TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DISSERTATION

Analysis of various parameters in labelled trees and tree-like structures

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

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

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Dipl.-Ing. Dipl.-Ing. Georg Seitz
0426015

Meiselstraße 8/2/14
1150 Wien

Wien, im November 2011

Kurzfassung

In der vorliegenden Dissertation analysieren wir diverse Parameter in markierten Bäumen (d.h. in Bäumen, in denen die Knoten mit unterschiedlichen Zahlen beschriftet sind) und allgemeiner in baumartigen Graphen.

In Bäumen betrachten wir zwei Parameter, die schon ausgiebig in endlichen Zahlenfolgen (vor allem in Permutationen) studiert wurden, nämlich die Anzahl der Inversionen und die Anzahl der lokalen Minima. Wir analysieren das Verhalten dieser Kenngrößen in zufälligen Bäumen unterschiedlicher markierter Baumfamilien.

Des weiteren betrachten wir eine Klasse von Graphen, die als „k-trees“ bezeichnet wird. Das sind Graphen mit einer baumartigen Struktur, die man durch einen randomisierten Graph-Evolutions-Prozess konstruieren kann, und die als einfaches Modell für das Wachstum komplexer Netzwerke angesehen werden können. In diesen Graphen studieren wir Parameter, die bereits in gewöhnlichen Bäumen analysiert wurden, nämlich die Anzahl der Vorfahren und Nachkommen von Knoten und die Gradverteilung.

Wir erhalten weitgehend sehr präzise Resultate über die exakte Verteilung dieser Parameter für jede fixe Größe der betrachteten zufälligen Objekte. Außerdem können wir das Grenzverhalten jedes dieser Parameter charakterisieren, wenn die Größe der betrachteten Objekte gegen unendlich geht.

Diese Dissertation basiert auf Forschungsarbeiten, die gemeinsam mit Alois Panholzer verfasst wurden, siehe [PS10a, PS10b, PS11, BPS]. Die Arbeit wurde durch den Österreichischen Wissenschaftsfond FWF, Forschungsprojekt S9608, “Combinatorial Analysis of Data Structures and Tree-Like Structures”, unterstützt.

Abstract

In this thesis, we analyse various parameters in labelled trees (i.e. in trees in which the nodes are labelled with distinct integers) and, more generally, in labelled tree-like graph structures.

In trees, we consider two parameters which have already been extensively studied in finite sequences (especially in permutations), namely the number of inversions and the number of local minima. We analyse the behaviour of these quantities in random trees of different labelled tree families.

Moreover, we consider a class of graphs known as k-trees. These are graphs with a tree-like structure, which can be constructed by a certain randomised graph evolution process, and which can be seen as a simple model for the growth of complex networks. In these graphs, we study parameters which have already been analysed in ordinary trees, namely the number of ancestors and descendants of nodes and the degree distribution.

To a large extent we obtain very precise results on the exact distribution of these quantities if the size of the considered random objects is fixed. Apart from that, we can characterize the limiting behaviour of each of these quantities as the size of the considered objects tends to infinity.

This thesis is based on research papers jointly written with Alois Panholzer, see [PS10a, PS10b, PS11, BPS]. The research has been supported by the Austrian Science Foundation FWF, research project S9608, "Combinatorial Analysis of Data Structures and Tree-Like Structures".

Acknowledgements

I would like to express my thanks to those people who contributed to this thesis in one way or another.

First of all, I would like to thank Alois Panholzer, who played an important role in awakening my interest in discrete mathematics and combinatorics. Without him, I would certainly not have started my Ph.D. studies in the first place. Furthermore, I would like to thank Ilse Fischer, who agreed without hesitation to review this thesis.

Next, I want to mention my colleagues (in alphabetical order) Marie-Louise "Mimi" Bruner, Veronika "Veri" Kraus, Benoît "Benni" Loridant, and Johannes "Hannes" Morgenbesser, who never let me run out of sweets and assured that my caffeine level was always high enough. Apart from that, I should note that Corollary 5.3.1 is based on an idea by Hannes.

Moreover, I want to thank my parents, who financed my diploma studies and thus made it possible for me to finance my Ph.D. studies myself. Last but not least, I'd like to put forward my girlfriend Sabine Palatin, who provided the necessary non-mathematical contrast to my studies, and yet always smiles and nods when I try to tell her about another interesting mathematical fact :-).

Contents

CHAPTER 1

Introduction

Trees appear in various different contexts, in practical applications as well as in theoretical considerations. For example, they appear as data structures in computer science, as game models in game theory, and in biology they are used to represent evolutionary relationships among various species. It is therefore not surprising that trees themselves are a major subject of current research. In this thesis, we present our contribution to the topic, namely the analysis of some parameters in trees and, more generally, in certain tree-like structures known as k-trees. In the following, we give a short overview of our work.

In Chapter 2, we collect the preliminaries which will be needed throughout this thesis. Since almost all of our results will be obtained by means of generating functions, we introduce the required tools and auxiliary results which will be used when working with those functions, such as the symbolic method and singularity analysis. Moreover, we collect some probabilistic tools, namely some facts on probability and moment generating functions and criteria for the weak convergence of random variables.

In Chapter 3, we study the number of *inversions* in random trees. The notion of inversions was originally introduced for permutations $p_1 p_2 \dots p_n$, where an inversion denotes a pair of entries (p_i, p_j) such that $i < j$ and $p_i > p_j$. The number of inversions in permutations appears very naturally in the analysis of certain sorting algorithms, since it counts the number of pairs whose relative order has to be changed in order to obtain the increasingly sorted sequence $1\,2\,3 \dots n$. In analogy to such increasing sequences, one can define *increasing trees* to be those rooted labelled trees in which the labels along each path starting at the root form an increasing sequence. Moreover, in analogy to inversions in permutations, an inversion in a tree denotes a pair of labels (ℓ_i, ℓ_j), such that ℓ_i lies on the path from ℓ_j to the root and $\ell_i > \ell_j$. The number of inversions then measures the degree to which a tree is *not* increasing.

In our studies, we determine the limiting behaviour of the total number of inversions in a random tree, where we use as tree model any family of the large class of so-called *simply generated tree families*. Ordered and unordered trees, d-ary trees, cyclic trees, and Motzkin trees are some well-known instances of this class. Apart from the total number of inversions, we also study the number of inversions *induced by a specific label* j, i.e. the number of inversions of the form (i, j), where i lies on the path from node j to the root and $i > j$. We fully characterize the limiting behaviour of this quantity in a random tree of size n of a simply generated tree family, not only for fixed j but also under the assumption that $j = j(n)$ grows with n. Moreover, we compute the exact distribution of this parameter for fixed j and n in some special instances.

In Chapter 4, we consider so-called *increasing k-trees*. These are graphs with a tree-like structure, which can be constructed by a certain randomised graph evolution process, and which can be seen as a simple model for the growth of complex networks.

We analyse different parameters in these k-trees, which are based on "inheritance relations" between the nodes in these growing networks. One such parameter is the number of *descendants* of a node v, i.e. the number of nodes which have joined the network directly or indirectly via v. Another considered parameter is the number of *ancestors* of a node v, i.e. the number of nodes of which v is a descendant. Moreover, we analyse the *node degree* of a node v, i.e. the number of nodes which are directly connected to v. This will also give us a corollary on the so-called *clustering coefficient* (which is considered as an important parameter in the study of real-world networks) in increasing k-trees.

We study all these parameters in three different families of increasing k-trees, which are constructed by different rules for the evolution process. The first rule may be called "uniform attachment" and assumes that for every node that newly joins the network each of the possible locations to which it can connect is chosen equally likely. For the second rule, which is called "preferential attachment" (or "success breeds success"), one assumes that a possible location which has already been chosen by many nodes will have a higher probability to acquire even more new nodes. The third rule assumes a "saturation property", which means that there is a maximum number of nodes which can connect to each location, and the probability that a location is chosen by a new node decreases with increasing saturation.

For each of these growth rules we compute the exact distribution of all mentioned parameters both for a fixed node v and a randomly chosen node, and additionally determine their limiting behaviour.

In Chapter 5, we consider another quantity in trees, namely the number of *local minima*. A local minimum in a tree is a node v with label $l(v)$ with the property that the label $l(w)$ of each neighbour (i.e. adjacent node) w of v satisfies $l(w) > l(v)$. Like the number of inversions, also this parameter has already been considered for permutations, and it is natural to extend this study to trees. As tree models we consider both ordered and unordered trees and study the number of local minima in a random tree of size n of the respective family. We both characterize the exact distribution of this quantity for fixed n and its limiting behaviour for $n \to \infty$.

Apart from these probabilistic considerations, we also show that there are interesting connections between the number of trees of size n with m local minima and certain other combinatorial quantities involving the same tree families, namely the number of unordered trees of size n with m leaf nodes, and the number of so-called *up-down alternating trees* of size n in the respective tree family.

CHAPTER 2

Preliminaries

2.1 Generating functions

Throughout this thesis we will work with *generating functions*: If $(a_n)_{n\geq 0}$ is a sequence of real numbers, then the *(ordinary) generating function* of $(a_n)_{n\geq 0}$ is the (formal) power series

$$A(z) = \sum_{n\geq 0} a_n z^n. \tag{2.1}$$

Furthermore, the (formal) power series

$$\hat{A}(z) = \sum_{n\geq 0} a_n \frac{z^n}{n!}, \tag{2.2}$$

is called *exponential generating function* of the sequence $(a_n)_{n\geq 0}$.

2.1.1 Combinatorial constructions and the symbolic method

In our studies, the numbers a_n in (2.1) and (2.2) will mostly be given by the sizes $a_n = |\mathcal{A}_n|$ of sets $\mathcal{A}_n$ of combinatorial objects. The *symbolic method* from [FS09] is a powerful tool in this context, which allows to translate certain combinatorial constructions directly to equations for generating functions.

We consider classes of combinatorial objects, in which each object α has a *size* $|\alpha|$ (which is just a natural number assigned to that object, e.g., if the considered objects are graphs, then $|\alpha|$ could for example denote the number of nodes of α). If $\mathcal{A}$ is such a class, we denote the class of all objects of $\mathcal{A}$ of size n by $\mathcal{A}_n$. Furthermore we assume that the objects in $\mathcal{A}$ are labelled, i.e. that each object of size n of $\mathcal{A}$ contains n "atoms" (e.g., the nodes in a graph), onto which the numbers $1, \dots, n$ are distributed (in [FS09], such objects are called *well-labelled*). We sometimes also consider objects in which the atoms are labelled with arbitrary distinct integers. These objects will then be called *weakly labelled* objects.

The exponential generating function $A(z)$ *associated with the class* $\mathcal{A}$ is the exponential generating function of the sequence $(|\mathcal{A}_n|)_{n\geq 0}$, i.e.

$$A(z) = \sum_{n\geq 0} |\mathcal{A}_n| \frac{z^n}{n!} = \sum_{\alpha\in\mathcal{A}} \frac{z^{|\alpha|}}{|\alpha|!}.$$

In the following, we collect some of the constructions from [FS09] (called "admissible constructions") which make it possible to build complex labelled classes from simpler ones, and which allow for a direct translation into equations for the associated exponential generating functions:

- Disjoint union, which we denote by $\dot\cup$: If $\mathcal{A}$ and $\mathcal{B}$ are two disjoint labelled classes, then $\mathcal{A} \,\dot\cup\, \mathcal{B}$ just denotes the union (in the set-theoretic sense) of $\mathcal{A}$ and $\mathcal{B}$. However, $\dot\cup$ can also be defined for not necessarily disjoint classes: If $\mathcal{A}$ and $\mathcal{B}$ are two arbitrary labelled classes, then $\mathcal{A} \,\dot\cup\, \mathcal{B}$ denotes the union of two *disjoint copies* of $\mathcal{A}$ and $\mathcal{B}$. Such disjoint copies can, e.g., be constructed by choosing two distinct colors, and painting the elements of $\mathcal{A}$ with the first color, and the elements of $\mathcal{B}$ with the second one.

- Labelled product: The labelled product $\beta * \gamma$ of two labelled objects β and γ is a set of objects (each of size $|\beta| + |\gamma|$), defined by

  $$\beta * \gamma := \{(\beta', \gamma') \mid (\beta', \gamma') \text{ is well-labelled, } \rho(\beta') = \beta, \rho(\gamma') = \gamma\},$$

  where ρ relabels each weakly labelled object of size n such that exactly the integers $\{1, \dots, n\}$ are used, while preserving the relative order of all labels. The labelled product $\mathcal{B} * \mathcal{C}$ of two labelled classes $\mathcal{B}$ and $\mathcal{C}$ is defined by

  $$\mathcal{B} * \mathcal{C} := \bigcup_{\beta\in\mathcal{B},\gamma\in\mathcal{C}} (\beta * \gamma).$$

- Sequences: We let $\textsc{Seq}(\mathcal{B})$ denote the sequence class of $\mathcal{B}$, which is defined by
$$\textsc{Seq}(\mathcal{B}) := \{\epsilon\} \,\dot\cup\, \mathcal{B} \,\dot\cup\, \mathcal{B} * \mathcal{B} \,\dot\cup\, \mathcal{B} * \mathcal{B} * \mathcal{B} \,\dot\cup\, \ldots = \bigcup_{k\geq 0} \mathcal{B}^k,$$
where ϵ denotes the empty sequence.

- Sets: We let $\textsc{Set}(\mathcal{B})$ denote the set class of $\mathcal{B}$, which is defined as the quotient $\textsc{Seq}(\mathcal{B})/\mathbf{R}$, where the equivalence relation $\mathbf{R}$ identifies sequences of equal length whenever the components of one are a permutation of the components of the other.

- Cycles: We let $\textsc{Cyc}(\mathcal{B})$ denote the cycle class of $\mathcal{B}$, which is defined as the quotient $(\textsc{Seq}(\mathcal{B}) \setminus \{\epsilon\})/\mathbf{R}$, where the equivalence relation $\mathbf{R}$ identifies sequences of equal length whenever one can be obtained from the other by cyclically shifting the components of the other (i.e. $(\beta_1, \ldots, \beta_k)\mathbf{R}(\gamma_1, \ldots, \gamma_\ell)$ iff $(\beta_s, \ldots, \beta_k, \beta_1, \ldots, \beta_{s-1}) = (\gamma_1, \ldots, \gamma_\ell)$ for some s).

- Boxed product: The boxed product $\mathcal{B}^\square * \mathcal{C}$ denotes the subclass of $\mathcal{B} * \mathcal{C}$ which consists of those elements where the smallest label lies in the $\mathcal{B}$ component.

The following theorem states how the combinatorial constructions above can directly be translated to equations for the associated generating functions:

Theorem 2.1.1 (Symbolic method [FS09])**.** *Let $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ be classes of labelled objects, and $A(z)$, $B(z)$, $C(z)$, respectively, the associated exponential generating functions. Then the following holds:*

$$\begin{array}{lll}
\textit{If } \mathcal{A} = \mathcal{B} \,\dot\cup\, \mathcal{C}, & \textit{then} & A(z) = B(z) + C(z), \\
\textit{if } \mathcal{A} = \mathcal{B} * \mathcal{C}, & \textit{then} & A(z) = B(z) \cdot C(z), \\
\textit{if } \mathcal{A} = \textsc{Seq}(\mathcal{B}) \;\textit{ and } \mathcal{B}_0 = \emptyset, & \textit{then} & A(z) = \dfrac{1}{1 - B(z)}, \\
\textit{if } \mathcal{A} = \textsc{Set}(\mathcal{B}) \;\textit{ and } \mathcal{B}_0 = \emptyset, & \textit{then} & A(z) = \exp(B(z)), \\
\textit{if } \mathcal{A} = \textsc{Cyc}(\mathcal{B}) \;\textit{ and } \mathcal{B}_0 = \emptyset, & \textit{then} & A(z) = \log\left(\dfrac{1}{1 - B(z)}\right), \\
\textit{if } \mathcal{A} = \mathcal{B}^\square * \mathcal{C}, & \textit{then} & \mathrm{D}_z A(z) = (\mathrm{D}_z B(z)) \cdot C(z),
\end{array}$$

where $\mathrm{D}_z := \frac{\partial}{\partial z}$.

However, in our studies we will mostly be interested in enumerating combinatorial objects of a class $\mathcal{A}$ not only according to their size but in addition according to some parameter $\chi = (\chi_1, \dots, \chi_d)$, which is just a function which maps every element $\alpha \in \mathcal{A}$ to a d-tuple $\chi(\alpha) = (\chi_1(\alpha), \dots, \chi_d(\alpha))$ of natural numbers (in our studies, we will mostly have $d = 1$).

Given a combinatorial class $\mathcal{A}$ and a d-dimensional parameter χ on $\mathcal{A}$, the exponential generating function of the pair $\langle \mathcal{A}, \chi \rangle$ is defined by

$$A(z, \mathbf{u}) := \sum_{\alpha \in \mathcal{A}} \mathbf{u}^{\chi(\alpha)} \frac{z^{|\alpha|}}{|\alpha|!},$$

where $\mathbf{u} := (u_1, \dots, u_d)$ and $\mathbf{u}^{(k_1, \dots, k_d)} := u_1^{k_1} \cdots u_d^{k_d}$. One also says that the variable z in $A(z, \mathbf{u})$ marks the size whereas u_j marks χ_j.

When working with such multivariate generating functions, one can use an extension of the symbolic method. We restrict our attention to so-called *compatible* parameters, that is, parameters whose values are invariant under order-preserving relabellings of objects (i.e. parameters which do not depend on the absolute value of labels, but may for example depend on their relative order). Moreover, we will only consider *inherited* parameters: Given three pairs of combinatorial classes with d-dimensional parameters $\langle \mathcal{A}, \chi \rangle$, $\langle \mathcal{B}, \xi \rangle$, and $\langle \mathcal{C}, \zeta \rangle$, the parameter χ is said to be inherited in the following cases:

- If $\mathcal{A} = \mathcal{B} \dot{\cup} \mathcal{C}$, then χ is inherited from ξ and ζ iff

$$\chi(\alpha) = \begin{cases} \xi(\alpha), & \text{if } \alpha \in \mathcal{B}, \\ \zeta(\alpha) & \text{if } \alpha \in \mathcal{C}, \end{cases}$$

  for all $\alpha \in \mathcal{A}$.

- If $\mathcal{A} = \mathcal{B} * \mathcal{C}$, then χ is inherited from ξ and ζ iff its value is obtained additively from ξ and ζ, i.e.

$$\chi((\beta, \gamma)) = \xi(\beta) + \zeta(\gamma),$$

  for all $(\beta, \gamma) \in \mathcal{A}$.

- If $\mathcal{A}$ is $\textsc{Seq}(\mathcal{B})$, $\textsc{Set}(\mathcal{B})$, or $\textsc{Cyc}(\mathcal{B})$, then χ is inherited from ξ iff its value is obtained additively from the values of ξ on components.

Using the above definitions, the statement of Theorem 2.1.1 can in the case of compatible inherited parameters be directly extended to multivariate generating functions:

Theorem 2.1.2 (Symbolic method with inherited parameters [FS09])**.** *Let $\langle \mathcal{A}, \chi \rangle$, $\langle \mathcal{B}, \xi \rangle$, and $\langle \mathcal{C}, \zeta \rangle$ be three combinatorial classes with d-dimensional parameters. If χ is inherited from ζ and (as the case may be) from ζ, then the implications in Theorem 2.1.1 hold analogously for the associated multivariate generating functions $A(z, \mathbf{u})$, $B(z, \mathbf{u})$, and $C(z, \mathbf{u})$.*

2.1.2 Operators

In order to simplify notation when working with power series, we define the following operators:

- The operator N_z evaluates $G(z)$ at $z = 0$, i.e. $\mathrm{N}_z G(z) := G(0)$.
- The operator U_z evaluates $G(z)$ at $z = 1$, i.e. $\mathrm{U}_z G(z) := G(1)$.
- $\mathrm{D}_z := \frac{\partial}{\partial z}$ is the differential operator with respect to z.
- The operator Z just multiplies a power series with z, i.e. $\mathrm{Z}G(z) := zG(z)$ (for variables other than z we will never use an operator of this kind).

Note that we mark the operators N, U, and D with the corresponding variable as a subindex, such that we can savely apply them to multivariate power series (e.g. $\mathrm{U}_q G(z, q) := G(z, 1)$).

2.1.3 Extraction of coefficients

Having computed the generating function $A(z)$ associated to some combinatorial class $\mathcal{A}$, one is of course interested in the coefficients of $A(z)$. Throughout this thesis, we denote by $[z^n]$ the operator which extracts the coefficient of z^n from a (formal) power series, i.e. $[z^n]A(z) = a_n$, if $A(z) = \sum_{k \geq 0} a_k z^k$. If the radius of convergence of a power series $A(z)$ is positive, then the coefficient $[z^n]A(z)$ can in principle be obtained from *Cauchy's integral formula*,

$$[z^n]A(z) = \frac{1}{2\pi i} \oint \frac{A(z)}{z^{n+1}} dz, \tag{2.3}$$

where the integral is taken in counter-clockwise direction along a simple closed curve around the origin which lies completely inside the circle of convergence of $A(z)$. However, in our studies we will mostly be able to extract coefficients by using one of the following well-known power series expansions,

$$(1+z)^\alpha = \sum_{n\geq 0} \binom{\alpha}{n} z^n, \qquad \frac{1}{(1-z)^{\alpha+1}} = \sum_{n\geq 0} \binom{n+\alpha}{n} z^n,$$
$$\exp(z) = \sum_{n\geq 0} \frac{z^n}{n!}, \qquad \log\left(\frac{1}{1-z}\right) = \sum_{n\geq 1} \frac{z^n}{n},$$

and use (2.3) only in order to arrive at one of these expansions by applying suitable substitutions. Furthermore, the following bivariate generating function, which involves the (unsigned) Stirling numbers of the first kind $\left[{i \atop j}\right]$, will occur (cf., e.g., [FS09]):

$$\frac{1}{(1-z)^v} = \sum_{i\geq 0} \sum_{j=0}^{i} \left[{i \atop j}\right] \frac{z^i}{i!} v^j, \tag{2.4}$$

On some occasions, we will also deal with generating functions $T(z)$ which are implicitly given by a functional equation of the form $T(z) = z\varphi(T(z))$. In order to extract coefficients of such a function, the following theorem proves useful:

Theorem 2.1.3 (Lagrange's inversion formula [FS09])**.** *Let $T(z)$ and $\varphi(x)$ be formal power series which satisfy $T(z) = z\varphi(T(z))$ and $[x^0]\varphi(x) \neq 0$. Then one has*

$$[z^n]f(T(z)) = \begin{cases} \frac{1}{n}[T^{n-1}]f'(x)(\varphi(T))^n, & n > 0 \\ [T^0]f(T), & n = 0, \end{cases}$$

for every formal power series $f(x)$.

2.1.4 Singularity analysis

Singularity analysis is a technique which allows to extract asymptotic expansions of the coefficients of power series which have isolated singularities on the boundary of their disc of convergence. If the power series under consideration has a unique dominant singularity ρ, one may of course use the scaling rule

$$[z^n]f(z) = \rho^{-n}[z^n]f(\rho z),$$

and hence it is sufficient to consider the case where the singularity is at 1. In this case, the following theorem applies:

Theorem 2.1.4 (Big-Oh transfer [FO90])**.** *Let $R > 1$ and $0 < \phi < \frac{\pi}{2}$, and assume that $f(z)$ is analytic in the domain*

$$\Delta = \Delta(\phi, R) := \{z \mid |z| < R, z \neq 1, |\operatorname{Arg}(z-1)| > \phi\}$$

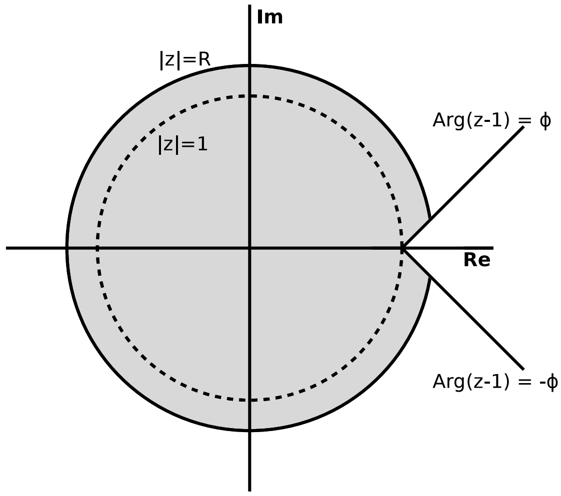

Figure 2.1: Sketch of a Δ-domain.

(see Figure 2.1). Furthermore, assume that, as $z \to 1$ in Δ,

$$f(z) = \mathcal{O}(|1-z|^{\alpha}),$$

for some constant $\alpha \in \mathbb{R}$. Then, as $n \to \infty$,

$$[z^n]f(z) = \mathcal{O}\big(n^{-\alpha-1}\big).$$

A direct consequence of this is the following:

Theorem 2.1.5 (Singularity analysis [FO90]). *Assume that $f(z)$ is analytic in the domain Δ from Theorem 2.1.4, and that, as $z \to 1$ in Δ,*

$$f(z) = \sum_{j=0}^{m} c_j (1-z)^{\alpha_j} + \mathcal{O}\big(|1-z|^{A}\big),$$

for real numbers $\alpha_0 \leq \alpha_1 \leq \ldots \leq \alpha_m < A$ and a sequence of complex numbers $(c_j)_{0 \leq j \leq m}$. Then, as $n \to \infty$,

$$[z^n]f(z) = \sum_{j=0}^{m} c_j \binom{n-\alpha_j-1}{n} + \mathcal{O}\big(n^{-A-1}\big).$$

A similar theorem applies for functions with a finite number of dominant singularities, in which case one can analyse each singularity separately and then add up all contributions. For details, we refer to [FS09].

In our studies we will sometimes need to differentiate functions of which only singular expansions are available. On these occasions, the following theorem comes in handy:

Theorem 2.1.6 (Differentiation of singular expansions [FFK05]). *Assume that $f(z)$ satisfies the conditions of Theorem 2.1.5. Then, for each $r \in \mathbb{N}$, the r-th derivate $f^{(r)}(z)$ of $f(z)$ is also analytic in Δ and admits an expansion obtained through term-by-term differentiation:*

$$f^{(r)}(z) = (-1)^r \sum_{j=0}^{m} c_j \frac{\Gamma(\alpha_j + 1)}{\Gamma(\alpha_j + 1 - r)} (1-z)^{\alpha_j - r} + \mathcal{O}\big(|1-z|^{A-r}\big).$$

2.2 Probabilistic tools

In this section we collect some probabilistic tools (cf., e.g., [Bil84]) which we will use throughout this thesis.

2.2.1 Probabilities, moments, and cumulants

Let X be a discrete random variable supported by $\mathbb{N}_0$. Then we define the following three generating functions which are closely related to the distribution of X:

- The function $p_X(z) := \mathbb{E}\big(z^X\big)$ is called *probability generating function* of X. This name derives from the fact that $p_X(z)$ is the (ordinary) generating function of the sequence of probabilities $(\mathbb{P}\{X = \ell\})_{\ell \geq 0}$, i.e.

  $$p_X(z) = \sum_{\ell \geq 0} \mathbb{P}\{X = \ell\}\, z^\ell.$$

  If $p_X(z)$ exists in a neighbourhood of $z = 1$, then the *factorial moments* $\mathbb{E}(X^{\underline{n}})$ of X (where $x^{\underline{0}} := 1$, and $x^{\underline{n}} := x(x-1)\cdots(x-n+1)$, for $n \geq 1$) can be computed using $p_X(z)$, via

  $$\mathbb{E}(X^{\underline{n}}) = \mathrm{U}_z \mathrm{D}_z^n p_X(z). \tag{2.5}$$

- The *moment generating function* $m_X(t)$ of X is defined by

  $$m_X(t) := \mathbb{E}\big(\mathrm{e}^{tX}\big) = p_X(\mathrm{e}^t) = \sum_{\ell \geq 0} \mathbb{P}\{X = \ell\}\, \mathrm{e}^{t\ell}.$$

Its name is justified by the fact that, if $m_X(t)$ exists in a neighbourhood of $t = 0$, then

$$\mathrm{N}_t \mathrm{D}_t^n m_X(t) = \mathbb{E}(X^n)\,,$$

that is,

$$m_X(t) = \sum_{\ell \geq 0} \mathbb{E}\big(X^\ell\big)\, \frac{t^\ell}{\ell!},$$

i.e. $m_X(t)$ is the exponential generating function of the sequence of moments $(\mathbb{E}\big(X^\ell\big))_{\ell \geq 0}$.

- The function

  $$C_X(t) := \log(m_X(t)) = \log\left(\mathbb{E}\big(\mathrm{e}^{tX}\big)\right),$$

  is called *cumulant generating function* of X. The coefficient $\kappa_\ell(X)$ in the expansion

  $$C_X(t) = \sum_{\ell \geq 1} \kappa_\ell(X)\, \frac{t^\ell}{\ell!},$$

  is the *ℓ-th cumulant of* X. In particular, the first two cumulants of X are the expected value and the variance of X, respectively, i.e. one has $\kappa_1(X) = \mathbb{E}(X)$ and $\kappa_2(X) = \mathbb{V}(X)$.

  For the sum $X + Y$ of two independent random variables X and Y, there holds $\kappa_\ell(X + Y) = \kappa_\ell(X) + \kappa_\ell(Y)$ for all $\ell \in \mathbb{N}$. Moreover, if c is a constant, one has $\kappa_\ell(cX) = c^\ell \kappa_\ell(X)$ for all $\ell \in \mathbb{N}$.

We will use the following relations between (ordinary) moments, factorial moments, and cumulants:

- The ordinary moments $\mathbb{E}(X^r)$ of every random variable X can of course be computed by a linear combination of its factorial moments $\mathbb{E}\big(X^{\underline{j}}\big)$. More precisely,

  $$\mathbb{E}(X^r) = \sum_{j=0}^{r} \left\{ {r \atop j} \right\} \mathbb{E}\big(X^{\underline{j}}\big)\,, \tag{2.6}$$

  where the numbers $\left\{ {r \atop j} \right\}$ are the Stirling numbers of the second kind.

- Each cumulant $\kappa_\ell(X)$ of X can be expressed as a polynomial in the moments $\mathbb{E}(X)\,, \mathbb{E}(X^2)\,, \ldots, \mathbb{E}\big(X^\ell\big)$, and vice versa. For details, we refer to [Bil84].

2.2.2 Convergence in distribution

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables, and $(F_n(x))_{n\in\mathbb{N}}$ the respective distribution functions. Furthermore, let X be another random variable with distribution function $F(x)$. The sequence $(X_n)_{n\in\mathbb{N}}$ is said to *converge in distribution* to X iff

$$\lim_{n\to\infty} F_n(x) = F(x),$$

for each $x \in \mathbb{R}$ at which F is continuous. We then also say that $(X_n)_{n\in\mathbb{N}}$ *converges weakly* to X, and we write

$$X_n \xrightarrow{(d)} X.$$

2.2.2.1 Tools for proving convergence in distribution

In our studies we will mainly use the *method of moments* in order to show convergence in distribution of a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables to another random variable X, i.e. we will show that for all $r \in \mathbb{N}$ the sequence of r-th moments $(\mathbb{E}(X_n^r))_{n\in\mathbb{N}}$ converges to $\mathbb{E}(X^r)$. Given that the distribution of X is uniquely determined by its moments, the convergence in distribution then follows directly from the following theorem:

Theorem 2.2.1 (Fréchet and Shohat [Loè77])**.** *Let X and $X_1, X_2, \ldots$ be random variables. If $\lim_{n\to\infty} \mathbb{E}(X_n^r) = \mathbb{E}(X^r)$ for all $r \in \mathbb{N}$ and the sequence of moments $(\mathbb{E}(X^r))_{r\in\mathbb{N}}$ uniquely determines the distribution of X, then $X_n \xrightarrow{(d)} X$.*

Regarding the question whether a sequence of moments uniquely determines a distribution, we will make use of the following sufficient condition:

Lemma 2.2.2 (Uniqueness of distribution [Bil84])**.** *Let μ be a probability measure on $\mathbb{R}$ having finite moments α_k of all orders. If the power series $\sum_{k\geq 0} \alpha_k \frac{t^k}{k!}$ has a positive radius of convergence, then μ is the only probability measure with the sequence of moments $(\alpha_k)_{k\in\mathbb{N}}$.*

In our studies, we will mostly show the convergence of moments in an indirect way, by using one of the following two ideas:

- Since our studies will mainly be based on probability generating functions, it will often be easier to study the asymptotics of the factorial moments $\mathbb{E}(X_n^{\underline{r}})$ (which can be obtained from the probability generating function of X_n using equation (2.5)) instead of studying the ordinary moments $\mathbb{E}(X_n^r)$ directly.

But, as already noted, the ordinary and factorial moments are related by equation (2.6) (where, in particular, ${r \brace r} = 1$). Hence, if we have $\mathbb{E}(X_n^{\underline{r}}) \gg \mathbb{E}\Big(X_n^{\underline{r-1}}\Big)$ for all $r \in \mathbb{N}$ (which is always satisfied in our studies), it follows that the ordinary and factorial moments are asymptotically equivalent, i.e. for all $r \in \mathbb{N}$ holds

$$\mathbb{E}(X_n^r) = \mathbb{E}(X_n^{\underline{r}}) + \mathcal{O}\Big(\mathbb{E}\Big(X_n^{\underline{r-1}}\Big)\Big), \quad \text{for } n \to \infty.$$

We will make frequent use of this fact.

- Since the cumulants and moments of each random variable X are polynomially related (see Section 2.2.1), the following holds: $(\mathbb{E}(X_n^r))_{n\in\mathbb{N}}$ converges to $\mathbb{E}(X^r)$ for all $r \in \mathbb{N}$ iff $(\kappa_r(X_n))_{n\in\mathbb{N}}$ converges to $\kappa_r(X)$ for all $r \in \mathbb{N}$. Moreover, the sequence of moments $(\mathbb{E}(X^r))_{r\in\mathbb{N}}$ uniquely determines the distribution of X iff the sequence of cumulants $(\kappa_r(X))_{r\in\mathbb{N}}$ does. Hence, instead of proving the convergence of all moments, one can always choose to show the convergence of all cumulants in order to apply Theorem 2.2.1. This is particularly useful for proving convergence to the normal distribution, which has very simple cumulants (cf. Appendix A.4).

Another useful theorem, which can often be applied in order to show that a sequence of discrete random variables is after standardization asymptotically normally distributed, is the following:

Theorem 2.2.3 (Hwang's quasi-powers theorem [FS09])**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of discrete random variables supported by $\mathbb{N}_0$, with probability generating functions $(p_n(v))_{n\in\mathbb{N}}$. Assume that, uniformly in a fixed complex neighbourhood of $v = 1$, for sequences $\beta_n, \kappa_n \to \infty$, there holds*

$$p_n(v) = A(v) \cdot B(v)^{\beta_n} \left(1 + \mathcal{O}\left(\frac{1}{\kappa_n}\right)\right),$$

where $A(v)$, $B(v)$ are analytic at $v = 1$ and $A(1) = B(1) = 1$. Assume finally that $B(v)$ satisfies the so-called "variability condition",

$$B''(1) + B'(1) - B'(1)^2 \neq 0.$$

Under these conditions, the mean and variance of X_n satisfy

$$\mathbb{E}(X_n) = \beta_n B'(1) + A'(1) + \mathcal{O}\big(\kappa_n^{-1}\big),$$
$$\mathbb{V}(X_n) = \beta_n(B''(1) + B'(1) - B'(1)^2) + A''(1) + A'(1) - A'(1)^2 + \mathcal{O}\big(\kappa_n^{-1}\big).$$

Furthermore, X_n is after standardization asymptotically normally distributed, with speed of convergence $\mathcal{O}\Big(\kappa_n^{-1} + \beta_n^{-1/2}\Big)$, i.e.

$$\mathbb{P}\left\{\frac{X_n - \mathbb{E}(X_n)}{\sqrt{\mathbb{V}(X_n)}} \le x\right\} = \Phi(x) + \mathcal{O}\left(\frac{1}{\kappa_n} + \frac{1}{\sqrt{\beta_n}}\right),$$

where $\Phi(x)$ is the distribution function of the standard normal distribution.

We close this section with the following lemmata, which will be valuable when deriving some of our limiting distribution results:

Lemma 2.2.4 (Stirling's formula [AS64])**.** *For fixed $\epsilon > 0$, the Gamma function $\Gamma(z)$ satisfies, in the range $|\operatorname{Arg}(z)| < \frac{\pi}{2} - \epsilon$,*

$$\log\left(\Gamma(z)\right) = \left(z - \frac{1}{2}\right)\log z - z + \frac{\log 2\pi}{2} + \mathcal{O}\left(\frac{1}{|z|}\right), \quad \text{for } |z| \to \infty.$$

From this, one gets the following useful lemma:

Lemma 2.2.5. *For fixed α and β, there holds*

$$\frac{\Gamma(n+\alpha)}{\Gamma(n+\beta)} = n^{\alpha-\beta}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right), \quad \text{for } n \to \infty \text{ in } \mathbb{N}.$$

Moreover, if only β is fixed, the $\mathcal{O}$-bound is uniform for bounded α.

CHAPTER 3

Inversions in labelled tree families

3.1 Introduction

In this chapter we consider *rooted trees* T in which the nodes are *labelled* with distinct integers of $\{1, \ldots, |T|\}$, where $|T|$ is the size (i.e. the number of nodes) of T. An *inversion* in a tree T is a pair (i, j) of nodes (we will always identify each node with its label), such that $i > j$ and i lies on the unique path from the root node $\operatorname{root}(T)$ of T to j (see Figure 3.1 for an example). Let us denote by $\operatorname{inv}(T)$ the number of inversions in T.

In [GYS99, MR68] studies concerning the number of inversions in some important combinatorial tree families $\mathcal{T}$ have been given by introducing so-called *tree inversion polynomials*. They shall be defined as follows:

$$J_n(q) := \sum_{T \in \mathcal{T} : |T| = n} q^{\operatorname{inv}(T)}.$$

Actually, unlike in our studies, in [GYS99, MR68] the authors exclusively considered trees with the root node labelled 1. Thus, in order to avoid confusion, we introduce also the slightly modified polynomials

$$\hat{J}_n(q) := \sum_{\substack{T \in \mathcal{T}: \\ |T| = n \text{ and } \operatorname{root}(T) = 1}} q^{\operatorname{inv}(T)}.$$

For *unordered trees* (cf. Section 3.2.1), Mallows and Riordan [MR68] could give an explicit formula for a suitable generating function of the corresponding tree

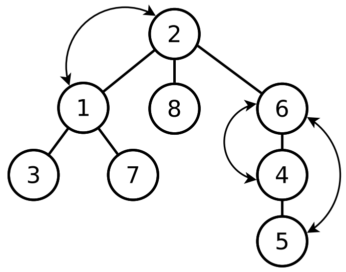

Figure 3.1: A tree with 3 inversions.

inversion polynomials:

$$\exp\left(\sum_{n\geq 1}(q-1)^{n-1}\hat{J}_n(q)\frac{t^n}{n!}\right) = \sum_{n\geq 0} q^{\binom{n}{2}}\frac{t^n}{n!}.$$

Gessel et al. [GYS99] considered $\hat{J}_n(q)$ for three other tree families:

- *Ordered trees* (we will formally define this tree family later, cf. Section 3.2.1)
- *Cyclic trees* (also, cf. Section 3.2.1)
- *Plane trees:* these are equivalence classes of ordered trees, where rearrangements of the subtrees of the root node lead to a tree of the same class.

Unlike for unordered trees, no explicit formulæ for a suitable generating function of the tree inversion polynomial of the latter tree families could be given, but the authors provide exact and asymptotic results for the evaluations of $\hat{J}_n(q)$ for the special values $q = 0, 1, -1$. In particular, $\hat{J}_n(0)$ enumerates so-called *increasing trees* (cf., e.g., [BFS92, PP07]), i.e. trees, where each child node has a label larger than its parent node.

Besides these studies it seems natural to ask, for a given combinatorial family $\mathcal{T}$ of trees, questions about the "typical behaviour" of the number of inversions in a tree $T \in \mathcal{T}$ of size n. In a probabilistic setting we may introduce a random variable I_n, which counts the number of inversions of a *random tree* of size n, i.e. a tree chosen uniformly at random from all trees of the family $\mathcal{T}$ of size n. Of course, this more probabilistic point of view and the before-mentioned combinatorial approach are closely related. Let us denote by T_n the number of trees of $\mathcal{T}$ of size n. Then there holds

$$J_n(q) = T_n \sum_{k\geq 0} \mathbb{P}\{I_n = k\}\, q^k,$$

i.e. the probability generating function $p_n(q) := \sum_{k\geq 0} \mathbb{P}\{I_n = k\} q^k$ of the random variable I_n is simply given by $p_n(q) = \frac{J_n(q)}{T_n}$, and it holds $T_{n,k} = T_n \mathbb{P}\{I_n = k\}$ for the number $T_{n,k} := [q^k] J_n(q)$ of trees of size n with exactly k inversions.

The main concern of this chapter is to describe the asymptotic behaviour of the random variable I_n for various important tree families by proving limiting distribution results. In our studies of I_n we use as tree-models *simply generated tree families* (see Section 3.2), which are tree families in which each tree T has a *weight* $w(T)$, and when speaking about a random tree of size n we assume that each tree T in $\mathcal{T}$ of size n is chosen with a probability proportional to its weight. For many important tree families $\mathcal{T}$ (such as binary trees, ordered trees, unordered trees, cyclic trees, cf. Section 3.2.1) the total weight

$$T_n := \sum_{\substack{T \in \mathcal{T}: \\ |T| = n}} w(T),$$

of all trees of size n is a natural number and can be intepreted as the number of trees of size n in $\mathcal{T}$. In these cases, choosing each (weighted) tree with a probability proportional to its weight corresponds to choosing each of the T_n trees of $\mathcal{T}$ of size n with the same probability $\frac{1}{T_n}$.

As a main result we can show, provided the degree-weight sequence satisfies certain mild growth conditions (which are all satisfied for the before-mentioned tree families), that, after a suitable normalization of order $n^{\frac{3}{2}}$, I_n converges in distribution to the Airy distribution (see Appendix A.6 for some details on this distribution). For the particular tree family of unordered trees this limiting distribution result for I_n has already been shown by Flajolet et al. in [FPV98] during their analysis of a linear probing hashing algorithm by using close relations between the insertion costs of this algorithm and the number of inversions in unordered trees. We note that we show convergence in distribution, thus obtaining asymptotic results for $\mathbb{P}\{I_n \leq x n^{\frac{3}{2}}\}$ or alternatively for the sums $\sum_{k \leq x n^{\frac{3}{2}}} T_{n,k}$, with $x \in \mathbb{R}^+$, but we do not obtain local limit laws, i.e. results concerning the behaviour of the probabilities $\mathbb{P}\{I_n = k\}$ or the numbers $T_{n,k}$ itself.

Besides this "global study" of the number of inversions in a random tree we are additionally interested in the contribution to this quantity induced by a specific label j, i.e. in a "local study". To do this we introduce random variables $I_{n,j}$, which count the number of inversions of the kind (i, j), with $i > j$ an ancestor of j, in a random tree of size n. Of course, one could also introduce "local inversion polynomials" $J_{n,j}(q) := T_n \sum_{k\geq 0} \mathbb{P}\{I_{n,j} = k\} q^k$. Note that $I_n = \sum_{j=1}^{n} I_{n,j}$, but the random variables $I_{n,j}$ are highly dependent. In our studies we describe the asymptotic behaviour of the random variable $I_{n,j}$, depending on the growth of $j = j(n)$ with respect to n. In particular, we obtain that for the main portion of labels, i.e. for $j \ll n - \sqrt{n}$, $I_{n,j}$ converges, after suitable normalization of order $\sqrt{n}$, in distribution to a Rayleigh distribution (cf. Section A.5). If $n - j \sim \rho\sqrt{n}$ or $n - j = o(\sqrt{n})$ then the behaviour changes. Apart from asymptotic results, we can for two tree families, namely ordered and unordered trees, also give explicit formulæ for the probabilities $\mathbb{P}\{I_{n,j} = k\}$.

We remark that the asymptotic results obtained for inversions in trees are completely different from the corresponding ones for permutations of a set $\{1, 2, \ldots, n\}$. It is well-known (see, e.g., [LP03]) that the total number of inversions in a random permutation of size n is asymptotically normally distributed with expectation and variance of order n^2 and n^3, respectively. Trivially, the number of inversions of the kind (i, j), with $i > j$ an element to the left of j, in a random permutation of size n is uniformly distributed on $\{0, 1, \ldots, n - j\}$.

