Title
Aspects of Exchangeable Partitions and Trees

Author
Haulk, Christopher Kenneth

Publication Date
2011

Peer reviewed|Thesis/dissertation

Aspects of Exchangeable Partitions and Trees

by

Christopher Kenneth Haulk

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

in

Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor James W. Pitman, Chair
Professor David J. Aldous
Professor Lawrence Craig Evans
Professor Steven N. Evans

Spring 2011

Aspects of Exchangeable Partitions and Trees

Abstract

Aspects of Exchangeable Partitions and Trees

by

Christopher Kenneth Haulk

Doctor of Philosophy in Statistics

University of California, Berkeley

Professor James W. Pitman, Chair

Exchangeability – the probabilistic symmetry meaning “invariance under the action of the symmetric group,” or less formally, “irrelevance of labels or indices” – has been the subject of continuing interest to probabilists and statisticians since de Finetti’s celebrated characterization of infinite exchangeable sequences of Bernoulli random variables as mixtures of IID sequences. The topic of this dissertation is exchangeability as it pertains to random partitions and trees.

The main result is a de Finetti-type theorem characterizing a class of exchangeable trees called *hierarchies* which arise in connection with fragmentation processes and hierarchical clustering problems. The other results are somewhat related in that they involve consideration of moments of statistics of exchangeable partitions or trees. One of these concerns random trees with leaves labeled by consecutive natural numbers which are exchangeable in the sense that deterministic permutation of the leaf labels does not change the distribution of the tree. In such trees, the set of interleaf distances is exchangeable, and so, for example, the distance between leaf 1 and leaf 2 is equal in distribution to the distance between leaf 2 and leaf 3. Distributional constraints of this type arising from exchangeability can be used to characterize “finite dimensional marginals” of well-understood trees such as the Brownian CRT. As an application we show that the Brownian CRT is the scaling limit of uniform random hierarchies. Another result is the characterization of the two-parameter family of Ewens-Pitman partitions by a kind of *deletion* property: speaking loosely, the Ewens-Pitman family is the class of exchangeable partitions in which the block containing 1 carries no information about the rest of the partition.

To my parents.

Contents

List of Figures

Acknowledgments

My deep gratitude to Jim Pitman for advising me.

I thank my dissertation committee for guidance and many interesting courses over the last five years.

I thank my instructors in the Berkeley Statistics Department for my education.

I thank Jennifer Lee and Karl Rohe for all the good times.

I thank my parents and my brother for their unconditional love and support.

Chapter 1

Introduction

1.1 Exchangeability

A finite sequence $(X_i, 1 \leq i \leq n)$ of random variables is said to be *exchangeable* if for every permutation σ of $[n] := \{1, \ldots, n\}$ there is the following equality in distribution,

$$(X_1, \ldots, X_n) \stackrel{d}{=} (X_{\sigma(1)}, \ldots, X_{\sigma(n)}), \tag{1.1}$$

and likewise an infinite sequence $(X_i, i \geq 1)$ of random variables is said to be exchangeable if (1.1) holds for all $n \geq 1$ and all permutations σ of $[n]$ [40]. Less formally, random variables are exchangeable if the order in which they are presented is irrelevant or if labels are unimportant. This latter formulation makes plain the appeal of exchangeability for modeling purposes, for it is often tempting to regard the order in which data is presented as irrelevant [38]. The key fact about exchangeable sequences is that they are *mixtures of IID sequences.* That is, every infinite exchangeable sequence is derived as if by first sampling a realization μ_0 of a random probability distribution μ and then sampling a realization of an IID sequence with common distribution μ_0. Bruno de Finetti's Theorem expresses this notion more precisely.

Theorem 1 (de Finetti [30, 15], Hewitt & Savage [59])**.** *If $(X_i, i \geq 1)$ is an infinite sequence of exchangeable random variables then on the same probability space as (X_i) there is random probability distribution μ called the* directing measure of (X_i) *for which for every $n \geq 1$ and every sequence $f_1, \ldots, f_n$ of bounded measurable functions there is the following almost sure equality,*

$$\mathbb{E}\left[f_1(X_1) \ldots f_n(X_n) \mid \mu\right] = \prod_{i=1}^{n} \int_{-\infty}^{\infty} f_i(x)\ \mu(dx).$$

Furthermore, the sequence of empirical distributions of $(X_1, \ldots, X_n)$ converges to μ in distribution almost surely; i.e., with δ_x denoting the Dirac mass at x,

$$\mu = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \quad a.s.$$

where the limit is taken in the distributional sense. The measure μ is therefore measurable with respect to the tail sigma field $\bigcap_{n\geq 1}\sigma(X_n, X_{n+1}, \ldots)$ of (X_i).

For extensive discussion of exchangeability and related topics see see [11].

1.2 Exchangeable combinatorial structures

The notion of exchangeability has been generalized considerably in the 85 years since de Finetti first opened the discussion on this topic. Notably there are now de Finetti-type characterizations of infinite exchangeable random arrays [10], partially ordered sets [64], graphs [33], total orders [62, 50], and partitions [70]. *Exchangeability* for these objects means invariance in distribution for some obvious or natural action of the symmetric group [6]. The case for exchangeable partitions is illustrative: a random partition $\Pi = \{B_1, B_2, \ldots\}$ of $\mathbb{N}$ is said to be *exchangeable* if for every finite permutation σ of $\mathbb{N}$, there is the distributional equality

$$\Pi \stackrel{d}{=} \sigma(\Pi) := \{\{\sigma(b) : b \in B\} : B \in \Pi\},$$

i.e. if for every bijection $\sigma : \mathbb{N} \mapsto \mathbb{N}$ that fixes all but finitely many natural numbers the partition derived by relabeling the contents of the sets or *blocks* that constitute Π by σ is equal in law to Π. The de Finetti-type characterization of exchangeable random partitions, provided by Kingman, may be stated succinctly as follows:

Theorem 2 (Kingman [70])**.** *If Π is an exchangeable partition of $\mathbb{N}$ then Π is equal in distribution to a partition derived from an independent and identically distributed sequence $(U_i, i \geq 1)$ of uniform[0,1] random variables and a random open subset $\mathcal{U}$ of the unit interval by putting natural numbers i and j in the same block if and only if either $i = j$ or U_i and U_j fall in the same connected component of $\mathcal{U}$.*

By a *de Finetti-type characterization* is meant a theorem asserting that every distribution of an exchangeable random object can be expressed as a mixture over simpler distributions. Uniqueness of the mixture representation is often asserted. So, in the case of infinite sequences of exchangeable random variables, the class of simpler distributions is the class of distributions of IID sequences, and in the case of exchangeable partitions of $\mathbb{N}$, these simpler distributions are distributions of partitions derived as in Theorem 2 from a sequence of IID uniform random variables (U_i) and a *nonrandom* open set $\mathcal{U}$ by putting natural numbers i and j in the same block if and only if U_i and U_j lie in the same connected component of $\mathcal{U}$.

1.3 Exchangeable hierarchies

A *hierarchy* (a.k.a. total partition, laminar family, phylogeny) on a set S is a collection $\mathcal{H}$ of subsets of S satisfying two conditions:

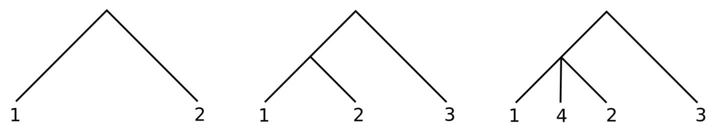

Figure 1.1: Examples of graphs of hierarchies.

- $\varnothing \in \mathcal{H}$, $S \in \mathcal{H}$, and for all $s \in \mathcal{H}$, $\{s\} \in \mathcal{H}$
- If $A, B \in \mathcal{H}$ then either $A \cap B$ equals A or B or $\varnothing$.

The *leaves* of a hierarchy on a set S are the singleton subsets of S. Hierarchies are associated to certain trees, as suggested by Figure 1.1 and explained in detail in Chapter 2. Hierarchies appear as solutions of *hierarchical clustering* problems [1], and in probability hierarchies appear in connection with fragmentation and coagulation processes [23]. In these contexts it is natural to consider *exchangeable* random hierarchies, that is, random hierarchies whose distributions are unchanged by permutations of the leaves. More precisely, if $\mathcal{H}$ is a random hierarchy on a (nonrandom) finite set S then $\mathcal{H}$ is exchangeable if and only if there is the equality in distribution

$$\mathcal{H} \overset{d}{=} \{\{\sigma(s) : s \in H\} : H \in \mathcal{H}\}$$

for every permutation σ of S. A sequence $(\mathcal{H}_n, n \geq 1)$ of hierarchies is *consistent* if for all $n \geq 1$, $\mathcal{H}_n$ is a hierarchy on $[n]$ and the restriction of $\mathcal{H}_{n+1}$ to $[n]$ equals $\mathcal{H}_n$, i.e. if

$$\mathcal{H}_n = \mathcal{H}_{n+1}\Big|_{[n]} := \{H \cap [n] : H \in \mathcal{H}_{n+1}\}.$$

An *exchangeable hierarchy on* $\mathbb{N}$ is a random sequence $(\mathcal{H}_n)$ of consistent hierarchies for which for all $n \geq 1$, $\mathcal{H}_n$ is exchangeable. In Chapter 2 we prove the following de Finetti-type representation theorem for exchangeable hierarchies. A few definitions are necessary to understand the statement of the Theorem: firstly, a *real tree* is a tree-like metric space – a precise definition following [42] is given in Chapter 2 but the essential properties of a real tree are that it be path-connected and loop-free, so that between any two points in the space there is a unique geodesic path; secondly, *weighted, rooted* real tree (T, p) is a real tree T that equipped with a distinguished element ρ called root and a probability measure p on the Borel subsets of T; lastly, for a point x in a in rooted real tree T, the *fringe subtree of* T *rooted at* x is the set $F_x(T) := \{y \in T :$ the geodesic path from y to root ρ passes through $x\}$.

Theorem 3. *If $\mathcal{H}$ is an exchangeable hierarchy on $\mathbb{N}$ then $\mathcal{H}$ is equal in distribution to a hierarchy $\mathcal{H}'$ derived from a random weighted rooted real tree (T, p) and a sequence of random* *elements* $(t_1, t_2, \ldots)$ *of* T *that are independent and identically distributed with distribution* p*, conditionally given* (T, p)*, as follows,*

$$\mathcal{H}' := \{\{n \in \mathbb{N} : t_n \in F_x(T)\} : x \in T\} \cup \{S, \varnothing\} \cup \{\{n\} : n \in \mathbb{N}\}. \tag{1.2}$$

Therefore (1.2) in Theorem 3 says that the fringe subtrees of the weighted rooted real tree (T, p) correspond to elements of the exchangeable hierarchy $\mathcal{H}'$ of that theorem. This is discussed in detail in Chapter 2.

1.4 The deletion property for exchangeable partitions

One way to get new partitions from old is by deleting blocks and remapping. To be more explicit, suppose that $\pi = \{B_1, B_2, \ldots\}$ is a partition of $\mathbb{N}$ into at least two blocks $B_1, B_2, \ldots$ that are ordered by least elements, so that $1 = \min B_1 < \min B_2 < \ldots$, and suppose that $\mathbb{N} \setminus B_1$ is an infinite set. Let F denote the unique increasing bijection that maps the set $\bigcup_{j \geq 2} B_j$ onto $\mathbb{N}$, and define a new partition π' by

$$\pi' := \{\{F(i) : i \in B_j\} : B_j \in \Pi \text{ and } j \geq 2\}. \tag{1.3}$$

Then π' is the partition of $\mathbb{N}$ derived by deleting the first block of π and mapping the resulting partition back to $\mathbb{N}$.

Say that a random partition $\Pi = \{B_1, B_2, \ldots\}$ of $\mathbb{N}$ *has the deletion property* if $\mathbb{N} \setminus B_1$ is almost surely an infinite set and if B_1 and Π' are independent, for Π' the partition of $\mathbb{N}$ derived from Π by deleting the first block as above.

It is easy to construct random partitions with the deletion property. For example: let (L_i) be a sequence of independent positive integer-valued random variables and let $\Pi = \{\{1 \ldots, L_1\}, \{L_1 + 1, \ldots, L_1 + L_2\}, \ldots\}$. It is less easy to construct *exchangeable* partitions with the deletion property. The partition of $\mathbb{N}$ into singletons is a trivial example of an exchangeable random partition with the deletion property. For a less trivial example, fix an integer $M \geq 1$ and let (X_i) be a sequence of IID random variables with $\mathbb{P}(X_i = k) = M^{-1}$ for $k = 1, \ldots M$. Then form the *coupon-collector's partition* by declaring i and j to be in the same block if and only if $X_i = X_j$.

In Chapter 3 we characterize the class of exchangeable random partitions of $\mathbb{N}$ with the deletion property. Aside from the pure singleton partition and the coupon collector's partition, the only exchangeable partitions of $\mathbb{N}$ with the deletion property are members of the two-parameter Ewens-Pitman family of partitions. The distribution of an Ewens-Pitman(α, θ) partition Π is best described by its marginals: if Π_n denotes the restriction of Π to $[n]$ and $\{B_1 \ldots, B_k\}$ is a fixed partition of $[n]$, then

$$\mathbb{P}(\Pi_n = \{B_1, \ldots, B_k\}) = \frac{(\alpha + \theta) \ldots ((k-1)\alpha + \theta)}{(\theta + 1) \ldots (\theta + n - 1)} \prod_{i=1}^{k} (1 - \alpha) \ldots (\#B_i - 1 + \alpha). \tag{1.4}$$

Here, (α, θ) is a pair of real numbers satisfying the condition $0 \leq \alpha \leq 1$ and $\theta > 0$, or $\alpha < 0$ and $\theta = -\alpha M$ for some integer M, the numerator in the ratio is understood to equal 1 if the number k of blocks of $\{B_1, \ldots, B_k\}$ equals 1, and $\#B_i$ denotes the size of the set B_i. This characterization theorem was first published in [46], following Pitman [82, 83]. For earlier work, see [69], where Kingman characterizes exchangeable partitions Π with the deletion property and having the additional property that for Π' derived by deleting the first block of Π, $\Pi' \stackrel{d}{=} \Pi$, as the class of partitions of the form (1.4) for $\alpha = 0, \theta > 0$.

1.5 Characterization of the Brownian CRT and application to uniform hierarchies

A random real tree is a random metric space satisfying a few properties that informally mean that the space is "tree-like." The *Brownian continuum random tree* (CRT) is perhaps the best example of a random real tree. It can be defined recursively as follows: let (X_i) denote the sequence of interpoint distances of a Poisson point process of rate $t\,dt$ on $[0, \infty)$; so in particular X_1 has Rayleigh distribution. Let T_1 be a space isomorphic to a line segment of length X_1. This space T_1 comes equipped with the uniform probability measure which is simply normalized length measure on T_1. Now, recursively for integers $k \geq 1$, assuming T_k has been defined,

- select a point uniformly at random from T_k according to normalized length measure
- graft to the chosen point a branch of length X_{k+1}
- call the resulting space T_{k+1} and define normalized length measure on T_{k+1} in the obvious way.

As in [2] this "intrinsic" or coordinate-free construction can be made explicit by embedding the trees (T_k) as subsets of ℓ_1, the Banach space of absolutely summable real sequences, since this space affords an infinite number of "orthogonal" directions in which new branches can lie without risk of unwanted collisions or intersections with other parts of the growing tree. The Brownian CRT is the closure of the union $\bigcup_{k \geq 1} T_k$, and for $k \geq 1$ the tree T_k is known as the k^{th} *marginal* of the Brownian CRT.

The construction above is called a *line-breaking* construction of a random real tree, because informally speaking the tree is constructed from broken-up bits of the line $[0, \infty)$ [2, 3, 4]. It is possible to augment the trees (T_k) appearing in this construction with leaf labels; that is, integer labels on the endpoints of the segments that are recursively grafted onto the growing tree. So, the segment T_1 has endpoints labeled root and 1, say, and when a new branch is added to T_1 to produce T_2, the free endpoint of this new branch may be labeled 2; and so on, recursively, so that for $k \geq 1$ the tree T_k has endpoints or *leaves* labeled root, $1, 2, \ldots, k$. It is a remarkable feature of this construction that these leaf labels are *exchangeable*: for every $k \geq 1$, if σ is a permutation of $[k]$ (or of $\{\text{root}, 1, \ldots, k\}$) and $\sigma(T_k)$ is the tree derived by permuting the leaf labels by σ, then $\sigma(T_k)$ and T_k are equal in distribution.

In Chapter 4 we prove that the marginals of the Brownian CRT – that is, the family (T_k) of trees appearing in the construction of the Brownian CRT above – is the only family of trees with labeled leaves

- that is derived by a line breaking construction, i.e. by a process of sequentially adding branches whose endpoints have consecutive integer labels
- for which the places where new branches are attached are chosen uniformly at random according to normalized length measure
- for which the first branch length X_1 has Rayleigh distribution
- and for which leaf-labels are exchangeable.

In fact, this characterization of the marginals of the Brownian CRT will follow as a corollary of a theorem with somewhat larger scope.

As an application of this characterization, we prove that the limits of *uniform random hierarchies* – trees of the type appearing in Figure 1.1 – are the marginals of the Brownian CRT. More explicitly: for $n \geq 1$ let $T(n)$ be a random tree selected uniformly at random from the set of all rooted, nonplanar (unoriented) trees with n leaves and no internal vertices of degree two except possibly the root, and edges of length $n^{-1/2}2(\log(2)-1)^{1/2}$. For $1 \leq k \leq n$ let $T_k(n)$ denote the subtree of $T(n)$ spanned by the root and the leaves $1, 2, \ldots, k$. Then for fixed k, as $n \to \infty$, $T_k(n)$ converges in distribution to the k^{th} marginal of the Brownian CRT.

Chapter 2

Exchangeable Hierarchies

A *hierarchy* on a set S, also called a *total partition of* S, is a collection $\mathcal{H}$ of subsets of S such that $S \in \mathcal{H}$, each singleton subset of S belongs to $\mathcal{H}$, and if $A, B \in \mathcal{H}$ then $A \cap B$ equals either A or B or $\varnothing$. Every exchangeable random hierarchy of positive integers has the same distribution as a random hierarchy $\mathcal{H}$ associated as follows with a random real tree $\mathcal{T}$ equipped with root element 0 and a random probability distribution p on the Borel subsets of $\mathcal{T}$: given $(\mathcal{T}, p)$, let $t_1, t_2, \ldots$ be independent and identically distributed according to p, and let $\mathcal{H}$ comprise all singleton subsets of $\mathbb{N}$, and every subset of the form $\{j : t_j \in F_x\}$ as x ranges over $\mathcal{T}$, where F_x is the fringe subtree of $\mathcal{T}$ rooted at x. There is also the alternative characterization: every exchangeable random hierarchy of positive integers has the same distribution as a random hierarchy $\mathcal{H}$ derived as follows from a random hierarchy $\mathscr{H}$ on $[0, 1]$ and a family (U_j) of IID uniform [0,1] random variables independent of $\mathscr{H}$: let $\mathcal{H}$ comprise all sets of the form $\{j : U_j \in B\}$ as B ranges over the members of $\mathscr{H}$.

