $\lambda_j = \mathbb{E}(\tilde{D}_{n,j})$.

$$d_{\mathrm{TV}}(\mathcal{L}(\tilde{D}_{n,j}), \mathrm{Po}(\lambda_j)) \leq \frac{1}{\lambda_n}\Big(\frac{\pi^2}{6}+1\Big) \leq \frac{1}{\log j}\Big(\frac{\pi^2}{6}+1\Big). \tag{7.55}$$

Hence we get

$$d_{\mathrm{TV}}(\mathcal{L}(\tilde{D}_{n,j}), \mathrm{Po}(\lambda_j)) \to 0 \quad \text{and} \quad \lambda_j \to \infty, \tag{7.56}$$

and consequently

$$\frac{\tilde{D}_{n,j} - \lambda_j}{\sqrt{\lambda_j}} \xrightarrow{(d)} \mathcal{N}(0,1). \tag{7.57}$$

$\square$

Proof of Theorem 43 - Direct proof. From the proof of Theorem 41 one obtains

$$p_{n,j}(v) = v(\alpha+1)^{j-2} \frac{\left(\frac{(v+1)\alpha}{\alpha+1} + j - 2\right)^{\underline{j-2}}}{\prod_{k=2}^{j-1}(k(\alpha+1)-2)} = v \frac{\Gamma(\frac{(v+1)\alpha}{\alpha+1} + j - 2)\Gamma(2 - \frac{2}{\alpha+1})}{\Gamma(\frac{(v+1)\alpha}{\alpha+1})\Gamma(j - \frac{2}{\alpha+1})}. \tag{7.58}$$

Let $\mu_j = \frac{\alpha}{\alpha+1}\log j \sim \mathbb{E}(\tilde{D}_{n,j})$ and $\sigma_j^2 = \mu_j \sim \mathbb{V}(\tilde{D}_{n,j})$. Then the moment generating function $\mathcal{M}_{n,j}(t)$ of $\tilde{D}^*_{n,j} := (\tilde{D}_{n,j} - \mu_j)/\sigma_j$ is given by

$$\mathcal{M}_{n,j}(t) := \mathbb{E}(e^{t\tilde{D}^*_{n,j}}) = e^{-\frac{\mu_{n,j}}{\sigma_{n,j}}t}\mathbb{E}(e^{\frac{\tilde{D}_{n,j}}{\sigma_{n,j}}t}) = e^{-\frac{\mu_{n,j}}{\sigma_{n,j}}t}p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}) = e^{-\sigma_{n,j}t}p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}). \tag{7.59}$$

The usage of Stirling's formula for the Gamma function

$$\Gamma(z) = \left(\frac{z}{e}\right)^z \frac{\sqrt{2\pi}}{\sqrt{z}}\left(1 + \frac{1}{12z} + \frac{1}{288z^2} + \mathcal{O}(\frac{1}{z^3})\right), \tag{7.60}$$

leads to the expansion

$$p_{n,j}(v) = vj^{\frac{(v+1)\alpha}{\alpha+1} - \frac{2\alpha}{\alpha+1}} \frac{\Gamma(\frac{2\alpha}{\alpha+1})}{\Gamma(\frac{(v+1)\alpha}{\alpha+1})}\left(1 + \mathcal{O}(\frac{1}{j})\right) = ve^{\frac{\alpha+1}{\alpha}\mu_j\left(\frac{(v+1)\alpha}{\alpha+1} - \frac{2\alpha}{\alpha+1}\right)} \frac{\Gamma(\frac{2\alpha}{\alpha+1})}{\Gamma(\frac{(v+1)\alpha}{\alpha+1})}\left(1 + \mathcal{O}(\frac{1}{j})\right). \tag{7.61}$$

Setting

$$v = e^{\frac{t}{\sigma_n}} = 1 + \frac{t}{\sigma_j} + \frac{t^2}{2\mu_j} + \mathcal{O}(\frac{1}{\sigma_j^3}) \tag{7.62}$$

one gets

$$p_{n,j}(e^{\frac{t}{\sigma_n}}) = e^{\frac{\alpha+1}{\alpha}\mu_j\left(\frac{t\alpha}{(\alpha+1)\sigma_j} + \frac{t^2\alpha}{\mu_j(\alpha+1)}\right)}\left(1 + \mathcal{O}(\frac{\log\log j}{\sqrt{\log j}})\right). \tag{7.63}$$

Hence

$$e^{-\sigma_{n,j}t}p_{n,j}(e^{\frac{t}{\sigma_{n,j}}}) = e^{-\sigma_{n,j}t}e^{\mu_j\left(\frac{t}{\sigma_j} + \frac{t^2}{\mu_j}\right)}\left(1 + \mathcal{O}(\frac{\log\log j}{\sqrt{\log j}})\right) = e^{\frac{t^2}{2}}\left(1 + \mathcal{O}(\frac{\log\log j}{\sqrt{\log j}})\right). \tag{7.64}$$

$\square$

7.5 Distances

Let $\tilde{\Delta}_{n,j}$ denote the random variable which counts the desistance between node n and node j in a random scale free tree of size n. Further $\tilde{D}_{n,j}$ denotes as before the depth of node j in a random scale free tree of size n. We denote with $\Delta_{n,j}$ and $D_{n,j}$ the corresponding random variables in GPORTS. We define the trivariate generating functions of $\tilde{\Delta}_{n,j}$ as follows.

$$\tilde{M}(z,u,v) := \sum_{k\geq 1}\sum_{j\geq 1}\sum_{m\geq 0}\mathbb{P}\{\tilde{\Delta}_{k+j,j} = m\}\tilde{T}_{k+j}\frac{z^{j-1}}{(j-1)!}\frac{u^{k-1}}{(k-1)!}v^m. \tag{7.65}$$

It holds the following relation:

$$\begin{aligned}\frac{\partial}{\partial z}\tilde{M}(z,u,v) &= \tilde{\varphi}'\big(T(z+u)\big)M(z,u,v) + v^2 N(z,u,v)N(z+u,0,v)\tilde{\varphi}''\big(T(z+u)\big),\\ &= \varphi\big(T(z+u)\big)M(z,u,v) + v^2 N(z,u,v)N(z+u,0,v)\varphi'\big(T(z+u)\big),\end{aligned} \tag{7.66}$$

where the corresponding generating functions of the GPORTs $M(z,u,v)$ is given by the following Lemma.

Lemma 20. *For GPORTs we get*

$$\begin{aligned}M(z,u,v) &= \frac{v\alpha\big(1-\frac{v(\alpha+1)}{\alpha(2v-1)+1}\big)}{(1-(\alpha+1)u)^{\frac{\alpha}{\alpha+1}(v-1)+1}(1-(\alpha+1)(z+u))^{\frac{\alpha}{\alpha+1}}}\\ &+ \frac{\alpha(\alpha+1)v^2(1-(\alpha+1)u)^{\frac{\alpha}{\alpha+1}(v-1)}}{(\alpha(2v-1)+1)(1-(\alpha+1)(z+u))^{\frac{\alpha}{\alpha+1}(2v-1)+1}};\end{aligned} \tag{7.67}$$

Note that for PORTs ($\alpha = 1$) the generating function of the probabilities $\mathbb{P}\{\Delta_{n,j} = m\}$ simplifies to

$$M(z,u,v) = \frac{v(1-2u)^{\frac{v-1}{2}}}{(1-2(z+u))^{v+\frac{1}{2}}}. \tag{7.68}$$

It is possible to explicitly calculate $\tilde{M}(z,u,v)$ via

$$\tilde{M}(z,u,v) = \int_0^z \Big(\tilde{\varphi}'\big(T(t+u)\big)M(t,u,v) + v^2 N(t,u,v)N(t+u,0,v)\tilde{\varphi}''\big(T(t+u)\big)\Big)dt + \frac{\partial}{\partial u}\tilde{N}(u,0,v), \tag{7.69}$$

but we omit the lengthy formula.

Proof of Theorem 44 and Theorem 45. The part $j = 1$ is already covered by Theorem 41. For obtaining the probability generating function $p_{n,j}(v) = p^{[1]}_{n,j}(v) + p^{[2]}_{n,j}(v)$ one extract coefficients from $\frac{\partial}{\partial z}\tilde{M}(z,u,v)$ given by (7.66) in two steps.

$$\begin{aligned}p^{[1]}_{n,j}(v) &= \frac{(j-2)!(n-j-1)!}{\tilde{T}_n}[z^{j-2}u^{n-j-1}]\varphi\big(T(z+u)\big)M(z,u,v)\\ &= \frac{(j-2)!(n-j-1)!}{\tilde{T}_n}[z^{j-2}u^{n-j-1}]\Big(\frac{v\alpha\big(1-\frac{v(\alpha+1)}{\alpha(2v-1)+1}\big)}{(1-(\alpha+1)u)^{\frac{\alpha}{\alpha+1}(v-1)+1}(1-(\alpha+1)(z+u))^{\frac{2\alpha}{\alpha+1}}}\\ &\quad+\frac{\alpha(\alpha+1)v^2(1-(\alpha+1)u)^{\frac{\alpha}{\alpha+1}(v-1)}}{(\alpha(2v-1)+1)(1-(\alpha+1)(z+u))^{\frac{\alpha}{\alpha+1}2v+1}}\Big).\end{aligned} \tag{7.70}$$

After reading off the coefficients one obtains

$$\begin{aligned}p^{[1]}_{n,j}(v) &= \frac{(j-2)!(n-j-1)!\alpha v(\alpha+1)^{n-3}}{\tilde{T}_n}\binom{\frac{(v+1)\alpha}{\alpha+1}+n-3}{n-j-1}\times\\ &\quad\times\Big[\big(1-\frac{v(\alpha+1)}{\alpha(2v-1)+1}\big)\binom{\frac{2v\alpha}{\alpha+1}+j-3}{j-2} + \frac{v(\alpha+1)}{\alpha(2v-1)+1}\binom{\frac{2v\alpha}{\alpha+1}+j-2}{j-2}\Big]\\ &= \Big(\prod_{k=j+1}^{n-1}\big(\frac{l(\alpha+1)-2-\alpha}{l(\alpha+1)-2} + v\frac{\alpha}{l(\alpha+1)-2}\big)\Big)p^{[1]}_{j+1,j}(v),\end{aligned} \tag{7.71}$$

where

$$p^{[1]}_{j+1,j}(v) = \frac{\alpha v}{j(\alpha+1)-2}\Big(1-\frac{v(\alpha+1)}{\alpha(2v-1)+1}\Big)$$

$$+\frac{v^2\alpha(\alpha+1)}{\big(j(\alpha+1)-2\big)\big(\alpha(2v-1)+1\big)}\prod_{k=3}^{j}\frac{l(\alpha+1)-2+2\alpha(v-1)}{(l-1)(\alpha+1)-2}. \quad (7.72)$$