3.2 Simply generated trees

Families of simply generated trees were introduced by Meir and Moon in [MM78]. Many important combinatorial tree families such as, e.g., labelled unordered trees (also called Cayley trees), binary trees, labelled cyclic trees (also called mobile trees) and ordered trees (also called planted plane trees), can be considered as special instances of simply generated trees. In the following, we recall how simply generated tree families are defined in the labelled context, and then collect some well-known auxiliary results. In the following, the term “tree” always denotes a labelled tree if not otherwise stated.

A class $\mathcal{T}$ of (labelled) simply generated trees is defined in the following way: One chooses a sequence $(\varphi_\ell)_{\ell\geq 0}$ (the so-called *degree-weight sequence*) of nonnegative real numbers with $\varphi_0 > 0$. Using this sequence, the *weight* $w(T)$ of each ordered tree (i.e. each rooted tree, in which the children of each node are ordered from left to right) is defined by

$$w(T) := \prod_{v\in T} \varphi_{d(v)},$$

where $v \in T$ means that v is a node of T and $d(v)$ denotes the number of children of v (i.e. the out-degree of v). The family $\mathcal{T}$ associated to the degree-weight sequence $(\varphi_\ell)_{\ell\geq 0}$ then consists of all trees T together with their weights.

We let

$$T_n := \sum_{|T|=n} w(T)$$

denote the the *total weight* of all trees of size n in $\mathcal{T}$, and let $T(z)$ be the exponential generating function of $(T_n)_{n\geq 1}$, i.e.

$$T(z) := \sum_{n\geq 1} T_n \frac{z^n}{n!}.$$

Then it follows that $T(z)$ satisfies the (formal) functional equation

$$T(z) = z\varphi(T(z)), \tag{3.1}$$

where the degree-weight generating function $\varphi(t)$ is defined via $\varphi(t) := \sum_{\ell\geq 0} \varphi_\ell t^\ell$. We remark that each simply generated tree family $\mathcal{T}$ can also be defined by a formal equation of the form

$$\mathcal{T} = \bigcirc * \varphi(\mathcal{T}), \tag{3.2}$$

where $\bigcirc$ denotes a node, $*$ is the combinatorial product of labelled objects, and $\varphi(\mathcal{T})$ is a certain substituted structure. Hence, the functional equation (3.1) can be obtained directly from the combinatorial construction of $\mathcal{T}$ using the symbolic method (cf. Section 2.1.1). Furthermore, as already noted in the introduction of this chapter, $(T_n)_{n\geq 1}$ is for many important simply generated tree families a sequence of natural numbers, and then the total weight T_n can be interpreted as the number of trees of size n in $\mathcal{T}$. We now give several examples where this is the case.

3.2.1 Examples

- Binary trees can be defined combinatorially as follows:

$$\mathcal{T} = \bigcirc * (\{\square\} \,\dot\cup\, \mathcal{T}) * (\{\square\} \,\dot\cup\, \mathcal{T}).$$

  Here, $\square$ denotes an empty subtree and $\dot\cup$ is the disjoint union. This formal equation expresses that each binary tree consists of a root node and a left and a right subtree, each of which is either a binary tree or empty. The formal equation for $\mathcal{T}$ can directly be translated into a functional equation for $T(z)$, namely

$$T(z) = z(1 + T(z))^2.$$

Hence, binary trees are the simply generated tree family defined by $\varphi(t) = (1+t)^2$, i.e. by the degree-weight sequence $\varphi_\ell = \binom{2}{\ell}$, $\ell \geq 0$. More generally, d-ary trees ($d \geq 2$), which can be described combinatorially as

$$\mathcal{T} = \bigcirc * \underbrace{(\{\Box\} \dot\cup \mathcal{T}) * \cdots * (\{\Box\} \dot\cup \mathcal{T})}_{d \text{ times}},$$

are the simply generated tree family defined by $\varphi(t) = (1+t)^d$, i.e. by the degree-weight sequence $\varphi_\ell = \binom{d}{\ell}$, $\ell \geq 0$.

- Ordered trees are rooted trees, in which the children of each node are linearly ordered. Thus, combinatorially speaking, each ordered tree consists of a root node and a sequence of ordered trees,

$$\mathcal{T} = \bigcirc * \textsc{Seq}(\mathcal{T}) = \bigcirc * \big(\{\Box\} \dot\cup \mathcal{T} \dot\cup \mathcal{T}^2 \dot\cup \mathcal{T}^3 \dot\cup \dots\big).$$

From this one gets the functional equation

$$T(z) = \frac{z}{1 - T(z)}, \tag{3.3}$$

i.e. $\varphi(t) = \frac{1}{1-t}$. Of course, this corresponds to the degree-weight sequence $\varphi_\ell = 1$, $\ell \geq 0$.

- Unordered trees are rooted trees in which there is no order on the children of any node. Hence, each unordered tree consists of a root node and a set of unordered trees, which can be written formally as

$$\mathcal{T} = \bigcirc * \textsc{Set}(\mathcal{T}) = \bigcirc * \left(\{\Box\} \dot\cup \mathcal{T} \dot\cup \frac{\mathcal{T}^2}{2!} \dot\cup \frac{\mathcal{T}^3}{3!} \dot\cup \dots\right).$$

This leads to the functional equation

$$T(z) = z \exp(T(z)),$$

i.e. one has $\varphi(t) = \exp(t)$, or equivalently $\varphi_\ell = 1/\ell!$, $\ell \geq 0$.

- Cyclic trees may be considered as equivalence classes of ordered trees, where cyclic rearrangements of the subtrees of nodes lead to a tree of the same class. Hence, each cyclic tree is either a single root node or it consists of a root node and a (non-empty) cycle of unordered trees, which can be written formally as

$$\mathcal{T} = \bigcirc \dot\cup \bigcirc * \textsc{Cyc}(\mathcal{T}) = \bigcirc * \big(\{\Box\} \dot\cup \textsc{Cyc}(\mathcal{T})\big).$$

This leads to the functional equation

$$T(z) = z\left(1 + \log\left(\frac{1}{1 - T(z)}\right)\right),$$

i.e. one has $\varphi(t) = 1 + \log\left(\frac{1}{1-t}\right)$, or equivalently $\varphi_0 = 1$ and $\varphi_\ell = 1/\ell$, $\ell \geq 1$.

3.2.2 Auxiliary results

We now collect some known results (see, e.g., [FS09, Pan04a]) on the function $T(z)$ satisfying (3.1). First note that in general $T(z)$ and $\varphi(t)$ must be regarded as formal power series, because they do not need to have a positive radius of convergence, and then (3.1) must be understood as a formal equation. Thus, in order to analyze properties of simply generated tree families using analytic methods, we will need to make certain assumptions on φ. In particular, we will assume that $\varphi(t)$ has a positive radius of convergence R, and that there exists a minimal positive solution $\tau < R$ of the equation

$$t\varphi'(t) = \varphi(t). \tag{3.4}$$

If we define

$$d := \gcd\{k : \varphi_k > 0\}, \tag{3.5}$$

then it follows that (3.4) has exactly d solutions of smallest modulus, which are given by $\tau_j = \omega^j \tau$, for $0 \leq j \leq d-1$, where $\omega = \exp\left(\frac{2\pi \mathrm{i}}{d}\right)$. From the implicit function theorem it follows that the equation $z = \frac{t}{\varphi(t)}$ is not invertible in any neighbourhood of $t = \tau_j$, for $0 \leq j \leq d-1$. This leads to d dominant singularities of $T(z)$ at $z = \rho_j$, where $\rho_j = \omega^j \rho$, $\rho = \frac{\tau}{\varphi(\tau)}$.

For our purpose, it is important to note that under the above assumptions, $T(z)$ is amenable to singularity analysis (cf. Section 2.1.4), i.e. there are constants $\eta > \rho$ and $0 < \phi < \pi/2$ such that $T(z)$ is analytic in the domain

$$\{z \in \mathbb{C} : |z| < \eta, z \neq \rho_j, |\operatorname{Arg}(z - \rho_j)| > \phi, \text{ for all } 0 \leq j \leq d-1\}.$$

The local expansion of $T(z)$ around the singularity $z = \rho_j$ is given by

$$T(z) = \tau_j - \omega^j \sqrt{\frac{2\varphi(\tau)}{\varphi''(\tau)}} \sqrt{1 - \frac{z}{\rho_j}} + \mathcal{O}(\rho_j - z). \tag{3.6}$$

Using singularity analysis and summing up the contributions of the d dominant singularities, one obtains

$$\frac{T_n}{n!} = [z^n]T(z) = \frac{d\sqrt{\varphi(\tau)}}{\sqrt{2\pi\varphi''(\tau)}\rho^n n^{\frac{3}{2}}}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right), \tag{3.7}$$

for $n \equiv 1 \mod d$. If $n \not\equiv 1 \mod d$, one has of course $T_n = 0$, because in this case each plane tree of size n has weight zero.

In our analysis, we will further make use of the functions $\varphi^{(k)}(T(z))$ (where $\varphi^{(k)}(t)$ is the k-th derivative of $\varphi(t)$). Each of these functions has d dominant singularities at $z = \rho_j$, $0 \le j \le d-1$, and complies with the requirements for singularity analysis. Around $z = \rho_j$, one has the expansion

$$\varphi^{(k)}(T(z)) = \varphi^{(k)}(\tau_j) - \varphi^{(k+1)}(\tau_j)\omega^j\sqrt{\frac{2\tau}{\rho\varphi''(\tau)}}\sqrt{1 - \frac{z}{\rho_j}} + \mathcal{O}(\rho_j - z)\,, \tag{3.8}$$

and we will especially make use of the expansion

$$z\varphi'(T(z)) = 1 - \sqrt{2\rho\tau\varphi''(\tau)}\sqrt{1 - \frac{z}{\rho_j}} + \mathcal{O}(\rho_j - z)\,. \tag{3.9}$$

3.3 Results

Let $\mathcal{T}$ be the labelled family of simply generated trees associated to a degree-weight sequence $(\varphi_\ell)_{\ell\ge0}$, where the function $\varphi(t) := \sum_{\ell\ge0}\varphi_\ell t^\ell$ has positive radius of convergence R, and equation (3.4) has a minimal positive solution $\tau < R$. Furthermore, let $\rho = \frac{\tau}{\varphi(\tau)}$ (recall the definitions of Section 3.2.2). Then the following holds:

Theorem 3.3.1 (Global behaviour)**.** *The random variable I_n, which counts the total number of inversions in a random tree of size n of $\mathcal{T}$ is, after proper normalization, asymptotically Airy distributed:*
There holds $\mathbb{E}(I_n) \sim c_\varphi\sqrt{\pi}n^{3/2}$, where $c_\varphi = \frac{1}{\sqrt{8\rho\tau\varphi''(\tau)}}$, and

$$\frac{I_n}{c_\varphi n^{3/2}} \xrightarrow{(d)} I,$$

where I is an Airy distributed random variable.

Theorem 3.3.2 (Local behaviour)**.** *The random variable $I_{n,j}$, which counts the number of inversions of the kind (i,j), with $i > j$ an ancestor of j, in a random tree of size n of $\mathcal{T}$ has, depending on the growth of $1 \le j = j(n) \le n$, the following asymptotic behaviour.*

- *Region $n - j \gg \sqrt{n}$: $I_{n,j}$ is, after proper normalization, asymptotically Rayleigh distributed:*
$$\frac{\sqrt{n}}{n-j} I_{n,j} \xrightarrow{(d)} X_\sigma,$$
where X_σ is a Rayleigh distributed random variable with parameter $\sigma := \frac{1}{\sqrt{\rho\tau\varphi''(\tau)}}$.

- *Region $n-j \sim \alpha\sqrt{n}$, with $\alpha \in \mathbb{R}^+$: $I_{n,j}$ converges in distribution to a discrete random variable Y_γ, with*
$$\mathbb{P}\{Y_\gamma = k\} = \frac{\gamma^k}{k!} \int_0^\infty x^{k+1} e^{-\frac{x^2}{2} - \gamma x} dx, \quad k \in \mathbb{N}_0,$$
where $\gamma := \frac{\alpha}{\sqrt{\rho\tau\varphi''(\tau)}}$.

- *Region $n - j \ll \sqrt{n}$: $I_{n,j}$ converges in distribution to a random variable with all its mass concentrated at 0, i.e. $I_{n,j} \xrightarrow{(d)} 0$.*

Furthermore, for ordered and unordered trees the exact distribution of $I_{n,j}$ can be stated explicitly. It holds that
$$\mathbb{P}\{I_{n,j} = k\} = \frac{(j-1)!(n-j)!}{n^{n-1}} \sum_{\ell=n-j-k}^{n-k-1} \binom{\ell}{n-j-k} \binom{n-\ell-1}{k} \frac{(n-\ell)n^{\ell-1}}{\ell!}, \tag{3.10}$$
for unordered trees, and
$$\mathbb{P}\{I_{n,j} = k\} = \frac{1}{\binom{n-1}{j-1}\binom{2(n-1)}{n-1}} \sum_{\ell=n-j-k}^{n-k-1} \binom{\ell}{n-j-k} \binom{2n-2}{\ell} \binom{n-\ell-1}{k} \frac{2n-1-2\ell}{2n-1-\ell}, \tag{3.11}$$
for ordered trees ($1 \le j \le n$ and $0 \le k \le n-j$).

3.3.1 Examples

Before we prove these results, we apply them to our example tree families from Section 3.2.1:

- Binary trees: From the equation $2t(t+1) = t\varphi'(t) = \varphi(t) = (t+1)^2$ we get the positive solution $\tau = 1$, and hence $c_\varphi = \frac{1}{2}$ and $\sigma = \sqrt{2}$. Thus, if we let I_n denote the number of inversions in a random binary tree of size n, then $\frac{2I_n}{n^{3/2}}$ converges in distribution to an Airy distributed random variable. Furthermore, for the number $I_{n,j}$ of inversions in a random binary tree of size n induced by node j, it holds that $\frac{\sqrt{n}}{n-j} I_{n,j}$ converges, for $n - j \gg \sqrt{n}$, in distribution to a Rayleigh distributed random variable with parameter $\sqrt{2}$.

- Ordered trees: The equation $\frac{t}{(1-t)^2} = \frac{1}{1-t}$ yields $\tau = \frac{1}{2}$, and further $c_\varphi = \frac{1}{4}$ and $\sigma = \frac{1}{\sqrt{2}}$. Hence, for the number I_n of inversions in a random ordered tree of size n, it holds that $\frac{4I_n}{n^{3/2}}$ is asymptotically Airy distributed. Furthermore, the normalized number of inversions $\frac{\sqrt{n}}{n-j} I_{n,j}$ induced by node j, is, for $n-j \gg \sqrt{n}$, asymptotically Rayleigh distributed with parameter $1/\sqrt{2}$.

- Unordered trees: Here, one has $\tau = 1$ and thus $c_\varphi = \frac{1}{\sqrt{8}}$ and $\sigma = 1$. This shows that $\frac{\sqrt{8} I_n}{n^{3/2}}$ converges in distribution to an Airy distributed random variable and that $\frac{\sqrt{n}}{n-j} I_{n,j}$ converges, for $n - j \gg \sqrt{n}$, in distribution to a Rayleigh distributed random variable with parameter 1.

- Cyclic trees: The positive real solution of the equation $\frac{t}{1-t} = 1 + \log\left(\frac{1}{1-t}\right)$ is numerically given by $\tau \approx 0.682155$. One further gets $c_\varphi = \frac{\sqrt{1-\tau}}{\sqrt{8}} \approx 0.199325$ and $\sigma = \sqrt{1-\tau} \approx 0.563776$. Thus, $\frac{I_n}{c_\varphi n^{3/2}}$ converges in distribution to an Airy distributed random variable, and $\frac{\sqrt{n}}{n-j} I_{n,j}$ converges, for $n - j \gg \sqrt{n}$, in distribution to a Rayleigh distributed random variable with parameter σ.

3.4 Proof of the result concerning the global behaviour

3.4.1 Short overview of the proof

We prove our result given in Theorem 3.3.1 by using the method of moments, i.e. we show that the moments of I_n converge (after proper normalization) to the moments of the Airy distribution. Since this distribution is uniquely determined by its moments, the convergence result then follows directly from the Theorem of Fréchet and Shohat (Theorem 2.2.1). To start with, we do not study the random variable I_n directly, but consider a closely related random variable $\hat{I}_n$. Using the same tree decomposition which was used in [GYS99], we then obtain a q-difference-differential equation for a suitably chosen generating function which encodes the distribution of $\hat{I}_n$. From this equation, we can "pump" the moments of $\hat{I}_n$ using techniques from [FPV98] and singularity analysis, and finally transfer our result to I_n. Note that we will in the following make frequent use of the operators acting on generating functions which we defined in Section 2.1.2.

3.4.2 Introduction of $\hat{I}_n$ and generating functions

We let $\hat{\mathcal{T}}$ be the subset of $\mathcal{T}$ which consists exactly of those trees in which the root has label 1. Obviously, the total weight of trees of size n in $\hat{\mathcal{T}}$ is then given by $\frac{T_n}{n}$. Also note that each tree in $\hat{\mathcal{T}}$ has the nice property that the root is not part of any inversion. Hence, the total number of inversions of each tree can just be obtained by summing up the contributions of the individual subtrees of the root. This fact will later be useful when we translate a decomposition of the trees in $\hat{\mathcal{T}}$ to generating functions.

We let $\hat{I}_n$ denote the number of inversions in a random tree of size n in $\hat{\mathcal{T}}$, where each element of $\hat{\mathcal{T}}$ of size n is chosen with a probability proportional to its weight. Furthermore, we introduce the generating function

$$F(z,q) := \sum_{n\geq 1}\sum_{k\geq 0} \mathbb{P}\left\{\hat{I}_n = k\right\} \frac{T_n}{n} q^k \frac{z^n}{n!}. \tag{3.12}$$

Note that $n![z^n q^k]F(z,q) = \mathbb{P}\left\{\hat{I}_n = k\right\} \frac{T_n}{n}$ is the total weight of all trees of size n in $\hat{\mathcal{T}}$ which contain exactly k inversions. Moreover, observe that $\mathrm{U}F(z,q) = F(z,1)$ is just the exponential generating function of $\left(\frac{T_n}{n}\right)_{n\geq 1}$, and hence we have the relation

$$\mathrm{Z}\mathrm{D}_z\mathrm{U}F(z,q) = T(z), \tag{3.13}$$

which we will use frequently. We further introduce the functions

$$f_r(z) := \mathrm{U}\mathrm{D}_q^r F(z,q),$$

which are generating functions of the factorial moments $\mathbb{E}\Big((\hat{I}_n)^{\underline{r}}\Big)$ of $\hat{I}_n$, in the sense that

$$f_r(z) = \sum_{n\geq 1} \mathbb{E}\Big(\hat{I}_n^{\underline{r}}\Big) \frac{T_n}{n} \frac{z^n}{n!}. \tag{3.14}$$

Clearly, we can recover the r-th factorial moment of $\hat{I}_n$ from (3.14) by

$$\mathbb{E}\Big(\hat{I}_n^{\underline{r}}\Big) = \frac{[z^n] f_r(z)}{[z^n] f_0(z)},$$

but as we will see later, it is more convenient to use

$$\mathbb{E}\Big(\hat{I}_n^{\underline{r}}\Big) = \frac{[z^n] z f_r'(z)}{[z^n] z f_0'(z)} = \frac{[z^n] z f_r'(z)}{[z^n] T(z)}, \tag{3.15}$$

where the second equality follows from (3.13).

3.4.3 The q-difference-differential equation for $F(z,q)$

It turns out that $F(z,q)$ satisfies a certain equation involving a q-difference operator H which is very similar to the one Flajolet, Poblete and Viola used in [FPV98] in their analysis of linear probing hashing. In our case, we define H by

$$\mathrm{H}G(z,q) := \frac{G(z,q) - G(qz,q)}{1-q}.$$

Using this, we get:

Lemma 3.4.1. *The function $F(z,q)$ defined by* (3.12) *satisfies*

$$\mathrm{D}_z F(z,q) = \varphi(\mathrm{H}F(z,q)). \tag{3.16}$$

Proof. This equation can be obtained by establishing mutually dependent recurrences for the total weights $T_{n,k} := \mathbb{P}\{I_n = k\}\, T_n$ and $\hat{T}_{n,k} := \mathbb{P}\Big\{\hat{I}_n = k\Big\} \frac{T_n}{n}$ of trees of size n with k inversions in $\mathcal{T}$ and $\hat{\mathcal{T}}$, respectively. However, we choose to give a combinatorial argument.

In order to derive (3.16), we establish relations between $\mathcal{T}$ and $\hat{\mathcal{T}}$, which can be translated into functional equations for $F(z,q) = \sum_{n\geq 1}\sum_{k\geq 0} \hat{T}_{n,k} q^k \frac{z^n}{n!}$ and

$$T(z,q) := \sum_{n\geq 1}\sum_{k\geq 0} T_{n,k} q^k \frac{z^n}{n!}.$$

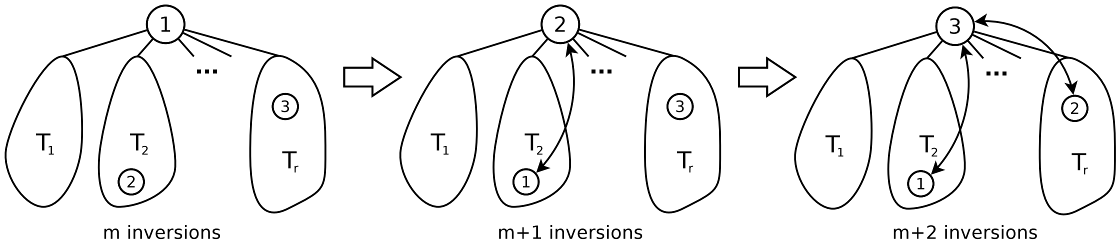

Figure 3.2: Adding inversions by exchanging labels.

For this purpose, we consider the sets $\mathcal{T}_n$ and $\hat{\mathcal{T}}_n$, which contain exactly the trees of size n of $\mathcal{T}$ and $\hat{\mathcal{T}}$, respectively. Clearly, $\mathcal{T}_n$ can be partitioned into n disjoint subsets $\mathcal{T}_n^{(1)}, \mathcal{T}_n^{(2)} \ldots, \mathcal{T}_n^{(n)}$, where $\mathcal{T}_n^{(j)}$ contains exactly those trees in which the root is labelled by j (in particular, $\mathcal{T}_n^{(1)} = \hat{\mathcal{T}}_n$). Now consider the bijective mapping between $\hat{\mathcal{T}}_n$ and $\mathcal{T}_n^{(2)}$ which is obtained by just swapping the labels 1 and 2 in each tree, and leaving all other labels and the structure of each tree unchanged. Since this mapping does not alter the relative order of any pair of nodes except $(1, 2)$, it clearly holds that each tree of $\hat{\mathcal{T}}_n$ with m inversions is mapped to a tree of $\mathcal{T}_n^{(2)}$ with $m + 1$ inversions. Repeating this argument (see Figure 3.2), we see that the trees in $\hat{\mathcal{T}}_n$ with m inversions can bijectively be mapped to the trees with $m+j-1$ inversions in $\mathcal{T}_n^{(j)}$.

This leads for the generating functions $F(z, q)$ and $T(z, q)$ to the equation

$$T(z,q) = \sum_{n\geq 1}\sum_{k\geq 0} \hat{T}_{n,k} \underbrace{(1+\ldots+q^{n-1})}_{\frac{1-q^n}{1-q}} q^k \frac{z^n}{n!} = \mathrm{H}F(z,q). \tag{3.17}$$

Next, remember that $\mathcal{T}$ is defined by the formal equation $\mathcal{T} = \bigcirc * \varphi(\mathcal{T})$ (cf. Section 3.2), and that $\hat{\mathcal{T}}$ consists exactly of those trees of $\mathcal{T}$ in which the root has label 1. It thus follows that $\hat{\mathcal{T}}$ satisfies the formal equation

$$\hat{\mathcal{T}} = ①^{\Box} * \varphi(\mathcal{T}), \tag{3.18}$$

which involves a boxed product (cf. Section 2.1.1). Due to the observation that the root node ① of any tree in $\hat{\mathcal{T}}$ does not contribute to the number of inversions, equation (3.18) can be translated by an application of the symbolic method to the differential equation

$$\mathrm{D}_z F(z,q) = \varphi(T(z,q)).$$

Using equation (3.17), we thus obtain (3.16). $\square$

3.4.4 Application of the pumping method

In order to extract expressions for the functions $f_r(z)$ as defined in (3.14) from (3.16), we use the pumping method from [FPV98]. This method basically rests on the idea of applying the operator UD_q^r to the given functional equation involving H, and using a "commutation rule" for the operators UD_q^r and H. Since our operator H is slightly different from the one in [FPV98], we will first establish the suitable commutation rule for our case.

Lemma 3.4.2. *The operator* $\mathrm{UD}_q^j\mathrm{H}$ *satisfies the operator equation*

$$\mathrm{UD}_q^j\mathrm{H} = \sum_{s=0}^{j}\binom{j}{s}\frac{1}{s+1}\mathrm{Z}^{s+1}\mathrm{D}_z^{s+1}\mathrm{UD}_q^{j-s}. \tag{3.19}$$

Proof. Since all occurring operators are linear, it suffices to show that the two sides of the equation coincide when applied to a function of the form $G(z,q) = q^k z^n$. Remember that $\mathrm{H}(q^k z^n) = q^k(1+q+\cdots+q^{n-1})z^n$, and thus we have

$$\begin{aligned}
\mathrm{UD}_q^j\mathrm{H}(q^k z^n) &= z^n\sum_{i=0}^{n-1}\mathrm{UD}_q^j(q^k q^i) = z^n\sum_{i=0}^{n-1}\sum_{s=0}^{j}\binom{j}{s}k^{\underline{j-s}}i^{\underline{s}} \\
&= z^n\sum_{s=0}^{j}\binom{j}{s}k^{\underline{j-s}}s!\sum_{i=0}^{n-1}\binom{i}{s} = z^n\sum_{s=0}^{j}\binom{j}{s}k^{\underline{j-s}}s!\binom{n}{s+1} \\
&= \sum_{s=0}^{j}\binom{j}{s}k^{\underline{j-s}}\frac{n^{\underline{s+1}}z^n}{s+1} = \sum_{s=0}^{j}\binom{j}{s}\frac{1}{s+1}\mathrm{Z}^{s+1}\mathrm{D}_z^{s+1}\mathrm{UD}_q^{j-s}(q^k z^n).
\end{aligned}$$

□

Using (3.19), we can now establish a recurrence for the derivatives $f_r'(z)$ of the factorial moment generating functions:

Lemma 3.4.3. *The factorial moment generating functions* $f_r(z)$ *satisfy, for* $r \geq 1$,

$$\begin{aligned}
f_r'(z) = \frac{1}{1-z\varphi'(T(z))}\Bigg(&\varphi'(T(z))\sum_{t=1}^{r}\binom{r}{t}\frac{1}{t+1}z^{t+1}f_{r-t}^{(t+1)}(z) \\
&+\sum_{(k_1,\ldots,k_{r-1})\in B_r}\Bigg[\frac{r!}{k_1!\cdots k_{r-1}!}\varphi^{(k_1+\ldots+k_{r-1})}(T(z)) \\
&\cdot\prod_{m=1}^{r-1}\left(\frac{1}{m!}\sum_{s=0}^{m}\binom{m}{s}\frac{1}{s+1}z^{s+1}f_{m-s}^{(s+1)}(z)\right)^{k_m}\Bigg]\Bigg),
\end{aligned} \tag{3.20}$$

where $B_1 := \emptyset$*, and*

$$B_r := \left\{(k_1, k_2, \ldots, k_{r-1}) \in \mathbb{N}_0^{r-1} : k_1 + 2k_2 + \ldots + (r-1)k_{r-1} = r\right\},$$

for $r \geq 2$*.*

Proof. We apply UD_q^r to (3.16) and express $\mathrm{D}_q^r \varphi(\mathrm{H}F(z,q))$ using Faá di Bruno's formula for higher derivatives of composite functions (cf., e.g., [Rio58]),

$$\mathrm{D}_q^r f(g(q)) = \sum_{(k_1,\ldots,k_r)\in A_r} \frac{r!}{k_1! \cdots k_r!} (\mathrm{D}_q^{(k_1+\ldots+k_r)} f)(g(q)) \prod_{m=1}^{r} \left(\frac{\mathrm{D}_q^m g(q)}{m!}\right)^{k_m},$$

where $A_r := \{(k_1, k_2, \ldots, k_r) \in \mathbb{N}_0^r : k_1 + 2k_2 + \ldots + rk_r = r\}$. We then obtain the claimed result by applying Lemma 3.4.2, solving for $f_r'(z)$ and using the fact that

$$\mathrm{U}\varphi(\mathrm{H}F(z,q)) = \varphi(\mathrm{UH}F(z,q)) = \varphi(\mathrm{ZD}_z\mathrm{U}F(z,q)) = \varphi(T(z)),$$

which follows from (3.19) and (3.13). $\square$

3.4.5 Application of singularity analysis

We now investigate the singular behaviour of the functions in (3.20) in order to compute the factorial moments $\mathbb{E}\left(\hat{I}_n^{\underline{r}}\right)$ asymptotically. In the following, we carry out only the computations for the case $d = 1$ (where d is defined by (3.5) and thus gives the number of dominant singularities of the functions considered). The general case runs completely analogous: when applying singularity analysis, one just has to take care of the contributions of all d singularities and add them.

In a first step, we want to find an asymptotic formula for the expected value $\mathbb{E}\left(\hat{I}_n\right)$. From Lemma 3.4.3, we have

$$f_1'(z) = \frac{\varphi'(T(z))\frac{z^2}{2} f_0''(z)}{1 - z\varphi'(T(z))},$$

and using the fact that $f_0''(z) = T'(z)\varphi'(T(z)) = \frac{\varphi(T(z))\varphi'(T(z))}{1 - z\varphi'(T(z))}$, which is easily obtained by differentiating (3.1) and (3.13), we get

$$f_1'(z) = \frac{1}{2} \frac{z(\varphi'(T(z)))^2 T(z)}{(1 - z\varphi'(T(z)))^2}.$$

Note that $z\varphi'(T(z)) \neq 1$ for $|z| \leq \rho$, which can be seen by differentiating (3.1), and thus $f_1'(z)$ inherits the dominant singularity at $z = \rho$ from $T(z)$ and $\varphi'(T(z))$. Using the expansions (3.6) and (3.9), we find

$$\begin{aligned} zf_1'(z) &= \frac{1}{2}\frac{\left(1+\mathcal{O}\left((\rho-z)^{1/2}\right)\right)^2 \tau \left(1+\mathcal{O}\left((\rho-z)^{1/2}\right)\right)}{2\rho\tau\varphi''(\tau)\left(1-\frac{z}{\rho}\right)(1+\mathcal{O}((\rho-z)^{1/2}))}, \\ &= \frac{1}{4\rho\varphi''(\tau)\left(1-\frac{z}{\rho}\right)}\left(1+\mathcal{O}\left((\rho-z)^{1/2}\right)\right), \quad z \to \rho. \end{aligned} \tag{3.21}$$

By applying basic singularity analysis, this immediately yields

$$[z^n]zf_1'(z) = \frac{1}{4\rho^{n+1}\varphi''(\tau)}\left(1+\mathcal{O}\left(n^{-1/2}\right)\right).$$

Now, using (3.15) and (3.7), we get the expected value

$$\begin{aligned} \mathbb{E}\left(\hat{I}_n\right) &= \frac{[z^n]zf_1'(z)}{[z^n]T(z)} = \frac{1}{\rho}\sqrt{\frac{\pi}{8\varphi(\tau)\varphi''(\tau)}}n^{3/2}\left(1+\mathcal{O}\left(n^{-1/2}\right)\right) \\ &= c_\varphi\sqrt{\pi}n^{3/2}\left(1+\mathcal{O}\left(n^{-1/2}\right)\right), \end{aligned}$$

where c_φ is defined as in Theorem 3.3.1.

We will now consider $f_r'(z)$ for general r. It turns out that all $f_r'(z)$ have a unique dominant singularity at $z = \rho$. The singular expansions around this point are given in the following lemma.

Lemma 3.4.4. *For $r \geq 1$, each $f_r'(z)$ has a unique dominant singularity at $z = \rho$, where the expansion*

$$zf_r'(z) = c_\varphi^r\sqrt{\frac{\varphi(\tau)}{2\varphi''(\tau)}}\frac{2C_r}{\left(1-\frac{z}{\rho}\right)^{(3r-1)/2}}\left(1+\mathcal{O}\left((\rho-z)^{1/2}\right)\right), \quad z \to \rho, \tag{3.22}$$

holds. Here, the constants C_r are the same which appear in the moments of the Airy distribution (cf. equation (A.4) *in Appendix A.6).*

Proof. One easily checks that in the case $r = 1$ equation (3.22) coincides with (3.21). For $r > 1$ we proceed by induction, following the inductive definition (A.4) of the constants C_r. So let $r > 1$ and assume that (3.22) holds for all functions $f_j(z)$ with $1 \leq j < r$. By the rules for singular differentiation (see Theorem 2.1.6) we then also have the following singular expansions for the k-th derivatives $f_j^{(k)}$ of the functions $f_j(z)$, for all $k \geq 1$ and $1 \leq j < r$:

$$zf_j^{(k)}(z) = c_\varphi^r \sqrt{\frac{\varphi(\tau)}{2\varphi''(\tau)}} \frac{2C_j \cdot \left(\frac{3j-1}{2}\right)^{\overline{k-1}}}{\rho^{k-1}\left(1-\frac{z}{\rho}\right)^{(3j-3+2k)/2}} \left(1+\mathcal{O}\big((\rho-z)^{1/2}\big)\right), \quad z \to \rho, \tag{3.23}$$

where $x^{\overline{k}}$ denotes the k-th rising factorial of x, i.e. $x^{\overline{0}} := 1$, and, for $k \geq 1$, $x^{\overline{k}} := x(x+1)\cdots(x+k-1)$. From this, one concludes that the dominant contributions in (3.20) can only arise from the terms corresponding to $t=1$ and $s=0$, i.e.

$$\begin{aligned} zf_r'(z) = \frac{z}{1-z\varphi'(T(z))} \Bigg(& \varphi'(T(z))\frac{r}{2}z^2 f_{r-1}''(z) \\ & + \sum_{(k_1,\ldots,k_{r-1})\in B_r} \frac{r!}{k_1!\cdots k_{r-1}!}\varphi^{(k_1+\ldots+k_{r-1})}(T(z)) \prod_{m=1}^{r-1}\left(\frac{zf_m'(z)}{m!}\right)^{k_m}\Bigg) \\ & \cdot \left(1+\mathcal{O}\big((\rho-z)^{1/2}\big)\right). \end{aligned}$$

Now note that

$$\prod_{m=1}^{r-1}\left(\frac{1}{m!}zf_m'(z)\right)^{k_m} = \mathcal{O}\left(\frac{1}{(\rho-z)^{(3r-(k_1+\ldots+k_{r-1}))/2}}\right),$$

which implies that the dominant terms in the remaining sum correspond to those $(k_1,\ldots,k_{r-1}) \in B_r$ with $k_1+\ldots+k_{r-1}=2$, and we thus get

$$\begin{aligned} zf_r'(z) = \frac{z}{1-z\varphi'(T(z))} \Bigg(& \varphi'(T(z))\frac{r}{2}z^2 f_{r-1}''(z) \\ & + \sum_{s=1}^{r-1}\frac{r!}{2}\varphi''(T(z))z^2\frac{f_s'(z)}{s!}\frac{f_{r-s}'(z)}{(r-s)!}\Bigg)\left(1+\mathcal{O}\big((\rho-z)^{1/2}\big)\right). \end{aligned}$$

Now, expanding the occurring functions using (3.8), (3.9) and (3.23), we obtain after some simplifications

$$
\begin{aligned}
zf_r'(z) = \frac{\rho}{\sqrt{2\rho\tau\varphi''(\tau)}\sqrt{1-\frac{z}{\rho}}} &\left(\frac{r}{2} c_\varphi^{r-1} \sqrt{\frac{\varphi(\tau)}{2\varphi''(\tau)}} \frac{2C_{r-1} \cdot \frac{3(r-1)-1}{2}}{\rho\left(1-\frac{z}{\rho}\right)^{(3(r-1)+1)/2}} \right. \\
&\left. + \frac{1}{2} \sum_{s=1}^{r-1} \binom{r}{s} \varphi''(\tau) c_\varphi^s c_\varphi^{r-s} \left(\frac{\varphi(\tau)}{2\varphi''(\tau)} \right) \frac{2C_s \cdot 2C_{r-s}}{\left(1-\frac{z}{\rho}\right)^{(3s-1)/2+(3(r-s)-1)/2}} \right) \\
&\cdot \left(1+\mathcal{O}\big((\rho-z)^{1/2}\big)\right) \\
= c_\varphi^r \sqrt{\frac{\varphi(\tau)}{2\varphi''(\tau)}} &\frac{(3r-4)rC_{r-1} + \sum_{s=1}^{r-1}\binom{r}{s}C_sC_{r-s}}{\left(1-\frac{z}{\rho}\right)^{(3r-1)/2}} \left(1+\mathcal{O}\big((\rho-z)^{1/2}\big)\right), \\
= c_\varphi^r \sqrt{\frac{\varphi(\tau)}{2\varphi''(\tau)}} &\frac{2C_r}{\left(1-\frac{z}{\rho}\right)^{(3r-1)/2}} \left(1+\mathcal{O}\big((\rho-z)^{1/2}\big)\right), \quad z \to \rho.
\end{aligned}
$$

□

Lemma 3.4.4 can now be used in order to compute the moments of $\hat{I}_n$ asymptotically:

Lemma 3.4.5. *The random variable $\hat{I}_n$ satisfies*

$$
\mathbb{E}\left(\hat{I}_n^r\right) = \frac{2\sqrt{\pi} c_\varphi^r n^{3r/2}}{\Gamma(\frac{3r-1}{2})} C_r \left(1+\mathcal{O}\big(n^{-1/2}\big)\right).
$$

Proof. By singularity analysis, it follows from Lemma 3.4.4 that

$$
[z^n] z f_r'(z) = 2 c_\varphi^r \sqrt{\frac{\varphi(\tau)}{2\varphi''(\tau)}} \frac{n^{\frac{3r-1}{2}-1}}{\rho^n \Gamma(\frac{3r-1}{2})} C_r \left(1+\mathcal{O}\big(n^{-1/2}\big)\right),
$$

and together with (3.15) and (3.7) this shows

$$
\mathbb{E}\left(\hat{I}_n^{\underline{r}}\right) = \frac{[z^n] z f_r'(z)}{[z^n]T(z)} = \frac{2\sqrt{\pi} c_\varphi^r n^{3r/2}}{\Gamma(\frac{3r-1}{2})} C_r \left(1+\mathcal{O}\big(n^{-1/2}\big)\right).
$$

Using the relation (2.6) between the factorial moments and the ordinary moments, we obtain that $\mathbb{E}\left(\hat{I}_n^r\right) = \mathbb{E}\left(\hat{I}_n^{\underline{r}}\right) + \mathcal{O}\left(\mathbb{E}\left(\hat{I}_n^{\underline{r-1}}\right)\right)$, and hence we get the desired result. □

3.4.6 Transfer of the result to I_n

We now transfer the result for $\hat{I}_n$ to the random variable I_n, which counts the total number of inversions in a random tree of size n of $\mathcal{T}$. In fact, we prove that the moments of $\hat{I}_n$ and I_n coincide asymptotically:

Lemma 3.4.6. *The random variable I_n satisfies*

$$\mathbb{E}(I_n^r) = \frac{2\sqrt{\pi} c_\varphi^r n^{3r/2}}{\Gamma(\frac{3r-1}{2})} C_r \left(1 + \mathcal{O}\big(n^{-1/2}\big)\right). \tag{3.24}$$

Proof. The relation between $\mathcal{T}$ and $\hat{\mathcal{T}}$ (compare equation (3.17)) directly translates to the following relation between the moments of I_n and $\hat{I}_n$:

$$\mathbb{E}(I_n^r) = \frac{1}{n}\left(\mathbb{E}\Big(\hat{I}_n^r\Big) + \mathbb{E}\Big((\hat{I}_n + 1)^r\Big) + \ldots + \mathbb{E}\Big((\hat{I}_n + n - 1)^r\Big)\right).$$

From this, one deduces

$$\begin{aligned}\mathbb{E}(I_n^r) &= \mathbb{E}\Big(\hat{I}_n^r\Big) + \frac{1}{n}\sum_{\ell=0}^{r-1}\binom{r}{\ell}\underbrace{(1^{r-\ell} + 2^{r-\ell} + \ldots + (n-1)^{r-\ell})}_{\mathcal{O}(n^{r-\ell+1})}\underbrace{\mathbb{E}\Big(\hat{I}_n^\ell\Big)}_{\mathcal{O}\left(n^{\frac{3\ell}{2}}\right)} \\ &= \mathbb{E}\Big(\hat{I}_n^r\Big) + \mathcal{O}\left(n^{\frac{3r-1}{2}}\right),\end{aligned}$$

and hence (3.24) follows directly from Lemma 3.4.5. □

By comparing (3.24) with μ_r in (A.3), we conclude that the moments of the normalized random variable $\frac{I_n}{c_\varphi n^{3/2}}$ converge to the moments of the Airy distribution. Due to Lemma A.6.1, the convergence in distribution of $\frac{I_n}{c_\varphi n^{3/2}}$ to an Airy distributed random variable thus follows directly from the Theorem of Fréchet and Shohat (Theorem 2.2.1).

This finishes the proof of our result on the total number of inversions in random trees.

3.5 Proofs of the results concerning the local behaviour

3.5.1 The generating functions approach

A main ingredient in the proof of Theorem 3.3.2 concerning the behaviour of the random variable $I_{n,j}$ is to introduce and study a suitable generating function for the probabilities $\mathbb{P}\{I_{n,j} = k\}$, which reflects in a simple way the recursive description of a tree as a root node and its subtrees. It turns out that the following trivariate generating function is appropriate:

$$N(z,u,q) := \sum_{m\ge 0}\sum_{j\ge 1}\sum_{k\ge 0} \mathbb{P}\{I_{m+j,j} = k\}\, T_{m+j} \frac{z^{j-1}}{(j-1)!}\frac{u^m}{m!} q^k. \tag{3.25}$$

Proposition 3.5.1. *The generating function $N(z,u,q)$ is given by the following explicit formula:*

$$N(z,u,q) = \frac{\varphi(T(z+u))}{1-(z+uq)\varphi'(T(z+u))}. \tag{3.26}$$

Proof. We will show the functional equation

$$N(z,u,q) = \varphi(T(z+u)) + z\varphi'(T(z+u))N(z,u,q) + uq\varphi'(T(z+u))N(z,u,q), \tag{3.27}$$

which is equivalent to (3.26). To do this we introduce specifically tricoloured trees: in each tree $T \in \mathcal{T}$ exactly one node is coloured red, all nodes with a label smaller than the red node are coloured white, whereas all nodes with a label larger than the red node are coloured black. Let us denote by $\mathcal{T}_C$ the family of all such tricoloured trees. Then in the generating function $N(z,u,q)$ the variable z encodes the white nodes, the variable u encodes the black nodes, whereas q encodes the black ancestors of the red node, i.e.

$$N(z,u,q) = \sum_{T_C \in \mathcal{T}_C} w(T_C) \frac{z^{\sharp\text{ white}}}{(\sharp\text{ white})!}\frac{u^{\sharp\text{ black}}}{(\sharp\text{ black})!} q^{\sharp\text{ black ancestors of red}}.$$

Since the black nodes as well as the white nodes are labelled it is appropriate to use a double exponential generating function.