2.1 Background

Definition 1. *A* hierarchy *on a finite set S is a collection $\mathcal{H}$ of subsets of S such that*

(a) if $A, B \in \mathcal{H}$ then $A \cap B$ equals either A or B or $\varnothing$, and

(b) $S \in \mathcal{H}$, $\{s\} \in \mathcal{H}$ for all $s \in S$, and $\varnothing \in \mathcal{H}$.

Hierarchies are known by several other names, including *total partitions* and *laminar families*. For brevity we use the term *hierarchy* throughout the chapter. If $\mathcal{H}$ is a hierarchy on a finite set S and $S_0 \subseteq S$ then the *restriction of* $\mathcal{H}$ *to* S_0 is the hierarchy on S_0 defined as follows:

$$\mathcal{H}\Big|_{S_0} := \{H \cap S_0 : H \in \mathcal{H}\}. \tag{2.1}$$

Definition 2. *A* hierarchy on $\mathbb{N}$ *is a sequence* $(\mathcal{H}_n, n \geq 1)$ *where for each* n*,* $\mathcal{H}_n$ *is a hierarchy on* $[n]$*, and for every* n*,* $\mathcal{H}_n = \mathcal{H}_{n+1}\Big|_{[n]}$.

Less formally, a hierarchy describes a scheme for recursively partitioning a set S into finer and finer subsets, down to singletons. If S is finite, this has an elementary meaning: S is partitioned into some set of blocks, then recursively: each non-singleton block that remains is partitioned into further blocks, until only singletons remain, and the hierarchy is the entire collection of sets that ever appear in this process. If S is infinite, matters can be more complex: a continuous recursive process of splitting may be involved, as in Bertoin's theory of self-similar or homogeneous fragmentation processes [20, 21] which have a natural regenerative structure. Alternatively, a hierarchy describes a process of coalescence, wherein the singleton subsets of S recursively coagulate to reconstitute the set S. We emphasize that *time* plays no role in our definition of a hierarchy: a hierarchy $\mathcal{H}$ encodes the *contents* of the blocks of some process of fragmentation (or coagulation), but does not include any additional information about the order in which these blocks appear in this fragmentation (or coagulation) process.

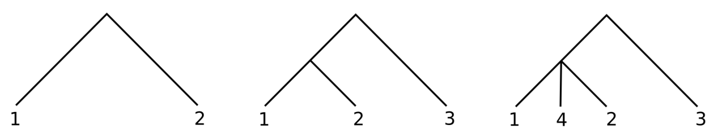

Figure 2.1: The tree on the right is the graph of $\mathcal{H} = \{\{1, 2, 4\}\} \cup \Xi([n])$. The other trees are the graphs of $\mathcal{H}\Big|_{[2]}$ and $\mathcal{H}\Big|_{[3]}$. The trivial hierarchy $\Xi([n])$ is defined at (2.3).

Hierarchies on $[n]$ are in bijective correspondence with certain trees. Explicitly, if $\mathtt{T}$ is a tree

- with n leaves, each labeled by a distinct element of $[n]$,
- and having a distinguished vertex called the root, which is not a leaf,
- with no internal vertices of degree two, except possibly the root
- and no edge lengths or planar embedding

then the map

$$\mathtt{T} \overset{g}{\mapsto} \{\{j \in [n] : \nu \text{ on path from leaf } j \text{ to root}\} : \nu \in V(\mathtt{T})\} \cup \{\varnothing\}$$

sends $\mathtt{T}$ to a hierarchy on $[n]$. Here, $V(\mathtt{T})$ denotes the set of vertices of $\mathtt{T}$ (including root and leaves). The map g is a bijection, and we we say that $\mathtt{T}$ is the *graph* of the hierarchy $g(\mathtt{T})$.

Random hierarchies of both finite and infinite sets arise naturally in a number of applications, including stochastic models for phylogenetic trees [71, 78, 12, 37, 37, 8, 7, 72, 91, 72], processes of fragmentation and coalescence [52, 5, 17, 18, 19, 22, 23, 24, 25, 26, 35, 39, 41, 47, 52, 56], and statistics and machine learning [1, 28, 58, 76, 27]. In these applications, the object of common interest is a rooted tree which describes evolutionary relationships (in the case of phylogentic trees) or the manner in which an object fragments into smaller pieces (in the case of models of fragmentation) or some notion of class membership (in the case of hierarchical clustering). Such trees sometimes have edges equipped with lengths that measure the time between speciations, or the amount of time between fragmentation events, or some measure of dissimilarity or distance between classes, but the hierarchies we consider correspond with trees of this type *without* edge lengths.

Permutations act on hierarchies by relabeling the contents of constituent sets: if $\mathcal{H}$ is a hierarchy on $[n]$ and σ a permutation of $[n]$, then

$$\sigma(\mathcal{H}) := \{\{\sigma(h) : h \in H\} : H \in \mathcal{H}\}.$$

An *exchangeable hierarchy on* $\mathbb{N}$ is a random hierarchy $(\mathcal{H}_n, n \geq 1)$ on $\mathbb{N}$ for which for every n and every permutation σ of $[n]$ there is the distributional equality

$$\sigma(\mathcal{H}_n) \stackrel{d}{=} \mathcal{H}_n. \tag{2.2}$$

The main result in this chapter is a de Finetti-type characterization of exchangeable random hierarchies on $\mathbb{N}$. Theorem 4 states that every exchangeable random hierarchy $\mathcal{H}$ on $\mathbb{N}$ is derived as if by sampling IID points $(t_j, j \geq 1)$ from a random measure μ supported by a random real tree $\mathcal{T}$: the blocks of $\mathcal{H}_n$ are the sets of the form $\{j \in [n] : t_j \in F_x\}$ as x ranges over $\mathcal{T}$, where F_x is the fringe subtree of $\mathcal{T}$ rooted at x. Real trees are tree-like metric spaces that are briefly discussed in Section 2.3.1; for a more complete treatment see [42] and references therein. Theorem 6 is an alternate characterization: every exchangeable hierarchy is derived as if from a sequence (U_j) of IID uniform[0,1] random variables and an independent random hierarchy $\mathscr{H}$ on [0,1]: the blocks of $\mathcal{H}_n$ are the sets of the form $\{j \in [n] : U_j \in B\}$ as B ranges over elements of $\mathscr{H}$. That $\mathscr{H}$ is a *random hierarchy on [0,1]* means simply that $\mathscr{H}$ is a random collection of subsets of $[0, 1]$ that satisfies (a) and (b) of Definition 1 with [0,1] in place of S. For some measure theory details concerning random hierarchies on [0,1] see the Remark at the end of Section 2.5.

As indicated in [16], an exchangeable hierarchy $(\mathcal{H}_n)$ of the set of positive integers $\mathbb{N}$ is generated by each of Bertoin's homogeneous fragmentation processes, and associated with each of Bertoin's homogeneous fragmentations there is a one-parameter family of self-similar fragmentations, each obtained from the homogeneous fragmentation by a suitable family of random time changes, and each generating the same random hierarchy $(\mathcal{H}_n)$ on $\mathbb{N}$. For more on fragmentation processes, see Chapter 3. An attractive feature of the self-similar fragmentations of index $\alpha < 0$ is that each sample path of such a fragmentation is associated with a compact real tree [55]. The sample paths of Kingman's coalescent [71] can likewise be naturally identified with a compact real tree [41]. There has been considerable interest in describing real tree limits of discrete trees with edge-lengths [86, 57, 56, 92, 35], and Theorem 4 of this chapter is in a similar vein.

This work forms part of a growing list of characterizations of infinite exchangeable combinatorial objects by de Finetti-type theorems. For example, Kingman characterized exchangeable partitions of $\mathbb{N}$ [70], Donelly and Joyce and Gnedin characterized composition structures [50, 36], Janson characterized exchangeable posets [64] and Hirth characterized exchangeable ordered trees [60], about which we will say a few words below. Many related de Finetti-type theorems are known [10, 31, 32, 33, 45, 48, 49, 65, 79, 63], and there are excellent treatments in [11, 67] of related material. Such de Finetti-type results are often proved via reverse martingale convergence arguments, similar in spirit to the modern approach to de Finetti's theorem in [40, Chapter 4]. Alternate approaches use harmonic analysis [61, 62, 89, 90] or isometries of L^2 [13], or Choquet theory [59]. The results of this chapter are proved using a third approach, the key idea of which is to encode an exchangeable hierarchy using a binary array, show that this array inherits exchangeability from the hierarchy, and apply well-known characterization theorems for arrays. A similar approach was first used by Aldous, who simplified of Kingman's proof characterizing of exchangeable partitions of $\mathbb{N}$ by encoding such partitions as exchangeable sequences of real random variables [11].

There are several papers on related topics. In [4, Theorem 3] it is shown that if $(\mathcal{R}(k), k \geq 1)$ is a consistent family of exchangeable trees with edge lengths that is *leaf-tight* then $(\mathcal{R}(k), k \geq 1)$ is derived as if by sampling from a random real tree. (Aldous also assumes that his trees are binary, but this assumption is not essential to his proof.) Since a hierarchy on $\mathbb{N}$ corresponds to a sequence of consistent trees *without* edge lengths, the main result of this chapter can be seen as a variation on this result of Aldous, showing that leaf-tightness (and indeed any pre-defined notion of distance) is not needed to obtain a de Finetti type theorem for trees with exchangeable leaves.

In [25], it is shown that every *exchangeable $\mathcal{P}$-coalescent process* corresponds to a unique *flow of bridges*. An exchangeable $\mathcal{P}-$coalescent process is a Markov process $(\Pi_t, t \geq 0)$ whose state space $\mathcal{P}$ is the set of partitions of $\mathbb{N}$, for which Π_t is an exchangeable partition of $\mathbb{N}$ for every $t \geq 0$ whose increments are independent and stationary, if the notion of "increments" of a $\mathcal{P}$-valued function is properly understood. This provides a de Finetti-type characterization of exchangeable coalescents. One may "forget" time by setting $\mathcal{H} := \{B \subset \mathbb{N} : B \in \Pi_t \text{ for some } t > 0\} \cup \{\mathbb{N}\}$ and thereby obtain an exchangeable hierarchy $\mathcal{H}$ on $\mathbb{N}$ (the notation $B \in \Pi_t$ means that B is a block in the partition Π_t). The results of Bertoin and Le Gall in [25] therefore provide a de Finetti-type characterization of hierarchies that arise in this manner from exchangeable coalescents. Due to the stationary, independent increments property, this class of hierarchies is far from including every exchangeable hierarchy, so the present work may be seen as extending the results of Bertoin and Le Gall.

Haas and Miermont [55] provide a de Finetti-type representation of self-similar fragmentations of index $\alpha < 0$ that have no erosion or sudden loss of mass in terms of continuum trees $(\mathcal{T}, p)$ as follows: every such fragmentation $(F(t), t \geq 0)$ is derived as if from a continuum tree (T, p) by setting $F(t)$ equal to the decreasing sequence of masses of connected components of $\{t \in \mathcal{T} : \mathrm{ht}(v) > t\}$ where $\mathrm{ht}(t)$ denotes the distance from t to the root of $\mathcal{T}$. This is proved by introducing a family $(R(k), k \geq 1)$ of trees derived from an associated $\mathcal{P}$-fragmentation (Π_t) whose sequence of ranked limit frequencies equals $(F(t))$. Distances in these trees $R(k)$ are related to times between dislocations in $F(t)$, and by using self-similarity the *leaf-tight* criterion of [4] is checked. The existence of the representing tree $(\mathcal{T}, p)$ is then a consequence of the aforementioned theorem of Aldous. This provides a de Finetti-type theorem for self-similar fragmentations.

In [60], Hirth considers *exchangeable ordered trees*, which in our terms are exchangeable hierarchies $\mathcal{H}$ on $\mathbb{N}$ for which every element $B \in \mathcal{H}$ besides $\mathbb{N}$ there is an associated pair of nonnegative integer-valued times (N_B, M_B) which are the times at which B is "born" and the times at which B "dies." There is also a partial order on such blocks B that is unimportant for our purposes. At the instant of its death, B gives birth to subsets born at that instant, whose union is B. Hirth provides a de Finetti-type characterization of exchangeable ordered trees using harmonic analysis techniques. Our hierarchies are more general than Hirth's trees, since there is no "discrete time" associated to the elements of a hierarchy. Our results may therefore be seen as an extension of Hirth's result using probabilistic techniques instead of harmonic analysis.

2.2 Results

This section provides some basic definitions and a statement of the main results of the chapter.

For arbitrary sets S we define

$$\Xi(S) := \{S\} \cup \{\{s\} : s \in S\} \cup \{\varnothing\}. \tag{2.3}$$

We call $\Xi([n])$ the *trivial hierarchy on* $[n]$; it is the smallest hierarchy on $[n]$ and we will refer to it numerous times throughout the chapter.

If $\mathcal{T}$ is a rooted real tree and $(t_n, n \in \mathbb{N})$ a deterministic or random sequence of points of $\mathcal{T}$, *the hierarchy derived from* $\mathcal{T}$ *and* $(t_n, n \in \mathbb{N})$ is the sequence $(\mathcal{H}_n, n \geq 1)$ defined by

$$\mathcal{H}_n := \{\{j \in [n] : t_j \in F_x(\mathcal{T})\} : x \in \mathcal{T}\} \cup \Xi([n]) \tag{2.4}$$

where $F_x(\mathcal{T})$ is the *fringe subtree of* $\mathcal{T}$ *rooted at* x,

$$F_x(\mathcal{T}) = \{y \in \mathcal{T} : x \text{ is in the geodesic path in } \mathcal{T} \text{ from } y \text{ to the root of } \mathcal{T}\}, \tag{2.5}$$

Real trees are tree-like metric spaces discussed in more detail in Section 2.3.1.

Recall that a random measure p is said to *direct* a family $(t_n, n \in \mathbb{N})$ of random elements if conditionally given p, $(t_n, n \in \mathbb{N})$ is an IID family with distribution p.

Theorem 4. *If $(\mathcal{H}_n, n \geq 1)$ is an exchangeable hierarchy on $\mathbb{N}$ then there is a $(\mathcal{H}_n)$-measurable triple $((\mathcal{H}'_n, n \geq 1), (\mathcal{T}, p), (t_1, t_2, \ldots))$, where $\mathcal{T}$ is a random real tree, p is a random probability measure with support contained in $\mathcal{T}$ almost surely, $(t_1, t_2, \ldots)$ is an exchangeable sequence directed by p, and $(\mathcal{H}'_n)$ is a hierarchy both equal in distribution to $(\mathcal{H}_n)$ and equal almost surely to the hierarchy derived from $\mathcal{T}$ and the samples $(t_1, t_2, \ldots)$.*

Theorem 4 is the main result of the chapter, proved in Section 2.4, where we explicitly construct the pair $(\mathcal{T}, p)$. One of the issues in this construction is how *lengths* in $\mathcal{T}$ are defined. Our main device for defining lengths is the concept of most recent common ancestor.

Definition 3. *If $\mathcal{H}_n$ is a hierarchy on $[n]$ and $i, j \in [n]$, then the* most recent common ancestor (MRCA) of i and j, *denoted $(i \wedge j)_n$, is the intersection of all elements of $\mathcal{H}_n$ that contain both i and j,*

$$(i \wedge j)_n := \bigcap_{G \in \mathcal{H}_n : i, j \in G} G, \tag{2.6}$$

so, e.g., $(i \wedge i)_n = \{i\}$ if $i \leq n$. Later in the chapter, when considering hierarchies on $\mathbb{N}$, we will wish to consider sequences of the form $(i \wedge j)_1, (i \wedge j)_2, (i \wedge j)_3, \ldots$, and in order to make sense of this notion we adopt the convention that if one of i or j is not in $[n]$, then $(i \wedge j)_n := \varnothing$. If $(\mathcal{H}_n, n \geq 1)$ is hierarchy on $\mathbb{N}$, then we denote by $(i \wedge j)$ the MRCA of i and j in $(\mathcal{H}_n)$, which is the following set,

$$(i \wedge j) = \bigcup_{n \geq 1} (i \wedge j)_n \tag{2.7}$$

where $(i \wedge j)_n$ is the MRCA of i and j in $\mathcal{H}_n$. When no confusion seems possible, we sometimes drop parentheses and subscripts from MRCAs to improve legibility. Also, when discussing more than one hierarchy, e.g. $(\mathcal{G}_n)$ and $(\mathcal{H}_n)$, we may write $(i \wedge j)_{\mathcal{G}_n}$ or $(i \wedge j)_{\mathcal{H}_n}$ to denote the MRCA of i and j in $\mathcal{G}_n$ or in $\mathcal{H}_n$.

To presage later developments, the family of indicators $\big(1(k \in (i \wedge j)), k \notin \{i, j\}\big)$ is exchangeable, so the limit

$$1 - \lim_{n \to \infty} \frac{1}{n} \#\{k \in [n] : k \in (i \wedge j)\}$$

exists almost surely. Also, the MRCA of i and j corresponds to a particular vertex in the graph of $\mathcal{T}_n$: the unique vertex found both in the path from root to leaf i and in the path from from root to leaf j that is at maximal graph distance from the root. This vertex has a counterpart ν, say, in the tree $\mathcal{T}$ of Theorem 4, and as will be made clear in the proof of that theorem, the distance from root to ν will the limit displayed above.

For comparison with Theorem 4, we state a version of Kingman's representation theorem for exchangeable partitions. Some preliminary definitions are necessary. Suppose that $\mathscr{P}$ is a fixed or random partition of $[0, 1]$ and that $(U_n, n \geq 1)$ is an IID sequence of uniform[0,1] random variables independent of $\mathscr{P}$. We say that a random partition Π of $\mathbb{N}$ is *derived as if by uniform sampling from* $\mathscr{P}$ if Π is equal in distribution to the partition of $\mathbb{N}$ which puts i and j in the same block if and only if U_i and U_j lie in the same block of $\mathscr{P}$. (We disregard for the moment the measure-theoretic details concerning *random* partitions of $[0,1]$.)