The p.g.f. $p^{[1]}_{j+1,j}(v)$ can be further simplified as follows. By defining

$$a_j=\frac{1}{\big(\alpha(2v-1)+1\big)}\prod_{k=3}^{j}\frac{l(\alpha+1)-2+2\alpha(v-1)}{(l-1)(\alpha+1)-2}, \quad (7.73)$$

and using the recurrence

$$a_j=\Big(1+\frac{\alpha(2v-1)+1}{(j-1)(\alpha+1)-2}\Big)a_{j-1}, \quad (7.74)$$

one gets the relation

$$a_j=\sum_{k=3}^{j}\frac{\alpha(2v-1)+1}{(k-1)(\alpha+1)-2}a_{k-1}+a_2=\frac{1}{2\alpha}\sum_{k=2}^{j-1}\prod_{l=3}^{k}\frac{l(\alpha+1)-2+2\alpha(v-1)}{l(\alpha+1)-2}+\frac{1}{\alpha(2v-1)+1}. \quad (7.75)$$

This leads to

$$p^{[1]}_{j+1,j}(v) \;=\; \frac{\alpha v}{j(\alpha+1)-2}+\frac{v^2(\alpha+1)}{2\big(j(\alpha+1)-2\big)}\sum_{k=2}^{j-1}\prod_{l=3}^{k}\Big(\frac{l(\alpha+1)-2-2\alpha}{l(\alpha+1)-2}+v\frac{2\alpha}{l(\alpha+1)-2}\Big). \quad (7.76)$$

For $\alpha=1$ we get a simpler expression for $p^{[1]}_{j+1,j}(v)$:

$$p^{[1]}_{j+1,j}(v)=\frac{v}{2}\prod_{k=3}^{j}\Big(\frac{k(\alpha+1)-2-2\alpha}{k(\alpha+1)-2}+v\frac{2\alpha}{k(\alpha+1)-2}\Big) \quad (7.77)$$

For the second summand one obtains

$$\begin{aligned}p^{[2]}_{n,j}(v)&=\frac{(j-2)!(n-j-1)!}{\tilde T_n}[z^{j-2}u^{n-j-1}]v^2N(z,u,v)N(z+u,0,v)\varphi'\big(T(z+u)\big)\\&=\frac{(j-2)!(n-j-1)!}{\tilde T_n}[z^{j-2}u^{n-j-1}]\frac{v^2\alpha\big(1-(\alpha+1)u\big)^{\frac{(v-1)\alpha}{\alpha+1}}}{\big(1-(\alpha+1)(z+u)\big)^{\frac{2v\alpha}{\alpha+1}+1}}\\&=\frac{(j-2)!(n-j)!v^2\alpha(\alpha+1)^{n-3}}{\tilde T_n}\binom{\frac{(v+1)\alpha}{\alpha+1}+n-3}{n-j-1}\binom{\frac{2v\alpha}{\alpha+1}+j-2}{j-2}.\end{aligned} \quad (7.78)$$

This can be simplified to

$$p^{[2]}_{n,j}(v)=\Big(\prod_{k=j+1}^{n-1}\big(\frac{l(\alpha+1)-2-\alpha}{l(\alpha+1)-2}+v\frac{\alpha}{l(\alpha+1)-2}\big)\Big)p^{[2]}_{j+1,j}(v), \quad (7.79)$$

where

$$p^{[2]}_{j+1,j}(v)=\frac{v^2}{2}\prod_{k=3}^{j}\Big(\frac{k(\alpha+1)-2-2\alpha}{k(\alpha+1)-2}+v\frac{2\alpha}{k(\alpha+1)-2}\Big). \quad (7.80)$$

This proves the distribution law. □

Proof of Theorem 46. For the probabilities $\mathbb{P}\{\tilde\Delta_{n,j}=m\}$ one proceeds by extracting coefficients in (7.71) and (7.78). Since the basic computations concerning (7.71) are a bit lengthy they are not presented.

$$\begin{aligned}
&\frac{(j-2)!(n-j-1)!}{\tilde{T}_n}[z^{j-2}u^{n-j-1}v^m]v^2 N(z,u,v)N(z+u,0,v)\varphi'\big(T(z+u)\big)\\
&=\frac{(j-2)!(n-j)!\alpha(\alpha+1)^{n-3}}{\tilde{T}_n}[v^m]v^2\binom{\frac{(v+1)\alpha}{\alpha+1}+n-3}{n-j-1}\binom{\frac{2v\alpha}{\alpha+1}+j-2}{j-2}\\
&=\frac{(j-2)!(n-j)!\alpha(\alpha+1)^{n-3}}{\tilde{T}_n}\sum_{l=0}^{m-2}[v^l]\binom{\frac{(v+1)\alpha}{\alpha+1}+n-3}{n-j-1}[v^{m-2-l}]\binom{\frac{2v\alpha}{\alpha+1}+j-2}{j-2}.
\end{aligned} \tag{7.81}$$

By using (7.54) one gets

$$\begin{aligned}
&\frac{(j-2)!(n-j-1)!}{\tilde{T}_n}[z^{j-2}u^{n-j-1}v^m]v^2 N(z,u,v)N(z+u,0,v)\varphi'\big(T(z+u)\big)\\
&=\frac{(j-2)!(n-j)!\alpha^{m-1}(\alpha+1)^{n-m-1}}{\prod_{k=2}^{n-1}(k(\alpha+1)-2)}\sum_{k=0}^{m-2}2^{m-2-k}\Big(\sum_{l=0}^{n-j-1}\frac{\left[{l \atop k}\right]}{l!}\binom{\frac{\alpha}{\alpha+1}+n-3-l}{n-j-1-l}\Big)\Big(\sum_{l=0}^{j-2}\frac{\left[{l \atop m-2-k}\right]}{l!}\Big).
\end{aligned} \tag{7.82}$$

□

Proof of Theorem 47. For $2 \le j \le n-1$ the expectation and the variance can be obtained as follows. We basically use

$$\begin{aligned}
\mathbb{E}(\tilde{\Delta}_{n,j}) &= \frac{(j-2)!(n-j-1)!}{\tilde{T}_n}[z^{j-2}u^{n-j-1}]E_v D_v\frac{\partial}{\partial z}\tilde{M}(z,u,v),\\
\mathbb{E}(\tilde{\Delta}_{n,j}^2) &= \frac{(j-2)!(n-j-1)!}{\tilde{T}_n}[z^{j-2}u^{n-j-1}]E_v D_v^2\frac{\partial}{\partial z}\tilde{M}(z,u,v),
\end{aligned} \tag{7.83}$$

and $\mathbb{V}(\tilde{\Delta}_{n,j}) = \mathbb{E}(\tilde{\Delta}_{n,j}^2) + \mathbb{E}(\tilde{\Delta}_{n,j}) - \mathbb{E}(\tilde{\Delta}_{n,j})^2$. One also makes use of the relations

$$\begin{aligned}
[z^n]\frac{\log\left(\frac{1}{1-z}\right)}{(1-z)^{x+1}} &= \binom{n+x}{n}(H_{n+x}-H_x),\\
[z^n]\frac{\log^2\left(\frac{1}{1-z}\right)}{(1-z)^{x+1}} &= \binom{n+x}{n}\big((H_{n+x}-H_x)^2-(H_{n+x}^{(2)}-H_x^{(2)})\big).
\end{aligned} \tag{7.84}$$

□

Part III

Weighted parameters in increasing trees

Chapter 8

Weighted depths and distances in increasing trees

8.1 Introduction

Most of the usually considered parameters in rooted labelled trees like height, width, node degree, depth, subtree size, etc., do not depend on the actual labelling of the trees. E.g. the node degrees of the roots in the trees in Figure 8.1 are both 4, although nodes with different labels contribute to the node degree.

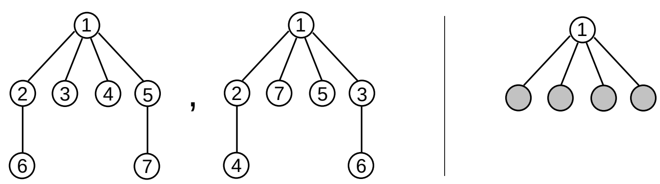

Figure 8.1: Two size 7 recursive trees with root degree 4.

Aguech, Lasmar and Mahmoud [1] have recently studied a weighted version of the depth in random binary search trees. Our results were inspired by their work. We will present two weighted extensions of parameters in rooted labelled trees.

8.1.1 Node weights

First we simply take the labelling of the node into account. For example for the weighted node degree we count now the labels of the nodes contributing to the original node degree instead of counting only the number of nodes. Parameters like height or depth can also be generalized this way by counting nodes instead of edges.

8.1.2 Edge weights

Another interesting generalization is to equip the edges with weights. It seems to be most natural to define the edge-weights as follows. If there is an edge between node v_j labelled j and node v_k labelled k in a labelled rooted tree T of size $n \geq \max\{j, k\}$, $e = (v_j, v_k)$, then we define the weight w_e of the edge e as $w_e := |j - k|$. If there is no edge $e = (v_j, v_k)$ then $w_e := 0$. Hence we get a weighted $n \times n$ adjacency matrix $(w_{i,j})_{1 \leq i,j \leq n}$ with entries $w_{i,j} = w_{e=(v_i,v_j)}$.

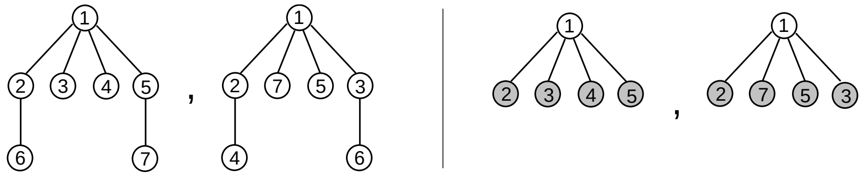

Figure 8.2: Two size 7 recursive trees. The root of the first tree has weighted node degree 14, whereas the root of the second tree has weighted node degree 17.

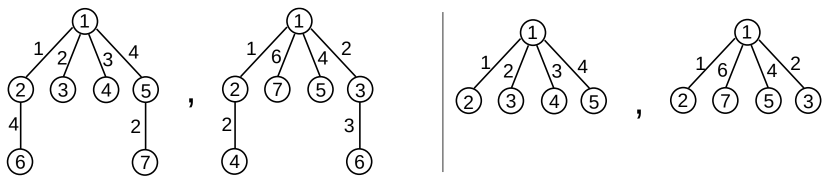

Figure 8.3: Two size 7 recursive trees. The root of the first tree has edge-weighted node degree 10, whereas the root of the second tree has edge-weighted node degree 13.

Remark 17. The concept of edge weights is not at all artificial. For a size n random recursive tree there is a nice combinatorial interpretation of the random variable S_n, counting the sum of all edge weights w_e. We observe that the contribution of the weight w_e, where the edge e is generated by attaching node $k+1$ at any of the former k nodes in a size k random recursive tree, is uniformly distributed on $1, 2, \ldots, k$. Hence S_n satisfies the following distribution law.

$$S_n \stackrel{(d)}{=} \mathrm{U}_1 + \mathrm{U}_2 + \cdots + \mathrm{U}_{n-1}, \tag{8.1}$$

where U_k denotes a uniform distribution on $1, 2, \ldots, k$. Thus S_n counts the number of inversions in a random permutation $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ of size n with the constraint that $\sigma_n = 1$. But this is just the number of inversions of a random permutation of $\{1, 2, \ldots, n-1\}$ shifted by $n-1$.

