

Scaling limits of random trees

by

Douglas Paul Rizzolo

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Mathematics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

James W. Pitman, Chair
Fraydoun Rezakhanlou
Nouredine El Karoui

Spring 2012

Scaling limits of random trees

Copyright 2012
by
Douglas Paul Rizzolo

Abstract

Scaling limits of random trees

by

Douglas Paul Rizzolo

Doctor of Philosophy in Mathematics

University of California, Berkeley

James W. Pitman, Chair

We investigate scaling limits of several types of random trees. The study of scaling limits of random trees was initiated by Aldous in the early 1990's. One of the primary goals of his work was to study the uniform distribution on trees with n vertices as n grew to infinity, as well as uniform distributions on other combinatorially motivated models of trees with n vertices. We are motivated by studying combinatorial models of random trees with n leaves as n goes to infinity. Conditioning on the number of leaves rather than the number of vertices has significant consequences in terms of what techniques are applicable. We deal with these issues by developing a general theory for scaling limits of Markov branching trees whose size is given by their number of vertices with out-degree in a fixed set. This general theory is then applied to obtain scaling limits of Galton-Watson trees conditioned on their number of vertices with out-degree in a fixed set. We also show that many combinatorial models of trees with n leaves can be realized as Galton-Watson trees (or probabilistic transforms thereof) conditioned to have n leaves.

Contents

Contents	i
List of Figures	ii
1 Introduction	1
1.1 Background	1
1.2 A motivating example	2
1.3 Organization	4
2 Random trees with a given number of leaves	6
2.1 Schröder's problems	6
2.2 Combinatorial models and Galton-Watson trees	8
2.3 Explicit computations using analytic combinatorics	17
3 Markov Branching Trees	28
3.1 Introduction	28
3.2 Markov branching trees	28
3.3 Trees as metric measure spaces	29
3.4 Convergence of Markov branching trees	34
4 Scaling limits of Galton-Watson trees	37
4.1 Introduction	37
4.2 Galton-Watson trees	38
4.3 The partition at the root	44
4.4 Convergence of Galton-Watson trees	52
Bibliography	53

List of Figures

1.1	A fragmentation tree with 6 leaves	3
2.1	Binary word bracketings and rooted ordered binary trees for $n = 4$	7
2.2	Binary set bracketings and rooted unordered leaf-labeled binary trees for $n = 4$.	7
2.3	Set bracketings and rooted unordered leaf-labeled trees for $n = 3$	8
4.1	A colored tree and its image under $\vee$	42

Acknowledgments

This work was supported in part by NSF Grant No. 0806118 and in part by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400.

I would like to thank my advisor, Jim Pitman, for his help throughout my time at Berkeley as well as my dissertation committee for feedback on drafts of this dissertation.

Chapter 1

Introduction

1.1 Background

The study of scaling limits of random trees was initiated by Aldous in the early 1990's in his three paper Continuum Random Tree series [1, 2, 3]. One of the primary goals of these papers was to study the uniform distribution on trees with n vertices as n grew to infinity, as well as uniform distributions on other combinatorially motivated models of trees with n vertices. The first paper, [1], introduced the notion of scaling limits of random trees and considered scaling limits of uniformly random labeled unordered trees. The limiting objects were called continuum random trees, which are random metric spaces that often have additional structure, such as a measure. The results for uniformly random labeled unordered trees were extended in [3] to obtain scaling limits for critical finite variance Galton-Watson trees conditioned to have n vertices. Conditioned Galton-Watson trees are natural to consider in this context because many combinatorial models of random trees with n vertices can be recognized as conditioned Galton-Watson trees. One of the central ideas put forth in these papers was that it is profitable to consider random trees as random metric spaces and to examine their scaling limits in an appropriate topology, typically a variant of the Hausdorff topology on compact subsets of a metric space.

Since the initial contributions of Aldous, the field of scaling limits of random trees has been developed by many authors using a wide variety of techniques. Initially, and still, substantial focus has been on extending the results and refining the techniques, for studying Galton-Watson trees conditioned to have n vertices. For example, in [10] in 2003 Duquesne extended the results of [3] to obtain scaling limits for conditioned Galton-Watson trees whose offspring distribution was in the domain of attraction of a stable law with index $\alpha \in (1, 2]$. At the same time Marckert and Mokkadem in [20] obtained a relatively elementary proof of the scaling limit for conditioned Galton-Watson trees whose offspring distribution had exponential moments. More recently, Le Gall combined the ideas of [20] with excursion theory for Brownian motion in [19] to obtain a relatively simple proof of the scaling limit of finite variance Galton-Watson trees. Many other authors have contributed to the study of

Galton-Watson trees conditioned have n vertices and we do not attempt to give a complete list.

At the same time as the theory for Galton-Watson trees conditioned to have n vertices was being developed, another line of research was building connections between continuum random trees and coalescent and fragmentation processes. The first connection to coalescent processes was made in [3], which showed that Kingman's coalescent can be viewed as a continuum random tree. Then, Aldous and Pitman showed in [4] that the standard additive coalescent could be constructed from a continuum random tree. In [14] Haas and Miermont showed that many continuum random trees of interest could be constructed as the genealogical trees of self-similar fragmentation processes. In this context of genealogical trees of coalescent and fragmentation processes, the trees under consideration have n leaves rather than n vertices. The first general result for fragmentation trees with n leaves was obtained in 2008 in [16] under a consistency assumption, which was recently removed in [15]. The models in this context are generally motivated by phylogenetic models, rather than the combinatorial problems that motivated the study of Galton-Watson trees.

In this dissertation, we are motivated by models that blend these two areas. In particular, we are motivated by natural combinatorial models of trees with n leaves. We will use tools from both areas, and develop some new tools as well, to obtain scaling limits for these trees.

1.2 A motivating example

As a motivating example consider, as prelude to Chapter 2, the following problem. Suppose you take the set $[n] := \{1, \dots, n\}$, and partition it into nonempty sets $B_1, \dots, B_k$ with $k \geq 2$. Then take each set B_i with more than two elements and partition it into at least two sets. Continue this process until you have only singletons. If $\mathbf{B}_n$ is the collection of sets that appeared in this process, then there is a natural rooted tree structure T_n on $\mathbf{B}_n$. Namely, $\mathbf{B}_n$ is the set of vertices, $[n]$ is the root and if $A, B \in \mathbf{B}_n$, then there is an edge connected A and B if $A \subsetneq B$ and there is no $C \in \mathbf{B}_n$ such that $A \subsetneq C \subsetneq B$ (or the equivalent condition is satisfied with A and B reversed). A tree T_n derived in this fashion is called a fragmentation tree (with n leaves). See Figure 1.2 for an example of a fragmentation tree with 6 leaves. Note that we only need to keep track of the rooted tree structure and the leaf labels, since the sets that make up the internal nodes of the tree can be recovered from this information.

From both a combinatorial perspective and from the perspective of trees derived from fragmentation processes, it is natural to ask: If T_n is selected uniformly at random from the set of fragmentation trees with n leaves, what does T_n look like for large n ? In particular, can we derive a scaling limit for T_n in the sense discussed above? While T_n is a tree derived from a fragmentation processes, indeed it even falls into the family of Gibbs fragmentation trees [21], it lacks a consistency property that until recently has been essential for studying the asymptotics of these types of trees [16].

In our efforts to ascertain the asymptotic properties of T_n , we will be led to develop several different areas of the general theory of scaling limits of random trees. We start by

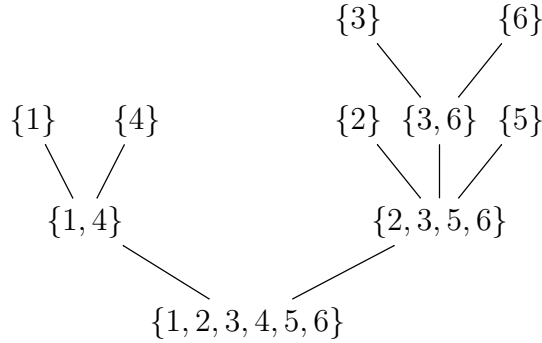


Figure 1.1: A fragmentation tree with 6 leaves

investigating the combinatorial properties of the set of fragmentation trees with n leaves and the law of T_n . This leads to two primary insights. The first is that the (exponential) generating function for the number of fragmentation trees with n leaves satisfies a nice functional equation. If we let c_n be the number fragmentation trees with n leaves and define

$$C(z) = \sum_{n \geq 1} c_n \frac{z^n}{n!},$$

then $C(z) = z + G(C(z))$ where $G(z) = \sum_{i \geq 2} z^i / i!$. See Section 2.2 for details. Generating function relations of this form have only appeared sparingly in the literature, but it allows for the machinery of analytic combinatorics to be applied to the study of T_n . In Section 2.3 we use the tools from analytic combinatorics to prove several asymptotic properties of T_n , such as the limiting distribution of the (appropriately normalized) height of a leaf chosen uniformly at random from T_n .

The second insight is that T_n is closely related, by a probabilistic transformation, to a particular Galton-Watson tree conditioned to have n leaves. Specifically, suppose that $\tilde{T}$ is distributed like a Galton-Watson tree with offspring distribution ξ given by

$$\xi_0 = \frac{2 \log(2) - 1}{\log(2)}, \quad \xi_1 = 0, \quad \text{and } \xi_j = \frac{(\log(2))^{j-1}}{j!} \text{ for } j \geq 2,$$

and $\tilde{T}_n$ is distributed like $\tilde{T}$ conditioned to have n leaves. We then have that T_n is equal in distribution to the tree obtained from $\tilde{T}_n$ by labeling its leaves from left to right by an independent uniform permutation of $[n]$ and forgetting the order structure of $\tilde{T}_n$. This is a special case of general result proved in Corollary 2. While scaling limits of Galton-Watson trees conditioned to have n vertices are well studied, Galton-Watson trees conditioned to have n leaves do not appear to have been studied before.

This connection to Galton-Watson trees allows for the translation of results about Galton-Watson trees into results about fragmentation trees. Thus, to obtain scaling limits of T_n , we may switch to studying Galton-Watson trees conditioned have n leaves. The advantage of this

is that Galton-Watson trees are naturally rooted ordered trees (as opposed to fragmentation trees which are labeled but unordered) and are nicely encoded by excursions of random walks. Nonetheless, the classical work on scaling limits of conditioned Galton-Watson trees relies heavily on the fact that the conditioning is on the number of vertices in the tree and not on the number of leaves. We deal with this difficulty by utilizing recent advances in the study of Markov branching trees [15]. The framework of Markov branching trees is inherently quite technical, so we will delay discussing it in detail until Section 3.2. In Chapter 3, we develop the theory of Markov branching trees to the point where we can, in Chapter 4, prove the following theorem for Galton-Watson trees.

Theorem. *Let T be a critical Galton-Watson tree with offspring distribution ξ such that $0 < \sigma^2 = \text{Var}(\xi) < \infty$ and let $A \subseteq \{0, 1, 2, \dots\}$ contain 0. Suppose that for sufficiently large n the probability that T has exactly n vertices with out-degree in A is positive, and for such n let T_n^A be T conditioned to have exactly n vertices with out-degree in A , considered as a rooted unordered tree with edge lengths 1 and the uniform probability distribution $\mu_{\partial_A T_n^A}$ on its vertices with out-degree in A . Then*

$$\frac{1}{\sqrt{n}} T_n^A \xrightarrow{d} \frac{2}{\sigma \sqrt{\xi(A)}} T^{Br},$$

where the convergence is with respect to the rooted Gromov-Hausdorff-Prokhorov topology and T^{Br} is the Brownian continuum random tree.

Taking $A = \{0\}$ in this theorem, we obtain scaling limits for Galton-Watson trees conditioned on their number of leaves, which in turn gives the scaling limit of the uniform fragmentation tree T_n that motivated this line of inquiry. When $A = \mathbb{N}$ we recover the classical case of scaling limits for Galton-Watson trees conditioned on their number of vertices (in fact, in this case our proof specializes to the proof given in [15]). For $A \neq \mathbb{N}$ the result is new. Note, however, that after a draft of the article [29] was posted on the arXiv, similar results were obtained using different methods by Kortchemski in [18].

1.3 Organization

The organization of this thesis is as follows. In Chapter 2 we introduce a general combinatorial model for random trees with n leaves. Our model is the natural adaption of the simply generated trees first introduced in [22], which are trees with n vertices, to the setting of trees with n leaves. Special cases of our model have appeared previously in the literature, for example in [12], typically under the name of uniform hierarchies. As such, no systematic study of the combinatorial properties of this model exists in the literature. The main purpose of Chapter 2 is to develop these properties and to prove relationships between this model and other models previously studied in the literature. We also show how the framework of analytic combinatorics, which has been a staple in the study of simply generated trees, can

be applied to our models of trees with n leaves. Most of the results of this chapter appear on the arXiv in a paper co-authored by Pitman and Rizzolo [28].

In Chapter 3 we introduce the notion of Markov branching trees and their scaling limits. The limiting objects are certain random compact pointed metric measure spaces that are commonly called continuum trees. We take some time to carefully formalize the notion of a metric space of compact pointed metric measure spaces because there is some confusion in the literature over how this is done. Our notion of Markov branching trees is a generalization of the Markov branching trees in [15] and the main result of Chapter 3 is an extension of the convergence results in [15] to our case.

Finally, in Chapter 4 we address scaling limits of conditioned Galton-Watson trees. In order to fit Galton-Watson trees conditioned on their number of vertices with out-degree in $A \subseteq \mathbb{N}$ (with $0 \in A$) into the framework for scaling limits of Markov branching trees that we develop, we also need to generalize the classical Otter-Dwass formula. To accomplish this, we are led to a class of deterministic transformations of the set of rooted ordered trees, interesting in their own right, that leave the family of Galton-Watson laws invariant. With the help of these transformations we will show that the number of vertices of a Galton-Watson tree with out-degree in A is distributed like the total number of vertices in a Galton-Watson tree with a related offspring distribution. Most of the results of Chapters 3 and 4 appear in a paper by Rizzolo [29] on the arXiv.

Chapter 2

Random trees with a given number of leaves

2.1 Schröder's problems

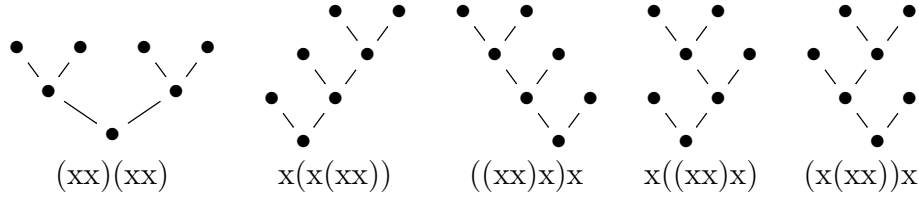
The example we gave in the introduction of uniform fragmentation trees is really a probabilistic variant of one of four related questions in enumerative combinatorics. In his now classic paper [30], Schröder posed four combinatorial problems about bracketings of words and sets: how many binary bracketings are there of a word of length n ? how many bracketings are there of a word of length n ? how many binary bracketings are there of a set of size n ? and how many bracketings are there of a set of size n ? These questions are well studied and [31] gives a good account of the solutions. The problem of uniform fragmentation trees is equivalent to the problem of, given a uniform pick from the bracketings of a set of size n , what does it look like? The same question can be asked in the setting of the other three problems as well. To answer these questions, we will use the well known correspondence of bracketings described above to various types of trees, which we now describe.

The first problem: The correspondence is best illustrated by example. For $n = 4$ the binary word bracketings are

$$(xx)(xx) \quad x(x(xx)) \quad ((xx)x)x \quad x((xx)x) \quad (x(xx))x.$$

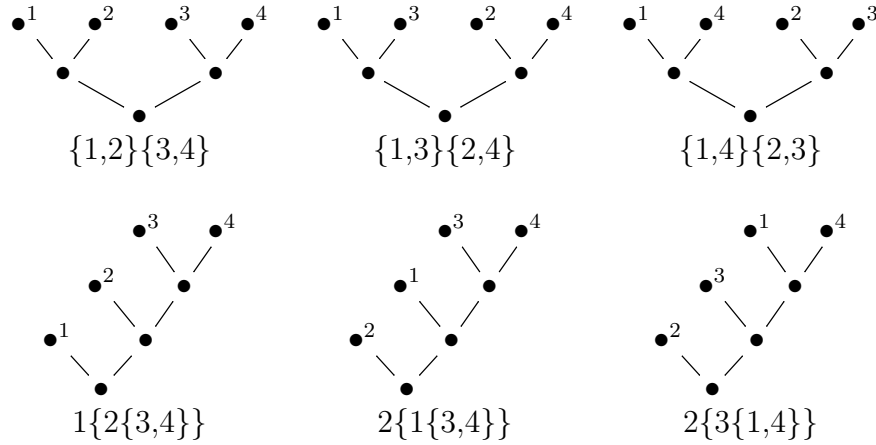
A binary bracketing of a word with n letters corresponds to rooted ordered binary tree with n leaves in a natural way. This is most easily described if we put brackets around the entire word and each letter, which are left out of our example because they are visually cumbersome. The tree corresponding to a bracketing is constructed recursively. A single bracketed letter is a leaf. For a word with more than one letter, the bracketing of the whole word is the root. Attached as subtrees to the root are, in order of appearance, the trees corresponding to the maximal proper bracketed subwords. For $n = 4$, this is illustrated by Figure 2.1.

It is worth noting that these trees are in bijection with rooted ordered trees with n vertices, but this correspondence is not as natural as the one above.


 Figure 2.1: Binary word bracketings and rooted ordered binary trees for $n = 4$

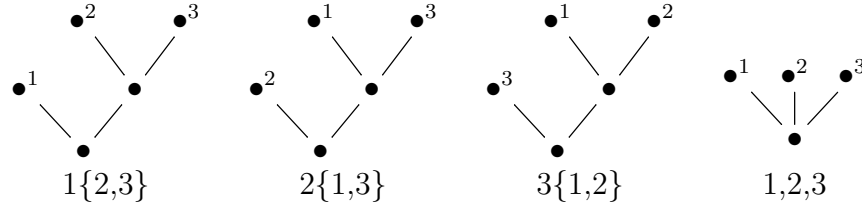
The second problem: General word bracketings are defined similarly to binary word bracketings and correspond to rooted ordered trees with n leaves and no vertices with out degree equal to one.

The third problem: The trees associated to binary set bracketings are constructed similarly to those associated to binary word bracketings. They are rooted, unordered, leaf-labeled binary trees. Figure 2.2 shows a sample of the correspondence for $n = 4$ (for $n = 4$ there are 15 bracketings, so showing the whole correspondence is unwieldy).


 Figure 2.2: Binary set bracketings and rooted unordered leaf-labeled binary trees for $n = 4$

The fourth problem: General set bracketings are defined similarly to binary set bracketings and correspond to rooted unordered leaf-labeled trees with n leaves and no vertices with out degree equal to one. In the literature, these trees are also called fragmentation trees [16] and hierarchies [12]. The correspondence for $n = 3$ is in Figure 2.3.

Scaling limits of uniform picks from the trees appearing in the first and third problems are well studied. A uniform pick from rooted ordered binary trees with n leaves has the same distribution as a Galton-Watson tree with offspring distribution $\xi_0 = \xi_2 = 1/2$ conditioned to have $2n - 1$ vertices. Thus it falls within the scope of the results in [3]. Similarly, a


 Figure 2.3: Set bracketings and rooted unordered leaf-labeled trees for $n = 3$

uniform pick from rooted unordered leaf-labeled binary tree's with n leaves is a uniform binary fragmentation tree with n leaves, and scaling limits of these are studied in [16]. In this thesis we present a unified approach that is able to handle all four of these types of trees simultaneously. In this chapter we develop the combinatorial theory necessary to work with these types of trees and show how tools from analytic combinatorics can be used to begin to obtain asymptotic results. We also draw a connection to certain Gibbs fragmentation trees, which were originally studied in [21]. The scaling limits for these trees are obtained in Theorem 23.

2.2 Combinatorial models and Galton-Watson trees

In this section we develop several combinatorial and probabilistic models of trees. There are two primary types of trees we will be dealing with in the sequel: rooted ordered unlabeled trees and rooted unordered leaf-labeled trees. Combinatorial relations between rooted ordered unlabeled trees and rooted unordered labeled trees are well known when the size of a tree is its number of vertices (see e.g. [27, 2, 12, 8]). In this section we develop analogous relations when the size of a tree is its number of leaves. Particularly important for us is Corollary 2, which relates Schröder's problems to particular Galton-Watson trees conditioned on their number of leaves.