A random partition Π of $\mathbb{N}$ is said to be *exchangeable* if the random array $(\mathbf{p}(i,j), i,j \in \mathbb{N})$ defined by

$$\mathbf{p}(i,j) = \begin{cases} 1 & i \text{ and } j \text{ are in same block of } \Pi \\ 0 & \text{else} \end{cases}$$

is exchangeable, meaning that for every $n \geq 1$ and every permutation σ of $[n]$

$$\Big(\mathbf{p}(\sigma(i),\sigma(j)), i,j \in [n]\Big) \stackrel{d}{=} \Big(\mathbf{p}(i,j), i,j \in [n]\Big). \tag{2.8}$$

The following is a weak version of Kingman's representation theorem for such partitions.

Theorem 5 ([70])**.** *If Π is an exchangeable partition of $\mathbb{N}$ then there is a Π-measurable random partition $\mathscr{P}$ of $[0,1]$ for which Π is derived as if by uniform sampling from $\mathscr{P}$.*

A stronger version of this theorem is stated in Section 2.6.2. It is natural to ask whether Theorem 4 might be reformulated to resemble Theorem 5, and such a reformulation is indeed possible. Continuing to disregard measure-theoretic details, say that a random collection $\mathscr{H}$ of subsets of $[0,1]$ is a *random hierarchy on [0,1]* if conditions (a) and (b) of Definition 1 hold with [0,1] in place of S. We say that $(\mathcal{H}_n)$ is *derived as if by uniform sampling from* $\mathscr{H}$ if $(\mathcal{H}_n)$ is equal in distribution to the sequence of hierarchies $(\mathcal{H}'_n)$ defined by

$$\mathcal{H}'_n = \{\{j \in [n] : U_j \in B\} : B \in \mathscr{H}\},$$

where (U_n) is a sequence of IID uniform random variables independent of $\mathscr{H}$.

Theorem 6. *If $(\mathcal{H}_n)$ is an exchangeable hierarchy on $\mathbb{N}$, then there is an $(\mathcal{H}_n)$-measurable random hierarchy $\mathscr{H}$ on $[0,1]$ for which $(\mathcal{H}_n)$ is derived as if by uniform sampling from $\mathscr{H}$.*

Theorem 6 is proved in Section 2.5 as a Corollary of Theorem 4. The rest of the chapter is organized as follows. Section 2.3 contains three subsections of definitions, well-known results, and elementary propositions needed for the proof of Theorem 4. Section 2.4 contains a proof of Theorem 4. Some complementary discussion and miscellaneous results may be found in Section 2.6.

2.3 Preliminaries

2.3.1 Real trees and hierarchies derived from real trees

Definition 4. *A* segment *of a metric space X is the image of an isometry $\alpha : [a,b] \mapsto X$. The endpoints of the the segment are $\alpha(a)$ and $\alpha(b)$. A* real tree *is a metric space $(\mathcal{T}, d)$ for which* *(a) for every pair x, y of distinct elements of $\mathcal{T}$ there is a unique segment with endpoints x and y, denoted $[[x, y]]$,*

(b) if two segments of $\mathcal{T}$ intersect in a single point, and this point is an endpoint of both, then the union of these two segments is again a segment,

(c) If a segment contains distinct points u, v then it contains $[[u, v]]$,

(d) if the intersection of two segments contains at least two distinct points, then this intersection is a segment.

A real tree is rooted *if there is a distinguished element of $\mathcal{T}$ called root. Every real tree we will discuss will assumed to be rooted, with root denoted* 0*. Furthermore, we define* $[[x, x]] = \{x\}$.

The uniqueness required in part (a) implies that real trees are *loop free*. In fact, parts (c) and (d) of Definition 4 follow from parts (a) and (b). For more regarding real trees see [42]. The following example, however, provides sufficient background on real trees to understand the proof of Theorem 4.

Example 2.3.1 (Line-breaking and a random real tree, following Aldous)**.** *Let ℓ_1 denote the Banach space of absolutely summable real sequences, and let $\mathbf{e}_i$ denote the ith element of the usual basis, so that $\mathbf{e}_1 = (1, 0, 0, \ldots)$, $\mathbf{e}_2 = (0, 1, 0, 0, \ldots)$, and so on, and let (L_n) be a sequence of positive numbers. We define a family of real trees as follows: first, let $u_1 = (0, 0, \ldots)$ and let*

$$\mathcal{T}_1 = u_1 + \mathbf{e}_1[0, L_1] := \{(0, 0, \ldots) + \mathbf{e}_1 x : 0 \le x \le L_1\}.$$

Next, select a point u_2 from $\mathcal{T}_1$ and let

$$\mathcal{T}_2 = \mathcal{T}_1 \cup (u_2 + \mathbf{e}_2[0, L_2]) := \mathcal{T}_1 \cup \{u_2 + \mathbf{e_2} x : 0 \le x \le L_2\}.$$

We continue recursively: supposing $\mathcal{T}_k$ has been defined, we select a point u_{k+1} from $\mathcal{T}_k$ and set

$$\mathcal{T}_{k+1} = \mathcal{T}_k \cup (u_{k+1} + \mathbf{e}_{k+1}[0, L_{k+1}]) \,,$$

and let $\mathcal{T}$ be the closure of the union $\bigcup_{n \ge 1} \mathcal{T}_k$. The tree $\mathcal{T}_k$ is therefore built up by "gluing together" k line segments, and if we endow $\mathcal{T}$ with the ℓ_1 metric the geodesic paths in $\mathcal{T}_k$ flow along these line segments as one would expect.

The idea of using the natural basis of ℓ_1 in order to obtain a countable family of "orthogonal directions" in which to grow the new branch of $\mathcal{T}_k$, is due to Aldous [4].

To get a random *real tree, simply randomize the construction above. For example, let (L_k) be the interarrival times of a Poisson process of on $\mathbb{R}_{\ge 0}$ of rate $t\,dt$, and for $k \ge 2$ select u_k according to normalized length measure on $\mathcal{T}_k$. The resulting tree is Aldous's Brownian continuum random tree.*

We have defined in (2.4) the *hierarchy derived from* $\mathcal{T}$ *and a sequence* $(t_n, n \in \mathbb{N})$ *of points of* $\mathcal{T}$, but to make the definition precise we need to define the *fringe subtree of* $\mathcal{T}$ *rooted at a point* $x \in \mathcal{T}$, a concept used informally at (2.5).

Definition 5. *If* $\mathcal{T}$ *is a real tree and* x *point of* $\mathcal{T}$*, then the* fringe subtree of $\mathcal{T}$ rooted at x *is the set*

$$F_x(\mathcal{T}) := \{y \in \mathcal{T} : x \in [[0, y]]\}.$$

Proposition 1. *Let* $\mathcal{T}$ *be a real tree. Then for* $x, y \in \mathcal{T}$*, either* $F_x(\mathcal{T}) \subset F_y(\mathcal{T})$*, or* $F_y(\mathcal{T}) \subset F_x(\mathcal{T})$*, or* $F_x(\mathcal{T}) = F_y(\mathcal{T})$.

Proof. We claim that for all points $x, y, t \in \mathcal{T}$,

(i) if $x \in [[0, y]]$ and $y \in [[0, t]]$ then $x \in [[0, t]]$, and

(ii) if $x \notin [[0, y]]$ and $y \notin [[0, x]]$ then $F_x(\mathcal{T}) \cap F_y(\mathcal{T}) = \varnothing$.

If x, y, t are distinct non-root elements of $\mathcal{T}$ then (i) above follows from two applications of Part (c) of Definition 4. Likewise, if x, y, t are distinct non-root elements of $\mathcal{T}$ and $x \notin [[0, y]]$ and $y \notin [[0, x]]$, and $t \in F_x(\mathcal{T}) \cap F_y(\mathcal{T})$, then the segments $[[t, x]] \cup [[x, 0]]$ and $[[t, y]] \cup [[y, 0]]$ are distinct (y is not in the first, x is not in the second), and since these segments have the same endpoints we arrive at a contradiction with Part (a) of Definition 4, and (ii) follows. If x, y, t are not distinct or one if one or more of these is the root of $\mathcal{T}$, one may easily argue by cases. □

Corollary 1. *If* $\mathcal{T}$ *is a real tree and* $(t_j, j \geq 1)$ *a sequence of points of* $\mathcal{T}$ *then the sequence* $(\mathcal{H}_n)$ *defined by*

$$\mathcal{H}_n := \{\{j \in [n] : t_j \in F_x(\mathcal{T})\} : x \in \mathcal{T}\} \cup \Xi([n]),$$

is a hierarchy. Here, $\Xi([n])$ *is the trivial hierarchy on* $[n]$ *defined at (2.3).*

2.3.2 Random hierarchies: details

In this section we prove the following elementary proposition and show that hierarchies on $\mathbb{N}$ are in bijective correspondence with certain binary arrays.

Proposition 2. *1. If* $n \geq 1$ *and* $\mathcal{H}_n$ *is a hierarchy on* $[n]$*, then*

$$\mathcal{H}_n = \{(i \wedge j)_n : i, j \in [n]\} \cup \Xi([n])$$

where $\Xi([n])$ *denotes the trivial hierarchy on* $[n]$ *and* $(i \wedge j)_n$ *the MRCA of* i *and* j *in* $\mathcal{H}_n$.

2. If $(\mathcal{H}_n, n \geq 1)$ *is a hierarchy on* $\mathbb{N}$ *then for every* n

$$(i \wedge j) \cap [n] = (i \wedge j)_n,$$

where $(i \wedge j)$ *and* $(i \wedge j)_n$ *denote the MRCAs of* i *and* j *in* $(\mathcal{H}_n)$ *and* $\mathcal{H}_n$*, respectively.*

Proof. 1. Note that the subset of $\mathcal{H}_n$ consisting of sets that contain i is totally ordered by inclusion, by part (a) of Definition 1. The smallest member of this class that contains j is then $(i \wedge j)_n$. This shows that

$$\mathcal{H}_n \supseteq \{(i \wedge j)_n : i, j \in [n]\} \cup \Xi([n]).$$

To prove the reverse inclusion, fix $x \in \mathcal{H}_n$ and $i \in x$. The class $\{(i \wedge j)_n : j \in x\}$ is totally ordered by inclusion, with maximal element $(i \wedge j')_n$, say. Then for all $k \in x$, $k \in (i \wedge k)_n \subseteq (i \wedge j')_n$, so $x \subseteq (i \wedge j')_n$. On the other hand, $i, j' \in x$ and therefore $(i \wedge j')_n \subseteq x$. This proves the reverse inclusion.

2. By consistency of the sequence $(\mathcal{H}_n)$, for every $n \geq \max\{i, j\}$,

$$[n] \cap \bigcap_{G \in \mathcal{H}_{n+1} : \{i,j\} \subseteq G} G = \bigcap_{G \in \mathcal{H}_n : \{i,j\} \subseteq G} G.$$

It follows that $(i \wedge j)_n \subseteq (i \wedge j)_{n+1}$ for every positive n (recall $(i \wedge j)_n = \varnothing$ if $\max\{i, j\} \geq n$). The second assertion follows from this and the fact $(i \wedge j)_n \subseteq [n]$.

$\square$

Proposition 2 shows that if $(\mathcal{H}_n)$ is a hierarchy on $\mathbb{N}$, then the class $\{(i \wedge j) : i, j \in \mathbb{N}\}$ contains complete information about $(\mathcal{H}_n)$, where $(i \wedge j)$ denotes the MRCA of i and j in $(\mathcal{H}_n)$. More explicitly, if the MRCA of i and j in $(\mathcal{H}_n)$ is known, then by restriction we obtain for every n the MRCA of i and j in $\mathcal{H}_n$, and $\mathcal{H}_n$ consists precisely of such MRCAs. The collection $\{(i \wedge j) : i, j \in \mathbb{N}\}$ can be conveniently encoded by the following array,

$$\mathtt{A}_{\mathcal{H}}(i, j, k) = \begin{cases} 1, & \text{if } k \in (i \wedge j) \\ 0, & \text{if } k \notin (i \wedge j) \end{cases} \qquad (i, j, k \in \mathbb{N}), \tag{2.9}$$

Such an array has two notable properties:

(a) For all triples $i, j, k \in \mathbb{N}$, $\mathtt{A}(i, j, k) = \mathtt{A}(j, i, k)$; also $\mathtt{A}(i, j, j) = 1$, and furthemore $\mathtt{A}(i, i, k) = 1$ if and only if $i = k$.

(b) For all pairs i, j and m, n of elements of $\mathbb{N}$, either the two sets

$$\{k \in S : \mathtt{A}(i, j, k) = 1\} \text{ and } \{k \in S : \mathtt{A}(m, n, k) = 1\}$$

are disjoint, or they are equal, or one of them contains the other.

Property (a) follows from symmetry of the roles of i and j in (2.6), and from the fact that $(i \wedge i) = \{i\}$. Property (b) follows from part (b) of Definition 1.

Proposition 3. *The correspondence (2.9) between hierarchies and binary arrays* $\mathtt{A} : \mathbb{N}^3 \mapsto \{0, 1\}$ *having properties (a) and (b) directly above, is bijective.*

The proof of this proposition is elementary and is therefore omitted.

2.3.3 Exchangeable Compositions

A *composition* of a set S is a partition of S together with a total order on blocks of this partition. Starting from such a pair one obtains a binary array $\mathtt{R}$ by setting $\mathtt{R}(i,j) = 1$ if either i and j are in the same block of the partition, or the block containing i precedes the block containing j, and otherwise setting $\mathtt{R}(i,j) = 0$, for all pairs $i, j \in S$. A binary array so derived necessarily has the following four properties, which hold for all $i, j, k \in S$.

(i) $\mathtt{R}(i,i) = 1$

(ii) if $\mathtt{R}(i,j) = 0$ then $\mathtt{R}(i,j) = 1$

(iii) if $\mathtt{R}(i,j) = 1$ and $\mathtt{R}(j,k) = 1$ then $\mathtt{R}(i,k) = 1$

(iv) if $\mathtt{R}(i,j) = 0$ and $\mathtt{R}(j,k) = 0$ then $\mathtt{R}(i,k) = 0$

Conversely, starting from a such an array $\mathtt{R}$ one may define an equivalence relation $\sim$ on S by

$$i \sim j \text{ if and only if } \mathtt{R}(i,j) = \mathtt{R}(j,i) = 1;$$

then the equivalence classes of $\sim$ form a partition of S, and we may totally order these classes by declaring that $[i]$ precedes $[j]$ if and only if $\mathtt{R}(i,j) = 1$, for all pairs i, j in S. This correspondence between a composition of a set S and a binary array $\mathtt{R}$ is obviously bijective. By (i)-(iv) above, the map

$$\mathtt{R} \mapsto \{(i,j) \in S^2 : R(i,j) = 1\}$$

sets up a bijective correspondence between such binary arrays $\mathtt{R}$ and binary relations on S that are reflexive, total, transitive, and whose complements are also transitive. Such relations need not be antisymmetric, and therefore need not be total orders, but every total order is such a relation. By abuse of notation we may use $i \mathrel{\mathtt{R}} j$ and $\mathtt{R}(i,j) = 1$ interchangeably.

We will find it more convenient to work with arrays than with totally ordered set partitions or binary relations, so for our purposes, a *composition of a set S* will mean a binary array $\mathtt{R} : S \times S \mapsto \{0,1\}$ for which properties (i)-(iv) above hold. If S is a finite or countably infinite set then an *exchangeable composition on S* is a random composition $\mathtt{R}$ for which for every finite subset S_0 of S and every permutation σ of S_0, there is the distributional equality

$$\Big(\mathtt{R}(\sigma(i), \sigma(j)), i, j \in S_0\Big) = \Big(\mathtt{R}(i,j), i, j \in S_0\Big).$$

Theorem 7 below is a de Finetti-type characterization of exchangeable compositions, originally given in [50, Theorem 11] and [36, Theorem 5]. Before stating the theorem we must say a few words about *left-uniformization*.

By the *left-uniformization* F_* of a distribution F, we mean the image of F via the map $x \mapsto F_l(x)$ from $\mathbb{R}$ to $[0,1]$, where F_l denotes the left-continuous version of the distribution function of F. That is,

$$\begin{aligned} F_*[0,a] &= \mathbb{P}(F_l(X) \le a) \text{ for } X \text{ with distribution specified by} \\ &\quad \mathbb{P}(X \le x) = F(-\infty, x], \text{ and } F_l(x) = \lim_{w \uparrow x} F(-\infty, w] = F(-\infty, x). \end{aligned}$$

It is well-known that if F is a continuous distribution function, then F_* is the uniform distribution on $[0,1]$. More generally, if the discrete part of F has atoms of magnitude f_i and locations x_i, where $f_i \ge 0$ and $\sum_i f_i \le 1$, then F_* is characterized by the following three properties:

- (i) the distribution F_* has an atom of magnitude f_i at $u_i \in [0,1]$, where $u_i = F(-\infty, x_i)$, for each i;
- (ii) the distribution F_* places no mass on the interval $I_i := (u_i, u_i + f_i)$, for each i;
- (iii) the continous component of F_* is the restriction of Lebesgue measure on $[0,1]$ to the complement of $\cup_i I_i$.

We say that F is *left-uniformized* if $F_* = F$.

Theorem 7 ([50, Theorem 11] and [36, Theorem 5])**.** *If* $\mathtt{R}$ *is an exchangeable composition of* $\mathbb{N}$ *then the limit*

$$X_j = \lim_{m \to \infty} \frac{1}{m} \#\{n \in \{1, \dots, m\} : \mathtt{R}(j,n) = 0\} \tag{2.10}$$

exists almost surely for every $j \in \mathbb{N}$*. The family* $(X_j, j \in \mathbb{N})$ *so defined is exchangeable, and the directing measure of the family is left-uniformized with probability one. Furthermore, almost surely for all pairs* j,k*,* $\mathtt{R}(j,k) = 1$ *if and only if* $X_j \le X_k$*.*

Sketch of proof. For every $j \ge 1$, the family $(Y_n^j, n \ge 1)$ defined by

$$Y_n^j := \mathtt{R}(j, n') \quad n \in \mathbb{N}, \quad n' := \begin{cases} n & \text{if } n < j \\ n+1 & \text{if } n \ge j \end{cases}$$

is exchangeable, and $X_j = \lim_{m \to \infty} m^{-1} \sum_{k=1}^m Y_k^j$. The a.s. existence of the limit in (2.10) is therefore a consequence of de Finetti's theorem. Part of checking that the family (X_j) has the asserted properties involves showing that if $\mathtt{R}(j,i) = 0$ then $X_i < X_j$, and similar arguments using exchangeable sequences derived from $\mathtt{R}$ shows that this implication holds almost surely. The remainder of the argument is straightforward.