Remark 18. It is of course possible to go a step further by defining even more general (f-weighted) parameters by introducing a weight function $f(x)$ which acts on the labels (node weights) or on the labels of adjacent pairs of nodes (edge weights). By choosing $f(x) \equiv 1$ we regain the ordinary parameters. The choice $f(x) = x$ leads to the node-weighted and edge-weighted parameters. Hence choosing $f(x) = x^\alpha$ would provide a transition between ordinary and weighted parameters.

It was already seen in the work of Aguech, Lasmar and Mahmoud [1] that a weighted parameter likely leads to Dickman limiting distribution. We will also encounter the Dickman distribution and generalizations of it. We will collect now some facts about the Dickman distribution.

8.1.3 The Dickman distribution and generalizations

The Dickman function $\rho(u)$, which appears in analytic number theory, is defined as the continuous solution of the differential-difference equation

$$u\rho'(u) + \rho(u-1) = 0 \quad (u > 1), \tag{8.2}$$

with initial condition $\rho(u) = 1$ for $0 \le u \le 1$ and with $\rho(u)$ differentiable on $(1, \infty)$. It is known that the Dickman function is positive and decreasing over the whole interval $(1, \infty)$. Note that

$$\int_0^\infty \rho(v)\mathrm{d}v = e^\gamma, \tag{8.3}$$

where γ denotes Euler's constant. For simplicity of reference, we call the distribution with the density function $e^{-\gamma}\rho(v)$ the Dickman distribution. It is known that a Dickman distributed random variable satisfies the following distributional fixed-point identity

$$X \stackrel{(d)}{=} \mathrm{U}(1+X), \tag{8.4}$$

where U denotes a uniform distribution on $[0,1)$.

Penrose and Wade introduced in [69] the *generalized Dickman distribution*. Given $\theta > 0$, a random variable X has a generalized Dickman distribution with shape parameter θ, or

$$X \stackrel{(d)}{=} \mathrm{GD}(\theta), \tag{8.5}$$

if it satisfies the distributional fixed-point identity

$$X \stackrel{(d)}{=} \mathrm{U}^{1/\theta}(1+X), \tag{8.6}$$

where U is uniform on $(0,1]$ and is independent of the X on the right. Some other known properties of the generalized Dickman distribution stated in [69] are listed as follows.

- If $X \stackrel{(d)}{=} \mathrm{GD}(\theta)$, then the Laplace transform of X is given by

$$\psi(t) = \mathbb{E}(e^{-tX}) = \exp\Big(\theta\int_0^t \frac{e^{-s}-1}{s}\mathrm{d}s\Big), \quad t \in \mathbb{R}. \tag{8.7}$$

- The $\mathrm{GD}(\theta)$ distribution is infinitely divisible.
- If $X \stackrel{(d)}{=} \mathrm{GD}(\theta)$, then the moments $\mathbb{E}(X^k)$ satisfy $\mathbb{E}(X^0) = 1$ and for integer $k \geq 1$

$$\mathbb{E}(X^k) = \frac{\theta}{k}\sum_{j=0}^{k-1}\binom{k}{j}\mathbb{E}(X^j). \tag{8.8}$$

For $\theta = 1$ the $\mathrm{GD}(\theta)$ distribution is just the ordinary Dickman distribution.
For more properties of the $\mathrm{GD}(\theta)$ distribution see the paper of Penrose and Wade [69].

8.2 The weighted depth in increasing trees

As already mentioned, the depth of node v, also called the level of a node v, is usually measured by the number of edges lying on the unique path from the root to node v. For labelled rooted trees we consider a generalization of the depth. Let $W_{n,j}$ denote the *weighted depth* of node j in a size $n \geq j$ random grown simple increasing tree, which is the sum of the labels of the nodes encountered on the path from j to the root labelled 1, where $W_{1,1} = 1$. For instance if the nodes labelled $\lambda_1, \dots, \lambda_k$, with $1 = \lambda_1 < \lambda_2 < \cdots < \lambda_k = j$, are visited on the unique path from j to the root then the weighted depth equals $\sum_{l=1}^{k}\lambda_k$.
Further we consider the *weighted distance* $\mathcal{W}_{n,j}$ between node j and node n in a size n random grown simple increasing tree which is the sum of the labels of the nodes encountered on the path from n to the node labelled j (hence $\mathcal{W}_{n,1} = W_{n,n}$).
E.g. in the tree of Figure 8.2, $W_{9,1} = 1$, $W_{9,2} = 3$, $W_{9,3} = 4$, $W_{9,4} = 8$, $W_{9,5} = 6$, $W_{9,6} = 12$, $W_{9,7} = 8$, $W_{9,8} = 20$ and $W_{9,9} = 17$; and for instance $\mathcal{W}_{9,4} = 24$, $\mathcal{W}_{9,8} = 36$.
We will show that $W_{n,j}$ and $\mathcal{W}_{n,j}$, appropriately scaled, lead to generalized Dickman distributions. We use several results concerning ordinary depths and distances, see the work of Dobrow and Smythe [18] and [45]. We will recollect the distribution laws of the ordinary depth and distance as studied in Chapter 6.

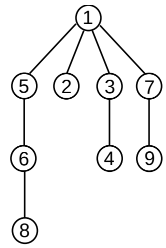

Figure 8.4: A size 9 recursive tree.

8.2.1 Results for grown simple families of increasing trees

Theorem 50. *The weighted depth of node j in a size $n \geq j$ random grown simple increasing tree $W_{n,j}$ admits the following distribution law.*

$$W_{n,j} \stackrel{(d)}{=} W_{j,j} \stackrel{(d)}{=} j \oplus \bigoplus_{k=1}^{j-1} \mathcal{B}_k, \tag{8.9}$$

where $\mathcal{B}_k = \mathcal{B}_k(c_1, c_2)$ are a sequence of of independent random variables such that

$$\mathbb{P}(\mathcal{B}_k = k) = \frac{1 + \frac{c_2}{c_1}}{k + \frac{c_2}{c_1}} \quad \textit{and} \quad \mathbb{P}(\mathcal{B}_k = 0) = 1 - \frac{1 + \frac{c_2}{c_1}}{k + \frac{c_2}{c_1}}. \tag{8.10}$$

Note that $\mathcal{B}_k \stackrel{(d)}{=} k \cdot \mathbb{1}(A_k)$, where again A_k denotes the event that node k is on the path from the root to node j. Further the $\mathcal{B}_k$'s are independent.

Corollar 16. *The expectation and the variance of the random variable $W_{n,j}$ are given as follows.*

$$\begin{aligned}
\mathbb{E}(W_{n,j}) &= (1 + \frac{c_2}{c_1})(j-1) - \frac{c_2}{c_1}(1 + \frac{c_2}{c_1})(H_{j-1+\frac{c_2}{c_1}} - H_{\frac{c_2}{c_1}}),\\
\mathbb{V}(W_{n,j}) &= \frac{1 + \frac{c_2}{c_1}}{2} j(j-1) - (1 + \frac{c_2}{c_1})(1 + 2\frac{c_2}{c_1})(j-1) + (1 + \frac{c_2}{c_1})\frac{c_2}{c_1}(2 + 3\frac{c_2}{c_1})(H_{j-1+\frac{c_2}{c_1}} - H_{\frac{c_2}{c_1}})\\
&\quad - \frac{c_2^2}{c_1^2}(1 + \frac{c_2}{c_1})^2(H^{(2)}_{j-1+\frac{c_2}{c_1}} - H^{(2)}_{\frac{c_2}{c_1}}).
\end{aligned} \tag{8.11}$$

Theorem 51. *The limiting distribution of the random variable $\frac{W_{n,j}-j}{j}$ is a generalized Dickman distribution with parameter $\theta = 1 + \frac{c_2}{c_1} > 0$.*

$$\lim_{j\to\infty} \frac{W_{n,j} - j}{j} \stackrel{(d)}{=} \mathrm{GD}(1 + \frac{c_2}{c_1}), \tag{8.12}$$

or equivalently let $\psi_j(t) := \mathbb{E}(e^{-t(W_{n,j}-j)})$, then

$$\lim_{j\to\infty} \psi_j(\frac{t}{j}) = \exp\Big(\big(1 + \frac{c_2}{c_1}\big) \int_0^t \frac{e^{-v} - 1}{v} dv\Big), \tag{8.13}$$

for constants c_1, c_2 as in (1).

Corollar 17. *For the three most prominent tree families we obtain the following result.*

- ***Recursive Trees*** *($c_1 = 1$, $c_2 = 0$): The limit distribution of $\frac{W_{n,j}-j}{j}$ is Dickman,* GD(1).

$$\lim_{j\to\infty} \mathbb{P}\{\frac{W_{n,j} - j}{n} < x\} = e^{-\gamma} \int_0^x \rho(v) dv, \quad x > 0. \tag{8.14}$$

- ***Binary Increasing trees*** *($c_1 = c_2 = 1$) : The limit distribution of $\frac{W_{n,j}-j}{j}$ is the convolution of two Dickman distributions,* GD(2).

$$\lim_{j\to\infty} \mathbb{P}\{\frac{W_{n,j}-j}{n} < x\} = e^{-2\gamma}\int_0^x \rho(v)\rho(x-v)dv, \quad x > 0. \tag{8.15}$$

- ***Plane oriented recursive trees*** *($c_1 = 2, c_2 = -1$): The limit distribution of $\frac{W_{n,j}-j}{j}$ a generalized Dickman distribution* GD(θ) *with parameter $\theta = \frac{1}{2}$.*

Theorem 52. *The weighted distance $\mathcal{W}_{n,j}$ between node j and node n in a size $n \geq j$ random grown simple increasing tree admits the following distribution law.*

$$\mathcal{W}_{n,j} \stackrel{(d)}{=} \big(\mathcal{W}_{j+1,j} - (j+1)\big) \oplus \bigoplus_{k=j+1}^{n-1} \mathcal{B}_k \oplus n, \tag{8.16}$$

where $\mathcal{B}_k = \mathcal{B}_k(c_1, c_2)$ are a sequence of of independent random variables such that

$$\mathbb{P}(\mathcal{B}_k = k) = \frac{1+\frac{c_2}{c_1}}{k+\frac{c_2}{c_1}} \quad \text{and} \quad \mathbb{P}(\mathcal{B}_k = 0) = 1 - \frac{1+\frac{c_2}{c_1}}{k+\frac{c_2}{c_1}}. \tag{8.17}$$

Note that $\mathcal{B}_k \stackrel{(d)}{=} k \cdot \mathbb{1}(A_k)$, where again A_k denotes the event that node k is on the path from node n to node j. Further the $\mathcal{B}_k$'s and $\mathcal{W}_{j+1,j}$ are independent.

Theorem 53. *For plane oriented recursive trees and tree families admitting $c_1 = -2c_2$ we obtain the following distribution law for $\mathcal{W}_{j+1,j}$.*

$$\mathcal{W}_{j+1,j} \stackrel{(d)}{=} (2j+1) \oplus \bigoplus_{k=1}^{j-1} \mathcal{C}_k \stackrel{(d)}{=} (2j+1) \oplus \bigoplus_{k=1}^{j-1} k \cdot \mathbb{1}(\mathcal{A}_k), \tag{8.18}$$

where $\mathcal{C}_k$ are a sequence of independent random variables such that

$$\mathbb{P}(\mathcal{C}_k = k) = \frac{2}{2k+1} \quad \text{and} \quad \mathbb{P}(\mathcal{C}_k = 0) = 1 - \frac{2}{2k+1}. \tag{8.19}$$

$\mathcal{A}_k$ denotes the event that node k is on the path from node $j+1$ to node j and the $\mathbb{1}(\mathcal{A}_k)$'s are mutually independent.