As auxiliary family we consider specifically bicoloured trees: the nodes in each tree $T \in \mathcal{T}$ are coloured black and white in a way such that each white node has a label smaller than any black node (i.e. all nodes up to a certain label are coloured white, whereas all remaining nodes are coloured black). Let us denote by $\mathcal{T}_B$ the set of all such bicoloured trees. The double exponential generating function of bicoloured trees,

$$B(z,u) = \sum_{T_B \in \mathcal{T}_B} w(T_B) \frac{z^{\sharp\text{ white}}}{(\sharp\text{ white})!}\frac{u^{\sharp\text{ black}}}{(\sharp\text{ black})!},$$

can be computed easily. It holds:

$$\begin{aligned}
B(z,u) &= \sum_{n\geq 1(0)} \sum_{T\in\mathcal{T}:|T|=n} w(T) \sum_{m=0}^{n} \frac{z^{n-m}}{(n-m)!}\frac{u^m}{m!} \\
&= \sum_{n\geq 0}\sum_{m=0}^{n} \frac{z^{n-m}}{(n-m)!}\frac{u^m}{m!} \sum_{T\in\mathcal{T}:|T|=n} w(T) \\
&= \sum_{n\geq 0} T_n \sum_{m=0}^{n} \frac{z^{n-m}}{(n-m)!}\frac{u^m}{m!} = \sum_{n\geq 0(1)} \frac{T_n}{n!}(z+u)^n = T(z+u).
\end{aligned} \tag{3.28}$$

Now we consider the decomposition of a tricoloured tree $T_C \in \mathcal{T}_C$ into the root node $\mathrm{root}(T_C)$ and its $\ell \geq 0$ subtrees $T_1, \dots, T_\ell$. Note that then the degree-weight of the root node is given by φ_ℓ. Three cases may occur.

- Case 1: The root node is the red node. Then the red node does not have any black ancestors and all of the subtrees $T_1, \dots, T_\ell$ are, after order preserving relabellings, specifically bicoloured trees, i.e. elements of $\mathcal{T}_B$.
- Case 2: The root node is a white node. Then the red node is contained in one of the ℓ subtrees. Let us denote this subtree by T_s. After an order preserving relabelling T_s is itself an element of $\mathcal{T}_C$, while all remaining subtrees are, after order preserving relabellings, elements of $\mathcal{T}_B$. Moreover, the number of black ancestors of the red node in T_C is the same as the number of black ancestors of the red node in the subtree T_s.
- Case 3: The root node is a black node. Again the red node is contained in one of the ℓ subtrees, which we call T_s. After an order preserving relabelling T_s is an element of $\mathcal{T}_C$, whereas all remaining subtrees are, after order preserving relabellings, elements of $\mathcal{T}_B$. In this case the number of black ancestors of the red node in T_C is one more than the number of black ancestors of the red node in the subtree T_s.

By considering all tricoloured trees of $\mathcal{T}_C$, and taking into account (3.28), the above decomposition leads to

$$\begin{aligned}
N(z,u,q) &= \sum_{\ell\geq 0} \varphi_\ell\big(T(z+u)\big)^\ell + z\sum_{\ell\geq 0} \ell\varphi_\ell\big(T(z+u)\big)^{\ell-1} N(z,u,q) \\
&\quad + uq\sum_{\ell\geq 0} \ell\varphi_\ell\big(T(z+u)\big)^{\ell-1} N(z,u,q) \\
&= \varphi(T(z+u)) + z\varphi'(T(z+u))N(z,u,q) + uq\varphi'(T(z+u))N(z,u,q),
\end{aligned}$$

which is exactly the claimed equation (3.27) for $N(z,u,q)$. ☐

3.5.2 Computation of the factorial moments

Starting with the explicit formula for the trivariate generating function $N(z,u,q)$ given in Proposition 3.5.1 we will compute the r-th factorial moments of $I_{n,j}$. According to the definition (3.25) of $N(z,u,q)$ one obtains:

$$\begin{aligned}\mathbb{E}\big(I_{n,j}^{\underline{r}}\big) &= \frac{(j-1)!(n-j)!}{T_n}[z^{j-1}u^{n-j}]\mathrm{U}_q\mathrm{D}_q^r N(z,u,q)\\ &= \frac{(j-1)!\,(n-j)!\,r!}{T_n}[z^{j-1}u^{n-j-r}]\frac{\left(\varphi'(T(z+u))\right)^r\varphi(T(z+u))}{\left(1-(z+u)\varphi'(T(z+u))\right)^{r+1}}.\end{aligned}$$

Since for any power series $g(x)$ holds

$$[z^a u^b]g(z+u) = \binom{a+b}{a}[z^{a+b}]g(z), \tag{3.29}$$

one further obtains the following expression, which will be the starting point for our asymptotic considerations:

$$\begin{aligned}\mathbb{E}\big(I_{n,j}^{\underline{r}}\big) &= \frac{(j-1)!\,(n-j)!\,r!}{T_n}\binom{n-r-1}{j-1}[z^{n-r-1}]\frac{\left(\varphi'(T(z))\right)^r\varphi(T(z))}{\left(1-z\varphi'(T(z))\right)^{r+1}}\\ &= \frac{(j-1)!\,(n-j)!\,r!}{T_n}\binom{n-r-1}{j-1}[z^n]\frac{\left(z\varphi'(T(z))\right)^r T(z)}{\left(1-z\varphi'(T(z))\right)^{r+1}}.\end{aligned}\tag{3.30}$$

In order to evaluate $\mathbb{E}\big(I_{n,j}^{\underline{r}}\big)$ asymptotically we use the local expansions (3.6) and (3.9) and apply singularity analysis. Again for simplicity in presentation we will only carry out the computations for the case that the functions involved have $d=1$ dominant singularities (see Section 3.2.2); for $d>1$ one just has to add the contributions of all these singularities.

We obtain (for r arbitrary, but fixed):

$$\begin{aligned}[z^n]\frac{\left(z\varphi'(T(z))\right)^r T(z)}{\left(1-z\varphi'(T(z))\right)^{r+1}} &= [z^n]\frac{\left(1+\mathcal{O}\big((\rho-z)^{1/2}\big)\right)^r\tau\left(1+\mathcal{O}\big((\rho-z)^{1/2}\big)\right)}{\left(\sqrt{2\rho\tau\varphi''(\tau)}\sqrt{1-\frac{z}{\rho}}+\mathcal{O}(\rho-z)\right)^{r+1}}\\ &= [z^n]\frac{\tau}{\left(2\rho\tau\varphi''(\tau)\right)^{\frac{r+1}{2}}\left(1-\frac{z}{\rho}\right)^{\frac{r+1}{2}}}\left(1+\mathcal{O}\big((\rho-z)^{1/2}\big)\right)\\ &= \frac{\tau\, n^{\frac{r-1}{2}}}{\left(2\rho\tau\varphi''(\tau)\right)^{\frac{r+1}{2}}\rho^n\Gamma\left(\frac{r+1}{2}\right)}\left(1+\mathcal{O}\big(n^{-1/2}\big)\right).\end{aligned}$$

Together with the asymptotic formula for T_n given in (3.7), we obtain from (3.30) after simple computations:

$$\begin{aligned}\mathbb{E}\big(I_{n,j}^{\underline{r}}\big) &= \frac{(n-j)!r!(n-r-1)!\sqrt{2\pi\varphi''(\tau)}n^{\frac{3}{2}}\tau n^{\frac{r-1}{2}}}{(n-r-j)!n!\sqrt{\varphi(\tau)}(2\rho\tau\varphi''(\tau))^{\frac{r+1}{2}}\Gamma(\frac{r+1}{2})}\left(1+\mathcal{O}\big(n^{-1/2}\big)\right)\\ &= \frac{r!\sqrt{\pi}}{(2\rho\tau\varphi''(\tau))^{\frac{r}{2}}\Gamma(\frac{r+1}{2})}\frac{(n-j)^{\underline{r}}}{n^{\frac{r}{2}}}\left(1+\mathcal{O}\big(n^{-1/2}\big)\right).\end{aligned}$$

Finally, we use the duplication formula (cf. [AS64]) for the Gamma function

$$\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(\frac{r}{2}+1\right) = \frac{r!\sqrt{\pi}}{2^r},$$

and thus obtain the following expansion of the r-th factorial moment of $I_{n,j}$, which holds uniformly for all $1 \le j \le n$:

$$\mathbb{E}\big(I_{n,j}^{\underline{r}}\big) = \frac{2^{\frac{r}{2}}\Gamma\big(\frac{r}{2}+1\big)}{(\rho\tau\varphi''(\tau))^{\frac{r}{2}}}\frac{(n-j)^{\underline{r}}}{n^{\frac{r}{2}}}\left(1+\mathcal{O}\big(n^{-1/2}\big)\right). \tag{3.31}$$

3.5.3 Application of the method of moments

The asymptotic behaviour of the moments of $I_{n,j}$ for $n \to \infty$ depending on the growth of $j = j(n)$ can be obtained easily from the uniform expansion (3.31). An application of the method of moments shows then the limiting distribution results stated in Theorem 3.3.2.

3.5.3.1 Region $n - j \gg \sqrt{n}$

For this region it holds

$$\frac{(n-j)^{\underline{r}}}{n^{\frac{r}{2}}} = \frac{(n-j)^r}{n^{\frac{r}{2}}}\left(1+\mathcal{O}\big((n-j)^{-1}\big)\right) = \frac{(n-j)^r}{n^{\frac{r}{2}}}\left(1+\mathcal{O}\big(n^{-1/2}\big)\right),$$

which implies the following expansion for the factorial moments:

$$\mathbb{E}\big(I_{n,j}^{\underline{r}}\big) = \frac{2^{\frac{r}{2}}\Gamma\big(\frac{r}{2}+1\big)}{(\rho\tau\varphi''(\tau))^{\frac{r}{2}}}\frac{(n-j)^r}{n^{\frac{r}{2}}}\left(1+\mathcal{O}\left(n^{-\frac{1}{2}}\right)\right). \tag{3.32}$$

Together with equation (2.6), which connects the factorial and the ordinary moments, we obtain the following asymptotic expansion for the r-th moments of $I_{n,j}$:

$$\mathbb{E}\big(I_{n,j}^{r}\big) = \frac{2^{\frac{r}{2}}\Gamma(\frac{r}{2}+1)}{(\rho\tau\varphi''(\tau))^{\frac{r}{2}}}\frac{(n-j)^r}{n^{\frac{r}{2}}}\left(1+\mathcal{O}\left(\frac{\sqrt{n}}{n-j}\right)\right).$$

Thus we obtain for each fixed r and $n \to \infty$:

$$\mathbb{E}\Big(\Big(\frac{\sqrt{n}}{n-j} I_{n,j}\Big)^r\Big) \to \Big(\frac{1}{\sqrt{\rho\tau\varphi''(\tau)}}\Big)^r 2^{\frac{r}{2}}\Gamma\Big(\frac{r}{2}+1\Big),$$

i.e. the moments of $\frac{\sqrt{n}}{n-j} I_{n,j}$ converge to the moments of a Rayleigh distributed random variable with parameter $\sigma = \frac{1}{\sqrt{\rho\tau\varphi''(\tau)}}$. An application of the Theorem of Fréchet and Shohat (Theorem 2.2.1) now shows the corresponding limiting distribution result of Theorem 3.3.2.

3.5.3.2 Region $n - j \sim \alpha\sqrt{n}$, $\alpha \in \mathbb{R}^+$

Also for this region the asymptotic expansion (3.32) of the r-th factorial moments computed above holds and one further gets

$$\mathbb{E}\big(I_{n,j}^{\underline{r}}\big) \to \frac{2^{\frac{r}{2}}\Gamma\big(\frac{r}{2}+1\big)}{(\rho\tau\varphi''(\tau))^{\frac{r}{2}}}\alpha^r = \Big(\frac{\alpha}{\sqrt{\rho\tau\varphi''(\tau)}}\Big)^r 2^{\frac{r}{2}}\,\Gamma\Big(\frac{r}{2}+1\Big). \tag{3.33}$$

Our aim is now to find the distribution of a random variable Y whose factorial moments $\mathbb{E}(Y^{\underline{r}})$ are given by the right-hand side of (3.33) (a-priori we don't know whether this distribution is unique, but as it turns out, it is). This distribution can luckily be "guessed" by assuming that the probability generating function $p_Y(z)$ of Y exists in a neighbourhood of $z = 1$, in which case one can use

$$\begin{aligned}\sum_{r\ge0}\mathbb{E}(Y^{\underline{r}})\,\frac{z^r}{r!} &= \sum_{r\ge0}\sum_{k\ge0}\mathbb{P}\{Y=k\}\,k^{\underline{r}}\frac{z^r}{r!} = \sum_{k\ge0}\mathbb{P}\{Y=k\}\sum_{r\ge0}\binom{k}{r}z^r\\ &= \sum_{k\ge0}\mathbb{P}\{Y=k\}\,(z+1)^k = p_Y(z+1).\end{aligned}$$

We omit the computation of $p_Y(z)$, instead we directly state the result in the following lemma:

Lemma 3.5.2. *Let Y_γ, with $\gamma > 0$, be a discrete random variable with distribution*

$$\mathbb{P}\{Y_\gamma = k\} = \frac{\gamma^k}{k!}\int_0^\infty x^{k+1}e^{-\frac{x^2}{2}-\gamma x}dx, \quad \text{for } k \in \mathbb{N}_0.$$

Then it holds that the r-th factorial moments of Y_γ are given as follows:

$$\mathbb{E}\big(Y_\gamma^{\underline{r}}\big) = \gamma^r 2^{\frac{r}{2}}\Gamma\Big(\frac{r}{2}+1\Big).$$

Moreover, the distribution of Y_γ is uniquely defined by its sequence of moments.

Proof. For $r \geq 0$ (the case $r = 0$ shows that the probabilities sum up to 1, i.e. they define indeed a probability distribution) we get:

$$\begin{aligned}
\mathbb{E}\big(Y_\gamma^{\underline{r}}\big) &= \sum_{k\geq 0} k^{\underline{r}} \frac{\gamma^k}{k!} \int_0^\infty x^{k+1} e^{-\frac{x^2}{2}-\gamma x} dx = \int_0^\infty e^{-\frac{x^2}{2}-\gamma x} \gamma^r x^{r+1} \sum_{k\geq r} \frac{(\gamma x)^{k-r}}{(k-r)!} dx \\
&= \int_0^\infty e^{-\frac{x^2}{2}-\gamma x} \gamma^r x^{r+1} e^{\gamma x} dx = \gamma^r \int_0^\infty x^{r+1} e^{-\frac{x^2}{2}} dx = 2^{\frac{r}{2}} \gamma^r \int_0^\infty u^{\frac{r}{2}} e^{-u} du \\
&= \gamma^r \, 2^{\frac{r}{2}} \, \Gamma\Big(\frac{r}{2}+1\Big) .
\end{aligned}$$

In order to show that the sequence of moments uniquely characterizes the distribution we consider the moment generating function $m_{Y_\gamma}(s) := \mathbb{E}\big(\mathrm{e}^{sY_\gamma}\big)$ of Y_γ,

$$\begin{aligned}
m_{Y_\gamma}(s) &= \sum_{k\geq 0} \mathbb{P}\{Y_\gamma = k\}\, e^{ks} = \sum_{k\geq 0} \frac{(\mathrm{e}^s\gamma)^k}{k!} \int_0^\infty x^{k+1} e^{-\frac{x^2}{2}-\gamma x} dx \\
&= \int_0^\infty x e^{-\frac{x^2}{2}-\gamma x} \sum_{k\geq 0} \frac{(\mathrm{e}^s \gamma x)^k}{k!} dx = \int_0^\infty x e^{-\frac{x^2}{2}-\gamma x + \gamma \mathrm{e}^s x} dx.
\end{aligned}$$

Clearly, this function exists in a real neighbourhood of $s = 0$ (actually it exists for all real s), which implies (by Lemma 2.2.2) that the corresponding distribution is uniquely defined by its moments. $\square$

Since the r-th factorial moments (and thus also the ordinary moments) of $I_{n,j}$ converge to the corresponding moments of Y_γ, with $\gamma = \frac{\alpha}{\sqrt{\rho\tau\varphi''(\tau)}}$, an application of the Theorem of Fréchet and Shohat now shows the limiting distribution result stated in Theorem 3.3.2 for this case.

3.5.3.3 Region $n - j \ll \sqrt{n}$

From (3.31) one easily gets that $\mathbb{E}\big(I_{n,j}^{\underline{r}}\big) \to 0$, for $r \geq 1$, which, by an application of the Theorem of Fréchet and Shohat, shows $I_{n,j} \xrightarrow{(d)} 0$ as stated in the corresponding part of Theorem 3.3.2.

3.5.4 Explicit formulæ for probabilities

For some special tree families it is possible to obtain explicit formulæ for the probabilities $\mathbb{P}\{I_{n,j} = k\}$ by extracting coefficients from the trivariate generating function $N(z, u, q)$ as given in (3.26).

3.5.4.1 Unordered trees

For unordered trees ($\varphi(t) = \mathrm{e}^t$) the generating function $N(z, u, q)$ is given by the following expression:

$$N(z,u,q) = \frac{\mathrm{e}^{T(z+u)}}{1-(z+uq)\mathrm{e}^{T(z+u)}}, \quad \text{with } T(z) = z\mathrm{e}^{T(z)}.$$

From this one gets

$$\begin{aligned}
[q^k]N(z,u,q) &= \frac{\mathrm{e}^{T(z+u)}}{1-z\mathrm{e}^{T(z+u)}}[q^k]\frac{1}{1-\frac{u\mathrm{e}^{T(z+u)}}{1-z\mathrm{e}^{T(z+u)}}q} = u^k\left(\frac{\mathrm{e}^{T(z+u)}}{1-z\mathrm{e}^{T(z+u)}}\right)^{k+1} \\
&= u^k\sum_{\ell\geq 0}\binom{\ell+k}{k}\mathrm{e}^{(k+\ell+1)T(z+u)}z^\ell,
\end{aligned}$$

and further

$$[z^{j-1}u^{n-j}q^k]N(z,u,q) = \sum_{\ell=0}^{j-1}\binom{\ell+k}{k}[z^{j-\ell-1}u^{n-j-k}]\mathrm{e}^{(k+\ell+1)T(z+u)}.$$

Now, using (3.29), this leads to

$$[z^{j-1}u^{n-j}q^k]N(z,u,q) = \sum_{\ell=0}^{j-1}\binom{\ell+k}{k}\binom{n-k-\ell-1}{j-\ell-1}[z^{n-k-\ell-1}]\mathrm{e}^{(k+\ell+1)T(z)}. \tag{3.34}$$

In order to proceed, we use Lagrange's inversions formula (Theorem 2.1.3). Since $T(z)$ satisfies the equation $z = \frac{T(z)}{\mathrm{e}^{T(z)}}$, we have, for $a, b \in \mathbb{N}$,

$$[z^a]\mathrm{e}^{bT(z)} = \frac{b}{a}[T^{a-1}]\mathrm{e}^{bT}\mathrm{e}^{aT} = \frac{b(a+b)^{a-1}}{a!}. \tag{3.35}$$

Note that the last expression actually gives the correct coefficient also in the case $a = 0$. Hence, using (3.35) in (3.34), we get

$$[z^{j-1}u^{n-j}q^k]N(z,u,q) = \sum_{\ell=0}^{j-1}\binom{\ell+k}{k}\binom{n-k-\ell-1}{j-\ell-1}\frac{(k+\ell+1)n^{n-k-\ell-2}}{(n-k-\ell-1)!}.$$

By reversing the order of summation and using the relation

$$\mathbb{P}\{I_{n,j} = k\} = \frac{(j-1)!(n-j)!}{T_n}[z^{j-1}u^{n-j}q^k]N(z,u,q),$$

as well as the well-known formula $T_n = n^{n-1}$ for the number of unordered labelled trees (which can also be derived from (3.35)), we thus finally obtain the formula for the exact probabilities $\mathbb{P}\{I_{n,j} = k\}$ given in Theorem 3.3.2.

3.5.4.2 Ordered trees

For ordered trees ($\varphi(t) = \frac{1}{1-t}$), the generating function $N(z,u,q)$ is given by

$$N(z,u,q) = \frac{1 - T(z+u)}{(1 - T(z+u))^2 - z - uq}, \quad \text{with } T(z) = \frac{z}{1 - T(z)}.$$

From this one can again extract the coefficient of $z^{j-1}u^{n-j}q^k$ in order to obtain the required probabilities. We omit the necessary computations since they are very similar to the ones in the previous case, but we stated the corresponding result in (3.11).

CHAPTER 4

Increasing k-trees

4.1 Introduction

k-trees, also called k-dimensional trees, are certain simple graphs that have been introduced and studied first by Beineke and Pippert [BP69], Moon [Moo69], and Harary and Palmer [HP68]. These graphs owe their name to the fact that they allow a recursive description analogous to trees: a k-tree T is either a k-clique (i.e. a complete connected graph with k nodes) or there exists a node u, which is incident to exactly k edges that connect this node to all of the nodes of a k-clique, such that, when removing u and the k incident edges from T, the remaining graph is itself a k-tree.

Since the pioneering studies mentioned before various families of k-trees (such as, e.g., labelled and unlabelled, ordered and unordered ones) have been introduced and considered. Very recently an interesting random graph model based on k-trees has been proposed in [Gao09] and some quantities (the degree distribution and distance parameters) have been analyzed in [Gao09] and [DHBS10], respectively. The model can be described by a simple graph evolution process. Starting with a k-clique (the so-called root clique) of nodes (the so-called root nodes) labelled by $0_1, 0_2, \dots, 0_k$, successively the nodes with labels $1, 2, \dots, n$ are inserted, where in each step the new node j is attached to all of the nodes of an already existing k-clique. In each insertion step one uses a “uniform attachment”-rule, i.e. one of the already existing k-cliques is chosen uniformly at random to attach the new node.

When considering the special instance $k = 1$ the resulting graphs are of course trees and this “uniform attachment”-rule leads to an important tree model called recursive trees, which is of interest in combinatorics and probability theory, see [Drm09, MS95]. For trees also other probabilistic growth rules have been introduced leading to further important tree models. In particular when applying “preferential attachment” (or “success breeds success”), i.e. carrying out the insertion process where in each step the probability that a new node will be attached to an already existing node x is proportional to one plus the number of nodes already attached to x, one obtains the model of plane recursive trees (also called plane-oriented recursive trees, or non-uniform recursive trees), see [MS95]. Furthermore, when applying a “saturation”-rule, where there is a maximum d for the number of children that can be attached to a node, and the probability that a new node will be attached to an already existing node x is proportional to the difference between the maximum possible number d and the actual number of nodes already attached to x, this leads to tree models known in combinatorics as d-ary increasing trees.

It seems natural to also apply these growth rules to generate k-trees, which will lead then to random k-tree models different from the one introduced in [Gao09]. To be precise, for the “preferential attachment”-rule we modify the graph evolution process for k-trees described above, such that in each insertion step the following probabilistic growth rule will be applied: the probability that a new node will be attached to an already existing k-clique is proportional to one plus the number of nodes that have been previously attached to this k-clique. Furthermore, for the “saturation”-rule we modify the graph evolution process by applying in each insertion step the following probabilistic growth rule: the probability that a new node will be attached to an already existing k-clique is proportional to the difference between the maximum possible number d of children and the actual number of nodes already attached to the considered k-clique.

Thus we will deal with three different random k-tree models, which we will call:

- unordered increasing k-trees (they correspond to “uniform attachment”),
- ordered increasing k-trees (they correspond to “preferential attachment”),
- d-ary increasing k-trees (they correspond to the “saturation”-rule).

These notions will be justified in the following section.

4.2 Combinatorial description of increasing k-trees

We now give a combinatorial description of the considered k-tree models. We will here always consider rooted k-trees, which means that in each k-tree one k-clique is distinguished as the root clique. The nodes contained in the root clique are called root nodes, whereas the remaining nodes are non-root nodes. For the k-tree models studied in this work we will also call the non-root nodes "inserted nodes". Then, apart from the edges connecting the root nodes with each other, this induces a natural orientation on the edges: For each non-root node, we can distinguish between ingoing edges (coming from the direction of the root clique) and outgoing edges, which also defines the in-degree $d^-(u)$ and the out-degree $d^+(u)$ of a node u. For a root node we will only define the out-degree. It is immediate from the definition that each non-root node u has exactly k ingoing edges, and these edges connect u with a k-clique $K = \{w_1, \ldots, w_k\}$. We might then say that u is a child of the k-clique K or that u is attached to K and that $w_1, \ldots, w_k$ are the parents of u. For the degree $d(u)$ of a node u it holds that $d(u) = d^+(u) + k$ for a non-root node and $d(u) = d^+(u) + k - 1$ for a root node. We also define the out-degree $d^+(K)$ of a k-clique K as the number of children of K.

For our purpose of modelling different growth rules it is important to introduce the following three variants of rooted k-trees:

- Unordered k-trees: one assumes that to each k-clique there is attached a set of children.
- Ordered k-trees: one assumes that to each k-clique there is attached a sequence of children, i.e. the children of each k-clique are linearly ordered and one might speak about the first, second, etc. child of a k-clique.
- d-ary k-trees: they can be considered as ordered k-trees, where each k-clique has exactly d positions at which a child might be attached or not (thus there are exactly $\binom{d}{l}$ different ways in which the sequence of $0 \leq l \leq d$ nodes $w_1, w_2, \ldots, w_l$ can be attached to a k-clique K in this linear order).

Furthermore, we introduce specific labellings of the nodes of k-trees, which might be called increasing labellings (in analogy to the corresponding term for trees, see, e.g., [BFS92]). Given a (unordered, ordered, d-ary) k-tree with n non-root nodes we label the set of root nodes by $\{0_1, \ldots, 0_k\}$, whereas the non-root nodes

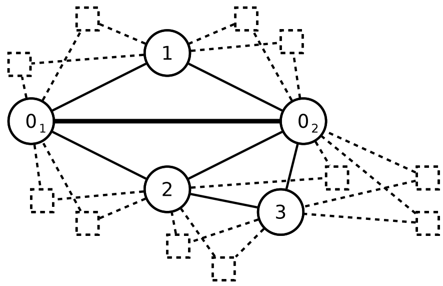

Figure 4.1: A binary increasing 2-tree of size 3. Boxes represent empty slots.

are labelled by $\{1, \ldots, n\}$ in such a way that the label of a node is always larger than the labels of all its parent nodes (of course, in this context the value of 0_ℓ, $1 \leq \ell \leq k$, is defined as 0).

Then the following graph families:

- $\mathcal{T}^{[u]} := \mathcal{T}^{[u]}(k)$: "unordered increasing k-trees",
- $\mathcal{T}^{[o]} := \mathcal{T}^{[o]}(k)$: "ordered increasing k-trees",
- $\mathcal{T}^{[d]} := \mathcal{T}^{[d]}(k)$: "$d$-ary increasing k-trees",

can be described combinatorially as the family of all increasingly labelled unordered k-trees, ordered k-trees, and d-ary k-trees, respectively. It is apparent from the definition, see [Drm09], that for $k = 1$ one gets the tree families of recursive trees, plane recursive trees, and d-ary increasing trees, respectively. We remark that it is often appropriate to add at each position in a d-ary increasing k-tree, where no child has been attached, a so-called "external node" (which does not get any label) that represents an "empty slot".

Throughout our work we use the convention that the size $|T|$ of a k-tree T is given by the number of non-root nodes. Thus the k-tree consisting only of the root clique $K_0 = \{0_1, \ldots, 0_k\}$ has size 0. Examples of k-trees from the families considered are given in Figures 4.1 and 4.2.

4.3 The number of increasing k-trees

For each of the families $\mathcal{T}^{[u]}$, $\mathcal{T}^{[o]}$ and $\mathcal{T}^{[d]}$, respectively, let T_n denote the number of different increasing k-trees of size n (we do not explicitly express the dependence

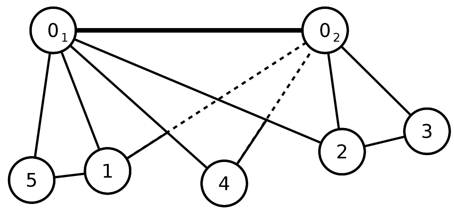

Figure 4.2: Two different ordered increasing 2-trees of size 5 (the order on the children of each 2-clique is expressed by drawing the nodes “in front of each other“). Regarded as unordered increasing 2-trees these are just two different drawings of the same object.

on k, which is of course always given). There exist simple enumeration formulæ for T_n as is shown next.

Proposition 4.3.1. *The number T_n of different increasing k-trees of size n is, for $n \in \mathbb{N}_0$, given as follows:*

$$T_n = \begin{cases} \prod_{\ell=0}^{n-1} \big(1 + k\ell\big) = n! k^n \binom{n-1+\frac{1}{k}}{n}, & \textit{family } \mathcal{T}^{[u]}, \\ \prod_{\ell=0}^{n-1} \big(1 + (k+1)\ell\big) = n!(k+1)^n \binom{n-1+\frac{1}{k+1}}{n}, & \textit{family } \mathcal{T}^{[o]}, \\ \prod_{\ell=0}^{n-1} \big(d + (kd-1)\ell\big) = n!(kd-1)^n \binom{n-1+\frac{d}{kd-1}}{n}, & \textit{family } \mathcal{T}^{[d]}. \end{cases} \tag{4.1}$$

Proof. We consider the three graph families separately.

- Family $\mathcal{T}^{[u]}$: Obviously it holds $T_0 = T_1 = 1$. To get an enumeration formula for T_n we observe that when inserting a node into an unordered increasing k-tree this always increases the number of possible ways of attaching a further node by k due to the newly generated k-cliques. Thus there are always $1 + k(n-1)$ possible ways of inserting node n into an unordered increasing k-tree of size $n-1$. Since each unordered increasing k-tree of size n is uniquely obtained from an unordered increasing k-tree of size $n-1$ and the insertion of node n in a possible way, it holds that $T_n = (1 + k(n-1))T_{n-1}$, which shows that the number of different unordered increasing k-trees of size n is given by equation (4.1).

- Family $\mathcal{T}^{[o]}$: Again it holds $T_0 = T_1 = 1$. When inserting a node into an ordered increasing k-tree this always increases the number of possible ways of attaching a further node by $k+1$: k due to the newly generated k-cliques and a further one due to a new available position at the parent k-clique of the newly inserted node. Thus there are always $1+(k+1)(n-1)$ possible ways of inserting node n into an ordered increasing k-tree of size $n-1$, which implies $T_n = (1+(k+1)(n-1))T_{n-1}$ and equation (4.1).

- Family $\mathcal{T}^{[d]}$: It holds $T_0 = 1$ and $T_1 = d$. When inserting a node into a d-ary increasing k-tree this always increases the number of possible ways of attaching a further node by $kd-1$: one gets kd new empty slots due to the newly generated k-cliques, but one previously empty slot is now occupied by the new node. Thus there are always $d+(kd-1)(n-1)$ possible ways of inserting node n into a d-ary increasing k-tree of size $n-1$, which implies $T_n = (d+(kd-1)(n-1))T_{n-1}$ and equation (4.1).

□

4.3.1 "Top-down" computation of T_n

In the proof of Proposition 4.3.1 we computed the numbers T_n of unordered, ordered and d-ary increasing k-trees, respectively, in a "bottom-up" fashion, i.e. by inductively considering the number of possible ways in which node n can be attached to an increasing k-tree of size $n-1$. We now present an alternative approach which works "top-down", i.e. by describing how an increasing k-tree of size n can be decomposed into smaller increasing k-trees. We will later use the same approach when we consider the degree distribution in random increasing k-trees (Section 4.8).

Let $\mathcal{T}$ be one of the families $\mathcal{T}^{[u]}$, $\mathcal{T}^{[o]}$ or $\mathcal{T}^{[d]}$. We construct an auxiliary family $\tilde{\mathcal{T}} := \tilde{\mathcal{T}}^{[u]}$, $\tilde{\mathcal{T}}^{[o]}$ or $\tilde{\mathcal{T}}^{[d]}$, respectively, in the following way: $\tilde{\mathcal{T}}$ contains all increasingly labelled k-trees which can be created from the $(k+1)$-clique consisting of the nodes $\{0_1, 0_2, \ldots 0_k, 1\}$ (which we will denote by Δ) and k elements $T_1, \ldots, T_k$ of $\mathcal{T}$ by

- relabelling the non-root nodes of $T_1, \ldots, T_k$ in an order-preserving way, such that exactly the labels $2, \ldots, n$ are used, where $n := |T_1| + \cdots + |T_k| + 1$, and
- identifying the root-cliques of the T_is each with one of the k-cliques of Δ which contain node 1.

In other words, $\tilde{\mathcal{T}}$ is constructed from Δ and $\mathcal{T}$ by using a boxed product (cf. Section 2.1.1):

$$\tilde{\mathcal{T}} = \Delta^{\square} * \mathcal{T}^k. \tag{4.2}$$

Note that $\tilde{\mathcal{T}}^{[u]}$ and $\tilde{\mathcal{T}}^{[o]}$ are just the families of unordered and ordered increasing k-trees, respectively, in which the root-clique has exactly one child. In contrast, $\tilde{\mathcal{T}}^{[d]}$ is not a sub-family of $\mathcal{T}^{[d]}$: In each element of $\tilde{\mathcal{T}}^{[d]}$, the root clique has exactly one child but no "empty slots".

From the combinatorial description in Section 4.2 one gets the following formal equations, which describe how $\mathcal{T}^{[u]}$, $\mathcal{T}^{[o]}$ and $\mathcal{T}^{[d]}$ can be constructed from $\tilde{\mathcal{T}}^{[u]}$, $\tilde{\mathcal{T}}^{[o]}$ and $\tilde{\mathcal{T}}^{[d]}$, respectively ($\square$ denotes a k-tree of size 0, i.e. it is a placeholder for an empty slot):

$$\mathcal{T}^{[u]} = \textsc{Set}\left(\tilde{\mathcal{T}}^{[u]}\right), \tag{4.3}$$

$$\mathcal{T}^{[o]} = \textsc{Seq}\left(\tilde{\mathcal{T}}^{[o]}\right), \tag{4.4}$$

$$\mathcal{T}^{[d]} = (\{\square\} \,\dot\cup\, \tilde{\mathcal{T}}^{[d]})^d. \tag{4.5}$$

We can now apply the symbolic method in order to re-obtain the enumeration result given in Proposition 4.3.1. Let us denote by T_n the number of unordered, ordered or d-ary increasing k-trees of size n, respectively, and by $\tilde{T}_n$ the number of objects of size n in the corresponding auxiliary family $\tilde{\mathcal{T}}$. Furthermore, we define

$$T(z) = \sum_{n\geq 0} T_n \frac{z^n}{n!}, \quad \text{and}$$
$$\tilde{T}(z) = \sum_{n\geq 1} \tilde{T}_n \frac{z^n}{n!}.$$

By an application of the formal method to the formal equations (4.2)-(4.5) one gets the following system of equations:

$$\tilde{T}'(z) = T(z)^k, \quad \tilde{T}(0) = 0, \tag{4.6}$$

$$T(z) = \begin{cases} \exp\left(\tilde{T}(z)\right), & \text{family } \mathcal{T}^{[u]}, \\ \frac{1}{1-\tilde{T}(z)}, & \text{family } \mathcal{T}^{[o]}, \\ (1+\tilde{T}(z))^d, & \text{family } \mathcal{T}^{[d]}. \end{cases} \tag{4.7}$$

Inserting (4.7) into (4.6), we obtain in all three cases a simple ordinary differential equation for $\tilde{T}(z)$, which can be solved by separation of variables. The solution is given by

$$\tilde{T}(z) = \begin{cases} \frac{1}{k}\log\left(\frac{1}{1-kz}\right), & \text{family } \mathcal{T}^{[u]}, \\ 1-(1-(k+1)z)^{\frac{1}{k+1}}, & \text{family } \mathcal{T}^{[o]}, \\ \frac{1}{(1-(kd-1)z)^{\frac{1}{kd-1}}} - 1, & \text{family } \mathcal{T}^{[d]}, \end{cases}$$

and by inserting this into (4.7), we obtain

$$T(z) = \begin{cases} \frac{1}{(1-kz)^{\frac{1}{k}}}, & \text{family } \mathcal{T}^{[u]}, \\ \frac{1}{(1-(k+1)z)^{\frac{1}{k+1}}}, & \text{family } \mathcal{T}^{[o]}, \\ \frac{1}{(1-(kd-1)z)^{\frac{d}{kd-1}}}, & \text{family } \mathcal{T}^{[d]}. \end{cases} \tag{4.8}$$

From this, the formulæ for the numbers T_n as given in (4.1) are easily re-discovered via the relation $T_n = n![z^n]T(z)$.

4.4 Relation to the considered growth rules

When studying parameters in unordered, ordered and d-ary increasing k-trees we always assume the “random increasing k-tree model” of the corresponding family, which means that we assume that each of the T_n increasing k-trees of size n of the family considered appears with the same probability. It remains to show that this combinatorial description of the random k-tree models indeed coincides with the probabilistic growth rules discussed in the introduction.

Proposition 4.4.1. *The following graph evolution process generates unordered, ordered, and d-ary increasing k-trees, respectively, uniformly at random:*

- *Step* 0*: start with the root clique labelled by* $0_1, 0_2, \dots, 0_k$.
- *Step* n*: the node with label* n *is attached to any* k*-clique* K *with out-degree* $d^+(K)$ *in the already grown* k*-tree of size* $n-1$ *with a probability* $p(K)$ *given as follows:*

$$p(K) = \begin{cases} \dfrac{1}{1+k(n-1)}, & \textit{family } \mathcal{T}^{[u]}, \\ \dfrac{d^+(K)+1}{1+(k+1)(n-1)}, & \textit{family } \mathcal{T}^{[o]}, \\ \dfrac{d-d^+(K)}{d+(kd-1)(n-1)}, & \textit{family } \mathcal{T}^{[d]}. \end{cases} \tag{4.9}$$

Proof. Again we consider the three graph families separately.

- Family $\mathcal{T}^{[u]}$: each unordered increasing k-tree of size n can be obtained in a unique way by attaching node n to one of the $1 + k(n-1)$ existing k-cliques of an unordered increasing k-tree of size $n-1$. Thus the "uniform attachment"-rule generates k-trees of this family uniformly at random.

- Family $\mathcal{T}^{[o]}$: when a k-clique K in an ordered increasing k-tree has ℓ children, i.e. $d^+(K) = \ell$, then there are always exactly $\ell+1$ possible ways of attaching a new node to K, namely as the first child, second, child, $\dots$, $(\ell+1)$-th child, leading to different ordered increasing k-trees. Since there are exactly $1+(k+1)(n-1)$ possibilities of inserting node n into an ordered increasing k-tree of size $n-1$, the stated "preferential attachment"-rule generates k-trees of this family uniformly at random.

- Family $\mathcal{T}^{[d]}$: when a k-clique K in a d-ary increasing k-tree has ℓ children, i.e. $d^+(K) = \ell$, then there are always exactly $d-\ell$ possible ways of attaching a new node to K, namely at one of $d-\ell$ empty slots, leading to different d-ary increasing k-trees. Since there are exactly $d+(kd-1)(n-1)$ possibilities (= the total number of empty slots) of inserting node n into a d-ary increasing k-tree of size $n-1$, the stated "saturation"-rule generates k-trees of this family uniformly at random.

□

4.5 Considered quantities

We will study various quantities in our three k-tree models, which shall give some insight into the structure of these random objects, namely the number of ancestors, the number of descendants, and the out-degree as well as the local clustering coefficient of the nodes. We will now define what we mean by these notions. Note that in the rest of this chapter we always identify each node with its label.

- We call node y an ancestor of node x in a k-tree T if there exists a path $0_\ell = w_0, w_1, \dots, w_r = x$ in T with the property that each node w_i has been inserted after node w_{i-1}, $1 \le i \le r$, and which contains node y; we might call such a path an increasing path, since it holds that the labels of this path are forming an increasing sequence, i.e. $w_0 < w_1 < \cdots < w_r$ (again, in this context the value of a root node 0_ℓ, $1 \le \ell \le k$, is defined as 0).

- We call node y a descendant of node x exactly if node x is an ancestor of y. We also want to mention the following equivalent, but recursive definition of the term descendant, which is advantageous for our analysis: node y is a descendant of node x, iff
  - either $x = y$, or
  - y is a child of a descendant of x.
- As we already defined in the beginning of Section 4.2, the out-degree $d^+(x)$ of node x is the number of children of x.
- The local clustering coefficient has been introduced by [WS98] and is considered as an important parameter in the study of real-world networks. The local clustering coefficient $C_G(u)$ of a node u in a graph $G(V, E)$ is defined as the proportion of edges between neighbours of u divided by the number of edges between the neighbours that could possibly exist. Formally, $C_G(u)$ is given by

$$C_G(u) := \begin{cases} \frac{|\{(x,y)\in E(G) \mid x,y\in N(u)\}|}{\binom{d(u)}{2}}, & \text{if } d(u) \geq 2, \\ 0, & \text{if } d(u) = 0 \text{ or } d(u) = 1, \end{cases} \tag{4.10}$$

  where $E(G)$ denotes the set of edges in G and $N(u)$ the set of neighbours (i.e. adjacent nodes) of u.

4.5.1 Relation to existing work

Before we start with our analysis, we want to mention that the quantities we are considering have already been thoroughly studied in the case of trees (i.e. in the case $k = 1$).

First of all, for trees there is a unique path from the root node to an arbitrary node x and the number of ancestors of node x is exactly one plus the number of edges contained in this path, i.e. it is one plus the root-to-node distance, also called the depth, of node x. Thus for trees this quantity has been studied extensively. In particular limiting distribution results for the number of ancestors of a random node and of the last inserted node, respectively, in a random tree of size n have been obtained for recursive trees, plane-oriented recursive trees and d-ary increasing trees, see, e.g., [DS96, MS95, PP07]. Although, for $k \geq 2$, the number of ancestors of x is not closely related to the ordinary distance, i.e. the shortest-path distance, between root nodes and node x, we remark that the analogy to trees allow an interpretation of this quantity as a kind of "clique-distance": the number of ancestors of node x is exactly k plus the number of $(k+1)$-cliques contained in the smallest sub-k-tree containing the root clique and node x. Furthermore, it is not difficult to see that for each inserted node x the number of ancestors of x is given by k plus the number of edges in a longest increasing path starting at a root node 0_ℓ and ending at x.

The number of descendants of node x in a tree can be described easily as the size of the subtree rooted at node x. However, for k-trees with $k \geq 2$, there does not seem to exist such a simple description. There exist various studies concerning descendants in increasing trees, see, e.g., [KP06, Pro96] for results concerning the number of descendants of nodes in random trees for random tree models mentioned before.

Of course, also the node degrees in random increasing trees have already been studied, for example in [KP07].

In our work we will give a precise distributional analysis of the mentioned quantities in a random k-tree of size n for all growth models described above. We will not only provide results for random nodes, but a main emphasis is given on describing the behaviour of the parameters for the j-th inserted node in a random k-tree of size n, depending on the growth of $j = j(n)$. The results describe quite well aspects of the local behaviour of the nodes during the graph evolution process.

Since various tree models are contained in our k-tree models as the special instance $k = 1$, our studies mostly generalize the corresponding results for these tree families. But note that for some quantities (e.g., for the out-degree of the nodes in unordered increasing k-trees), the case $k = 1$ shows a behaviour which is qualitatively different from the cases where $k \geq 2$. For better readability, and since the case $k = 1$ has already been thoroughly studied, we will thus always assume that $k \geq 2$ in our analysis.

As a final remark in this section, we want to mention that the special instance $d = 1$ of d-ary increasing k-trees leads to network models which have been introduced previously and are known as random Apollonian networks, see, e.g., [AM08].

4.5.2 Considered random variables

In the following, the random variable $D_{n,j}$, with $n \geq j \geq 1$, counts the number of descendants of node j in a random increasing k-tree of size n. Furthermore the random variable $\overline{D}_n$ counts the number of descendants of a randomly selected inserted node in a random increasing k-tree of size $n \geq 1$, i.e. $\overline{D}_n \stackrel{(d)}{=} D_{n,U_n}$, where $U_n \stackrel{(d)}{=} \text{Uniform}(\{1, 2, \ldots, n\})$ denotes a discrete random variable that is uniformly distributed on $\{1, \ldots, n\}$. The random variable A_n counts the number of ancestors of node n in a random increasing k-tree of size $n \geq 1$. We note that it is not necessary to study the random variable $A_{n,j}$ counting the number of ancestors of node j in a random increasing k-tree of size n separately, since it holds $A_{n,j} \stackrel{(d)}{=} A_{j,j} = A_j$, which is a direct consequence of the graph evolution process for the k-tree models considered. Furthermore, the random variable $\overline{A}_n$ counts the number of ancestors of a randomly selected inserted node in a random increasing k-tree of size $n \geq 1$, i.e. $\overline{A}_n \stackrel{(d)}{=} A_{U_n}$, where $U_n \stackrel{(d)}{=} \text{Uniform}(\{1, 2, \ldots, n\})$. The random variable $O_{n,j}$, with $n \geq j \geq 1$, denotes the out-degree of node j in a random increasing k-tree of size n, and $\bar{O}_n$ the out-degree of a randomly selected node in a random increasing k-tree of size n. Finally, the random variable C_n denotes the local clustering coefficient of a randomly selected node (amongst the root nodes and the inserted nodes) in a random increasing k-tree of size n.