$\square$

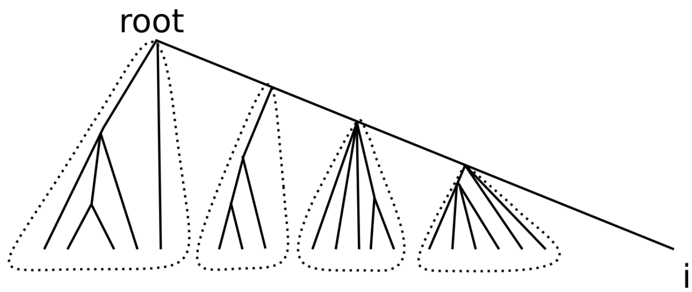

Figure 2.2: The ith spinal spinal composition associated to a hierarchy is the partition of leaves of the hierarchy into blocks according to attachment point on the spinal path from root to leaf i, together with the following ordering on these blocks: block s precedes block t if the attachment point for block s is nearer the root than the attachment point for block t.

2.3.4 Spinal Compositions

Definition 6. *If $(\mathcal{H}_n, n \geq 1)$ is a hierarchy on $\mathbb{N}$ and i an element of S, the i^{th}* spinal composition *of $\mathbb{N} \setminus \{i\}$ is the binary array $\mathtt{R}_i$ defined by*

$$\mathtt{R}_i(j,k) = \mathtt{A}_{\mathcal{H}}(i,j,k) \quad (j,k \in \mathbb{N} \setminus \{i\}) \tag{2.11}$$

where $\mathtt{A}$ is the binary array associated to $(\mathcal{H}_n)$ defined at (2.9).

The i^{th} spinal composition of $\mathbb{N} \setminus \{i\}$ associated to a hierarchy $(\mathcal{H}_n, n \geq 1)$ can be described less formally as follows in terms of the graph of $\mathcal{H}_n$ defined in Section 2.1.

> For $1 \leq i,j,k \leq n$, draw the path from root to leaf i in the graph of $\mathcal{H}_n$. Traverse the vertices of this path starting at the root and moving towards i, and keep track of which vertices contain j and which contain k. If every vertex that contains j also contains k, then $R_i(j,k) = 1$, otherwise $R_i(j,k) = 0$.

See Figure 2.2 for a depiction of a spinal composition.

It is easily checked that $\mathtt{R}_i$ so defined is a composition of $\mathbb{N} \setminus \{i\}$. Furthermore, if $(\mathcal{H}_n)$ is an exchangeable hierarchy on $\mathbb{N}$ then $\mathtt{R}_i$ is an exchangeable composition of $\mathbb{N} \setminus \{i\}$. A version of Theorem 7 then holds, showing the existence of [0,1]-valued random variables

$$X_j^i := \begin{cases} \lim_{m\to\infty} \frac{1}{m} \#\{n \in [m] \setminus \{i\} : \mathtt{R}_i(j,n) = 0\} & j \in \mathbb{N} \setminus \{i\} \\ 1 & j = i \end{cases}. \tag{2.12}$$

The random variables $(X_j^i, j \in \mathbb{N} \setminus \{i\})$ are exchangeable and have a driving measure that is left-uniformized almost surely, and for $j,k \in \mathbb{N} \setminus \{i\}$, $\mathtt{R}_i(j,k) = 1(X_j^i \leq X_k^i)$ holds a.s. We call these variables (X_j^i) *spinal variables.*

Proposition 4. *Let $(\mathcal{H}_n)$ be an exchangeable hierarchy on $\mathbb{N}$, let $\mathtt{A}_{\mathcal{H}}$ and $(\mathtt{R}_i, i \in \mathbb{N})$ be the binary array associated to $(\mathcal{H}_n)$ as by (2.9) and the family of spinal compositions associated to $(\mathcal{H}_n)$, and for every $i \in \mathbb{N}$ let $(X^i_j, j \in \mathbb{N} \setminus \{i\})$ denote the family of spinal variables associated to $\mathtt{R}_i$ by (2.12). Then for all $i, j, k \in \mathbb{N}$ with $i \notin \{j, k\}$ there is the almost sure equality of events,*

$$\{X^i_j \leq X^i_k\} = \{\mathtt{R}_i(j,k) = 1\} = \{\mathtt{A}_{\mathcal{H}}(i,j,k) = 1\} = \{(i \wedge k) \subseteq (i \wedge j)\}, \tag{2.13}$$

where for $i, j \in \mathbb{N}$, $(i \wedge j)$ denotes the MRCA of i and j in $(\mathcal{H}_n)$. Also,

$$X^i_j = \lim_{m\to\infty} \frac{1}{m} \#\{n \in \{1, \dots, m\} : n \notin (i \wedge j)\} \tag{2.14}$$

holds with probability one for all distinct $i, j \in S$. Finally, with probability one, for distinct i, j, k, l in S,

(i) $X^i_j = X^j_i$ *if* $i \neq j$,

(ii) $(i \wedge j) = \{m \in S : X^i_m \geq X^i_j$ *or* $m = i\}$,

(iii) $X^i_k < X^j_k$ *implies* $X^i_k = X^i_j$.

Proof. The almost sure equalities in (2.13) are immediate consequences of definitions; (2.13) simply collects them in one place for easy reference. For (2.14) we note that for every distinct triple i, j, k of distinct elements of S, there is the almost sure equality of events

$$\{\mathtt{R}_i(j,n) = 0\} = \{n \notin (i \wedge j)\}. \tag{2.15}$$

which is immediate from (2.9) and (2.11). Now (2.14) follows from (2.15) and (2.12).

Assertion (i) follows from (2.14) and the fact that $(i \wedge j) = (j \wedge i)$. Assertion (ii) follows from (2.13).

For assertion (iii), suppose that $X^i_k < X^j_k$, then from (i) and (2.14) we have $(j \wedge k) \subsetneq (i \wedge k)$. We will show that $(i \wedge k) = (i \wedge j)$, by (2.14) this is enough for (iii). Already it is plain that $(i \wedge j) \subseteq (i \wedge k)$; it will suffice to show that $k \in (i \wedge j)$. We proceed by cases: since $(i \wedge k) \cap (i \wedge j) \neq \varnothing$, either $(i \wedge k) \subseteq (i \wedge j)$, in which case we are done, or $(i \wedge j) \subseteq (i \wedge k)$. So assume that $(i \wedge j) \subseteq (i \wedge k)$.

- If $(j \wedge k) \subseteq (i \wedge j)$, then $k \in (i \wedge j)$ and we are done.
- If $(i \wedge j) \subseteq (j \wedge k)$, then $(i \wedge k) \subseteq (j \wedge k)$, which is absurd, since $(j \wedge k) \subsetneq (i \wedge k)$.

Because $(j \wedge k) \cap (i \wedge j) \neq \varnothing$, one of the two bulleted cases above must obtain, and we conclude that $k \in (i \wedge j)$ as desired.

$\square$

2.4 Proof of Theorem 4

Let $(\mathcal{H}'_n, n \geq 1)$ be an exchangeable hierarchy on $\mathbb{N}$. For reasons that will soon become clear, it will be much more convenient to work with a hierarchy on $\mathbb{Z}$ rather than on $\mathbb{N}$. Therefore fix an arbitrary bijection $b : \mathbb{N} \mapsto \mathbb{Z}$, and for every $n \geq 1$ set

$$\mathcal{H}_n := \{\{b(k) : k \in (i \wedge j)\} \cap [\pm n] : i, j \in \mathbb{N}\} \cup \Xi([\pm n]), \tag{2.16}$$

where $[\pm n] := \{-n, \ldots, 0, \ldots, n\}$ and Ξ is defined as at (2.3), and $(i \wedge j)$ is the MRCA of i and j in $(\mathcal{H}'_n)$. Then $\mathcal{H}_n$ is a hierarchy on $[\pm n]$ and $\mathcal{H}_{n+1}\Big|_{[\pm n]} = \mathcal{H}_n$ for every $n \geq 1$. We still need the notion of MRCA in $\mathcal{H}_n$ and in $(\mathcal{H}_n)$. Happily, Definition 3 makes sense in the present context with obvious minimal changes, e.g. reading $[\pm n]$ for $[n]$.

We will also need some auxiliary hierarchies, defined as follows.

Definition 7. *For integers $i < 0$, $k < 0$, and $n \geq 1$, and Ξ defined at (2.3), and $(i \wedge l)$ the MRCA of i and l in $(\mathcal{H}_n)$ we set*

$$\mathcal{H}_n^i := \{(i \wedge l) \cap [n] : l \in \mathbb{Z}\} \cup \Xi([n]), \tag{2.17}$$

$$\mathcal{G}_n^k := \{(i \wedge l) : i \in \{-1, \ldots, k\}, l \in \mathbb{Z}\} \cup \Xi([n]) = \bigcup_{i=-1}^{k} \mathcal{H}_n^i. \tag{2.18}$$

$$\mathcal{G}_n := \{(i \wedge l) : i < 0, l \in \mathbb{Z}\} \cup \Xi([n]) = \bigcup_{i=-1}^{-\infty} \mathcal{H}_n^i. \tag{2.19}$$

It is easily checked that $\mathcal{H}_n^i$, $\mathcal{G}_n^k$, and $\mathcal{G}_n$ defined above are hierarchies on $[n]$. We now outline of the proof of Theorem 4.

(i) We define for every $k \leq 1$ a random tree $\mathcal{T}_k$ and a sequence $(t_j^k, j \geq 1)$ of random elements of $\mathcal{T}_k$. Both the tree and the samples are contained in ℓ_1, the Banach space of absolutely summable real sequences. We define another tree $\mathcal{T}$ and samples $(t_j, j \geq 1)$ by

$$\mathcal{T} := \mathrm{cl} \bigcup_{k \leq 1} \mathcal{T}_k, \qquad t_j := \lim_{k \to -\infty} t_j^k, \tag{2.20}$$

where cl denotes ℓ_1-closure, and the limits exist almost surely. Both $(t_j^k, j \geq 1)$ for $k \geq 1$ and $(t_j, j \geq 1)$ are exchangeable, and for the measure p of Theorem 4 we take the directing measure of the sequence $(t_1, t_2, \ldots)$. For $k < -1$ we let p_k denote the directing measure of (t_j^k). These random measures (p_k) are not used in the proof of Theorem 4, but see Figure 2.3 for an image of how p_k and p_{k-1} are related.

(ii) We show that $\mathcal{G}_n$ is the hierarchy derived from $\mathcal{T}$ and the samples $(t_1, \ldots, t_n)$, almost surely for all n. To do this, we first take the following intermediate step:

(ii a) We show that $\mathcal{G}_n^k$ is derived from $\mathcal{T}_k$ and the samples $(t_1^k, \ldots, t_n^k)$, almost surely for all $n \geq 1$, $k \leq -1$. After taking this intermediate step, we establish the assertion of (ii) by taking a limit as $k \to -\infty$.

(iii) We show that the hierarchy $(\mathcal{G}_n, n \geq 1)$ is equal in distribution to the hierarchy $(\mathcal{H}'_n, n \geq 1)$ on $\mathbb{N}$ with which we started.

Steps (ii) and (iii), taken together, prove Theorem 4. The proof of Theorem 4 occupies the remainder of this section, and is broken into parts according the outline above.

2.4.1 Part (i)

Our main tool for constructing the real tree $\mathcal{T}$ and the samples $(t_1, t_2, \ldots)$ of Theorem 4 is the collection of $[0,1]$-valued spinal variables associated to spinal compositions, i.e. the family $(X_j^i, i, j \in \mathbb{Z}, i \neq j)$ defined by

$$X_j^i := \lim_{m \to \infty} \frac{1}{2m} \#\{n \in [\pm m] : n \notin (i \wedge j)\} \quad (i, j \in \mathbb{Z}, i \neq j) \tag{2.21}$$

where $(i \wedge j)$ denotes the MRCA of i and j in $(\mathcal{H}_n)$. We adopt the convention that $X_n^n \equiv 1$ for $n \in \mathbb{Z}$. Obviously Proposition 4 remains true in this context with minimal changes. It is worth emphasizing at this point that superscripts i and k on X_j^i's and t_j^i's, $\mathcal{G}_n^k$'s and $\mathcal{I}_n^k$'s (to be defined later) will be *negative*, and when taking limits we send k to $-\infty$ rather than ∞.

Definition 8. *Let* $(\mathbf{e}_j, j \geq 1)$ *be the natural basis of* ℓ_1*, so that* $\mathbf{e}_1 = (1, 0, 0, \ldots)$*,* $\mathbf{e}_2 = (0, 1, 0, 0, \ldots)$*, etc., and for* $m \geq 1$ *let* π_m *denote the orthogonal projection onto span*$\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$*, so that* $\pi_m((x_1, x_2, \ldots)) = (x_1, \ldots, x_k, 0, 0, \ldots)$*, and* $\pi_0((x_1, x_2, \ldots)) = (0, 0, \ldots)$*.*

Following Aldous [4], for $x \in \ell_1$ let $[[0, x]]_{sp}$ denote the path that proceeds from 0 to x along successive directions, for which $[[0, x]]_{sp}$ equals the closure of $[[0, x]]_{sp}^\circ$, where

$$[[0, x]]_{sp}^\circ := \bigcup_{m \geq 0} \{t\pi_m(x) + (1-t)\pi_{m+1}(x) : 0 \leq t \leq 1\} \tag{2.22}$$

Observe that $[[0, x]]_{sp}$ differs from $[[0, x]]_{sp}^\circ$ only when $x = (x_1, x_2, \ldots)$ does not terminate in zeros, i.e when $x_j > 0$ for infinitely many j, and in this case the set difference $[[0, x]]_{sp} \backslash [[0, x]]_{sp}^\circ$ consists of the singleton $\{x\}$.

Definition 9. *Let* $(X_j^i, i, j \in \mathbb{N}, i \neq j)$ *be the spinal variables defined in (2.21). For all* $j \geq 1$*, set* $t_j^{-1} = \mathbf{e}_1 X_j^{-1}$ *and for every* $k \leq -2$ *set*

$$t_j^k := \mathbf{e}_1 X_j^{-1} + \sum_{l=2}^{k} \mathbf{e}_l \max\{0, X_j^l - \max\{X_j^{-1}, \ldots, X_j^{l-1}\}\} \quad (j \geq 1). \tag{2.23}$$

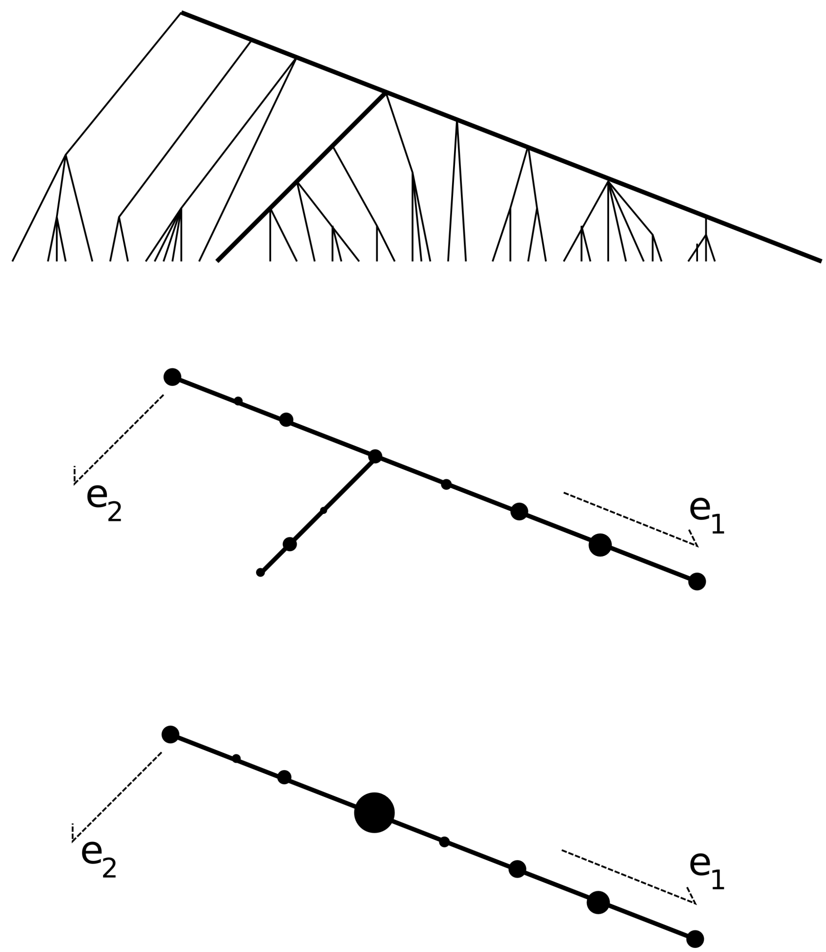

Figure 2.3: At top is shown the graph of $\mathcal{H}_n$ with leaf labels erased. The bold paths are the spinal paths to leaves -1 and -2, respectively. In the middle, $(\mathcal{T}_{-2}, p_{-2})$ is shown. The arrows indicate the ℓ_1 basis directions, and atoms of p_{-2} are represented by black circles or *beads* on $\mathcal{T}_2$, with circle size corresponding to atom size. At bottom is shown $(\mathcal{T}_{-1}, p_{-1})$. Note that (T_{-2}, p_{-2}) is derived from $(\mathcal{T}_{-1}, p_{-1})$ by "crushing" a bead on $\mathcal{T}_{-1}$ into fragments and stringing the crushed bead fragments out in the $\mathbf{e}_2$ direction.

Define a family of trees $(\mathcal{T}_k, k \leq -1)$ *using the samples* $(t_j^k, j \geq 1, k \leq -1)$ *as follows:*

$$\mathcal{T}_k = cl \bigcup_{j \geq 1} [[0, t_j^k]]_{sp},$$

where cl denotes closure in ℓ_1*. For every* $k \leq -1$ *let* d_k *be the* ℓ_1*-metric on* $\mathcal{T}_k$*, and let* $\mathcal{T}_k$ *be rooted at* $0 \in \ell_1$*.*

It is easily checked that for every $k \leq -1$, $(t_j^k, j \geq 1)$ is an exchangeable family. Observe that by definition, $||t_j^k|| = \max\{X_j^{-1}, \ldots, X_j^k\} \leq 1$.

Definition 9 of the samples (t_j^k) can be described as follows: once the samples (t_j^k) and tree $\mathcal{T}_k$ have been defined, to define (t_j^{k-1}) we select a subset of samples among those remaining and push these out in the $\mathbf{e}_{|k-1|}$-direction, orthogonal to $\mathcal{T}_k$ (this subset may possibly be empty). The next proposition shows that every one of these samples is selected from the *same spot* on $\mathcal{T}_k$; that is, $\mathcal{T}_{k-1}$ is derived by adding a *single* branch to $\mathcal{T}_k$ (or perhaps not adding a branch at all).