We briefly give an account of the formal constructions of the trees we will be considering. Fix a countably infinite set S ; we will consider the vertex sets of all graphs discussed to be subsets of S . Let $\mathcal{T}_n$ denote the set of rooted unordered trees with n leaves (where the root is considered a leaf if and only if it is the only vertex in the tree) whose leaves are labeled by $\{1, 2, \dots, n\}$. More precisely, we consider the set $\mathcal{T}_n^S$ of all trees whose vertex sets are contained in S that have a distinguished root and n leaves, whose leaves are labeled by $\{1, 2, \dots, n\}$ and set $\mathcal{T}_n = \mathcal{T}_n^S / \sim$ where $t \sim s$ if there is a root and label preserving isomorphism from t to s . This is the only time we shall go through this formal construction, but all other sets of trees we discuss should be considered as formally constructed in an analogous fashion. We also let $\mathcal{T} = \cup_{n \geq 1} \mathcal{T}_n$. We let $\mathcal{T}_n^{(o)}$ be the set of rooted ordered unlabeled trees with n leaves and $\mathcal{T}^{(o)} = \cup_{n \geq 1} \mathcal{T}_n^{(o)}$.

We will be proving analogous results for trees in $\mathcal{T}$ and $\mathcal{T}^{(o)}$, so analogous that the only

difference in the statements will be the superscript (o) . To avoid repetition we will use $\mathcal{T}^*$ and $\mathcal{T}_n^*$ when we do not want to specify whether we are in $\mathcal{T}$, $\mathcal{T}^{(o)}$, $\mathcal{T}_n$ or $\mathcal{T}_n^{(o)}$. That is, in a given result you may replace all the $*$'s by nothing or (o) . For a tree $t \in \mathcal{T}^*$, we define $|t|$ to be the number of leaves in t and $\#t$ to be the number of vertices in t .

Probabilities on trees

Let $\zeta = (\zeta_i)_{i \geq 0}$ be a sequence of real numbers. We may then define the weight of a tree $t \in \mathcal{T}^*$ to be

$$w_\zeta(t) = \prod_{v \in t} \zeta_{\deg(v)}.$$

Here and throughout, $\deg(v)$ is the out degree of v , i.e., the number of children of v . We will assume the following conditions:

Condition 1. (i) $\zeta_i \geq 0$ for all i , (ii) $\zeta_0 > 0$, and (iii) for each n we have $\sum_{t \in \mathcal{T}_n^*} w_\zeta(t) < \infty$.

Observe that part (iii) of this condition is necessary because we do not require $\zeta_1 = 0$ and, consequently, it is possible for $\mathcal{T}_n^*$ to contain infinitely many trees with positive weight. For each n such that $w_\zeta(t) > 0$ for some $t \in \mathcal{T}_n^*$ we may define a probability measure on $\mathcal{T}_n^*$ by

$$Q_n^{\zeta^*}(t) = \frac{w_\zeta(t)}{\sum_{s \in \mathcal{T}_n^*} w_\zeta(s)}.$$

We wish to consider generating functions, but we want an ordinary generating function for $\mathcal{T}^{(o)}$ and an exponential generating function for $\mathcal{T}$. In order to do this all at once, for $z \in \mathbb{C}$, we define $y_n(z) = z^n/n!$ and $y_n^{(o)}(z) = z^n$, both for $n \geq 0$, and we use y_n^* in the same fashion as $\mathcal{T}^*$. The weighted generating function induced on $\mathcal{T}^*$ by ζ with the weights defined above is

$$C_\zeta^*(z) = \sum_{t \in \mathcal{T}^*} w_\zeta(t) y_{|t|}^*(z).$$

Let $G_{\zeta,*}(z) = \sum_{i=1}^{\infty} \zeta_i y_i^*(z)$

Theorem 1. C_ζ^* satisfies the functional equation

$$C_\zeta^*(z) = \zeta_0 z + G_{\zeta,*}(C_\zeta^*(z)), \quad (2.1)$$

in the sense of formal power series.

This is very straightforward when $*$ = (o) , but in the case when $*$ is nothing, it is somewhat technical to deal with the labeling rigorously. Thus we include the proof for the case when $*$ is nothing.

Proof. Suppose $T \in \mathcal{T}$ and let v be a vertex of T . Define T_v to be the subtree of T whose vertices are v together with the vertices above v in T such that v is distinguished as the root and leaves of T_v are labeled as they were in T . Order the vertices of height 1 in T by least label; that is let $\{v_1, \dots, v_k\}$ be the vertices of height 1 indexed such that the smallest leaf label in T_{v_i} is smaller than the smallest leaf label in T_{v_j} when $i < j$. With this ordering, we will denote the root-subtree T_{v_i} by T_i . Also, we denote by B^{T_i} the set of labels of leaves in T_i . Observe that a tree is uniquely determined by its root-subtrees.

More precisely, we define an ordered k -forest of rooted leaf-labeled trees to be an ordered k -tuple $(t_1, \dots, t_k)$ of rooted leaf-labeled trees where $(B^{t_1}, \dots, B^{t_k})$ is an ordered partition of $[|t_1| + \dots + |t_k|]$. Let $\mathcal{F}_k$ be the set of all labeled k -forests of rooted leaf-labeled trees and let $\mathcal{F}_k^o$ be the subset of $\mathcal{F}_k$ of k -forests such that the least element of B^{t_i} is less than the least element of B^{t_j} when $i < j$. The map $F_k(T) = (T_1, \dots, T_k)$ is a bijection between trees in $\mathcal{T}$ with root degree k and $\mathcal{F}_k^o$. Furthermore, letting S_k be the symmetric group on k elements we have a bijection $J_k : S_k \times \mathcal{F}_k^o \rightarrow \mathcal{F}_k$ given by

$$J_k(\sigma; (T_1, \dots, T_k)) = (T_{\sigma(1)}, \dots, T_{\sigma(k)}).$$

Suppose that T is a rooted unordered tree with n leaves whose leaves are labeled by $B \subseteq \mathbb{N}$. The leaves of T are then naturally ordered by the size of their labels. A reduced tree of T is the tree $\tilde{T} \in \mathcal{T}_n$ for which there is a rooted graph isomorphism $\Lambda : T \rightarrow \tilde{T}$ that preserves the order of the leaves. It is immediate that reduced trees exist, are unique, and there is a bijection between trees labeled by B and their reduced trees. Let $\mathcal{P}_{n_1, \dots, n_k}^n$ be the set of ordered partitions of n into blocks of sizes $(n_1, \dots, n_k)$. Then, for each $x \in \mathcal{T}^k$ we have, a bijection

$$H_x : \{(t_1, \dots, t_k) \in \mathcal{F}_k : (\tilde{t}_1, \dots, \tilde{t}_k) = x\} \rightarrow \mathcal{P}_{|x_1|, \dots, |x_k|}^{|x|},$$

given by

$$H_x(t_1, \dots, t_k) = (B^{t_1}, \dots, B^{t_k}).$$

With this notation we can, in excruciating detail, make the following computation

$$\begin{aligned}
 C_\zeta(z) &= \zeta_0 z + \sum_{r=1}^{\infty} \sum_{\{T \in \mathcal{T} \mid \deg(\text{root}(T))=r\}} w_\zeta(T) \frac{z^{|T|}}{|T|!} \\
 &= \zeta_0 z + \sum_{r=1}^{\infty} \zeta_r \left(\sum_{\{T \in \mathcal{T} \mid \deg(\text{root}(T))=r\}} \prod_{i=1}^r w_\zeta(T_i) \frac{z^{|T|}}{|T|!} \right) \\
 &= \zeta_0 z + \sum_{r=1}^{\infty} \zeta_r \left(\sum_{(t_1, \dots, t_r) \in \mathcal{F}_r^o} \prod_{i=1}^r w_\zeta(t_i) \frac{z^{|t_1| + \dots + |t_r|}}{(|t_1| + \dots + |t_r|)!} \right) \\
 &= \zeta_0 z + \sum_{r=1}^{\infty} \frac{\zeta_r}{r!} \left(\sum_{(\sigma; (t_1, \dots, t_r)) \in S_r \times \mathcal{F}_r^o} \prod_{i=1}^r w_\zeta(t_{\sigma(i)}) \frac{z^{|t_{\sigma(1)}| + \dots + |t_{\sigma(r)}|}}{(|t_{\sigma(1)}| + \dots + |t_{\sigma(r)}|)!} \right) \\
 &= \zeta_0 z + \sum_{r=1}^{\infty} \frac{\zeta_r}{r!} \left(\sum_{(t_1, \dots, t_r) \in \mathcal{F}_r} \prod_{i=1}^r w_\zeta(t_i) \frac{z^{|t_1| + \dots + |t_r|}}{(|t_1| + \dots + |t_r|)!} \right) \\
 &= \zeta_0 z + \sum_{r=1}^{\infty} \frac{\zeta_r}{r!} \left(\sum_{x=(x_1, \dots, x_r) \in \mathcal{T}^r} \sum_{\{(t_1, \dots, t_k) \in \mathcal{F}_k : (\tilde{t}_1, \dots, \tilde{t}_k) = x\}} \prod_{i=1}^r w_\zeta(t_i) \frac{z^{|t_1| + \dots + |t_r|}}{(|t_1| + \dots + |t_r|)!} \right) \\
 &= \zeta_0 z + \sum_{r=1}^{\infty} \frac{\zeta_r}{r!} \left(\sum_{(x_1, \dots, x_r) \in \mathcal{T}^r} \left(\sum_{i=1}^r |x_i| \right) \prod_{i=1}^r w_\zeta(x_i) \frac{z^{|x_1| + \dots + |x_r|}}{(|x_1| + \dots + |x_r|)!} \right) \\
 &= \zeta_0 z + \sum_{r=1}^{\infty} \frac{\zeta_r}{r!} C_\zeta(z)^r = z + G_\zeta(C_\zeta(z)).
 \end{aligned} \tag{2.2}$$

□

Note, implicit in our computation is the following relation:

$$C_\zeta^{(r)}(z) = \sum_{\{T \in \mathcal{T} \mid \deg(\text{root}(T))=r\}} w_\zeta(T) \frac{z^{|T|}}{|T|!} = \frac{\zeta_r}{r!} C_\zeta(z)^r, \tag{2.3}$$

where $C_\zeta^{(r)}(z)$ is the weighted generating function for trees with root-degree r .

Our interest is in the measures Q_n^ζ and, in particular, we would like to find a Galton-Watson tree T such that Q_n^ζ is the law of T conditioned to have n leaves. Recall that if $(\xi_i)_{i \geq 0}$ is a distribution on $\mathbb{Z}_+$ with mean less than or equal to one and $\xi_0 > 0$, a Galton-Watson tree with offspring distribution ξ is a random element T of $\mathcal{T}^{(o)}$ with law

$$\mathbb{P}(T = t) = \prod_{v \in t} \xi_{\deg(v)}.$$

T is called critical if ξ has mean equal to one. This leads to the notion of tilting, which is similar to exponential tilting for Galton-Watson trees conditioned on their number of vertices.

Proposition 1. *Suppose that ζ satisfies Condition 1 and suppose that $a, b > 0$. Define $\tilde{\zeta}$ by*

$$\tilde{\zeta}_0 = a\zeta_0 \quad \text{and} \quad \tilde{\zeta}_i = b^{i-1}\zeta_i \quad \text{for } i \geq 1.$$

Then $Q_n^{\zeta^*} = Q_n^{\tilde{\zeta}^*}$ for all $n \geq 1$.

Proof. This follows immediately from the computation that, for $t \in \mathcal{T}_n^*$, $w_{\tilde{\zeta}}(t) = a^n b^{n-1} w_{\zeta}(t)$. $\square$

The above result is the equivalent of exponential tilting for trees conditioned on their number of vertices. A consequence of this is that we can find a Galton-Watson tree T such that $Q_n^{\zeta^{(o)}}$ is the law of T conditioned to have n leaves if we can find $a, b > 0$ such that

$$a\zeta_0 + \frac{G_{\zeta, (o)}(b)}{b} = 1.$$

Furthermore, T will be critical if $G_{\zeta, (o)}'(b) = 1$. An immediate consequence of this is the following corollary.

Corollary 1. *Let $\xi^u = (\xi_i^u)_{i=0}^\infty$ be the probability distribution defined by*

$$\xi_0^u = 2 - \sqrt{2} \approx 0.5858, \quad \xi_1^u = 0, \quad \text{and} \quad \xi_i^u = \left(\frac{2 - \sqrt{2}}{2} \right)^{i-1} \approx (0.2929)^{i-1} \quad \text{for } i \geq 2.$$

Note that ξ^u has mean 1 and variance $4\sqrt{2}$. Let T be a Galton-Watson tree with offspring distribution ξ^u . Then the law of T conditioned to have n leaves is uniform on the subset of $\mathcal{T}_n^{(o)}$ of trees with no vertices of out degree one.

Proof. The proof follows immediately from the discussion above by noting that, if $\zeta_i = 1$ for $i \neq 1$ and $\zeta_1 = 0$ then $Q_n^{\zeta^{(o)}}$ is uniform on $\mathcal{T}_n^{(o)}$. Explicitly, the distribution ξ^u is found by solving $G'_{\zeta, (o)}(b) = 1$, setting $a = (b - G_{\zeta, (o)}(b))/b$, and tilting as in Proposition 1. $\square$

Given the similarities in the constructions of Q_n^ζ and $Q_n^{\zeta^{(o)}}$, there should be a natural way to go back and forth between them.

Proposition 2. *Suppose that ζ satisfies Condition 1 for $* = (o)$. Define $\hat{\zeta}$ by $\hat{\zeta}_n = n!\zeta_n$. Suppose that T is distributed like $Q_n^{\zeta^{(o)}}$ and let U be a uniformly random ordering of $\{1, \dots, n\}$ independent of T . Define $\hat{T} \in \mathcal{T}_n$ to be the tree obtained from T by labeling the leaves of T by U and forgetting the ordering of T . Then $\hat{T}$ is distributed like $Q_n^{\hat{\zeta}}$.*

Results of this type connecting plane and labeled trees where the size of a tree is given by the number of its vertices can be traced back to [17, 24, 25]. See [27] for a more complete history. Our proposition is analogous to an implicit discussion in [2, Section 2.1] as well as Theorem 7.1 in [27], which considered the case where the size of a tree is given by the number of its vertices. To prove this proposition, we will need some notation. For a rooted ordered tree x let $\text{shape}(x)$ be the rooted unordered tree obtained by forgetting the order on x . Similarly, for $t \in \mathcal{T}$, $\text{shape}(t)$ is defined to be the rooted unlabeled tree obtained from forgetting the labeling of t . For $t \in \mathcal{T}$, $x \in \mathcal{T}_n^{(o)}$, and a rooted unordered tree y define $\#\text{labels}_t(x)$ to be the number of ways to label the leaves of x such that when you forget the order on x you get t and $\#\text{ordered}(y)$ to be the number of ordered trees whose shape is y . Observe that $\#\text{labels}_t(x)$ depends only on $\text{shape}(x)$, so we will abuse our notation and write $\#\text{labels}_t(\text{shape}(x))$.

Proof. Fix $t \in \mathcal{T}_n$. Observe that

$$\mathbb{P}(\hat{T} = t) = \sum_{x \in \mathcal{T}_n^{(o)}} \mathbb{P}(T = x) \mathbb{P}(\hat{T} = t | T = x).$$

Furthermore, observe that

$$\mathbb{P}(\hat{T} = t | T = x) = \frac{\#\text{labels}_t(\text{shape}(x))}{n!},$$

Observe that $\#\text{labels}_t(\text{shape}(x)) = 0$ unless $\text{shape}(t) = \text{shape}(x)$. Furthermore, $\mathbb{P}(T = x)$ depends only on $\text{shape}(x)$, and is given by

$$\mathbb{P}(T = x) = \frac{\prod_{v \in \text{shape}(x)} \zeta_{\deg(v)}}{\sum_{s \in \mathcal{T}_n^{(o)}} w_\zeta(s)}.$$

Consequently we have

$$\mathbb{P}(\hat{T} = t) = \frac{\#\text{ordered}(\text{shape}(t)) \prod_{v \in \text{shape}(t)} \zeta_{\deg(v)} \frac{\#\text{labels}_t(\text{shape}(t))}{n!}}{\sum_{s \in \mathcal{T}_n^{(o)}} w_\zeta(s)}. \quad (2.4)$$

But

$$\#\text{ordered}(\text{shape}(t)) \#\text{labels}_t(\text{shape}(t)) = \prod_{v \in \text{shape}(t)} (\deg(v)!). \quad (2.5)$$

This is because both sides count the number of distinct leaf-labeled ordered trees that equal t upon forgetting their order. On the left hand side, you pick a ordered tree and the label it and, on the right hand side, you label an unordered tree with the appropriate shape and then order the children of each vertex.

Therefore we have

$$\mathbb{P}(\hat{T} = t) = \frac{w_\zeta(t)}{n! \sum_{s \in \mathcal{T}_n^{(o)}} w_\zeta(s)}.$$

The last step is to observe that

$$n! \sum_{s \in \mathcal{T}_n^{(o)}} w_\zeta(s) = \sum_{s \in \mathcal{T}_n} w_{\hat{\zeta}}(s).$$

This is because for $s \in \mathcal{T}_n^{(o)}$, there are $n!$ rooted ordered leaf-labeled trees whose ordered tree is s upon forgetting the labeling, so the left hand side is the weighted number of rooted ordered leaf-labeled trees with n leaves. Furthermore, we have already noted above that for $s \in \mathcal{T}_n$, there are $\prod_{v \in s} (\deg(v)!)^n$ rooted ordered leaf-labeled trees whose labeled tree is s upon forgetting the ordering. Thus the right hand side is also the weighted number of rooted ordered leaf-labeled trees with n leaves. Note that this step also shows that $\hat{\zeta}$ satisfies Condition 1 for $*$ being nothing. $\square$

Combining with tilting, we have the following corollary.

Corollary 2. *Let ζ satisfy Condition 1 with $*$ being nothing and $\zeta_0 = 1$. Suppose there exist $r > 0$ and $s > 0$ satisfying $s = r + G_\xi(s)$ and $G'_\xi(s) \leq 1$. Define $\xi = (\xi_i)_{i=0}^\infty$ by $\xi_0 = rs^{-1}$ and $\xi_j = s^{j-1}\zeta_j/j!$ for $j \geq 1$. Note that ξ is a probability distribution on $\mathbb{Z}_+$. Let T be a Galton-Watson tree with offspring distribution ξ and construct $\hat{T}$ by labeling the leaves of T uniformly at random with $\{1, \dots, |T|\}$, independently of T and forgetting the order of T . Then $\mathbb{P}(\hat{T} \in \cdot | |T| = n) = Q_n^\zeta(\cdot)$ for all $n \geq 1$ such that Q_n^ζ is defined. Furthermore, for n such that Q_n^ζ is not defined, $P(|T| = n) = 0$.*

Schröder's problems

In this section we record which of the trees above correspond to the trees that appear in Schröder's problems. The proofs of the claims here are simple applications of the results in Section 2.2.

The first problem: The trees here are uniform binary rooted ordered unlabeled trees. We can obtain these by taking $*$ = (o) and $\zeta_0 = \zeta_2 = 1$ and $\zeta_i = 0$ for $i \notin \{0, 2\}$. Letting ξ be the probability distribution given by $\xi_0 = \xi_2 = 1/2$ and T be a Galton-Watson tree with offspring distribution ξ , we have that T conditioned to have n leaves is a uniform binary rooted ordered unlabeled tree with n leaves. Also note that T is critical and the variance of ξ is equal to one.

The second problem: These are uniform rooted ordered trees with no vertices of out degree one. These were dealt with in Corollary 1

The third problem: These are uniform binary unordered leaf-labeled trees. We can obtain these by taking $*$ to be nothing and $\zeta_0 = \zeta_2 = 1$ and $\zeta_i = 0$ for $i \notin \{0, 2\}$. In this case, if T is the Galton-Watson tree defined in the first problem and $\hat{T}$ is defined as in Corollary 2, then $\hat{T}$ conditioned to have n leaves is a uniform binary unordered leaf-labeled tree with n leaves.