Remark 19. Note that a representation of the form

$$\mathcal{W}_{j+1,j} \stackrel{(d)}{=} (2j+1) \oplus \sum_{k=1}^{j-1} k\mathbb{1}(\mathcal{A}_k), \tag{8.20}$$

can always be achieved, but the indicator variables $\mathbb{1}(\mathcal{A}_k)$ are *not* mutually independent for grown simple families of increasing trees satisfying $c_1 \neq -2c_2$. Hence for this case we cannot further specify the distribution law of $\mathcal{W}_{j+1,j}$ at the moment.

Theorem 54. *For fixed j and $n \to \infty$ the limiting distribution of the weighted distance $\mathcal{W}_{n,j}$ between node j and node n in a size n grown simple increasing tree is a generalized Dickman distribution with parameter $\theta = 1 + \frac{c_2}{c_1} > 0$.*

$$\lim_{n\to\infty} \frac{\mathcal{W}_{n,j} - j - n}{n} \stackrel{(d)}{=} \mathrm{GD}(1 + \frac{c_2}{c_1}), \tag{8.21}$$

or equivalently $\mathbb{E}(e^{-t(\mathcal{W}_{n,j}-n-j)}) = \psi_n(t)$.

$$\lim_{j\to\infty} \psi_n(\frac{t}{n}) = \exp\Big(\big(1+\frac{c_2}{c_1}\big)\int_0^t \frac{e^{-v}-1}{v}dv\Big), \tag{8.22}$$

for constants c_1, c_2 as in (1).

Theorem 55. *The limiting distribution of the weighted distance $\mathcal{W}_{n,j}$ between node j and node n in a grown simple families of increasing tree satisfying $c_1 = -2c_2$ depends on the growth of j.*

- *The region $j = o(n)$: The limiting distribution is a generalized Dickman distribution with parameter $\theta = 1/2$.*
$$\lim_{n\to\infty} \frac{\mathcal{W}_{n,j} - j - n}{n} \overset{(d)}{=} \mathrm{GD}(\frac{1}{2}), \tag{8.23}$$
- *The region $j = \mu n$, with $0 < \mu < 1$. The limiting distribution can be characterized via its Laplace transform. Let $\mathbb{E}(e^{-t(\mathcal{W}_{n,j}-n-j)}) = \psi_n(t)$.*
$$\lim_{n\to\infty} \psi_n(\frac{t}{n}) = \exp\Big(\frac{1}{2}\int_0^{\mu t} \frac{e^{-v}-1}{v} dv + \int_{\mu t}^{t} \frac{e^{-v}-1}{v} dv\Big), \tag{8.24}$$
- *The region $l = n - j = o(n)$: The limiting distribution is a Dickman distribution.*
$$\lim_{n\to\infty} \frac{\mathcal{W}_{n,j} - j - n}{n} \overset{(d)}{=} \mathrm{GD}(1), \tag{8.25}$$

8.2.2 Deriving the distribution laws

We already encountered all the ingredients for proving Theorem 50 and 52 in Chapter 6. Dobrow and Smythe [18] already provided a full characterization of the depth which we reproved in Corollary 10:

Lemma 21 (Dobrow & Smythe)**.** *The depth of node j in a size $n \geq j$ random grown simple increasing tree $D_{n,j}$ admits the following distribution law.*

$$D_{n,j} \overset{(d)}{=} D_{j,j} \overset{(d)}{=} \bigoplus_{k=1}^{j-1} B_k, \tag{8.26}$$

where $B_k \overset{(d)}{=} \mathrm{Be}(p_k)$ with $p_k = \frac{1+\frac{c_2}{c_1}}{k+\frac{c_2}{c_1}}$ for $1 \leq k \leq j-1$. Note that $B_k \overset{(d)}{=} \mathrm{Be}(p_k) \overset{(d)}{=} \mathbb{1}(A_k)$, where A_k denotes the event that node k is on the path from the root to node j. Further the B_k's are independent.

Hence we immediately obtain Theorem 50 because the structure is independent of the weights.

Remark 20. The characterization of grown simple families of increasing trees $p_k = \frac{1+\frac{c_2}{c_1}}{k+\frac{c_2}{c_1}}$ is due to Panholzer and Kuba.

Dobrow and Smythe [18] also analyzed the distribution law of $\Delta_{n,j}$.

Lemma 22 (Dobrow & Smythe)**.** *The distance of between node j and node n in a random grown simple increasing tree $\Delta_{n,j}$ admits the following distribution law.*

$$\Delta_{n,j} \overset{(d)}{=} \Delta_{j+1,j} \oplus \bigoplus_{k=j+1}^{n-1} B_k, \tag{8.27}$$

where $B_k \overset{(d)}{=} \mathrm{Be}(p_k)$ with $p_k = \frac{1+\frac{c_2}{c_1}}{k+\frac{c_2}{c_1}}$ for $j+1 \leq k \leq n-1$.

Further the distribution law of $\Delta_{j+1,j}$ is completely characterized by Theorem 36 which is stated here as a Lemma.

Lemma 23 (Panholzer & Kuba)**.** *The distribution law of $\Delta_{j+1,j}$ is given as follows. For $c_1 \neq -2c_2$:*

$$\Delta_{j+1,j} \overset{(d)}{=} \sum_{k=1}^{\eta_j} \tilde{B}_k, \tag{8.28}$$

where $\tilde{B}_k \overset{(d)}{=} \mathrm{Be}(\tilde{p_k})$, $\tilde{p}_0 = \tilde{p}_1 = 1$ and $\tilde{p}_k = \frac{1}{k-1+\frac{c_2}{c_1}}$ for $3 \leq k \leq j$. Further $\mathbb{P}\{\eta_j = 1\} = \frac{1+\frac{c_2}{c_1}}{j+\frac{c_2}{c_1}}$ and $\mathbb{P}\{\eta_j = m\} = \frac{1}{j+\frac{c_2}{c_1}}$ for $2 \leq m \leq j$. For $c_1 = -2c_2$ we find an even simpler decomposition:

$$\Delta_{j+1,j} \overset{(d)}{=} \bigoplus_{k=1}^{j} \tilde{B}_k, \tag{8.29}$$

where $\tilde{B}_1 \overset{(d)}{=} 1$ and $\tilde{B}_k \overset{(d)}{=} \mathrm{Be}(\frac{2}{2k-1})$ for $2 \leq k \leq j$.

Now observe that by directly translating Lemma 22 the weight of node $j+1$ is counted one time too much. This leads to $\mathcal{W}_{j+1,j} - (j+1)$.

8.2.3 Deriving the limiting distribution for the weighted depth

The proof of Theorem 51 is an extension of a result due to Hwang and Tsai [35].

Lemma 24 (Hwang & Tsai)**.** *The limiting distribution of the random variable*

$$X_j \overset{(d)}{=} \bigoplus_{k=1}^{j} \mathcal{B}_k(1,0), \tag{8.30}$$

where the $\mathcal{B}_k$'s are defined as in Theorem 50, is for $j \to \infty$ asymptotically Dickman.

$$\lim_{j\to\infty} \mathbb{P}\{\frac{X_j}{n} < x\} = e^{-\gamma} \int_0^x \rho(v)dv, \quad (x > 0). \tag{8.31}$$

We will follow the proof of Lemma 24 in [35] in order to prove Theorem 51. Let $\psi_j(t) := \mathbb{E}(e^{-t(W_{n,j}-j)})$ denote the Laplace transform of the shifted r.v. $W_{n,j} - j$.

$$\phi_j(t) = \mathbb{E}(e^{-t(W_{n,j}-j)}) = \prod_{1\leq k\leq j-1} \frac{k-1+e^{-tk}(1+\frac{c_2}{c_1})}{k+\frac{c_2}{c_1}} = \prod_{1\leq k\leq j-1} \Big(1 + \frac{(e^{-tk}-1)(1+\frac{c_2}{c_1})}{k+\frac{c_2}{c_1}}\Big). \tag{8.32}$$

It suffices, by Lévy's continuity theorem, and (8.7) to show that

$$\lim_{j\to\infty} \phi_j(\frac{t}{j}) = \lim_{j\to\infty} \mathbb{E}(e^{-t\frac{(W_{n,j}-j)}{j}}) = \exp\Big((1+\frac{c_2}{c_1})\int_0^t \frac{e^{-v}-1}{v}\mathrm{d}v\Big), \tag{8.33}$$

for finite and real t. Now

$$\begin{aligned}\phi_j(\frac{t}{j}) &= \exp\big(\log(\phi_j(\frac{t}{j}))\big) = \exp\Big(\sum_{k=1}^{j-1} \log\big(1 + \frac{(e^{-tk}-1)(1+\frac{c_2}{c_1})}{k+\frac{c_2}{c_1}}\big)\Big) \\ &= \exp\Big((1+\frac{c_2}{c_1})\sum_{k=1}^{j-1} \frac{e^{\frac{-kt}{j}}-1}{k+\frac{c_2}{c_1}} + R_j(t)\Big),\end{aligned} \tag{8.34}$$

where

$$R_j(t) := \sum_{l\geq 2} \frac{(-1)^{l-1}}{l} \sum_{k=1}^{j-1} \frac{\big((1+\frac{c_2}{c_1})\big)^l \big(e^{\frac{-kt}{j}}-1\big)^l}{\big(k+\frac{c_2}{c_1}\big)^l} = \mathcal{O}(\frac{|t^2|}{j}). \tag{8.35}$$

Now we apply the Euler-MacLaurin summation formula shows

$$\sum_{k=1}^{j-1} \frac{e^{\frac{-kt}{j}} - 1}{k + \frac{c_2}{c_1}} = \int_0^t \frac{e^{-v} - 1}{v} \mathrm{d}v + \mathcal{O}(\frac{|t^2|}{j}). \tag{8.36}$$

8.2.4 Deriving the limiting distribution for the weighted distances

We only sketch the proofs for the weighted distances since they are similar to Subsection 8.2.3. Let $\psi_n(t) := \mathbb{E}(e^{-t(\mathcal{W}_{n,j}-j-n)})$ denote the Laplace transform of the shifted r.v. $\mathcal{W}_{n,j} - j - n$.

$$\begin{aligned} \phi_n(t) = \mathbb{E}(e^{-t(\mathcal{W}_{n,j}-j-n)}) &= \mathbb{E}(e^{-t(\mathcal{W}_{j+1,j}-j)}) \prod_{j+1 \le k \le n-1} \frac{k - 1 + e^{-tk}(1 + \frac{c_2}{c_1})}{k + \frac{c_2}{c_1}} \\ &= \mathbb{E}(e^{-t(\mathcal{W}_{j+1,j}-j)}) \prod_{j+1 \le k \le n-1} \Big(1 + \frac{(e^{-tk} - 1)(1 + \frac{c_2}{c_1})}{k + \frac{c_2}{c_1}}\Big). \end{aligned} \tag{8.37}$$