Proposition 5. *For every* $k \leq -1$*, the set* $\{\pi_{|k|}(t_j^{k-1}) : t_j^{k-1} \neq t_j^k\}$ *is either a singleton or the empty set.*

Proof. Suppose that $X_j^{k-1} > \max\{X_j^1, \ldots, X_j^k\}$. Then for every $i \in \{-1, \ldots, k\}$, $X_j^i < X_j^{k-1}$. Thus by part (iii) of Proposition 4, for every $i \in \{-1, \ldots, -k\}$, $X_j^i = X_{k-1}^i$. We have shown that

$$j \in \{\pi_k(t_j^{k-1}) : t_j^{k-1} \neq t_j^k\} \text{ implies } (X_j^1, \ldots, X_j^k) = (X_{k-1}^1, \ldots, X_{k-1}^k)$$

and we note that t_j^k is determined by $(X_j^1, \ldots, X_j^k)$ to conclude that $\{\pi_{|k|}(t_j^{k-1}) : t_j^{k-1} \neq t_j^k\}$ is a singleton. On the other hand, on the event that $X_j^{k-1} \leq \max\{X_j^1, \ldots, X_j^k\}$, for every $j \leq 0$ then $t_j^{k+1} = t_j^k$ for all j, and the set in question is empty. □

From the definition of $\mathcal{T}_k$ it can be seen that $\mathcal{T}_k$ is a real tree with probability one. It follows from Proposition 5 that $\mathcal{T}_k$ is furthermore a real tree derived by a line-breaking construction, like the tree in the Example in Section 2.3.1.

Examples

The following two examples are not part of the proof of Theorem 4 but together with Figure 2.3 they may help the reader visualize the construction of the tree $\mathcal{T}$.

Example 2.4.1. *Let* $(U_n, n \in \mathbb{Z})$ *be a family of IID uniform[0,1] random variables, and let*

$$\mathcal{H}_n := \{\{j \in [\pm n] : U_j \geq x\} : 0 \leq x \leq 1\} \cup \Xi([\pm n])$$

Following the construction above it can be seen that

$$\mathcal{T}_1 := \mathbf{e}_1[0, U_{-1}],$$

that p_{-1} *is length measure on* $\mathbf{e}_1[0, U_{-1})$*, and that* $\mathbf{e}_1 U_{-1}$ *is an atom of* p_{-1} *of size* $1 - U_{-1}$*. Now let* $k_1 = -1$ *and define a sequence* $(k_m, m \geq 1)$ *recursively by* $k_{m+1} := \max\{i < 0 : U_i > U_{k_m}\}$*. Then* $(\mathcal{T}_{-1}, p_{-1}) = \ldots = (\mathcal{T}_{k_2+1}, p_{k_2+1})$ *and*

$$\mathcal{T}_{k_2} := \mathcal{T}_1 \cup (\mathbf{e}_1 U_{-1} + \mathbf{e}_{k_2}[0, U_{k_2} - U_{k_1}]),$$

i.e., T_{k_2} *is an isometric embedding of* $[0, U_{k_2}]$ *in* ℓ_1*, with a* kink *or bend at the image of* U_{-1} *in* ℓ_1*. The measure* p_{k_2} *is the sum of two measures: length measure on* $\mathcal{T}_{k_2}$*, and an atom of size* $1 - U_{k_2}$ *at the "end"* $\mathbf{e}_{k_1} U_{k_1} + \mathbf{e}_{k_2} U_{k_2}$*. In general,* $\mathcal{T}_{k_m}$ *is a an isometric embedding of* $[0, U_{k_m}]$ *into* ℓ_1 *with* $|k_m| - 1$ *kinks, and* p_k *is length measure on* $\mathcal{T}_{k_m}$ *plus an atom of size* $1 - U_{k_m}$ *at the end of* T_{k_m}*. The limit tree* $\mathcal{T}$ *is an isometric copy of* $[0, 1]$*, embedded in* ℓ_1*, and* p *is length measure on* $\mathcal{T}$*. The tree* $\mathcal{T}$ *has one* leaf *(nonroot element whose removal does not disconnect the space), and this leaf has p-measure 0.*

Example 2.4.2. *Let* $\widehat{\mathcal{H}}$ *denote the following collection of subsets of* $[0, 1]$*,*

$$\widehat{\mathcal{H}} := \bigcup_{n \geq 1} \left\{ \left(\frac{j}{2^n}, \frac{j+1}{2^n} \right) : 0 \leq j \leq 2^n - 1 \right\} \cup \Xi([0, 1]).$$

Let $(U_n, n \in \mathbb{Z})$ *be a family of IID uniform[0,1] random variables, and let*

$$\mathcal{H}_n := \{\{j \in [\pm n] : U_j \in B\} : B \in \widehat{\mathcal{H}}\} \cup \Xi([\pm n])$$

Following the construction above it can be seen that

$$\mathcal{T}_1 := \mathbf{e}_1[0, 1],$$

and that p_{-1} *is purely atomic. The atoms can be described thusly: with* $f_1 := 0$ *and* $f_n := \sum_{j=1}^{n-1} 2^{-n}$*, the atoms of* p_{-1} *are at the locations* $\{\mathbf{e}_1 f_n\}_{n \geq 1}$*, and* $p_{-1}(\{\mathbf{e}_n\}) = 2^{-n-1}$*.*

In fact, for every $k < 0$*, the measure* p_k *is purely atomic. The atoms can be visualized as* beads *on the strings (segments) that constitute* $\mathcal{T}_k$*. To create the next tree* $\mathcal{T}_{k-1}$*, one of the atoms of* p_k *is selected with probability proportional to size and crushed into a sequence of smaller atoms, which are then strung out on the new string, respecting left-uniformization except at location of the crushed atom. More explicitly, suppose that* $\mathcal{T}_k$ *has been defined, and that the selected atom* x *has* p_k*-mass* 2^{-m} *for some* m*. It will follow from the construction that the distance from* x *to* $0 \in \ell_1$ *is* $1 - 2^{-m+1}$*, and that for some finite increasing sequence* $1 \leq i_1 < i_2 \ldots < i_{j-1} < i_j = m$*,* x *looks as follows,*

$$x = (f_{i_1}, 0, f_{i_2} - f_{i-1}, 0, 0, 0, f_{i_3} - f_{i_2}, 0, 0, \ldots, f_{i_j} - f_{i_{j-1}}, 0, 0, 0, \ldots),$$

say; i.e. x *is derived by thinning the vector* $(f_{i_1}, f_{i_2} - f_{i-1}, f_{i_3} - f_{i_2}, \ldots, f_{i_j} - f_{i_{j-1}})$ *with zeros. Suppose that* $f_{i_j} - f_{i_{j-1}}$ *is found in the* l*th coordinate of* x*; this indicates that the branch on* *which atom x is found was added at the lth step of the construction. Now, to create the next tree $\mathcal{T}_{k-1}$, set*

$$\mathcal{T}_{k-1} = \mathcal{T}_k \cup x + \mathbf{e}_{|k|+1}[0, 2^{-m+1}],$$

ie. add a new branch at x in the $\mathbf{e}_{|k|+1}$-direction so that the total distance from root to tip of the new branch is 1. Note that for every point y in the new branch, there will be $|k| + 1 - l$ zeros between the penultimate and final nonzero entries of the y; this explains "zero-thinning".

The new measure p_{k-1} equals p_k on $\mathcal{T}_k \setminus \{x\}$. The atom x is crushed, and $p_{k-1}(x) = 0$; crushed bits of x are strung out on the new branch, so that p_{k-1} has atoms at the following locations,

$$(f_{i_1}, 0, f_{i_2} - f_{i-1}, 0, 0, 0, f_{i_3} - f_{i_2}, 0, 0, \ldots, f_{i_j} - f_{i_{j-1}}, 0, \ldots, 0, 2^n - 2^m, 0, 0, \ldots) \quad (n > m)$$

ie. at $x + \mathbf{e}_{|k|+1}(2^n - 2^m)$ for every $n \geq m$, and $p_{k-1}(x + \mathbf{e}_{|k|+1}(2^n - 2^m)) = 1 - 2^{-n-1}$.

The limit tree $\mathcal{T}$ has uncountably many leaves. The measure p is supported on these leaves, and on the set

$$\{x \in \ell_1 : \pi_j(x) \neq x \text{ for all } j \geq 1\},$$

because for every $j \geq 1$, all atoms on $\mathcal{T}_{-j}$ are eventually selected, crushed, and strung out, off of the set $\{x \in \ell_1 : \pi_j(x) = x\}$. It can also be seen that p is diffuse (i.e. nonatomic), because every atom is eventually crushed into smaller atoms, so no atom of positive mass can remain in the limit.

2.4.2 Part (ii)a

For $n \geq 1$ and $k \leq -1$, let $\mathcal{I}_n^k$ denote the hierarchy derived from $\mathcal{T}_k$ and the samples $(t_1^k, \ldots, t_n^k)$, that is,

$$\mathcal{I}_n^k := \{\{j \in [n] : x \in [[0, t_j^k]]_{sp}\} : x \in \mathcal{T}_k\} \cup \Xi([n]).$$

Proposition 6. *For all positive integers n and k, $\mathcal{G}_n^k = \mathcal{I}_n^k$ almost surely.*

Several intermediate results are needed to prove Proposition 6.

Lemma 1. *Let $\mathcal{H}$ be a hierarchy on a finite set S, and suppose that $i \in B \in \mathcal{H}$. Then almost surely there is $j \in S$ such that $B = (i \wedge j)$, where $(i \wedge j)$ denotes the MRCA of i and j in $\mathcal{H}$. As a corollary, if every element of $\mathcal{H}$ contains i, then*

$$\{(j \wedge l) : j, l \in S, j \neq l\} = \{(j \wedge s) : j \neq s\} \quad a.s. \tag{2.24}$$

Proof. Fix $B \in \mathcal{H}$. By the argument for Proposition 2,

$$B = \bigcup_{j \in B} (s \wedge j).$$

Since the members of $\{s \wedge j : j \in B\}$ all have the point s in common, by part (a) of Definition 1 they are totally ordered by inclusion. Therefore there is some maximal element j' of B for which $B = (s \wedge j')$. $\square$

Lemma 2. *With $\mathcal{H}_n^i$ as in Definition 7,*

$$\{B \cap [n] : i \in B \in \mathcal{H}_n\} = \mathcal{H}_n^i = \{\{j \in [n] : X_j^i \geq x\} : 0 \leq x \leq 1\}$$

holds almost surely for all $n \geq 1$ and $i < 0$.

Proof. This follows from Lemma 1, Proposition 2, and part (ii) of Proposition 4. $\square$

Lemma 3. *For all $k < 0$, and $x \in \mathcal{T}_k$,*

$$\{j > 0 : t_j^{k-1} \in F_x(\mathcal{T}_{k-1})\} = \{j \geq 0 : t_j^k \in F_x(\mathcal{T}_k)\}. \tag{2.25}$$

holds almost surely.

Proof. If t_j^{k-1} is in $F_x(\mathcal{T}_{k-1})$ for $x \in \mathcal{T}_k$ then $x \in [[0, t_j^{k-1}]]_{sp}$, so $x = \pi_{|k|}(x) \in \pi_{|k|}\Big([[0, t_j^{k-1}]]_{sp}\Big) = [[0, t_j^k]]_{sp}$, so $t_j^k \in F_x(\mathcal{T}_k)$. On the other hand, $[[0, t_j^k]]_{sp} \subseteq [[0, t_j^{k-1}]]_{sp}$ so $t_j^k \in F_x(\mathcal{T}_k)$ implies $t_j^{k-1} \in F_x(\mathcal{T}_{k-1})$. $\square$

Proof of Proposition 6. Since $t_j^1 := \mathbf{e}_1 X_j^1$ for $j \geq 1$, for $k = -1$ the assertion is covered by Lemma 2. We proceed by induction on k and argue by cases. Throughout, $(i \wedge j)$ denotes the MRCA of i and j in $(\mathcal{H}_n)$ and $(i \wedge j)_n$ denotes the MRCA of i and j in $\mathcal{H}_n$.

The first case is that the following event occurs: $\mathcal{T}_k = \mathcal{T}_{k-1}$ and $t_j^{k-1} = t_j^k$ for every $j \geq 1$, or otherwise put, there is $i \in \{-1, \ldots, -k\}$ so that $X_j^i \geq X_j^{k-1}$ for all $j \geq 1$. Noting that $(i \wedge j)_n \cap (k-1 \wedge j)_n \neq \varnothing$ for sufficiently large n, from (2.14) and $X_j^i \geq X_j^{k-1}$ we have $(i \wedge j) \subseteq (k-1 \wedge j)$. Thus $i \in (k-1 \wedge j)$, so

$$\{(k-1 \wedge j)_n \cap [n] : j \in [n]\} \subseteq \{(i \wedge j)_n \cap [n] : j \in [n]\},$$

so by Lemma 1,

$$\mathcal{H}_n^{k-1} \subseteq \mathcal{H}_n^i,$$

and it follows that $\mathcal{G}_n^{k-1} = \mathcal{G}_n^k$. Since $\mathcal{I}_n^k = \mathcal{I}_n^{k-1}$ in this case, by the induction hypothesis, $\mathcal{G}_n^k = \mathcal{I}_n^k$.

The second case is that the following event occurs: $\mathcal{T}_k \subsetneq \mathcal{T}_{k-1}$, or otherwise put, for some j, $X_j^{k-1} > \max\{X_j^1, \ldots, X_j^k\}$. It is enough to show two inclusions,

$$\mathcal{I}_n^{k-1} \subseteq \mathcal{G}_n^{k-1} \quad \text{and} \quad \mathcal{H}_n^{k-1} \subseteq \mathcal{I}_n^{k-1}$$

to conclude that $\mathcal{I}_n^{k-1} = \mathcal{G}_n^{k-1}$.

For the first inclusion, we claim that for every point x in the "new branch" $\mathcal{T}_{k-1} \setminus \mathcal{T}_k$, the set $\{j \in [n] : t_j^{k-1} \in F_x(\mathcal{T}_{k-1})\}$ of $\mathcal{I}_n^{k-1}$ is also in $\mathcal{G}_{k-1}$. Therefore fix such x and assume without loss of generality that the set in question is nonempty. If x happens to equal $t_{j_0}^{k-1}$ for some $j_0 \in [n]$ then set $x' = x$, otherwise proceed along the new branch in the 'outward'/away-from-zero/increasing-norm direction until encountering the first sample $t_{j_0}^{k-1}$ with $j_0 \in [n]$, and set $x' = t_j^{k-1}$. More precisely, define x' by

$$x' = \text{ an element of } \{t_j^{k-1} : t_j^{k-1} \in F_x(\mathcal{T}_{k-1}), j \in S_0\} \text{ with minimal } \ell_1\text{-norm,}$$

Then $\{j \in [n] : t_j^{k-1} \in F_x(\mathcal{T}_{k-1})\} = \{j \in [n] : t_j^{k-1} \in F_{x'}(\mathcal{T}_{k+1})\}$. According to (2.23), $\{j \in [n] : t_j^{k-1} \in F_{x'}(\mathcal{T}_{k-1})\} = \{j \in [n] : X_j^{k-1} \geq X_{j_0}^{k-1}\}$ for every element j_0 of $\{t_j^{k-1} : t_j^{k+1} \in F_x(\mathcal{T}_{k-1}), j \in [n]\}$ that has minimal ℓ_1 norm among members of this set. According to Proposition 4(ii), $\{j \in [n] : X_j^{k-1} \geq X_{j_0}^{k-1}\} = (k-1 \cap j_0)_n \cap [n]$, which is an element of $\mathcal{G}_n^{k-1}$. The claim is proved, and in conjunction with the induction hypothesis and Lemma 3 the first inclusion follows.

For the second inclusion, note that if $(k-1 \wedge l)$ contains $i \in \{-1, \ldots, k+1\}$ then $(k-1 \wedge l) \cap [n]$ appears in $\mathcal{G}_n^k$ and hence in $\mathcal{G}_n^{k-1}$ by Lemma 2 and Lemma 3. On the other hand, if $(k-1 \wedge l)$ is disjoint from $\{-1, \ldots, k+1\}$ then

$$\begin{aligned}(k-1 \wedge l) \cap [n] = \{j \in [n] : X_j^k \geq X_j^l\} = \{j \in [n] : t_j^k \in F_x(\mathcal{T}_k)\} \text{ where}\\ x \text{ is the unique point of } \mathcal{T}_k \setminus \mathcal{T}_{k+1} \text{ at distance } X_l^k \text{ from root.}\end{aligned}$$

The second inclusion follows, and we conclude that $\mathcal{I}_n^{k-1} = \mathcal{G}_n^{k-1}$.

The two inclusions taken together show that on the event $\mathcal{T}_k \subsetneq \mathcal{T}_{k+1}$, we have $\mathcal{I}_n^{k-1} = \mathcal{G}_n^{k-1}$ almost surely. This completes the inductive proof.

□

2.4.3 Part (ii)

Proposition 7. *$\mathcal{T}_{-1} \subseteq \mathcal{T}_{-2} \subseteq \ldots$ almost surely, and the limits*

$$t_j := \lim_{k \to -\infty} t_j^k \quad (j \geq 1)$$

exist almost surely and are members of $\mathcal{T} := cl \bigcup_{k \leq -1} \mathcal{T}_k$, where cl denotes ℓ_1-closure.

Proof. By (2.21) the spinal variables (X_j^i) variables take values in $[0, 1]$ almost surely. Observe that by definition, $||t_j^k|| = \max\{X_j^{-1}, \ldots, X_j^k\} \leq 1$, and $\pi_{|k|}(t_j^{k-1}) = t_j^k$. The assertions of the proposition follow from definitions and these two facts. □

Let $\mathcal{I}_n$ be the hierarchy derived from $\mathcal{T}$ and the samples $(t_1, \ldots, t_n)$, i.e.

$$\mathcal{I}_n := \{\{j \in [n] : t_j \in F_x(\mathcal{T})\} : x \in \mathcal{T}\} \cup \Xi([n])$$

where $\Xi([n])$ is the trivial hierarchy on $[n]$.

Proposition 8. *For every positive integer n, $\mathcal{I}_n = \mathcal{G}_n$ almost surely.*

We need the following lemma.