Note that we will not explicitly express the dependence of the random variables on k in our notation, although this dependence is of course always given.

4.6 Ancestors

4.6.1 Results

Theorem 4.6.1. *The random variable A_n, which counts the number of ancestors of node n in a random increasing k-tree of size n, has the following exact distribution:*

$$\mathbb{P}\{A_n = m + k\} = \begin{cases} \frac{1}{\binom{n-1+\frac{1}{k}}{n-1}} \sum_{i=m}^{n} \frac{{i-1 \brack m-1}}{(i-1)!} \binom{n-i-1+\frac{1}{k}}{n-i}, & \textit{family } \mathcal{T}^{[u]}, \\ \frac{k^{m-1}}{(k+1)^{m-1}\binom{n-1+\frac{1}{k+1}}{n-1}} \sum_{i=m}^{n} \frac{{i-1 \brack m-1}}{(i-1)!} \binom{n-i-1+\frac{2}{k+1}}{n-i}, & \textit{family } \mathcal{T}^{[o]}, \\ \frac{(kd)^{m-1}}{(kd-1)^{m-1}\binom{n-1+\frac{d}{kd-1}}{n-1}} \sum_{i=m}^{n} \frac{{i-1 \brack m-1}}{(i-1)!} \binom{n-i-1+\frac{d-1}{kd-1}}{n-i}, & \textit{family } \mathcal{T}^{[d]}, \end{cases}$$

for $n \geq m \geq 1$, where the numbers ${i \brack j}$ are the (unsigned) Stirling numbers of the first kind.

Furthermore, A_n admits the following decomposition into a sum of independent random variables:

$$A_n = (k+1) \oplus \mathbb{1}(\mathcal{C}_{n,1}) \oplus \mathbb{1}(\mathcal{C}_{n,2}) \oplus \cdots \oplus \mathbb{1}(\mathcal{C}_{n,n-1}),$$

where $\mathcal{C}_{n,j}$ denotes the event that node j is an ancestor of node n. It holds that the indicator variables of the events $\mathcal{C}_{n,j}$, $1 \le j \le n-1$, are Bernoulli distributed random variables with success probabilities independent of n: $\mathbb{1}(\mathcal{C}_{n,j}) \stackrel{(d)}{=} \text{Bernoulli}(p_j)$, with

$$p_j = \mathbb{P}\{\mathcal{C}_{n,j}\} = \begin{cases} \frac{k}{kj+1}, & \textit{family } \mathcal{T}^{[u]}, \\ \frac{k}{(k+1)j+1}, & \textit{family } \mathcal{T}^{[o]}, \\ \frac{kd}{(kd-1)j+d}, & \textit{family } \mathcal{T}^{[d]}. \end{cases}$$

Theorem 4.6.2. *The random variable A_n is, for $n \to \infty$, asymptotically normally distributed:*

$$\mathbb{P}\left\{ \frac{A_n - \mathbb{E}(A_n)}{\sqrt{\mathbb{V}(A_n)}} \le x \right\} = \Phi(x) + \mathcal{O}\left(\frac{1}{\sqrt{\log n}} \right),$$

where $\Phi(x)$ denotes the distribution function of the standard normal distribution. The expectation $\mathbb{E}(A_n)$ and the variance $\mathbb{V}(A_n)$ satisfy

$$\mathbb{E}(A_n) = \begin{cases} \log n + \mathcal{O}(1), & \textit{family } \mathcal{T}^{[u]}, \\ \frac{k}{k+1} \log n + \mathcal{O}(1), & \textit{family } \mathcal{T}^{[o]}, \\ \frac{kd}{kd-1} \log n + \mathcal{O}(1), & \textit{family } \mathcal{T}^{[d]}, \end{cases}$$

and $\mathbb{V}(A_n) = \mathbb{E}(A_n) + \mathcal{O}(1)$.

Theorem 4.6.3. *The random variable $\overline{A}_n$, which counts the number of ancestors of a randomly selected inserted node in a random increasing k-tree of size n, is, for $n \to \infty$, asymptotically normally distributed:*

$$\mathbb{P}\left\{ \frac{\overline{A}_n - \mathbb{E}(\overline{A}_n)}{\sqrt{\mathbb{V}(\overline{A}_n)}} \le x \right\} = \Phi(x) + \mathcal{O}\left(\frac{1}{\sqrt{\log n}} \right),$$

where the expectation $\mathbb{E}(\overline{A}_n)$ and the variance $\mathbb{V}(\overline{A}_n)$ satisfy $\mathbb{E}(\overline{A}_n) = \mathbb{E}(A_n) + \mathcal{O}(1)$ and $\mathbb{V}(\overline{A}_n) = \mathbb{E}(\overline{A}_n) + \mathcal{O}(1)$.

4.6.2 Proofs of the results

4.6.2.1 Derivation of the exact distribution

We derive the exact distribution of A_n by finding a recurrence relation for the probabilities $\mathbb{P}\{A_n = m\}$. Consider, for $n \ge 2$, the event

$$\mathcal{C}_n := [\text{node } n \text{ is a child of node } n-1].$$

By the law of total probability, it holds that

$$\mathbb{P}\{A_n = m\} = \mathbb{P}\{\mathcal{C}_n\}\,\mathbb{P}\{A_n = m \mid \mathcal{C}_n\} + \mathbb{P}\{\mathcal{C}_n^c\}\,\mathbb{P}\{A_n = m \mid \mathcal{C}_n^c\}\,,$$

where $\mathbb{P}\{\mathcal{A} \mid \mathcal{B}\}$ denotes the conditional probability of event $\mathcal{A}$, given event $\mathcal{B}$, and $\mathcal{C}_n^c$ is the complement of event $\mathcal{C}_n$. Now we use the following two observations:

- If node n is a child of a node x, then it has exactly one ancestor more than x. It follows in particular that

$$\mathbb{P}\{A_n = m \mid \mathcal{C}_n\} = \mathbb{P}\{A_{n-1} = m-1 \mid \mathcal{C}_n\}\,.$$

- Consider the set of (unordered, ordered or d-ary) increasing k-trees with the property that node n is not a child of node $n-1$. Clearly this set can be mapped onto itself bijectively by swapping the nodes n and $n-1$. This implies

$$\mathbb{P}\{A_n = m \mid \mathcal{C}_n^c\} = \mathbb{P}\{A_{n-1} = m \mid \mathcal{C}_n^c\}\,.$$

Furthermore, the number of ancestors of node $n-1$ is of course independent from the event $\mathcal{C}_n$, i.e. we have $\mathbb{P}\{A_{n-1} = m-1 \mid \mathcal{C}_n\} = \mathbb{P}\{A_{n-1} = m-1\}$ and $\mathbb{P}\{A_{n-1} = m \mid \mathcal{C}_n^c\} = \mathbb{P}\{A_{n-1} = m\}$. Hence we get the desired recurrence for the probabilities:

$$\mathbb{P}\{A_n = m\} = \mathbb{P}\{\mathcal{C}_n\}\,\mathbb{P}\{A_{n-1} = m-1\} + \mathbb{P}\{\mathcal{C}_n^c\}\,\mathbb{P}\{A_{n-1} = m\}\,, \tag{4.11}$$

for $n > 1$ and $m \geq 1$. Of course, the probabilities $\mathbb{P}\{\mathcal{C}_n\}$ in this recurrence depend on the family of increasing k-trees we are considering and, due to the graph evolution process stated in Proposition 4.4.1, they are given as follows:

$$\mathbb{P}\{\mathcal{C}_n\} = \begin{cases} \frac{k}{k(n-1)+1}, & \text{family } \mathcal{T}^{[u]}, \\ \frac{k}{(k+1)(n-1)+1}, & \text{family } \mathcal{T}^{[o]}, \\ \frac{kd}{(kd-1)(n-1)+d}, & \text{family } \mathcal{T}^{[d]}. \end{cases}$$

Before treating recurrence (4.11) we note that the above considerations show, for $n > 1$, a decomposition

$$A_n = \tilde{A}_{n-1} \oplus \mathbb{1}(\mathcal{C}_n) = \tilde{A}_{n-1} \oplus \mathbb{1}(\mathcal{C}_{n,n-1}), \quad \text{with } \tilde{A}_{n-1} \stackrel{(d)}{=} A_{n-1}, \tag{4.12}$$

and $\mathcal{C}_{n,n-1}$ defined as in Theorem 4.6.1. Here the random variable $\tilde{A}_{n-1}$ counts the number of ancestors of node $n-1$ in the increasing k-tree $\tilde{T}$ (of size $n-1$)

obtained by starting with a random increasing k-tree T of size n and applying the following simple construction: (i) if n is a child of $n-1$, i.e. event $\mathcal{C}_n$ occurs for T, then $\tilde{T}$ is obtained by removing node n from T; (ii) if n is not a child of $n-1$, then $\tilde{T}$ is obtained from T by interchanging the nodes n and $n-1$, and removing node n afterwards. Iterating the decomposition (4.12) and taking into account that $\tilde{A}_1 = A_1 = (k+1)$ (the ancestors of node 1 are the k root nodes $0_1, \dots, 0_k$ and node 1 itself) immediately shows the corresponding result in Theorem 4.6.1.

Now we continue our computations leading to an exact formula for the probabilities $\mathbb{P}\{A_n = m\}$ by treating (4.11) (of course, one could alternatively study further the decomposition of A_n into independent random variables to show these results).

Family $\mathcal{T}^{[u]}$ (unordered increasing k-trees): In the case of unordered increasing k-trees, (4.11) reads

$$\mathbb{P}\{A_n = m\} = \frac{k}{k(n-1)+1}\mathbb{P}\{A_{n-1} = m-1\} + \frac{k(n-2)+1}{k(n-1)+1}\mathbb{P}\{A_{n-1} = m\}\,.$$

Multiplying this equation by T_n, we get the following recurrence for the number $T_{n,m} := T_n \mathbb{P}\{A_n = m\}$ of increasing k-trees of size n in which node n has m ancestors:

$$T_{n,m} = kT_{n-1,m-1} + (k(n-2)+1)T_{n-1,m}, \tag{4.13}$$

for $n > 1$ and $m \geq 1$. Furthermore, we have the initial values $T_{1,k+1} = 1$, and $T_{1,m} = 0$ for $m \neq k+1$ (since there is exactly one unordered increasing k-tree of size 1, and in this k-tree node 1 has $k+1$ ancestors). In order to solve this recurrence, we introduce the generating function

$$T(z,v) := \sum_{n \geq 1} \sum_{m \geq 1} T_{n,m} \frac{z^{n-1}}{(n-1)!} v^m. \tag{4.14}$$

We multiply (4.13) by $\frac{z^{n-2}}{(n-2)!}v^m$ and sum up for $n \geq 2$ and $m \geq 1$. The left-hand side of (4.13) then gives

$$\sum_{n \geq 2} \sum_{m \geq 1} T_{n,m} \frac{z^{n-2}}{(n-2)!} v^m = \sum_{n \geq 2} \sum_{m \geq 1} (n-1) T_{n,m} \frac{z^{n-2}}{(n-1)!} v^m = T_z(z,v),$$

and the right-hand side sums up to

$$\begin{aligned}
&\sum_{n\geq 2}\sum_{m\geq 1(2)} kT_{n-1,m-1}\frac{z^{n-2}}{(n-2)!}v^m + \sum_{n\geq 2}\sum_{m\geq 1}(k(n-2)+1)T_{n-1,m}\frac{z^{n-2}}{(n-2)!}v^m \\
&= kv\sum_{n\geq 1}\sum_{m\geq 1} T_{n,m}\frac{z^{n-1}}{(n-1)!}v^m + \sum_{n\geq 1}\sum_{m\geq 1}(k(n-1)+1)T_{n,m}\frac{z^{n-1}}{(n-1)!}v^m \\
&= kvT(z,v) + kzT_z(z,v) + T(z,v).
\end{aligned}$$

Furthermore, we have $T(0,v) = \sum_{m\geq 1} T_{1,m}v^m = v^{k+1}$. Hence, $T(z,v)$ satisfies

$$(kz-1)T_z(z,v) + (kv+1)T(z,v) = 0, \qquad T(0,v) = v^{k+1},$$

or equivalently

$$T_z(z,v) = \frac{kv+1}{1-kz}T(z,v), \qquad T(0,v) = v^{k+1},$$

This is an ordinary differential equation with respect to z which can be solved easily, and hence we find that $T(z,v)$ is given by

$$T(z,v) = v^{k+1}\exp\left(\int_0^z \frac{kv+1}{1-k\zeta}d\zeta\right) = \frac{v^{k+1}}{(1-kz)^{v+\frac{1}{k}}}. \tag{4.15}$$

In order to extract coefficients from (4.15) we use the expansion (2.4) for $\frac{1}{(1-z)^v}$ which involves the (unsigned) Stirling numbers of the first kind. With this we obtain

$$\begin{aligned}
\frac{T_{n,m+k}}{(n-1)!} &= [z^{n-1}v^{m+k}]T(z,v) = [z^{n-1}v^{m-1}]\frac{1}{(1-kz)^{v+\frac{1}{k}}} \\
&= k^{n-1}[z^{n-1}]\left(\frac{1}{(1-z)^{\frac{1}{k}}}\sum_{i\geq m-1}\begin{bmatrix} i \\ m-1\end{bmatrix}\frac{z^i}{i!}\right) \\
&= k^{n-1}\sum_{i\geq m-1}\frac{\begin{bmatrix} i \\ m-1\end{bmatrix}}{i!}[z^{n-i-1}]\frac{1}{(1-z)^{\frac{1}{k}}} \\
&= k^{n-1}\sum_{i=m-1}^{n-1}\frac{\begin{bmatrix} i \\ m-1\end{bmatrix}}{i!}\binom{n-i-2+\frac{1}{k}}{n-i-1}.
\end{aligned}$$

Hence, using the relation $\mathbb{P}\{A_n = m+k\} = \frac{T_{n,m+k}}{T_n}$ and the formula for T_n given in Theorem 4.6.1, we obtain the exact probabilities $\mathbb{P}\{A_n = m+k\}$ for the family

$\mathcal{T}^{[u]}$:

$$\begin{aligned}\mathbb{P}\{A_n = m+k\} &= \frac{(n-1)!k^{n-1}}{n!k^n\binom{n-1+\frac{1}{k}}{n}} \sum_{i=m-1}^{n-1} \frac{\left[{i \atop m-1}\right]}{i!} \binom{n-i-2+\frac{1}{k}}{n-i-1} \\ &= \frac{1}{\binom{n-1+\frac{1}{k}}{n-1}} \sum_{i=m}^{n} \frac{\left[{i-1 \atop m-1}\right]}{(i-1)!} \binom{n-i-1+\frac{1}{k}}{n-i}.\end{aligned}$$

Family $\mathcal{T}^{[o]}$ (ordered increasing k-trees): For ordered increasing k-trees, equation (4.11) becomes

$$\mathbb{P}\{A_n = m\} = \\ \frac{k}{(k+1)(n-1)+1}\mathbb{P}\{A_{n-1} = m-1\} + \frac{(k+1)(n-2)+2}{(k+1)(n-1)+1}\mathbb{P}\{A_{n-1} = m\},$$

for $n > 1$ and $m \geq 1$. Multiplication by T_n leads to the recurrence

$$T_{n,m} = kT_{n-1,m-1} + ((k+1)(n-2)+2)T_{n-1,m}, \tag{4.16}$$

for $n > 1$ and $m \geq 1$. Moreover, one has the initial values $T_{1,k+1} = 1$, and $T_{1,m} = 0$ for $m \neq k+1$ (since there is exactly 1 ordered increasing k-tree of size n, and in this k-tree node 1 has $k+1$ ancestors). Like in the previous case, we solve this equation using the generating function $T(z,v)$ defined by (4.14). By multiplying (4.16) with $\frac{z^{n-2}}{(n-2)!}v^m$ and summing up, we find that in this case $T(z,v)$ satisfies the following ordinary differential equation with respect to z:

$$((k+1)z-1)T_z(z,v) + (kv+2)T(z,v) = 0, \qquad T(0,v) = v^{k+1}.$$

Of course, this equation can be solved as easily as in the previous case, and we find the solution

$$T(z,v) = \frac{v^{k+1}}{(1-(k+1)z)^{\frac{kv+2}{k+1}}}. \tag{4.17}$$

Again we extract coefficients using (2.4), and obtain:

$$\begin{aligned}
\frac{T_{n,m+k}}{(n-1)!} &= [z^{n-1}v^{m+k}]T(z,v) = [z^{n-1}v^{m-1}]\frac{1}{(1-(k+1)z)^{\frac{kv+2}{k+1}}} \\
&= \frac{(k+1)^{n-1}k^{m-1}}{(k+1)^{m-1}}[z^{n-1}]\left(\frac{1}{(1-z)^{\frac{2}{k+1}}}\sum_{i\ge m-1}\begin{bmatrix} i \\ m-1\end{bmatrix}\frac{z^i}{i!}\right) \\
&= (k+1)^{n-m}k^{m-1}\sum_{i\ge m-1}\frac{\begin{bmatrix} i \\ m-1\end{bmatrix}}{i!}[z^{n-i-1}]\frac{1}{(1-z)^{\frac{2}{k+1}}} \\
&= (k+1)^{n-m}k^{m-1}\sum_{i=m-1}^{n-1}\frac{\begin{bmatrix} i \\ m-1\end{bmatrix}}{i!}\binom{n-i-2+\frac{2}{k+1}}{n-i-1}.
\end{aligned}$$

From this, the formula for the probabilities $\mathbb{P}\{A_n = m+k\}$ for the family $\mathcal{T}^{[u]}$ as stated in Theorem 4.6.1 can easily be derived using Theorem 4.6.1.

Family $\mathcal{T}^{[d]}$ (d-ary increasing k-trees): In the case of d-ary increasing k-trees, equation (4.11) reads

$$\mathbb{P}\{A_n = m\} = \frac{kd}{(kd-1)(n-1)+d}\mathbb{P}\{A_{n-1} = m-1\} + \frac{(kd-1)(n-2)+d-1}{(kd-1)(n-1)+d}\mathbb{P}\{A_{n-1} = m\},$$

for $n > 1$ and $m \ge 1$. By multiplication with T_n, this leads for the numbers $T_{n,m}$, which count d-ary increasing k-trees of size n in which node n has m ancestors, to the recurrence

$$T_{n,m} = kdT_{n-1,m-1} + ((kd-1)(n-2)+d-1)T_{n-1,m},$$

for $n > 1$ and $m \ge 1$. Furthermore, one has the initial values $T_{1,k+1} = d$, and $T_{1,m} = 0$ for $m \ne k+1$ (since there are exactly d d-ary increasing k-trees of size 1, and node 1 has $k+1$ ancestors in each of them). Proceeding as in the previous two cases, we get the differential equation

$$((kd-1)z-1)T_z(z,v) + (kdv+d-1)T(z,v) = 0, \qquad T(0,v) = dv^{k+1},$$

which yields the solution

$$T(z,v) = \frac{dv^{k+1}}{(1-(kd-1)z)^{\frac{kdv+d-1}{kd-1}}}. \tag{4.18}$$

By extracting coefficients as in the previous two cases, one obtains the formula for $\mathbb{P}\{A_n = m+k\}$ given in Theorem 4.6.1 for the family $\mathcal{T}^{[d]}$.

4.6.2.2 Derivation of the limiting distributions results

We derive the limiting distributions for the number of ancestors directly from the exact results.

Family $\mathcal{T}^{[u]}$ (unordered increasing k-trees): From (4.15) we can compute the probability generating function $p_n(v) := \sum_{m\geq 0} \mathbb{P}\{A_n = m\}\, v^m$ of A_n as follows:

$$\begin{aligned} p_n(v) &= \frac{(n-1)!}{T_n}[z^{n-1}]T(z,v) = \frac{(n-1)!}{n!k^n\binom{n-1+\frac{1}{k}}{n}} v^{k+1} k^{n-1} \binom{n-2+v+\frac{1}{k}}{n-1} \\ &= v^{k+1}\frac{\binom{n-2+v+\frac{1}{k}}{n-1}}{\binom{n-1+\frac{1}{k}}{n-1}} = \frac{v^{k+1}\Gamma\big(n-1+v+\frac{1}{k}\big)\,\Gamma\big(1+\frac{1}{k}\big)}{\Gamma\big(v+\frac{1}{k}\big)\,\Gamma\big(n+\frac{1}{k}\big)}. \end{aligned} \tag{4.19}$$

From this we get, by an application of Lemma 2.2.5, the asymptotic expansion

$$p_n(v) = \frac{v^{k+1}\Gamma\big(1+\frac{1}{k}\big)}{\Gamma\big(v+\frac{1}{k}\big)} n^{v-1}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right), \quad \text{for } n\to\infty, \tag{4.20}$$

which holds uniformly for v in a complex neighbourhood of 1. Clearly, (4.20) can be written in the form

$$p_n(v) = A(v)B(v)^{\beta_n}\left(1+\mathcal{O}\big(\kappa_n^{-1}\big)\right), \quad \text{for } n\to\infty, \tag{4.21}$$

with $B(v) = \exp(v-1)$, $\beta_n = \log n$ and $\kappa_n = n$. This is exactly a situation, where Hwang's quasi-powers theorem (Theorem 2.2.3) can be applied to show that the centered and normalized random variable $\frac{A_n - \mathbb{E}(A_n)}{\sqrt{\mathbb{V}(A_n)}}$ converges in distribution to the standard normal distribution (with rate of convergence $\mathcal{O}\Big(\frac{1}{\sqrt{\log n}}\Big)$), and that the mean $\mathbb{E}(A_n)$ and the variance $\mathbb{V}(A_n)$ of A_n satisfy

$$\begin{aligned} \mathbb{E}(A_n) &= \beta_n B'(1) + \mathcal{O}(1) = \log n + \mathcal{O}(1)\,, \\ \mathbb{V}(A_n) &= \beta_n(B''(1) + B'(1) - (B'(1))^2) + \mathcal{O}(1) = \log n + \mathcal{O}(1)\,. \end{aligned}$$

This completes the proof of the corresponding result in Theorem 4.6.2.

Family $\mathcal{T}^{[o]}$ (ordered increasing k-trees): In analogy to the computations for the family $\mathcal{T}^{[u]}$, we extract the probability generating function $p_n(v)$ of A_n from

(4.17). We find that for ordered increasing k-trees $p_n(v)$ is given by

$$\begin{aligned}
p_n(v) &= \frac{(n-1)!}{T_n}[z^{n-1}]T(z,v) \\
&= \frac{(n-1)!}{n!(k+1)^n\binom{n-1+\frac{1}{k+1}}{n}} v^{k+1}(k+1)^{n-1}\binom{n-2+\frac{kv+2}{k+1}}{n-1} \\
&= v^{k+1}\frac{\binom{n-2+\frac{kv+2}{k+1}}{n-1}}{\binom{n-1+\frac{1}{k+1}}{n-1}} = \frac{v^{k+1}\Gamma\left(n-1+\frac{kv+2}{k+1}\right)\Gamma\left(1+\frac{1}{k+1}\right)}{\Gamma\left(\frac{kv+2}{k+1}\right)\Gamma\left(n+\frac{1}{k+1}\right)} \\
&= \frac{v^{k+1}\Gamma\left(1+\frac{1}{k+1}\right)}{\Gamma\left(\frac{kv+2}{k+1}\right)} n^{\frac{k}{k+1}(v-1)}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right), \quad \text{for } n\to\infty,
\end{aligned} \tag{4.22}$$

where we have again used Lemma 2.2.5 in the last step. Hence, also in this case $p_n(v)$ satisfies an equation of the form (4.21), which holds uniformly for v in a neighbourhood of 1, where now $B(v) = \exp\left(\frac{k}{k+1}(v-1)\right)$, $\beta_n = \log n$ and $\kappa_n = n$. Thus, by another application of Hwang's quasi-powers theorem, we get

$$\mathbb{E}(A_n) = \beta_n B'(1) + \mathcal{O}(1) = \frac{k}{k+1}\log n + \mathcal{O}(1)\,,$$

$$\mathbb{V}(A_n) = \beta_n(B''(1) + B'(1) - (B'(1))^2) + \mathcal{O}(1) = \frac{k}{k+1}\log n + \mathcal{O}(1)\,,$$

and the claimed convergence result for the centered and normalized random variable $\frac{A_n-\mathbb{E}(A_n)}{\sqrt{\mathbb{V}(A_n)}}$ for family $\mathcal{T}^{[o]}$.

Family $\mathcal{T}^{[d]}$ (d-ary increasing k-trees): Analogously to the other two cases we compute $p_n(v)$ from (4.18):

$$\begin{aligned}
p_n(v) &= \frac{(n-1)!}{T_n}[z^{n-1}]T(z,v) \\
&= \frac{(n-1)!}{n!(kd-1)^n\binom{n-1+\frac{d}{kd-1}}{n}} dv^{k+1}(kd-1)^{n-1}\binom{n-2+\frac{kdv+d-1}{kd-1}}{n-1} \\
&= v^{k+1}\frac{\binom{n-2+\frac{kdv+d-1}{kd-1}}{n-1}}{\binom{n-1+\frac{d}{kd-1}}{n-1}} = \frac{v^{k+1}\Gamma\left(n-1+\frac{kdv+d-1}{kd-1}\right)\Gamma\left(1+\frac{d}{kd-1}\right)}{\Gamma\left(\frac{kdv+d-1}{kd-1}\right)\Gamma\left(n+\frac{d}{kd-1}\right)} \\
&= \frac{v^{k+1}\Gamma\left(1+\frac{d}{kd-1}\right)}{\Gamma\left(\frac{kdv+d-1}{kd-1}\right)} n^{\frac{kd}{kd-1}(v-1)}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right), \quad \text{for } n\to\infty.
\end{aligned} \tag{4.23}$$

Hence, $p_n(v)$ is of the form (4.21) with $B(v) = \exp\left(\frac{kd}{kd-1}(v-1)\right)$, $\beta_n = \log n$ and $\kappa_n = n$ in this case. Thus, another application of the quasi-powers theorem gives the result in Theorem 4.6.2 for the family $\mathcal{T}^{[d]}$.

4.6.2.3 Derivation of the results for a randomly selected node

To obtain the results for $\overline{A}_n$, i.e. the number of ancestors of a randomly selected node in a random increasing k-tree of size n, as stated in Theorem 4.6.3, we use

$$\begin{aligned}\mathbb{P}\{\overline{A}_n = m\} &= \sum_{j=1}^{n} \mathbb{P}\{\overline{A}_n = m \mid \text{node } j \text{ is chosen}\} \, \mathbb{P}\{\text{node } j \text{ is chosen}\} \\ &= \frac{1}{n} \sum_{j=1}^{n} \mathbb{P}\{A_j = m\},\end{aligned}$$

and the exact results for A_j.

Family $\mathcal{T}^{[u]}$ (unordered increasing k-trees): We consider the probability generating function of $\overline{A}_n$,

$$\overline{p}_n(v) := \sum_{m \ge 0} \mathbb{P}\{\overline{A}_n = m\} \, v^m, \tag{4.24}$$

and obtain, by using (4.19):

$$\overline{p}_n(v) = \frac{1}{n} \sum_{j=1}^{n} p_j(v) = \frac{v^{k+1}}{n} \sum_{j=1}^{n} \frac{\binom{j-2+v+\frac{1}{k}}{j-1}}{\binom{j-1+\frac{1}{k}}{j-1}}. \tag{4.25}$$

To carry out the summation we consider the expression

$$\sum_{j=1}^{n} \frac{\binom{j-1+\alpha}{j-1}}{\binom{j-1+\beta}{j-1}} = \frac{1}{\binom{\alpha}{\alpha-\beta}} \sum_{j=0}^{n-1} \binom{j+\alpha}{\alpha-\beta},$$

for real α, β, where we assume that $-\alpha$, $-\beta$, and $\beta - \alpha \notin \mathbb{N}$. Using the formula

$$\sum_{j=0}^{n} \binom{j+a}{b} = \binom{n+a+1}{b+1} - \binom{a}{b+1}, \tag{4.26}$$

which is a direct consequence of the representation $\binom{j+a}{b} = \binom{j+1+a}{b+1} - \binom{j+a}{b+1}$, which yields a telescoping sum, one easily obtains

$$\begin{aligned}\sum_{j=1}^{n} \frac{\binom{j-1+\alpha}{j-1}}{\binom{j-1+\beta}{j-1}} &= \frac{1}{\binom{\alpha}{\alpha-\beta}} \left(\binom{n+\alpha}{\alpha-\beta+1} - \binom{\alpha}{\alpha-\beta+1} \right) \\ &= \frac{\beta}{\alpha-\beta+1} \left(\frac{\binom{n+\alpha}{n}}{\binom{n+\beta-1}{n}} - 1 \right).\end{aligned} \tag{4.27}$$

A direct application of (4.27) to (4.25) (via $\alpha = v - 1 + \frac{1}{k}$, $\beta = \frac{1}{k}$) gives the following exact formula for $\overline{p}_n(v)$:

$$\overline{p}_n(v) = \frac{v^k}{kn}\left(\frac{\binom{n+v-1+\frac{1}{k}}{n}}{\binom{n-1+\frac{1}{k}}{n}} - 1\right).$$

From this one gets, by applying Lemma 2.2.5, the following asymptotic formula (which holds, for arbitrarily small $\epsilon > 0$, uniformly in a neighbourhood of $v = 1$):

$$\begin{aligned}\overline{p}_n(v) &= \frac{v^k\Gamma\big(1+\frac{1}{k}\big)}{\Gamma\big(v+\frac{1}{k}\big)} n^{v-1}\left(1+\mathcal{O}\big(n^{-1+\epsilon}\big)\right)\\ &= A(v)B(v)^{\beta_n}\left(1+\mathcal{O}\big(\kappa_n^{-1}\big)\right),\end{aligned}$$

with $B(v) = \exp(v-1)$, $\beta_n = \log n$ and $\kappa_n = n^{1-\epsilon}$. Hwang's quasi-powers theorem now shows the corresponding part of Theorem 4.6.3.

Family $\mathcal{T}^{[o]}$ (ordered increasing k-trees): We compute the probability generating function $\overline{p}_n(v)$ defined by (4.24) from the exact expressions for the functions $p_j(v)$ which we found in (4.22):

$$\overline{p}_n(v) = \frac{1}{n}\sum_{j=1}^{n} p_j(v) = \frac{v^{k+1}}{n}\sum_{j=1}^{n}\frac{\binom{j-2+\frac{kv+2}{k+1}}{j-1}}{\binom{j-1+\frac{1}{k+1}}{j-1}} = \frac{v^{k+1}}{n(kv+1)}\left(\frac{\binom{n-1+\frac{kv+2}{k+1}}{n}}{\binom{n-1+\frac{1}{k+1}}{n}} - 1\right).$$

Note that we have used (4.27) with $\alpha = \frac{kv+2}{k+1} - 1$ and $\beta = \frac{1}{k+1}$ in the last step. From this one gets, again by an application of Lemma 2.2.5, the following asymptotic formula, which holds (for arbitrary $\epsilon > 0$) uniformly in a neighbourhood of $v = 1$:

$$\overline{p}_n(v) = \frac{v^{k+1}\Gamma\big(\frac{1}{k+1}\big)}{(kv+1)\Gamma\big(\frac{kv+2}{k+1}\big)} n^{\frac{k}{k+1}(v-1)}\left(1+\mathcal{O}\big(n^{-1+\epsilon}\big)\right).$$

Hence we can apply Theorem 2.2.3 with $B(v) = \exp\big(\frac{k}{k+1}(v-1)\big)$, $\beta_n = \log n$ and $\kappa_n = n^{1-\epsilon}$, which proves the result in Theorem 4.6.3 for the family $\mathcal{T}^{[o]}$.

Family $\mathcal{T}^{[d]}$ (d-ary increasing k-trees): Analogously to the previous two cases, we compute $\overline{p}_n(v)$ using (4.23). We carry out the summation using (4.27) with

$\alpha = \frac{kdv+d-1}{kd-1} - 1$ and $\beta = \frac{d}{kd-1}$, and then apply Lemma 2.2.5, in order to obtain

$$\begin{aligned}
\overline{p}_n(v) &= \frac{1}{n}\sum_{j=1}^{n} p_j(v) = \frac{v^{k+1}}{n}\sum_{j=1}^{n} \frac{\binom{n-2+\frac{kdv+d-1}{kd-1}}{n-1}}{\binom{n-1+\frac{d}{kd-1}}{n-1}} \\
&= \frac{dv^{k+1}}{n(kdv-1)}\left(\frac{\binom{n-1+\frac{kdv+d-1}{kd-1}}{n}}{\binom{n-1+\frac{d}{kd-1}}{n}} - 1\right) \\
&= \frac{dv^{k+1}\Gamma\big(\frac{d}{kd-1}\big)}{(kdv-1)\Gamma\big(\frac{kdv+d-1}{kd-1}\big)} n^{\frac{kd}{kd-1}(v-1)} \left(1+\mathcal{O}\big(n^{-1+\epsilon}\big)\right).
\end{aligned}$$

Hence, Hwang's quasi-powers theorem is applicable with $B(v) = \exp\big(\frac{kd}{kd-1}(v-1)\big)$, $\beta_n = \log n$ and $\kappa_n = n^{1-\epsilon}$. This finishes the proof of Theorem 4.6.3.

4.7 Descendants

4.7.1 Results

Theorem 4.7.1. *The random variable $D_{n,j}$, which counts the number of descendants of node j in a random increasing k-tree of size n, has the following exact distribution:*

$$\mathbb{P}\{D_{n,j} = m\} = \begin{cases} \frac{\binom{n-m-1+\frac{1}{k}}{n-m-j+1}}{\binom{n-1+\frac{1}{k}}{n-j}}, & \textit{family } \mathcal{T}^{[u]}, \\ \frac{\binom{m-1-\frac{1}{k+1}}{m-1}\binom{n-m-1+\frac{2}{k+1}}{n-m-j+1}}{\binom{n-1+\frac{1}{k+1}}{n-j}}, & \textit{family } \mathcal{T}^{[o]}, \\ \frac{\binom{m-1+\frac{1}{kd-1}}{m-1}\binom{n-m-1+\frac{d-1}{kd-1}}{n-m-j+1}}{\binom{n-1+\frac{d}{kd-1}}{n-j}}, & \textit{family } \mathcal{T}^{[d]}, \end{cases}$$

for $n \geq j \geq 1$ and $m \geq 1$.

Theorem 4.7.2. *The limiting distribution behaviour of $D_{n,j}$ is, for $n \to \infty$ and depending on the growth of j, characterized as follows:*

- *The region for j fixed: The normalized random variable $\frac{D_{n,j}}{n}$ is asymptotically* Beta(α, β) *distributed with parameters*

$$\alpha = \begin{cases} 1, & \textit{family } \mathcal{T}^{[u]}, \\ \frac{k}{k+1}, & \textit{family } \mathcal{T}^{[o]}, \\ \frac{kd}{kd-1}, & \textit{family } \mathcal{T}^{[d]}, \end{cases} \quad \textit{and} \quad \beta = \begin{cases} j-1+\frac{1}{k}, & \textit{family } \mathcal{T}^{[u]}, \\ j-1+\frac{2}{k+1}, & \textit{family } \mathcal{T}^{[o]}, \\ j-1+\frac{d-1}{kd-1}, & \textit{family } \mathcal{T}^{[d]}, \end{cases}$$

i.e. $\frac{D_{n,j}}{n} \xrightarrow{(d)} D_j$, *where the moments of* D_j *are, for* $s \geq 0$, *given as follows:*

$$\mathbb{E}\big(D_j^s\big) = \begin{cases} s!\frac{\Gamma\left(j+\frac{1}{k}\right)}{\Gamma\left(s+j+\frac{1}{k}\right)}, & \textit{family } \mathcal{T}^{[u]}, \\ \frac{\Gamma\left(s+\frac{k}{k+1}\right)\Gamma\left(j+\frac{1}{k+1}\right)}{\Gamma\left(\frac{k}{k+1}\right)\Gamma\left(s+j+\frac{1}{k+1}\right)}, & \textit{family } \mathcal{T}^{[o]}, \\ \frac{\Gamma\left(s+\frac{kd}{kd-1}\right)\Gamma\left(j+\frac{d}{kd-1}\right)}{\Gamma\left(\frac{kd}{kd-1}\right)\Gamma\left(s+j+\frac{d}{kd-1}\right)}, & \textit{family } \mathcal{T}^{[d]}. \end{cases}$$

- *The region for* j *small:* $j \to \infty$ *such that* $j = o(n)$*: The normalized random variable* $\frac{j}{n}D_{n,j}$ *is asymptotically* $\mathrm{Gamma}(\alpha, \theta)$ *distributed with parameters*

$$\alpha = \begin{cases} 1, & \textit{family } \mathcal{T}^{[u]}, \\ \frac{k}{k+1}, & \textit{family } \mathcal{T}^{[o]}, \\ \frac{kd}{kd-1}, & \textit{family } \mathcal{T}^{[d]}, \end{cases} \qquad \textit{and} \qquad \theta = 1,$$

(for the family $\mathcal{T}^{[u]}$ *this is the* $\mathrm{Exp}(1)$ *distribution), i.e.* $\frac{j}{n}D_{n,j} \xrightarrow{(d)} D$, *where the moments of* D *are, for* $s \geq 0$, *given as follows:*

$$\mathbb{E}(D^s) = \begin{cases} s!, & \textit{family } \mathcal{T}^{[u]}, \\ \frac{\Gamma\left(s+\frac{k}{k+1}\right)}{\Gamma\left(\frac{k}{k+1}\right)}, & \textit{family } \mathcal{T}^{[o]}, \\ \frac{\Gamma\left(s+\frac{kd}{kd-1}\right)}{\Gamma\left(\frac{kd}{kd-1}\right)}, & \textit{family } \mathcal{T}^{[d]}. \end{cases}$$

- *The central region for* j*:* $j \to \infty$ *such that* $j \sim \rho n$, *with* $0 < \rho < 1$*: The shifted random variable* $D_{n,j} - 1$ *is asymptotically negative binomial distributed* $\mathrm{NegBin}(r, p)$ *with parameters*

$$r = \begin{cases} 1, & \textit{family } \mathcal{T}^{[u]}, \\ \frac{k}{k+1}, & \textit{family } \mathcal{T}^{[o]}, \\ \frac{kd}{kd-1}, & \textit{family } \mathcal{T}^{[d]}, \end{cases} \qquad \textit{and} \qquad p = \rho,$$

(for the family $\mathcal{T}^{[u]}$ *this is the* $\mathrm{Geom}(\rho)$ *distribution), i.e.* $D_{n,j} - 1 \xrightarrow{(d)} D_\rho$, *where the probability mass function of* D_ρ *is given by*

$$\mathbb{P}\{D_\rho = m\} = \begin{cases} \rho(1-\rho)^m, & \textit{family } \mathcal{T}^{[u]}, \\ \binom{m-\frac{1}{k+1}}{m}\rho^{\frac{k}{k+1}}(1-\rho)^m, & \textit{family } \mathcal{T}^{[o]}, \\ \binom{m+\frac{1}{kd-1}}{m}\rho^{\frac{kd}{kd-1}}(1-\rho)^m, & \textit{family } \mathcal{T}^{[d]}, \end{cases}$$

for $m \in \mathbb{N}_0$.

- *The region for j large: $j \to \infty$ such that $n - j = o(n)$: It holds that $\mathbb{P}\{D_{n,j} = 1\} \to 1$.*

Theorem 4.7.3. *The random variable $\overline{D}_n$, which counts the number of descendants of a randomly selected inserted node in a random increasing k-tree of size n, has the following exact distribution:*

$$\mathbb{P}\{\overline{D}_n = m\} = \begin{cases} \frac{(kn-km+m+1)\binom{n-m-1+\frac{1}{k}}{n-m}}{m(m+1)\binom{n-1+\frac{1}{k}}{n-1}}, & \text{family } \mathcal{T}^{[u]}, \\ \frac{k(n+\frac{1}{k+1})}{(k+1)(m+\frac{k}{k+1})(m-\frac{1}{k+1})n} \\ \quad - \frac{\binom{m-1-\frac{1}{k+1}}{m-1}}{n\binom{n-1+\frac{1}{k+1}}{n}} \sum_{\ell=0}^{m-1} \frac{\binom{m-1}{\ell}(-1)^\ell}{(k+1)(\ell+1)+k} \binom{n-\ell-2+\frac{2}{k+1}}{n}, & \text{family } \mathcal{T}^{[o]}, \\ \frac{kd(n+\frac{d}{kd-1})}{(kd-1)(m+\frac{kd}{kd-1})(m+\frac{1}{kd-1})n} \\ \quad - \frac{d\binom{m-1+\frac{1}{kd-1}}{m-1}}{n\binom{n-1+\frac{d}{kd-1}}{n}} \sum_{\ell=0}^{m-1} \frac{\binom{m-1}{\ell}(-1)^\ell}{(kd-1)(\ell+1)+kd} \binom{n-\ell-2+\frac{d-1}{kd-1}}{n}, & \text{family } \mathcal{T}^{[d]}, \end{cases}$$

for $n \geq m \geq 1$.

For $n \to \infty$, $\overline{D}_n$ converges in distribution to a discrete random variable $\overline{D}$, i.e. $\overline{D}_n \xrightarrow{(d)} \overline{D}$, where the distribution of $\overline{D}$ is given as follows, with $m \geq 1$:

$$\mathbb{P}\{\overline{D} = m\} = \begin{cases} \frac{1}{m(m+1)}, & \text{family } \mathcal{T}^{[u]}, \\ \frac{k}{(k+1)(m+\frac{k}{k+1})(m-\frac{1}{k+1})}, & \text{family } \mathcal{T}^{[o]}, \\ \frac{kd}{(kd-1)(m+\frac{kd}{kd-1})(m+\frac{1}{kd-1})}, & \text{family } \mathcal{T}^{[d]}. \end{cases}$$

4.7.2 Proofs of the results

4.7.2.1 Derivation of the exact distribution

Family $\mathcal{T}^{[u]}$ (unordered increasing k-trees): In order to get a suitable description of the random variable $D_{n,j}$ we consider the graph evolution process of k-trees as described in Proposition 4.4.1:

Clearly, it holds that in any unordered increasing k-tree T of size j the node with label j has exactly one descendant (namely itself). Furthermore, there are exactly k possible ways of adding a new node x to T such that the number of descendants of j increases, namely by attaching x to one of the k-cliques which contain j.

Now, we use the following observation: each new descendant x of j increases the number of possible locations for subsequent descendants of j by exactly k

(since k new k-cliques are created when x is added to the k-tree). Thus, if in an unordered increasing k-tree of size n node $j \geq 1$ has m descendants then there are exactly mk possible locations at which a new descendant of j can be attached, whereas attaching a node to any of the $k(n-m)+1$ other locations will keep the number of descendants of j unchanged.