Hence

$$\phi_n(\frac{t}{n}) = \mathbb{E}(e^{-\frac{t(\mathcal{W}_{j+1,j}-j)}{n}}) \exp\Big((1 + \frac{c_2}{c_1}) \sum_{k=j+1}^{n-1} \frac{(e^{-tk} - 1)}{k + \frac{c_2}{c_1}}) + R_{n,j}(t)\Big), \tag{8.38}$$

where it can be seen that

$$R_{n,j}(t) := \mathcal{O}(\frac{|t^2|}{n}). \tag{8.39}$$

We use again Euler-MacLaurin summation to show that

$$\sum_{k=j+1}^{n-1} \frac{(e^{-\frac{tk}{n}} - 1)}{k + \frac{c_2}{c_1}} = \int_0^t \frac{e^{-v} - 1}{v} \mathrm{d}v + \mathcal{O}(\frac{|t^2|}{n}). \tag{8.40}$$

For grown simple families of increasing trees satisfying $c_1 = -2c_2$ we get

$$\phi_n(t) = \mathbb{E}(e^{-t(\mathcal{W}_{n,j}-j-n)}) = \Big(\prod_{1 \le k \le j-1} \Big(1 + \frac{(e^{-tk} - 1)}{k + \frac{1}{2}}\Big)\Big)\Big(\prod_{j+1 \le k \le n-1} \Big(1 + \frac{(e^{-tk} - 1)\frac{1}{2}}{k - \frac{1}{2}}\Big)\Big), \tag{8.41}$$

and further

$$\phi_n(\frac{t}{n}) = \exp\Big(\sum_{k=1}^{j-1} \frac{(e^{-\frac{tk}{n}} - 1)}{k + \frac{c_2}{c_1}} + \frac{1}{2} \sum_{k=j+1}^{n-1} \frac{(e^{-\frac{tk}{n}} - 1)}{k + \frac{c_2}{c_1}} + \hat{R}_{n,j}(t) + \tilde{R}_{n,j}(t)\Big). \tag{8.42}$$

An application of Euler-MacLaurin summation provides then Theorem 55.

Remark 21. The edge weighted variant of the depth and the distance shows a quite different behavior. Let $K_{n,j} = \sum_{e \in E} w_e \mathbb{1}(A_e)$ denote the random variable counting the edge-weighted depth of node j in a size n grown simple increasing tree, where A_e denotes the event that the edge e is on the path from j to the root. For $f(x) \equiv 1$ one would obtain the ordinary depth. Further we denote with $\mathcal{K}_{n,j}$ the random variable counting the distance measured by the sum of edge-weights on the unique path from j to n. The r. v. $K_{n,j}$ satisfies the following distribution.

$$K_{n,j} \overset{(d)}{=} K_{j,j} \overset{(d)}{=} j - 1. \tag{8.43}$$

Further one can easily show that

$$\mathcal{K}_{n,j} \overset{(d)}{=} (n - j - 1) \oplus \mathcal{K}_{j+1,j}. \tag{8.44}$$

The distribution law of $\mathcal{K}_{j+1,j}$ is non-degenerate. For recursive trees it can be specified by basic manipulation and usage of the probabilities $\mathbb{P}\{j <_c k\}$, already stated in [18], which give the probability that node j is attached (a child of) of node k. For recursive trees we get

$$\mathbb{P}\{\mathcal{K}_{j+1,j} = 2m-1\} = \begin{cases} \mathbb{P}\{j+1 <_c j\} = \frac{1}{j}, & m = 1, \\ \frac{1}{(j+2-m)(j+1-m)}, & m = 2, \ldots, j. \end{cases} \tag{8.45}$$

Chapter 9

A thorough study of the node degree in recursive trees

9.1 Introduction

Let $X_{n,j}$ denote the random variable which counts the node degree (i. e. the out-degree) of the node labelled j in a size n random recursive tree. This r. v. was studied in Chapter 5 Section 5.5 ([47]) for the family grown simple families of increasing trees which contains recursive trees as a special instance. The results in Chapter 5 Section 5.5 are not sufficient to obtain results for the weighted node degrees in recursive trees. We will use a simple probabilistic approach to extend the results of Chapter 5 Section 5.5 for recursive tree. We will generalize the results of [47] by specifying the limit distribution of the random vector $\mathbf{X}_{n;\mathbf{j}_r} = (X_{n;j_1}, \dots, X_{n;j_r})$, counting the node degree of the nodes $j_1, \dots, j_r$, and the random variable $X_{n;j_1,\dots,j_r} = \sum_{k=1}^{r} X_{n,j_k}$, counting the sum of the outdegrees of nodes $j_1, \dots, j_r$ in a random size n recursive tree. It turns out that for fixed $j_1, \dots, j_r$ the limiting distribution is asymptotically gaussian. For $X_{n;j_1,\dots,j_r} = \sum_{k=1}^{r} X_{n,j_k}$ we give limiting distribution results for the full region $1 \le j_1 < \dots < j_r \le n$. Further we are able to give results for weighted versions of the node degree which are defined as follows. Let $\mathfrak{X}_{n,j}$ denote the r. v. counting the labels of the nodes attached to node j in a random size n recursive tree, where each node contributing to the node degree is weighted by its label. We can specify the limit distribution of the random vector $\boldsymbol{\mathfrak{X}}_{n;\mathbf{j}_r} = (\mathfrak{X}_{n;j_1}, \dots, \mathfrak{X}_{n;j_r})$, counting the weighted node degree of the nodes $j_1, \dots, j_r$ in a random size n recursive tree, and the random variable $\mathfrak{X}_{n;j_1,\dots,j_r} = \sum_{k=1}^{r} \mathfrak{X}_{n,j_k}$. It turns out that the limiting distributions involve Dickman's infinitely divisible distribution as the limit.

9.2 Results

9.2.1 Result concerning the ordinary node degree

Theorem 56. *The probability generating function of the random vector $\mathbf{X}_{n;\mathbf{j}_r} = (X_{n;j_1}, \dots, X_{n;j_r})$, counting the node degree of the nodes $j_1, \dots, j_r$ in a random size n recursive tree, is for $n \ge j_r > \dots > j_1 \ge 1$ given as follows.*

$$p_{n;\mathbf{j}_r}(\mathbf{v}_r) = \prod_{i=1}^{r} \prod_{k=j_i}^{j_{i+1}-1} \Big(\frac{k-i}{k} + \sum_{l=1}^{i} \frac{v_l}{k} \Big) = \prod_{i=1}^{r} \frac{\binom{j_{i+1}-1-i+\sum_{l=1}^{i} v_l}{j_{i+1}-1}}{\binom{j_i-1-i+\sum_{l=1}^{i} v_l}{j_i-1}}, \tag{9.1}$$

where $j_{r+1} = n$.

Theorem 57. *The covariance of the random variables $X_{n;j_1}$ and $X_{n;j_2}$, counting the node degree of nodes* *labelled j_1 and j_2 in a random recursive tree of size n is given by*

$$\mathrm{Cov}(X_{n;j_1}, X_{n;j_2}) = -(H^{(2)}_{n-1} - H^{(2)}_{j_2-1}). \tag{9.2}$$

Let $\mathbf{X}^*_{n;\mathbf{j}_r}$ denote the random vector $(X^*_{n;j_1}, \dots, X^*_{n;j_r})$ with components

$$X^*_{n;j_i} = \frac{X_{n;j_i} - \mu_{n;j_i}}{\sigma_{n;j_i}}, \quad \text{for} \quad 1 \le i \le r, \tag{9.3}$$

with $\mu_{n;j_i} = \sigma^2_{n;j_i} = \log n - \log j_i \sim \mathbb{E}(X_{n;j_i})$.

Theorem 58. *For $\max\{j_1, \dots, j_r\} = o(n)$ the random vector $\mathbf{X}^*_{n;\mathbf{j}_r}$ converges in distribution to r independent gaussian distributions*

$$\mathbf{X}^*_{n;\mathbf{j}_r} \xrightarrow{(d)} \mathbf{X_r} \qquad \mathbb{P}\{\mathbf{X_r} \le \mathbf{m}_r\} = \Phi(m_1)\Phi(m_2)\dots\Phi(m_r). \tag{9.4}$$

Let $X_{n;j_1,\dots,j_r} = \sum_{k=1}^r X_{n,j_k}$ denote the random variable counting the sum of the outdegrees of nodes $j_1, \dots, j_r$.

Corollar 18. *The r. v. $X_{n;j_1,\dots,j_r} = \sum_{k=1}^r X_{n,j_k}$ counting the sum of the outdegrees of nodes $j_1, \dots, j_r$ in a random size n recursive tree satisfies the following distribution law.*

$$Y_{n;j_1,\dots,j_r} \overset{(d)}{=} \sum_{i=1}^{r} \sum_{k=j_i}^{j_{i+1}-1} B_k^{[i]} \overset{(d)}{=} \sum_{i=1}^{r} \sum_{k=j_i}^{j_{i+1}-1} \mathbb{1}(\mathcal{A}_k), \tag{9.5}$$

where $j_r = n-1$ and $B_k^{[i]} \overset{(d)}{=} \mathrm{Be}(\frac{i}{k})$. $\mathcal{A}_k$ denotes the event that node k is joined to any node $j_1, \dots, j_r$. The events $\mathcal{A}_k$ are mutually independent.

Corollar 19. *For fixed $r \ge 1$ and fixed $j_1, \dots, j_r$ the expectation and the variance of $X_{n;j_1,\dots,j_r}$ is given as follows.*

$$\begin{aligned}
\mathbb{E}(X_{n;j_1,\dots,j_r}) &= r(H_{n-1} - H_{j_r-1}) + \sum_{i=1}^{r} i(H_{j_{i+1}-1} - H_{j_i-1}), \\
\mathbb{V}(X_{n;j_1,\dots,j_r}) &= r(H_{n-1} - H_{j_r-1}) - r^2(H^{(2)}_{n-1} - H^{(2)}_{j_r-1}) \\
&\quad + \sum_{i=1}^{r} i(H_{j_{i+1}-1} - H_{j_i-1}) - \sum_{i=1}^{r} i^2(H^{(2)}_{j_{i+1}-1} - H^{(2)}_{j_i-1}).
\end{aligned} \tag{9.6}$$

We get the following generalization of the theorem concerning recursive trees in [47].