Lemma 4. *For every pair u, v of positive integers, $(u \wedge v) \cap \{-1, -2, \dots\}$ is nonempty with probability one. Here, $(u \wedge v)$ denotes the MRCA of u and v in $(\mathcal{H}_n)$.*

Proof. Define a family $(W_j, j \in \mathbb{Z} \setminus \{i\})$ by

$$W_j = \begin{cases} 1 & \text{if } (u \wedge j) = \{u, j\} \\ 0 & \text{otherwise} \end{cases} \qquad (j \in \mathbb{Z} \setminus \{u, v\}).$$

For distinct integers j_1, j_2, the event $\{W_{j_1} = W_{j_2} = 1\}$ is null set, because $W_{j_1} = W_{j_2} = 1$ implies that $\{j_1, u\}$ and $\{j_2, u\}$ are both members of $\mathcal{H}_n$ for all sufficiently large n, contradicting part (b) of Definition 1. Therefore there is almost surely at most one 1 in the sequence (W_j). Since (W_j) is easily seen to be exchangeable, by de Finetti's theorem $W_j = 0$ almost surely for all j.

It follows that $(\mathtt{A}_H(u, v, j), j \in \mathbb{Z} \setminus \{i, j\})$ is a family of Bernoulli variables with at least one 1, almost surely (see (2.9) for the definition of $\mathtt{A}$). Since $(\mathtt{A}_H(u, v, j))$ is an exchangeable family, the conclusion follows from de Finetti's theorem.

□

Proof of Proposition 8. Set $\mathcal{T}^\circ := \bigcup_{k<-1} \mathcal{T}_k$ and $\partial\mathcal{T} := \mathcal{T} \setminus \mathcal{T}^\circ$. By Proposition 6 and Lemma 3 it follows that

$$\mathcal{G}_n = \bigcup_{k \le -1} \mathcal{G}_n^k = \bigcup_{k \le -1} \mathcal{I}_n^k \subseteq \{\{j \in [n] : t_j \in F_x(\mathcal{T})\} : x \in \mathcal{T}^\circ\} \cup \Xi([n])$$

holds for every positive integer n. It remains to establish that

$$\{j \in [n] : t_j \in F_x(\mathcal{T})\} \subseteq \mathcal{G}_n \tag{2.26}$$

holds for every $x \in \partial\mathcal{T}$. If the set in (2.26) is empty or a singleton it is in $\mathcal{G}_n$ by definition, therefore without loss of generality suppose $\{u, v\} \subseteq \{j \in [n] : t_j \in F_x(\mathcal{T})\}$ for some distinct pair $u, v \in [n]$ and $x \in \partial\mathcal{T}$. We will derive a contradiction.

We claim first that given these assumptions, $t_u = x = t_v$ almost surely. To see this, note that since $x \in \partial\mathcal{T}$, $x = (x_1, x_2, \dots)$ does not terminate in zeros, ie. $x_l \neq 0$ infinitely often. It follows that t_u does not terminate in zeros, i.e. $t_u \in \partial\mathcal{T}$, for otherwise x could not be in $[[0, t_u]]_{sp}$. Now, by definitions it follows that the only point of $[[0, t_u]]_{sp}$ that does not terminate in zeros is t_u itself, so since $x \in [[0, t_u]]_{sp}$ (because $t_u \in F_x(\mathcal{T})$) we must have $x = t_u$, and similarly for t_v.

Since $x = t_u = \lim_{k \to -\infty} t_u^k$ is in $\partial\mathcal{T}$, it follows that there is a subsequence k_m of $\{-1, -2, \dots\}$ for which

$$||t_u^{k_1}|| < ||t_u^{k_2}|| \dots . \tag{2.27}$$

Let $(k_m, m \geq 1)$ be the subsequence of $\{-1, -2, \ldots\}$ consisting of the times at which $(X_u^i, i \leq -1)$ exceeds its past maximum,

$$i_1 = -1 \quad i_{m+1} := \max\{i < k_m : X_u^i > X_u^{k_m}\}.$$

Since $||t_u^k|| = \max\{X_u^{-1}, \ldots, X_u^k\}$, this sequence $(k_m, m \leq 1)$ is well-defined. Now by (2.14) it follows that

$$(k_1 \wedge u) \supsetneq (k_2 \wedge u) \supsetneq \ldots$$

is a strictly decreasing nested family of sets. We claim that $v \in \bigcap_{m \geq 1}(k_m \wedge u)$. This is apparent from the proof of Proposition 6, where it is shown that

$$(k_m \wedge u) \cap [n] = \{j \in [n] : t_j^{k_m} \in F_{t_u^{k_m}}(\mathcal{T}_k)\},$$

since $t_v^{k_m} = \pi_{|k_m|}(t_v) = \pi_{|k_m|}(t_u) = t_u^{k_m}$.

Since u and v are both contained in $\bigcap_{m \geq 1}(k_m \wedge u)$, it follows that $(u \wedge v) \subseteq \bigcap_{m \geq 1}(k_m \wedge u)$. From Proposition 4 there is then with probability one a negative number – we can let i denote the maximum such number – for which $i \in \bigcap_{m \geq 1}(k_m \wedge u)$. It follows that $(i \wedge u) \subsetneq (k_m \wedge u)$ for all m, so that $X_u^i \geq X_u^{k_m}$ for every m by (2.21), contradicting the definition of (k_m).

We have obtained the desired contradiction. It follows that for every fixed $x \in \partial\mathcal{T}$, the set $\{j \in [n] : t_j \in F_x(\mathcal{T})\}$, if nonempty, is with probability one a singleton and therefore an element of $\mathcal{G}_n$. Now observe that

$$\{\{j \in [n] : t_j \in F_x(\mathcal{T})\} : x \in \mathcal{T}\} \cup \Xi([n]) = \{\{j \in [n] : t_j \in F_{t_i}(\mathcal{T})\} : i \in [n]\} \cup \Xi([n]) \quad a.s.$$

Proposition 8 follows. □

Remark. The proof of Proposition 8 shows that the restriction of p to the set $\mathcal{T} \setminus \bigcup_{k<-1} \mathcal{T}_k$ is diffuse, i.e. nonatomic.

2.4.4 Part (iii)

Proposition 9. *The hierarchies $(\mathcal{G}_n, n \geq 1)$ and $(\mathcal{H}'_n, n \geq 1)$ are equal in distribution.*

It should perhaps be pointed out again that $(\mathcal{H}'_n, n \geq 1)$ is the hierarchy on $\mathbb{N}$ with which we started, i.e. with which we defined the hierarchy $(\mathcal{H}_n)$ on $\mathbb{Z}$. We will need the following lemma:

Lemma 5. *The following equality holds almost surely,*

$$\{(i \wedge j) \cap [\pm n] : j \in \mathbb{Z}, i < 0\} = \{(l \wedge j) \cap [\pm n] : l, j \in \mathbb{Z}\},$$

where $(i \wedge j)$ and $(l \wedge j)$ denote MRCAs in $(\mathcal{H}_n)$.

Proof. We need only prove the "$\supseteq$" direction of the equality. Fix l, j in $\mathbb{Z}$, and let μ_j be the directing measure of the exchangeable sequence $(X_n^j, n \in \mathbb{Z} \setminus \{j\})$. There are three cases to consider.

- X_l^j is an atom of the directing measure μ_j. Then there is almost surely some negative integer i for which $X_i^j = X_l^j$. Then by Proposition 4 part (ii),

$$(i \wedge j) = \{k \in \mathbb{Z} : X_k^j \geq X_i^j\} = \{k \in \mathbb{Z} : X_k^j \geq X_l^j\} = (j \wedge l)$$

  and the claim follows.

- X_l^j is not an atom of μ_j. Recalling the discussion of left-uniformization preceding Theorem 7, it can be seen that with probability 1 there is some negative integer i for which $\max\{X_k^j :\in [\pm n], k \notin (j \wedge l)\} < X_i^j < X_l^j$. Then

$$(i \wedge j) \cap [\pm n] = \{m \in [\pm n] : X_m^j \geq X_i^j\} = \{m \in [\pm n] : X_m^j \geq X_l^j\} = (l \wedge j) \cap [\pm n].$$

The third case, on which we need not linger, is that a probability zero event occurs, e.g. X_l^j lies outside the support of μ_j. $\square$

Lemma 6. *There is the following equality in distribution for all n,*

$$\mathcal{H}_n' \stackrel{d}{=} \{(l \wedge j)_{\mathcal{H}} \cap [n] : l, j \in [n] \cup \Xi([n])$$

where $(l \wedge j)_{\mathcal{H}}$ is the MRCA of l and j in $(\mathcal{H}_n)$.

Proof. Let us say that (2.16) defines $(\mathcal{H}_n)$ as *image of* $(\mathcal{H}_n')$ *under* b, and write $(\mathcal{H}_n) = b((\mathcal{H}_n'))$ to express this succinctly. Let c be a bijection from $\mathbb{Z}$ to $\mathbb{Z}$, and let $(\widehat{\mathcal{H}}_n) = c((\mathcal{H}_n))$ be the image of $(\mathcal{H}_n)$ under c. By exchangeability of $(\mathcal{H}_n')$, there is the following equality in distribution,

$$(\widehat{\mathcal{H}}_n) := c((\mathcal{H}_n)) \stackrel{d}{=} (\mathcal{H}_n) := b((\mathcal{H}_n'))$$

which holds for all fixed bijections $c : \mathbb{Z} \mapsto \mathbb{Z}$ and $b : \mathbb{N} \mapsto \mathbb{Z}$. It follows that

$$\widehat{\mathcal{H}}_n\Big|_{[n]} \stackrel{d}{=} \mathcal{H}_n\Big|_{[n]}.$$

Now choose c so that $c(b(j)) = j$ for $j = 1, \ldots, n$. It is straightforward to check that $\widehat{\mathcal{H}}_n\Big|_{[n]} = \mathcal{H}_n'$ almost surely for this choice of c. Note finally that

$$\mathcal{H}_n\Big|_{[n]} = \{(l \wedge j) \cap [n] : l, j \in [n]\} \cup \Xi([n])$$

by Proposition 2. This establishes the claim. $\square$

Proof of Proposition 9. From Lemma 5, Proposition 2, and Lemma 6 we have

$$\begin{aligned}
\mathcal{G}_n &= \{(i \wedge j)_{\mathcal{H}} \cap [n] : i \leq 0, j \in \mathbb{Z}\} \cup \Xi([n]) \\
&= \{(i \wedge j)_{\mathcal{H}} \cap [n] : i, j \in \mathbb{Z}\} \cup \Xi([n]) \\
&= \{(i \wedge j) \cap [n] : i, j \in [n]\} \cup \Xi([n]) \\
&\stackrel{d}{=} \mathcal{H}'_n,
\end{aligned}$$

where $(i \wedge j)_{\mathcal{H}}$ denotes the MRCA of i and j in $(\mathcal{H}_n)$. □

2.5 Proof of Theorem 6

Proof. Suppose WLOG that $(\mathcal{H}_n)$ is the hierarchy derived from a real tree $\mathcal{T}$ and an exchangeable family $(t_j, j \geq 1)$ of random elements of $\mathcal{T}$ having directing measure p. We may further suppose that $\mathcal{T}$ is embedded in ℓ_1 by a stick-breaking procedure as in the proof of Theorem 4 or as in Example of Section 2.3.1. For $k \in \mathbb{N}$ let π_k be the orthogonal projection onto the the span of the first k standard basis elements of ℓ_1,

$$\pi_k((x_1, x_2, \ldots)) = (x_1, \ldots, x_k, 0, 0, \ldots).$$

We will define a map $\xi : [0,1] \mapsto \mathcal{T}$ such that for every $k \geq 1$, and every point $x \in \mathcal{T}_k := \{\pi_k(x) : x \in \mathcal{T}\}$,

(a) $\xi^{-1}(F_x(\mathcal{T}))$ is an interval,

(b) the Lebesgue measure of $\xi^{-1}(F_x(\mathcal{T}))$ equals $p(F_x(\mathcal{T}))$.

To that end, for $k \geq 1$ let p_k be the image of p under π_k. The branches of $\mathcal{T}_k$ can be visualized as strings, and atoms of p_k can be visualized as *beads* on these strings. With this imagery, $(\mathcal{T}_{k+1}, p_{k+1})$ is derived as if by selecting a bead of p_k, crushing this bead into a series of smaller beads, and then drawing these smaller beads out onto the new string $\mathcal{T}_{k+1} \setminus \mathcal{T}_k$, possibly leaving some mass at the location of the crushed atom. (There is also the possibility that $(\mathcal{T}_{k+1}, p_{k+1}) = (\mathcal{T}_k, p_k)$, but this may be ignored.) Let $\xi_1 : [0,1] \mapsto \mathcal{T}_1$ be

- an increasing map, meaning that $x < y$ implies $||\xi_1(x)|| < ||\xi_1(y)||$
- such that p_1 is the image of Lebesgue measure under ξ_1.

Now, $\mathcal{T}_2$ is derived as if by selecting an atom of p_1, crushing it, and stringing the crushed bits in the $\mathbf{e}_2$ direction. If a is the selected atom of p_1, then $\xi_1^{-1}(\{a\})$ is an interval $(u, v]$ or $[u, v]$ in [0,1]. We may therefore define a modification ξ_2 of ξ_1, in such a way that

- ξ_2 agrees with ξ_1 off of $(u, v]$ (or $[u, v]$ as the case may be)

- ξ_2 sends (u, v) onto $a \cup \mathcal{T}_2 \setminus \mathcal{T}_1$
- $u \leq s < t \leq v$ implies $||\xi_2(s)|| \leq ||\xi_2(t)||$
- p_2 is the image of Lebesgue measure under ξ_2.

With this type of construction we can establish the existence of a family of maps $(\xi_k, k \geq 1)$, such that

1. for all $x \in [0, 1]$, $\pi_k(\xi_{k+1}(x)) = \xi_k(x)$
2. if $\xi_k(x) = \xi_k(y)$ and $x < y$, then $\xi_{k+1}(x) < \xi_{k+1}(y)$
3. p_k is the image of Lebesgue measure under ξ_k.

It is straightforward to show that the limit $\xi := \lim_{k\to\infty} \xi_k$ exists Lebsegue a.e. and has the asserted properties (a) and (b). Now set

$$\mathscr{H} := \left\{ \xi^{-1}(F_x(\mathcal{T})) : x \in \bigcup_k \mathcal{T}_k \right\} \cup \Xi([0, 1]).$$

For (U_j) an IID sequence of uniform[0,1] random variables independent of $\mathcal{T}$ and $n \geq 1$ let

$$\mathcal{H}'_n := \{\{j \in [n] : U_j \in B\} : B \in \mathscr{H}$$

and let

$$\mathcal{H}''_n := \{\{j \in [n] : \xi(U_j) \in F_x(\mathcal{T})\} : x \in \mathcal{T}\} \cup \Xi([n]).$$

It is easily seen that $\mathcal{H}'_n = \mathcal{H}''_n$ almost surely. Also, conditionally given $(\mathcal{T}, p)$, the sequence $(\xi(U_1), \ldots, \xi(U_n))$ is an IID sequence of points with common distribution p. An argument such as can be found in the proof of Proposition 8 shows that if $x \in \mathcal{T} \setminus \bigcup_k \mathcal{T}_k$ then $\{j \in [n] : \xi(U_j) \in F_x(\mathcal{T})\}$ is with probability one either empty or a singleton. It follows that

$$(\mathcal{H}''_n) \stackrel{d}{=} (\mathcal{H}_n).$$

□

2.6 Complements

2.6.1 Properties of p and $(\mathcal{H}_n)$

Let $\mathscr{H}$ denote the following class of subsets of the closed interval [0,3],

$$\begin{aligned} \mathscr{H} : &= \{(0,1),(1,2),(2,3)\} \cup \left\{ \bigcup_{n\geq 1} \left\{ \left(\frac{j}{2^n}, \frac{j+1}{2^n}\right) : 0 \leq j \leq 2^n - 1 \right\} \right\} \\ &\cup \{(2,x) : 2 < x < 3\} \cup \Xi([0,3]). \end{aligned}$$

Let $(U_n, n \geq 1)$ be an iid sequence of Uniform[0,3] random variables, and define an exchangeable hierarchy on $\mathbb{N}$ by

$$\mathcal{H}_n := \{\{j \in [n] : U_j \in B\} : B \in \mathscr{H}\}. \tag{2.28}$$

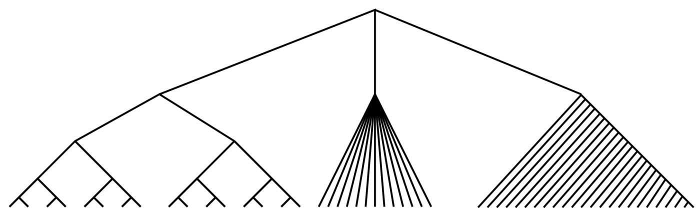

Figure 2.4: Graph of the hierarchy defined by (2.28) with leaf labels omitted.

Figure 2.4 shows the graph $\mathtt{T}_n$ of $\mathcal{H}_n$ for large n, omitting leaf labels. Let us describe a few key features of this graph $\mathtt{T}_n$ and relate them to $\mathscr{H}$.

- The root of $\mathtt{T}_n$ has degree three. The three vertices v_1, v_2, v_3 connected to the root correspond to the three subintervals $(0,1), (1,2), (2,3)$ of $[0,3]$ contained in $\mathscr{H}$.
- The graph of $\mathtt{T}_n$ exhibits recursive binary splitting below the vertex v_1; this is a consequence of the recursive binary splitting of $(0,1)$ in $\mathscr{H}$.
- The graph of $\mathtt{T}_n$ looks star-like or broomstick-like below v_2; this is because $\mathscr{H}$ contains no nonsingleton subsets of (1,2).
- The graph of $\mathtt{T}_n$ looks like a comb or a caterpillar below v_3; this is because $\mathscr{H}$ contains a family of subsets of $(2,3)$ of the form $(2,x)$ for x in a dense subset of $(2,3)$.

From this example, one might make the following naive conjecture.

> **Naive Conjecture:** The three phenomena exhibited by $(\mathcal{H}_n)$ and its graph – infinite recursive splitting, finite splitting, and comblike erosion – are the basic building blocks out of which every exchangeable hierarchy is made.

However, we have difficulty seeing how to make this conjecture more precise: comblike erosion can be interspersed with recursive splitting, splits need not be binary, and a countable family of splits can precede another countable family of splits, and it may be that this latter family of splits is not well-ordered by containment. It is easy to imagine pathological examples of hierarchies.

In lieu of a precise form of the conjecture, we offer the following propositions, which in conjunction with Theorem 4 represent an effort at proving something like the Naive Conjecture. But first, supposing that $(\mathcal{H}_n)$ is a hierarchy on $\mathbb{N}$, we set

$$\alpha_n(j) := \bigcap_{\substack{G \in \mathcal{H}_n \\ j \in G, G \neq \{j\}}} G, \qquad (n \geq j)$$

and call $\alpha_n(j)$ the *parent* of j in $\mathcal{H}_n$; it is easily checked that $\alpha_n(j) \in \mathcal{H}_{n+j}$ for all $n, j \geq 1$.