The fourth problem: These are uniform rooted unordered leaf-labeled trees with no vertices with out-degree 1. We can obtain these by taking $*$ to be nothing and $\zeta_1 = 0$ and $\zeta_i = 1$ for $i \neq 1$. We define a probability distribution ξ by

$$\xi_0 = \frac{2 \log(2) - 1}{\log(2)}, \quad \xi_1 = 0, \quad \text{and } \xi_j = \frac{(\log(2))^{j-1}}{j!} \text{ for } j \geq 2.$$

Note that ξ has mean 1 and variance $\text{Var}(\xi) = 2 \log 2$. Letting T be a Galton-Watson tree with offspring distribution ξ and defining $\hat{T}$ is as in Corollary 2, we have that $\hat{T}$ conditioned to have n leaves is a uniform unordered leaf-labeled tree with no vertices of out degree one and n leaves.

Gibbs trees

Above we saw a natural way to put probability measures on $\mathcal{T}_n$ that are concentrated on fragmentation trees (the trees appearing in Schröder's fourth problem); namely, take $\zeta_1 = 0$. Another natural type of probability to put on fragmentation trees is a *Gibbs model*, which we now describe. First, we need to set up the natural framework in which to view fragmentation trees. The idea is that, while in Schröder's fourth problem we have an arbitrary set bracketing, for fragmentations we recursively partition a set. This dynamic view of constructing a set bracketing makes Gibbs models quite natural.

Definition 1 ([21]). *A fragmentation of the finite set B is a collection $\mathbf{t}_B$ of non-empty subsets of B such that*

1. $B \in \mathbf{t}_B$
2. *If $\#B \geq 2$ then there is a partition of B into $k \geq 2$ parts $B_1, \dots, B_k$, called the children of B , such that*

$$\mathbf{t}_B = \{B\} \cup \mathbf{t}_{B_1} \cup \dots \cup \mathbf{t}_{B_k},$$

where $\mathbf{t}_{B_i}$ is a fragmentation of B_i .

We can naturally consider $\mathbf{t}_B$ as a tree whose vertices are the elements of $\mathbf{t}_B$ and whose edges are defined by the parent-child relationship. Considering the properties of such a tree leads naturally to the following definition of a *fragmentation tree* on B .

Definition 2. *A fragmentation tree T on n leaves is a rooted tree such that*

1. *The root of T does not have degree 1,*
2. *T has no non-root vertices of degree 2,*
3. *The leaves of T are labeled by a set B with $\#B = n$. We denote the label of a leaf v by $\ell(v)$.*

The idea of the Gibbs model is that, at each step in the fragmentation the next step is distributed according to multiplicative weights depending on the block sizes. We first take a sequence $\{\alpha_n\}$, $\alpha_n \geq 0$ of weights and a Gibbs weight, which is a function $g : \mathbb{Z}_+ \rightarrow \mathbb{R}_+$ with $g(0) = 0$ and $g(1) > 0$. Then, for $n \geq 2$, define a normalization constant

$$Z(n) = \sum_{\{B_1, \dots, B_k\}} \alpha_k \prod_{j=1}^k g(\#B_j),$$

where the sum is over unordered partitions of $[n]$ into at least two elements. Whenever we write a formula like this, we assume that each block B_i is nonempty. Now, assuming $Z(n) > 0$, define the probability of a partition of $[n]$ by

$$P_n^{g, \alpha}(B_1, \dots, B_k) = \frac{\alpha_k \prod_{j=1}^k g(\#B_j)}{Z(n)}.$$

The probability of a fragmentation X of $[n]$ is then defined as

$$P_n^{g, \alpha}(X) = \prod_{B \in X} P_n^{g, \alpha}(B_1, \dots, B_k),$$

where $B_1, \dots, B_k$ are the children of B . Using the correspondence between fragmentations and fragmentation trees, for $T_n \in \mathcal{T}_n$, we define $P_n^{g, \alpha}(T_n)$ to be $P_n^{g, \alpha}(X)$ where X is the fragmentation determined by T_n . The probabilistic properties of Gibbs models are studied in [21].

Theorem 2. *Suppose that ζ satisfies Condition 1 with $*$ being nothing and $\zeta_1 = 0$. Define $\alpha_k = \zeta_k$ and $g(k) = k![z^k]C_\zeta(z)$. Then $Z(n) = g(n)$ and $Q_n^\zeta = P_n^{g, \alpha}$. Furthermore, given a nonnegative weight sequence α and a Gibbs weight g such that $Z(n) = g(n)$, there is a ζ satisfying Condition 1 with $*$ being nothing and $\zeta_1 = 0$ such that $Q_n^\zeta = P_n^{g, \alpha}$.*

Proof. Arguing similarly as in (2.1), we see that for $n \geq 2$

$$Z(n) = \sum_{\{B_1, \dots, B_k\}} \alpha_k \prod_{j=1}^k g(\#B_j) = \sum_{k=2}^{\infty} \frac{\alpha_k}{k!} \sum_{\substack{(n_1, \dots, n_k) \in \mathbb{N}^k \\ n_1 + \dots + n_k = n}} \binom{n}{n_1, \dots, n_k} \prod_{j=1}^k g(n_j) = n![z^n]C_\zeta(z).$$

Consequently we have $Z(n) = g(n)$. Using this, one proves inductively that $P_n^{g, \alpha}(T_n) = Q_n^\zeta(T_n)$. Furthermore, observe that the condition $Z(n) = g(n)$ implies that there is a weight sequence $(\zeta_i)_{i \geq 0}$ from which the fragmentation model can be derived in the above manner; just take $\zeta_0 = g(1)$, $\zeta_1 = 0$, and $\zeta_k = \alpha_k$ for $k \geq 2$. $\square$

When we have $Z(n) = g(n)$, the model is called a *combinatorial* Gibbs model. This is justified by the fact that, in this case, $Z(n)$ (and thus $g(n)$) is the weighted number of trees

with n leaves. For example, if we let $g(n)$ be the number of fragmentation trees with n leaves, and $\alpha_k = 1$ for $k \geq 2$, we then see that

$$Z(n) = \sum_{\{B_1, \dots, B_k\}} \prod_{j=1}^k g(\#B_j).$$

The right hand side of this equation is just the sum over partitions at the root of a fragmentation tree with n leaves of the number of fragmentation trees with that partition at the root, which is precisely the number of fragmentation trees with n leaves. That is, $Z(n) = g(n)$.

Note that combinatorial Gibbs models are a generalization of the hierarchies studied in [12] and, as previously observed, a special case of the Gibbs models introduced in [21].

2.3 Explicit computations using analytic combinatorics

In this section we demonstrate that the models of random trees we have been discussing are amenable to the methods of analytic combinatorics. In particular, we will use tools from analytic combinatorics to obtain asymptotic results about several natural statistics of these trees. This analytic approach is based on considering the asymptotics of generating functions. The primary source for asymptotics in general is [12], which develops the theory with extensive examples.

Our main goal in this section is to develop the general framework of additive functionals for leaf-labeled trees whose size is counted by their number of leaves. We use this to compute a number of asymptotic results, such as the height of a uniformly randomly chosen leaf, the number of leaves or vertices at a fixed level, and the degree of the root. These computations are meant to be illustrative and by no means exhaust the power of analytic combinatorics framework. Indeed, it seems that most of the techniques used to study simple varieties of trees (see [12] for a summary of the extensive work in this area) have close analogs that will provide results about the trees we are considering here.

Analytic background

We summarize some of the fundamental results here, but make no attempt to prove them. The approach is based on the asymptotics of several universal functions. Recall that if $f(z)$ is either a formal power series, $[z^n]f(z)$ denotes the coefficient of z^n . Similarly, if $f : \mathbb{C} \rightarrow \mathbb{C}$ is analytic at 0 then $[z^n]f(z)$ denotes the coefficient of z^n in the power series expansion of f at 0.

Proposition 3. *Let $f(z) = (1 - z)^{1/2}$, $g(z) = (1 - z)^{-1/2}$, and $h(z) = (1 - z)^{-1}$. Then $[z^n]f(z) \sim -1/2\sqrt{\pi n^3}$, $[z^n]g(z) \sim 1/\sqrt{n\pi}$, and $[z^n]h(z) = 1$.*

To use these classical results we need a special type of analyticity called Δ -analyticity, which we now define.

Definition 3 (Definition VI.I p. 389 [12]). *Given two numbers ϕ and R with $R > 1$ and $0 < \phi < \pi/2$, the open domain $\Delta(\phi, R)$ is defined as*

$$\Delta(\phi, R) = \{z \mid |z| < R, z \neq 1, |\arg(z - 1)| > \phi\}.$$

For a complex number ζ a domain D is a Δ -domain at ζ if there exist ϕ and R such that $D = \zeta\Delta(\phi, R)$. A function is Δ -analytic if it is analytic on a Δ -domain.

Let

$$\mathcal{S} = \{(1 - z)^{-\alpha} \lambda(z)^\beta \mid \alpha, \beta \in \mathbb{C}\} \quad \text{where} \quad \lambda(z) \equiv \frac{1}{z} \log \frac{1}{1 - z}.$$

Theorem 3 (Theorem VI.4 p. 393 [12]). *Let $f(z)$ be a function analytic at 0 with a singularity at ζ , such that $f(z)$ can be continued to a domain of the form $\zeta\Delta_0$, for a Δ -domain Δ_0 . Assume that there exist two function σ and τ , where σ is a (finite) linear combination of elements of $\mathcal{S}$ and $\tau \in \mathcal{S}$, so that*

$$f(z) = \sigma(z/\zeta) + O(\tau(z/\zeta)) \quad \text{as } z \rightarrow \zeta \text{ in } \zeta\Delta_0.$$

Then the coefficients $f_n = [z^n]f(z)$ of $f(z)$ satisfy the asymptotic estimate

$$f_n = \zeta^{-n} \sigma_n + O(\zeta^{-n} \tau_n^*),$$

where $\tau_n^* = n^{a-1}(\log n)^b$, if $\tau(z) = (1 - z)^{-a} \lambda(z)^b$.

Occasionally we will also need to deal with derivatives and the next theorem shows us how this is done.

Theorem 4 (Theorem VI.8 p. 419 [12]). *Let $f(z)$ be Δ -analytic with singular expansion near its singularity of the simple form*

$$f(z) = \sum_{j=0}^J c_j (1 - z)^{a_j} + O((1 - z)^A),$$

with $a_j \geq 0$ for all j . Then, for each integer $r > 0$, the derivative $f^{(r)}(z)$ is Δ -analytic. The expansion of the derivative at its singularity is obtained through term by term differentiation:

$$\frac{d^r}{dz^r} f(z) = (-1)^r \sum_{j=0}^J c_j \frac{\Gamma(a_j + 1)}{\Gamma(a_j + 1 - r)} (1 - z)^{a_j - r} + O((1 - z)^{A-r}).$$

The generating functions we will work with fall into the smooth implicit-function schema, which provides a way to derive coefficient asymptotics from functional equations.

Definition 4 (Definition VII.4 p. 467 [12]). Let $y(z)$ be a function analytic at 0, $y(z) = \sum_{n \geq 0} y_n z^n$, with $y_0 = 0$ and $y_n \geq 0$. The function is said to belong to the smooth implicit-function schema if there exists a bivariate function $G(z, w)$ such that

$$y(z) = G(z, y(z)),$$

where $G(z, w)$ satisfies the following conditions.

- (i) $G(z, w) = \sum_{m, n \geq 0} g_{m, n} z^m w^n$ is analytic in a domain $|z| < R$ and $|w| < S$, for some $R, S > 0$.
- (ii) The coefficients of G satisfy $g_{m, n} \geq 0$, $g_{0, 0} = 0$, $g_{0, 1} \neq 1$, and $g_{m, n} > 0$ for some m and for some $n \geq 2$.
- (iii) There exist two numbers r and s such that $0 < r < R$ and $0 < s < S$, satisfying the system of equations

$$G(r, s) = s, \quad G_w(r, s) = 1, \quad \text{with } r < R, \ s < S,$$

which is called the characteristic system.

Definition 5 (Definition IV.5 p. 266 [12]). Consider the formal power series $f(z) = \sum f_n z^n$. The series f is said to admit span d if for some r

$$\{f_n\}_{n=0}^{\infty} \subseteq r + d\mathbb{Z}_+.$$

The largest span is the period of f . If f has period 1, then f is aperiodic.

With this definition, we get the following theorem. It is worth noting that this result appears in several places in the literature. We give the version that appears as Theorem VII.3 on page 468 of [12]. In that source it is footnoted that many statements occurring previously in the literature contained errors, so caution is advised.

Theorem 5 (Theorem VII.3 p. 468 [12]). Let $y(z)$ belong to the smooth implicit-function schema defined by $G(z, w)$, with (r, s) the positive solution of the characteristic system. Then $y(z)$ converges at $z = r$, where it has a square root singularity,

$$y(z) \underset{z \rightarrow r}{=} s - \gamma \sqrt{1 - z/r} + O(1 - z/r), \quad \gamma \equiv \sqrt{\frac{2rG_z(r, s)}{G_{ww}(r, s)}},$$

the expansion being valid in a Δ -domain. If, in addition, $y(z)$ is aperiodic, then r is the unique dominant singularity of y and the coefficients satisfy

$$[z^n]y(z) \underset{n \rightarrow \infty}{=} \frac{\gamma}{2\sqrt{\pi n^3}} r^{-n} (1 + O(n^{-1})).$$

We will also need the following theorem.

Theorem 6 ((A special case of) Theorem IX.16 p. 709 [12]). *Let $H(z)$ be Δ -continuable and of the form $H(z) = \sigma - h(1 - z/\rho)^{1/2} + O(1 - z/\rho)$ and let $k_n = x_n n^{1/2}$ for x_n in any compact subinterval of $(0, \infty)$. Then*

$$[z^n]H(z)^{k_n} \sim \sigma^{k_n} \rho^{-n} \frac{h k_n}{2\sigma \sqrt{\pi n^3}} \exp\left(-\frac{h^2 k_n^2}{4\sigma^2 n}\right).$$

Restricting the generality

So far we have been considering a very general situation. However, in what follows we will be doing computations that are tedious to do in full generality. Consequently, we will restrict the generality. In particular, we will let $\zeta = (\zeta_i)_{i \geq 0}$ be a sequence of non-negative weights such that $\zeta_0 = 1$, $\zeta_1 = 0$, $\gcd\{k : \zeta_k \neq 0\} = 1$, and

$$G_\zeta(z) = \sum_{j=2}^{\infty} \zeta_j \frac{z^j}{j!}$$

is entire. These conditions can be relaxed, but doing so makes the analysis more difficult.

Proposition 4. *With ζ as above, C_ζ (defined in Section 2.2) belongs to the smooth implicit-function schema with $G(z, w) = z + G_\zeta(w)$. Furthermore, in the case where ζ corresponds to Schröder's third problem $(r, s) = (1/2, 1)$ and in the case of the fourth problem, we have $(r, s) = (2 \log(2) - 1, \log(2))$. Additionally*

$$[z^n]C_\zeta(z) \sim \frac{\gamma}{2\sqrt{\pi n^3}} r^{-n}, \quad \gamma = \sqrt{\frac{2r}{G_\zeta''(s)}}.$$

Proof. All that really needs to be checked is that the characteristic system has a positive solution. For $G(z, w) = z + G_\zeta(w)$, the characteristic system is $s = r + G_\zeta(s)$ and $G_\zeta'(s) = 1$. Using that G_ζ is entire, $G_\zeta'(0) = 0$ and $G_\zeta'(+\infty) = +\infty$, and G_ζ' is increasing on $\mathbb{R}_+$, the intermediate value theorem yields $s > 0$. An easy computation yields that $G_\zeta(s) < s G_\zeta'(s) = s$, so $r > 0$ as well. $\square$

Root Degrees

As our first application, we consider the distribution of root degree. For n such that $\mathcal{T}_n$ is nonempty, let Π_n be the partition of $[n]$ given by the labels of vertices of T_n of height 1. The main result of this section is that the number of blocks in Π_n converges in distribution without normalization.

Theorem 7. *Let T_n have law Q_n^ζ and let X be distributed like $P(X = k) = \zeta_k s^{k-1} / (k-1)!$. Then $\#\Pi_n \xrightarrow{d} X$.*

Proof. To see that the definition of the distribution of X is valid, note that s is the unique positive number such that $\zeta'(s) = 1$ and consequently

$$\sum_{k=2}^{\infty} \zeta_k \frac{s^{k-1}}{(k-1)!} = G'_\zeta(s) = 1.$$

Recall that, in equation (2.3), we derived the equation

$$C_\zeta^{(k)}(z) = \zeta_k \frac{C_\zeta(z)^k}{k!}.$$

Consequently for $n \geq 2$,

$$P(\#\Pi_n = k) = \zeta_k \mathbf{1}(k \in [2, n]) \frac{[z^n]C_\zeta(z)^k}{k! [z^n]C_\zeta(z)}.$$

By Theorem 5 and Proposition 4 we have that

$$C_\zeta(z) \underset{z \rightarrow r}{=} s - \gamma \sqrt{1 - z/r} + O(1 - z/r), \quad \gamma = \sqrt{\frac{2r}{G''_\zeta(s)}}. \quad (2.6)$$

Therefore

$$C_\zeta(z)^k = s^k - s^{k-1} k \gamma \sqrt{1 - z/r} + O(1 - z/r).$$

Consequently we have that

$$P(\#\Pi_n = k) = \zeta_k \mathbf{1}(k \in [0, n]) \frac{[z^n]C_\zeta(z)^k}{k! [z^n]C_\zeta(z)} \sim \zeta_k \frac{s^{k-1}}{(k-1)!}.$$

The result for the uniform cases follows from the fact that $s = \log(2)$ in this case. $\square$

The height of a random leaf

Let H_n be the height of a randomly chosen leaf from a tree in $\mathcal{T}_n$. Specifically, to get H_n , we choose a tree T_n from $\mathcal{T}_n$ according to P_n and then choose a leaf uniformly at random from T_n . Our main result in this section is the following theorem.

Theorem 8.

$$\frac{\lambda}{\sqrt{n}} H_n \xrightarrow{d} \text{Rayleigh}(1),$$

for $\lambda = \sqrt{G''_\zeta(s)r}$. In Schröder's third problem $\lambda = 1/\sqrt{2}$, and in the fourth problem $\lambda = \sqrt{4 \log(2) - 2}$.

Our approach will be that of additive functionals, whose theory we now develop. We parallel the development of these functions in [12], p. 457. Their work was done for simple varieties of trees whose size was determined by the number of vertices. Here we work with trees whose size is determined by the number of leaves.

For a rooted unordered tree t whose leaves are labeled by $B \subseteq \mathbb{N}$, let $\tilde{t} \in \mathcal{T}$ be the tree that results from relabeling the leaves of t by the unique increasing bijection from B to $\{1, 2, \dots, |B|\}$ (where $|B|$ is the cardinality of B). Suppose we have functions $\xi, \theta, \psi : \mathcal{T} \rightarrow \mathbb{R}$ satisfying the relation

$$\xi(t) = \theta(t) + \sum_{j=1}^{\deg(t)} \psi(\tilde{t}_j),$$

where $\deg(t)$ is the root degree of t and the $\{t_j\}$ are the root subtrees of t ordered in increasing order of the leaf with the smallest label. Letting $\bullet$ denote the tree on one leaf, we note that $\deg(\bullet) = 0$, so in particular $\xi(\bullet) = \theta(\bullet)$. Define the exponential generating functions

$$\Xi(z) = \sum_{t \in \mathcal{T}} \xi(t) w(t) \frac{z^{|t|}}{|t|!}, \quad \Theta(z) = \sum_{t \in \mathcal{T}} \theta(t) w(t) \frac{z^{|t|}}{|t|!}, \quad \text{and} \quad \Psi(z) = \sum_{t \in \mathcal{T}} \psi(t) w(t) \frac{z^{|t|}}{|t|!}.$$

Our results make use of the following lemma, which is a relation of formal power series.