Theorem 59. *We get the following characterization of $X_{n;j_1,\dots,j_r}$ depending on r and the growth of the j_i, $1 \le i \le r$.*

- *Region: fixed $r \ge 1$, fixed $j_1, \dots, j_r$: The distribution of normalized and centralized $X_{n;j_1,\dots,j_r}$ is asymptotically gaussian.*

$$\frac{X_{n;j_1,\dots,j_r} - \mathbb{E}(X_{n;j_1,\dots,j_r})}{\sqrt{\mathbb{V}(X_{n;j_1,\dots,j_r})}} \xrightarrow{(d)} \mathcal{N}(0,1). \tag{9.7}$$

- *Region: fixed $r \ge 1$, $0 < j_1 = \rho_1 n < \dots < j_r = \rho_r n < n$: The distribution of $X_{n;j_1,\dots,j_r}$ is asymptotically Poisson.*

$$X_{n;j_1,\dots,j_r} \xrightarrow{(d)} X_{\rho_1,\dots,\rho_r} \overset{(d)}{=} \mathrm{Poi}(-\log(\rho_1 \dots \rho_r)), \tag{9.8}$$

where $\mathrm{Po}(-\log(\rho_1 \dots \rho_r))$ denotes a Poisson distribution with parameter $\lambda = -\log(\rho_1 \dots \rho_r)$,

$$\mathbb{P}\{X_{\rho_1,\dots,\rho_r} = m\} = \rho_1 \dots \rho_r \frac{(-\log(\rho_1 \dots \rho_r))^m}{m!}. \tag{9.9}$$

- *Region: $r = n+1-l$, fixed l: The distribution of $X_{n;j_1,\dots,j_{n+1-l}}$ satisfies*

$$\frac{X_{n;j_1,\dots,j_{n+1-l}}}{n} \xrightarrow{(d)} 0, \tag{9.10}$$

moreover let $\{\lambda_1,\dots,\lambda_l\} = \{1,\dots,n\} \setminus \{j_1,\dots,j_{n+1-l}\}$, then the expectation and the variance are given by

$$\begin{aligned}\mathbb{E}(X_{n;j_1,\dots,j_{n+1-l}}) &= n-1-\sum_{i=1}^{l} i(H_{\lambda_{i+1}-1} - H_{\lambda_i-1}),\\ \mathbb{V}(X_{n;j_1,\dots,j_{n+1-l}}) &= \sum_{i=1}^{l} i(H_{\lambda_{i+1}-1} - H_{\lambda_i-1}) - \sum_{i=1}^{l} i^2(H^{(2)}_{\lambda_{i+1}-1} - H^{(2)}_{\lambda_i-1}),\end{aligned} \tag{9.11}$$

where $\lambda_{l+1} := n$.

Let $X_{n,r}^{[R]}$ denote the random variable which counts the out-degree of $r \geq 1$ randomly chosen nodes in a random size n recursive tree.

Corollar 20. *$X_{n,r}^{[R]} \xrightarrow{(d)} X_r$, where the probabilities $\mathbb{P}\{X_r = m\}$ are given by*

$$\mathbb{P}\{X_r = m\} = \frac{\binom{m+r-1}{m}}{2^{m+r}}, m \geq 0. \tag{9.12}$$

9.2.2 Result concerning the weighted node degree

Corollar 21. *The probability generating function of the random vector $\mathfrak{X}_{n;\mathbf{j}_r} = (\mathfrak{X}_{n;j_1},\dots,\mathfrak{X}_{n;j_r})$, counting the weighted node degrees of the nodes $j_1,\dots,j_r$ in a random size n recursive tree, is for $n \geq j_r > \dots > j_1 \geq 1$ given as follows.*

$$p_{n;\mathbf{j}_r}(\mathbf{v}_r) = \prod_{i=1}^{r} \prod_{k=j_i}^{j_{i+1}-1} \Big(\frac{k-i}{k} + \sum_{l=1}^{i} \frac{v_l^k}{k}\Big), \tag{9.13}$$

where $j_{r+1} = n$.

Theorem 60. *Let $\psi_n(\mathbf{t}_r)$ denote the Laplace transform of the random vector $\mathfrak{X}_{n;\mathbf{j}_r}$. For fixed $j_1,\dots,j_r$ $\mathfrak{X}_{n;\mathbf{j}_r} n^{-1}$ converges in distribution to a distribution characterized by its Laplace transform*

$$\lim_{n\to\infty} \psi_n(\frac{\mathbf{t}_r}{n}) = \exp\Big(\int_0^1 \frac{\sum_{l=1}^{r} e^{-t_l v} - r}{v} dv\Big). \tag{9.14}$$

Corollar 22. *For fixed $j_1,\dots,j_r$ the random variable $\mathfrak{X}_{n;j_1,\dots,j_r} = \sum_{k=1}^{r} \mathfrak{X}_{n;j_k}$ is asymptotically a generalized Dickman distribution with parameter r*

$$\frac{\mathfrak{X}_{n;j_1,\dots,j_r}}{n} \xrightarrow{(d)} \mathrm{GD}(r). \tag{9.15}$$

Corollar 23. *For $r = 1$ we determine the limiting distribution of $\mathfrak{X}_{n;j}$ depending on the growth of j.*

- *Region $j = o(n)$: $\frac{\mathfrak{X}_{n;j}}{n}$ is asymptotically Dickman distributed (GD(1)).*

$$\lim_{j\to\infty} \mathbb{P}\{\frac{\mathfrak{X}_{n;j}}{n} < x\} = e^{-\gamma} \int_0^x \rho(v) dv, \quad x > 0. \tag{9.16}$$

- *Region $j = \mu n$, $0 < \mu < 1$: The limiting distribution of $\frac{\mathfrak{X}_{n;j}}{n}$ can be characterized via its Laplace* *transform. Let* $\mathbb{E}(e^{-t\mathfrak{X}_{n;j}}) = \psi_n(t)$.

$$\lim_{n\to\infty} \psi_n(\frac{t}{n}) = \exp\Big(\int_{\mu t}^{t} \frac{e^{-v}-1}{v} dv\Big). \tag{9.17}$$

The limit distribution is infinitely divisible.

- *Region* $l = n - j = o(n)$*:*

$$\frac{\mathfrak{X}_{n;j}}{n} \xrightarrow{(d)} 0. \tag{9.18}$$

9.3 Proofs

9.3.1 Deriving the probability generating function

Proof of Theorem 56. Theorem 56 is a generalization of a result implicitly given in [47]:

Lemma 25. *The random variable* $X_{n,j}$*, which counts the degree of node* j *in a random recursive tree of size* n *satisfies the following distribution law.*

$$X_{n,j} \overset{(d)}{=} \bigoplus_{k=j}^{n-1} B_k \overset{(d)}{=} \bigoplus_{k=j+1}^{n} \mathbb{1}(A_k), \tag{9.19}$$

where the B_k *are Bernoulli distributed random variables* $B_k \overset{(d)}{=} \mathrm{Be}(\frac{1}{k})$*, for* $j \le k \le n-1$*.* A_k *denotes the event that node* k *is attached to node* j*.*

We will use induction on n to prove Theorem 56. For $r = 1$ this is just Corollary 25. The increments are at any stage k, $1 \le k \le n$, and any $r \ge 1$ independent of the actual values of the node degrees $\mathbf{X}_{k-1;j_1\dots,j_r}$. It only depends on $j_1, \dots, j_r$ and k.
When a new node $k+1$ is inserted in a size k recursive tree with $j_i \le k < j_{i+1}$, there are already i nodes $j_1, \dots, j_i$ of interest in the size k tree. Hence with probability $\frac{i}{k}$ it attaches to any node $j_1, \dots, j_i$ and with probability $(k-i)/k$ it attaches to any node in $\{1, \dots, k\} \setminus \{j_1, \dots, j_i\}$. If $n \ge j_r$ node $n+1$ attaches with probability $\frac{r}{n}$ to any node $j_1, \dots, j_r$ and with probability $(n-r)/n$ it attaches to a node in $\{1, \dots, n\} \setminus \{j_1, \dots, j_r\}$. Formally this can be expressed as

$$\begin{aligned}
\mathbb{P}\{\mathbf{X}_{n+1;\mathbf{j}_r} = \mathbf{m}_r\} &= \sum_{l=1}^{r} \mathbb{P}\{\mathbf{X}_{n+1;\mathbf{j}_r} = \mathbf{m}_r | n+1 <_c j_l\}\mathbb{P}\{n+1 <_c j_l\} \\
&\quad + \mathbb{P}\{\mathbf{X}_{n+1;\mathbf{j}_r} = \mathbf{m}_r | n+1 \not<_c j_1, \dots, j_r\}\mathbb{P}\{n+1 \not<_c j_1, \dots, j_r\} \\
&= \sum_{l=1}^{r} \mathbb{P}\{X_{n;j_1} = m_1, \dots, X_{n;j_l} = m_l - 1, \dots, X_{n;j_r} = m_r\}\frac{1}{n} + \frac{n-r}{n}\mathbb{P}\{\mathbf{X}_{n;\mathbf{j}_r} = \mathbf{m}_r\},
\end{aligned} \tag{9.20}$$

where we use the notation $n <_c k$ for the event that node n is a child of node k and consequently $n \not<_c k$ for the event that node n is not a child of node k. □

Corollary 18 follows immediately by setting $v = v_1 = \dots = v_l$ in Theorem 56. Also Corollary 21 is easily obtained by replacing v_l by v_l^k in Theorem 56.

9.3.2 Deriving the covariances and the limiting distribution

Proof of Theorem 57. To obtain the covariances of $X_{n;j_g}X_{n;j_h}$ we use

$$\begin{aligned} p_{n;\mathbf{j}_r}(\mathbf{v}_r) &= \prod_{i=1}^{r} \frac{\binom{j_{i+1}-1-i+\sum_{l=1}^{i} v_l}{j_{i+1}-1}}{\binom{j_i-1-i+\sum_{l=1}^{i} v_l}{j_i-1}} = \prod_{i=1}^{r} \frac{\binom{j_{i+1}-1-i+\sum_{l=1}^{i} v_l}{j_{i+1}-j_i}}{\binom{j_{i+1}-1}{j_i-1}} \\ &= \prod_{i=1}^{r} \frac{[u^{j_{i+1}-j_i}]}{\binom{j_{i+1}-1}{j_i-1}} \frac{1}{(1-u)^{\sum_{l=1}^{i} v_l+j_i-i}}. \end{aligned} \tag{9.21}$$

Hence we get

$$\begin{aligned} &\mathbb{E}(X_{n;j_g}X_{n;j_h}) = E_{\mathbf{v}_r}D_{v_g}D_{v_h}p_{n;j_1,\dots,j_r}(v_1,\dots,v_r) = E_{\mathbf{v}_r}D_{v_g}D_{v_h}\prod_{i=1}^{r} \frac{[u^{j_{i+1}-j_i}]}{\binom{j_{i+1}-1}{j_i-1}} \frac{1}{(1-u)^{\sum_{l=1}^{i} v_l+j_i-i}} \\ &= \Big(E_{\mathbf{v}_r}\prod_{i=1}^{g-1} \frac{[u^{j_{i+1}-j_i}]}{\binom{j_{i+1}-1}{j_i-1}} \frac{1}{(1-u)^{\sum_{l=1}^{i} v_l+j_i-i}}\Big)\Big(E_{\boldsymbol{v}}D_{v_g}D_{v_h}\prod_{i=g}^{r-1} \frac{[u^{j_{i+1}-j_i}]}{\binom{j_{i+1}-1}{j_i-1}} \frac{1}{(1-u)^{\sum_{l=1}^{i} v_l+j_i-i}}\Big) \\ &= \sum_{i=g}^{r}(H_{j_{i+1}-1}-H_{j_i-1})\sum_{\substack{l=h\\ l\neq i}}^{r}(H_{j_{l+1}-1}-H_{j_l-1}) + \sum_{i=h}^{r}\big((H_{j_{i+1}-1}-H_{j_i-1})^2-(H^{(2)}_{j_{i+1}-1}-H^{(2)}_{j_i-1})\big) \\ &= (H_{j_h-1}-H_{j_g-1})(H_{n-1}-H_{j_h-1}) + (H_{n-1}-H_{j_h-1})^2 - \sum_{i=h}^{r}(H_{j_{i+1}-1}-H_{j_i-1})^2 \\ &\quad + \sum_{i=h}^{r}\big((H_{j_{i+1}-1}-H_{j_i-1})^2-(H^{(2)}_{j_{i+1}-1}-H^{(2)}_{j_i-1})\big) \\ &= (H_{n-1}-H_{j_g-1})(H_{n-1}-H_{j_h-1}) - (H^{(2)}_{n-1}-H^{(2)}_{j_h-1}). \end{aligned} \tag{9.22}$$