Also, for $(\mathcal{H}_n)$ a hierarchy on $\mathbb{N}$ and $i, j \in \mathbb{N}$ we write $i \preceq j$ if either $i = j$ or for all $n \geq \max\{i, j\}$,

- $\alpha_n(i) \subsetneq (i \wedge j)_n = \alpha_n(j)$, and
- if u is in $\{i\} \cup (i \wedge j)_n \setminus \alpha_n(i)$ and $v \in [n]$ then

$$\alpha_n(u) = \alpha_n(v) \text{ implies } u = v.$$

Less formally, we write $i \preceq j$ for distinct $i, j \in \mathbb{N}$ if for every $n \geq \max\{i, j\}$, the graph of $\mathcal{H}_n$ looks like a comb in the neighborhood of i and j, and i is "lower down" in this comb than is j. Next, let Π be the partition of $\mathbb{N}$ derived by putting i and j in the same block if and only if either $i \preceq j$ or $j \preceq i$. We say that Π is the *comb-partition* of $(\mathcal{H}_n)$ and the blocks of Π are the *comb components* of $(\mathcal{H}_n)$. It is easily checked that if $(\mathcal{H}_n)$ is an exchangeable random hierarchy on $\mathbb{N}$ then the comb-partition is exchangeable.

Proposition 10. *Suppose that $(\mathcal{T}, p)$ is a random weighted real tree, that (t_j) an exchangeable sequence directed by p, and that $(\mathcal{H}_n)$ is the exchangeable hierarchy on $\mathbb{N}$ derived from $\mathcal{T}$ and (t_j). On the event that*

- *there is a segment $[[u, v]]$ of $\mathcal{T}$ that is oriented towards the root of $\mathcal{T}$, with v further from the root, meaning that $[[u, v]] \subseteq [[0, v]]$*
- *such that $[[u, v]]$ does not sprout any branches of positive p-mass, meaning that for all x in the support of p with $x \notin [[u, v]]$,*

$$u \in [[0, x]] \text{ implies } v \in [[0, x]]$$

- *such that $p_a([[u, v]]) = 0$ and $p_d([[u, v]]^\circ) > 0$, where p_a and p_d are the atomic and diffuse components of p, respectively, and $[[u, v]]^\circ$ denotes the interior of $[[u, v]]$ for the topology of $\mathcal{T}$*

the set $\{j : t_j \in [[u, v]]^\circ\}$ is a subset of one of the comb-components of $(\mathcal{H}_n)$. Conversely, on the event that distinct positive integers i and j lie in the same comb-component of $(\mathcal{H}_n)$, there is with probability one a segment $[[u, v]]$ of $\mathcal{T}$ having the properties above for which $t_i, t_j \in [[u, v]]^\circ$.

The statement of the proposition is obvious from definitions.

Proposition 11. *Suppose that $(\mathcal{T}, p)$ is a random weighted real tree, that (t_j) an exchangeable sequence directed by p, and that $(\mathcal{H}_n)$ is the exchangeable hierarchy on $\mathbb{N}$ derived from $\mathcal{T}$ and (t_j). On the event that a is an atom of p, for all distinct pairs u, v in the set $B = \{j : t_j = a\}$, $(u \wedge v) = B$, where $(u \wedge v)$ denotes the MRCA of u and v in the hierarchy derived from (t_j) and $\mathcal{T}$. Furthermore, t_j is an atom of p if and only if*

$$0 < \lim_{n\to\infty} \frac{1}{n} \#\{k \in [n] : \alpha_m(j) = \alpha_m(k) \text{ for all } m \geq \max\{j,k\}\},$$

and if t_j is an atom of p then $p(\{t_j\})$ equals this limit above almost surely.

The proof of this proposition is elementary and is therefore omitted.

Remark. The atomic and diffuse parts p_a and p_d of the random measure p of Theorem 4 are "invariants," loosely speaking, of the exchangeable hierarchy $(\mathcal{H}_n)$ of that theorem. More formally, p_a and p_d are measurable functions of p, within the standard abstract setup for random measures [66, Chapter 1].

2.6.2 Tail measurability and open problems

If $(\mathcal{H}_n)$ is a random hierarchy on $\mathbb{N}$, define the *tail sigma field* of $(\mathcal{H}_n)$ as follows,

$$\text{tail}(\mathcal{H}_n) = \bigcap_{n\geq 1} \sigma\left(\mathcal{H}_{n+1}\Big|_{\{n+1\}}, \mathcal{H}_{n+2}\Big|_{\{n+1,n+2\}}, \ldots\right).$$

If $(\mathcal{H}_n)$ is an exchangeable hierarchy, then the pair $(\mathcal{T}, p)$ of Theorem 4 is not tail-measurable for the following reason. Let $\pi_1(x) = (x_1, 0, 0, \ldots)$ for $x \in \ell_1$. Then the image of p under π_1 is the directing measure for the exchangeable spinal variables $(X_{b^{-1}(j)}^{b^{-1}(1)})$ defined by (2.21), where b is the bijection mentioned at the beginning of Section 2.4, but neither these spinal variables nor their directing measure are tail measurable. On the other hand,

Proposition 12. *The distribution of the pair $(\mathcal{T}, p)$ of Theorem 4 is measurable with respect to tail$(\mathcal{H}_n)$.*

Proof. This is a direct consequence of the fact that the bijection b mentioned at the beginning of Section 2.4 can be chosen arbitrarily. □

Instead of proving the assertion that $(\mathcal{T}, p)$ is not tail$(\mathcal{H}_n)$-measurable, we offer the following analogy using exchangeable partitions. Suppose that $\mathscr{U}$ is a random open subset of $[0,1]$ having Lebesgue measure one, and let (U_n) and (V_n) be independent IID sequences of uniform[0,1] random variables, jointly independent of $\mathscr{U}$. Form an exchangeable partition Π of $\mathbb{N}$ by putting i and j in the same block of Π if U_i and U_j fall in the same connected component of $\mathscr{U}$, and index the blocks $\{B_1, B_2, \ldots\}$ of Π by least elements, so that $1 = \min B_1 < \min B_2 < \ldots$. Then the limits

$$P_i = P_i(\Pi) = \lim_{n\to\infty} \frac{1}{n} \# B_i \cap [n]$$

exist almost surely, and P_i is the width of the interval of $\mathscr{U}$ containing $U_{\min B_i}$. Now form another open subset $\mathscr{U}'$ by placing intervals of widths P_i in left-to-right order,

$$\mathscr{U}' := (0, P_1) \cup \bigcup_{n\geq 1} (P_1 + \ldots, P_n, P_1 + \ldots + P_{n+1}).$$

Then $\mathscr{U}'$ is not measurable with respect to the tail of Π,

$$\mathrm{tail}(\Pi) = \bigcap_{n\geq 1} \sigma \left(\Pi \Big|_{\{n,n+1,\ldots\}} \right),$$

because $P_1(\Pi)$ is not measurable with respect to tail(Π), but $P_1(\Pi)$ equals almost surely the length of the connected component of $\mathscr{U}'$ whose left-endpoint is zero. Our weighted tree $(\mathcal{T}, p)$ is very much like the open subset $\mathcal{U}'$.

Continuing this discussion, it is evident that if $(P_1^\downarrow, P_2^\downarrow, \ldots)$ is the sequence of P_i's ranked in nonincreasing order, and

$$\mathscr{U}_{\text{ranked}} := (0, P_1^\downarrow) \cup \bigcup_{n\geq 1} (P_1^\downarrow + \ldots, P_n^\downarrow, P_1^\downarrow + \ldots + P_{n+1}^\downarrow), \tag{2.29}$$

then $\mathscr{U}_{\text{ranked}}$ is measurable with respect to tail(Π). Deterministically reranking the components of $\mathscr{U}'$ in nonincreasing order effectively erases the information contained in $\mathscr{U}'$ but not contained in tail(Π). Obviously, the fact that the resulting order is by decreasing length is immaterial; any deterministic ordering will do.

We may now state the previously-promised stronger version of Kingman's theorem and some open problems.

Theorem 8 ([70]). *Suppose that Π is an exchangeable partition of $\mathbb{N}$ and that the probability space supports a sequence (U_n) of IID uniform[0,1] random variables independent of Π. Then there is a Π-measurable random open subset $\mathscr{U}$ of [0,1] such that if Π' is the partition defined by*

$$\{i \text{ and } j \text{ in same block of } \Pi'\} = \{U_i \text{ and } U_j \text{ in same component of } \mathscr{U}, \text{ or } i = j\}.$$

then there is the equality of joint distributions

$$(\Pi, \mathscr{U}) \stackrel{d}{=} (\Pi', \mathscr{U}) \tag{2.30}$$

Question 1 Define an equivalence relation $\sim$ on laws of weighted real trees, writing $\mathscr{L}(\mathcal{T},p) \sim \mathscr{L}(\mathcal{T}',p')$ if and only if $(\mathcal{H}_n) \stackrel{d}{=} (\mathcal{H}'_n)$ for $(\mathcal{H}_n)$ and $(\mathcal{H}'_n)$ exchangeable hierarchies derived by sampling from $(\mathcal{T},p)$ and $(\mathcal{T}',p')$, respectively.

Is there a nice way of telling whether or not $\mathscr{L}(\mathcal{T},p) \sim \mathscr{L}(\mathcal{T}',p')$? Speaking loosely, it should be possible to *prune away* tree branches of $\mathcal{T}$ that carry no p-mass, and also stretch segments of $\mathcal{T}$ arbitrarily, and not change the equivalence class of $\mathscr{L}(\mathcal{T},p)$. Purely topological considerations are not quite enough to settle this question: suppose that $\mathcal{T}_1$ is the tree [0,1] rooted at 1 and p_1 is Lebesgue measure on [0,1], and suppose that $\mathcal{T}_2$ is is the half line $[0,\infty)$ rooted at 0, and p_2 is the exponential(1) distribution on $\mathcal{T}_2$. Then $\mathcal{T}_1$ and $\mathcal{T}_2$ are not homeomorphic, but $\mathscr{L}(\mathcal{T},p) \sim \mathscr{L}(\mathcal{T}',p')$.

Question 2 Is there a nice way to select from each equivalence class of $\sim$ above a unique representative of that equivalence class? Such a recipe would be akin to reordering component intervals of open subsets of [0,1], as discussed above. By nice we mean *measurable*, and you can pick the sigma fields, but the goal is to have an analogy of the strong version of Kingman's theorem involving an equality of joint distributions, as in (2.30).

Question 3 Repeat the previous questions in the context of Theorem 6, i.e. with hierarchies on [0,1] instead of weighted real trees in ℓ_1.

I thank David Aldous, Steve Evans, and Matthias Winkel for helpful discussion of the contents of this chapter.

Chapter 3

Ewens-Pitman partitions

Let B_1 be the first block – i.e., the block containing 1 – of a random partition Π of $\mathbb{N}$. If the complement $\mathbb{N} \setminus B_1$ of B_1 is infinite one may form a new partition Π' of $\mathbb{N}$ as $\Pi' := \{\{F(b) : b \in B\} : B \in \Pi, B \neq B_1\}$ for F the unique increasing bijection sending $\mathbb{N} \setminus B_1$ to $\mathbb{N}$. This new partition Π' is the *partition derived from* Π *by deleting the first block*, and we say that Π *has the deletion property* if Π' and B_1 are independent. The two-parameter family of Ewens-Pitman partitions, described below, is a family of random partitions with the deletion property.

The proof of the following theorem is the focus of this chapter. The random variables (W_i) appearing below will be defined later; for now we anticipate the well-known result which follows from Kingman's theorem 10 that if Π is an exchangeable partition of $\mathbb{N}$ with first block B_1 then the limit

$$\lim_{n\to\infty} \frac{1}{n} \#\{j \in [n] : j \in B_1\}$$

exists almost surely: W_1 is by definition this limit, and the event $\{W_1 < 1\}$ is almost surely equal to the event that the complement of B_1 is infinite, which is the event that the partition derived from Π by deleting the first block is a well-defined partition of $\mathbb{N}$.

Theorem 9. *If* Π *is an exchangeable partition such that* $\mathbb{P}(W_1 < 1) = 1$ *and* Π *has the deletion property, then one of three cases obtains: (i) with probability one* Π *has exactly two blocks, or (ii) with probability one* Π *is the partition of* $\mathbb{N}$ *into singletons, or (iii)* Π *is a member of the Ewens-Pitman family of exchangeable partitions.*

The proof of Theorem 9 may be found in Section 3.5. The proof may also be found in [46] on which this chapter is based. A variant of the deletion property was studied by Kingman, who characterized the class of partitions Π with the additional property that for Π' derived by deleting the first block of Π, $\Pi \stackrel{d}{=} \Pi'$: the class of such partitions is exactly subset of the Ewens-Pitman(α, θ) family for which the α parameter equals zero [69].

As will be made clear below, the problem of characterizing all partitions with the deletion property is intimately related to the problem of characterizing the class of functions p

- whose domain is the set $\bigcup_{k\geq 2}\mathbb{N}^k$ of integer compositions, and whose range is $[0,1]$
- that satisfy the addition rule

$$p(\lambda_1,\ldots,\lambda_k)=p(\lambda_1+1,\ldots,\lambda_k)+\ldots+p(\lambda_1,\ldots,\lambda_k+1)+p(\lambda_1,\ldots,\lambda_k) \tag{3.1}$$

  and the symmetry rule

$$p(\lambda_1,\ldots,\lambda_k)=p(\lambda_{\sigma(1)},\ldots,\lambda_{\sigma(k)}) \tag{3.2}$$

  for every sequence $(\lambda_1,\ldots,\lambda_k)$ and permutation σ of $\{1,\ldots,k\}$
- that have the normalization condition $p(1)=1$ and the additional property

$$\limsup_{n\to\infty} p(n)=0$$

- and that factorize as

$$p(\lambda_1,\ldots,\lambda_k)=g(\lambda_1+\ldots+\lambda_k,\lambda_1)p'(\lambda_2,\ldots,\lambda_k)$$

  for a function g and another nonnegative symmetric, additive function p'.

This characterization problem has an "algebraic" character, and after proving Theorem 9 by probabilistic arguments we give an independent algebraic proof using this reformulation of the problem. This latter approach can be generalized to solve a problem posed in [57] that arose in the study of homogeneous fragmentation processes, which are partition-of-$\mathbb{N}$-valued Markov processes $(\Pi(t),t\geq 0)$ having stationary "increments" [23]. Disregarding a few details, every such process is characterized by a symmetric, additive function p as above that describes the transition rates for the Markov process $\Pi_n(t)$ that is the restriction of $\Pi(t)$ to $[n]$, except that in this setting p need only be defined on compositions having more than one part – $p(n)$ may not be defined for $n\geq 1$ – and instead of $p(1)=1$ we only require that $p(1,1)<\infty$. Such functions p are called exchangeable partition rate functions (EPRFs), and each EPRF determines the distribution of a homogeneous fragmentation process [20]. The algebraic proof of Theorem 11 can be extended to characterize all EPRFs that factorize as above: this is done in Section 3.8.

The first several sections of this chapter are expository and largely based on [81, 82, 83].

3.1 Exchangeable partitions

A partition of $[n]=\{1,\ldots,n\}$ is a collection $\{B_1,\ldots,B_k\}$ of pairwise disjoint sets called *blocks* whose union is $[n]$, that are conventionally labeled in order of minimal elements, so that $1 = \min B_1 < \ldots < \min B_k$. A sequence $(\pi_n, n \geq 1)$ of partitions is said to be *consistent* if for all $n \geq 1$,

$$\pi_n = \pi_{n+1}\Big|_{[n]} := \{B \cap [n+1] : B \in \pi_{n+1}\}$$

i.e. if π_{n+1} is formed either by adding $n+1$ to a block of π_n or by adding the singleton block $\{n+1\}$ to Π_n. Consistent sequences of partitions are in bijective correspondence with partitions of $\mathbb{N}$: if π is a partition of $\mathbb{N}$ then π corresponds to the sequence (π_n) defined by $\pi_n := \pi\Big|_{[n]}$. Partitions π of $\mathbb{N}$ and consistent sequences of partitions (π_n) are often spoken of interchangeably, and because of this correspondence there is no harm in doing this.

A random partition Π_n of $[n]$ is said to be *exchangeable* if for every permutation σ of $[n]$ there is the equality in distribution

$$\Pi_n \stackrel{d}{=} \sigma(\Pi_n) := \{\{\sigma(b) : b \in B\} : B \in \Pi_n\}, \tag{3.3}$$

i.e. if relabeling the contents of the blocks of Π_n by σ produces a partition equal in distribution to Π_n. Likewise, a random partition (Π_n) of $\mathbb{N}$ is exchangeable if (3.3) holds for all $n \geq 1$.

For positive integers n a *composition of* n is a sequence $\lambda = (\lambda_1, \ldots, \lambda_k)$ of positive integers with sum n. The length $k = k_\lambda$ of λ may be regarded as a function of λ. If (Π_n) is an exchangeable partition of $\mathbb{N}$ then for all $n \geq 1$ and any partition π_n of $[n]$ there is the equality

$$\mathbb{P}(\Pi_n = \pi_n) = \mathbb{P}(\Pi_n = \sigma(\pi_n))$$

indeed, this is an equivalent form of (3.3). Consequently, the probability $\mathbb{P}(\Pi_n = \pi_n)$ only depends on π_n through its block sizes. The *exchangeable partition probability function* (EPPF) associated to (Π_n) is the [0,1]-valued function p of integer compositions defined by

$$p(\lambda_1, \ldots, \lambda_k) := \mathbb{P}(\Pi_n = \{B_1, \ldots, B_k\})$$

for any (and every) partition $\{B_1, \ldots, B_k\}$ of $[n]$ into blocks of size $\lambda_i = \#B_i$, $i = 1, \ldots, k$, $n \geq 1$ [82]. The EPPF of an exchangeable partition is *symmetric, additive, and normalized* as discussed in the introductory remarks at the beginning of this chapter around (3.2) and (3.1). The symmetry property of p follows from exchangeability of the partition Π_n, additivity of p follows from consistency of (Π_n) as n varies, and the normalization property follows from the fact that there is only one partition of the set [1].

Conversely, any [0,1]-valued function p of integer compositions with these three properties is the EPPF of exactly one exchangeable partition.