Hence, if we count by $T_{n,j,m} := T_n \mathbb{P}\{D_{n,j} = m\}$ the number of unordered increasing k-trees of size n in which node j has m descendants, we immediately get, by distinguishing whether node n is a descendant of j or not, the following recurrence:

$$T_{n,j,m} = (k(n-m-1)+1)\, T_{n-1,j,m} + k(m-1)T_{n-1,j,m-1}, \tag{4.28}$$

for $n > j \geq 1$ and $m \geq 1$. Moreover, one clearly has the boundary values $T_{j,j,1} = T_j$, for $j \geq 1$, and $T_{j,j,m} = 0$, for $m > 1$. In order to solve this recurrence, we introduce the generating function

$$T^{[j]}(z,v) := \sum_{n \geq j} \sum_{m \geq 1} T_{n,j,m} \frac{z^{n-j}}{(n-j)!} v^m. \tag{4.29}$$

Multiplying (4.28) by $\frac{z^{n-j-1}}{(n-j-1)!} v^m$ and summing up for $n > j$ and $m \geq 1$ leads then to the following linear first order partial differential equation:

$$(kz-1)T_z^{[j]}(z,v) + kv(v-1)T_v^{[j]}(z,v) + (kj+1)T^{[j]}(z,v) = 0, \quad T^{[j]}(0,v) = T_j v. \tag{4.30}$$

This PDE can be solved by an application of the method of characteristics (see Appendix B). For this purpose, we consider the system of characteristic differential equations for (4.30), i.e.

$$\dot{z} = kz - 1, \quad \dot{v} = kv(v-1), \tag{4.31}$$

where we regard $z = z(t)$ and $v = v(t)$ as functions of a parameter t. A first integral of (4.31) can be found by considering the phase differential equation

$$\frac{dz}{dv} = \frac{kz-1}{kv(v-1)}.$$

This differential equation can be solved easily by separation of variables: From

$$\int \frac{dz}{kz-1} = \int \frac{dv}{kv(v-1)} = \frac{1}{k} \int -\frac{1}{v} + \frac{1}{v-1} dv,$$

one obtains

$$\frac{1}{k}\log|kz-1| = \frac{1}{k}\log\left|\frac{v-1}{v}\right| + const,$$

and thus

$$\frac{v(1-kz)}{1-v} = const.$$

Hence, a first integral $\zeta(z,v)$ of (4.31) is given by

$$\zeta(z,v) = \frac{v(1-kz)}{1-v}.$$

Now, in order to solve (4.30), we use the coordinate transform

$$\zeta = \zeta(z,v) = \frac{v(1-kz)}{1-v}, \quad \eta = \eta(z,v) = 1-kz,$$

or equivalently

$$z = z(\zeta,\eta) = \frac{1-\eta}{k}, \quad v = v(\zeta,\eta) = \frac{\zeta}{\zeta+\eta}.$$

By setting $\tilde{T}^{[j]}(\zeta,\eta) := T^{[j]}(z(\zeta,\eta), v(\zeta,\eta))$, the corresponding transformation of (4.30) is given as follows (note that the terms containing $\tilde{T}^{[j]}_\zeta$ cancel, compare Appendix B, and that $\eta_v = 0$):

$$\begin{aligned}
0 &= (kj+1)T^{[j]}(z,v) + (kz-1)T^{[j]}_z(z,v) + kv(v-1)T^{[j]}_v(z,v) \\
&= (kj+1)\tilde{T}^{[j]}(\zeta,\eta) + (kz-1)\left(\tilde{T}^{[j]}_\zeta(\zeta,\eta)\frac{\partial\zeta}{\partial z} + \tilde{T}^{[j]}_\eta(\zeta,\eta)\frac{\partial\eta}{\partial z}\right) \\
&\quad + kv(v-1)\left(\tilde{T}^{[j]}_\zeta(\zeta,\eta)\frac{\partial\zeta}{\partial v} + \tilde{T}^{[j]}_\eta(\zeta,\eta)\frac{\partial\eta}{\partial v}\right) \\
&= (kj+1)\tilde{T}^{[j]}(\zeta,\eta) + k\eta\tilde{T}^{[j]}_\eta(\zeta,\eta).
\end{aligned}$$

Hence, we obtain the following (surprisingly simple) ordinary differential equation with respect to η:

$$\tilde{T}^{[j]}_\eta(\zeta,\eta) = -\frac{kj+1}{k\eta}\tilde{T}^{[j]}(\zeta,\eta).$$

This immediately yields

$$\tilde{T}^{[j]}(\zeta,\eta) = C(\zeta)\eta^{-j-\frac{1}{k}},$$

where C is a continuous function, and thus the general solution of (4.30) is given by

$$T^{[j]}(z,v) = \frac{C\left(\frac{v(1-kz)}{1-v}\right)}{(1-kz)^{j+\frac{1}{k}}}. \tag{4.32}$$

It remains to adapt C to the boundary condition given in (4.30). By setting $z=0$ in (4.32), we get

$$T_j v = T^{[j]}(0,v) = C\left(\frac{v}{1-v}\right),$$

and by substituting $v = \frac{x}{1+x}$ we find

$$C(x) = \frac{T_j x}{1+x}.$$

Hence we finally obtain the solution:

$$T^{[j]}(z,v) = \frac{T_j \frac{v(1-kz)}{1-v}}{\left(1+\frac{v(1-kz)}{1-v}\right)(1-kz)^{j+\frac{1}{k}}} = \frac{T_j v}{(1-vkz)(1-kz)^{j-1+\frac{1}{k}}}. \tag{4.33}$$

From this, we can easily compute the numbers $T_{n,j,m}$ by extracting the proper coefficients:

$$\begin{aligned}\frac{T_{n,j,m}}{(n-j)!} &= [z^{n-j}v^m]T^{[j]}(z,v) = T_j[z^{n-j}]\frac{(kz)^{m-1}}{(1-kz)^{j-1+\frac{1}{k}}}\\ &= T_j k^{n-j}[z^{n-m-j+1}]\frac{1}{(1-z)^{j-1+\frac{1}{k}}} = T_j k^{n-j}\binom{n-m-1+\frac{1}{k}}{n-m-j+1}.\end{aligned}$$

Thus, using the relation $\mathbb{P}\{D_{n,j}=m\} = \frac{T_{n,j,m}}{T_n}$ and the formula for T_n from Proposition 4.3.1, one obtains the desired formula for the probabilities $\mathbb{P}\{D_{n,j}=m\}$:

$$\begin{aligned}\mathbb{P}\{D_{n,j}=m\} &= \frac{T_j k^{n-j}(n-j)!}{T_n}\binom{n-m-1+\frac{1}{k}}{n-m-j+1}\\ &= \frac{j!k^j\binom{j-1+\frac{1}{k}}{j}k^{n-j}(n-j)!}{n!k^n\binom{n-1+\frac{1}{k}}{n}}\binom{n-m-1+\frac{1}{k}}{n-m-j+1}\\ &= \frac{\Gamma\left(j+\frac{1}{k}\right)(n-j)!}{\Gamma\left(n+\frac{1}{k}\right)}\binom{n-m-1+\frac{1}{k}}{n-m-j+1}\\ &= \frac{\binom{n-m-1+\frac{1}{k}}{n-m-j+1}}{\binom{n-1+\frac{1}{k}}{n-j}}.\end{aligned}$$

Family $\mathcal{T}^{[o]}$ (ordered increasing k-trees): Let now $T_{n,j,m}$ denote the number of ordered increasing k-trees of size n in which node j has m descendants. In order to find a recurrence for $T_{n,j,m}$ we can use an argumentation analogous to the one for the previously considered family $\mathcal{T}^{[u]}$:

In any ordered increasing k-tree of size j node j has exactly one descendant, and there are k possible locations at which a new descendant can be attached. Furthermore, each new descendant increases the possible locations for subsequent descendants by $k+1$. This is because when x is added to the k-tree then k new k-cliques are created, and, apart from that, the k-clique to which x is attached gets one additional spot in the linear order of its children. Consequently, in each increasing k-tree of size n in which node j has m descendants, there are $(k+1)m-1$ locations at which a new descendant of j can be attached, whereas attaching a new node at any of the remaining $(k+1)(n-m)+2$ locations doesn't affect the number of descendants of node j. These considerations lead for the numbers $T_{n,j,m}$ to the recurrence relation

$$T_{n,j,m} = ((k+1)(n-m-1)+2)\, T_{n-1,j,m} + ((k+1)(m-1)-1)\, T_{n-1,j,m-1},$$

for $n > j \geq 1$ and $m \geq 1$, with initial values $T_{j,j,1} = T_j$, for $j \geq 1$, and $T_{j,j,m} = 0$, for $m > 1$. Defining the generating function $T^{[j]}(z,v)$ as in (4.29) and proceeding as for the family $\mathcal{T}^{[u]}$, we obtain the PDE

$$\begin{aligned}((k+1)z-1)T_z^{[j]}(z,v) &+ (k+1)v(v-1)T_v^{[j]}(z,v) \\ &+ ((k+1)j+2-v)T^{[j]}(z,v) = 0, \quad T^{[j]}(0,v) = T_j v.\end{aligned}$$

This PDE can be solved as well by using the method of characteristics. Since the computations are completely analogous to the ones for the family $\mathcal{T}^{[u]}$, we confine ourselves to stating the intermediary results:

A first integral of the system of characteristic equations $\dot{z} = (k+1)z - 1$, $\dot{v} = (k+1)v(v-1)$, is given by

$$\zeta(z,v) = \frac{v(1-(k+1)z)}{1-v}.$$

Using the coordinate transform $\zeta = \zeta(z,v)$, $\eta = \eta(z,v) = 1-(k+1)z$ (equivalently, $z(\zeta,\eta) = \frac{1-\eta}{k+1}$, $v(\zeta,\eta) = \frac{\zeta}{\zeta+\eta}$), one gets that the function $\tilde{T}^{[j]}(\zeta,\eta) := T^{[j]}(z(\zeta,\eta), v(\zeta,\eta))$ satisfies the following ordinary differential equation:

$$\tilde{T}_\eta^{[j]}(\zeta,\eta) = \left(\frac{\zeta}{(k+1)\eta(\zeta+\eta)} - \frac{(k+1)j+2}{(k+1)\eta}\right)\tilde{T}^{[j]}(\zeta,\eta).$$

Solving this equation and transforming it back to (z,v)-coordinates gives

$$T^{[j]}(z,v) = \frac{(1-v)^{\frac{1}{k+1}} C\left(\frac{v(1-(k+1)z)}{1-v}\right)}{(1-(k+1)z)^{j+\frac{2}{k+1}}}, \tag{4.34}$$

where C is a continuous function. C can be determined from the boundary condition $T^{[j]}(0,v) = T_j v$ by setting $z=0$ in (4.34). One gets

$$C(x) = \frac{T_j x}{(1+x)^{\frac{k}{k+1}}},$$

and this finally yields the solution

$$T^{[j]}(z,v) = \frac{T_j v}{(1-(k+1)vz)^{\frac{k}{k+1}}(1-(k+1)z)^{j-1+\frac{2}{k+1}}}.$$

From this, we compute the numbers $T_{n,j,m}$ by extracting the proper coefficients:

$$\begin{aligned}
\frac{T_{n,j,m}}{(n-j)!} &= [z^{n-j}v^m]T^{[j]}(z,v) = T_j \binom{m-1-\frac{1}{k+1}}{m-1}[z^{n-j}]\frac{((k+1)z)^{m-1}}{(1-(k+1)z)^{j-1+\frac{2}{k+1}}} \\
&= T_j(k+1)^{n-j}\binom{m-1-\frac{1}{k+1}}{m-1}[z^{n-m-j+1}]\frac{1}{(1-(k+1)z)^{j-1+\frac{2}{k+1}}} \\
&= T_j(k+1)^{n-j}\binom{m-1-\frac{1}{k+1}}{m-1}\binom{n-m-1+\frac{2}{k+1}}{n-m-j+1}.
\end{aligned}$$

Now, again using the relation $\mathbb{P}\{D_{n,j}=m\} = \frac{T_{n,j,m}}{T_n}$ and the formula for T_n from Theorem 4.3.1, we can compute the probabilities $\mathbb{P}\{D_{n,j}=m\}$:

$$\begin{aligned}
\mathbb{P}\{D_{n,j}=m\} &= \frac{T_j(k+1)^{n-j}(n-j)!}{T_n}\binom{m-1-\frac{1}{k+1}}{m-1}\binom{n-m-1+\frac{2}{k+1}}{n-m-j+1} \\
&= \frac{j!\binom{j-1+\frac{1}{k+1}}{j}(n-j)!}{n!\binom{n-1+\frac{1}{k+1}}{n}}\binom{m-1-\frac{1}{k+1}}{m-1}\binom{n-m-1+\frac{2}{k+1}}{n-m-j+1} \\
&= \frac{\binom{m-1-\frac{1}{k+1}}{m-1}\binom{n-m-1+\frac{2}{k+1}}{n-m-j+1}}{\binom{n-1+\frac{1}{k+1}}{n-j}}.
\end{aligned}$$

This proves the part of Theorem 4.7.1 concerning the family $\mathcal{T}^{[o]}$.

Family $\mathcal{T}^{[d]}$ (d-ary increasing k-trees): Like in the previous cases, we first establish a recurrence relation for the numbers $T_{n,j,m}$: Clearly, it holds that in any d-ary increasing k-tree of size j node j has exactly one descendant (namely itself). Furthermore, in any such k-tree, there are exactly kd locations at which a descendant of node j can be attached, since node j is contained in k different k-cliques, each of which has d empty slots. Moreover, we observe that each new descendant

x of j increases the number of possible locations for subsequent descendants by exactly $kd-1$, namely by creating k new k-cliques (each with d empty slots) but at the same time occupying one of the existing locations. Consequently, given any d-ary increasing k-tree of size n in which node j has m descendants, there are $(kd-1)m+1$ slots at which a new descendant of node j can be attached, and $(kd-1)(n-m)+d-1$ possible locations for a new node which will not be a descendant of j. Hence, for the numbers $T_{n,j,m}$ of d-ary increasing k-trees of size n in which node j has m descendants, we get the recurrence relation

$$\begin{aligned} T_{n,j,m} &= ((kd-1)(n-m-1)+d-1)\, T_{n-1,j,m} \\ &\quad + ((kd-1)(m-1)+1)\, T_{n-1,j,m-1}, \end{aligned}$$

for $n > j \geq 1$ and $m \geq 1$, with $T_{j,j,1} = T_j$, for $j \geq 1$, and $T_{j,j,m} = 0$, for $m > 1$. In this case, one gets that the generating function $T^{[j]}(z,v)$ defined by (4.29) satisfies the PDE

$$\begin{aligned} &((kd-1)z-1)T_z^{[j]}(z,v) + (kd-1)v(v-1)T_v^{[j]}(z,v) \\ &\qquad + ((kd-1)j+d-1+v)T^{[j]}(z,v) = 0, \quad T^{[j]}(0,v) = T_j v. \end{aligned}$$

Like in the previous cases, one can solve this equation by an application of the method of characteristics: A first integral for the system of characteristic equations $\dot{z} = (kd-1)z-1$, $\dot{v} = (kd-1)v(v-1)$, is given by

$$\zeta(z,v) = \frac{v(1-(kd-1)z)}{1-v}.$$

We use the coordinate transform $\zeta = \zeta(z,v)$, $\eta = \eta(z,v) = 1-(kd-1)z$ (equivalently, $z(\zeta,\eta) = \frac{1-\eta}{kd-1}$, $v(\zeta,\eta) = \frac{\zeta}{\zeta+\eta}$), which leads for the function $\tilde{T}^{[j]}(\zeta,\eta) := T^{[j]}(z(\zeta,\eta), v(\zeta,\eta))$ to the following ordinary differential equation:

$$\tilde{T}_\eta^{[j]}(\zeta,\eta) = \left(-\frac{\zeta}{(kd-1)\eta(\zeta+\eta)} - \frac{(kd-1)j-d+1}{(kd-1)\eta}\right)\tilde{T}^{[j]}(\zeta,\eta).$$

This equation can easily be solved, and after transformation to (z,v)-coordinates one obtains

$$T^{[j]}(z,v) = \frac{C\left(\frac{v(1-(kd-1)z)}{1-v}\right)\left(\frac{1-(kd-1)z}{1-v}\right)^{\frac{1}{kd-1}}}{(1-(kd-1)z)^{j+\frac{d-2}{kd-1}}},$$

where C is a continuous function. Using the boundary condition $T^{[j]}(0,v) = T_j v$, C can be computed like in the previous two cases, and this finally leads to the solution

$$T^{[j]}(z,v) = \frac{T_j v}{(1-(kd-1)vz)^{\frac{kd}{kd-1}}(1-(kd-1)z)^{j-1+\frac{d-1}{kd-1}}}.$$

We extract the numbers $T_{n,j,m}$ like in the previous two cases, which gives

$$\begin{aligned}\frac{T_{n,j,m}}{(n-j)!} &= [z^{n-j}v^m]T^{[j]}(z,v)\\ &= T_j(kd-1)^{n-j}\binom{m-1+\frac{1}{kd-1}}{m-1}\binom{n-m-1+\frac{d-1}{kd-1}}{n-m-j+1},\end{aligned}$$

from which one easily derives the formula for the exact probabilities $\mathbb{P}\{D_{n,j}=m\}$ as given in Theorem 4.7.1 by using the relation $\mathbb{P}\{D_{n,j}=m\} = \frac{T_{n,j,m}}{T_n}$ and the formula for T_n from Proposition 4.3.1.

4.7.2.2 Derivation of the limiting distributions results

We derive the limiting distribution results for $D_{n,j}$ claimed in Theorem 4.7.2 directly from the exact results obtained in Section 4.7.2.1. We concentrate on the derivations for the family $\mathcal{T}^{[u]}$. The computations for the remaining cases are only sketched, since they are quite similar.

Family $\mathcal{T}^{[u]}$ (unordered increasing k-trees):

- *The region for j fixed:*

  We use the method of moments in order to show the convergence result for $\frac{D_{n,j}}{n}$: Since the probability generating function

  $$p_{n,j}(v) := \sum_{m\ge 0}\mathbb{P}\{D_{n,j}-1=m\}\,v^m,$$

  of the shifted random variable $D_{n,j}-1$ is given by

  $$p_{n,j}(v) = \frac{(n-j)!}{T_n}[z^{n-j}]\frac{T^{[j]}(z,v)}{v},$$

  where $T^{[j]}(z,v)$ is defined as in (4.29), one can obtain the s-th factorial moments $\mathbb{E}((D_{n,j}-1)^{\underline{s}})$ via

  $$\mathbb{E}((D_{n,j}-1)^{\underline{s}}) = \frac{(n-j)!}{T_n}[z^{n-j}]\mathrm{U}_v\mathrm{D}_v^s\frac{T^{[j]}(z,v)}{v}.$$

Hence, by using the formula for $T^{[j]}(z,v)$ from (4.33), we get

$$\begin{aligned}\mathbb{E}((D_{n,j}-1)^{\underline{s}}) &= \frac{(n-j)!\;T_j}{T_n}[z^{n-j}]\frac{1}{(1-kz)^{j-1+\frac{1}{k}}}\mathrm{U}_v\mathrm{D}_v^s\frac{1}{1-vkz} \\ &= \frac{(n-j)!\;T_j}{T_n}[z^{n-j}]\frac{s!(kz)^s}{(1-kz)^{s+j+\frac{1}{k}}} \\ &= s!\frac{k^{n-j}(n-j)!\;T_j}{T_n}[z^{n-j-s}]\frac{1}{(1-z)^{s+j+\frac{1}{k}}} \\ &= s!\frac{k^{n-j}(n-j)!\;T_j}{T_n}\binom{n-1+\frac{1}{k}}{n-j-s} = s!\frac{\binom{n-1+\frac{1}{k}}{n-j-s}}{\binom{n-1+\frac{1}{k}}{n-j}}.\end{aligned}$$

Simple manipulations of the last expression yield

$$\mathbb{E}((D_{n,j}-1)^{\underline{s}}) = s!\frac{\Gamma\big(j+\frac{1}{k}\big)}{\Gamma\big(s+j+\frac{1}{k}\big)}(n-j)^{\underline{s}}, \tag{4.35}$$

leading (for j arbitrary, but fixed) to the following asymptotic expansion of the (shifted) s-th factorial moments:

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

Now note that the ordinary s-th moments $\mathbb{E}\big(D_{n,j}^s\big)$ of $D_{n,j}$ can be computed from the factorial moments $\mathbb{E}\Big(D_{n,j}^{\underline{j}}\Big)$ via a linear combination of the form

$$\mathbb{E}\big(D_{n,j}^s\big) = \mathbb{E}((D_{n,j}-1)^{\underline{s}}) + \sum_{i=0}^{s-1} c_{s,i}\mathbb{E}\big((D_{n,j}-1)^{\underline{i}}\big)\,, \tag{4.36}$$

with computable constants $c_{s,i}$. Thus we also have the asymptotic expansion

$$\mathbb{E}\big(D_{n,j}^s\big) = s!\frac{\Gamma\big(j+\frac{1}{k}\big)}{\Gamma\big(s+j+\frac{1}{k}\big)}n^s + \mathcal{O}\big(n^{s-1}\big)\,.$$

This shows that

$$\lim_{n\to\infty}\mathbb{E}\bigg(\Big(\frac{D_{n,j}}{n}\Big)^s\bigg) = s!\frac{\Gamma\big(j+\frac{1}{k}\big)}{\Gamma\big(s+j+\frac{1}{k}\big)},$$

i.e. the s-th moments of $\frac{D_{n,j}}{n}$ converge to the s-th moments of a random variable which is Beta(α,β) distributed with $\alpha=1$ and $\beta=j-1+\frac{1}{k}$. The convergence in distribution claimed in Theorem 4.7.2 now follows directly from the Theorem of Fréchet and Shohat (Theorem 2.2.1).

- *The region for j small: $j \to \infty$ such that $j = o(n)$:*

  Like in the case for fixed j we use the method of moments: Since $j \to \infty$, we can apply Lemma 2.2.5 to the factor $\frac{\Gamma\left(j+\frac{1}{k}\right)}{\Gamma\left(s+j+\frac{1}{k}\right)}$ in (4.35). This immediately yields

$$\begin{aligned}\mathbb{E}((D_{n,j}-1)^{\underline{s}}) &= s!\frac{(n-j)^s}{j^s}\left(1+\mathcal{O}\left(\frac{1}{j}\right)\right)\left(1+\mathcal{O}\left(\frac{1}{n-j}\right)\right)\\ &= s!\left(\frac{n}{j}\right)^s\left(1+\mathcal{O}\left(\frac{1}{j}\right)\right)\left(1+\mathcal{O}\left(\frac{j}{n}\right)\right).\end{aligned}$$

  As before, this leads via (4.36) also to an asymptotic expansion for the s-th moments $\mathbb{E}\left(D_{n,j}^s\right)$, and we finally obtain

$$\lim_{n\to\infty}\mathbb{E}\left(\left(\frac{j}{n}D_{n,j}\right)^s\right) = s!.$$

  By using the Theorem of Fréchet and Shohat, this proves the convergence in distribution of $\frac{j}{n}D_{n,j}$ to an Exp(1) distributed random variable.

- *The central region for j: $j \to \infty$ such that $j \sim \rho n$, with $0 < \rho < 1$:*

  We use the exact expressions for the probabilities $\mathbb{P}\{D_{n,j} = m\}$ given in Theorem 4.7.1, which can be written as

$$\mathbb{P}\{D_{n,j}-1=m\} = \frac{\Gamma\left(n-m-1+\frac{1}{k}\right)}{\Gamma\left(n+\frac{1}{k}\right)}\frac{\Gamma\left(j+\frac{1}{k}\right)}{\Gamma\left(j-1+\frac{1}{k}\right)}\frac{\Gamma(n-j+1)}{\Gamma(n-j-m+1)}.$$

  Since we have $n \to \infty$, $j \to \infty$ and $n-j \to \infty$, we can apply Lemma 2.2.5 three times and get, for each fixed $m \in \mathbb{N}_0$,

$$\begin{aligned}\mathbb{P}\{D_{n,j}-1=m\} &= \frac{j(n-j)^m}{n^{m+1}}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)\\ &= \frac{\rho n(n-\rho n)^m}{n^{m+1}}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)(1+o(1))\\ &= \rho(1-\rho)^m(1+o(1)),\end{aligned}$$

  where we have used that in the considered region holds $\mathcal{O}\left(\frac{1}{j}\right) = \mathcal{O}\left(\frac{1}{n}\right)$, $\mathcal{O}\left(\frac{1}{n-j}\right) = \mathcal{O}\left(\frac{1}{n}\right)$, and $j = \rho n + o(\rho n) = \rho n(1+o(1))$. Hence, the probability mass function of $D_{n,j}-1$ converges pointwise to the probability mass function of a Geom(ρ) distributed random variable, which proves the convergence in distribution in this region as stated in Theorem 4.7.2.

- *The region for j large: $j \to \infty$ such that $\tilde{j} := n - j = o(n)$:*

  From Theorem 4.7.1 we directly obtain the result stated in Theorem 4.7.2:

$$\mathbb{P}\{D_{n,j} = 1\} = \frac{\binom{n-2+\frac{1}{k}}{n-j}}{\binom{n-1+\frac{1}{k}}{n-j}} = \frac{j-1+\frac{1}{k}}{n-1+\frac{1}{k}} = \frac{n-1+\frac{1}{k}+\tilde{j}}{n-1+\frac{1}{k}} = 1 + \mathcal{O}\Big(\frac{\tilde{j}}{n}\Big).$$

Family $\mathcal{T}^{[o]}$ (ordered increasing k-trees): The computations for this family are completely analogous to the ones for $\mathcal{T}^{[u]}$. First, one extracts the coefficient of z^{n-j} from the function $U_v D_v^s \frac{T^{[j]}(z,v)}{v}$ in order to obtain the auxiliary result

$$\mathbb{E}((D_{n,j}-1)^{\underline{s}}) = s! \frac{\binom{s-\frac{1}{k+1}}{s}\binom{n-1+\frac{1}{k+1}}{n-j-s}}{\binom{n-1+\frac{1}{k+1}}{n-j}}. \tag{4.37}$$

From this one eventually gets, for j fixed,

$$\begin{aligned}\mathbb{E}\big(D_{n,j}^s\big) &= s!\binom{s-\frac{1}{k+1}}{s}\frac{\Gamma\big(j+\frac{1}{k+1}\big)}{\Gamma\big(s+j+\frac{1}{k+1}\big)}n^s + \mathcal{O}\big(n^{s-1}\big)\\ &= \frac{\Gamma\big(s+\frac{k}{k+1}\big)\,\Gamma\big(j+\frac{1}{k+1}\big)}{\Gamma\big(\frac{k}{k+1}\big)\,\Gamma\big(s+j+\frac{1}{k+1}\big)}n^s + \mathcal{O}\big(n^{s-1}\big),\end{aligned}$$

which, by Theorem 2.2.1, implies the convergence in distribution of $\frac{D_{n,j}}{n}$ to a Beta$(\frac{k}{k+1}, j-1+\frac{2}{k+1})$ distributed random variable. For the region $j \to \infty$, $j = o(n)$, one can use (4.37) in order to show that

$$\mathbb{E}\big(D_{n,j}^s\big) = \frac{\Gamma\big(s+\frac{k}{k+1}\big)}{\Gamma\big(\frac{k}{k+1}\big)}\left(\frac{n}{j}\right)^s\left(1+\mathcal{O}\Big(\frac{1}{j}\Big)\right)\left(1+\mathcal{O}\Big(\frac{j}{n}\Big)\right),$$

and thus another application of the Theorem of Fréchet and Shohat yields the convergence in distribution of $\frac{j}{n}D_{n,j}$ to a Gamma$(\frac{k}{k+1}, 1)$ distributed random variable. For the region where $j \sim \rho n$, $0 < \rho < 1$, one uses the exact formula for the probabilities $\mathbb{P}\{D_{n,j} - 1 = m\}$ given in Theorem 4.7.1, and derives

$$\begin{aligned}\mathbb{P}\{D_{n,j}-1=m\} &= \binom{m-\frac{1}{k+1}}{m}\frac{j^{\frac{k}{k+1}}(n-j)^m}{n^{m+\frac{k}{k+1}}}\left(1+\mathcal{O}\Big(\frac{1}{n}\Big)\right)\\ &= \binom{m+\frac{k}{k+1}-1}{m}\rho^{\frac{k}{k+1}}(1-\rho)^m\,(1+o(1)).\end{aligned}$$

This then proves that $D_{n,j} - 1 \xrightarrow{(d)} D_\rho$, where D_ρ is NegBin$(\frac{k}{k+1}, \rho)$ distributed. Finally, for the region where $n - j = o(n)$, one simply uses the formula for $\mathbb{P}\{D_{n,j} = 1\}$ in order to show that $\mathbb{P}\{D_{n,j} = 1\} \to 1$ in this case.

Family $\mathcal{T}^{[d]}$ (d-ary increasing k-trees): The computations for d-ary increasing k-trees do not significantly differ from the ones for the previous two cases. Again, we first compute the s-th factorial moments

$$\mathbb{E}((D_{n,j}-1)^{\underline{s}}) = s!\frac{\binom{s+\frac{1}{kd-1}}{s}\binom{n-1+\frac{d}{kd-1}}{n-j-s}}{\binom{n-1+\frac{d}{kd-1}}{n-j}},$$

from which asymptotic expansions for the s-th moments $\mathbb{E}\big(D_{n,j}^s\big)$ can be derived easily. Then we use the method of moments in order to prove the limit laws for fixed j, and for $j \to \infty$, $j = o(n)$. For the limiting distribution results in the cases $j \sim \rho n$ $(0 < \rho < 1)$ and $n - j = o(n)$ one can directly study the formula for the probabilities $\mathbb{P}\{D_{n,j} = m\}$ as given in Theorem 4.7.1.

4.7.2.3 Derivation of the results for a randomly selected node

To obtain our results for $\overline{D}_n$, i.e. the number of descendants of a randomly selected node in a random increasing k-tree of size n, as stated in Theorem 4.7.3, we use $\overline{D}_n \overset{(d)}{=} D_{n,U_n}$, with $U_n \overset{(d)}{=} \text{Uniform}(\{1,2,\dots,n\})$, and the exact results for $D_{n,j}$ given in Theorem 4.7.1.

Family $\mathcal{T}^{[u]}$ (unordered increasing k-trees): Clearly, the exact distribution of $\overline{D}_n$ can be obtained by summation via

$$\mathbb{P}\big\{\overline{D}_n = m\big\} = \frac{1}{n}\sum_{j=1}^{n}\mathbb{P}\{D_{n,j} = m\}\,. \tag{4.38}$$

To get a closed formula we consider an alternative representation for the exact probabilities of $D_{n,j}$ (note that we use here the definition $z! := \Gamma(z+1)$ also for a non-integer z):

$$\begin{aligned}
\mathbb{P}\{D_{n,j} = m\} &= \frac{\binom{n-m-1+\frac{1}{k}}{n-m-j+1}}{\binom{n-1+\frac{1}{k}}{n-j}} = \frac{(n-m-1+\frac{1}{k})!(n-j)!}{(n-m-j+1)!(n-1+\frac{1}{k})!}\left(j-1+\frac{1}{k}\right) \\
&= \frac{(n-m+\frac{1}{k})!(n-j)!}{(n-m-j+1)!(n-1+\frac{1}{k})!} - \frac{(n-m-1+\frac{1}{k})!(n-j)!}{(n-m-j)!(n-1+\frac{1}{k})!} \\
&= \frac{\binom{n-j}{m-1}}{\binom{n-1+\frac{1}{k}}{m-1}} - \frac{\binom{n-j}{m}}{\binom{n-1+\frac{1}{k}}{m}}, \quad \text{for } n \geq j \geq 1 \text{ and } m \geq 1.
\end{aligned}$$

Using this and the standard binomial identity $\sum_{j=0}^{n} \binom{j}{m} = \binom{n+1}{m+1}$, with $n, m \in \mathbb{N}_0$, (4.38) sums up to

$$\mathbb{P}\{\overline{D}_n = m\} = \frac{1}{n} \left(\frac{\binom{n}{m}}{\binom{n-1+\frac{1}{k}}{m-1}} - \frac{\binom{n}{m+1}}{\binom{n-1+\frac{1}{k}}{m}} \right), \tag{4.39}$$

which, after simple manipulations, leads to the exact formula stated in Theorem 4.7.3.

In order to obtain the limiting distribution of $\overline{D}_n$, we use (4.39) and apply Stirling's formula. We then get, for arbitrary but fixed $m \geq 1$:

$$\begin{aligned}
\mathbb{P}\{\overline{D}_n = m\} &= \frac{\Gamma(n)\,\Gamma\big(n-m+1+\frac{1}{k}\big)}{m\Gamma\big(n+\frac{1}{k}\big)\,\Gamma(n-m+1)} - \frac{\Gamma(n)\,\Gamma\big(n-m+\frac{1}{k}\big)}{(m+1)\Gamma\big(n+\frac{1}{k}\big)\,\Gamma(n-m)} \\
&= \frac{n^{\frac{1}{k}}\left(1+\mathcal{O}\big(\frac{1}{n}\big)\right)}{mn^{\frac{1}{k}}\left(1+\mathcal{O}\big(\frac{1}{n}\big)\right)} - \frac{n^{\frac{1}{k}}\left(1+\mathcal{O}\big(\frac{1}{n}\big)\right)}{(m+1)n^{\frac{1}{k}}\left(1+\mathcal{O}\big(\frac{1}{n}\big)\right)} \\
&= \frac{1}{m(m+1)}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right).
\end{aligned}$$

Hence, the probability mass function of $\overline{D}_n$ converges pointwise to the probability mass function of a discrete random variable $\overline{D}$, with

$$\mathbb{P}\{\overline{D} = m\} = \frac{1}{m(m+1)}, \quad \text{for } m \geq 1.$$

This completes the proof of Theorem 4.7.3 for the family $\mathcal{T}^{[u]}$.

Family $\mathcal{T}^{[o]}$ (ordered increasing k-trees): Here one does not obtain a closed formula for the probabilities

$$\mathbb{P}\{\overline{D}_n = m\} = \frac{1}{n}\sum_{j=1}^{n} \mathbb{P}\{D_{n,j} = m\} = \frac{\binom{m-1-\frac{1}{k+1}}{m-1}}{n} \sum_{j=1}^{n} \frac{\binom{n-m-1+\frac{2}{k+1}}{n-m-j+1}}{\binom{n-1+\frac{1}{k+1}}{n-j}}. \tag{4.40}$$

However, in order to characterize the limiting distribution of $\overline{D}_n$, it is advantageous to transform (4.40) into a sum in which the summation boundaries don't depend on n. In order to do this, we use the following idea: Consider the expression

$$\frac{\binom{n+\alpha-m}{n-m-j}}{\binom{n+\beta}{n-j}} = \frac{(n+\alpha-m)!}{(n+\beta)!}(n-j)^{\underline{m}}\frac{(j+\beta)!}{(j+\alpha)!}, \tag{4.41}$$

for real α, β, where we assume that $-\alpha$, $-\beta$, and $\alpha - \beta \notin \mathbb{N}_0$ (again, we set $z! := \Gamma(z+1)$ also for a non-integer z). Now, write

$$(n-j)^{\underline{m}} = \sum_{\ell=0}^{m} a_\ell (j+\alpha)^{\underline{\ell}}, \tag{4.42}$$

where the coefficients a_ℓ should not depend on j. Note that such a representation is always possible, since $(n-j)^{\underline{m}}$ is a polynomial in j of degree m, and the polynomials $(j+\alpha)^{\underline{\ell}}$, $\ell = 0 \ldots m$, are linearly independent. Given this representation, it follows that

$$\sum_{j=1}^{n} (n-j)^{\underline{m}} \frac{(j+\beta)!}{(j+\alpha)!} = \sum_{j=1}^{n} \sum_{\ell=0}^{m} a_\ell \frac{(j+\beta)!}{(j+\alpha-l)!} = \sum_{\ell=0}^{m} a_\ell (\beta-\alpha+l)! \sum_{j=1}^{n} \binom{j+\beta}{\beta-\alpha+l}.$$

The inner sum of the last expression can then be computed using (4.26), and hence we have transformed a sum over (4.41) for $j = 1 \ldots n$ into a sum in which the summation boundaries are independent of n.

Now let us make this idea concrete: The constants a_ℓ in (4.42) can be computed using the difference calculus (see Appendix C). For this purpose, let Δ_j denote the difference operator with respect to j. By applying, for fixed $r \in \{0, \ldots, m\}$, Δ_j^r to (4.42) and then setting $j = -\alpha$, one obtains

$$\begin{aligned} r! a_r &= \left[\Delta_j^r (n-j)^{\underline{m}}\right]_{j=-\alpha} = \left[(-1)^r m^{\underline{r}} (n-j-r)^{\underline{m-r}}\right]_{j=-\alpha} \\ &= (-1)^r m^{\underline{r}} (n+\alpha-r)^{\underline{m-r}}, \end{aligned}$$

and hence (4.42) reads explicitly as

$$(n-j)^{\underline{m}} = \sum_{\ell=0}^{m} \binom{m}{\ell} (-1)^\ell (n+\alpha-\ell)^{\underline{m-\ell}} (j+\alpha)^{\underline{\ell}}.$$

By applying this to (4.41), we get after simple manipulations:

$$\frac{\binom{n+\alpha-m}{n-m-j}}{\binom{n+\beta}{n-j}} = \sum_{\ell=0}^{m} \binom{m}{\ell} (-1)^\ell \frac{\binom{j+\beta}{\beta-\alpha+\ell}}{\binom{n+\beta}{\beta-\alpha+\ell}}. \tag{4.43}$$

Now, we sum up (4.43) for $j = 1 \ldots n$, change the summation order and compute the inner sum using (4.26):

$$\begin{aligned}
\sum_{j=1}^{n} \frac{\binom{n+\alpha-m}{n-m-j}}{\binom{n+\beta}{n-j}} &= \sum_{\ell=0}^{m} \binom{m}{\ell}(-1)^{\ell} \frac{1}{\binom{n+\beta}{\beta-\alpha+\ell}} \sum_{j=1}^{n} \binom{j+\beta}{\beta-\alpha+\ell} \\
&= \sum_{\ell=0}^{m} \binom{m}{\ell}(-1)^{\ell} \frac{1}{\binom{n+\beta}{\beta-\alpha+\ell}} \left(\binom{n+1+\beta}{\beta-\alpha+\ell+1} - \binom{1+\beta}{\beta-\alpha+\ell+1} \right) \\
&= \sum_{\ell=0}^{m} \binom{m}{\ell}(-1)^{\ell} \frac{n+\beta+1}{\ell+\beta-\alpha+1} \\
&\quad - \frac{1}{\binom{n+\beta}{n}} \sum_{\ell=0}^{m} \binom{m}{\ell}(-1)^{\ell} \frac{\beta+1}{\ell+\beta-\alpha+1} \binom{n+\alpha-\ell}{n}.
\end{aligned}$$

The first sum of the last expression can be simplified by using the identity (C.3) (cf. Appendix C) which is also obtained via the difference calculus. Hence, after finally shifting m by 1, we obtain the following formula:

$$\begin{aligned}
&\sum_{j=1}^{n} \frac{\binom{n+\alpha-m+1}{n-m-j+1}}{\binom{n+\beta}{n-j}} = \\
&\quad = \frac{n+\beta+1}{m\binom{m+\beta-\alpha}{m}} - \frac{1}{\binom{n+\beta}{n}} \sum_{\ell=0}^{m-1} \binom{m-1}{\ell}(-1)^{\ell} \frac{\beta+1}{\ell+\beta-\alpha+1} \binom{n+\alpha-\ell}{n}.
\end{aligned} \tag{4.44}$$

A direct application of (4.44) to (4.40) (via $\alpha = \frac{2}{k+1} - 2$, $\beta = \frac{1}{k+1} - 1$) now shows after some simple manipulations the exact formula for the distribution of $\overline{D}_n$ as stated in Theorem 4.7.3.

Applying Stirling's formula to this expression for the probabilities $\mathbb{P}\{\overline{D}_n = m\}$ shows that, for arbitrary but fixed $m \geq 1$, the second part of the formula is asymptotically negligible, i.e. of order $\mathcal{O}\left(n^{-2+\frac{1}{k+1}}\right)$, and hence one gets:

$$\mathbb{P}\{\overline{D}_n = m\} = \frac{k}{(k+1)(m+\frac{k}{k+1})(m-\frac{1}{k+1})} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

Thus the proof of the theorem for the family $\mathcal{T}^{[o]}$ is completed.

Family $\mathcal{T}^{[d]}$ (d-ary increasing k-trees): In the case of d-ary increasing k-trees, the distribution of $\overline{D}_n$ can be computed analogously via

$$\mathbb{P}\{\overline{D}_n = m\} = \frac{1}{n} \sum_{j=1}^{n} \mathbb{P}\{D_{n,j} = m\} = \frac{\binom{m-1+\frac{1}{kd-1}}{m-1}}{n} \sum_{j=1}^{n} \frac{\binom{n-m-1+\frac{d-1}{kd-1}}{n-m-j+1}}{\binom{n-1+\frac{d}{kd-1}}{n-j}}. \tag{4.45}$$

To this, we apply (4.44) with $\alpha = \frac{d-1}{kd-1} - 2$ and $\beta = \frac{d}{kd-1} - 1$, which gives the formula for the exact distribution of $\overline{D}_n$ as stated in Theorem 4.7.3.

For the limiting distribution, one again applies Stirling's formula to the exact expression for the probabilities $\mathbb{P}\{\overline{D}_n = m\}$, and obtains

$$\mathbb{P}\{\overline{D}_n = m\} = \frac{kd}{(kd-1)(m+\frac{kd}{kd-1})(m+\frac{1}{kd-1})}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right).$$

This completes the proof of Theorem 4.7.3 for the family $\mathcal{T}^{[d]}$.

4.8 Degree distribution

4.8.1 Results

In the following, whenever we state results on the family $\mathcal{T}^{[d]}$, we assume that not both $k = 2$ and $d = 1$. Since this is a special case in which the distribution of the out-degree shows a completely different behaviour, we chose to omit it here.

Theorem 4.8.1. *The random variable $O_{n,j}$, which counts the out-degree of node j in a random increasing k-tree of size n, has the following exact distribution:*

$$\mathbb{P}\{O_{n,j} = m\} = \begin{cases} \frac{\binom{m+\frac{1}{k-1}}{m}}{\binom{n-\frac{k-1}{k}}{n-j}} \sum_{\ell=0}^{m}(-1)^{\ell}\binom{m}{\ell}\binom{n-\frac{(k-1)\ell+2k-1}{k}}{n-j}, & \text{family } \mathcal{T}^{[u]}, \\ \frac{1}{\binom{n-\frac{k}{k+1}}{n-j}} \sum_{\ell=0}^{m}(-1)^{\ell}\binom{m}{\ell}\binom{n-\frac{k\ell+2k}{k+1}}{n-j}, & \text{family } \mathcal{T}^{[o]}, \\ \frac{\binom{m+\frac{d+1}{(k-1)d-1}}{m}}{\binom{n-\frac{(k-1)d-1}{kd-1}}{n-j}} \sum_{\ell=0}^{m}(-1)^{\ell}\binom{m}{\ell}\binom{n-\frac{((k-1)d-1)\ell+(2k-1)d-1}{kd-1}}{n-j}, & \text{family } \mathcal{T}^{[d]}, \end{cases}$$

for $1 \le j \le n$ and $0 \le m \le n-j$.