The expectation and variance was already derived in [47].

$$\mathbb{E}(X_{n,j}) = H_{n-1}-H_{j-1}, \quad \mathbb{V}(X_{n,j}) = H_{n-1}-H_{j-1}-(H^{(2)}_{n-1}-H^{(2)}_{j-1}). \tag{9.23}$$

□

Proof of Theorem 58. Let $\mathbf{t}_r = (t_1,\dots,t_r)$ and $\boldsymbol{\sigma}_{n,r} = (\sigma_{n,j_1},\dots,\sigma_{n,j_r})$. Further we use the shorthand notations $\mathbf{t}_r\mathbf{X}^*_{n;\mathbf{j}_r} = \sum_{k=1}^{r} t_k X^*_{n;j_k}$ and $\mathbf{t}_r\boldsymbol{\sigma}_{n,r} = \sum_{k=1}^{r} t_k\sigma_{n,j_k}$. The characteristic function $\varphi_r(\mathbf{t}_r)$ of $\mathbf{X}^*_{n;\mathbf{j}_r}$

$$\varphi_r(\mathbf{t}_r) = \mathbb{E}(\exp\big(i\mathbf{t}_r\mathbf{X}^*_{n;\mathbf{j}_r}\big)) = \exp\big(-i\mathbf{t}_r\boldsymbol{\sigma}_{n,r}\big)p_{n;\mathbf{j}_r}(e^{i\frac{X_{n;j_1}}{\sigma_{n,j_1}}},\dots,e^{i\frac{X_{n;j_r}}{\sigma_{n,j_r}}}) \tag{9.24}$$

By using Stirling's formula for the gamma function

$$\Gamma(z) = \Big(\frac{z}{e}\Big)^z\frac{\sqrt{2\pi}}{\sqrt{z}}\Big(1+\frac{1}{12z}+\frac{1}{288z^2}+\mathcal{O}(\frac{1}{z^3})\Big), \tag{9.25}$$

one gets for fixed $v_1,\dots,v_r$, note that $j_{r+1} := n$,

$$\begin{aligned} p_{n;\mathbf{j}_r}(\mathbf{v}_r) &= \prod_{i=1}^{r} \frac{\binom{j_{i+1}-1-i+\sum_{l=1}^{i} v_l}{j_{i+1}-j_i}}{\binom{j_{i+1}-1}{j_i-1}} \\ &= \exp\Big(\log n(\sum_{l=1}^{r} v_l - r)\Big)\Big(\prod_{i=1}^{r-1} \frac{\binom{j_{i+1}-1-i+\sum_{l=1}^{i} v_l}{j_{i+1}-j_i}}{\binom{j_{i+1}-1}{j_i-1}}\Big)\Big(1+\mathcal{O}(\frac{1}{n})\Big). \end{aligned} \tag{9.26}$$

Using for fixed $j_1, \ldots, j_r$ the expansions

$$e^{\frac{it}{\sigma_{n,j_l}}} - 1 = \frac{it}{\sqrt{\log n}} - \frac{t^2}{2\log n} + \mathcal{O}\Big(\frac{1}{\log^{\frac{3}{2}} n}\Big), \tag{9.27}$$

one obtains

$$p_{n;\mathbf{j}_r}\big(e^{i\frac{X_{n;j_1}}{\sigma_{n,j_1}}}, \ldots, e^{i\frac{X_{n;j_r}}{\sigma_{n,j_r}}}\big) = \exp\Big(\log n\big(\sum_{l=1}^{r}\big(\frac{it_l}{\sqrt{\log n}} - \frac{t_l^2}{\log n}\big)\big)\Big)\Big(1 + \mathcal{O}\big(\frac{\log\log n}{\log n}\big)\Big). \tag{9.28}$$

Hence

$$\exp\big(-i\mathbf{t}_r\boldsymbol{\sigma}_{n,r}\big)p_{n;\mathbf{j}_r}\big(e^{i\frac{X_{n;j_1}}{\sigma_{n,j_1}}}, \ldots, e^{i\frac{X_{n;j_r}}{\sigma_{n,j_r}}}\big) = \exp\Big(-\sum_{k=1}^{r}\frac{t_k^2}{2}\Big)\Big(1 + \mathcal{O}\big(\frac{\log\log n}{\log n}\big)\Big). \tag{9.29}$$

□

Remark 22. The proof of the general case $\max\{j_1, \ldots, j_r\} = o(n)$ follows easily by similar arguments. See also [47] where this was carried out in the case $r = 1$.

Proof of Theorem 59. We only sketch this proof. It is a straight forward extension of the proof of Theorem 21 (the case $r = 1$).

- Region: fixed $r \geq 1$, fixed $j_1, \ldots, j_r$: Poisson approximation or by application of Lévy's continuity theorem for the moment generating function.
- Region: fixed $r \geq 1$, $0 < j_1 = \rho_1 n < \cdots < j_r = \rho_r n < n$: Apply Stirling's formula for the Gamma function to the probability generating function of $X_{n;j_1,\ldots,j_r}$.
- Region: $r = n + 1 - l$, fixed l; Method of Moments.

□

Proof of Corollary 20. One may use the limiting distribution results for the central region of $X_{n;j_1,\ldots,j_r}$, i. e. $0 < j_1 = \rho_1 n < \cdots < j_r = \rho_r n < n$. Due to $X_{n;j_1,\ldots,j_r} \xrightarrow{(d)} X_{\rho_1,\ldots,\rho_r}$ the discrete r. v. X_r can be obtained via

$$\begin{aligned}\mathbb{P}\{X_r = m\} &= \int_0^1 \ldots \int_0^1 \mathbb{P}\{X_{\rho_1,\ldots,\rho_r} = m\} d\rho_1 \ldots d\rho_r \\ &= \int_0^1 \ldots \int_0^1 \rho_1 \ldots \rho_r \frac{(-\log(\rho_1 \ldots \rho_r))^m}{m!} d\rho_1 \ldots d\rho_r.\end{aligned} \tag{9.30}$$

By using

$$\int_0^1 \rho_i(\log(\rho_i))^{k_i} d\rho_i = \frac{k_i!(-1)^{k_i}}{2^{k_i+1}}, \tag{9.31}$$

one further gets

$$\begin{aligned}\mathbb{P}\{X_r = m\} &= \frac{(-1)^m}{m!} \sum_{\substack{k_1+\cdots+k_r=m \\ k_i \geq 0}} \binom{m}{k_1, \ldots, k_r} \int_0^1 \ldots \int_0^1 \rho_1 \ldots \rho_r (\log(\rho_1))^{k_1} \ldots (\log(\rho_r))^{k_r} d\rho_1 \ldots d\rho_r \\ &= \frac{(-1)^m}{m!} \sum_{\substack{k_1+\cdots+k_r=m \\ k_i \geq 0}} \binom{m}{k_1, \ldots, k_r} \frac{k_1!k_2!\ldots k_r!(-1)^{k_1+\cdots+k_r}}{2^{k_1+\cdots+k_r+r}} = \frac{1}{2^r} \sum_{\substack{k_1+\cdots+k_r=m \\ k_i \geq 0}} 1 \\ &= \frac{\binom{m+r-1}{m}}{2^{m+r}}.\end{aligned} \tag{9.32}$$

□

Remark 23. One can immediately obtain results or edge-weighted variants of the node degree. The results and the proofs are analogous to the proof of the weighted depth and distance and are omitted here.

Remark 24. It seems interesting to extend the studies of weighted variants of the node degree to arbitrary grown simple families of increasing trees. Furthermore an analysis of weighted variants of the subtree sizes should be done. A martingale approach seems to be very promising.

Bibliography

[1] R. Aguech, N. Lasmar and H. Mahmoud, Extremal Weighted Path Lengths in Random Binary Search Trees. submitted 2006.

[2] D. Aldous, The Continuum Random Tree II: An Overview, *London Mathematical Society Lecture Note Series* 167, 23–70, 1991.

[3] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, ninth edition, Dover, New York, 1970.

[4] R. Arratia, L. Goldstein, and L. Gordon, Poisson approximation and the Chen-Stein method, *Stat. Sci.* 5, 403434, 1990.

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

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

[7] B. Bollobás and O. M. Riordan, Mathematical results on scale-free random graphs, in *Handbook of graphs and networks*, 1–34, Wiley-VCH, Weinheim, 2003.

[8] D. Borwein, J. M. Borwein and R. Girgensohn, Explicit evaluation of Euler sums, *Proc. Edinburgh Math. Soc.* (2) 38:2, 277-294, 1995.

[9] J. M. Borwein, R. Girgensohn, Evaluation of triple Euler sums, *Electronic Journal of Combinatorics* 3, research paper #23, 1996.

[10] M. Bousquet-Mélou, Limit laws for embedded trees. Applications to the integrated superbrownian excursion, *Random Structures & Algorithms*, to appear.

[11] D. Y. C. Chan, B. D. Hughes, A. S. Leong and W. J. Reed. Stochastically evolving networks. *Physical Review E* 68 06624-106624-24, 2003.

[12] C. Christophi and H. Mahmoud, The oscillatory distribution of distances in random tries, *Annals of Applied Probability* 15, 1536–1564, 2005.

[13] K. L. Chung, *A course in probability theory*, Academic Press, San Diego, 1974.

[14] J. H. Curtiss, A note on the theory of moment generating functions. *Annals of Math. Stat.* 13-4, 430433, 1942.

[15] L. Devroye and R. Neininger, Distances and finger search in random binary search trees, *SIAM Journal on Computing* 33, 647–658, 2004.

[16] R. Dobrow, On the distribution of distances in recursive trees. *Journal of Applied Probability* 33, 749-757, 1996.

[17] R. P. Dobrow and J. A. Fill, Total path length for random recursive trees. *Combin. Probab. Comput.* 8 317 -333, 1999.

[18] R. Dobrow and R. Smythe, Poisson approximations for functionals of random trees. *Random Structures & Algorithms* 9, 79-92, 1996.

[19] M. Drmota, Combinatorics and Asymptotics on trees, *Random Struct. Algorithms* 10, 103-124, 1997.

[20] Feng Q., H. Mahmoud and A. Panholzer. Phase changes in subtree varieties in random recursive and binary search trees.
submitted, 2006.

[21] J. A. Fill, N. Kapur and A. Panholzer, Destruction of very simple trees, submitted, 2004.

[22] P. Flajolet, J. A. Fill and N. Kapur, Singularity Analysis, Hadamard Products, and Tree Recurrences, *Journal of Computational and Applied Mathematics* 174, 271–313, 2005.

[23] P. Flajolet and M. Noy, Analytic combinatorics of non-crossing configurations,
Discrete Mathematics, 204:203229, 1999.