Lemma 1. *We have the relation*

$$\Xi(z) = \Theta(z) + G'_\zeta(C_\zeta(z))\Psi(z). \tag{2.7}$$

In the purely recursive case where $\xi \equiv \psi$ we have

$$\Xi(z) = \frac{\Theta(z)}{1 - G'_\zeta(C_\zeta(z))} = C'_\zeta(z)\Theta(z). \tag{2.8}$$

Proof. We clearly have

$$\Xi(z) = \Theta(z) + \tilde{\Psi}(z), \quad \text{where} \quad \tilde{\Psi}(z) = \sum_{t \in \mathcal{T}} \left(w(t) \frac{z^{|t|}}{|t|!} \sum_{j=1}^{\deg(t)} \psi(\tilde{t}_j) \right).$$

Decomposing by root degree and using that $\xi(\bullet) = \theta(\bullet)$, we have

$$\begin{aligned}
 \tilde{\Psi}(z) &= \sum_{r \geq 1} \sum_{\deg(t)=r} \zeta_r \prod_{j=1}^r w(\tilde{t}_j) \frac{z^{|t_1|+\dots+|t_r|}}{(|t_1|+\dots+|t_r|)!} (\psi(\tilde{t}_1) + \dots + \psi(\tilde{t}_r)) \\
 &= \sum_{r \geq 1} \left[\frac{\zeta_r}{r!} \sum_{(t_1, \dots, t_r) \in \mathcal{T}^r} \prod_{j=1}^r w(t_j) \binom{|t_1|+\dots+|t_r|}{|t_1|, \dots, |t_r|} \frac{z^{|t_1|+\dots+|t_r|}}{(|t_1|+\dots+|t_r|)!} (\psi(t_1) + \dots + \psi(t_r)) \right] \\
 &= \sum_{r \geq 1} \left[\frac{\zeta_r}{r!} \sum_{(t_1, \dots, t_r) \in \mathcal{T}^r} \prod_{j=1}^r w(t_j) \frac{z^{|t_1|+\dots+|t_r|}}{|t_1|! \dots |t_r|!} (\psi(t_1) + \dots + \psi(t_r)) \right] \\
 &= \sum_{r \geq 1} \frac{\zeta_r}{(r-1)!} C_\zeta(z)^{r-1} \Psi(z) \\
 &= G'_\zeta(C_\zeta(z)) \Psi(z).
 \end{aligned}$$

This yields (2.7). In the recursive case, we have $\Xi(z) = \Theta(z) + G'_\zeta(C_\zeta(z))\Xi(z)$. Solving for $\Xi(z)$ gives the first equality in (2.8). To get the second, we differentiate (2.1) to get $C'_\zeta(z) = 1 + G'_\zeta(C_\zeta(z))C'_\zeta(z)$. Solving for $C'_\zeta(z)$ gives $C'_\zeta(z) = 1/(1 - G'_\zeta(C_\zeta(z)))$, from which the second equality in (2.8) is immediate. $\square$

Two immediate applications are to counting the weighted numbers of leaves and vertices of a given height.

Theorem 9. *The expected number of leaves at height k converges to $G''_\zeta(s)rk$ and the expected number of nodes at height k converges to $sG''_\zeta(s)k + 1$.*

Proof. Let $\xi_k(t)$ be the number of leaves of height k in t , so that $\xi_k(t)w(t)$ is the weighted number of leaves of height k . Define $\Xi_k = \sum_t \xi_k(t)w(t)z^{|t|}/|t|!$. For $k \geq 1$ we apply the lemma with $\xi = \xi_k$, $\theta = 0$ and $\psi = \xi_{k-1}$ to obtain

$$\Xi_k(z) = G'_\zeta(C_\zeta(z))\Xi_{k-1}(z),$$

which easily yields

$$\Xi_k(z) = [G'_\zeta(C_\zeta(z))]^k \Xi_0(z) = z [G'_\zeta(C_\zeta(z))]^k.$$

Letting $\Lambda_k(z)$ be the generating function for the weighted number of vertices of height k we similarly get

$$\Lambda_k(z) = [G'_\zeta(C_\zeta(z))]^k \Lambda_0(z) = C_\zeta(z) [G'_\zeta(C_\zeta(z))]^k.$$

Using these forms, we are able to compute asymptotics. Expanding G'_ζ about s , we have that $G'_\zeta(z) = 1 + G''_\zeta(s)(z - s) + O((z - s)^2)$. Plugging in the asymptotic expansion of C_ζ we get from Proposition 4 and Theorem 5 and doing some algebra, we have

$$G'_\zeta(C_\zeta(z)) = 1 - G''_\zeta(s)\gamma\sqrt{1 - z/r} + O(1 - z/r). \quad (2.9)$$

Hence, using that $(1 - z)^k = 1 - kz + O(z^2)$, we see that

$$[G'_\zeta(C_\zeta(z))]^k = 1 - G''_\zeta(s)k\gamma\sqrt{1 - z/r} + O(1 - z/r).$$

Thus, using Theorem 3, we have

$$[z^n]\Xi_k(z) = [z^n]z [G'_\zeta(C_\zeta(z))]^k \sim \frac{G''_\zeta(s)k\gamma}{2r^{n-1}\sqrt{\pi n^3}},$$

and, similarly with a bit more algebra,

$$[z^n]\Lambda_k(z) = [z^n]C_\zeta(z) [G'_\zeta(C_\zeta(z))]^k \sim \frac{\gamma(skG''_\zeta(s) + 1)}{2r^n\sqrt{\pi n^3}}.$$

Using the result on p. 474 of [12], that $[z^n]C_\zeta(z) \sim \gamma/2r^n\sqrt{\pi n^3}$, we find that

$$E_{\mathcal{T}_n}(\xi_k) = \frac{n![z^n]\Xi_k(z)}{n![z^n]C_\zeta(z)} \sim G''_\zeta(s)rk.$$

Letting $\zeta_k : \mathcal{T} \rightarrow \mathbb{Z}$ be the number of nodes of height k in t , we have that

$$E_{\mathcal{T}_n}(\zeta_k) = \frac{n![z^n]\Lambda_k(z)}{n![z^n]C_\zeta(z)} \sim sG''_\zeta(s)k + 1.$$

□

The proof of Theorem 8 is similar, but we make use of Theorem 6 for the asymptotics.

Proof of Theorem 8. Let $\{k_n\}$ be a sequence of integers varying such that $ck_n/n^{1/2} \rightarrow x \in (0, \infty)$ for some $c > 0$. By Theorem 6 and equation (2.9) we see that

$$[z^n][G'_\zeta(C_\zeta(z))]^{k_n} \sim r^{-n} \frac{G''_\zeta(s)\gamma}{2\sqrt{\pi n^3}} k_n \exp\left(-\frac{G''_\zeta(s)^2\gamma^2 k_n^2}{4n}\right). \quad (2.10)$$

Therefore

$$[z^n]\Xi_{k_n}(z) \sim r^{-(n-1)} \frac{G''_\zeta(s)\gamma}{2\sqrt{\pi n^3}} k_{n-1} \exp\left(-\frac{G''_\zeta(s)^2\gamma^2 k_{n-1}^2}{4(n-1)}\right).$$

This yields

$$\begin{aligned} E_{\mathcal{T}_n}(\xi_{k_n}) &\sim G''_\zeta(s)rk_{n-1} \exp\left(-\frac{G''_\zeta(s)^2\gamma^2 k_{n-1}^2}{4(n-1)}\right) \\ &= G''_\zeta(s)rk_{n-1} \exp\left(-\frac{G''_\zeta(s)rk_{n-1}^2}{2(n-1)}\right). \end{aligned}$$

Note that $P(H_n = k) = E_{\mathcal{T}_n}(\xi_k)/n$. Observe that $\{k_n\}$ satisfies the hypotheses of the above theorem. Consequently, we have that

$$\begin{aligned} \frac{\sqrt{n}}{c} P_n \left(\frac{c}{\sqrt{n}} H_n = \frac{c}{\sqrt{n}} k_n \right) &= \frac{\sqrt{n}}{c} P(H_n = k_n) \\ &= \frac{1}{c\sqrt{n}} E_{\mathcal{T}_n}(\xi_{k_n}) \\ &\sim \frac{1}{c^2} G''_{\zeta}(s) r \frac{ck_{n-1}}{\sqrt{n}} \exp \left(-\frac{G''_{\zeta}(s)r}{2c^2} \frac{c^2 k_{n-1}^2}{(n-1)} \right) \\ &\rightarrow \frac{G''_{\zeta}(s)r}{c^2} x \exp \left(-\frac{G''_{\zeta}(s)r}{2c^2} x^2 \right). \end{aligned}$$

The proof is finished by an application of a standard corollary of Scheffé's theorem (see Theorem 3.3 in [7] for an idea of the proof, just adapted for a distribution on $(0, \infty)$) and choosing $c = \sqrt{G''_{\zeta}(s)r}$. $\square$

In addition to proving convergence in distribution we can prove convergence of the first moment.

Theorem 10. $E_{\mathcal{T}_n} H_n \sim \sqrt{\frac{\pi}{2rG''_{\zeta}(s)}} n^{1/2}$.

The approach is to first compute the expected sum of the heights of the leaves of a tree.

Theorem 11. Let $\phi(t)$ be the sum of the heights of the leaves of t . Then $E_{\mathcal{T}_n} \phi \sim \sqrt{\frac{\pi}{2rG''_{\zeta}(s)}} n^{3/2}$.

Proof. Observe that

$$\phi(t) = |t| + \sum_{j=1}^{\deg(T)} \phi(\tilde{t}_j).$$

Let $\Phi(z)$ be the exponential generating function associated with ϕ . Applying Lemma 1, we have

$$\Phi(z) = z(C'_{\zeta}(z))^2.$$

By Theorem 4 we have

$$C'_{\zeta}(z) = \frac{\gamma}{2r} (1 - z/r)^{-1/2} + O(1).$$

Consequently,

$$(C'_{\zeta}(z))^2 = \frac{\gamma^2}{4r^2} (1 - z/r)^{-1} + ((1 - z/r)^{-1/2}).$$

Therefore

$$E_{\mathcal{T}_n} \phi = \frac{n![z^{n-1}](C'_{\zeta}(z))^2}{n![z^n]C_{\zeta}(z)} \sim \frac{\frac{\gamma^2 r^{-n+1}}{4r^2}}{\frac{\gamma}{2\sqrt{\pi n^3}} r^{-n}} = \frac{\gamma\sqrt{\pi}}{2r} n^{3/2} = \sqrt{\frac{\pi}{2rG''_{\zeta}(s)}} n^{3/2}.$$

$\square$

Proof of Theorem 10. Simply observe that $E_{\mathcal{T}_n} H_n = \frac{1}{n} E_{\mathcal{T}_n} \phi$. $\square$

The Number of Nodes and Node Degrees

We can also derive distributions for the out-degree of a random node. First we need the asymptotic number of nodes.

Theorem 12. *The expected number of nodes in T_n is asymptotic to sn/r and the expected number of internal nodes is asymptotic to $(s-r)n/r$.*

Proof. Letting $\xi(t)$ be the number of nodes in t we apply the recursive case of Lemma 1 with $\theta \equiv 1$ to get that $\Xi(z) = C'_\zeta(z)C_\zeta(z)$. From Theorem 4 we find that

$$C'_\zeta(z) = \frac{\gamma}{2r} \left(1 - \frac{z}{r}\right)^{-1/2} + O(1),$$

and consequently for any $k \geq 1$ (we will need the $k \neq 1$ case later)

$$C'_\zeta(z)C_\zeta(z)^k = \frac{\gamma s^k}{2r} \left(1 - \frac{z}{r}\right)^{-1/2} + O(1).$$

It now follows from Proposition 3 that

$$[z^n]C'_\zeta(z)C_\zeta(z)^k \sim \frac{\gamma s^k}{2r^{n+1}\sqrt{\pi n}}.$$

From this, we find that the expected number of nodes in a tree is

$$E_{\mathcal{T}_n}(\xi) = \frac{n![z^n]\Xi(z)}{n![z^n]C_\zeta(z)} \sim \frac{s}{r}n.$$

Since the number of leaves is always n , the expected number of internal nodes, $\nu(t)$ is then

$$E_{\mathcal{T}_n}(\nu) = E_{\mathcal{T}_n}(\xi) - n \sim \frac{s-r}{r}n.$$

$\square$

The above result is of interest in statistical classification theory (see e.g., [32, 9]), where the number of internal nodes is the number of classification stages.

We now turn to the out-degree of a random node. The first question is: What do we mean by a random node? An interpretation we would like to answer is, suppose that we choose a tree T_n according to P_n , and then choose a node uniformly from the nodes of T_n , what is the probability that the out-degree is k ? Unfortunately, current methods do not allow us to answer this question. Rather, we consider the following set up. For $t \in \mathcal{T}$, let $\mathcal{N}_t$ be the set of nodes of t and define

$$\mathcal{N}_n = \coprod_{t \in \mathcal{T}_n} \mathcal{N}_t.$$

We define the probability $P_n^{\mathcal{N}}$ on $\mathcal{N}_n$ by

$$P_n^{\mathcal{N}}(A, t) = \frac{w(t)}{n![z^n]\Xi(z)}.$$

That is, the probability of a node A in t is the weighted number of occurrences of A over the weighted number of nodes in trees of size n . We will find the out-degree distribution for nodes chosen from $\mathcal{N}_n$ according to $P_n^{\mathcal{N}}$.

Theorem 13. *Let D_n be the out-degree of a random node chosen from $\mathcal{N}_n$ according to $P_n^{\mathcal{N}}$. Then $P_n^{\mathcal{N}}(D_n = k | D_n > 0) \rightarrow \frac{\zeta_k s^k}{G_\zeta(s)k!}$.*

Proof. Let $\xi_k(t)$ be the number of nodes of t of out-degree k . Observe that

$$\xi_k(t) = \mathbf{1}(\deg(t) = k) + \sum_{j=1}^{\deg(t)} \xi_k(\tilde{t}_j).$$

Hence we are in the recursive case of Lemma 1. Since $\Theta(z) = C_\zeta^{[k]}(z)$, we have, for $k \geq 2$,

$$\Xi_k(z) = C'_\zeta(z)\Theta(z) = \frac{\zeta_k}{k!}C'_\zeta(z)C_\zeta(z)^k$$

and $\Xi_0(z) = zC'_\zeta(z)$. Consequently for $k \geq 2$

$$[z^n]\Xi_k(z) \sim \frac{\zeta_k \gamma s^k}{2r^{n+1}k!\sqrt{n\pi}},$$

and

$$[z^n]\Xi_0(z) \sim \frac{\gamma}{2r^n\sqrt{\pi n}}.$$

Therefore for $k \geq 2$,

$$P_n^{\mathcal{N}}(D_n = k) = \mathbf{1}(k \leq n) \frac{n![z^n]\Xi_k(z)}{n![z^n]\Xi(z)} \sim \frac{\zeta_k s^{k-1}}{k!},$$

and

$$P_n^{\mathcal{N}}(D_n = 0) = \frac{n![z^n]\Xi_0(z)}{n![z^n]\Xi(z)} \sim \frac{r}{s}.$$

Notice that

$$\frac{r}{s} + \sum_{k=2}^{\infty} \frac{\zeta_k s^{k-1}}{k!} = \frac{r}{s} + \frac{G_\zeta(s)}{s} = 1,$$

thus giving us convergence in distribution of the out-degree. Since internal nodes are exactly the nodes with out-degree greater than 0, we have that the probability that a randomly chosen internal node has degree $k \geq 2$ is

$$P_n^{\mathcal{N}}(D_n = k | D_n > 0) = \frac{P_n^{\mathcal{N}}(D_n = k)}{P_n^{\mathcal{N}}(D_n > 0)} \rightarrow \frac{\frac{\zeta_k s^{k-1}}{k!}}{1 - \frac{r}{s}} = \frac{\zeta_k s^k}{(s-r)k!} = \frac{\zeta_k s^k}{G_\zeta(s)k!}.$$

□

Chapter 3

Markov Branching Trees

3.1 Introduction

In this chapter, we introduce Markov branching trees and their scaling limits. We generalize the notion of Markov branching trees in [15] to allow for the construction of random trees with a given number of nodes whose out-degree are in a given set $A \subseteq \mathbb{Z}^+$ such that $0 \in A$. We then introduce the limiting objects, which are random compact metric measure spaces that are constructed as the genealogical trees of certain fragmentation trees. Finally, in Theorem 18 we prove scaling limits for Markov branching trees based on the fluctuations of the size of the subtrees attached to the root as the number of vertices with out-degree in A goes to infinity. Theorem 18 will be one of the cornerstones of our analysis of scaling limits of conditioned Galton-Watson trees in Chapter 4 and thus also, by the results of Chapter 2, for all of the types of trees considered in this thesis.

3.2 Markov branching trees

In this section we extend the notion of Markov branching trees developed in [15], where Markov branching trees were constructed separately in the cases $A = \{0\}$ and $A = \mathbb{Z}^+$. Here we give a construction for general A such that $0 \in A$. Let $\mathcal{T}^u$ be the set of rooted unordered trees considered up to root preserving isomorphism. If t is in $\mathcal{T}^{(o)}$ or $\mathcal{T}^u$ and $v \in t$ is a vertex, the out-degree of v is the number of vertices in t that are both adjacent to v and further from the root than v with respect to the graph metric. The out-degree of v will simply be denoted by $\deg(v)$, since we will only ever discuss out-degrees. Fix a set $A \subseteq \mathbb{Z}^+$ such that $0 \in A$ and for t in $\mathcal{T}^{(o)}$ or $\mathcal{T}^u$ define $\#_A t$ to be the number of vertices in t whose out-degree is in A . Furthermore, we define $\mathcal{T}_{A,n}^{(o)}$ and $\mathcal{T}_{A,n}^u$ by

$$\mathcal{T}_{A,n}^{(o)} = \{t \in \mathcal{T}^{(o)} : \#_A t = n\} \quad \text{and} \quad \mathcal{T}_{A,n}^u = \{t \in \mathcal{T}^u : \#_A t = n\}.$$

Let $\bar{\mathcal{P}}_n$ be the set of partitions of n and, for $\lambda \in \bar{\mathcal{P}}_n$, let $p(\lambda)$ be the number of nonzero blocks in λ and $m_j(\lambda)$ the number of blocks in λ equal to j . For convenience, we take

$\bar{\mathcal{P}}_1 = \{\emptyset, (1)\}$ and define $p(\emptyset) = -1$. Define $\bar{\mathcal{P}}_1^A = \bar{\mathcal{P}}_1$ and for $n \geq 2$, define $\bar{\mathcal{P}}_n^A$ by

$$\bar{\mathcal{P}}_n^A = \{\lambda \in \bar{\mathcal{P}}_n : p(\lambda) \notin A\} \cup \{\lambda \in \bar{\mathcal{P}}_{n-1} : p(\lambda) \in A\}.$$

Let (n_k) be an increasing sequence of integers. A sequence $(q_{n_k})_{k \geq 1}$, such that q_{n_k} is a probability measure on $\bar{\mathcal{P}}_{n_k}^A$, is called *compatible* if for each k , q_{n_k} is concentrated on partitions $\lambda = (\lambda_1, \dots, \lambda_p)$ such that q_{λ_i} is defined for all i . Our goal is to construct a sequence of laws $(\mathbf{P}_{n_k}^q)_{k=1}^\infty$ such that $\mathbf{P}_{n_k}^q$ is a law on $\mathcal{T}_{A, n_k}^u$ and such that the subtrees above a vertex are conditionally independent given the degree of that vertex. Consequently, we suppose further that q_1 is defined, $q_{n_k}((n_k)) < 1$ if $1 \notin A$ and $q_2((1)) = 1$ if $1 \in A$.

Remark 1. *Note that these assumptions put nontrivial restrictions on the compatible sequences we consider. For example, if $1 \notin A$ and $2 \in A$ then q_2 cannot exist because $\mathcal{P}_2^A = \{(2)\}$ and we are supposing that $q_2((2)) < 1$. In terms of the trees we are considering this is to be expected because if a tree has root degree 2 then it has at least 2 leaves and, as a result, cannot possibly have exactly 2 vertices with out-degree in A .*

Define $\mathbf{P}_1^q$ to be the law of the path with a root attached to a leaf by a path with G edges where $G = 0$ if $1 \in A$ and has the geometric distribution $\mathbb{P}(G = j) = q_1(\emptyset)(1 - q_1(\emptyset))^j$, $j \geq 0$ if $1 \notin A$. For $k \geq 2$, $\mathbf{P}_{n_k}^q$ is defined as follows: Choose $\Lambda \in \bar{\mathcal{P}}_{n_k}^A \setminus \{(n_k)\}$ according to $q_{n_k}(\cdot | \bar{\mathcal{P}}_{n_k}^A \setminus \{(n_k)\})$ and independently choose G' with $G' = 1$ if $1 \in A$ and G' has a geometric distribution

$$\mathbb{P}(G' = j) = (1 - q_{n_k}((n_k)))q_{n_k}((n_k))^{j-1}, j \geq 1,$$

if $1 \notin A$. Let $(T_1, T_2, \dots, T_{p(\Lambda)})$ be a vector of trees, independent of G' , such that the T_i are independent and T_i has distribution $\mathbf{P}_{\Lambda_i}^q$ and let T be the tree that results from attaching the roots of the T_i to the same new vertex and then if $G' = 1$ call this vertex the root, and otherwise attach that vertex to a new root by a path with $G' - 1$ edges. $\mathbf{P}_{n_k}^q$ is defined to be the law of T . An easy induction shows that $\mathbf{P}_{n_k}^q$ is concentrated on the set of unordered rooted trees with exactly n_k vertices whose out-degree is in A .