Theorem 4.8.2. *The limiting distribution behaviour of $O_{n,j}$ is, for $n \to \infty$ and depending on the growth of j, characterized as follows:*

- *The region for j fixed: The normalized random variable*

$$\begin{aligned} &n^{-\frac{k-1}{k}}O_{n,j}, && \text{family } \mathcal{T}^{[u]}, \\ &n^{-\frac{k}{k+1}}O_{n,j}, && \text{family } \mathcal{T}^{[o]}, \\ &n^{-\frac{(k-1)d-1}{kd-1}}O_{n,j}, && \text{family } \mathcal{T}^{[d]}, \end{aligned}$$

respectively, converges in distribution to a random variable O_j, which is fully characterized by its moments. The s-th moments of O_j are, for $s \geq 0$, given by

$$\mathbb{E}\big(O_j^s\big) = \begin{cases} \frac{\Gamma\big(s+\frac{k}{k-1}\big)\Gamma\big(j+\frac{1}{k}\big)}{\Gamma\big(\frac{k}{k-1}\big)\Gamma\big(j+\frac{(k-1)s+1}{k}\big)}, & \text{family } \mathcal{T}^{[u]}, \\ s!\frac{\Gamma\big(j+\frac{1}{k+1}\big)}{\Gamma\big(j+\frac{ks+1}{k+1}\big)}, & \text{family } \mathcal{T}^{[o]}, \\ \frac{\Gamma\big(s+\frac{kd}{(k-1)d-1}\big)\Gamma\big(j+\frac{d}{kd-1}\big)}{\Gamma\big(\frac{kd}{(k-1)d-1}\big)\Gamma\big(j+\frac{((k-1)d-1)s+d}{kd-1}\big)}, & \text{family } \mathcal{T}^{[d]}. \end{cases}$$

- *The region for j small: $j \to \infty$ such that $j = o(n)$. The normalized random variable*

$$\begin{aligned} &\left(\frac{j}{n}\right)^{\frac{k-1}{k}} O_{n,j}, && \text{family } \mathcal{T}^{[u]}, \\ &\left(\frac{j}{n}\right)^{\frac{k}{k+1}} O_{n,j}, && \text{family } \mathcal{T}^{[o]}, \\ &\left(\frac{j}{n}\right)^{\frac{(k-1)d-1}{kd-1}} O_{n,j}, && \text{family } \mathcal{T}^{[d]}, \end{aligned}$$

converges weakly to the $\mathrm{Gamma}(\kappa, \theta)$ *distribution with parameters*

$$\kappa = \begin{cases} \frac{k}{k-1}, & \text{family } \mathcal{T}^{[u]}, \\ 1, & \text{family } \mathcal{T}^{[o]}, \\ \frac{kd}{(k-1)d-1}, & \text{family } \mathcal{T}^{[d]}, \end{cases} \qquad \text{and} \qquad \theta = 1.$$

For the family $\mathcal{T}^{[o]}$, this limiting distribution is the $\mathrm{Exp}(1)$ *distribution.*

- *The central region for j: $j \to \infty$ such that $j \sim \rho n$, with $0 < \rho < 1$: The random variable $O_{n,j}$ converges in distribution to a random variable O_ρ which is* $\mathrm{NegBin}(r, p)$ *distributed with parameters*

$$r = \begin{cases} \frac{k}{k-1}, & \text{family } \mathcal{T}^{[u]}, \\ 1, & \text{family } \mathcal{T}^{[o]}, \\ \frac{kd}{(k-1)d-1}, & \text{family } \mathcal{T}^{[d]}, \end{cases} \qquad \text{and} \qquad p = \begin{cases} \rho^{\frac{k-1}{k}}, & \text{family } \mathcal{T}^{[u]}, \\ \rho^{\frac{k}{k+1}}, & \text{family } \mathcal{T}^{[o]}, \\ \rho^{\frac{(k-1)d-1}{kd-1}}, & \text{family } \mathcal{T}^{[d]}. \end{cases}$$

For the family $\mathcal{T}^{[o]}$, this is the $\mathrm{Geom}(\rho^{\frac{k}{k+1}})$ *distribution.*

- *The region for j large: $j \to \infty$ such that $n - j = o(n)$: It holds that $\mathbb{P}\{O_{n,j} = 0\} \to 1$.*

Theorem 4.8.3. *The random variable $\bar{O}_n$, which counts the out-degree of a randomly selected inserted node in a random increasing k-tree of size n, has the following exact distribution:*

$$\mathbb{P}\{\bar{O}_n = m\} = \begin{cases} \frac{\binom{m+\frac{1}{k-1}}{m}}{n\binom{n-\frac{k-1}{k}}{n}} \sum_{\ell=0}^{m} \frac{(-1)^\ell \binom{m}{\ell}}{(k-1)\ell+2k} \left(\binom{n+\frac{1}{k}}{n} - \binom{n-\frac{(k-1)\ell+2k-1}{k}}{n}\right), & \text{family } \mathcal{T}^{[o]}, \\ \frac{1}{n\binom{n-\frac{k}{k+1}}{n}} \sum_{\ell=0}^{m} \frac{(-1)^\ell \binom{m}{\ell}}{k\ell+2k+1} \left(\binom{n+\frac{1}{k+1}}{n} - \binom{n-\frac{k\ell+2k}{k+1}}{n}\right), & \text{family } \mathcal{T}^{[o]}, \\ \frac{\binom{m+\frac{d+1}{(k-1)d-1}}{m}}{n\binom{n-\frac{(k-1)d-1}{kd-1}}{n}} \sum_{\ell=0}^{m} \frac{(-1)^\ell \binom{m}{\ell} d}{((k-1)d-1)\ell+2kd-1} & \\ \quad \cdot \left(\binom{n+\frac{d}{kd-1}}{n} - \binom{n-\frac{((k-1)d-1)\ell+(2k-1)d-1}{kd-1}}{n}\right), & \text{family } \mathcal{T}^{[d]}, \end{cases}$$

for $0 \le m < n$. For $n \to \infty$, $\bar{O}_n$ converges in distribution to a discrete random variable $\bar{O}$, i.e. $\bar{O}_n \xrightarrow{(d)} \bar{O}$, with

$$\mathbb{P}\{\bar{O} = m\} = p_m := \begin{cases} \frac{k\binom{m+\frac{1}{k-1}}{m}}{(k-1)(m+1)\binom{m+2+\frac{2}{k-1}}{m+1}}, & \text{family } \mathcal{T}^{[u]}, \\ \frac{k+1}{k(m+1)\binom{m+2+\frac{1}{k}}{m+1}}, & \text{family } \mathcal{T}^{[o]}, \\ \frac{(kd-1)\binom{m+\frac{d+1}{(k-1)d-1}}{m}}{((k-1)d-1)(m+1)\binom{m+2+\frac{2d+1}{(k-1)d-1}}{m+1}}, & \text{family } \mathcal{T}^{[d]}, \end{cases}$$

for $m \in \mathbb{N}_0$. Since

$$p_m \sim \begin{cases} \frac{k\Gamma\left(\frac{2k}{k-1}\right)}{\Gamma\left(\frac{1}{k-1}\right)} m^{-2-\frac{1}{k-1}}, & \text{family } \mathcal{T}^{[u]}, \\ \frac{(k+1)\Gamma\left(2+\frac{1}{k}\right)}{k} m^{-2-\frac{1}{k}}, & \text{family } \mathcal{T}^{[o]}, \\ \frac{(kd-1)\Gamma\left(\frac{2kd-1}{(k-1)d-1}\right)}{((k-1)d-1)\Gamma\left(\frac{kd}{(k-1)d-1}\right)} m^{-2-\frac{d}{(k-1)d-1}}, & \text{family } \mathcal{T}^{[d]}, \end{cases}$$

for $m \to \infty$, this shows that $\bar{O}_n$ follows in all three cases asymptotically a power-law distribution.

Corollary 4.8.4. *Let the random variable C_n denote the local clustering coefficient of a random node in a random increasing k-tree of size n. Then the expected local*

k	$c_k = \lim_{n\to\infty} \mathbb{E}(C_n)$			
	$\mathcal{T}^{[u]}$	$\mathcal{T}^{[o]}$	$\mathcal{T}^{[d]}(d=3)$	$\mathcal{T}^{[d]}(d=10)$
2	$0.739208\ldots$	$0.793390\ldots$	$0.714285\ldots$	$0.732193\ldots$
3	$0.813194\ldots$	$0.843184\ldots$	$0.800402\ldots$	$0.809535\ldots$
10	$0.929089\ldots$	$0.933975\ldots$	$0.927289\ldots$	$0.928559\ldots$
50	$0.982496\ldots$	$0.982804\ldots$	$0.982391\ldots$	$0.982465\ldots$
100	$0.990800\ldots$	$0.990885\ldots$	$0.990771\ldots$	$0.990791\ldots$

Table 4.1: Numerical values for the limit c_k of the expected local clustering coefficient $\mathbb{E}(C_n)$ for small values of k.

clustering coefficient $\mathbb{E}(C_n)$ *converges to a constant, i.e.* $\lim_{n\to\infty} \mathbb{E}(C_n) = c_k$, *where*

$$c_k = \begin{cases} \sum_{m\ge 0} \frac{k\binom{m+\frac{1}{k-1}}{m}}{(m+1)(m+k)\binom{m+2+\frac{2}{k-1}}{m+1}} \left(2 - \frac{k-2}{m+k-1}\right), & \textit{family } \mathcal{T}^{[u]}, \\ \sum_{m\ge 0} \frac{(k+1)(k-1)}{k(m+1)(m+k)\binom{m+2+\frac{1}{k}}{m+1}} \left(2 - \frac{k-2}{m+k-1}\right), & \textit{family } \mathcal{T}^{[o]}, \\ \sum_{m\ge 0} \frac{(kd-1)(k-1)\binom{m+\frac{d+1}{(k-1)d-1}}{m}}{((k-1)d-1)(m+1)(m+k)\binom{m+2+\frac{2d+1}{(k-1)d-1}}{m+1}} \left(2 - \frac{k-2}{m+k-1}\right), & \textit{family } \mathcal{T}^{[d]}, \end{cases}$$

It further holds that $c_k \to 1$, *for* $k \to \infty$.

4.8.2 Proofs of the results

4.8.2.1 Derivation of the exact distribution

The claimed results can be shown using the same approach as for the distribution of the number of descendants of the nodes, i.e. by establishing a recurrence for the number $T_{n,j,m}$ of increasing k-trees of size n in which node j has out-degree m, and translating this recurrence into a partial differential equation for a suitable generating function. However, we choose a different approach in this case, which rests on the symbolic method (see Section 2.1.1). As always, we consider the three increasing k-tree families separately.

Family $\mathcal{T}^{[u]}$ (unordered increasing k-trees): Let us, as an auxiliary result, first consider the out-degree of a root node. We denote by $O_{n,0}$ the out-degree of node 0_1 in a random unordered increasing k-tree of size n, and by $T_{n,0,m} := \mathbb{P}\{O_{n,0} = m\}\, T_n$ the number of unordered increasing k-trees of size n in which node 0_1 has out-degree m. Furthermore, we consider the generating function

$$T^{[0]}(z,v) := \sum_{n\geq 0}\sum_{m\geq 0} T_{n,0,m}\frac{z^n}{n!}v^m. \tag{4.46}$$

This generating function can be computed using the symbolic method as follows: Consider the family $\mathcal{T}^{[u]}$ and the auxiliary family $\tilde{\mathcal{T}}^{[u]}$ (see Section 4.2) which contains exactly those elements of $\mathcal{T}^{[u]}$ in which the root-clique has exactly one child. As discussed in Section 4.2, one has the system of formal equations

$$\begin{aligned}\mathcal{T}^{[u]} &= \textsc{Set}\left(\tilde{\mathcal{T}}^{[u]}\right),\\ \tilde{\mathcal{T}}^{[u]} &= \Delta^{\square} * \left(\mathcal{T}^{[u]}\right)^k,\end{aligned}$$

where Δ denotes a $(k+1)$-clique consisting of k root-nodes and one child with label 1. Now note that, whenever an increasing k-tree $\mathcal{K}$ of $\mathcal{T}^{[u]}$ is constructed from a set $\{\mathcal{K}_1,\dots,\mathcal{K}_\ell\}$ of elements of $\tilde{\mathcal{T}}^{[u]}$, the out-degree of node 0_1 in $\mathcal{K}$ is just given by the sum of the out-degrees of 0_1 in $\mathcal{K}_1,\dots,\mathcal{K}_\ell$ (more precisely, by the sum of the out-degrees of those root-nodes of $\mathcal{K}_1,\dots,\mathcal{K}_\ell$, respectively, which are identified with 0_1 in the combinatorial construction). In $\tilde{\mathcal{T}}^{[u]}$ on the other hand, the out-degree of node 0_1 arises from one edge to node 1 in the component Δ, and from the out-degrees of 0_1 in those $k-1$ factors of $\left(\mathcal{T}^{[u]}\right)^k$ which are attached to Δ such that they contain 0_1 as a root-node.

Hence, if we define the generating function $\tilde{T}^{[0]}(z,v)$ for the family $\tilde{\mathcal{T}}^{[u]}$ analogously to (4.46), i.e.

$$\tilde{T}^{[0]}(z,v) := \sum_{n\geq 0}\sum_{m\geq 0} \tilde{T}_{n,0,m}\frac{z^n}{n!}v^m, \tag{4.47}$$

where $\tilde{T}_{n,0,m}$ denotes the number of objects of size n in $\tilde{\mathcal{T}}^{[u]}$, in which node 0_1 has out-degree m, we obtain, using the symbolic method, the following system of equations:

$$\begin{aligned}T^{[0]}(z,v) &= \exp\left(\tilde{T}^{[0]}(z,v)\right),\\ \tilde{T}^{[0]}_z(z,v) &= v\left(T^{[0]}(z,v)\right)^{k-1}T(z).\end{aligned} \tag{4.48}$$

We insert the first of these equations into the second one and use the known formula (4.8) for $T(z)$, which then yields

$$\tilde{T}^{[0]}_z(z,v) = \frac{v}{(1-kz)^{\frac{1}{k}}}\exp\left((k-1)\tilde{T}^{[0]}_z(z,v)\right), \quad \tilde{T}^{[0]}(0,v)=0,$$

where the given boundary value follows directly from the construction of $\tilde{\mathcal{T}}^{[u]}$. This ordinary differential equation can of course be solved by separation of variables: From

$$\int \exp\left((1-k)\tilde{T}^{[0]}\right) d\tilde{T}^{[0]} = \int \frac{vdz}{(1-kz)^{\frac{1}{k}}},$$

one gets

$$\exp\left((1-k)\tilde{T}^{[0]}\right) = \frac{v}{(1-kz)^{\frac{1}{k}-1}} + C(v),$$

where $C(v)$ is an arbitrary function. Using the initial condition $\tilde{T}^{[0]}(0,v) = 0$, one gets $C(v) = 1 - v$, and after simple manipulations this yields

$$\tilde{T}^{[0]}(z,v) = \frac{1}{k-1} \log\left(\frac{1}{1-v+v(1-kz)^{\frac{k-1}{k}}}\right).$$

By inserting this solution into the first equation of (4.48), we obtain

$$T^{[0]}(z,v) = \left(\frac{1}{1-v+v(1-kz)^{\frac{k-1}{k}}}\right)^{\frac{1}{k-1}}. \tag{4.49}$$

We now use this auxiliary result in order to determine a formula for the numbers $T_{n,j,m} := \mathbb{P}\{O_{n,j} = m\}\, T_n$ of unordered increasing k-trees of size n in which the inserted node j has out-degree m. It turns out that the following generating function is suitable for this problem:

$$T^{[j]}(z,v) := \sum_{n\geq j}\sum_{m\geq 0} T_{n,j,m} \frac{z^{n-j}}{(n-j)!} v^m. \tag{4.50}$$

We derive $T^{[j]}(z,v)$ for a fixed j by a combinatorial argument: Note that each unordered increasing k-tree T of size $n \geq j$ can in a unique way be constructed by the following procedure:

- Choose one of the T_j unordered increasing k-trees B of size j (one may think of B as the increasing k-tree of size j on which T is "based").
- Choose $kj+1$ elements $K_1, \ldots, K_{kj+1}$ of $\mathcal{T}^{[u]}$ which will be rooted at the $kj+1$ k-cliques of B (one might think of $K_1, \ldots, K_{kj+1}$ being "glued" onto the k-cliques of B, i.e. their root cliques are identified each with one k-clique of B). Of course, in order to obtain an element of size n of $\mathcal{T}^{[u]}$, the respective sizes of $K_1, \ldots, K_{kj+1}$ have to sum up to $n-j$, and an order-preserving

relabelling is necessary such that exactly the labels $\{j+1,\ldots,n\}$ are used by the inserted nodes of $K_1,\ldots,K_{kj+1}$. Note that the nodes in B are not relabelled in this process.

In this construction, exactly k of the elements $K_1,\ldots,K_{kj+1}$ contribute to the out-degree of node j, namely those which are rooted at one of the k-cliques which consist of node j and $k-1$ of j's parents. Hence, by an application of the symbolic method, this shows:

$$T^{[j]}(z,v) = T_j(T(z))^{k(j-1)+1}(T^{[0]}(z,v))^k.$$

From (4.8) and (4.49), we thus get

$$T^{[j]}(z,v) = \frac{T_j}{(1-kz)^{j-1+\frac{1}{k}}}\left(\frac{1}{1-v+v(1-kz)^{\frac{k-1}{k}}}\right)^{\frac{k}{k-1}}. \tag{4.51}$$

The probabilities $\mathbb{P}\{O_{n,j}=m\}$ can now be computed from (4.51). First, one extracts

$$\begin{aligned}
\frac{T_{n,j,m}}{(n-j)!} &= [z^{n-j}v^m]T^{[j]}(z,v)\\
&= k^{n-j}[z^{n-j}]\left(\frac{T_j}{(1-z)^{j-1+\frac{1}{k}}}[v^m]\left(\frac{1}{1-v\left(1-(1-z)^{\frac{k-1}{k}}\right)}\right)^{\frac{k}{k-1}}\right)\\
&= k^{n-j}T_j\binom{m+\frac{k}{k-1}-1}{m}[z^{n-j}]\frac{\left(1-(1-z)^{\frac{k-1}{k}}\right)^m}{(1-z)^{j-1+\frac{1}{k}}}\\
&= k^{n-j}T_j\binom{m+\frac{1}{k-1}}{m}\sum_{\ell=0}^{m}(-1)^\ell\binom{m}{\ell}[z^{n-j}]\frac{1}{(1-z)^{j-1-\frac{(k-1)\ell-1}{k}}}\\
&= k^{n-j}T_j\binom{m+\frac{1}{k-1}}{m}\sum_{\ell=0}^{m}(-1)^\ell\binom{m}{\ell}\binom{n-\frac{(k-1)\ell+2k-1}{k}}{n-j}.
\end{aligned}$$

Then, using the relation $\mathbb{P}\{O_{n,j}=m\}=\frac{T_{n,j,m}}{T_n}$ and formula (4.1) for T_n, one obtains for the family $\mathcal{T}^{[u]}$ the formula for the probabilities $\mathbb{P}\{O_{n,j}=m\}$ which is stated in Theorem 4.8.1.

Family $\mathcal{T}^{[o]}$ (ordered increasing k-trees): Like in the previous case, we first consider the out-degree of the root node 0_1. We let $T_{n,0,m} := \mathbb{P}\{O_{n,0}=m\}\,T_n$

denote the number of ordered increasing k-trees of size n in which node 0_1 has out-degree m, and define the generating function $T^{[0]}(z,v)$ as in (4.46). Again, we also consider the auxiliary family $\tilde{\mathcal{T}}^{[o]}$ which consists of all ordered increasing k-trees in which the root clique has exactly one child, and we define the generating function $\tilde{T}^{[0]}(z,v)$ for the numbers $\tilde{T}_{n,0,m}$ of objects of size n in $\tilde{\mathcal{T}}^{[o]}$ in which node 0_1 has out-degree m by (4.47). The families $\mathcal{T}^{[o]}$ and $\tilde{\mathcal{T}}^{[o]}$ satisfy

$$\mathcal{T}^{[o]} = \textsc{Seq}\left(\tilde{\mathcal{T}}^{[o]}\right),$$
$$\tilde{\mathcal{T}}^{[o]} = \Delta^{\Box} * \left(\mathcal{T}^{[o]}\right)^{k},$$

where again Δ denotes a $(k+1)$-clique consisting of k root-nodes and one child with label 1.

The out-degrees of 0_1 in $\mathcal{T}^{[o]}$ and $\tilde{\mathcal{T}}^{[o]}$ are related exactly like in the unordered case: Whenever an increasing k-tree $\mathcal{K}$ of $\mathcal{T}^{[o]}$ is constructed from a sequence $(\mathcal{K}_1,\ldots,\mathcal{K}_\ell)$ of elements of $\tilde{\mathcal{T}}^{[o]}$, the out-degree of node 0_1 in $\mathcal{K}$ is just given by the sum of the out-degrees of 0_1 in $\mathcal{K}_1,\ldots,\mathcal{K}_\ell$. In $\tilde{\mathcal{T}}^{[o]}$ on the other hand, the out-degree of node 0_1 is one (coming from the component Δ) plus the sum of out-degrees of 0_1 in those $k-1$ factors of $\left(\mathcal{T}^{[o]}\right)^k$ which are attached to Δ such that they contain 0_1 as a root-node. These considerations lead to the following system of equations for the generating functions $T^{[0]}(z,v)$ and $\tilde{T}^{[0]}(z,v)$:

$$\begin{aligned} T^{[0]}(z,v) &= \frac{1}{1-\tilde{T}^{[0]}(z,v)}, \\ \tilde{T}^{[0]}_z(z,v) &= v\left(T^{[0]}(z,v)\right)^{k-1} T(z). \end{aligned} \tag{4.52}$$

By inserting the first of these equations into the second one and using formula (4.8) for $T(z)$, we get the ordinary differential equation

$$\tilde{T}^{[0]}_z(z,v) = \frac{v}{(1-(k+1)z)^{\frac{1}{k+1}}}\left(\frac{1}{1-\tilde{T}^{[0]}(z,v)}\right)^{k-1}, \quad \tilde{T}^{[0]}(0,v)=0.$$

By separation of variables we obtain

$$\int \left((1-\tilde{T}^{[0]}\right)^{k-1} d\tilde{T}^{[0]} = \int \frac{v}{(1-(k+1)z)^{\frac{1}{k+1}}} dz,$$

which leads to

$$\left(1-\tilde{T}^{[0]}\right)^{k} = v(1-(k+1)z)^{\frac{k}{k+1}} + C(v),$$

where $C(v)$ is an arbitrary function. From the boundary condition $\tilde{T}^{[0]}(0, v) = 0$ one gets $C(v) = 1 - v$, and hence

$$\tilde{T}^{[0]}(z, v) = 1 - \left(1 - v + v(1 - (k+1)z)^{\frac{k}{k+1}}\right)^{\frac{1}{k}}.$$

Inserting this into the first equation of (4.52), we thus finally obtain

$$T^{[0]}(z, v) = \left(\frac{1}{1 - v + v(1 - (k+1)z)^{\frac{k}{k+1}}}\right)^{\frac{1}{k}}.$$

Like in the previous case, we use this auxiliary result in order to determine the generating function $T^{[j]}(z, v)$ defined by (4.51), where in this case $T_{n,j,m}$ denotes the number of ordered increasing k-trees of size n in which node j has out-degree m. Again, this can be done by a combinatorial argument:

Each ordered increasing k-tree T of size $n \geq j$ can uniquely be constructed in the following way:

- Choose one of the T_j ordered increasing k-trees B of size j.
- Choose $(k+1)j+1$ elements $K_1, \dots, K_{kj+1}$ of $\mathcal{T}^{[o]}$ whose respective sizes sum up to $n - j$. Relabel their inserted nodes in an order-preserving way such that exactly the labels $\{j+1, \dots, n\}$ are used, and identify their root cliques with the k-cliques of B, each at one of the $(k+1)j+1$ possible positions.

In this construction, exactly k of the elements $K_1, \dots, K_{(k+1)j+1}$ contribute to the out-degree of node j, namely those which are rooted at one of the k-cliques which consist of node j and $k-1$ of j's parents. This implies

$$T^{[j]}(z, v) = T_j (T(z))^{(k+1)j+1-k} (T^{[0]}(z, v))^k,$$

and hence, by using the known formulæ for T_j, $T^{[0]}(z, v)$ and $T(z)$,

$$T^{[j]}(z, v) = \frac{T_j}{(1 - (k+1)z)^{j-1+\frac{2}{k+1}} \left(1 - v + v(1 - (k+1)z)^{\frac{k}{k+1}}\right)}.$$

From this we compute

$$\begin{aligned}
\frac{T_{n,j,m}}{(n-j)!} &= [z^{n-j}v^m]T^{[j]}(z,v) = T_j[z^{n-j}]\frac{\left(1-(1-(k+1)z)^{\frac{k}{k+1}}\right)^m}{(1-(k+1)z)^{j-1+\frac{2}{k+1}}} \\
&= T_j\sum_{\ell=0}^{m}(-1)^\ell\binom{m}{\ell}[z^{n-j}]\frac{1}{(1-(k+1)z)^{j-1+\frac{2}{k+1}-\frac{k\ell}{k+1}}} \\
&= T_j(k+1)^{n-j}\sum_{\ell=0}^{m}(-1)^\ell\binom{m}{\ell}\binom{n-\frac{k\ell+2k}{k+1}}{n-j},
\end{aligned}$$

and hence the claimed formula for the probabilities $\mathbb{P}\{O_{n,j}=m\}$ follows directly from the relation $\mathbb{P}\{O_{n,j}=m\}=\frac{T_{n,j,m}}{T_n}$ and the known formula (4.1) for T_j.

Family $\mathcal{T}^{[d]}$ (d-ary increasing k-trees): Since this case is very similar to the previous two cases, we only sketch it briefly. Again, we first consider the out-degree of node 0_1, and define the numbers $T_{n,0,m}$ and the generating function $T^{[0]}(z,v)$ accordingly. Furthermore, we make use of the auxiliary family $\tilde{\mathcal{T}}^{[d]}$ as defined in Section 4.2 and the corresponding generating function $\tilde{T}^{[0]}(z,v)$. From the formal equations

$$\begin{aligned}
\mathcal{T}^{[d]} &= (\{\square\}\ \dot{\cup}\ \tilde{\mathcal{T}}^{[d]})^d, \\
\tilde{\mathcal{T}}^{[d]} &= \Delta^{\square} * \left(\mathcal{T}^{[d]}\right)^k,
\end{aligned}$$

one obtains, by considering how the out-degree of node 0_1 is compound in the combinatorial construction of $\mathcal{T}^{[d]}$ and $\tilde{\mathcal{T}}^{[d]}$, the following system of equations for $T^{[0]}(z,v)$ and $\tilde{T}^{[0]}(z,v)$:

$$\begin{aligned}
T^{[0]}(z,v) &= \left(1+\tilde{T}^{[0]}(z,v)\right)^d, \\
\tilde{T}^{[0]}_z(z,v) &= v\left(T^{[0]}(z,v)\right)^{k-1}T(z).
\end{aligned} \tag{4.53}$$

From this and the known formula (4.8) for $T(z)$ one gets the differential equation

$$\tilde{T}^{[0]}_z(z,v) = \frac{v\left(1+\tilde{T}^{[0]}(z,v)\right)^{(k-1)d}}{(1-(kd-1)z)^{\frac{d}{kd-1}}}, \quad \tilde{T}^{[0]}(0,v)=0,$$

which has (under the assumption that $k\neq 2$ or $d\neq 1$) the solution

$$\tilde{T}^{[0]}(z,v) = \left(\frac{1}{1-v+v(1-(kd-1)z)^{\frac{(k-1)d-1}{kd-1}}}\right)^{\frac{1}{(k-1)d-1}} - 1.$$

By inserting this into the first equation of (4.53), one thus obtains

$$T^{[0]}(z,v) = \left(\frac{1}{1 - v + v(1-(kd-1)z)^{\frac{(k-1)d-1}{kd-1}}} \right)^{\frac{d}{(k-1)d-1}}. \tag{4.54}$$

Now, we consider the numbers $T_{n,j,m}$ of d-ary increasing k-trees of size n in which node j has out-degree m, and define $T^{[j]}(z,v)$ like in (4.51). Again, we can find a formula for $T^{[j]}(z,v)$ by a combinatorial argument: Note that each d-ary increasing k-tree T of size $n \geq j$ can in a unique way be obtained by the following construction:

- Choose one of the T_j d-ary increasing k-trees B of size j.
- Choose k elements $K_1, \ldots, K_k$ of $\mathcal{T}^{[d]}$ and $(j-1)(kd-1)+d-1$ elements $K_1', \ldots, K_{(j-1)(kd-1)+d-1}'$ of $\{\square\} \cup \tilde{\mathcal{T}}^{[d]}$ ($\square$ denoting an increasing k-tree of size 0) whose sizes sum up to $n-j$, and relabel their inserted nodes in an order-preserving way such that exactly the labels $\{j+1, \ldots, n\}$ are used. Identify the root cliques of $K_1, \ldots, K_k$ each with one of the k-cliques of B which contain node j, and attach $K_1', \ldots, K_{(j-1)(kd-1)+d-1}'$ each to one of the free "slots" of the other k-cliques in B.

In this construction, exactly the elements $K_1, \ldots, K_k$ contribute to the out-degree of node j, and hence one obtains

$$T^{[j]}(z,v) = T_j (1+\tilde{T}(z))^{(j-1)(kd-1)+d-1} \left(T^{[0]}(z,v)\right)^k.$$

Thus, using formula (4.1) for T_n, (4.54) for $T^{[0]}(z,v)$ and (4.3.1) for $\tilde{T}(z)$, we find

$$T^{[j]}(z,v) = \frac{T_j}{(1-(kd-1)z)^{j-1+\frac{d-1}{kd-1}}} \left(\frac{1}{1 - v + v(1-(kd-1)z)^{\frac{(k-1)d-1}{kd-1}}} \right)^{\frac{kd}{(k-1)d-1}}.$$

From this, the probabilities $\mathbb{P}\{O_{n,j} = m\}$ can be computed like in the previous two cases simply via

$$\mathbb{P}\{O_{n,j} = m\} = \frac{T_{n,j,m}}{T_n} = \frac{(n-j)!}{T_n} [z^{n-j} v^m] T^{[j]}(z,v).$$

4.8.2.2 Derivation of the limiting distributions results

We derive the limiting behaviour of $O_{n,j}$ claimed in Theorem 4.8.2 directly from the exact results obtained in Section 4.8.2.1. We concentrate on the derivations for the family $\mathcal{T}^{[u]}$. The computations for the remaining cases are quite similar, thus we will only sketch them.

Family $\mathcal{T}^{[u]}$ (unordered increasing k-trees):

- *The region for j fixed:*

  We use the method of moments in order to show the claimed convergence result for $n^{-\frac{k-1}{k}} O_{n,j}$: Clearly, the probability generating function of $O_{n,j}$,

  $$p_{n,j}(v) := \sum_{m \ge 0} \mathbb{P}\{O_{n,j} = m\}\, v^m,$$

  is given by

  $$p_{n,j}(v) = \frac{(n-j)!}{T_n} [z^{n-j}] T^{[j]}(z,v),$$

  where $T^{[j]}(z,v)$ is the generating function of the numbers $T_{n,j,m}$ as defined in (4.50). Hence, we can obtain the s-th factorial moments $\mathbb{E}\big(O_{n,j}^{\underline{s}}\big)$ via

  $$\mathbb{E}\big(O_{n,j}^{\underline{s}}\big) = \frac{(n-j)!}{T_n} [z^{n-j}] \mathrm{U}_v \mathrm{D}_v^s T^{[j]}(z,v).$$

  Thus, by using formula (4.51) for $T^{[j]}(z,v)$, we find

  $$\begin{aligned}
  &\mathbb{E}\big(O_{n,j}^{\underline{s}}\big) = \\
  &= \frac{(n-j)!\; T_j}{T_n} [z^{n-j}] \frac{1}{(1-kz)^{j-1+\frac{1}{k}}} \mathrm{U}_v \mathrm{D}_v^s \left(\frac{1}{1-v+v(1-kz)^{\frac{k-1}{k}}}\right)^{\frac{k}{k-1}} \\
  &= \frac{(n-j)!\; T_j}{T_n} [z^{n-j}] \frac{s!\binom{s+\frac{1}{k-1}}{s}}{(1-kz)^{j-1+\frac{1}{k}}} \frac{\left(1-(1-kz)^{\frac{k-1}{k}}\right)^s}{(1-kz)^{1+\frac{k-1}{k}s}} \\
  &= \frac{(n-j)!\; T_j\; k^{n-j}}{T_n} [z^{n-j}] \sum_{\ell=0}^{s} (-1)^\ell \binom{s}{\ell} \frac{s!\binom{s+\frac{1}{k-1}}{s}}{(1-z)^{j+\frac{(k-1)s+1}{k}-\frac{k-1}{k}\ell}} \\
  &= \frac{s!\binom{s+\frac{1}{k-1}}{s}}{\binom{n-\frac{k-1}{k}}{n-j}} \sum_{\ell=0}^{s} (-1)^\ell \binom{s}{\ell} \binom{n+\frac{(k-1)(s-\ell-1)}{k}}{n-j}.
  \end{aligned} \tag{4.55}$$

  For $n \to \infty$ one further gets, by writing $\binom{n+\frac{(k-1)(s-\ell-1)}{k}}{n-j}$ and $\binom{n-\frac{k-1}{k}}{n-j}$ as a quotient of Gamma functions and using Lemma 2.2.5,

$$\begin{aligned}
&\mathbb{E}\big(O_{n,j}^{\underline{s}}\big) = \\
&= s!\binom{s+\frac{1}{k-1}}{s}\sum_{\ell=0}^{s}(-1)^{\ell}\binom{s}{\ell}\frac{\Gamma\Big(n+\frac{(k-1)(s-\ell)+1}{k}\Big)\,\Gamma\big(j+\frac{1}{k}\big)}{\Gamma\big(n+\frac{1}{k}\big)\,\Gamma\Big(j+\frac{(k-1)(s-\ell)+1}{k}\Big)} \\
&= s!\binom{s+\frac{1}{k-1}}{s}\sum_{\ell=0}^{s}(-1)^{\ell}\binom{s}{\ell}n^{\frac{(k-1)(s-\ell)}{k}}\frac{\Gamma\big(j+\frac{1}{k}\big)}{\Gamma\Big(j+\frac{(k-1)(s-\ell)+1}{k}\Big)}\left(1+\mathcal{O}\Big(\frac{1}{n}\Big)\right) \\
&= s!\binom{s+\frac{1}{k-1}}{s}n^{\frac{(k-1)s}{k}}\frac{\Gamma\big(j+\frac{1}{k}\big)}{\Gamma\Big(j+\frac{(k-1)s+1}{k}\Big)}\left(1+\mathcal{O}\Big(\frac{1}{n^{\frac{k-1}{k}}}\Big)\right).
\end{aligned} \tag{4.56}$$

Now, using the relation (2.6) between the ordinary and factorial moments, we get from (4.56) also the asymptotic expansion

$$\mathbb{E}\big(O_{n,j}^{s}\big) = s!\binom{s+\frac{1}{k-1}}{s}n^{\frac{(k-1)s}{k}}\frac{\Gamma\big(j+\frac{1}{k}\big)}{\Gamma\Big(j+\frac{(k-1)s+1}{k}\Big)}\left(1+\mathcal{O}\Big(\frac{1}{n^{\frac{k-1}{k}}}\Big)\right).$$

Hence, the moments of $n^{-\frac{k-1}{k}}O_{n,j}$ converge to the moments of a random variable O_j, i.e. $\lim_{n\to\infty}\mathbb{E}\Big(\Big(n^{-\frac{k-1}{k}}O_{n,j}\Big)^{s}\Big) = \mathbb{E}\big(O_j^s\big)$, where

$$\mathbb{E}\big(O_j^s\big) = s!\binom{s+\frac{1}{k-1}}{s}\frac{\Gamma\big(j+\frac{1}{k}\big)}{\Gamma\Big(j+\frac{(k-1)s+1}{k}\Big)} = \frac{\Gamma\big(s+\frac{k}{k-1}\big)\,\Gamma\big(j+\frac{1}{k}\big)}{\Gamma\big(\frac{k}{k-1}\big)\,\Gamma\Big(j+\frac{(k-1)s+1}{k}\Big)}.$$

Simple growth estimates for these moments show that the moment generating function $\sum_{s\geq 0}\mathbb{E}\big(O_j^s\big)\frac{t^s}{s!}$ has a positive radius of convergence, and by Lemma 2.2.2 we conclude that the sequence of moments $(\mathbb{E}\big(O_j^s\big))_{s\in\mathbb{N}}$ uniquely determines the distribution of O_j. Thus an application of the Theorem of Fréchet and Shohat (Theorem 2.2.1) proves the convergence in distribution of $n^{-\frac{k-1}{k}}O_{n,j}$ to O_j.

- *The region for j small: $j\to\infty$ such that $j = o\,(n)$:*

As in the case for fixed j we use the method of moments: We first write $\binom{n+\frac{(k-1)(s-\ell-1)}{k}}{n-j}$ and $\binom{n-\frac{k-1}{k}}{n-j}$ in the last expression of (4.55) as quotients of

Gamma functions, and then apply Lemma 2.2.5 twice,

$$\begin{aligned}
\mathbb{E}\big(O_{n,j}^{\underline{s}}\big) &= s!\binom{s+\frac{1}{k-1}}{s}\sum_{\ell=0}^{s}(-1)^\ell\binom{s}{\ell}\frac{\Gamma\big(j+\frac{1}{k}\big)\,\Gamma\Big(n+\frac{(k-1)(s-\ell)+1}{k}\Big)}{\Gamma\big(n+\frac{1}{k}\big)\,\Gamma\Big(j+\frac{(k-1)(s-\ell)+1}{k}\Big)}\\
&= s!\binom{s+\frac{1}{k-1}}{s}\sum_{\ell=0}^{s}(-1)^\ell\binom{s}{\ell}\frac{n^{\frac{(k-1)(s-\ell)}{k}}\,\big(1+\mathcal{O}\big(\frac{1}{n}\big)\big)}{j^{\frac{(k-1)(s-\ell)}{k}}\,\Big(1+\mathcal{O}\Big(\frac{1}{j}\Big)\Big)}\\
&= s!\binom{s+\frac{1}{k-1}}{s}\left(\frac{n}{j}\right)^{\frac{k-1}{k}s}\left(1+\mathcal{O}\left(\frac{1}{j}\right)\right)\left(1+\mathcal{O}\left(\left(\frac{j}{n}\right)^{\frac{k-1}{k}}\right)\right).
\end{aligned}$$

From this we get, again using the relation (2.6) between factorial and ordinary moments, also the asymptotic expansion

$$\mathbb{E}\big(O_{n,j}^{s}\big) = s!\binom{s+\frac{1}{k-1}}{s}\left(\frac{n}{j}\right)^{\frac{k-1}{k}s}\left(1+\mathcal{O}\left(\frac{1}{j}\right)\right)\left(1+\mathcal{O}\left(\left(\frac{j}{n}\right)^{\frac{k-1}{k}}\right)\right).$$

Hence, it follows that

$$\lim_{\substack{n,j\to\infty\\ j=o(n)}} \mathbb{E}\left(\left(\left(\frac{j}{n}\right)^{\frac{k-1}{k}} O_{n,j}\right)^s\right) = s!\binom{s+\frac{1}{k-1}}{s} = \frac{\Gamma\big(s+\frac{k}{k-1}\big)}{\Gamma\big(\frac{k}{k-1}\big)},$$

i.e. for each $s \in \mathbb{N}$ the s-th moment of $\big(\frac{j}{n}\big)^{\frac{k-1}{k}} O_{n,j}$ converges to the s-th moment of a Gamma$(\frac{k-1}{k}, 1)$ distributed random variable. Since the Gamma distribution is uniquely determined by its sequence of moments, the claimed convergence result thus follows directly from Theorem 2.2.1.

- *The central region for j: $j \to \infty$ such that $j \sim \rho n$, with $0 < \rho < 1$:*

  We use the exact expressions for the probabilities $\mathbb{P}\{O_{n,j} = m\}$ given in Theorem 4.8.1, which can be written as

$$\begin{aligned}
&\mathbb{P}\{O_{n,j} = m\} =\\
&= \binom{m+\frac{1}{k-1}}{m}\sum_{\ell=0}^{m}(-1)^\ell\binom{m}{\ell}\frac{\Gamma\Big(n+1-\frac{(k-1)\ell+2k-1}{k}\Big)\,\Gamma\big(j+1-\frac{k-1}{k}\big)}{\Gamma\big(n+1-\frac{k-1}{k}\big)\,\Gamma\Big(j+1-\frac{(k-1)\ell+2k-1}{k}\Big)}.
\end{aligned}$$

Since we have $n \to \infty$ and $j \to \infty$ we can apply Lemma 2.2.5 twice and then get, for arbitrary but fixed $m \in \mathbb{N}_0$,

$$\begin{aligned}
\mathbb{P}\{O_{n,j} = m\} &= \binom{m + \frac{1}{k-1}}{m} \sum_{\ell=0}^{m} (-1)^{\ell} \binom{m}{\ell} \left(\frac{j}{n}\right)^{\frac{(k-1)\ell}{k}+1} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \\
&= \binom{m + \frac{1}{k-1}}{m} \sum_{\ell=0}^{m} (-1)^{\ell} \binom{m}{\ell} \rho^{\frac{(k-1)\ell}{k}+1} (1 + o(1)) \\
&= \binom{m + \frac{k}{k-1} - 1}{m} \rho \left(1 - \rho^{\frac{k-1}{k}}\right)^{m} (1 + o(1)),
\end{aligned}$$

where we have used the fact that $j = \rho n (1 + o(1))$. Hence, for $n \to \infty$ and $j \to \infty$ such that $j \sim \rho n$ the probabilities $\mathbb{P}\{O_{n,j} = m\}$ converge to the probabilities $\mathbb{P}\{O_\rho = m\}$, where O_ρ is a random variable which is $\mathrm{NegBin}(r, p)$ distributed with parameters $r = \frac{k}{k-1}$ and $p = \rho^{\frac{k-1}{k}}$, which proves the convergence in distribution of $O_{n,j}$ to O_ρ as claimed in Theorem 4.8.2.

- *The region for j large: $j \to \infty$ such that $\tilde{j} := n - j = o(n)$:*

  From the exact formula for $\mathbb{P}\{O_{n,j} = 0\}$ in Theorem 4.8.1 we immediately obtain the result stated in Theorem 4.8.2:

$$\mathbb{P}\{O_{n,j} = 0\} = \frac{\binom{n-1-\frac{k-1}{k}}{n-j}}{\binom{n-\frac{k-1}{k}}{n-j}} = \frac{j - \frac{k-1}{k}}{n - \frac{k-1}{k}} = 1 - \frac{\tilde{j}}{n - \frac{k-1}{k}} = 1 + o(1).$$

Family $\mathcal{T}^{[o]}$ (ordered increasing k-trees): The computations for this family are completely analogous to the ones for $\mathcal{T}^{[u]}$. First, one extracts the coefficient of z^{n-j} from the function $U_v D_v^s T^{[j]}(z, v)$ in order to obtain the auxiliary result

$$\mathbb{E}\big(O_{n,j}^{\underline{s}}\big) = \frac{s!}{\binom{n-\frac{k}{k+1}}{n-j}} \sum_{\ell=0}^{s} (-1)^{\ell} \binom{s}{\ell} \binom{n + \frac{k(s-\ell-1)}{k+1}}{n-j}. \tag{4.57}$$

From this one eventually gets, for j fixed,

$$\mathbb{E}\big(O_{n,j}^{s}\big) = s! \frac{\Gamma\big(j + \frac{1}{k+1}\big)}{\Gamma\big(j + \frac{ks+1}{k+1}\big)} n^{\frac{k}{k+1}s} \left(1 + \mathcal{O}\left(\frac{1}{n^{\frac{k}{k+1}}}\right)\right),$$

which, by Theorem 2.2.1, implies the convergence in distribution of $n^{-\frac{k}{k+1}} O_{n,j}$ to a random variable O_j whose distribution is uniquely determined by the moments

$$\mathbb{E}\big(O_{n,j}^s\big) = s! \frac{\Gamma\big(j + \frac{1}{k+1}\big)}{\Gamma\big(j + \frac{ks+1}{k+1}\big)}.$$

For $j \to \infty$ such that $j = o(n)$, one gets from (4.57)

$$\lim_{\substack{n,j\to\infty \\ j=o(n)}} \mathbb{E}\left(\left(\left(\frac{j}{n}\right)^{\frac{k}{k+1}} O_{n,j}\right)^s\right) = s!,$$

i.e. the s-th moments of $\left(\frac{j}{n}\right)^{\frac{k}{k+1}} O_{n,j}$ converge to the s-th moments of an Exp(1) distributed random variable. Thus the convergence in distribution claimed in Theorem 4.8.2 follows again from the Theorem of Fréchet and Shohat.

For the region where $j \sim \rho n$, $0 < \rho < 1$, one studies the exact expressions for the probabilities $\mathbb{P}\{O_{n,j} = m\}$ and derives

$$\lim_{\substack{n,j\to\infty \\ j\sim\rho n}} \mathbb{P}\{O_{n,j} = m\} = \rho^{\frac{k}{k+1}} \left(1 - \rho^{\frac{k}{k+1}}\right)^m,$$

which proves the convergence in distribution of $O_{n,j}$ to a Geom$(\rho^{\frac{k}{k+1}})$ distributed random variable in this case.

Finally, for $n \to \infty$ and $n - j = o(n)$, one simply uses the exact formula for $\mathbb{P}\{O_{n,j} = 0\}$ in order to show that $\mathbb{P}\{O_{n,j} = 0\} \to 1$.

Family $\mathcal{T}^{[d]}$ (d-ary increasing k-trees): The convergence results given in Theorem 4.8.2 can be proven for this family completely analogously to the previous two cases. First, one considers the derivatives of $T^{[j]}(z, v)$ with respect to v evaluated at $v = 1$ in order to get the auxiliary result

$$\mathbb{E}\big(O_{n,j}^{\underline{s}}\big) = \frac{s!\binom{s+\frac{d+1}{(k-1)d-1}}{s}}{\binom{n-\frac{(k-1)d-1}{kd-1}}{n-j}} \sum_{\ell=0}^{s} (-1)^\ell \binom{s}{\ell} \binom{n + \frac{((k-1)d-1)(s-\ell-1)}{kd-1}}{n-j}.$$

From this one derives, for the regions where j is fixed or $j \to \infty$, $j = o(n)$, asymptotic formulæ for the moments $\mathbb{E}\big(O_{n,j}^s\big)$ in order to apply the method of moments. For the regions where $j \sim \rho n$, $0 < \rho < 1$, and $n - j = o(n)$, one studies directly the exact formula for the probabilities $\mathbb{P}\{O_{n,j} = m\}$ as given in Theorem 4.8.1.