[24] P. Flajolet and A. Odlyzko, The Average Height of Binary Trees and Other Simple Trees, *Journal of Computer and System Sciences* 25, 171–213, 1982.

[25] P. Flajolet and A. Odlyzko, Singularity Analysis of Generating Functions, *SIAM Journal on Discrete Mathematics* 3, 216–240, 1990.

[26] F. Flajolet, B. Salvy (1996), Euler Sums and Contour Integral Representations, *Journal of Experimental Mathematics* 7:1, 15-35, 1998.

[27] Philippe Flajolet und Robert Sedgewick, *An introduction to the Analysis of Algorithms.*
Addison-Wesley, Reading, MA, 1996 .

[28] 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.

[29] C. Goldschmidt and J. B. Martin, Random recursive trees and the Bolthausen-Sznitman coalescent, *Electronic Journal of Probability* 10, paper no. 21, 718–745, 2005.

[30] R. Graham, D. Knuth and O. Patashnik, *Concrete Mathematics*, Second Edition, Addison-Wesley, Reading, 1994.

[31] Daniel H. Greene und Donald E. Knuth, *Mathematics for the Analysis of Algorithms.*
Birkhäuser, Boston, zweite Auflage, 1982 .

[32] M. Hoffman, Multiple harmonic series, *Pacific Journal of Mathematics* 152, 275–290, 1992.

[33] Hu Z, Feng Q and Su C, Uniform recursive trees: Branching structure and simple random downward walk.
Journal of Mathematical Analysis and Applications 315, 225-243, 2006.

[34] H. -. Hwang and R. Neininger. Phase change of limit laws in the quicksort recurrences under varying toll functions.
SIAM Journal on Computing, 31:6, 1687-1722, 2002.

[35] H. -K. Hwang and T. -H. Tsai, Quickselect and Dickman function.
Combinatorics, Probability and Computing, 11:4, 353-371, 2002.

[36] S. Janson, Asymptotic degree distribution in random recursive trees, *Random Structures & Algorithms* 26, 69–83, 2005.

[37] S. Janson, Random cutting and records in deterministic and random trees, submitted, 2006.

[38] M. Javanian and M. Q. Vahidi-Asl, Isolating the most recent entry in a random recursive tree by random cuts, *Journal of Applied Mathematics & Computing* 16, 115–123, 2004.

[39] D. Knuth and A. Schönhage, The expected linearity of a simple equivalence algorithm, *Theoretical Computer Science* 6, 281–315, 1978.

[40] Donald E. Knuth, *Selected Papers on Analysis of Algorithms.* CLSI Publications, 2000 .

[41] Donald E. Knuth, *The Art of Computer Programming: Fundamental Algorithms, volume 1.* Addison-Wesley, Reading, Mass., 3rd edition, 1997.

[42] Donald E. Knuth, *The Art of Computer Programming: Sorting and Searching, volume 3.* Addison-Wesley, Reading, Mass., 2nd edition, 1998.

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

[44] M. Kuba and A. Panholzer, Isolating nodes in recursive trees, submitted, 2005.

[45] M. Kuba and A. Panholzer, On the distribution of distances between specified nodes in increasing trees, submitted, 2005.

[46] M. Kuba and A. Panholzer, Descendants in increasing trees, *Electronic Journal of Combinatorics*, 13-1, 8, 2006.

[47] M. Kuba and A. Panholzer, On the degree distribution of the nodes in increasing trees, submitted, 2005. available at `http://info.tuwien.ac.at/panholzer/nodedegree3.pdf`

[48] J. Lent and H. Mahmoud,*Average-case analysis of multiple quickselect: An algorithm for finding order statistics.* Statistics and Probability Letters, 1995.

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

[50] H. Mahmoud, *Evolution of random search trees*, Wiley, New York, 1992.

[51] H. Mahmoud and R. Neininger, Distribution of distances in random binary search trees, *Annals of Applied Probability* 13, 253–276, 2003.

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

[53] Conrado Martínez, *Partial Quicksort.* First ACM-SIAM Workshop on Analytic Algorithmics and Combinatorics, 2004, (to appear).

[54] A. Meir and J. W. Moon, Cutting down random trees, *Journal of the Australian Mathematical Society* 11, 313–324, 1970.

[55] A. Meir and J. W. Moon, Cutting down recursive trees, *Mathematical Biosciences* 21, 173–181, 1974.

[56] A. Meir and J.W. Moon, Climbing certain types of rooted trees I, *Proc. of the 5th british combinatorial conf.* 461-469, 1975.

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

[58] J. W. Moon, The distance between nodes in recursive trees, *London Mathematical Society Lecture Notes* 13, 125–132, 1974.

[59] T. F. Móri, On random trees. *Studia Sci. Math. Hungar.*, 39, 143155, 2002.

[60] T. F. Móri, The maximum degree of the Barabsi-Albert random tree, *Comb. Probab. Computing*, 14, 339348, 2005.

[61] D. Najock and C. C. Heyde. On the number of terminal vertices in certain random trees with an application to stemma construction in philology.
Journal of Applied Probability 19, 675680, , 1982.

[62] A. Panholzer. The climbing depth of random trees.
Random Structures & Algorithms, 26, 84–109, 2005.

[63] A. Panholzer, Cutting down very simple trees,
accepted for publication in Quaestiones Mathematicae.

[64] A. Panholzer, Non-crossing trees revisited: cutting down and spanning subtrees, in *Discrete Random Walks*, C. Banderier and C. Krattenthaler (eds.), *Discrete Mathematics and Theoretical Computer Science*, Proceedings AC, 265–276, 2003.

[65] A. Panholzer, Destruction of Recursive Trees, *Proceedings of the "Third Colloquium on Mathematics and Computer Science"*, M. Drmota et al (eds.), Birkhäuser, Basel, 267–280, 2004.

[66] A. Panholzer, The distribution of the size of the ancestor-tree and of the induced spanning subtree for random trees, *Random Structures & Algorithms* 25, 179–207, 2004.

[67] A. Panholzer and H. Prodinger, Bijections for Ternary Trees and Non-Crossing Trees,
Discrete Mathematics, 250:181–195, 2002.

[68] A. Panholzer and H. Prodinger, The level of nodes in increasing trees revisited, submitted, 2005. available at `http://info.tuwien.ac.at/panholzer/levelinc3.pdf`

[69] M .D. Penrose and A. R. Wade, Random minimal directed spanning trees and Dickman-type distributions.
Adv. in Appl. Probab. 36, 3, 691714, 2004.

[70] A. Rudas, B. Tóth and B. Valkó, Random Trees and General Branching Processes,
accepted for publication in Random Structures & Algorithms.

[71] J. Szymański, On the nonuniform random recursive tree, *Annals of Discrete Math.*, 33, 297-306, 1987.

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

List of used symbols

$n!$ *factorial*
$n(n-1)(n-2)\ldots 21 = \prod_{j=0}^{n-1}(n-j)$

$n!!$ *double factorial*
$n(n-2)!!$, $0!! = 1!! = 1$

$n^{\underline{k}}$ *falling factorials*
$n(n-1)(n-2)\ldots(n-k+1) = \prod_{j=0}^{k-1}(n-j)$

$n^{\overline{k}}$ *rising factorials*
$n(n+1)(n+2)\ldots(n+k-1) = \prod_{j=0}^{k-1}(n+j)$

H_n *n-th harmonic number*
$\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n} = \sum_{k=1}^{n} \frac{1}{k}$

$H_n^{(r)}$ *n-th harmonic number of r-st order*
$\frac{1}{1^r} + \frac{1}{2^r} + \cdots + \frac{1}{n^r} = \sum_{k=1}^{n} \frac{1}{k^r}$

D_z^k *k-th derivative with respect to z*
$\frac{\partial^k}{\partial z^k}$

E_z *evaluation at $z = 1$*
$(\,.\,)\big|_{z=1}$

$\zeta(z)$ *Riemann's zeta function*
$\sum_{n\geq 1} \frac{1}{n^z}$

$[z^n]f(z)$ *coefficient of z^n*
$f(z) = \sum_{n\geq 0} a_n z^n$, $[z^n]f(z) = a_n$

γ *Euler constant, (Euler-Mascheroni Constant)*
$\gamma = \lim_{n\to\infty}(H_n - \log n) = 0.5772156\ldots$

$\Psi(z)$ *Digamma function or Psi function is the logarithmic derivative of the gamma function*
$\Psi(z) = \frac{d}{dz}\log\big(\Gamma(z)\big)$, $\Psi(n) = H_{n-1} - \gamma$

$\left[{n \atop k}\right]$ *Stirling numbers of the first kind*

$\left\{{n \atop k}\right\}$ *Stirling numbers of the second kind*

$[\![A]\!]$ *Iverson's bracket for the predicate A*
$[\![A]\!] = 1$, if A is true and 0 otherwise

$\mathbb{1}(A)$ *indicator function of the event A*

$\oplus$; $\bigoplus$ *Sum of mutually independent random variables*

$\stackrel{(d)}{=}$, $\xrightarrow{(d)}$ *equality in distribution, convergence in distribution*

List of Figures

List of Tables

Lebenslauf

Ich, Markus Kuba, wurde am 11. November 1980 als Sohn von Huey-Shan Kuba, geborene Hwang, und Walter Kuba in Wien geboren. Von 1986 bis 1990 besuchte ich die Volksschule Graf-Starhemberg-Gasse - Wien 4, und von 1991 bis 1999 das Bundesrealgymnasium Waltergasse - Wien 4. Am 6. Juni 1999 maturierte ich dort mit Auszeichung.

Im Oktober 1999 immatrikulierte ich an der Technischen Universität Wien und inskribierte für das Studium der Technischen Mathematik, Studienzweig Computerwissenschaften. Nach dem Studiumjahr 1999 bis 2000 leiste ich von Oktober 2000 bis September 2001 den Zivildienst im Caritasheim St.Barbara - Wien 23 ab, und war aufgrund von Nichtinskription vom Studium beurlaubt. Oktober 2001 nahm ich mein Studium wieder auf und schloss es am 5.Oktober 2004 mit Auszeichung ab.

Von Oktober 2004 an inskribierte ich das Doktoratsstudium der technischen Wissenschaften. Von Mai 2005 bis Dezember 2005 arbeite ich als Forschungsassistent im FWF-Projekt P18009, Analyse von Datenstrukturen und baumartigen Strukturen, unter der Leitung von Prof. Alois Panholzer. Beginnend mit Jänner 2006 bin ich im Rahmen des National Research Network S9600 Analytic Combinatorics and Probabilistic Number Theory, Netzwerkleiter Prof. Michael Drmota, im Teilprojekt S9608 - Combinatorial Analysis of Data Structures and Tree-Like Structures, Projektleiter Prof. Alois Panholzer als Forschungsassistent tätig.

Wissenschaftliche Arbeiten (veröffentlicht)

• M. Kuba and A. Panholzer. Descendants in increasing trees
The electronic journal of combinatorics, Volume 13 (1), R8, 2006.

Wissenschaftliche Arbeiten (akzeptiert)
• M. Kuba, On Quickselect, Partial sorting and Multiple Quickselect.
akzeptiert zur Veröffentlichung in *Information Processing letters.*