To connect with [15], if $(n_k) = (1, 2, 3, \dots)$, the case $A = \{0\}$ corresponds to the $\mathbf{P}_n^q$ defined in [15] and the case $A = \mathbb{Z}^+$ corresponds to the $\mathbf{Q}_n^q$ defined in [15]. Other choices of A interpolate between these two extremes. A sequence $(T_{n_k})_{k \geq 1}$ such that for each k , T_{n_k} has law $\mathbf{P}_{n_k}^q$ for some choice of A and q (independent of n) is called a Markov branching family. For ease of notation, we will generally drop the subscript k and it will be implicit that we are only considering n for which the quantities discussed are defined.

3.3 Trees as metric measure spaces

The trees we have been talking about can naturally be considered as metric spaces with the graph metric. That is, the distance between two vertices is the number of edges on the path connecting them. Let (t, d) be a tree equipped with the graph metric. For $a > 0$, we define at to be the metric space (t, ad) , i.e. the metric is scaled by a . This is equivalent to saying

the edges have length a rather than length 1 in the definition of the graph metric. More, generally we can attach a positive length to each edge in t and use these in the definition of the graph metric. Moreover, the trees we are dealing with are rooted so we consider (t, d) as a pointed metric space with the root as the point. Moreover, we are concerned with the vertices whose out-degree is in A , so we attach a measure $\mu_{\partial_A t}$, which is the uniform probability measure on $\partial_A t = \{v \in t : \deg(v) \in A\}$. If we have a random tree T , this gives rise to a random pointed metric measure space $(T, d, \text{root}, \mu_{\partial_A T})$. To make this last concept rigorous, we need to put a topology on pointed metric measure spaces. This is hard to do in general, but note that the pointed metric measure spaces that come from the trees we are discussing are compact.

We would like to consider the set of equivalence classes of compact pointed metric measure spaces (equivalence here being up to point and measure preserving isometry). Unfortunately, we cannot do this directly if we are working in Zermelo-Fraenkel set theory with the axiom of choice (which we are). To see the problem, note that if C is a set, $\{C\}$ can be thought of as a one point metric space and we have immediately run into set-of-all-sets type problems. What we can do, however, is note that there is a set $\mathcal{M}$ such that every compact metric space is isometric to exactly one element of $\mathcal{M}$.

Theorem 14. *There is a set $\mathcal{M}$ such that every compact metric space is isometric to exactly one element of $\mathcal{M}$.*

Proof. It is clearly sufficient to note that every compact metric space embeds isometrically into $\ell^\infty(\mathbb{N})$. This is the content of Fréchet's embedding theorem, whose proof is simple enough that we recall it here for completeness. Let X be a compact metric space. This implies X is separable, so we can let $\{x_i\}_{i \geq 1}$ be a countable dense subset of X . Define $\phi : X \rightarrow \ell^\infty(\mathbb{N})$ by $\phi(x) = \{d(x, x_i) - d(x_i, x_1)\}_{i \geq 1}$. It is trivial to verify that ϕ is isometric. $\square$

Consequently, there exists a set (which we denote by $\mathcal{M}_w$) of compact pointed metric measure spaces such that every compact pointed metric measure space is equivalent to exactly one element of $\mathcal{M}_w$. We metrize $\mathcal{M}_w$ with the pointed Gromov-Hausdorff-Prokhorov metric (see [15]). Fix $(X, d, \rho, \mu), (X', d', \rho, \mu') \in \mathcal{M}_w$ and define

$$d_{\text{GHP}}(X, X') = \inf_{(M, \delta)} \inf_{\substack{\phi: X \rightarrow M \\ \phi': X' \rightarrow M}} [\delta(\phi(\rho), \phi'(\rho')) \vee \delta_H(\phi(X), \phi'(X')) \vee \delta_P(\phi_*\mu, \phi'_*\mu')],$$

where the first infimum is over metric spaces (M, δ) , the second infimum if over isometric embeddings ϕ and ϕ' of X and X' into M , δ_H is the Hausdorff distance on compact subsets of M , and δ_P is the Prokhorov distance between the pushforward $\phi_*\mu$ of μ by ϕ and the pushforward $\phi'_*\mu'$ of μ' by ϕ' . Again, the definition of this metric has potential to run into set-theoretic difficulties, but they are not terribly difficult to resolve in a similar fashion to how we resolved the problems with $\mathcal{M}_w$.

Proposition 5 (Proposition 1 in [15]). *The space $(\mathcal{M}_w, d_{\text{GHP}})$ is a complete separable metric space.*

An $\mathbb{R}$ -tree is a complete metric space (T, d) with the following properties:

- For $v, w \in T$, there exists a unique isometry $\phi_{v,w} : [0, d(v, w)] \rightarrow T$ with $\phi_{v,w}(0) = v$ to $\phi_{v,w}(d(v, w)) = w$.
- For every continuous injective function $c : [0, 1] \rightarrow T$ such that $c(0) = v$ and $c(1) = w$, we have $c([0, 1]) = \phi_{v,w}([0, d(v, w)])$.

If (T, d) is a compact $\mathbb{R}$ -tree, every choice of root $\rho \in T$ and probability measure μ on T yields an element (T, d, ρ, μ) of $\mathcal{M}_w$. With this choice of root also comes a height function $\text{ht}(v) = d(v, \rho)$. The leaves of T can then be defined as a point $v \in T$ such that v is not in $[[\rho, w[:= \phi_{\rho,w}([0, \text{ht}(w))])$ for any $w \in T$. The set of leaves is denoted $\mathcal{L}(T)$.

Definition 6. A continuum tree is an $\mathbb{R}$ -tree (T, d, ρ, μ) with a choice of root and probability measure such that μ is non-atomic, $\mu(\mathcal{L}(T)) = 1$, and for every non-leaf vertex w , $\mu\{v \in T : [[\rho, v]] \cap [[\rho, w]] = [[\rho, w]]\} > 0$.

The last condition says that there is a positive mass of leaves above every non-leaf vertex. We will usually just refer to a continuum tree T , leaving the metric, root, and measure as implicit. A continuum random tree (CRT) is an $(\mathcal{M}_w, d_{GHP})$ valued random variable that is almost surely a continuum tree. The continuum random trees we will be interested in are those associated with self-similar fragmentation processes.

Self-similar fragmentations

For any set B , let $\mathcal{P}_B$ be the set of countable partitions of B , i.e. countable collections of disjoint sets whose union is B . For $n \in \bar{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$, let $\mathcal{P}_n := \mathcal{P}_{[n]}$. Suppose that $\pi = (\pi_1, \pi_2, \dots) \in \mathcal{P}_n$ (here and throughout we index the blocks of π in increasing order of their least elements), and $B \subseteq \bar{\mathbb{N}}$. Define the restriction of π to B , denoted by $\pi|_B$ or $\pi \cap B$, to be the partition of $[n] \cap B$ whose elements are the blocks $\pi_i \cap B$, $i \geq 1$. We topologize $\mathcal{P}_n$ by the metric

$$d(\pi, \sigma) := \frac{1}{\inf\{i : \pi \cap [i] \neq \sigma \cap [i]\}}.$$

It is worth noting that this is, in fact, an ultra-metric, i.e. for $\pi^1, \pi^2, \pi^3 \in \mathcal{P}_n$, we have

$$d(\pi^1, \pi^2) \leq \max(d(\pi^1, \pi^3), d(\pi^2, \pi^3)).$$

Note that $(\mathcal{P}_n, d)$ is compact for all n .

Definition 7 (Definition 3.1 in [6]). Consider two blocks $B \subseteq B' \subseteq \mathbb{N}$. Let π be a partition of B with $\#\pi = n$ non-empty blocks ($n = \infty$ is allowed), and $\pi^{(\cdot)} = \{\pi^{(i)}, i = 1, \dots, n\}$ be a sequence in $\mathcal{P}_{B'}$. For every integer i , we consider the partition of the i -th block π_i of π induced by the i -th term $\pi^{(i)}$ of the sequence $\pi^{(\cdot)}$, that is,

$$\pi|_{\pi_i}^{(i)} = \left(\pi_j^{(i)} \cap \pi_i \mid j \in \mathbb{N} \right).$$

As i varies in $[n]$, the collection $\left\{ \pi_j^{(i)} \cap \pi_i : i, j \in \mathbb{N} \right\}$ forms a partition of B , which we denote by $\text{Frag}(\pi, \pi^{(\cdot)})$ and call the fragmentation of π by $\pi^{(\cdot)}$.

Note that Frag is Lipschitz continuous in the first variable, and continuous in an appropriate sense in the second. Also, if π is an exchangeable partition and $\pi^{(\cdot)}$ is a sequence of independent exchangeable partitions (also independent of π), then π and $\text{Frag}(\pi, \pi^{(\cdot)})$ are jointly exchangeable. See chapter 3 of [6] for both of these facts. We will use the Frag function to define the transition kernels of our fragmentation processes.

Define

$$\mathcal{S}^\downarrow = \left\{ (s_1, s_2, \dots) : s_1 \geq s_2 \geq \dots \geq 0, \sum_{i=1}^{\infty} s_i \leq 1 \right\},$$

and

$$\mathcal{S}_1 = \left\{ (s_1, s_2, \dots) \in [0, 1]^{\mathbb{N}} \mid \sum_{i=1}^{\infty} s_i \leq 1 \right\},$$

and endow both with the topology they inherit as subsets of $[0, 1]^{\mathbb{N}}$ with the product topology. Observe that $\mathcal{S}^\downarrow$ and $\mathcal{S}_1$ are compact. For a partition $\pi \in \mathcal{P}_\infty$, we define the asymptotic frequency $|\pi_i|$ of the i 'th block by

$$|\pi_i| = \lim_{n \rightarrow \infty} \frac{|\pi_i \cap [n]|}{n},$$

provided this limit exists. If all of the blocks of π have asymptotic frequencies, we define $|\pi| \in \mathcal{S}_1$ by $|\pi| = (|\pi_1|, |\pi_2|, \dots)$.

Definition 8 (Definition 3.3 in [6]). *Let $\Pi(t)$ be an exchangeable, càdlàg $\mathcal{P}_\infty$ -valued process such that $\Pi(0) = \mathbf{1}_{\mathbb{N}} := (\mathbb{N}, 0, \dots)$ such that*

1. $\Pi(t)$ almost surely possesses asymptotic frequencies $|\Pi(t)|$ simultaneously for all $t \geq 0$ and
2. if we denote by $B_i(t)$ the block of $\Pi(t)$ which contains i , then the process $t \mapsto |B_i(t)|$ has right-continuous paths.

We call Π a self-similar fragmentation process with index $\alpha \in \mathbb{R}$ if and only if, for every $t, t' \geq 0$, the conditional distribution of $\Pi(t + t')$ given $\mathcal{F}_t$ is that of the law of $\text{Frag}(\pi, \pi^{(\cdot)})$, where $\pi = \Pi(t)$ and $\pi^{(\cdot)} = (\pi^{(i)}, i \in \mathbb{N})$ is a family of independent random partitions such that for $i \in \mathbb{N}$, $\pi^{(i)}$ has the same distribution as $\Pi(t' | \pi_i)^\alpha$.

One important tool for studying self-similar fragmentations is the equivalent of the Lévy-Itô decomposition of Lévy processes. Suppose, for the moment, that Π is a self-similar fragmentation process with $\alpha = 0$ (these are also called homogeneous fragmentations). In

this case, it turns out that Π is a Feller process as is $\Pi_{|[n]}$ for every n . Thus it is natural to try to identify the jump rates of these processes. For $\pi \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}$, let

$$q_\pi = \lim_{t \rightarrow 0^+} \frac{1}{t} \mathbb{P}(\Pi_{|[n]}(t) = \pi).$$

By exchangeability, it is obvious that $q_\pi = q_{\sigma(\pi)}$ for every permutation σ of $[n]$. Less obvious, but still true, is that the law of Π is determined by the jump rates $\{q_\pi : \pi \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}, n \in \mathbb{N}\}$. Furthermore, there is a nice description of these rates. For $\pi \in \mathcal{P}_n$ and $n' \in \{n, n+1, \dots, \infty\}$, define

$$\mathcal{P}_{n', \pi} = \{\pi' \in \mathcal{P}_{n'} : \pi'_{|[n]} = \pi\}.$$

Proposition 6 (Propositions 3.2 and 3.3 in [6]). *Suppose we have a family $\{q_\pi : \pi \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}, n \in \mathbb{N}\}$. This family is the family of jump rates of some homogeneous fragmentation Π if and only if there is an exchangeable measure μ on $\mathcal{P}_\infty$ satisfying*

1. $\mu(\{\mathbf{1}_\mathbb{N}\}) = 0$ and,
2. $\mu(\{\pi \in \mathcal{P}_\infty : \pi_{|[n]} \neq \mathbf{1}_{[n]}\}) < \infty$ for every $n \geq 2$,

such that $\mu(\mathcal{P}_{\infty, \pi}) = q_\pi$. Furthermore, this correspondence is bijective, and we call μ the splitting rate of Π .

For $n \in \mathbb{N}$, let $\varepsilon^{(n)}$ be the partition of $\mathbb{N}$ with exactly two blocks, $\{n\}$ and $\mathbb{N} \setminus \{n\}$ and define

$$\epsilon = \sum_{n=1}^{\infty} \delta_{\varepsilon^{(n)}}.$$

For a measure ν on $\mathcal{S}^\downarrow$ such that $\nu(\{\mathbf{1}\}) = 0$ and $\int_{\mathcal{S}^\downarrow} (1 - s_1) \nu(ds) < \infty$, define a measure ρ_ν on $\mathcal{P}_\infty$ by

$$\rho_\nu(\cdot) = \int_{\mathcal{S}^\downarrow} \rho_s(\cdot) \nu(ds).$$

Theorem 15 (Theorem 3.1 in [6]). *Let μ be the splitting rate of a homogeneous fragmentation. Then there exists a unique $c \geq 0$ and a unique measure ν on $\mathcal{S}^\downarrow$ with $\nu(\{\mathbf{1}\}) = 0$ and $\int_{\mathcal{S}^\downarrow} (1 - s_1) \nu(ds) < \infty$, such that*

$$\mu = c\epsilon + \rho_\nu.$$

The interpretation of this is that c is the erosion coefficient, i.e. the rate at which mass is lost continuously, and ν is the dislocation measure, i.e. it measures the rate of macroscopic fragmentation. From this theorem, it is clear that every homogeneous fragmentation process is characterized by the pair (c, ν) . Given a homogeneous fragmentation $\Pi^0(t)$ with parameters $(0, \nu)$, and $\alpha < 0$, we can construct an α -self-similar fragmentation with parameters $(\alpha, 0, \nu)$ by a time change. Let $\pi^i(t)$ be the block of Π^0 that contains i at time t and define

$$T_i(t) = \inf \left\{ u \geq 0 : \int_0^u |\pi^i(r)|^{-\alpha} dr > t \right\}.$$

For $t \geq 0$, let $\Pi(t)$ be the partition such that i, j are in the same block of $\Pi(t)$ if and only if they are in the same block of $\Pi^0(T_i(t))$. Then $(\Pi(t), t \geq 0)$ is a self-similar fragmentation with characteristics $(\alpha, 0, \nu)$. See [5] for details.

We will need trees associated to fragmentations with characteristics $(\alpha, 0, \nu)$, where $\alpha < 0$ and $\nu(\sum_i s_i < 1) = 0$. What these assumptions tell us is that there is no continuous loss of mass due to erosion ($c = 0$), mass is not lost during macroscopic fragmentations ($\nu(\sum_i s_i < 1) = 0$), and the fragmentation eventually becomes the partition into singletons (Proposition 2 in [5], the rate of convergence is give in Proposition 14 in [13]). Henceforth, we let Π be such a self-similar fragmentation.

The tree associated with a fragmentation processes Π is a continuum random tree that keeps track of when blocks split apart and the sizes of the resulting blocks. For a continuum tree (T, μ) and $t \geq 0$, let $T_1(t), T_2(t), \dots$ be the tree components of $\{v \in T : \text{ht}(v) > t\}$, ranked in decreasing order of μ -mass. We call (T, μ) self-similar with index $\alpha < 0$ if for every $t \geq 0$, conditionally on $(\mu(T_i(t)), i \geq 1)$, $(T_i(t), i \geq 1)$ has the same law as $(\mu(T_i(t))^{-\alpha} T^{(i)}, i \geq 1)$ where the $T^{(i)}$'s are independent copies of T .

The following summarizes the parts of Theorem 1 and Lemma 5 in [14] that we will need.

Theorem 16. *Let Π be a $(\alpha, 0, \nu)$ -self-similar fragmentation with $\alpha < 0$ and ν as above and let $F := |\Pi|^\downarrow$ be its ranked sequence of asymptotic frequencies. There exists an α -self-similar CRT $(T_{-\alpha, \nu}, \mu_{-\alpha, \nu})$ such that, writing $F'(t)$ for the decreasing sequence of masses of the connected components of $\{v \in T_{-\alpha, \nu} : \text{ht}(v) > t\}$, the process $(F'(t), t \geq 0)$ has the same law as F . Furthermore, T_F is a.s. compact.*

The choice of where to put negative signs in the notation in the above theorem is to conform with the notation of [15].

Definition 9. *The Brownian CRT is the $-1/2$ -self-similar random tree with dislocation measure ν_2 given by*

$$\int_{\mathcal{S}^\downarrow} \nu_2(ds) f(s) = \int_{1/2}^1 \sqrt{\frac{2}{\pi s_1^3 (1-s_1)^3}} ds_1 f(s_1, 1-s_1, 0, 0, \dots).$$

It is denoted by $T_{1/2, \nu_2}$ or T^{Br} depending on whether we want to emphasize the connection to fragmentation processes or Brownian motion.

Since we will always have $c = 0$, we will drop it and for a measure ν satisfying the above conditions and $\gamma > 0$, we refer to $(-\gamma, \nu)$ as fragmentation pair, which is associated to a $(-\gamma, \nu)$ -self-similar fragmentation.

3.4 Convergence of Markov branching trees

We first recall some of the main results of [15]. Let $A \subseteq \mathbb{Z}^+$ contain 0 and let (q_n) be a compatible sequence of probability measures satisfying the conditions of Section 3.2. Define $\bar{q}_n$ to be the push forward of q_n onto $\mathcal{S}^\downarrow$ by $\lambda \mapsto \lambda / \sum_i \lambda_i$.

Theorem 17 (Theorems 1 and 2 in [15]). *Suppose that $A = \{0\}$ or $A = \mathbb{Z}^+$. Further suppose that there is a fragmentation pair $(-\gamma, \nu)$ with $0 < \gamma < 1$ and a function $\ell : (0, \infty) \rightarrow (0, \infty)$, slowly varying at ∞ (or $\gamma = 1$ and $\ell(n) \rightarrow 0$) such that, in the sense of weak convergence of finite measures on $\mathcal{S}^\downarrow$, we have*

$$n^\gamma \ell(n) (1 - s_1) \bar{q}_n(ds) \rightarrow (1 - s_1) \nu(ds).$$

Let T_n have law $\mathbf{P}_n^q$ and view T_n as a random element of $\mathcal{M}_w$ with the graph distance and the uniform probability measure $\mu_{\partial_A T_n}$ on $\partial_A T_n = \{v \in T_n : \deg v \in A\}$. Then we have the convergence in distribution

$$\frac{1}{n^\gamma \ell(n)} T_n \rightarrow T_{\gamma, \nu},$$

with respect to the rooted Gromov-Hausdorff-Prokhorov topology.

The case where $A = \{0\}$ this is a special case of Theorem 1 in [15] and the case $A = \mathbb{Z}^+$ is Theorem 2 in the same paper. The case $A = \mathbb{Z}^+$ is proved by reduction to the $A = \{0\}$ case. We extend this to the case of general A containing 0.

Theorem 18. *The conclusions of Theorem 17 are valid if the only assumption on $A \subseteq \mathbb{Z}^+$ is that $0 \in A$.*

As argued at the start of Section 4 in [15], we may assume that $q_1(\emptyset) = 1$. This is because each leaf is connected to the rest of the tree by a stalk of vertices with out-degree one and geometric length. Setting $q_1(\emptyset) = 1$ collapses these to be length one. Since these stalks are independent from one another, with probability approaching one, this costs $\log(n)$ in the Gromov-Hausdorff-Prokhorov metric. This is negligible since we are scaling by $(n^\gamma \ell(n))^{-1}$. Let t be a rooted unordered tree with n vertices whose out-degree is in A and let t° be the tree obtained from t by attaching a leaf to every non-leaf vertex of t whose out-degree is in A .