4.8.2.3 Derivation of the results for a randomly selected node

To obtain our results for $\bar{O}_n$, i.e. the out-degree of a randomly selected inserted node in a random increasing k-tree of size n, as stated in Theorem 4.8.3, we use $\bar{O}_n \stackrel{(d)}{=} O_{n,U_n}$, with $U_n \stackrel{(d)}{=} \text{Uniform}(\{1,2,\dots,n\})$, and the exact results for $O_{n,j}$ given in Theorem 4.8.1.

Family $\mathcal{T}^{[u]}$ (unordered increasing k-trees): We compute

$$\begin{aligned}
\mathbb{P}\{\bar{O}_n = m\} &= \frac{1}{n}\sum_{j=1}^{n}\mathbb{P}\{O_{n,j}=m\} \\
&= \frac{\binom{m+\frac{1}{k-1}}{m}}{n}\sum_{\ell=0}^{m}(-1)^{\ell}\binom{m}{\ell}\sum_{j=1}^{n}\frac{\binom{n-\frac{(k-1)\ell+2k-1}{k}}{n-j}}{\binom{n-\frac{k-1}{k}}{n-j}} \\
&= \frac{\binom{m+\frac{1}{k-1}}{m}}{n}\sum_{\ell=0}^{m}(-1)^{\ell}\frac{\binom{m}{\ell}}{\binom{n-\frac{k-1}{k}}{\frac{(k-1)\ell}{k}+1}}\sum_{j=1}^{n}\binom{j-\frac{k-1}{k}}{\frac{(k-1)\ell}{k}+1} \\
&= \frac{\binom{m+\frac{1}{k-1}}{m}}{n}\sum_{\ell=0}^{m}(-1)^{\ell}\frac{\binom{m}{\ell}}{\binom{n-\frac{k-1}{k}}{\frac{(k-1)\ell}{k}+1}}\left(\binom{n+\frac{1}{k}}{\frac{(k-1)\ell}{k}+2}-\binom{\frac{1}{k}}{\frac{(k-1)\ell}{k}+2}\right),
\end{aligned}$$

where we have used (4.26) in order to carry out the summation over j in the last step. From this one gets by simple manipulations the formula for the probabilities $\mathbb{P}\{\bar{O}_n = m\}$ which is stated in Theorem 4.8.3. For $n \to \infty$ one further gets, by an application of Stirling's formula,

$$\begin{aligned}
\mathbb{P}\{\bar{O}_n = m\} &= \binom{m+\frac{1}{k-1}}{m}\sum_{\ell=0}^{m}(-1)^{\ell}\binom{m}{\ell}\frac{1}{\frac{(k-1)\ell}{k}+2}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \\
&= \frac{k\binom{m+\frac{1}{k-1}}{m}}{(k-1)(m+1)\binom{m+2+\frac{2}{k-1}}{m+1}}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right),
\end{aligned}$$

where we have used (C.3) to compute the sum. Hence $\bar{O}_n$ converges in distribution to a random variable $\bar{O}$ with

$$\mathbb{P}\{\bar{O} = m\} = \frac{k\binom{m+\frac{1}{k-1}}{m}}{(k-1)(m+1)\binom{m+2+\frac{2}{k-1}}{m+1}},$$

for $m \in \mathbb{N}_0$. By another application of Stirling's formula, one finally gets that

$$\mathbb{P}\{\bar{O} = m\} \sim \frac{k\Gamma\left(\frac{2k}{k-1}\right)}{\Gamma\left(\frac{1}{k-1}\right)}m^{-(2+\frac{1}{k-1})}, \quad \text{for } m \to \infty,$$

which shows that $\bar{O}$ follows asymptotically a power-law distribution with parameter $2 + \frac{1}{k-1}$. We remark that this is in accordance with the result of Gao [Gao09].

Family $\mathcal{T}^{[o]}$ (ordered increasing k-trees) and Family $\mathcal{T}^{[d]}$ (d-ary increasing k-trees): The computations for these two families are completely analogous, hence we skip them.

4.8.2.4 Proof of Corollary 4.8.4

The crucial observation for analyzing the clustering coefficient in k-trees is the following:

Lemma 4.8.5. *For any k-tree T the local clustering coefficient $C_T(u)$ of a node u only depends on the degree $d(u)$ of u: For $d(u) \geq k \geq 2$, one has*

$$C_T(u) = \frac{2(k-1)}{d(u)} - \frac{(k-1)(k-2)}{d(u)(d(u)-1)}. \tag{4.58}$$

Proof. To show this we will, according to Definition (4.10), count the number $M(u)$ of edges between neighbours of u: Consider a node u in a k-tree. Then it always holds that $d(u) \geq k-1$. If $d(u) = k-1$ then the k-tree can only consist of a single k-clique, and thus all $k-1$ neighbours of u are connected with each other, which implies $M(u) = \binom{k-1}{2}$. In order to determine $M(u)$ when $d(u) \geq k$ we observe that in any k-tree holds that when increasing the degree of a node u by 1 then the number of edges between neighbours of u increases exactly by $k-1$. This holds since a new node w adjacent to u generates a k-clique, such that w is also adjacent to $k-1$ neighbours of u. Thus $M(u) = \binom{k-1}{2} + (k-1)(d(u)-k+1)$, for $d(u) \geq k-1$.

This shows that the clustering coefficient of every node u in any k-tree T is given by $C_T(u) = \frac{\binom{k-1}{2}+(k-1)(d(u)-k+1)}{\binom{d(u)}{2}}$, which is equivalent to formula (4.58). □

Due to the above lemma, the random variable C_n which denotes the local clustering coefficient of a random node in a random increasing k-tree of size n satisfies

$$C_n \stackrel{(d)}{=} \frac{2(k-1)}{\tilde{O}_n} - \frac{(k-1)(k-2)}{\tilde{O}_n(\tilde{O}_n-1)},$$

where $\tilde{O}_n$ denotes the degree of a randomly selected node (amongst the root nodes and the inserted nodes) in a random increasing k-tree of size n. Of course, if we denote by $\mathcal{B}_n$ the event that the chosen node is an inserted node, it holds that

$$\begin{aligned}\mathbb{P}\Big\{\tilde{O}_n = m\Big\} &= \mathbb{P}\Big\{\tilde{O}_n = m \mid \mathcal{B}_n\Big\}\,\mathbb{P}\{\mathcal{B}_n\} + \mathbb{P}\Big\{\tilde{O}_n = m \mid \mathcal{B}_n^c\Big\}\,\mathbb{P}\{\mathcal{B}_n^c\}\\ &= \mathbb{P}\big\{\bar{O}_n + k = m\big\}\,\mathbb{P}\{\mathcal{B}_n\} + \mathbb{P}\Big\{\tilde{O}_n = m \mid \mathcal{B}_n^c\Big\}\,\mathbb{P}\{\mathcal{B}_n^c\},\end{aligned}$$

for $m \ge k$. Due to the fact that $\lim_{n\to\infty} \mathbb{P}\{\mathcal{B}_n\} = 1$ and $\bar{O}_n \xrightarrow{(d)} \bar{O}$, this shows that $\tilde{O}_n \xrightarrow{(d)} \bar{O} + k$. Consequently, since the function $f(m) = \frac{2(k-1)}{m} - \frac{(k-1)(k-2)}{m(m-1)}$ is uniformly bounded for $m \ge k$, it immediately follows that

$$\lim_{n\to\infty} \mathbb{E}(C_n) = \lim_{n\to\infty} \mathbb{E}\Big(f(\tilde{O}_n)\Big) = c_k := \mathbb{E}\big(f(\bar{O}+k)\big) = \sum_{m\ge 0} \mathbb{P}\big\{\bar{O} = m\big\}\, f(m+k). \tag{4.59}$$

Using the result on the distribution of $\bar{O}$ (see Theorem 4.8.3) in each of the cases $\mathcal{T}^{[u]}$, $\mathcal{T}^{[o]}$ and $\mathcal{T}^{[d]}$, respectively, this leads to the expressions for c_k given in Corollary 4.8.4.

Finally, it is easy to see that $c_k \to 1$ as $k \to \infty$: Since in (4.59) all summands $\mathbb{P}\big\{\bar{O} = m\big\}\, f(m+k)$ are positive, we certainly have

$$\lim_{k\to\infty} c_k \ge \lim_{k\to\infty} \sum_{m=0}^{M} \mathbb{P}\big\{\bar{O} = m\big\}\, f(m+k),$$

for all $M \in \mathbb{N}$. Furthermore, for fixed m one clearly has

$$\lim_{k\to\infty} \mathbb{P}\big\{\bar{O} = m\big\}\, f(m+k) = \frac{1}{(m+1)(m+2)},$$

in all three cases, which implies

$$\lim_{k\to\infty} c_k \ge \lim_{M\to\infty} \sum_{m=0}^{M} \frac{1}{(m+1)(m+2)} = 1.$$

One the other hand, one trivially has $c_k \le 1$, and hence $c_k \to 1$.

CHAPTER 5

Local minima in trees

5.1 Introduction

Given a tree T which is labelled with distinct integers, a *local minimum* in T is a node v (as usual, we always identify each node with its label), which has the property that each neighbour (i.e. adjacent node) w of v satisfies $w > v$ (see Figure 5.1 for an example). Analogously, a node v in T is a local maximum if every neighbour w of v satisfies $w < v$. Of course, these definitions make sense for arbitrary trees, however, we will in this chapter restrict our attention to rooted ones. Moreover, as usual, we will assume that each tree T is labelled exactly with the integers $\{1, \ldots, |T|\}$, where $|T|$ denotes the number of nodes of T.

For the special case of permutations (which can of course be interpreted as "linear" labelled rooted trees), the number of local minima (and maxima) has

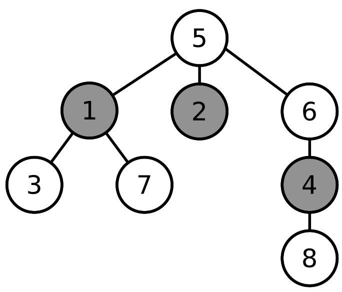

Figure 5.1: A tree with 3 local minima (marked gray).

already been studied extensively (in this context, local minima are mostly referred to as "troughs", and local maxima as "peaks"): In a series of papers published between 1879 and 1896 (described in [Net01]), André considered the number $P_{n,s}$ of permutations of length n which consist of exactly s monotone runs (e.g., the permutation $(3, 1, 4, 7, 6, 5, 2)$ consists of the monotone runs $(3, 1)$, $(1, 4, 7)$, and $(7, 6, 5, 2)$), and derived a recurrence formula for these numbers. This is of course closely related to counting permutations by their local minima and maxima, since $P_{n,s}$ equals the number of permutations of length n which have a total number of $s+1$ local minima and maxima. In random permutations, the total number of local minima and maxima is asymptotically normally distributed, which was asserted by David and Barton [DB62, Chapter 10] (although without detailed proof).

Carlitz [Car74] considered only the number of local maxima in permutations, and derived a generating function for the numbers $M(n, k)$ of permutations of length n with exactly k local maxima. Stigler [Sti86] studied the number of local maxima from the probabilistic point of view, and argued that this number can be used to estimate the correlation of serial data. Warren and Seneta [WS96] and Esseen [Ess84] later independently showed that the number of local maxima (and thus also the number of local minima) in random permutations of length n is for $n \to \infty$ asymptotically normally distributed with expectation and variance both of order n.

Studying the number of local minima (or maxima) also in trees seems to be a very natural extension of the above work. Thus, we will in this chapter consider the number of local minima in the two tree families of unordered and ordered trees (cf. Section 3.2.1), respectively. In particular, we consider the random variable M_n, which counts the number of local minima in a random tree of size n, where we always assume that each tree of size n of the respective tree family is chosen with the same probability. For both tree families we show that, in analogy to the corresponding result for permutations, M_n is for $n \to \infty$ asymptotically normally distributed with expectation and variance both of order n. However, we do not only obtain limiting distribution results, but can also characterize the exact probability distribution of M_n for fixed n. Note that studying M_n is equivalent to studying the number $\tilde{M}_n$ of local maxima in a random tree of size n: One clearly has $\tilde{M}_n \stackrel{(d)}{=} M_n$, since the relabelling $j \mapsto n + 1 - j$ maps each tree with m local minima to a tree with m local maxima and vice versa.

Apart from our probabilistic considerations, we will also show that there are interesting connections between the numbers $T_{n,m}$ of trees of size n with m local minima and certain other combinatorial quantities involving the same tree families, namely the number $L_{n,m}$ of unordered trees of size n with m leaves, and the number F_n of up-down alternating trees of size n in the respective tree family.

5.2 Results

5.2.1 Unordered trees

Theorem 5.2.1. *For unordered trees, the random variable M_n which counts the number of local minima of a random tree of size n has the following exact distribution:*

$$\mathbb{P}\{M_n = m\} = \frac{1}{n^{n-1}} \binom{n}{m} \sum_{j=0}^{n-m} (-1)^{n-m-j} \binom{n-m}{j} j^{n-1},$$

for $1 \le m \le n$.

Corollary 5.2.2. *The number $T_{n,m}$ of unordered trees with n nodes and m local minima is equal to the number $L_{n,m}$ of unordered trees with n nodes and m leaves.*

Corollary 5.2.3. *For unordered trees, the random variable M_n which counts the number of local minima of a random tree of size n is asymptotically normally distributed: There holds*

$$\mu_n := \mathbb{E}(M_n) \sim \frac{n}{\mathrm{e}}, \quad \sigma_n^2 := \mathbb{V}(M_n) \sim \left(\frac{1}{\mathrm{e}} - \frac{2}{\mathrm{e}^2}\right) n,$$

and

$$\frac{M_n - \mu_n}{\sigma_n} \xrightarrow{(d)} \mathcal{N}(0,1).$$

5.2.2 Ordered trees

Theorem 5.2.4. *For ordered trees, the random variable M_n which counts the number of local minima of a random tree of size n equals in distribution a sum of independent Bernoulli distributed random variables: There holds*

$$M_n \stackrel{(d)}{=} M_n^{(1)} \oplus M_n^{(2)} \oplus \cdots \oplus M_n^{(n)},$$

where $M_n^{(k)}$ is Bernoulli$(\frac{n-k}{n+k-2})$*-distributed for $1 \le k \le n$. Moreover, the random variables $M_n^{(k)}$ satisfy $M_n^{(k)} \stackrel{(d)}{=} \mathbb{1}(\mathcal{M}_n^{(k)})$, where $\mathcal{M}_n^{(k)}$ denotes the event that the node with label k in a tree of size n is a local minimum and $\mathbb{1}(\mathcal{M}_n^{(k)})$ is the indicator variable of this event.*

Theorem 5.2.5. *For ordered trees, the random variable M_n which counts the number of local minima of a random tree of size n is asymptotically normally distributed: There holds*

$$\mu_n := \mathbb{E}(M_n) \sim (-1 + 2\log 2)n, \quad \sigma_n^2 := \mathbb{V}(M_n) \sim (-4 + 6\log 2)n,$$

and

$$\frac{M_n - \mu_n}{\sigma_n} \xrightarrow{(d)} \mathcal{N}(0,1).$$

5.3 Proofs of the results

In our proofs, T_n will always denote the number of trees of size n in the respective tree family, i.e.

$$T_n = \begin{cases} n^{n-1}, & \text{unordered trees,} \\ (n-1)!\binom{2(n-1)}{n-1}, & \text{ordered trees,} \end{cases}$$

which is well-known. By $T_{n,m}$ we denote the number of trees of size n with exactly m local minima in the respective tree family. Moreover, we define the corresponding exponential generating functions $T(z) = \sum_{n\ge1} T_n \frac{z^n}{n!}$ and $T(z,v) = \sum_{n\ge1}\sum_{m\ge0} T_{n,m}\frac{z^n}{n!}v^m$. We also consider the polynomials $T_n(v) := \sum_{m\ge0} T_{n,m}v^m$, which contain the whole information of the distribution of the number M_n of local minima in a random tree of size n: The probability generating function $p_n(v)$ of M_n is clearly given by $\frac{1}{T_n}T_n(v)$.

5.3.1 Unordered trees

5.3.1.1 Exact distribution

In order to prove our results, we use the tree decomposition according to the node with the largest label n which is shown in Figure 5.2: Each tree of size n in which node n has r children is decomposed into the subtrees $T_1, \dots, T_r$ which are rooted at the children of n, and (if n is not the root of T) a tree T_0, which is the part of T lying "above" n, i.e. the part containing the root of T.

Clearly, for $n > 1$, the node with the largest label n cannot be a local minimum. Moreover, each node which lies in some component T_j is a local minimum in T_j iff it is a local minimum in T. Hence, the given decomposition leads to the following

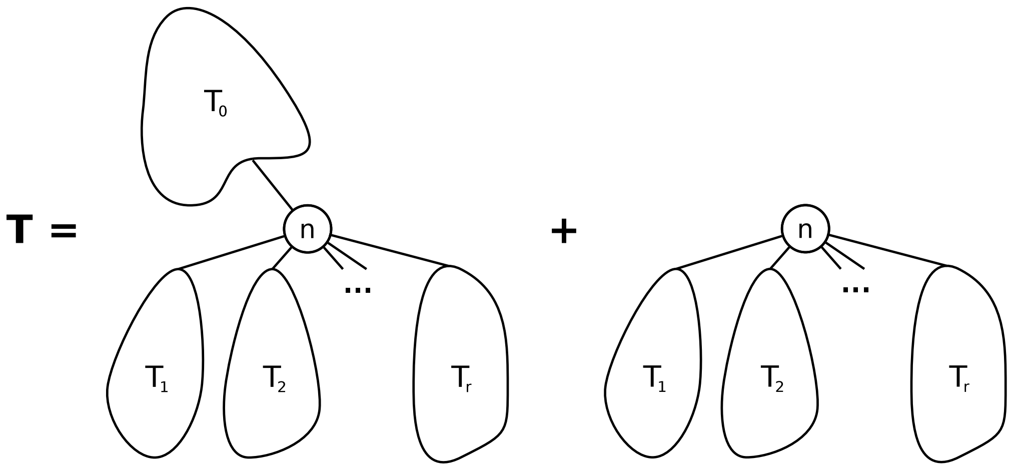

Figure 5.2: Decomposition of unordered trees with respect to the node with the largest label.

recursion for the polynomials $T_n(v) = \sum_{m\geq 0} T_{n,m} v^m$:

$$\begin{aligned} T_n(v) &= \sum_{r\geq 0} \sum_{\substack{n_0,n_1,\dots,n_r\geq 1,\\ n_0+\dots+n_r=n-1}} \binom{n-1}{n_0,\dots,n_r} n_0 T_{n_0}(v) \frac{T_{n_1}(v)\cdots T_{n_r}(v)}{r!} \\ &\quad + \sum_{r\geq 1} \sum_{\substack{n_1,\dots,n_r\geq 1,\\ n_1+\dots+n_r=n-1}} \binom{n-1}{n_1,\dots,n_r} \frac{T_{n_1}(v)\cdots T_{n_r}(v)}{r!}, \quad n\geq 2, \\ T_1(v) &= v. \end{aligned} \tag{5.1}$$

Here, the factors $\binom{n-1}{n_0,\dots,n_r}$ and $\binom{n-1}{n_1,\dots,n_r}$ count all possible ways to distribute the labels $\{1,\dots,n-1\}$ among the components $T_0,\dots,T_r$ and $T_1,\dots,T_r$, respectively, the factor $\frac{1}{r!}$ is needed because we are considering unordered trees (i.e. permuting $T_1,\dots,T_r$ leads to the same tree), and the additional factor n_0 in the first sum accounts for the fact that node n can be attached as a child to any of the n_0 nodes in T_0. Multiplying (5.1) by $\frac{z^{n-1}}{(n-1)!}$ and summing up for $n \geq 1$, one obtains the following differential equation for $T(z,v)$:

$$T_z(z,v) = z\mathrm{e}^{T(z,v)} T_z(z,v) + \mathrm{e}^{T(z,v)} + v - 1, \quad T(0,v) = 0,$$

where the given boundary value follows directly from the definition of $T(z,v)$. This is an exact differential equation whose solution is easily seen to be given by the functional equation

$$T(z,v) = z\left(\mathrm{e}^{T(z,v)} + v - 1\right). \tag{5.2}$$

In order to extract coefficients from this equation, we use Langrange's inversion formula (cf. Theorem 2.1.3). One gets, for $n \geq 1$,

$$
\begin{aligned}
[z^n]T(z,v) &= \frac{1}{n}[T^{n-1}](\mathrm{e}^T + v - 1)^n = \frac{1}{n}[T^{n-1}]\sum_{j=0}^{n}\binom{n}{j}\mathrm{e}^{jT}(v-1)^{n-j} \\
&= \sum_{j=0}^{n}\binom{n}{j}\frac{j^{n-1}}{n!}(v-1)^{n-j},
\end{aligned}
$$

and further, for $1 \leq m \leq n$,

$$
\begin{aligned}
T_{n,m} &= n![z^n v^m]T(z,v) = [v^m]\sum_{j=0}^{n}\binom{n}{j}j^{n-1}(v-1)^{n-j} \\
&= \sum_{j=0}^{n-m}(-1)^{n-m-j}\binom{n}{j}\binom{n-j}{m}j^{n-1} = \binom{n}{m}\sum_{j=0}^{n-m}(-1)^{n-m-j}\binom{n-m}{j}j^{n-1}.
\end{aligned}
$$

Since $\mathbb{P}\{M_n = m\} = \frac{T_{n,m}}{T_n} = \frac{T_{n,m}}{n^{n-1}}$, one thus gets the formula for the probabilities $\mathbb{P}\{M_n = m\}$ which is stated in Theorem 5.2.1.

5.3.1.2 The relation to the number of leaves

If we denote by $L_{n,m}$ the number of unordered trees of size n with exactly m leaf nodes (nodes with out-degree 0) and consider the generating function $L(z,v) = \sum_{n\geq 1}\sum_{m\geq 0} L_{n,m}\frac{z^n}{n!}v^m$ then it is easy to obtain (via the symbolic method) that $L(z,v)$ is given by the functional equation (5.2) (of course, this is well-known, cf., e.g., [DG99], where this is shown in a much more general context). This proves that $T_{n,m} = L_{n,m}$, as stated in Corollary 5.2.2. In the following, we present a bijective proof of this Corollary, which actually shows an even stronger result.

Bijective proof of Corollary 5.2.2: Let $\mathcal{C}_n$ be the set of unordered trees with n nodes. We present a bijective mapping $\phi_n : \mathcal{C}_n \to \mathcal{C}_n$ which maps each tree with m local minima to a tree with m leaves. However, it will be convenient to consider not only the set $\mathcal{C}_n$ of unordered trees of size n which are *well-labelled* (i.e. labelled with the integers $\{1,\ldots,n\}$), instead we consider the set $\bar{\mathcal{C}}_n$ of all *weakly labelled* unordered trees of size n, i.e. unordered trees in which the nodes are labelled by arbitrary distinct integers. We then present a bijective mapping $\bar{\phi}_n : \bar{\mathcal{C}}_n \to \bar{\mathcal{C}}_n$ with the property that each tree with m local minima is mapped to a tree with the same set of labels and m leaves. Thus, the required bijection ϕ_n will then simply be the restriction of $\bar{\phi}_n$ to $\mathcal{C}_n$.

We define the bijections $\bar{\phi}_n$ inductively. Of course, for $n = 1$ the bijection is trivial: Each tree of $\bar{\mathcal{C}}_1$ consists only of one node, which is both a local minimum and a leaf, and thus we simply let $\bar{\phi}_1$ be the identity on $\bar{\mathcal{C}}_1$. So let $n > 1$ and assume that we have already defined the bijections $\bar{\phi}_j, 1 \leq j < n$. We define $\bar{\phi}_n(T)$ for each $T \in \bar{\mathcal{C}}_n$ by the following procedure, where we denote by m the number of local minima in T, by R the label of T's root, and by L the largest label in T:

1) First, re-root T at the node labelled with L. Since this node cannot be a local minimum, the m local minima now lie in the subtrees rooted at the children of L.

2) For $1 \leq j < n$, apply $\bar{\phi}_j$ to every subtree rooted at a child of L which is of size j. Since the total number of local minima in these subtrees is m, the total number of leaves after this step will be m.

3) Swap the labels L and R (i.e. put label R back to the root). The resulting tree is $\bar{\phi}_n(T)$.

One easily checks that $\bar{\phi}_n$ is bijective, since by the induction hypothesis the mappings $\bar{\phi}_j, 1 \leq j < n$ are. Moreover, also by the induction hypothesis, the mappings $\bar{\phi}_j, 1 \leq j < n$ preserve the used set of labels, and thus so does $\bar{\phi}_n$. Hence, the restriction of $\bar{\phi}_n$ to $\mathcal{C}_n$ is a bijection $\phi_n : \mathcal{C}_n \to \mathcal{C}_n$ which maps each tree with m local minima to a tree with m leaves. This concludes the bijective proof of Corollary 5.2.2. For small values of n, the bijections ϕ_n are depicted in Figure 5.3.

We note that the bijection above actually proves an even stronger result: Given a tree $T \in \mathcal{C}_n$, let us denote the number of nodes in T with out-degree j (i.e. with j children) by $d_j^+(T)$ (in particular, $d_0^+(T)$ denotes the number of leaves of T). Moreover, let us write $l_j(T)$ for the number of nodes in T whose label is larger than the labels of exactly j of their neighbours (in particular, $l_0(T)$ gives the number of local minima of T). Then there holds that for arbitrary but fixed $j, m \in \mathbb{N}$, the above bijection $\bar{\phi}_n$ maps each tree $T \in \mathcal{C}_n$ with $l_j(T) = m$ to a tree $\bar{\phi}_n(T)$ with $d_j^+(\bar{\phi}_n(T)) = m$. This fact is trivially true for $n = 1$, and for $n > 1$ it follows by induction, since only one of the following two cases can occur:

- Case 1: The number of neighbours of the node with largest label L in T is different from j. Then, if $l_j(T) = m$, all of the m nodes whose label is larger than the labels of exactly j of their neighbours lie in one of the subtrees rooted at a child of L after step 1) of the construction of $\bar{\phi}_n$. By the induction hypothesis, these subtrees will after step 2) contain a total number

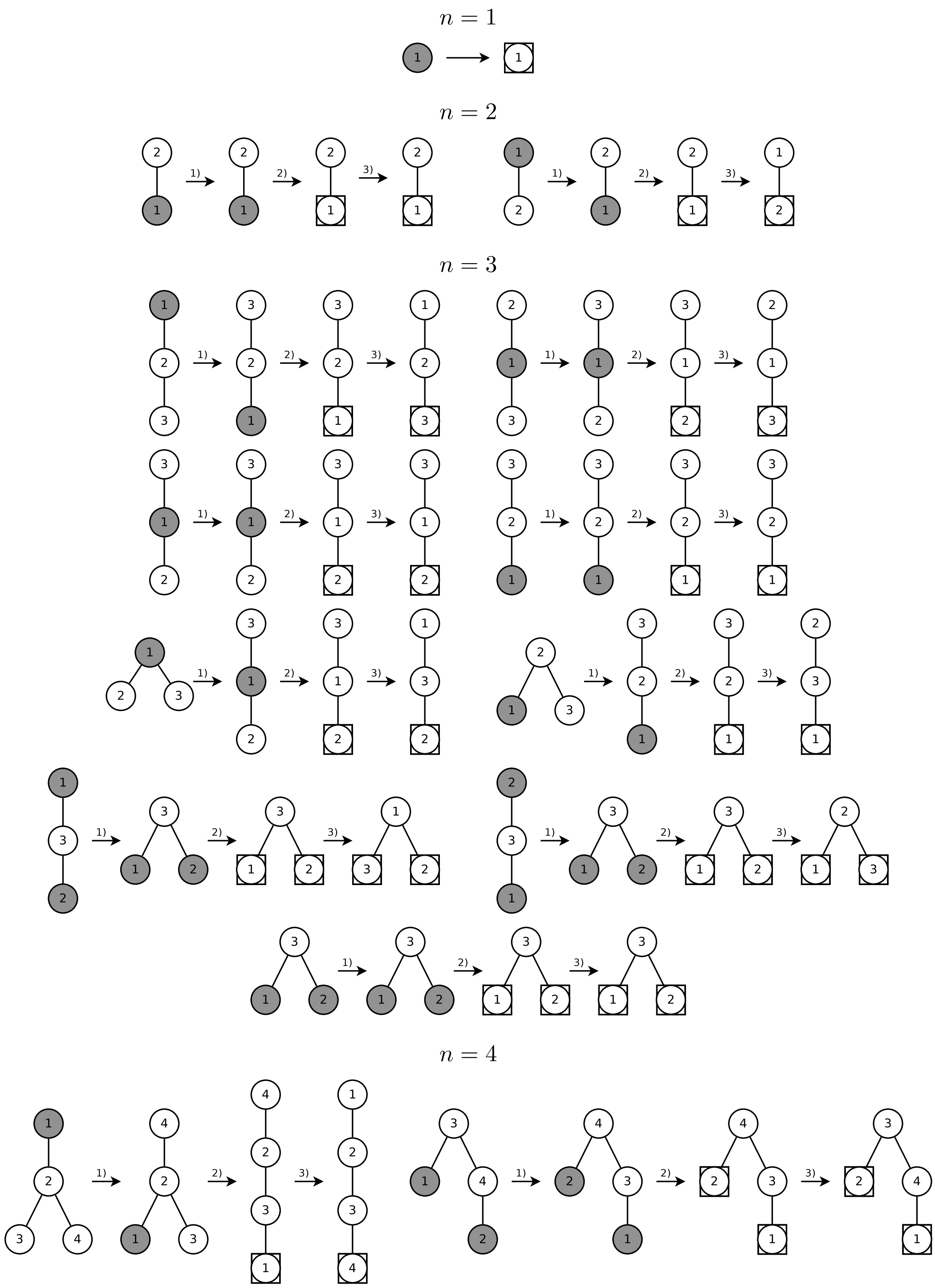

Figure 5.3: Application of ϕ_n to $\mathcal{C}_n$ ($n = 1, 2, 3$) and of ϕ_4 to two trees of $\mathcal{C}_4$. Local minima are marked gray, the resulting leaves are emphasized by boxes.

of m nodes with out-degree j. Moreover, the out-degree of node L (or R, respectively, after step 3)) is different from j. Hence $d_j^+(\bar{\phi}_n(T)) = m$.

- Case 2: The node with the largest label L in T has exactly j neighbours. Then node L is one of those $l_j(T)$ nodes whose label is larger than the labels of exactly j of their neighbours. Hence, if $l_j(T) = m$, then only $m-1$ nodes with the same property lie in the subtrees rooted at the children of L, and after step 2) in the construction of $\bar{\phi}_n$, these subtrees will contain a total number of $m-1$ nodes of out-degree j. In addition, node L itself (R, respectively, after step 3)) has out-degree j, and thus also in this case we see that $d_j^+(\bar{\phi}_n(T)) = m$.

Let us call the sequence $(d_0^+(T), \ldots, d_{n-1}^+(T))$ the *out-degree profile* of T, and $(l_0(T), \ldots, l_{n-1}(T))$ the *label relation profile*. Then the above considerations lead to the following corollary:

Corollary 5.3.1. *For arbitrary $n \in \mathbb{N}$ and $m_0, \ldots, m_{n-1} \in \mathbb{N}_0$, the number of unordered trees of size n with out-degree profile $(m_0, \ldots, m_{n-1})$ is equal to the number of unordered trees of size n with label relation profile $(m_0, \ldots, m_{n-1})$.*

5.3.1.3 Limiting distribution

In [DG99], it was shown (in a much more general context), that the number Z_n of leaves in a random unordered labelled tree of size n is asymptotically normally distributed with mean value $\mu_n := \mathbb{E}(Z_n) \sim \frac{n}{\mathrm{e}}$ and variance $\sigma_n^2 := \mathbb{V}(Z_n) \sim (\frac{1}{\mathrm{e}} - \frac{2}{\mathrm{e}^2})n$. From this and the relation between the number $T_{n,m}$ of unordered trees with m local minima and the number $L_{n,m}$ of unordered trees with m leaves, one immediately gets Corollary 5.2.3.

5.3.2 Ordered trees

5.3.2.1 Derivation of the exact distribution

Again, we use the tree decomposition according to the node with the largest label n. However, in the case of ordered trees, this decomposition is even simpler than in the unordered case (see Figure 5.4): If the total degree of node n in T is r, then removing n from T gives an *ordered forest* $(T_1, \ldots, T_r)$ of r trees. The local minima of T then all lie in these trees T_j, and thus this decomposition leads to the

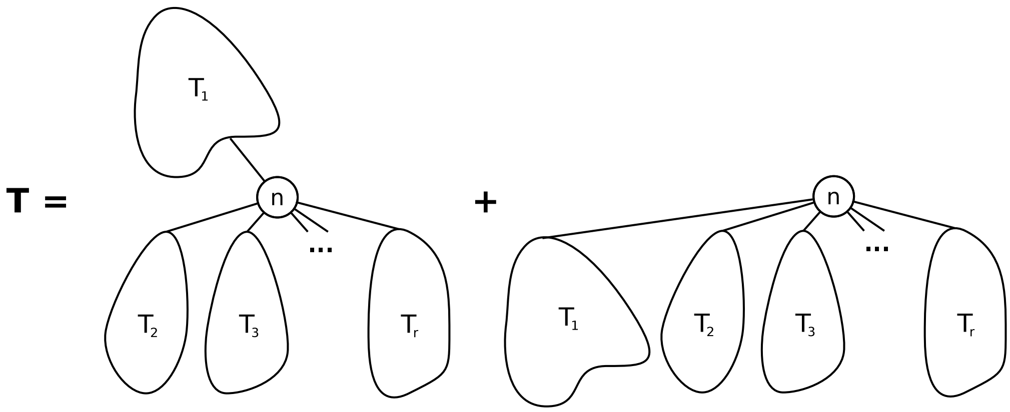

Figure 5.4: Decomposition of ordered trees with respect to the node with the largest label.

following recurrence for the polynomials $T_n(v) = \sum_{m \geq 0} T_{n,m} v^m$:

$$\begin{aligned} T_n(v) &= \sum_{r \geq 1} \sum_{\substack{n_1, \dots, n_r \geq 1, \\ n_1 + \dots + n_r = n-1}} \binom{n-1}{n_1, \dots, n_r} 2n_1 T_{n_1}(v) T_{n_2}(v) \cdots T_{n_r}(v), \quad n \geq 2, \\ T_1(v) &= v. \end{aligned} \tag{5.3}$$

Here, the factor $2n_1$ counts the different possible ways to connect node n to T_1: If T_1 has size n_1, then there are $2n_1 - 1$ possible positions at which n can be attached as a child of one of the nodes of the ordered tree T_1, and moreover one has the possibility to attach the root of T_1 to node n as its leftmost child.

Multiplying (5.3) by $\frac{z^{n-1}}{(n-1)!}$ and summing up for $n \geq 1$ leads to the following differential equation for the generating function $T(z, v)$:

$$T_z(z, v) = \frac{2z T_z(z, v)}{1 - T(z, v)} + v, \quad T(0, v) = 0. \tag{5.4}$$

We will solve this differential equation implicitly by finding functions $F(z, T)$ and $C(v)$ which satisfy $F(z, T(z, v)) = C(v)$. In order to do this, we first write (5.4) as

$$(T + 2z - 1) T_z + v(1 - T) = 0, \tag{5.5}$$

where $T := T(z, v)$. This differential equation is not exact, but we can find an integrating factor μ: Using the ansatz $\mu = \mu(T)$, one gets the differential equation

$$2\mu(T) = \frac{\partial(\mu(T)(T + 2z - 1))}{\partial z} = \frac{\partial(\mu(T) v (1 - T))}{\partial T} = \mu_T(T) v (1 - T) - v\mu(T),$$

for $\mu(T)$, for which one solution is given by $\mu(T) = (1-T)^{-1-\frac{2}{v}}$. By multiplying (5.5) by $\mu(T)$, we now obtain the exact differential equation

$$(2z(1-T)^{-1-\frac{2}{v}} - (1-T)^{-\frac{2}{v}})T_z + v(1-T)^{-\frac{2}{v}} = 0.$$

A potential function of this differential equation is straightforward to obtain and is given by

$$F(z,T) = zv(1-T)^{-\frac{2}{v}} + v\frac{(1-T)^{1-\frac{2}{v}}}{v-2}.$$

Hence, the general solution of (5.5) is given implicitly by

$$C(v) = F(z, T(z,v)) = zv(1-T(z,v))^{-\frac{2}{v}} + v\frac{(1-T(z,v))^{1-\frac{2}{v}}}{v-2},$$

where $C(v)$ is an arbitrary function. In order to solve (5.4), we now use the boundary condition $T(0,v) = 0$, which leads to

$$C(v) = \frac{v}{v-2}.$$

Rearranging the terms a bit, we obtain that the solution of (5.4) is implicitly given by the functional equation

$$z = \frac{1 - T(z,v) - (1-T(z,v))^{\frac{2}{v}}}{2-v}. \tag{5.6}$$

In order to extract the polynomials $T_n(v) = n![z^n]T(z,v)$ from this equation, we apply Lagrange's inversion formula (cf. Theorem 2.1.3). First, we write (5.6) in the following form:

$$z = \frac{T(z,v)}{\frac{(2-v)T(z,v)}{1-T(z,v)-(1-T(z,v))^{\frac{2}{v}}}}.$$

Note that, when expanding the denominator in the right hand side of this equation into a series of T, the constant term in this expansion does not vanish. Thus we can indeed apply Lagrange's inversion formula and obtain

$$\begin{aligned} [z^n]T(z,v) &= \frac{1}{n}[T^{n-1}]\left(\frac{(2-v)T}{1-T-(1-T)^{\frac{2}{v}}}\right)^n \\ &= \frac{(2-v)^n}{n}[T^{-1}]\frac{1}{\left(1-T-(1-T)^{\frac{2}{v}}\right)^n}. \end{aligned}$$

We now use Cauchy's integration formula and get

$$\begin{aligned}
[z^n]T(z,v) &= \frac{(2-v)^n}{n}\frac{1}{2\pi i}\oint \frac{dT}{\left(1-T-(1-T)^{\frac{2}{v}}\right)^n} \\
&= \frac{(2-v)^n}{n}\frac{1}{2\pi i}\oint \frac{1}{(1-T)^n}\frac{1}{\left(1-(1-T)^{\frac{2-v}{v}}\right)^n} dT,
\end{aligned}$$

where the integration is taken in counter-clockwise direction along a simple closed curve around the origin. In order to continue, we use the substitution

$$G = 1-(1-T)^{\frac{2-v}{v}},$$

which gives

$$1-T = (1-G)^{\frac{v}{2-v}} \quad \text{and} \quad dT = \frac{v}{2-v}(1-G)^{-\frac{2(1-v)}{2-v}}dG,$$

and thus

$$\begin{aligned}
[z^n]T(z,v) &= \frac{v(2-v)^{n-1}}{n}\frac{1}{2\pi i}\oint \frac{(1-G)^{-\frac{vn}{2-v}-\frac{2(1-v)}{2-v}}}{G^n}dG \\
&= \frac{v(2-v)^{n-1}}{n}[G^{n-1}]\frac{1}{(1-G)^{\frac{vn+2(1-v)}{2-v}}}.
\end{aligned}$$

Hence, one gets

$$[z^n]T(z,v) = \frac{v(2-v)^{n-1}}{n}\binom{n-1+\frac{v(n-1)}{2-v}}{n-1}$$

and

$$\begin{aligned}
T_n(v) = n![z^n]T(z,v) &= v(2-v)^{n-1}\prod_{k=1}^{n-1}\left(k+\frac{v(n-1)}{2-v}\right) \\
&= v\prod_{k=1}^{n-1}\left((n-k-1)v+2k\right).
\end{aligned}$$

It follows that the probability generating function $p_n(v) = \frac{1}{T_n}T_n(v)$ of the random variable M_n, which counts the number of local minima in a random ordered tree of size n, is given by

$$\begin{aligned}
p_n(v) &= \frac{v}{(n-1)!\binom{2(n-1)}{n-1}}\prod_{k=1}^{n-1}((n-k-1)v+2k) = \frac{v\prod_{k=2}^{n}((n-k)v+2(k-1))}{(2n-2)(2n-3)\cdots n} \\
&= \prod_{k=1}^{n}\left(\frac{n-k}{n+k-2}v+\frac{2(k-1)}{n+k-2}\right).
\end{aligned}$$

If we let, for $1 \leq k \leq n$, $M_n^{(k)}$ be a Bernoulli($\frac{n-k}{n+k-2}$)-distributed random variable, then the k-th factor in $p_n(v)$ is the probability generating function of $M_n^{(k)}$, and hence $M_n \stackrel{(d)}{=} M_n^{(1)} \oplus M_n^{(2)} \oplus \cdots \oplus M_n^{(n)}$, that is, M_n equals in distribution the direct sum of these random variables. This proves the first part of Theorem 5.2.4.

We now show that $M_n^{(k)} \stackrel{(d)}{=} \mathbb{1}(\mathcal{M}_n^{(k)})$, where $\mathcal{M}_n^{(k)}$ denotes the event that the node with label k in a tree of size n is a local minimum and $\mathbb{1}(\mathcal{M}_n^{(k)})$ is the indicator variable of this event, i.e. we prove that $\mathbb{P}\left\{\mathcal{M}_n^{(k)}\right\} = \frac{n-k}{n+k-2}$. We compute this probability by case distinction, i.e. we use the decomposition

$$\mathbb{P}\{\mathcal{M}_n^{(k)}\} = \sum_{m=1}^{n-1} \mathbb{P}\{\mathcal{M}_n^{(k)} \mid k \text{ has total degree } m\} \, \mathbb{P}\{k \text{ has total degree } m\} . \quad (5.7)$$

Here, one clearly has

$$\mathbb{P}\{\mathcal{M}_n^{(k)} \mid k \text{ has total degree } m\} = \frac{\binom{n-k}{m}}{\binom{n-1}{m}}, \quad (5.8)$$

which is simply seen by considering the possible choices for the labels of the neighbours of node k which make k a local minimum. Moreover, one has

$$\mathbb{P}\{k \text{ has total degree } m\} = \frac{D_{n,k,m}}{T_n},$$

where $D_{n,k,m}$ denotes the number of ordered trees of size n in which the node with label k has total degree m. This number is of course independent of the label k, and it is thus sufficient to compute the numbers $D_{n,n,m}$. For the function $D(z,v) := \sum_{n\geq 1} \sum_{m \geq 1} D_{n,n,m} \frac{z^n}{n!} v^m$ one gets, again by using the decomposition shown in Figure 5.4,

$$\frac{D_z(z,v)}{z} = \frac{2vT'(z)}{1-vT(z)}.$$

From this equation we can, for $n \geq 2$ and $m \geq 1$, extract the numbers $D_{n,n,m}$ in the following way: First, one has

$$\begin{aligned}
\frac{D_{n,n,m}}{(n-1)!} &= [z^{n-1}v^m] D_z(z,v) = [z^{n-2}v^m] \frac{D_z(z,v)}{z} = (n-1)[z^{n-1}v^m] \int_0^z \frac{D_\zeta(\zeta,v)}{\zeta} d\zeta \\
&= (n-1)[z^{n-1}v^m] \int_0^z \frac{2vT'(\zeta)}{1-vT(\zeta)} d\zeta = 2(n-1)[z^{n-1}v^m] \log\left(\frac{1}{1-vT(z)}\right) \\
&= 2(n-1)[z^{n-1}] \frac{T(z)^m}{m},
\end{aligned}$$

and by an application of Lagrange's inversion formula (remember that $T(z) = \frac{z}{1-T(z)}$, cf. equation (3.3)), one further obtains

$$\begin{aligned} \frac{D_{n,n,m}}{(n-1)!} &= 2(n-1)[T^{n-2}]\frac{1}{n-1}T^{m-1}\frac{1}{(1-T)^{n-1}} \\ &= 2[T^{n-m-1}]\frac{1}{(1-T)^{n-1}} = 2\binom{2n-m-3}{n-2}. \end{aligned} \tag{5.9}$$

Now, by using (5.8) and (5.9) in (5.7), the probabilities $\mathbb{P}\Big\{\mathcal{M}_n^{(k)}\Big\}$ can be computed as follows:

$$\begin{aligned} \mathbb{P}\{\mathcal{M}_n^{(k)}\} &= \sum_{m=1}^{n-1} \frac{\binom{n-k}{m}}{\binom{n-1}{m}} \frac{D_{n,n,m}}{T_n} = \sum_{m=1}^{n-k} \frac{\binom{n-k}{m}}{\binom{n-1}{m}} \frac{2(n-1)!\binom{2n-m-3}{n-2}}{(n-1)!\binom{2n-2}{n-1}} \\ &= \frac{1}{\binom{2n-3}{n+k-3}} \sum_{m=1}^{n-k} \binom{2n-m-3}{n+k-3} = \frac{\binom{2n-3}{n+k-2}}{\binom{2n-3}{n+k-3}} = \frac{n-k}{n+k-2}. \end{aligned}$$

Hence, we have shown that $M_n \stackrel{(d)}{=} M_n^{(1)} \oplus M_n^{(2)} \oplus \cdots \oplus M_n^{(n)}$, where $M_n^{(k)} \stackrel{(d)}{=} \mathbb{1}(\mathcal{M}_n^{(k)})$ and $\mathcal{M}_n^{(k)}$ denotes the event that the node with label k is a local minimum. This concludes the proof of Theorem 5.2.4.