Define the inclusion $\iota : \bar{\mathcal{P}}_{n-1} \rightarrow \bar{\mathcal{P}}_n$ by $\iota(\lambda) = (\lambda, 1)$. We now define a sequence q_n° of probability measures on $\bar{\mathcal{P}}_n$. Define $q_1^\circ(\emptyset) = 1$ and for $n \geq 2$,

$$q_n^\circ(\lambda) = \begin{cases} q_n(\lambda) & \text{if } \lambda \in \bar{\mathcal{P}}_n^A \setminus \iota(\bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}), \\ q_n(\lambda) + q_n(\lambda') & \text{if } \lambda \in \bar{\mathcal{P}}_n^A \text{ and } \lambda = \iota(\lambda') \text{ for some } \lambda' \in \bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}, \\ q_n(\lambda') & \text{if } \lambda \notin \bar{\mathcal{P}}_n^A \text{ and } \lambda = \iota(\lambda') \text{ for some } \lambda' \in \bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 2. *If T_n has distribution $\mathbf{P}_n^q$ then T_n° has distribution $\mathbf{P}_n^{q^\circ}$.*

Proof. We prove this by induction. The result is clear for $n = 1$. For $n \geq 2$, we condition on the root partition. Indeed, since in both T_n° and a tree with law $\mathbf{P}_n^{q^\circ}$ the subtrees attached to the root are independent given the root partition, by induction (and a little care about when the partition at the root is (n)), we need only check that the laws of the partitions at the root agree. This, however, is immediate from the construction of q_n° . $\square$

Therefore Theorem 18 is an immediate consequence of the following lemma.

Lemma 3. *If*

$$n^\gamma \ell(n)(1 - s_1) \bar{q}_n(d\mathbf{s}) \rightarrow (1 - s_1) \nu(d\mathbf{s}),$$

then

$$n^\gamma \ell(n)(1 - s_1) \bar{q}_n^\circ(d\mathbf{s}) \rightarrow (1 - s_1) \nu(d\mathbf{s}).$$

Proof. Let $f : \mathcal{S}^\downarrow \rightarrow \mathbb{R}$ be Lipschitz continuous (with respect to the uniform norm) with both the uniform norm and Lipschitz constant bounded by K . Observe that for $\lambda \in \bar{\mathcal{P}}_n$,

$$\left| f\left(\frac{\iota(\lambda)}{n+1}\right) - f\left(\frac{\lambda}{n}\right) \right| \leq K \sum_{i=1}^{p(\lambda)} \frac{\lambda_i}{n(n+1)} + \frac{K}{n+1} = \frac{2K}{n+1}.$$

Letting $g(\mathbf{s}) = (1 - s_1)f(\mathbf{s})$, we have

$$\begin{aligned} |\bar{q}_n^\circ(g) - \bar{q}_n(g)| &\leq \sum_{\lambda \in \bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}} q_n(\lambda) \left| \left(1 - \frac{\lambda_1}{n}\right) f\left(\frac{\iota(\lambda)}{n}\right) - \left(1 - \frac{\lambda_1}{n-1}\right) f\left(\frac{\lambda}{n-1}\right) \right| \\ &\leq \sum_{\lambda \in \bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}} q_n(\lambda) \left(\frac{K\lambda_1}{n(n-1)} + \frac{2K}{n} \right) \\ &\leq \frac{3K}{n}. \end{aligned}$$

Multiplying by $n^\gamma \ell(n)$, we see that this upper bound goes to 0 and the result follows. $\square$

Proof of Theorem 18. Note that, if $a > 0$, then $d_{\text{GHP}}(at, at^\circ) \leq a$. Consequently

$$d_{\text{GHP}}\left(\frac{1}{n^\gamma \ell(n)} T_n, \frac{1}{n^\gamma \ell(n)} T_n^\circ\right) \leq \frac{1}{n^\gamma \ell(n)} \rightarrow 0.$$

Since $(n^\gamma \ell(n))^{-1} T_n^\circ \rightarrow T_{\gamma, \nu}$ by Lemma 3 and Theorem 17, $(n^\gamma \ell(n))^{-1} T_n \rightarrow T_{\gamma, \nu}$ as well. $\square$

Chapter 4

Scaling limits of Galton-Watson trees

4.1 Introduction

As a consequence of Theorem 18 we obtain a new theorem for scaling limits of Galton-Watson trees. In the literature on scaling limits of random trees, much interest has been focused on limits of Galton-Watson trees conditioned on their total number of vertices. Furthermore, the techniques employed generally rely heavily on the fact that the conditioning is on the number of leaves. Using the framework of Markov branching trees, we are able to modify the conditioning to condition on the number of vertices with out-degree in a given set. In particular, we prove the following theorem.

Theorem 19. *Let T be a critical Galton-Watson tree with offspring distribution ξ such that $0 < \sigma^2 = \text{Var}(\xi) < \infty$ and let $A \subseteq \{0, 1, 2, \dots\}$ contain 0. Suppose that for sufficiently large n the probability that T has exactly n vertices with out-degree in A is positive, and for such n let T_n^A be T conditioned to have exactly n vertices with out-degree in A , considered as a rooted unordered tree with edge lengths 1 and the uniform probability distribution $\mu_{\partial_A T_n^A}$ on its vertices with out-degree in A . Then*

$$\frac{1}{\sqrt{n}} T_n^A \xrightarrow{d} \frac{2}{\sigma \sqrt{\xi(A)}} T_{1/2, \nu_2},$$

where the convergence is with respect to the rooted Gromov-Hausdorff-Prokhorov topology and $T_{1/2, \nu_2}$ is the Brownian continuum random tree.

In the case $A = \mathbb{Z}^+ = \{0, 1, 2, \dots\}$ we recover the classical result about the scaling limit of a Galton-Watson tree conditioned on its number of vertices first obtained in [3]. For other choices of A the result appears to be new. We note, however, that subsequent to the appearance of [29] on the arXiv, Kortchemski obtained similar results by different techniques in [18]. The condition that for sufficiently large n the probability that T has exactly n vertices with out-degree in A is positive is purely technical and could be dispensed with at the cost of chasing periodicity considerations through our computations. In addition

to generalizing the results of [15], the key to proving this theorem is a generalization of the classical Otter-Dwass formula, which we prove in Section 4.2. The Otter-Dwass formula (see [26]) has been an essential tool in several proofs that the Brownian continuum random tree is the scaling limit of Galton-Watson trees conditioned on their number of vertices, including the original proof in [3] as well as newer proofs in [19] and [15]. While we follow the approach in [15], our generalization of Otter-Dwass formula should allow for proofs along the lines of [3] and [19] as well. Furthermore, with our results here, it should be straightforward to prove the analogous theorem in the infinite variance case using the approach in [15].

We begin by proving our generalization of the Otter-Dwass formula and we then use this to analyze the asymptotics of the partition at the root of a Galton-Watson tree. Bringing this together with the results of Chapter 3, we give a proof of Theorem 19. We finish by incorporating the results of Chapter 2 to prove Theorem 23, which gives the precise statements for the scaling limits appearing in Schröder's problems.

4.2 Galton-Watson trees

We first recall the definition of a Galton-Watson tree. Let $\xi = (\xi_i)_{i \geq 0}$ be a probability distribution with mean less than or equal to 1, and assume that $\xi_1 < 1$. A Galton-Watson tree with offspring distribution ξ is a random element T of $\mathcal{T}^{(o)}$ with law

$$\text{GW}_\xi(t) = \mathbb{P}(T = t) = \prod_{v \in t} \xi_{\deg(v)}.$$

The fact that ξ has mean less than or equal to 1 implies that the right hand side defines an honest probability distribution on $\mathcal{T}^{(o)}$.

Otter-Dwass type formulae

In this section we develop a transformation of rooted ordered trees that takes Galton-Watson trees to Galton-Watson trees. This transformation is motivated by the observation that the number of leaves in a Galton-Watson tree is distributed like the progeny of a Galton-Watson tree with a related offspring distribution. This simple observation was first made in [23]. Let $\xi = (\xi_i)_{i \geq 0}$ be a probability distribution with mean less than or equal to 1, and assume that $\xi_1 < 1$. Let T be a Galton-Watson tree with offspring distribution ξ and let

$$C(z) = \sum_{i=1}^{\infty} \mathbb{P}(\#\{0\}T = i) z^i$$

be the probability generating function of the number of leaves of T . Furthermore, let

$$\phi(z) = \sum_{i=0}^{\infty} \xi_{i+1} z^i.$$

Decomposing by the root degree, we see that $C(z)$ satisfies the functional equation

$$C(z) = \xi_0 z + C(z)\phi(C(z)).$$

Note this is just a slightly modified form of the functional equation obtained in Theorem 1. Solving for $C(z)$ yields

$$C(z) = z \left(\frac{\xi_0}{1 - \phi(C(z))} \right). \quad (4.1)$$

Define

$$\theta(z) = \frac{\xi_0}{1 - \phi(z)}. \quad (4.2)$$

Observe that θ has nonnegative coefficients, $[z^0]\theta(z) = \xi_0/(1 - \xi_1)$ and $\theta(1) = 1$. Thus the coefficients of θ are a probability distribution, call it $\zeta = (\zeta_i)_{i \geq 0}$.

Proposition 7. *Let T be a Galton-Watson tree with offspring distribution ξ and let T' be a Galton-Watson tree with offspring distribution ζ where ξ and ζ are related as above. Then for all $k \geq 1$, $\mathbb{P}(\#_{\{0\}}T = k) = \mathbb{P}(\#_{\mathbb{Z}^+}T' = k)$. Also, T' is critical (subcritical) if and only if T is critical (subcritical).*

Proof. The computations above show that the probability generating functions for $\#_{\{0\}}T$ and $\#_{\mathbb{Z}^+}T'$ satisfy the same functional equation and the Lagrange inversion formula implies they have the same coefficients. The criticality claims follows from Equation (4.2), which can also be used to obtain higher moments of ζ . $\square$

Corollary 3. *Let $\mathcal{F}_n$ be an ordered forest of n independent Galton-Watson trees all with offspring distribution ξ . Let ζ be related to ξ as in Proposition 7. Let $(X_i)_{i \geq 1}$ be an i.i.d. sequence of ζ distributed random variables and let $S_k = \sum_{i=1}^k (X_i - 1)$. Let $\#_{\{0\}}\mathcal{F}_n$ denote the number of leaves in $\mathcal{F}_n$. Then for $1 \leq k \leq n$*

$$\mathbb{P}(\#_{\{0\}}\mathcal{F}_k = n) = \frac{k}{n} \mathbb{P}(S_n = -k).$$

Proof. This follows immediately from Proposition 7 and the Otter-Dwass formula (see [26]). $\square$

This relationship between T and T' can also be proved in a more probabilistic fashion. Indeed, by taking a more probabilistic approach we can get a more general result that includes the results in [23] as a special case. To prove the result in full generality, it is more convenient to work with the depth-first queue of T than with T itself.

For $x \in \mathbb{Z}^{\mathbb{N}} = \mathbb{Z}^{\{1,2,3,\dots\}}$, let $\tau_{-1} = \inf \{n : \sum_{i=1}^n x_i = -1\}$. Let $\mathcal{D}$ be the set of sequences of increments first-passage bridges from 0 to -1 in $\mathbb{Z}^{\mathbb{N}}$ that are bounded below by -1 . Formally,

$$\mathcal{D} = \{x \in \mathbb{Z}^{\mathbb{N}} : x_i \geq -1 \text{ for } i \geq 1, x_i = 0 \text{ for } i \geq \tau_{-1}(x), \tau_{-1}(x) < \infty\}.$$

For $t \in \mathcal{T}^{(o)}$ with n vertices, index the vertices V of t from 1 to n by order of appearance on the depth-first walk of t . Define $DQ(t) = (DQ_k(t))_{k=1}^\infty$ by $DQ_k(t) = \deg v_k - 1$ for $k \leq n$ and 0 for $k > n$, which are the increments of the depth-first queue of t . Note that $DQ(t) \in \mathcal{D}$. Furthermore $t \mapsto DQ(t)$ is a bijection from $\mathcal{T}^{(o)}$ to $\mathcal{D}$ (see e.g. [26]).

Let π_n be the projection onto the n 'th coordinate of $\mathbb{Z}^\mathbb{N}$ and let $\mathcal{F}_n = \sigma(\pi_k, k \leq n)$. Let θ_n be the shift $(\theta_n x)(i) = x(n+i)$. Let N' be a stopping time with respect to $(\mathcal{F}_n)$. Let $N^0 = 0$ and for $i \geq 1$ define $N^i = N^{i-1} + (N' \wedge \tau_{-1}) \circ \theta_{N^{i-1}}$. Define $\hat{x}$ by

$$\hat{x}(k) = \begin{cases} \sum_{i=N^{k-1}+1}^{N^k} x(i) & \text{if } N^k < \infty \\ 0 & \text{if } N^k = \infty. \end{cases}$$

Proposition 8. *If $x \in \mathcal{D}$, then $\hat{x} \in \mathcal{D}$.*

Proof. The only non-trivial part is to see that for each $x \in \mathcal{D}$ there exists k such that $N^k = \tau_{-1}$. Clearly $N^1 \leq \tau_{-1}$. Let $k = \max\{i : N^i \leq \tau_{-1}\}$. Suppose, for the sake of contradiction, that $N^k < \tau_{-1}$. We then have that $\sum_{i=1}^{N^k} x(i) \geq 0$, so $\sum_{i=N^{k+1}}^{\tau_{-1}} x(i) \leq -1$, so $N^{k+1} \leq \tau_{-1}$, which is our contradiction. $\square$

Combining these ideas, we obtain the following theorem.

Theorem 20. *Let ξ be a probability distribution on $\mathbb{Z}_+$ with $0 < \xi_0 < 1$. Suppose that T is a Galton-Watson tree with offspring distribution ξ . Let N' be a stopping time and let $\hat{T}$ be the tree determined by $\widehat{DQ}(T)$ by the bijection above. Let $X = (X_1, X_2, \dots)$ be a vector with i.i.d. entries distributed like $P(X_1 = k) = \xi_{k+1}$. Then $\hat{T}$ is a Galton-Watson tree whose offspring distribution is the law of $1 + \sum_{i=1}^{N^1(X)} X_i$. Furthermore, if $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}(N' \wedge \tau_{-1}) < \infty$ then*

$$\mathbb{E} \left(1 + \sum_{i=1}^{N^1(X)} X_i \right) = 1 + \mathbb{E}X_1 \mathbb{E}(N' \wedge \tau_{-1}),$$

and if, additionally, $\mathbb{E}X_1 = 0$ (i.e. T is critical) and $\text{Var}(X_1) = \sigma^2 < \infty$, then

$$\text{Var} \left(1 + \sum_{i=1}^{N^1(X)} X_i \right) = \sigma^2 \mathbb{E}(N' \wedge \tau_{-1}).$$

Proof. Define $R(X)$ to be the vector with $R(X)_k = X_k \mathbf{1}(k \leq \tau_{-1}(X))$. It is well known that $DQ(T) =_d R(X)$ and that the vectors $\{(X_{N^i+1}, \dots, X_{N^{i+1}})\}_{i=0}^\infty$ are i.i.d. Consequently $\hat{X}$ is the vector of increments of a random walk with jump distribution given by the law of $\sum_{i=1}^{N^1(X)} X_i$. Observing that $\widehat{DQ}(T) =_d \widehat{R(X)} = R(\hat{X})$ shows that $\hat{T}$ is a Galton-Watson tree with the appropriate offspring distribution. The last claims follow from Wald's equations. $\square$

Let us give a specific example of how the general theorem above may be applied. For a nonempty subset A of $\mathbb{Z}^+$ let $\mathcal{D}_n^A$ be the set of paths in $\mathcal{D}$ with exactly n terms no later than

$\tau_{-1}(x)$ in $A - 1$. When $A = \mathbb{Z}^+$, we just write $\mathcal{D}_n$. Note that for every $n \geq 1$, $t \mapsto DQ(t)$ restricts to a bijection from $\mathcal{T}_{A,n}^{(o)}$ to $\mathcal{D}_n^A$. We obtain the following plethora of Otter-Dwass type formulae.

Corollary 4. *Fix $A \subseteq \mathbb{Z}^+$ such that $0 \in A$ and define $N'(x) = \inf\{i : x_i + 1 \in A\}$. Define T , $\hat{T}$, and X as in Theorem 20 and let $\hat{X}_1, \hat{X}_2, \dots$ be i.i.d. distributed like $1 + \sum_{i=1}^{N^1(X)} X_i$. Then*

$$\mathbb{P}(\#_A T = n) = \mathbb{P}(\hat{T} = n) = \frac{1}{n} \mathbb{P}\left(\sum_{i=1}^n \hat{X}_i = -1\right).$$

The corresponding result for forests also holds. Furthermore, if T is critical with variance $0 < \sigma^2 < \infty$, then $\text{Var}(\hat{X}_i) = \sigma^2/\xi(A)$.

Proof. This follows from the observation that, with this N' , $x \in \mathcal{D}_n^A$ if and only if $\hat{x} \in \mathcal{D}_n$. The formula for the variance follows from the fact that N' is geometric with parameter $\xi(A)$. $\square$

We note that, in the context of Corollary 4, the same construction can be done directly on the trees without first passing to the depth-first queue, though setting up the formalism for the proof and the proof itself are slightly more involved. The idea is a lifeline construction. You proceed around the tree in the order of the depth-first walk and when you encounter a vertex whose degree is in A you label the edges and vertices on the path from the vertex to the root that are not yet labeled by that vertex. This labeled path can be considered the lifeline of the vertex. A new tree is constructed by letting the root be the first vertex encountered whose degree is in A and attaching vertices whose degree is in A to the earliest vertex whose lifeline touches its own. Going through the details of this helps make this transformation more concrete, so the case of $A = \{0\}$ is included below.

Suppose that $t \in \mathcal{T}_{\{0\},n}^{(o)}$ and label the leaves by the order they appear in the depth-first walk of t . We will now color all of the edges of t . Color every edge on the path from leaf 1 to the root with 1. Continuing in increasing order of their labels, color all edges on the path from leaf i to the root that are not colored with an element of $\{1, 2, \dots, i-1\}$ with i , until all edges are colored. Note that for any $1 \leq k \leq n$ the subtree spanned by leaves $\{1, \dots, k\}$ is colored by $\{1, \dots, k\}$ and an edge is colored by an element in $\{1, \dots, k\}$ if and only if it is in this subtree. Furthermore, the path from leaf k to any edge colored k contains only edges colored k . See Figure 4.1 for an example of such a coloring. Call two edges of t coincident if they share a common vertex.

Lemma 4. *If t is colored as above and $2 \leq j \leq n$, then there is exactly one edge colored j that is coincident to an edge with a smaller color.*

Proof. First we show existence. Consider the path from leaf j to the root. Let e be the last edge in this path that is not contained in the subtree spanned by leaves $\{1, \dots, j-1\}$. By construction this edge is colored j and is coincident to an edge colored by an element of $\{1, \dots, j-1\}$.

To see uniqueness, suppose that f is an edge with the desired properties. Then f is on the path from j to the root and f is coincident to an edge in the subtree spanned by leaves $\{1, \dots, j-1\}$. If f contains the root, then f is the last edge on the path from j to the root that is colored j , i.e. $f = e$. Otherwise, after f , we finish the path from j to the root within this subtree. Hence f is the last edge on the path from j to the root that is colored j and again $f = e$. $\square$

With t labeled as above we define a rooted plane tree with n vertices, called the life-line tree and denoted $\check{t}$, as follows. The vertex set of $\check{t}$ is $\{1, 2, \dots, n\}$, 1 is the root. Furthermore, if $i < j$, i is adjacent to j if i is the smallest number such that there exist coincident edges e_1, e_2 in t with e_1 colored i and e_2 colored j . Finally, the children of a vertex are ordered by the appearance of the corresponding leaves in the depth-first search of t . See Figure 4.1 for an example of this map.

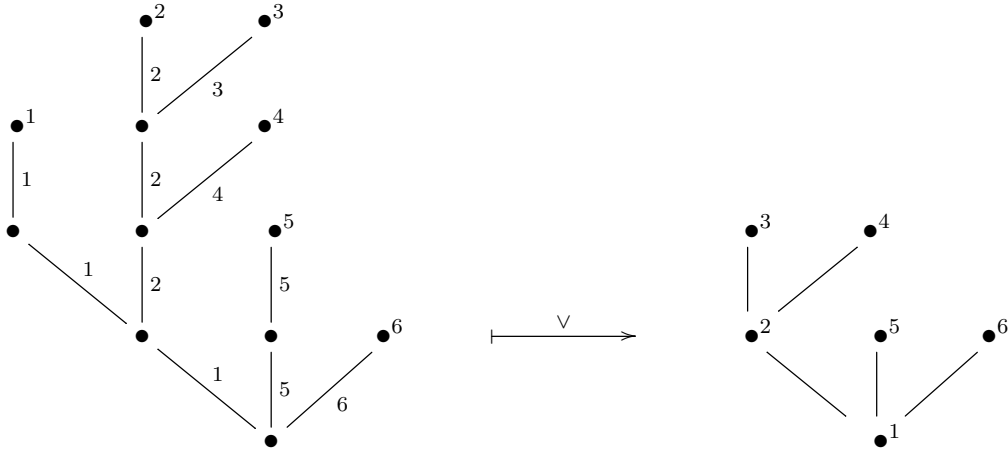


Figure 4.1: A colored tree and its image under V

Lemma 5. *The life-line tree is a tree.*

Proof. We must show that $\check{t}$ is connected and acyclic. Suppose that $\check{t}$ has at least two components. Let j be the smallest vertex not in the same component as 1. By Lemma 4, there exists $1 \leq i < j$ and coincident edges e_1, e_2 in t labeled i and j respectively. Thus i is adjacent to j , a contradiction.

Suppose that $\check{t}$ contains a cycle. Let j be the largest vertex in this cycle. Then j is adjacent to two smaller vertices, contradicting our definition of $\check{t}$. $\square$

Let $(v_1, \dots, v_k)$ be the list of vertices (ordered by order of appearance in the depth-first walk of t) in t that are children of vertices on the path from 1 to the root, but not actually on that path themselves. Let t_{v_i} be the plane subtree of t above v_i . It is then easily verified that $\check{t}$ is obtained by joining the trees $(\check{t}_{v_1}, \dots, \check{t}_{v_k})$ to a common root with their natural order (and renaming the vertices as they appear in the depth first walk).

Lemma 6. *Let T be a Galton-Watson tree with offspring distribution ξ . Condition T on the event that the first leaf on the depth-first walk of T has height n and that there are exactly k vertices in T that are children of vertices on the path from the root to the first leaf on the depth-first walk that are not on this path themselves. Let $v_1^n, \dots, v_k^n$ be these vertices (again in order of appearance) and let $T_{v_j^n}$ be the plane subtree of T above v_j^n . The collection $\{T_{v_j^n}\}_{j=1}^k$ is a collection of i.i.d Galton-Watson trees with common distribution T .*

Proof. Let $t_1, \dots, t_k$ be rooted ordered trees. Let U be the set of trees t such that the first leaf on the depth-first walk of t has height n and that there are exactly k vertices $v_1, \dots, v_k$ (listed in order of appearance on the depth-first walk) in t that are children of vertices on the path from the root to the first leaf on the depth-first walk that are not on this path themselves. Let $V \subseteq U$ be the set of trees such that t_j is the tree above v_j for all $1 \leq j \leq k$. Let W be the set sequences that appear as the sequence of degrees of vertices on the path from the root to the left-most leaf of a tree in U . Note that

$$\mathbb{P}(T \in U) = \sum_{(y_1, \dots, y_{n+1}) \in W} \prod_{i=1}^{n+1} \xi_{y_i}$$

since, given $(y_1, \dots, y_{n+1}) \in W$, there are k places to attach trees to the path from the root to the left-most leaf, and we sum over all ways of doing this. Consequently, we have

$$\begin{aligned} \mathbb{P}(T_{v_j^n} = t_j, \ 1 \leq j \leq k) &= \frac{1}{\mathbb{P}(U)} \sum_{t \in V} \mathbb{P}(T = t) \\ &= \frac{1}{\mathbb{P}(U)} \sum_{(y_1, \dots, y_{n+1}) \in W} \prod_{i=1}^{n+1} \xi_{y_i} \prod_{i=1}^k \mathbb{P}(T = t_i) \\ &= \prod_{i=1}^k \mathbb{P}(T = t_i). \end{aligned}$$

□

Theorem 21. *Let T be a Galton-Watson tree with offspring distribution ξ and define $\tilde{T}$ and T' as above (see Proposition 7 for T'). Then $\tilde{T} \stackrel{d}{=} T'$.*

Proof. Let t be a rooted plane tree and consider $\mathbb{P}(\tilde{T} = t)$. If t has one vertex, it is clear that $\mathbb{P}(\tilde{T} = t) = \mathbb{P}(T' = t)$. Suppose that the result is true for all trees with less than n vertices, and suppose that t has n vertices. Let $t_1, \dots, t_k$ be the subtrees of t attached to the root of t , listed in order of appearance of the depth-first walk of t . Let $A_{i,k}$ be the event that the first leaf to appear on the depth-first walk of T has height i and that there are k vertices in T that are children of vertices on the path from the root to the first leaf on the depth-first walk that are not on this path themselves. Let $v_1^i, \dots, v_k^i$ be these vertices (again in order of appearance) and let $T_{v_j^i}$ be the plane subtree of T above v_j^i . Lemma 6 shows that, for fixed

i , conditionally on $A_{i,k}$, the $T_{v_j^i}$ are i.i.d. distributed like T . From our discussion above, we have that conditionally on $A_{i,k}$, $\check{T} = t$ if and only if $\check{T}_{v_j^i} = t_j$ for all j . Since t_j has fewer than n vertices, the inductive hypothesis implies

$$\begin{aligned} \mathbb{P}(\check{T} = t) &= \sum_{i=1}^{\infty} \mathbb{P}(T \in A_{i,k}) \mathbb{P}(\check{T}_{v_j^i} = t_j \text{ for all } j \mid T \in A_i) \\ &= \left(\prod_{j=1}^k \mathbb{P}(T' = t_j) \right) \left(\sum_{i=1}^{\infty} \mathbb{P}(T \in A_{i,k}) \right). \end{aligned}$$

Hence it remains to show that

$$\sum_{i=1}^{\infty} \mathbb{P}(T \in A_{i,k}) = \zeta_k.$$

Let $(X_i)_{i=0}^{\infty}$ be i.i.d. distributed like ξ . We then have

$$\mathbb{P}(T \in A_{i,k}) = \mathbb{P}\left(X_0 = 0, \sum_{j=1}^i (X_j - 1) = k, X_j - 1 \geq 0 \text{ for } 1 \leq j \leq i\right) = \xi_0 [z^k] \phi(z)^i,$$

where for a power series $\psi(z)$, $[z^k]\psi$ is the coefficient of z^k . Thus we have

$$\sum_{i=1}^{\infty} \mathbb{P}(T \in A_{i,k}) = \xi_0 [z^k] \sum_{i=0}^{\infty} \phi(z)^i = [z^k] \frac{\xi_0}{1 - \phi(z)} = \zeta_k,$$

where the interchange of limits is justified by positivity of the coefficients involved and we may start the second sum at 0 since $k \geq 1$. $\square$

4.3 The partition at the root

Let $\xi = (\xi_i)_{i \geq 0}$ be a probability distribution with mean 1 and variance $0 < \sigma_1^2 < \infty$. Let T be a Galton-Watson tree with offspring distribution ξ (denote the law of T by GW_{ξ}). Let $A \subseteq \mathbb{Z}^+$ contain 0 and construct $\hat{T}$ as in Corollary 4. Then, by Theorem 20, $\hat{T}$ is a Galton-Watson tree. Let ζ be its offspring distribution. Again by Theorem 20, ζ has mean 1 and variance $\sigma^2 = \sigma_1^2 / \xi(A)$. Assume that for sufficiently large n , $\mathbb{P}(\#_A T = n) > 0$. Let T_n^A be T conditioned to have exactly n vertices with out-degree in A (whenever this conditioning makes sense).

For a t be rooted unordered tree with exactly n vertices with out-degree in A , let $\Pi^A(t)$ be the partition of n or $n - 1$ (depending on whether or not the degree of the root of t is in A) defined by the number of vertices with out-degree in A in the subtrees of t attached to the root.

Lemma 7. (i) Considered as an unordered tree, the law of T_n^A is equal to $\mathbf{P}_n^q$ where, for $n \geq 2$ such that T_n^A is defined, and $\lambda = (\lambda_1, \dots, \lambda_p) \in \bar{\mathcal{P}}_n^A$, we have

$$q_n(\lambda) = \mathbb{P}(\Pi^A(T_n^A) = \lambda) = \frac{p!}{\prod_{j \geq 1} m_j(\lambda)!} \xi(p) \frac{\prod_{i=1}^p \mathbb{P}(\#_A T = \lambda_i)}{\mathbb{P}(\#_A T = n)}.$$

(ii) Let $X_1, X_2, \dots$ be i.i.d. distributed like $\#_A T$ and $\tau_k = X_1 + \dots + X_k$. We have

$$\mathbb{P}(p(\Pi^A(T_n^A)) = p) = \xi(p) \frac{\mathbb{P}(\tau_p = n - \mathbf{1}(p \in A))}{\mathbb{P}(\tau_1 = n)},$$

and $\mathbb{P}(\Pi^A(T_n^A) \in \cdot \mid \{p(\Pi^A(T_n^A)) = p\})$ is the law of a non-increasing rearrangement of $(X_1, \dots, X_p)$ conditionally on $X_1 + \dots + X_p = n - \mathbf{1}(p \in A)$.

Proof. (i) Letting $c_\emptyset(T_n^A)$ be the root degree of T_n^A and $a_1, \dots, a_p \in \mathbb{N}$ with sum $n - \mathbf{1}(p \in A)$ we have

$$\mathbb{P}(c_\emptyset(T_n^A) = p, \#_A[(T_n^A)_i] = a_i, 1 \leq i \leq p) = \xi(p) \frac{\prod_{i=1}^p \mathbb{P}(\#_A T = a_i)}{\mathbb{P}(\#_A T = n)}. \quad (4.3)$$

Part (i) now follows by considering the number of sequences $(a_1, \dots, a_p)$ with the same decreasing rearrangement.

(ii) This follows from Equation (4.3). $\square$

To simplify notation, let q_n be the law of $\Pi^A(T_n^A)$ and let $\mathbf{1}_p = \mathbf{1}(p \in A)$. Let $(S_r, r \geq 0)$ be a random walk with step distribution $(\zeta_{i+1}, i \geq -1)$. By Corollary 4, we have

$$q_n(p(\lambda) = p) = \xi(p) \frac{\frac{p}{n - \mathbf{1}_p} \mathbb{P}(S_{n - \mathbf{1}_p} = -p)}{\frac{1}{n} \mathbb{P}(S_n = -1)} = \frac{n}{n - \mathbf{1}_p} \hat{\xi}(p) \frac{\mathbb{P}(S_{n - \mathbf{1}_p} = -p)}{\mathbb{P}(S_n = -1)}, \quad (4.4)$$

where $\hat{\xi}(p) = p\xi(p)$ is the size-biased distribution of ξ .

Define $\bar{q}_n$ to be the pushforward of q_n onto $\mathcal{S}^\downarrow$ by the map $\lambda \mapsto \lambda / \sum_i \lambda_i$.

For a sequence $(x_1, x_2, \dots)$ of non-negative numbers such that $\sum_i x_i < \infty$, let i^* be a random variable with

$$\mathbb{P}(i^* = i) = \frac{x_i}{\sum_{j \geq 1} x_j}.$$

The random variable $x_1^* = x_{i^*}$ is called a size-biased pick from $(x_1, x_2, \dots)$. Given i^* , we remove the i^* 'th entry from $(x_1, x_2, \dots)$ and repeat the process. This yields a random re-ordering $(x_1^*, x_2^*, \dots)$ of $(x_1, x_2, \dots)$ called the size-biased order (if ever no positive terms remain, the rest of the size-biased elements are 0). Similarly for a random sequence $(X_1, X_2, \dots)$ we define the size-biased ordering by first conditioning on the value of the sequence. For any non-negative measure μ on $\mathcal{S}^\downarrow$, define the size-biased distribution μ^* of μ by

$$\mu^*(f) = \int_{\mathcal{S}^\downarrow} \mu(ds) \mathbb{E}[f(\mathbf{s}^*)],$$

where $\mathbf{s}^*$ is the size-biased reordering of $\mathbf{s}$.

Define the measure ν_2 on $\mathcal{S}^\downarrow$ by

$$\int_{\mathcal{S}^\downarrow} \nu_2(d\mathbf{s}) f(\mathbf{s}) = \int_{1/2}^1 \sqrt{\frac{2}{\pi s_1^3 (1-s_1)^3}} ds_1 f(s_1, 1-s_1, 0, 0, \dots).$$

Theorem 22. *With the notation above,*

$$\lim_{n \rightarrow \infty} n^{1/2} (1-s_1) \bar{q}_n(d\mathbf{s}) = \frac{\sigma_1 \sqrt{\xi(A)}}{2} (1-s_1) \nu_2(d\mathbf{s}),$$

where the limit is taken in the sense of weak convergence of finite measures.

Proof. We follow the reductions in Section 5.1 of [15]. By Lemma 16 in [15] (which is a easy variation of Proposition 2.3 in [6]) it is sufficient to show that

$$\lim_{n \rightarrow \infty} n^{1/2} ((1-s_1) \bar{q}_n(d\mathbf{s}))^* = \frac{\sigma_1 \sqrt{\xi(A)}}{2} ((1-s_1) \nu_2(d\mathbf{s}))^*.$$

Note that for any finite non-negative measure μ on $\mathcal{S}^\downarrow$ and non-negative continuous function $f : \mathcal{S}_1 \rightarrow \mathbb{R}$ we have

$$((1-s_1) \mu(d\mathbf{s}))^*(f) = \int_{\mathcal{S}_1} \mu^*(d\mathbf{x}) (1 - \max \mathbf{x}) f(\mathbf{x}).$$

Consequently the theorem follows from the following Proposition. □

Proposition 9. *Let $f : \mathcal{S}_1 \rightarrow \mathbb{R}$ be continuous and let $g(\mathbf{x}) = (1 - \max \mathbf{x}) f(\mathbf{x})$. Then*

$$\sqrt{n} \bar{q}_n^*(g) \rightarrow \frac{\sigma_1 \sqrt{\xi(A)}}{\sqrt{2\pi}} \int_0^1 \frac{dx}{x^{1/2} (1-x)^{3/2}} g(x, 1-x, 0, \dots).$$

First note that, by linearity, we may assume that $f \geq 0$ and $\|f\|_\infty \leq 1$. We begin the proof of this Proposition with several Lemmas regarding the concentration of mass of $\bar{q}_n^*$. We also note that for the remainder of this section we are following Section 5.1 in [15] very closely with minor differences to account for our more general setting and we invoke Corollary 4 rather than the Otter-Dwass formula – and we get different intermediate constants than they get, but the end results are the same. Nonetheless, the full computation is worth including because it makes clear why the factor of $\sqrt{\xi(A)}$ appears in the scaling limit.

Lemma 8. *For every $\epsilon > 0$, $\sqrt{n} q_n(p(\lambda) > \epsilon \sqrt{n}) \rightarrow 0$ as $n \rightarrow \infty$.*

Proof. Observe that we have the local limit theorem (see e.g. Theorem 3.5.2 in [11])

$$\lim_{n \rightarrow \infty} \sup_{p \in \mathbb{Z}} \left| \sqrt{n} \mathbb{P}(S_n = -p) - \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{p^2}{2n\sigma^2}\right) \right| = 0. \quad (4.5)$$

Using this and Equation (4.4) we have $q_n(p(\lambda) = p) \leq C \hat{\xi}(p)$ for some C independent of n and p . Since ξ has finite variance, $\hat{\xi}$ has finite mean so $\hat{\xi}((k, \infty)) = o(k^{-1})$. The result follows. □

Note that an immediate consequence of the local limit theorem is that for any fixed $k > 0$ we have

$$\lim_{n \rightarrow \infty} \sup_{1 \leq p \leq kn^{1/2}} \left| \sigma \sqrt{2\pi n} \exp\left(\frac{p^2}{2n\sigma^2}\right) \mathbb{P}(S_n = -p) - 1 \right| = 0,$$

and this is often the result we are really using when we cite the local limit theorem.

Lemma 9. *For g as in Proposition 9 we have*

$$\lim_{\eta \downarrow 0} \limsup_{n \rightarrow \infty} \sqrt{n} \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 > 1-\eta\}}) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \sqrt{n} \bar{q}_n^*(\mathbf{1}_{\{x_1 < n^{-7/8}\}}) = 0.$$

Proof. Fix $\eta > 0$. Since we are assuming $\|f\|_\infty \leq 1$, we know that $|g(\mathbf{x})| \leq (1 - x_1)$. Observing that

$$\mathbb{P}(X_1^* = m | X_1 + \dots + X_p = n) = \frac{pm}{n} \frac{\mathbb{P}(X_1 = m) \mathbb{P}(X_2 + \dots + X_p = n - m)}{\mathbb{P}(X_1 + \dots + X_p = n)},$$

and using (ii) in Lemma 7 we see that $\sqrt{n} \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 > 1-\eta\}})$ is bounded above by $o(1)$ plus

$$n^{1/2} \sum_{1 \leq p \leq n^{1/2}} q_n(p(\lambda) = p) \sum_{(1-\eta)n \leq m_1} \left(1 - \frac{m_1}{n - \mathbf{1}_p}\right) \frac{pm_1}{n - \mathbf{1}_p} \frac{\mathbb{P}(X_1 = m_1) \mathbb{P}(\tau_{p-1} = n - m_1 - \mathbf{1}_p)}{\mathbb{P}(\tau_p = n - \mathbf{1}_p)},$$

where the $o(1)$ term is justified by Lemma 8 and our restriction to $1 \leq p \leq n^{1/2}$. Observe that Equation (4.5) implies that

$$\mathbb{P}(\tau_1 = n) = \frac{1}{n} \mathbb{P}(S_n = -1) \sim \frac{1}{\sigma \sqrt{2\pi}} n^{-3/2}.$$

Using Corollary 4, and again that $q_n(p(\lambda) = p) \leq C \hat{\xi}(p)$ for some C independent of n and p , we find that for large n , $\sqrt{n} \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 > 1-\eta\}})$ is bounded above by $o(1)$ plus

$$\begin{aligned} & C n^{1/2} \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) \\ & \times \sum_{(1-\eta)n \leq m_1 < n - \mathbf{1}_p} \left(1 - \frac{m_1}{n - \mathbf{1}_p}\right) \frac{m_1}{n - \mathbf{1}_p} m_1^{-3/2} \frac{n - \mathbf{1}_p}{n - m_1 - \mathbf{1}_p} \frac{\mathbb{P}(S_{n-m_1-\mathbf{1}_p} = -p + 1)}{\mathbb{P}(S_{n-\mathbf{1}_p} = -p)}. \end{aligned}$$

Simplifying we get

$$\sqrt{n} \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 > 1-\eta\}}) \leq o(1) + C \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) \frac{1}{n} \sum_{(1-\eta)n \leq m_1 \leq n - \mathbf{1}_p} \sqrt{\frac{n}{m_1}} \frac{\mathbb{P}(S_{n-m_1-\mathbf{1}_p} = -p + 1)}{\mathbb{P}(S_{n-\mathbf{1}_p} = -p)}.$$

Equation (4.5) implies that, for $1 \leq p \leq n^{1/2}$, $\sqrt{n} \mathbb{P}(S_{n-\mathbf{1}_p} = -p)$ and $\sqrt{n - m_1} \mathbb{P}(S_{n-\mathbf{1}_p} = -p + 1)$ are bounded below and above respectively for some constants independent of n and

p . Hence we have

$$\begin{aligned} \sqrt{n}\bar{q}_n^*(|g|\mathbf{1}_{\{x_1 > 1-\eta\}}) &\leq o(1) + C \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) \frac{1}{n} \sum_{(1-\eta)n \leq m_1 < n - \mathbf{1}_p} \frac{1}{\sqrt{\frac{m_1}{n} \left(1 - \frac{m_1}{n}\right)}} \\ &\leq o(1) + C \sum_{p=1}^{\infty} p^2 \xi(p) \frac{1}{n} \sum_{(1-\eta)n \leq m_1 < n - \mathbf{1}_p} \frac{1}{\sqrt{\frac{m_1}{n} \left(1 - \frac{m_1}{n}\right)}}. \end{aligned}$$

Note that the $m_1 = n$ term has been absorbed into the $o(1)$ term. The upper bound converges to $C \int_{1-\eta}^1 (x(1-x))^{-1/2} dx$, which goes to 0 as $\eta \rightarrow 0$. A little bit of care must be taken here since the integral is improper as a Riemann integral, however this is fine since the sums actually under approximate the integral in this case.