Remark: Naturally, one has $M_n = \mathbb{1}(\mathcal{M}_n^{(1)}) + \mathbb{1}(\mathcal{M}_n^{(2)}) + \ldots + \mathbb{1}(\mathcal{M}_n^{(n)})$. The fact that $M_n \stackrel{(d)}{=} M_n^{(1)} \oplus M_n^{(2)} \oplus \cdots \oplus M_n^{(n)}$ suggests that the events $\mathcal{M}_n^{(k)}$ may actually be independent, i.e. that one has $M_n = \mathbb{1}(\mathcal{M}_n^{(1)}) \oplus \mathbb{1}(\mathcal{M}_n^{(2)}) \oplus \ldots \oplus \mathbb{1}(\mathcal{M}_n^{(n)})$. We strongly believe that this is the case (it can also be verified computationally for small values of n), but unfortunately we were not able to prove this for general n.

5.3.2.2 Derivation of the limiting distribution result

We derive the limiting distribution result for M_n directly from the exact results: Since $M_n \stackrel{(d)}{=} M_n^{(1)} \oplus M_n^{(2)} \oplus \cdots \oplus M_n^{(n)}$, one has

$$\mu_n := \mathbb{E}(M_n) = \sum_{k=1}^{n} \mathbb{E}\big(M_n^{(k)}\big) = \sum_{k=1}^{m} \frac{n-k}{n+k-2}.$$

Thus, using the fact that, for $n \to \infty$,

$$\left| \frac{n-k}{n+k-2} - \frac{n-k}{n+k} \right| \to 0,$$

uniformly for $1 \leq k \leq n$, one gets

$$\lim_{n\to\infty} \frac{\mathbb{E}(M_n)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{n-k}{n+k-2} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{n-k}{n+k} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{1-\frac{k}{n}}{1+\frac{k}{n}}$$
$$= \int_0^1 \frac{1-x}{1+x} dx = -1 + 2\log 2.$$

Similarly, one gets that the variance $\sigma_n^2 := \mathbb{V}(M_n)$ satisfies

$$\lim_{n\to\infty} \frac{\mathbb{V}(M_n)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{V}\big(M_n^{(k)}\big) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{n-k}{n+k-2} \left(1 - \frac{n-k}{n+k-2}\right)$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{n-k}{n+k} \left(1 - \frac{n-k}{n+k}\right) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{1-\frac{k}{n}}{1+\frac{k}{n}} \left(1 - \frac{1-\frac{k}{n}}{1+\frac{k}{n}}\right)$$
$$= \int_0^1 \frac{1-x}{1+x} \left(1 - \frac{1-x}{1+x}\right) dx = -4 + 6\log 2.$$

In order to prove Theorem 5.2.5, we consider the cumulants $\kappa_i(M_n^*)$, $i \in \mathbb{N}$, of the centered and normalized random variable $M_n^* := \frac{M_n - \mu_n}{\sigma_n}$. The first two cumulants are of course simply given, for all $n \in \mathbb{N}$, by $\kappa_1(M_n^*) = \mathbb{E}(M_n^*) = 0$, and $\kappa_2(M_n^*) = \mathbb{V}(M_n^*) = 1$. For $m \geq 3$ we use the following observation: Clearly, it holds that for each fixed m the m-th cumulant of a Bernoulli(p)-distributed random variable X_p, which can be computed by

$$\kappa_m(X_p) = [t^m] \log\left(\mathbb{E}\big(\mathrm{e}^{tX_p}\big)\right) = [t^m] \log\left(1 + p(\mathrm{e}^t - 1)\right),$$

is a polynomial in p. Especially, $\kappa_m(X_p)$ is uniformly bounded for $0 \leq p \leq 1$. Thus, using the fact that $M_n \stackrel{(d)}{=} M_n^{(1)} \oplus M_n^{(2)} \oplus \cdots \oplus M_n^{(n)}$, we get

$$\lim_{n\to\infty} \kappa_m(M_n^*) = \lim_{n\to\infty} \frac{\kappa_m(M_n)}{\sigma_n^m} = \lim_{n\to\infty} \frac{\sum_{k=1}^{n} \kappa_m\left(M_n^{(k)}\right)}{\sigma_n^m} = \lim_{n\to\infty} \frac{\mathcal{O}(n)}{\sigma_n^m} = 0,$$

for each fixed $m \geq 3$, since $\sigma_n^m \gg n$ in this case.

Hence, we see that the cumulants of the centralized and normalized random variable M_n^* converge to the cumulants of the standard normal distribution, which proves the convergence in distribution as claimed in Theorem 5.2.5.

5.4 Relation to up-down alternating trees

Up-down alternating trees are rooted labelled trees in which the labels $v_1, v_2, v_3, \ldots$ on each path starting at the root satisfy $v_1 < v_2 > v_3 < v_4 > \ldots$. The number F_n of such trees of size n has been studied by Postnikov (for unordered trees, cf. [Pos97]), by Chauve, Dulucq and Rechnitzer (for ordered trees, cf. [CDR01]), and by Kuba and Panholzer (for more general tree families, cf. [KP10]). In particular, it has been shown that the exponential generating function $F(z) := \sum_{n\geq 1} F_n \frac{z^n}{n!}$ of these numbers satisfies

$$z = \frac{2F(z)}{1+\mathrm{e}^{F(z)}}, \tag{5.10}$$

for unordered trees, and

$$z = (1-F(z)) \log\left(\frac{1}{1-F(z)}\right), \tag{5.11}$$

for ordered trees. By comparing (5.10) with (5.2), and (5.11) with the limit of (5.6) for $v \to 2$, one obtains in both cases the relation

$$F(z) = T(\frac{z}{2}, 2).$$

In other words, one has

$$F_n = \sum_{m\geq 0} T_{n,m} 2^{m-n}, \tag{5.12}$$

and thus the number of up-down alternating (ordered or unordered) trees can be obtained from the distribution of the number of local minima. In the following, we give a combinatorial proof of this fact, and show that this holds analogously for many other tree families. The following theorem states that (5.12) even holds for every tree family which simply consists of all labelled copies of a fixed unlabelled ordered tree:

Theorem 5.4.1. *Let t be an arbitrary unlabelled ordered tree of size n. Denote by $f(t)$ the number of labelled copies of t which are up-down alternating, and by $g_m(t)$ the number of labelled copies of t with m local minima. Then there holds*

$$2^n f(t) = \sum_{m\geq 0} 2^m g_m(t). \tag{5.13}$$

Proof. We prove this result by presenting a bijection between two tree families $\mathcal{F}_t$, $\mathcal{G}_t$ of size $2^n f(t)$ and $\sum_{m\geq 0} 2^m g_m(t)$, respectively: We let $\mathcal{F}_t$ be the family of bicoloured up-down alternating copies of t, i.e. each tree of $\mathcal{F}_t$ is a copy of t in which the nodes are labelled with $\{1,\ldots,n\}$ in an up-down alternating fashion, and in addition each node has one of two possible colors (say, white or gray). Clearly, there are exactly $2^n f(t)$ trees of this kind. The family $\mathcal{G}_t$ on the other hand is constructed by taking all bicoloured labelled copies of t with the only restriction that in this family only the nodes which are local minima are allowed to be gray. Hence, $\mathcal{G}_t$ contains $\sum_{m\geq 0} 2^m g_m(t)$ elements.

We now prove that there is a bijection $\phi_t : \mathcal{G}_t \to \mathcal{F}_t$. However, it will be convenient to consider not only the sets $\mathcal{F}_t$ and $\mathcal{G}_t$ in which all trees are labelled with $\{1, \ldots, n\}$, instead we consider the sets $\bar{\mathcal{F}}_t$ and $\bar{\mathcal{G}}_t$ which are constructed in the same way as $\mathcal{F}_t$ and $\mathcal{G}_t$ from t, but allowing any arbitrary set of distinct integers as node labels (in particular, $\bar{\mathcal{F}}_t$ and $\bar{\mathcal{G}}_t$ are infinite, but this will cause no problems). We then present a bijection $\bar{\phi}_t : \bar{\mathcal{G}}_t \to \bar{\mathcal{F}}_t$ with the property that every element of $\bar{\mathcal{G}}_t$ is mapped to an element of $\bar{\mathcal{F}}_t$ with the same set of labels. Hence, the required bijection ϕ_t will then simply be the restriction of $\bar{\phi}_t$ to $\mathcal{G}_t$.

We construct the bijection $\bar{\phi}_t$ by induction on the size n of t. Of course, the case $n = 1$ is trivial, because $\bar{\mathcal{F}}_t = \bar{\mathcal{G}}_t$ in this case and we simply let $\bar{\phi}_t$ be the identity. So let $n > 1$ and assume that we have already constructed the bijections $\bar{\phi}_{t'}$ for all ordered trees t' of sizes smaller than n. We then define $\bar{\phi}_t(\bar{t})$ for every tree t of size n and every bicoloured labelled copy $\bar{t} \in \bar{\mathcal{G}}_t$ by the following procedure:

1) Let L be the largest label in $\bar{t}$. By the construction of $\bar{\mathcal{G}}_t$, L surely is a white node, since it cannot be a local minimum. Removing L from $\bar{t}$ yields an "ordered forest" $(\bar{t}_0, \ldots, \bar{t}_r)$ of $r+1$ trees, where $\bar{t}_0$ contains the root of t (or is empty, if L is the root of t) and $\bar{t}_1, \ldots, \bar{t}_r$ are the subtrees rooted at the first, second,..., r-th child of L ($r = 0$ is possible if L has no children). For $0 \leq j \leq r$, $\bar{t}_j$ is contained in a family $\bar{\mathcal{G}}_{t_j}$, where t_j is an unlabelled ordered tree of size smaller than n, and hence we can apply $\bar{\phi}_{t_j}$ to $\bar{t}_j$ in order to obtain an element of $\bar{\mathcal{F}}_{t_j}$. This yields an ordered forest of $r+1$ bicoloured up-down alternating trees $(\bar{\phi}_{t_0}(\bar{t}_0), \ldots, \bar{\phi}_{t_r}(\bar{t}_r))$.

2) We now have to distinguish two cases:

   a) If the depth of L in $\bar{t}$ (i.e. the number of edges in the path from the root of $\bar{t}$ to node L) is odd, we can simply "glue" the up-down alternating trees $(\bar{\phi}_{t_0}(\bar{t}_0), \ldots, \bar{\phi}_{t_r}(\bar{t}_r))$ back together using node L. That is, the trees $\bar{\phi}_{t_1}(\bar{t}_1), \ldots, \bar{\phi}_{t_r}(\bar{t}_r)$ are connected to L in exactly this linear order, and if L had a parent node v in $\bar{t}$ (i.e. $\bar{\phi}_{t_0}(\bar{t}_0)$ is non-empty), then L again becomes a child of v (at the same position as in $\bar{t}$). The resulting tree will then be an element of $\bar{\mathcal{F}}_t$. In order to mark that this first case has occurred, we now color node L gray, and we let the resulting tree be $\bar{\phi}_t(\bar{t})$.

b) If, on the other hand, the depth of L in $\bar{t}$ is even, we can not simply glue the trees together in order to obtain an element of $\bar{\mathcal{F}}_t$ like in the first case. Instead, we do the following:

   i. Replace $\bar{\phi}_{t_0}(\bar{t}_0)$ by $\bar{\phi}_{t_0}(\bar{t}_0)'$, which results from $\bar{\phi}_{t_0}(\bar{t}_0)$ by swapping the largest label with the smallest one, the largest but one with the smallest but one, and so on (the colors of each node are swapped together with the labels). $\bar{\phi}_{t_0}(\bar{t}_0)'$ is then clearly a bicoloured down-up alternating tree. Then, by glueing $(\bar{\phi}_{t_0}(\bar{t}_0)', \bar{\phi}_{t_1}(\bar{t}_1), \ldots, \bar{\phi}_{t_r}(\bar{t}_r))$ together like in a) we get a bicoloured down-up alternating tree.

   ii. Now, again by swapping the largest label with the smallest one, the largest but one with the smallest but one, and so on, we obtain an element of $\bar{\mathcal{F}}_t$, which we let be $\bar{\phi}_t(\bar{t})$. Node L remains white in this case.

Since, by the induction hypothesis, the mappings $\bar{\phi}_{t_j}, 0 \leq j \leq r$ are injective, so is $\bar{\phi}_t$. Moreover, since the the mappings $\bar{\phi}_{t_j}$ are surjective and both cases ("white" or "gray") for the node with the largest label of any element of $\bar{\mathcal{F}}_t$ are covered by step 2), the mapping $\bar{\phi}_t$ is also surjective. Furthermore, also by the induction hypothesis, each of the mappings $\bar{\phi}_{t_j}$ in step 1) preserves the used set of labels, and hence so does $\bar{\phi}_t$. Thus, as required, the restriction of $\bar{\phi}_t$ to $\mathcal{G}_t$ is a bijection $\phi_t : \mathcal{G}_t \to \mathcal{F}_t$. This concludes the proof of Theorem 5.4.1.

For some small trees t, the bijections ϕ_t are depicted in Figure 5.5. □

Corollary 5.4.2. *Let $\mathcal{T}$ be an arbitrary labelled family of simply generated trees, and denote by F_n the total weight of up-down alternating trees of size n in $\mathcal{T}$. Moreover, let $G_{n,m}$ be the total weight of trees in $\mathcal{T}$ which are of size n and contain exactly m local minima. Then there holds*

$$2^n F_n = \sum_{m \geq 0} 2^m G_{n,m}. \tag{5.14}$$

Hence, (5.14) *especially holds if F_n denotes the number of up-down alternating unordered, cyclic or d-ary trees of size n, and $G_{n,m}$ the number of unordered, cyclic or d-ary trees, respectively, of size n with m local minima.*

Proof. Since the weight $w(t)$ of any tree $t \in \mathcal{T}$ does not depend on the labelling of the nodes, one has $F_n = \sum_{t \in \mathcal{U}_n} w(t) f(t)$, and $G_{n,m} = \sum_{t \in \mathcal{U}_n} w(t) g_m(t)$, where $\mathcal{U}_n$ denotes the set of unlabelled ordered trees of size n and $f(t)$ and $g_m(t)$ are defined

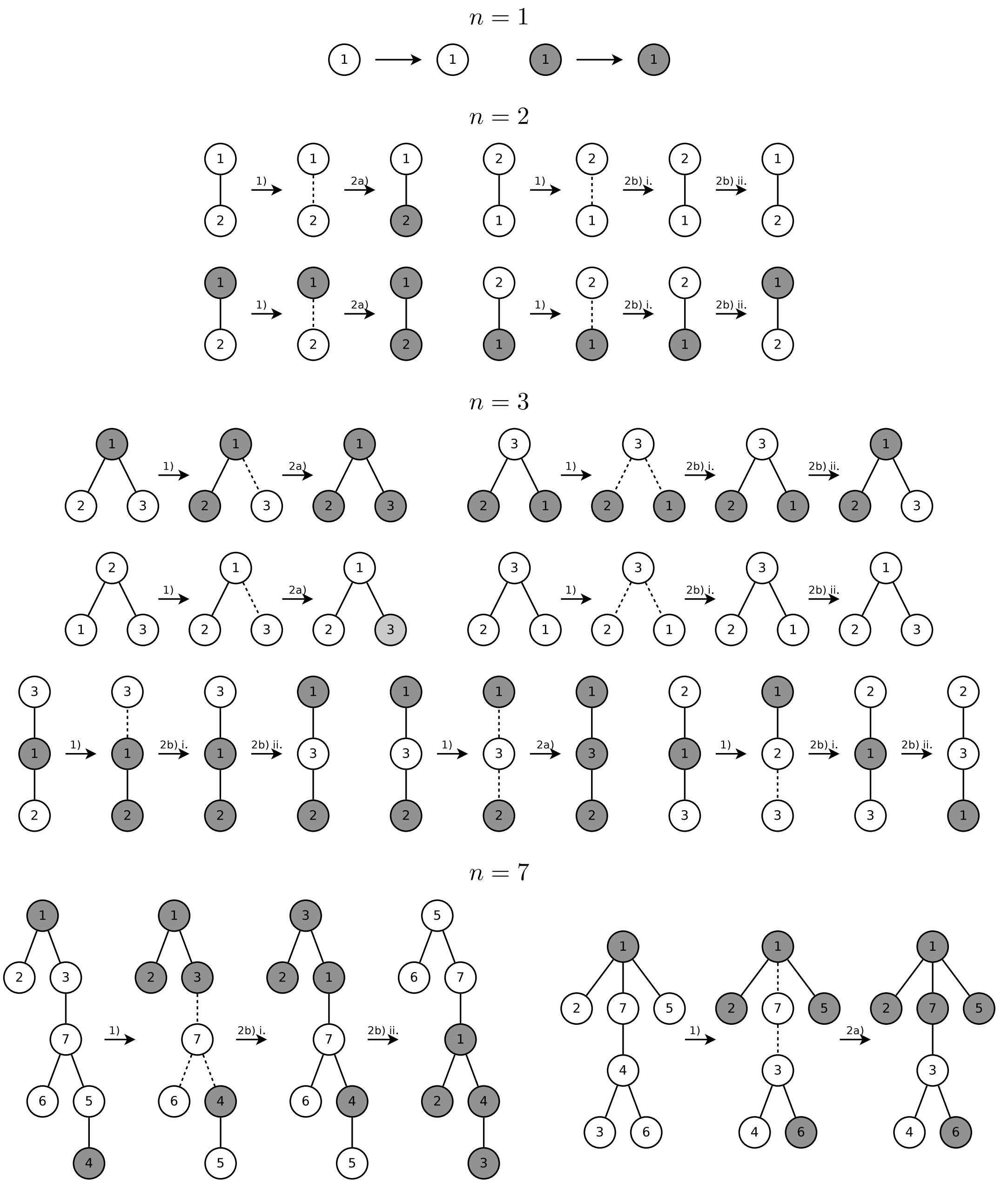

Figure 5.5: Application of ϕ_t to $\mathcal{G}_t$ for t of size $n = 1$ and $n = 2$ and some examples for sizes $n = 3$ and $n = 7$.

as in Theorem 5.4.1. For each $t \in \mathcal{U}_n$, we can use equation (5.13), and we thus obtain

$$\begin{aligned} 2^n F_n &= \sum_{t \in \mathcal{U}_n} w(t) 2^n f(t) = \sum_{t \in \mathcal{U}_n} w(t) \sum_{m \geq 0} 2^m g_m(t) = \sum_{m \geq 0} 2^m \sum_{t \in \mathcal{U}_n} w(t) g_m(t) \\ &= \sum_{m \geq 0} 2^m G_{n,m}. \end{aligned}$$

$\square$

APPENDIX A

Probability distributions

When studying the limiting behaviour of the random variables which we considered in this thesis, we encountered several common probability distributions, but also some more "exotic" ones like the Airy distribution. For reference we collect the definitions of these distributions in the form in which we used them (note that, e.g., for Gamma distribution and negative binomial distribution, also slightly differing formulations are used in the literature).

A.1 Gamma distribution, exponential distribution

The Gamma distribution with parameters $\kappa > 0$ and $\theta > 0$ (for short, Gamma(κ, θ) distribution) is the probability distribution of a random variable X with probability density function

$$f_X(x) = x^{\kappa-1} \frac{e^{-x/\theta}}{\theta^\kappa \, \Gamma(\kappa)}, \quad x \geq 0.$$

The moments of a Gamma$(\kappa,\, \theta)$ distributed random variable X are given by

$$\mathbb{E}(X^s) = \theta^s \frac{\Gamma(s+\kappa)}{\Gamma(\kappa)}, \quad s \in \mathbb{N},$$

and the Gamma$(\kappa,\, \theta)$ distribution is uniquely determined by these moments. In the special case $\kappa = 1$ the Gamma distribution is called Exponential distribution with parameter $\tilde{\theta} := 1/\theta$, which we write as Exp$(\tilde{\theta})$.

A.2 Beta distribution

A random variable X is Beta distributed with parameters $\alpha > 0$ and $\beta > 0$ (for short, $\mathrm{Beta}(\alpha, \beta)$ distributed), iff it has the probability density function

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du}, \quad x \in (0,1).$$

The moments of X are then given by

$$\mathbb{E}(X^s) = \frac{\Gamma(\alpha+\beta)\,\Gamma(\alpha+s)}{\Gamma(\alpha+\beta+s)\,\Gamma(\alpha)}, \quad s \in \mathbb{N},$$

and these moments uniquely determine the $\mathrm{Beta}(\alpha, \beta)$ distribution.

A.3 Negative binomial distribution, geometric distribution

A discrete random variable X is negative binomial distributed with parameters $r > 0$ and $p \in (0,1)$ (for short, $\mathrm{NegBin}(r,p)$ distributed), iff its probability mass function is given by

$$\mathbb{P}\{X = m\} = \binom{m+r-1}{m} p^r (1-p)^m, \quad m \in \mathbb{N}_0.$$

In the special case $r = 1$, this distribution is called Geometric distribution with parameter p, which we write as $\mathrm{Geom}(p)$.

A.4 Normal distribution

A random variable X is normally distributed with mean μ and standard deviation σ (for short, $\mathcal{N}(\mu, \sigma)$ distributed) iff it has the probability density function

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

The cumulants $\kappa_r(X)$ of a normally distributed random variable are given by $\kappa_1(X) = \mathbb{E}(X) = \mu$, $\kappa_2(X) = \mathbb{V}(X) = \sigma^2$, and $\kappa_r(X) = 0$, for $r \geq 3$, and these cumulants uniquely determine the $\mathcal{N}(\mu, \sigma)$-distribution. The distribution function of the *standard normal distribution* $\mathcal{N}(0,1)$ is denoted by $\Phi(x)$.

A.5 Rayleigh distribution

The Rayleigh distribution with parameter $\sigma > 0$ is the distribution of a random variable X with probability density function

$$f_X(x) = \frac{x}{\sigma^2}\, e^{-\frac{x^2}{2\sigma^2}}, \quad x > 0. \tag{A.1}$$

This distribution appears frequently when studying combinatorial objects, see, e.g., [FS09]. When applying the method of moments in order to establish some of our results, we use the following basic fact about the Rayleigh distribution:

Lemma A.5.1. *The Rayleigh distribution is uniquely determined by its sequence of r-th moments $(\mu_r)_{r\geq 1}$, which are given as follows:*

$$\mu_r := \mathbb{E}(X_\sigma^r) = \sigma^r\, 2^{\frac{r}{2}}\, \Gamma\Big(\frac{r}{2} + 1\Big)\,. \tag{A.2}$$

A.6 Airy distribution

The Airy distribution is the probability distribution of a random variable X with r-th moments

$$\mu_r := \mathbb{E}(X^r) = -\frac{\Gamma\big(-\frac{1}{2}\big)}{\Gamma\big(\frac{3r-1}{2}\big)} C_r, \tag{A.3}$$

where the constants C_r can inductively be defined by

$$2C_r = (3r-4)rC_{r-1} + \sum_{j=1}^{r-1} \binom{r}{j} C_j C_{r-j}, \quad r \geq 2, \quad C_1 = \frac{1}{2}. \tag{A.4}$$

The Airy distribution occurs as a limiting distribution in enumerative studies of many combinatorial objects, such as the area below lattice paths [Lou84], the area of staircase polygons [SRT10], sums of parking functions [KY03], and the costs of linear probing hashing algorithms [FPV98]. The following well-known result about the Airy distribution proves useful (see, e.g., [FL01], where one can find more details about the Airy distribution and some equivalent definitions).

Lemma A.6.1. *The Airy distribution is uniquely determined by its sequence of moments $(\mu_r)_{r\geq 1}$.*

APPENDIX B

The method of characteristics

The method of characteristics (cf., e.g., [Tay96]) is a technique for solving linear partial differential equations (linear PDEs). Since we use this method in our derivations, we shortly explain it here.

We only consider the simplest case of a homogeneous linear first order PDE for a function $f(x, y)$ of two variables, which is of the form

$$u_1(x, y)f_x(x, y) + u_2(x, y)f_y(x, y) + u_3(x, y)f(x, y) = 0. \tag{B.1}$$

The method of characteristics rests on the idea of transforming this equation to an ordinary differential equation along so-called *characteristic curves*. In order to find these characteristic curves, one considers the system of characteristic equations,

$$\begin{aligned} \dot{x} &= u_1(x, y), \\ \dot{y} &= u_2(x, y), \end{aligned} \tag{B.2}$$

where $x = x(t)$ and $y = y(t)$ are considered as functions of a new parameter t. Now assume that $\zeta(x, y)$ is a first integral of (B.2), i.e. $\zeta(x(t), y(t))$ is constant along each solution $(x(t), y(t))$ of (B.2). If we choose for each point (x_0, y_0) of the (x, y)-plane a solution $(x(t), y(t))$ of (B.2) which passes through (x_0, y_0), i.e. $(x_0, y_0) = (x(t_0), y(t_0))$ for some t_0, then it follows that

$$\zeta_x(x_0, y_0)u_1(x_0, y_0) + \zeta_y(x_0, y_0)u_2(x_0, y_0) = \zeta_x(x_0, y_0)\dot{x}(t_0) + \zeta_y(x_0, y_0)\dot{y}(t_0) = 0,$$

and hence

$$\zeta_x(x, y)u_1(x, y) + \zeta_y(x, y)u_2(x, y) \equiv 0. \tag{B.3}$$

Now, if we consider a coordinate transform $\zeta = \zeta(x, y)$, $\eta = \eta(x, y)$ (where we require that this transform is invertible, i.e. $x = x(\zeta, \eta)$, $y = y(\zeta, \eta)$, and continuously differentiable), equation (B.1) is transformed into the following PDE for $F(\zeta, \eta) := f(x(\zeta, \eta), y(\zeta, \eta))$:

$$\begin{aligned} 0 &= u_1 f_x + u_2 f_y + u_3 f \\ &= u_1(F_\zeta \zeta_x + F_\eta \eta_x) + u_2(F_\zeta \zeta_y + F_\eta \eta_y) + u_3 F, \end{aligned}$$

where $u_1 = u_1(x(\zeta, \eta), y(\zeta, \eta))$, $f_x = f_x(x(\zeta, \eta), y(\zeta, \eta))$, and so on. Due to equation (B.3), the terms containing F_ζ in this PDE cancel, and hence one obtains an ordinary differential equation for $F(\zeta, \eta)$ with respect to η. Solving this equation and transforming the result back to (x, y)-space thus yields the solution of (B.1).

In practice, one obtains the required first integral $\zeta(x, y)$ of (B.2) by considering the *phase differential equation*

$$\frac{dx}{dy} = \frac{u_1(x, y)}{u_2(x, y)}. \tag{B.4}$$

Given the general solution $x(y) = f(y, c)$ of (B.4) which depends on a parameter c, one then solves for c, i.e. $c = \zeta(x, y)$.

APPENDIX C

Difference calculus

Difference calculus (cf., e.g., [GKP94]) is based on the properties of the *difference operator* defined by

$$\Delta f(x) = f(x+1) - f(x).$$

We collect some of these properties.

- One of the most important properties of Δ is the fact that for the m-th falling factorials $x^{\underline{m}} := x(x-1)\cdots(x-m+1)$ the following equation holds:

$$\Delta x^{\underline{m}} = m x^{\underline{m-1}}, \quad \text{for } m \in \mathbb{N}. \tag{C.1}$$

- More generally, one easily gets by induction that $\Delta^r x^{\underline{m}} = m^{\underline{r}} x^{\underline{m-r}}$, for $r \in \mathbb{N}_0$.

- Δ commutes with the *α-shift operator* E_α ($\alpha \in \mathbb{R}$), defined by $\mathrm{E}_\alpha f(x) = f(x+\alpha)$, i.e. one has the operator equation

$$\Delta \mathrm{E}_\alpha = \mathrm{E}_\alpha \Delta, \quad \text{for } \alpha \in \mathbb{R}.$$

  As an example, we have $\Delta^r (x+\alpha)^m = \Delta^r \mathrm{E}_\alpha x^m = \mathrm{E}_\alpha \Delta^r x^m = m^{\underline{r}} (x+\alpha)^{\underline{m-r}}$.

An important field of application for difference calculus is the derivation of summation formulæ for certain sums involving factors of the form $(-1)^\ell \binom{m}{\ell}$. This is due to the fact that

$$\Delta^m f(x) = (-1)^m \sum_{\ell=0}^{m} (-1)^\ell \binom{m}{\ell} f(x+\ell), \quad \text{for } m \in \mathbb{N}_0. \tag{C.2}$$

Note that equation (C.2) can just be proven by induction on m, but a more elegant proof is the following [GKP94]:

Consider the shift operator $\mathrm{E} := \mathrm{E}_1$, defined by $\mathrm{E}f(x) = f(x+1)$. Clearly, one has the operator equation $\Delta = \mathrm{E} - \mathrm{I}$, where I denotes the identity operator. By using the standard binomial theorem for $(\mathrm{E} - \mathrm{I})^m$, it follows that

$$\Delta^m f(x) = (\mathrm{E} - \mathrm{I})^m f(x) = \sum_{\ell=0}^{m} \binom{m}{\ell} (-1)^{m-\ell} \mathrm{E}^\ell f(x),$$

which proves (C.2).

As an example application of (C.2), we prove a summation formula which we use in our studies (note that this example is taken directly from [GKP94]): We define the negative-exponent falling factorials $x^{\underline{-m}}$ by

$$x^{\underline{-m}} = \frac{1}{(x+1)(x+2)\cdots(x+m)}, \quad m \in \mathbb{N}.$$

One easily checks that with this definition equation (C.1) holds for all $m \in \mathbb{Z}$. Hence, by setting $f(x) = \frac{1}{x} = (x-1)^{\underline{-1}}$, equation (C.2) shows that

$$\begin{aligned} \sum_{\ell=0}^{m} \binom{m}{\ell} (-1)^\ell \frac{1}{\ell + x} &= (-1)^m \Delta^m (x-1)^{\underline{-1}} = (-1)^m (-1)^{\underline{m}} (x-1)^{\underline{-m-1}} \\ &= \frac{m!}{x(x+1)\cdots(x+m)} = \frac{1}{(m+1)\binom{m+x}{m+1}}, \end{aligned} \tag{C.3}$$

for $m \in \mathbb{N}_0$ and $x \notin \{-m, \dots, 0\}$.

Notation

$\mathbb{N}$	the set of natural numbers (without 0): $\mathbb{N} = \{1, 2, 3, \ldots\}$
$\mathbb{N}_0$	the set of natural numbers including 0: $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$
$\mathbb{Z}$	the set of integers: $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
$n!$	factorial of n ($n \in \mathbb{N}_0$): $0! = 1$, $n! = n \cdot (n-1)! = n(n-1)\cdots 1$, $n \in \mathbb{N}$
$\alpha^{\underline{k}}$	k-th falling factorial of α ($\alpha \in \mathbb{R}$, $k \in \mathbb{N}_0$): $\alpha^{\underline{0}} = 1$, $\alpha^{\underline{k}} = \alpha \cdot (\alpha-1)^{\underline{k-1}} = \alpha(\alpha-1)\cdots(\alpha-k+1)$, $k \in \mathbb{N}$
$\alpha^{\overline{k}}$	k-th rising factorial of α ($\alpha \in \mathbb{R}$, $k \in \mathbb{N}_0$): $\alpha^{\overline{0}} = 1$, $\alpha^{\overline{k}} = \alpha \cdot (\alpha+1)^{\underline{k-1}} = \alpha(\alpha+1)\cdots(\alpha+k-1)$, $k \in \mathbb{N}$
$\Gamma(x)$	gamma function
$\binom{\alpha}{\beta}$	binomial coefficient ($\alpha, \beta \in \mathbb{R}$): $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ ($n, k \in \mathbb{N}_0$, $k \leq n$); $\binom{\alpha}{k} = \frac{\alpha^{\underline{k}}}{k!}$ ($\alpha \in \mathbb{R}$, $k \in \mathbb{N}_0$); in general: $\binom{\alpha}{\beta} = \lim_{t \to 0} \frac{\Gamma(\alpha+t+1)}{\Gamma(\beta+1)\Gamma(\alpha-\beta+t+1)}$
$\mathrm{Arg}(z)$	argument of z: $\mathrm{Arg}(z) = \phi$ for $z = \lvert z\rvert \mathrm{e}^{\mathrm{i}\phi}$, $\phi \in (-\pi, \pi]$

$\left[{n \atop k}\right]$	(unsigned) Stirling numbers of the first kind
$\left\{{n \atop k}\right\}$	Stirling numbers of the second kind
$\lvert T\rvert$	size of the combinatorial object T
$*$	combinatorial product of labelled objects
$\dot\cup$	disjoint union
$\Box$	placeholder for an object of size 0
$[z^n]$	operator which extracts the coefficient of z^n from a power series: $A(z) = \sum_{n\in\mathbb{N}_0} a_n z^n \Rightarrow [z^n]A(z) = a_n$
D_x	differential operator with respect to x: $\mathrm{D}_x = \frac{\partial}{\partial x}$
N_x	operator which evaluates at $x = 0$: $\mathrm{N}_x(.) = (.)\rvert_{x=0}$
U_x	operator which evaluates at $x = 1$: $\mathrm{U}_x(.) = (.)\rvert_{x=1}$
Z	operator which multiplies with z: $\mathrm{Z}(.) = z \cdot (.)$
Δ	difference operator: $\Delta f(x) = f(x+1) - f(x)$
$\mathbb{E}(X)$	expected value of X
$\mathbb{V}(X)$	variance of X
$\kappa_r(X)$	r-th cumulant of X
$X \oplus Y$	sum of independent random variables

$\mathbb{1}(\mathcal{A})$	indicator variable of event $\mathcal{A}$
$\mathbb{P}\{\mathcal{A}\}$	probability of event $\mathcal{A}$
$\mathcal{A}^c$	complement of event $\mathcal{A}$
$\mathbb{P}\{\mathcal{A} \mid \mathcal{B}\}$	conditional probability of event $\mathcal{A}$, given event $\mathcal{B}$
$X_n \xrightarrow{(d)} X$	convergence in distribution of $(X_n)_{n\in\mathbb{N}}$ to X
$X \overset{(d)}{=} Y$	equality in distribution of X and Y
$\Phi(x)$	distribution function of the standard normal distribution

List of Figures

Bibliography

[AM08] M. Albenque and J.-F. Marckert. Some families of increasing planar maps. *Electronic Journal of Probability*, 13:1624–1671, 2008.

[AS64] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.

[BA99] A.-L. Bárabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[BFS92] F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. *Lecture Notes in Computer Science*, 581:24–48, 1992.

[Bil84] P. Billingsley. *Probability and Measure.* John Wiley & Sons, New York, second edition edition, 1984.

[BP69] L. W. Beineke and R. E. Pippert. The number of labeled k-dimensional trees. *Journal of Combinatorial Theory*, 6:200–205, 1969.

[BPS] C. Banderier, A. Panholzer, and G. Seitz. Three problems related to label patterns in trees. In preparation.

[BR03] B. Bollobás and O. M. Riordan. Mathematical results on scale-free random graphs. In *Handbook of graphs and networks*, pages 1–34, Weinheim, 2003. Wiley-VCH.

[Car74] L. Carlitz. Permutations and sequences. *Advances in Mathematics*, 14:92–120, 1974.

[CDR01] C. Chauve, S. Dulucq, and A. Rechnitzer. Enumerating alternating trees. *Journal of Combinatorial Theory, Series A*, 94(1):142–151, 2001.

[DB62] F. N. David and D. E. Barton. *Combinatorial Chance*. Griffin, London, 1962.

[DG99] M. Drmota and B. Gittenberger. The distribution of nodes of given degree in random trees. *Journal of Graph Theory*, 31:227–253, 1999.

[DHBS10] A. Darrasse, H.-K. Hwang, O. Bodini, and M. Soria. The connectivity-profile of random increasing k-trees. In *Proceedings of ANALCO'10: Workshop on Analytic Algorithmics and Combinatorics*, pages 99–106, 2010.

[Drm09] M. Drmota. *Random trees*. Springer, Vienna, 2009.

[DS96] R. P. Dobrow and R. T. Smythe. Poisson approximation for functionals of random trees. *Random Structures and Algorithms*, 9:79–92, 1996.

[DS09] A. Darrasse and M. Soria. Limiting distribution for distances in k-trees. In *Proceedings of the 20th International Workshop on Combinatorial Algorithms (IWOCA 2009)*, Lecture Notes in Computer Science, pages 170–182. Springer-Verlag, 2009.

[Ess84] C.-G. Esseen. On the application of the theory of probability to two combinatorial problems involving permutations. In *Probability theory, Proc. 7th Conf.*, pages 137–147, 1984.

[FFK05] J. A. Fill, P. Flajolet, and N. Kapur. Singularity analysis, hadamard products, and tree recurrences. *Journal of Computational and Applied Mathematics*, 174:271–313, 2005.

[FL01] P. Flajolet and G. Louchard. Analytic variations on the airy distribution. *Algorithmica*, 31:361–377, 2001.

[FO90] P. Flajolet and A. M. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3:216–240, 1990.

[FPV98] P. Flajolet, P. V. Poblete, and A. Viola. On the analysis of linear probing hashing. *Algorithmica*, 22:490–515, 1998.

[FS09] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[Gao09] Y. Gao. The degree distribution of random k-trees. *Theoretical Computer Science*, 410:688–695, 2009.

[GKP94] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete mathematics.* Addison-Wesley, Reading, second edition, 1994.

[GYS99] I. M. Gessel, Y.-N. Yeh, and B. E. Sagan. Enumeration of trees by inversions. *J. Graph Theory*, 19:435–459, 1999.

[HP68] F. Harary and E. Palmer. On acyclic simplicial complexes. *Mathematika*, 15:115–122, 1968.

[Hwa98] H.-K. Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, 19:329–343, 1998.

[Knu73] D. E. Knuth. *The art of computer programming*, volume 3. Addison-Wesley, Reading, 1973.

[KP06] M. Kuba and A. Panholzer. Descendants in increasing trees. *Electronic Journal of Combinatorics*, 13:Research paper 8, 2006.

[KP07] M. Kuba and A. Panholzer. On the degree distribution of the nodes in increasing trees. *Journal of Combinatorial Theory, Series A*, 114:597–618, 2007.

[KP10] M. Kuba and A. Panholzer. Enumeration results for alternating tree families. *European Journal of Combinatorics*, 31(7):1751–1780, 2010.

[KY03] J. P. S. Kung and C. Yan. Exact formulas for moments of sums of classical parking functions. *Advances in Applied Mathematics*, 31:415–441, 2003.

[Loè77] M. Loève. *Probability Theory I.* Springer, New York, 4th edition, 1977.

[Lou84] G. Louchard. Kac's formula, Lévy's local time and Brownian excursion. *Journal of Applied Probability*, 21:479–499, 1984.

[LP03] G. Louchard and H. Prodinger. The number of inversions in permutations: a saddle point approach. *Journal of Integer Sequences*, 6:Article 03.2.8, 2003.

[Mah92] H. M. Mahmoud. *Evolution of random search trees.* John Wiley & Sons, New York, 1992.

[MM78] A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, 30:997–1015, 1978.

[Moo69] J. W. Moon. The number of labeled k-trees. *Journal of Combinatorial Theory*, 6:196–199, 1969.

[MR68] C. Mallows and J. Riordan. The inverson enumerator for labeled trees. *Bulletin of the American Mathematical Society*, 74:92–94, 1968.

[MS95] H. M. Mahmoud and R. T. Smythe. A survey of recursive trees. *Theoretical Probability and Mathematical Statistics*, 51:1–37, 1995.

[Net01] E. Netto. *Lehrbuch der Combinatorik.* B. G. Teubner, Leipzig, 1901.

[Pan04a] A. Panholzer. The distribution of the size of the ancestor-tree and of the induced spanning subtree for random trees. *Random Structures and Algorithms*, 25:179–207, 2004.

[Pan04b] A. Panholzer. Distribution of the Steiner distance in generalized m-ary search trees. *Combinatorics, Probability & Computing*, 13:717–733, 2004.

[Pos97] A. Postnikov. Intransitive trees. *Journal of Combinatorial Theory, Series A*, 79(2):260–266, 1997.

[PP07] A. Panholzer and H. Prodinger. Level of nodes in increasing trees revisited. *Random Structures and Algorithms*, 31:203–226, 2007.

[Pro96] H. Prodinger. Descendants in heap ordered trees, or, a triumph of computer algebra. *Electronic Journal of Combinatorics*, 3:R29, 1996.

[PS10a] A. Panholzer and G. Seitz. Ancestors and descendants in evolving k-tree models. Submitted, 2010.

[PS10b] A. Panholzer and G. Seitz. Ordered increasing k-trees: Introduction and analysis of a preferential attachment network model. *Discrete Mathematics and Theoretical Computer Science*, Proceedings AM:549–564, 2010.

[PS11] A. Panholzer and G. Seitz. Limiting distributions for the number of inversions in labelled tree families. Accepted for publication in *Annals of Combinatorics*, 2011.

[Rio58] J. Riordan. *An Introduction to Combinatorial Analysis*. John Wiley & Sons, New York, 1958.

[SRT10] U. Schwerdtfeger, C. Richard, and B. Thatte. Area limit laws for symmetry classes of staircase polygons. *Combinatorics, Probability and Computing*, 19:441–461, 2010.

[Sti86] S. Stigler. Estimating serial correlation by visual inpsection of diagnostic plots. *The American Statistician*, 40(2):111–116, 1986.

[Tay96] M. E. Taylor. *Partial differential equations. Basic theory*, volume 23 of *Texts in Applied Mathematics*. Springer, New York, 1996.

[WS96] D. I. Warren and E. Seneta. Peaks and eulerian numbers in a random sequence. *Journal of Applied Probability*, 33(1):101–114, 1996.

[WS98] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

Lebenslauf

Ich wurde am 18. März 1985 in Wien geboren. Von 1991 bis 1995 besuchte ich die Volksschule Pfeilgasse in Wien 8, und von 1995 bis 2003 das Gymnasium Albertgasse, ebenfalls in Wien 8. Im Juni 2003 schloss ich die Matura mit ausgezeichnetem Erfolg ab. Von Oktober 2003 bis September 2004 leistete ich meinen Zivildienst in der Organisationsabteilung der Wiener Jugendzentren in Wien 21 ab.

Im Oktober 2004 inskribierte ich mich an der Technischen Universität Wien zum Studium der Technischen Mathematik (Studienzweig „Mathematik in den Computerwissenschaften"), welches ich im April 2009 mit ausgezeichnetem Erfolg abschloss. Während meiner Studienzeit wirkte ich als Tutor bei der Abhaltung verschiedener Lehrveranstaltungen für Informatik- und Mathematikstudierende mit und absolvierte drei einmonatige Sommer-Praktika bei Siemens Österreich in Wien 10.

Seit Juni 2009 bin ich im Rahmen des FWF National Research Network S9600, „Analytic Combinatorics and Probabilistic Number Theory" als Forschungsassistent im Teilprojekt S9608, „Combinatorial Analysis of Data Structures and Tree-Like Structures", Projektleiter Alois Panholzer, angestellt. Gleichzeitig inskribierte ich mich an der Technischen Universität Wien zum Doktoratsstudium der technischen Wissenschaften. Des weiteren absolvierte ich zwischen Juli 2010 und September 2011 das Informatik-Masterstudium „Computational Intelligence" mit Auszeichnung. Seit Jänner 2011 bin ich an der Technischen Universität Wien als Universitäts-Assistent angestellt.