The second limit can be proved in a similar fashion. Note that $\sqrt{n}\bar{q}_n^*(\mathbf{1}_{\{x_1 < n^{-7/8}\}})$ is bounded above by

$$\begin{aligned} &n^{1/2} \sum_{1 \leq p \leq n^{1/2}} q_n(p(\lambda) = p) \sum_{m_1 \leq n^{1/8}} \frac{pm_1}{n - \mathbf{1}_p} \frac{\mathbb{P}(X_1 = m_1) \mathbb{P}(\tau_{p-1} = n - m_1 - \mathbf{1}_p)}{\mathbb{P}(\tau_p = n - \mathbf{1}_p)} + o(1) \\ &\leq Cn^{-3/8} \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) \sum_{1 \leq m_1 \leq n^{1/8}} \frac{\mathbb{P}(\tau_{p-1} = n - m_1 - \mathbf{1}_p)}{\mathbb{P}(\tau_p = n - \mathbf{1}_p)} + o(1) \\ &\leq Cn^{-3/8} \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) \sum_{1 \leq m_1 \leq n^{1/8}} \frac{\mathbb{P}(S_{n-m_1-\mathbf{1}_p} = -p+1)}{\mathbb{P}(S_{n-\mathbf{1}_p} = -p)} + o(1) \\ &\leq Cn^{-1/4} \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) + o(1), \end{aligned}$$

where the last step is justified by the local limit theorem. $\square$

Lemma 10. *For every $\eta > 0$ we have*

$$\lim_{n \rightarrow \infty} \sqrt{n}\bar{q}_n^*(\mathbf{1}_{\{x_1+x_2 < 1-\eta\}}) = 0.$$

Proof. Fix $0 < \epsilon < 1$. Up to addition by an $o(1)$ term depending on ϵ we have that $\sqrt{n}\bar{q}_n^*(\mathbf{1}_{\{x_1+x_2 < 1-\eta\}})$ is bounded above by

$$\begin{aligned} &C\sqrt{n} \sum_{1 \leq p \leq \epsilon n^{1/2}} p \xi(p) \\ &\times \sum_{m_1+m_2 \leq (1-\eta)n} \frac{pm_1}{n - \mathbf{1}_p} \frac{(p-1)m_2}{n - m_1 - \mathbf{1}_p} \frac{\mathbb{P}(X_1 = m_1) \mathbb{P}(X_2 = m_2) \mathbb{P}(\tau_{p-2} = n - m_1 - m_2 - \mathbf{1}_p)}{\mathbb{P}(\tau_p = n - \mathbf{1}_p)}, \end{aligned}$$

where $m_1, m_2 \geq 1$. If $m_1 + m_2 \leq (1-\eta)n$ then $n - m_1 - m_2 \geq \eta n$ and, in particular, the quantity on the left goes to ∞ as n does. Consequently Corollary 4 and Equation (4.5) imply that

$$\frac{\mathbb{P}(\tau_{p-2} = n - m_1 - m_2 - \mathbf{1}_p)}{\mathbb{P}(\tau_p = n - \mathbf{1}_p)} \leq \frac{C}{\eta^{3/2}},$$

for C independent of $1 \leq p \leq \epsilon n^{1/2}$. Our assumption that $\epsilon < 1$ implies that C is independent of ϵ as well. Again using that $n^{3/2}\mathbb{P}(X_1 = n)$ is bounded we have

$$\begin{aligned} \sqrt{n}\bar{q}_n^*(\mathbf{1}_{\{x_1+x_2 < 1-\eta\}}) &\leq o(1) + \frac{C}{\eta^{5/2}\sqrt{n}} \sum_{1 \leq p \leq \epsilon n^{1/2}} p^3 \xi(p) \frac{1}{n^2} \sum_{m_1+m_2 \leq (1-\eta)n} \sqrt{\frac{n}{m_1}} \sqrt{\frac{n}{m_2}} \\ &\leq o(1) + \frac{C\epsilon}{\eta^{5/2}} \sum_{p=1}^{\infty} p^2 \xi(p) \int_0^1 \int_0^1 \frac{dx dy}{\sqrt{xy}}. \end{aligned}$$

Taking the lim sup as $n \rightarrow \infty$ and then letting $\epsilon \rightarrow 0$ yields the result. $\square$

Lemma 11. *There exists a function $\beta_\eta = o(\eta)$ as $\eta \downarrow 0$ such that*

$$\begin{aligned} \lim_{\eta \downarrow 0} \liminf_{n \rightarrow \infty} \sqrt{n}\bar{q}_n^*(g\mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\beta_\eta\}}) &= \lim_{\eta \downarrow 0} \limsup_{n \rightarrow \infty} \sqrt{n}\bar{q}_n^*(g\mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\beta_\eta\}}) \\ &= \frac{\sigma_1 \sqrt{\xi(A)}}{\sqrt{2\pi}} \int_0^1 \frac{g((x, 1-x, 0, \dots))}{x^{1/2}(1-x)^{3/2}} dx. \end{aligned}$$

Proof. Fix $\eta > 0$ and suppose that $\eta' \in (0, \eta)$. Using Lemmas 8 and 9 we decompose according to the events $\{p(\lambda) > \epsilon\sqrt{n}\}$ and $\{\mathbf{x} : x_1 \leq n^{-7/8}\}$ to get

$$\begin{aligned} \sqrt{n}\bar{q}_n^*(g\mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\eta'\}}) &= o(1) + \sqrt{n} \sum_{1 \leq p \leq \epsilon n^{1/2}} q_n(p(\lambda) = p) \\ &\times \sum_{\substack{n^{1/8} \leq m_1 \leq (1-\eta)(n-1_p) \\ (1-\eta')(n-1_p) \leq m_1+m_2 \leq n-1_p}} \mathbb{E}[g((m_1, m_2, X_3^*, \dots, X_p^*, 0, \dots)/(n-1_p)) | \tau_p = n-1_p, X_1^* = m_1, X_2^* = m_2] \\ &\times \frac{pm_1}{n-1_p} \frac{\frac{(p-1)m_2}{n-1_p}}{1 - \frac{m_1}{n-1_p}} \mathbb{P}(X_1 = m_1) \mathbb{P}(X_2 = m_2) \frac{\mathbb{P}(\tau_{p-2} = n - m_1 - m_2 - 1_p)}{\mathbb{P}(\tau_p = n - 1_p)}. \quad (4.6) \end{aligned}$$

Observe that, if $1 \geq x_1 + x_2 \geq 1 - \eta'$ and $x_1 \leq 1 - \eta$, then $x_2/(1 - x_1) \geq 1 - \eta'/\eta$ and $(1 - x_1)/x_2 \geq 1$.

Using the local limit theorem we observe that

$$\sup_{1 \leq p \leq \epsilon n^{1/2}} \left| \sigma \sqrt{2\pi n} \exp\left(\frac{p^2}{2n\sigma^2}\right) \mathbb{P}(S_n = -p) - \sigma \sqrt{2\pi n} \exp\left(\frac{1}{2n\sigma^2}\right) \mathbb{P}(S_n = -1) \right| \rightarrow 0$$

Consequently,

$$\sup_{1 \leq p \leq \epsilon n^{1/2}} \left| \frac{\frac{q_n(p(\lambda)=p)}{\hat{\xi}(p)}}{\exp\left(\frac{1-p^2}{2n\sigma^2}\right)} - 1 \right| = \sup_{1 \leq p \leq \epsilon n^{1/2}} \left| \frac{\sqrt{2\pi\sigma^2 n} \exp(p^2/2n\sigma^2) n \mathbb{P}(S_{n-1_p} = -p)}{\sqrt{2\pi\sigma^2 n} \exp(1/2n\sigma^2) (n-1_p) \mathbb{P}(S_n = -1)} - 1 \right| \rightarrow 0.$$

Thus, for sufficiently large n and small ϵ , we have

$$1 - \eta \leq \frac{q_n(p(\lambda) = p)}{\hat{\xi}(p)} \leq 1 + \eta,$$

for all $1 \leq p \leq \epsilon n^{1/2}$. Using the local limit theorem and Corollary 4 we have

$$\sup_{1 \leq p \leq \epsilon n^{1/2}} \left| p^{-1} \sigma \sqrt{2\pi} n^{3/2} \exp\left(\frac{p^2}{2n\sigma^2}\right) \mathbb{P}(\tau_p = n) - 1 \right| \rightarrow 0.$$

Thus, for sufficiently large n and small ϵ , we have

$$\frac{1 - \eta}{\sigma \sqrt{2\pi}} \leq p^{-1} n^{3/2} \mathbb{P}(\tau_p = n) \leq \frac{1 + \eta}{\sigma \sqrt{2\pi}},$$

for all $1 \leq p \leq \epsilon n^{1/2}$. We note in particular that $\tau_1 = X_1 =_d X_2$. Furthermore, for $n^{1/8} \leq m_1 \leq (1 - \eta)n$ and $m_1 + m_2 \geq (1 - \eta')n$ we have $m_2 \geq (\eta - \eta')n$ so that m_1 and m_2 go to infinity as n does. Thus, for large n (how large now depends on η') we have

$$\frac{(1 - \eta)^2}{2\pi\sigma^2} \leq (m_1 m_2)^{3/2} \mathbb{P}(X_1 = m_1) \mathbb{P}(X_2 = m_2) \leq \frac{(1 + \eta)^2}{2\pi\sigma^2}.$$

Now, recall that f is uniformly continuous on $\mathcal{S}_1$. Furthermore, on the set $\{\mathbf{x} \in \mathcal{S}_1 : x_1 + x_2 > 3/4\}$ we have $\max \mathbf{x} = x_1 \vee x_2$ and $\mathbf{x} \mapsto \max \mathbf{x}$ is thus uniformly continuous on this set. Therefore for $\eta' < (1/4) \wedge \eta^2$ sufficiently small we have

$$|g((m_1, m_2, m_3, \dots)/n) - g((m_1, n - m_1, 0, \dots)/n)| \leq \eta,$$

for every $(m_1, m_2, \dots)$ with sum n sufficiently large such that $m_1 + m_2 \geq (1 - \eta')n$. Take $\beta_\eta := \eta'$.

Given the symmetry of the bounds we have just established it is easy to see that the proofs for the lim sup and lim inf will be nearly identical, one using the upper bounds and the other the lower. We will only write down the proof for the lim inf. For sufficiently large n we have that, up to addition of an $o(1)$ term, $\sqrt{n} \bar{q}_n^* (g \mathbf{1}_{\{x_1 < 1 - \eta, x_1 + x_2 > 1 - \eta'\}})$ is bounded below by

$$\begin{aligned} & \frac{(1 - \eta)^3 (1 - \eta'/\eta)}{(1 + \eta)} \sum_{1 \leq p \leq \epsilon n^{1/2}} (p - 1) \hat{\xi}(p) \frac{1}{n - \mathbf{1}_p} \\ & \times \sum_{n^{1/8} \leq m_1 \leq (1 - \eta)(n - \mathbf{1}_p)} (g((m_1, n - m_1 - \mathbf{1}_p, 0, \dots)/(n - \mathbf{1}_p)) - \eta) \\ & \times \frac{1}{(m_1/(n - \mathbf{1}_p))^{1/2}} \frac{1}{(1 - m_1/(n - \mathbf{1}_p))^{3/2}} \frac{1}{\sigma \sqrt{2\pi}} \\ & \times \sum_{(1 - \eta')(n - \mathbf{1}_p) - m_1 \leq m_2 \leq n - m_1 - \mathbf{1}_p} \mathbb{P}(\tau_{p-2} = n - m_1 - m_2 - \mathbf{1}_p). \end{aligned}$$

Observe that this last sum is equal to $\sum_{m=0}^{\eta'(n - \mathbf{1}_p)} \mathbb{P}(\tau_{p-2} = m)$. By the local limit theorem, this can be made arbitrarily close to 1 independent of $1 \leq p \leq n^{1/2}$. Using the convergence

of Riemann sums (again care must be taken since the integral we get is improper), we have

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \sqrt{n} \bar{q}_n^*(g \mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\eta'\}}) \\ & \geq \frac{(1-\eta)^3(1-\eta'/\eta)}{1+\eta} \sum_{p=1}^{\infty} (p-1) \hat{\xi}(p) \int_0^{1-\eta} \frac{dx}{\sigma \sqrt{2\pi} x^{1/2} (1-x)^{3/2}} (g(x, 1-x, 0, \dots) - \eta) \end{aligned}$$

Letting $\eta \downarrow 0$ coupled with observing that $\sum_{p=1}^{\infty} (p-1) \hat{\xi}(p) = \sigma_1^2$ and recalling that $\sigma^2 = \sigma_1^2 / \xi(A)$ completes the proof. $\square$

Proof of Proposition 9. Observe that

$$|\bar{q}_n^*(g) - \bar{q}_n^*(g \mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\eta'\}})| \leq \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 \geq 1-\eta\}}) + \bar{q}_n^*(|g| \mathbf{1}_{\{x_1+x_2 \leq 1-\eta'\}}).$$

Fix $\epsilon > 0$ and apply Lemmas 9 and 11 to find η, η' such that

$$\sqrt{n} \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 \geq 1-\eta\}}) < \frac{\epsilon}{2}$$

and

$$\left| \sqrt{n} \bar{q}_n^*(g \mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\beta_\eta\}}) - \frac{\sigma_1 \sqrt{\xi(A)}}{\sqrt{2\pi}} \int_0^1 \frac{g(x, 1-x, 0, 0, \dots)}{x^{1/2} (1-x)^{3/2}} dx \right| \leq \frac{\epsilon}{2},$$

for large enough n . For this choice of η, η' and large n we have

$$\left| \sqrt{n} \bar{q}_n^*(g) - \frac{\sigma_1 \sqrt{\xi(A)}}{\sqrt{2\pi}} \int_0^1 \frac{g(x, 1-x, 0, 0, \dots)}{x^{1/2} (1-x)^{3/2}} dx \right| \leq \epsilon + \sqrt{n} \bar{q}_n^*(|g| \mathbf{1}_{\{x_1+x_2 \leq 1-\eta'\}}).$$

By Lemma 10 the upper bound goes to ϵ as $n \rightarrow \infty$, and the result follows. $\square$

As an immediate corollary of these results, we also identify the unnormalized limit of $\bar{q}_n$.

Corollary 5. $\bar{q}_n \xrightarrow{d} \delta_{(1,0,0,\dots)}$.

Proof. Taking $f \equiv 1$ in Proposition 9 gives $\bar{q}_n(1-s_1) \rightarrow 0$. Since L^1 convergence implies convergence in probability, it follows that for all $0 < \eta < 1$ we have $\bar{q}_n(s_1 \geq \eta) \rightarrow 1$. $\square$

Note that, as a consequence of Equation (4.4), we have $q_n(p(\lambda) = p) \rightarrow \hat{\xi}(p)$. Thus, while the degree of the root vertex may be large, only one of the trees attached to the root will have noticeable size.

4.4 Convergence of Galton-Watson trees

We are now prepared to prove Theorem 19, which, after all of our work above, is rather straightforward.

Proof. Lemma 7 shows that T_n^A has law $\mathbf{P}_n^q$ for a particular choice of $(q_n)_{n \geq 1}$. Theorem 22 then shows that the hypotheses of Theorem 18 are satisfied. $\square$

As a consequence of Theorem 19 and the results connecting Schröder's problems to Galton-Watson trees in Section 2.2, we obtain the following theorem.

Theorem 23. *For $i = 1, 2, 3, 4$, let T_n^i be a uniform random tree of the type appearing in Schröder's i 'th problem with n leaves. For each i and n equip T_n^i with the graph metric where edges have length one and the uniform probability measure on its leaves. We then have the following limits with respect to the rooted Gromov-Hausdorff-Prokhorov topology:*

$$\begin{aligned} (i) \quad \frac{1}{\sqrt{n}} T_n^1 &\xrightarrow{d} 2\sqrt{2} T^{Br} & (ii) \quad \frac{1}{\sqrt{n}} T_n^2 &\xrightarrow{d} \frac{\sqrt{2}}{2\sqrt{\sqrt{2}-1}} T^{Br} \\ (iii) \quad \frac{1}{\sqrt{n}} T_n^3 &\xrightarrow{d} 2\sqrt{2} T^{Br} & (iv) \quad \frac{1}{\sqrt{n}} T_n^4 &\xrightarrow{d} \frac{2}{\sqrt{4\log(2)-2}} T^{Br}, \end{aligned}$$

where T^{Br} is the Brownian continuum random tree.

We remark that parts (i) and (iii) were originally proven in [3] and [16] respectively. Parts (ii) and (iv) appear to be new.

Bibliography

- [1] David Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.
- [2] David Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991.
- [3] David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.
- [4] David Aldous and Jim Pitman. The standard additive coalescent. *Ann. Probab.*, 26(4):1703–1726, 1998.
- [5] Jean Bertoin. The asymptotic behavior of fragmentation processes. *J. Eur. Math. Soc. (JEMS)*, 5(4):395–416, 2003.
- [6] Jean Bertoin. *Random fragmentation and coagulation processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.
- [7] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [8] Michael Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability.
- [9] Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combin. Probab. Comput.*, 13(4-5):577–625, 2004.
- [10] Thomas Duquesne. A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.*, 31(2):996–1027, 2003.
- [11] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.
- [12] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

- [13] Bénédicte Haas. Loss of mass in deterministic and random fragmentations. *Stochastic Process. Appl.*, 106(2):245–277, 2003.
- [14] Bénédicte Haas and Grégory Miermont. The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electron. J. Probab.*, 9:no. 4, 57–97 (electronic), 2004.
- [15] Bénédicte Haas and Grégory Miermont. Scaling limits of Markov branching trees, with applications to Galton-Watson and random unordered trees. *arXiv:1003.3632v2 (to appear Ann. Probab.)*, 2010.
- [16] Bénédicte Haas, Grégory Miermont, Jim Pitman, and Matthias Winkel. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *Ann. Probab.*, 36(5):1790–1837, 2008.
- [17] V. F. Kolchin. Branching processes, random trees, and a generalized scheme of arrangements of particles. *Mathematical Notes*, 21:386–394, 1977. 10.1007/BF01788236.
- [18] Igor Kortchemski. Invariance principles for conditioned Galton-Watson trees. *arXiv:1110.2163v1*, 2011.
- [19] Jean-François Le Gall. Itô’s excursion theory and random trees. *Stochastic Process. Appl.*, 120(5):721–749, 2010.
- [20] Jean-François Marckert and Abdelkader Mokkadem. The depth first processes of Galton-Watson trees converge to the same Brownian excursion. *Ann. Probab.*, 31(3):1655–1678, 2003.
- [21] Peter McCullagh, Jim Pitman, and Mathias Winkel. Gibbs fragmentation trees. *Bernoulli*, 14:988–1002, 2008.
- [22] A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canad. J. Math.*, 30(5):997–1015, 1978.
- [23] Nariyuki Minami. On the number of vertices with a given degree in a Galton-Watson tree. *Adv. in Appl. Probab.*, 37(1):229–264, 2005.
- [24] Yu. L. Pavlov. The asymptotic distribution of maximum tree size in a random forest. *Theor. Probab. Appl.*, 22(3):509–520, 1978.
- [25] Yu. L. Pavlov. Limit distributions of the height of a random forest. *Teor. Veroyatnost. i Primenen.*, 28(3):449–457, 1983.
- [26] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002.

- [27] Jim Pitman. Enumerations of trees and forests related to branching processes and random walks. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 163–180. Amer. Math. Soc., Providence, RI, 1998.
- [28] Jim Pitman and Douglas Rizzolo. Schröder’s problems and scaling limits of random trees. *arXiv:1107.1760v1*, 2011.
- [29] Douglas Rizzolo. Scaling limits of Markov branching trees and Galton-Watson trees conditioned on the number of vertices with out-degree in a given set. *arXiv:1105.2528v1*, 2011.
- [30] E. Schröder. Vier combinatorische probleme. *Z. Math. Physik*, 15:361–376, 1870.
- [31] Richard P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999.
- [32] Bernard Van Cutsem. Combinatorial structures and structures for classification. *Comput. Statist. Data Anal.*, 23(1):169–188, 1996.