Proposition 13 (Pitman [82]). *If p is a [0,1]-valued function of integer compositions that is symmetric, additive, and normalized as above, then there is an exchangeable partition (Π_n) of $\mathbb{N}$ for which p is the EPPF of (Π_n).*

Proof. We construct the sequence (Π_n) sequentially using p and external randomness in the form of an IID sequence (U_i) of uniform random variables. Let $\Pi_1 = \{\{1\}\}$. Now suppose that Π_n has been defined for some n and that $\Pi_n = \{B_1, \ldots, B_k\}$ for some partition of $[n]$, and it remains to construct Π_{n+1}. For $j \in [k]$, define x_j by

$$x_j := p(\#B_1, \ldots, \#B_j + 1, \#B_k)/p(\#B_1, \ldots, \#B_k),$$

and set

$$x_{k+1} = p(\#B_1, \ldots, \#B_k, 1)/p(\#B_1, \ldots, \#B_k) = 1 - x_1 - \ldots - x_k.$$

Then let $A_{n+1} = \inf\{j \in [k+1] : x_1 + \ldots + x_j > U_{n+1}\}$, so that $\mathbb{P}(A_{n+1} = j \mid \Pi_n) = x_j$, and form Π_{n+1} by

- adding element $n+1$ block $B_{A(n+1)}$ if $A_{n+1} < k+1$
- and adding a new singleton block $\{n+1\}$ to Π_n if $A_{n+1} = k+1$.

It is easily seen that for the sequence of partitions so constructed, $\mathbb{P}(\Pi_n = \{B_1, \ldots, B_k\}) = p(\#B_1, \ldots, \#B_k)$ for any partition $\{B_1, \ldots, B_k\}$ of $[n]$ and all $n \geq 1$. Thus Π_n is exchangeable. The sequence (Π_n) is obviously consistent, and the claim follows.

□

3.1.1 Constructing exchangeable partitions

Let $\nabla := \{(s_i, i \geq 1) : s_1 \geq s_2 \geq \ldots \geq 0, \sum_{i \geq 1} s_i \leq 1\}$ be the set of nonnegative nonincreasing sequences of real numbers with sum at most one, and let $\mathbf{S} = (S_1, S_2, \ldots,)$ be a *random discrete distribution*, which is to say a random element of ∇ [73]. Furthermore let ν be the distribution of $\mathbf{S}$, and let (Z_i) be a sequence of integer-valued random variables conditionally independent given $\mathbf{S}$ with

$$\mathbb{P}(Z_i = j \mid \mathbf{S} = \mathbf{s}) = \begin{cases} s_j & \text{if } j > 0 \\ 1 - \sum_{j \geq 1} & \text{if } j = 0 \end{cases} \qquad (i \geq 1, j \geq 0). \tag{3.4}$$

Then (Z_j) is exchangeable, and by slight abuse of terminology of Aldous [11] we say that (Z_i) is *directed by* $\mathbf{S}$. A *PB(ν)-partition* is any partition equal in distribution to a partition derived as follows from exchangeable variables (Z_i) directed by $\mathbf{S}$: place natural numbers i and j in the same block if and only if $Z_i = Z_j \neq 0$. If Π is derived in this manner, all integers j for which $Z_j = 0$ belong to singleton blocks $\{j\}$ of Π. A *paintbox partition* is any random partition of $\mathbb{N}$ that is a PB(ν)-partition for some probability distribution ν on ∇.

If $\mathcal{U}$ is a random open subset of $[0,1]$ and $(U_i, i \geq 1)$ an IID sequence of uniform[0,1] random variables independent of $\mathcal{U}$ then one may form an exchangeable partition by stipulating that natural numbers i and j be in the same block if and only if U_i and U_j fall in the same connected component of $\mathcal{U}$. Less formally, one may regard [0,1] as a paintbox and the connected components of $\mathcal{U}$ as wells containing different colors of paint. Each integer i is to be dipped in the paint well in which U_i lies, and after all integers have been painted in this manner one derives a partition of $\mathbb{N}$ by color [81]. The two notions of paintbox partition are easily seen to be equivalent, for given a random open subset of $\mathcal{U}$ one may arrange the lengths of the connected components of $\mathcal{U}$ in nonincreasing order to obtain a random discrete distribution $\mathbf{S}$, and conversely given a random discrete distribution $\mathbf{S}$ one may form a random open subset by placing intervals of lengths S_i in left-to-right order in [0,1].

For $\mathbf{s} \in \nabla$ define a function of integer compositions $p_{\mathbf{s}}$ by

$$p_{\mathbf{s}}(\lambda_1, \ldots, \lambda_k) = \sum_{I: I \subseteq \mathcal{I}} (1 - s_1 - s_2 \ldots)^{\#\mathcal{I} - \#I} \sum_{\substack{f: f \text{ is injective} \\ f: I \cup \mathcal{M} \mapsto \mathbb{N}}} \prod_{i \in I \cup \mathcal{I}'} s_{f(i)}^{\lambda_i} \tag{3.5}$$

where $\mathcal{I}$ denotes the set $\{j \in [k] : \lambda_j = 1\}$ of indices of singletons, $\mathcal{M} = \{j \in [k] : \lambda_j > 1\}$ denotes the set of indices of nonsingletons, and the second sum is over all injective functions with the given domain and range. If $\sum_{i \geq 1} s_i = 1$, equation (3.5) simplifies:

$$p_{\mathbf{s}}(\lambda_1, \ldots, \lambda_k) = \sum_{\substack{f: f \text{ is injective} \\ f: [k] \mapsto \mathbb{N}}} \prod_{i \in [k]} s_{f(i)}^{\lambda_i} \qquad \text{if } \sum_{i \geq 1} s_i = 1. \tag{3.6}$$

The relevance of the function $p_{\mathbf{s}}$ is apparent from the following proposition:

Proposition 14. *If (Π_n) is a PB(ν) partition derived from a random discrete distribution $\mathbf{S}$ with distribution ν and $\pi_n = \{B_1, \ldots, B_k\}$ is a partition of $[n]$ with block sizes $(\#B_1, \ldots, \#B_k) = (\lambda_1, \ldots, \lambda_k) = \lambda$, then*

$$\mathbb{P}(\Pi = \pi_n \mid \mathbf{S}) = p_{\mathbf{S}}(\lambda) \tag{3.7}$$

and

$$p_\nu(\lambda_1, \ldots, \lambda_k) := \int_\nabla p_{\mathbf{s}}(\lambda_1, \ldots, \lambda_k)\, \nu(d\mathbf{s}) = \mathbb{P}(\Pi_n = \pi_n). \tag{3.8}$$

The proof is elementary and obvious from the definition of the random discrete distribution construction of an exchangeable partition. It may have been first noticed by Kingman [69].

3.1.2 Size-biasing

Suppose again that $\mathbf{S} = (S_1, S_2, \ldots)$ is a random discrete distribution and that the family (Z_i) of nonnegative integer-valued random variables is directed by $\mathbf{S}$ as in Section 3.1.1. Let $J_1 = 1$ and for $i \geq 2$ let $J_i = \inf\{n \geq J_{i-1} : X_n \notin \{X_{J(1)}, \ldots X_{J(n-1)}\} \text{ or } X_n = 0\}$. Then the sequence $(Y_1, Y_2, \ldots)$ defined by

$$Y_i = \begin{cases} S_{J(i)} \text{ if } J_i < \infty \\ 0 \text{ if } J_i = \infty \end{cases}$$

is known as a *size-biased permutation of* $\mathbf{S}$ [34]. The distribution of (Y_i) can be described informally in terms of an urn: put black marbles of weights $S_1, S_2, \dots$ in an urn and add an additional white marble having weight $1 - \sum_{i \geq 1} S_i$. Then recursively draw marbles from the urn with probability proportional to weight, and

- when a black ball is drawn, record its weight, and do not replace the marble in the urn
- when the white marble is drawn, record zero, and replace the white marble in the urn.

With the convention that the white marble is always the next marble to be drawn if it is the only marble remaining in the urn, this process can be continued indefinitely, and the distribution of (Y_i) is the same as the distribution of the sequence of numbers recorded by this urn process. More formally, if the positive members of $\mathbf{S}$ are strictly decreasing, the conditional distribution of (Y_i) given $\mathbf{S} = \mathbf{s}$ is easily seen to be

$$\mathbb{P}(Y_1 = Q_{i(1)}, \dots, Y_k = Q_{i(k)} \mid \mathbf{S}) = \prod_{j=1}^{k} S_{i(j)} \left(\prod_{j=2}^{k} (1 - Q_{i(1)} - \dots - Q_{i(j)}) \right)^{-1} 1(E) \quad a.s. \tag{3.9}$$

where for brevity we have written S_0 in place of $1 - \sum_{i \geq 1} S_i$ and where (Q_i) is defined by

$$Q_i := \begin{cases} S_i & \text{if } i \geq 1 \\ 0 & \text{if } i = 0 \end{cases},$$

and where E is the event that the nonzero members of the sequence $(i(1), \dots, i(j))$ are distinct positive integers. If $\mathbf{S}$ contains nonzero ties with positive probability, i.e. if $\mathbb{P}(S_i = S_{i+1} > 0) > 0$, then (3.9) must be modified by introducing a combinatorial factor. *Size-biasedness* is a distributional property: a sequence $(Y_i', i \geq 1)$ of nonnegative random variables with sum almost surely no greater than one is said to be a size-biased permutation of the sequence $\mathbf{S}$ derived by ranking (Y_i') in nonincreasing order, if the the joint distribution of $((Y_i'), \mathbf{S})$ is the same as the distribution of $((Y_i), \mathbf{S})$ for an exchangeable sequence (Y_i) directed by the discrete distribution $\mathbf{S}$.

For more developments of size-biasing see [51] and references therein.

3.2 Kingman's theorem

Kingman's proved a de Finetti-type characterization of exchangeable partitions:

Theorem 10 (Kingman [70]). *If* Π *is an exchangeable partition of* $\mathbb{N}$ *then* Π *is equal in distribution to a partition directed by a random discrete distribution.*

Proof. (Aldous [11]) . Without loss of generality let (U_i) be a sequence of IID uniform random variables independent of Π. For $j \in \mathbb{N}$ let L_j denote the minimal element of the block of Π containing j. Thus if $\Pi_5 = \{\{1,4\},\{2,5\},\{3\}\}$ then $(L_1, L_2, L_3, L_4, L_5) = (1,2,3,1,2)$. Next define a sequence (X_j) by $X_j = U_{L(j)}$. It is easily seen that because Π is exchangeable, (X_j) is exchangeable as well, and furthermore Π is equal to the partition derived from the sequence (X_i) by putting natural numbers i and j in the same block if and only if $X_i = X_j$.

Let μ be the directing random measure of (X_i), and let $\mathbf{S} = \mathbf{S}(\mu)$ be the sequence of atom sizes of μ ranked in nonincreasing order. By de Finetti's theorem the conditional distribution of (X_i) given $\mu = m$ is the distribution of a sequence of IID variables with common law m, and such a sequence can be coupled to a sequence of independent random variables (Z_i) with

$$\mathbb{P}(Z_i = k) = \begin{cases} s_k(\mu) & \text{if } k \geq 1 \\ 1 - \sum_{i\geq 1} s_i(\mu) & \text{if } k = 0 \end{cases}$$

in an obvious way so that the partition Π' derived by putting i and j in the same block if and only if $Z_i = Z_j \neq 0$ is equal to the partition derived by putting i and j in the same block if and only if $X_i = X_j$. Thus Π is equal in distribution to a partition Π' directed by a discrete distribution.

□

Corollary 2. *If $\Pi = \{B_1, B_2, \ldots\}$ is an exchangeable partition of $\mathbb{N}$, then for every block B_i of Π_n the limit*

$$P_i := \lim_{n\to\infty} \frac{1}{n} \# B_i \cap [n] \tag{3.10}$$

exists almost surely and equals zero if and only if B_i is a singleton block of Π. With the convention that $P_i = 0$ if Π has fewer than i blocks, the sequence (P_i) is a size-biased permutation of the sequence $\mathbf{S}$ derived by ranking (P_i) in nonincreasing order. Finally, if $\mathcal{C}$ denotes the class of singletons, $\mathcal{C} := \{j \in \mathbb{N} : \{j\} \in \mathbb{N}\}$, then $\mathcal{C}$ is almost surely empty if $\sum_{i\geq 1} P_i = 1$, and

$$\lim_{n\to\infty} \frac{1}{n} \# \mathcal{C} \cap [n] = 1 - \sum_{i\geq 1} P_i$$

holds almost surely.

The statement of the corollary is obvious for partitions directed by discrete distributions and so follows from Kingman's theorem [81]. The limits (P_i) are known as the *limit frequencies of* Π, or sometimes *limit frequencies of* Π *in order of discovery.* By convention $P_i := 0$ if Π has fewer than i blocks. For later reference we define the *residual limit frequencies* (W_i): if Π is an exchangeable partition with limit frequencies (P_i) then $W_1 = P_1$ and

$$W_i := \begin{cases} P_i/(1 - P_1 - \ldots - P_{i-1}) & \text{if } P_1 + \ldots + P_{i-1} < 1 \\ 0 & \text{if } P_1 + \ldots + P_{i-1} = 1 \end{cases}. \tag{3.11}$$

Proposition 15 (Kingman, Pitman [81, 82]). *If Π is an exchangeable partition of $\mathbb{N}$ with EPPF p and $\mathbf{S} = (S_i)$ denotes the random sequence derived from the limit frequencies (P_i) of Π by ranking these limits in nonincreasing order, then for every partition $\{B_1, \ldots, B_k\}$ of $[n]$ with $\lambda_i := \#B_i$, for $i \in [k]$, there are the following almost sure equalities,*

$$\mathbb{P}(\Pi_n = \{B_1, \ldots, B_k\} \mid \mathbf{S}) = p_{\mathbf{S}}(\lambda_1, \ldots, \lambda_k), \tag{3.12}$$

and

$$\mathbb{P}\left(\Pi_n = \{B_1, \ldots, B_k\} \mid (P_i)\right) = \prod_{i=1}^{k} P_i^{\lambda_i - 1} R_i \tag{3.13}$$

where $R_1 := 1$ and $R_i := 1 - P_1 - \ldots - P_{i-1}$ for $i \geq 1$. Additionally,

$$\mathbb{P}(\Pi_n = \{B_1, \ldots, B_k\} \mid (W_i)) = \prod_{i=1}^{k} W_i^{\lambda_i - 1} \overline{W}_i^{\Lambda_{i+1}} \tag{3.14}$$

for $\overline{W}_i := 1 - W_i$ and $\Lambda_i := \lambda_i + \ldots + \lambda_k$.

Proof. According to Kingman's theorem, without loss of generality it may be assumed that Π is a paintbox partition derived from a sequence of nonnegative integer-valued random variables (Z_i) directed by the discrete distribution $\mathbf{S}$ by putting i and j in the same block if and only if $Z_i = Z_j \neq 0$. Equation (3.12) then follows from Proposition 14. For an idea of how (3.13) may be obtained, let $J_i := \min B_i$ for $\{B_1, B_2, \ldots\}$ the sequence of blocks of the partition Π of $\mathbb{N}$, and observe that $P_i = S_{J_i}$ unless $J_i = 0$ in which case $P_i = 0$. Next note that for $m \geq k$,

$$\frac{\mathbb{P}(\Pi_n = \{B_1, \ldots, B_k\}, J_1 = j_1, \ldots, J_m = j_m \mid \mathbf{S})}{\mathbb{P}(J_1 = j_1, \ldots, J_m = j_m \mid \mathbf{S})} = \frac{\prod_{i=1}^{k} S_{j_i}^{n_i}}{\prod_{i=1}^{k} S_{j_i}/(1 - P_1 - \ldots - P_{i-1})},$$

holds almost surely, where for brevity we write S_0 in place of $1 - \sum_{i \geq 1} S_i$. The left-hand side of this display equals $\mathbb{P}(\Pi_n = \{B_1, \ldots, B_k\} \mid J_1 = j_1, \ldots, J_m = j_m, \mathbf{S})$. On taking expectation with respect to the sigma field generated by $(P_1, \ldots, P_m)$ we obtain

$$\mathbb{P}(\Pi_n = \{B_1, \ldots, B_k\} \mid P_1, \ldots P_m) = \prod_{i=1}^{k} P_i^{\lambda_i - 1} R_i,$$

now send m upwards to the number of blocks of Π to obtain the desired result. For more details see [82]. Finally, (3.14) is derived from (3.13) by the change-of-variables (3.11). $\square$

3.3 The two-parameter family

It was shown in [82] that for each pair of real parameters (α, θ) with

$$0 \leq \alpha < 1, \; \theta > -\alpha \tag{3.15}$$

the formula

$$p_{\alpha,\theta}(\lambda) := \frac{\prod_{i=1}^{k-1}(\theta + i\alpha)}{(\theta+1)_{n-1}} \prod_{j=1}^{k} (1-\alpha)_{\lambda_j - 1} \tag{3.16}$$

where $k = k_\lambda$, $n = n_\lambda$, and

$$(x)_n := x(x+1)\dots(x+n-1) = \frac{\Gamma(x+n)}{\Gamma(x)}$$

is a rising factorial, defines the EPPF of an exchangeable random partition of positive integers whose limit frequencies (P_i) in order of appearance as defined by (3.10) admit the *stick-breaking representation*

$$P_i = W_i \prod_{j=1}^{i-1} (1 - W_j) \tag{3.17}$$

for residual limits (W_j) such that

$$W_1, W_2, \dots \text{ are mutually independent} \tag{3.18}$$

with W_j having the Beta$(1-\alpha, \theta + j\alpha)$ distribution.

Formula 3.16 also defines an EPPF for (α, θ) in the range

$$\alpha < 0, \; \theta = -M\alpha \text{ for some } M \in \mathbb{N}, \tag{3.19}$$

in which case the stick-breaking representation (3.17) makes sense for $1 \leq k \leq M$, if the last factor W_M, which has "Beta$(1-\alpha, 0)$" distribution for these parameter values, is understood to be equal to 1 almost surely. The frequencies $(P_1, \dots, P_M)$ in this case are a size-biased random permutation of $(Q_1, \dots, Q_M)$ with the symmetric Dirichlet distribution with M parameters equal to $\kappa := -\alpha > 0$. It is well known that the Q_i can be constructed as $Q_i = \gamma_\kappa^{(i)}/\Sigma, 1 \leq i \leq M$, where $\Sigma = \sum_{i=1}^{M} \gamma_\kappa^{(i)}$ and the $\gamma_\kappa^{(i)}$ are independent and identically distributed copies of a gamma variable γ_κ with density

$$\mathbb{P}(\gamma_\kappa \in \mathrm{d}x) = \Gamma(\kappa)^{-1} x^{\kappa-1} e^{-x}\, dx \qquad (x > 0). \tag{3.20}$$

As shown by Kingman [69], the $(0, \theta)$ EPPF (3.16) for $\alpha = 0, \theta > 0$ arises in the limit of random sampling from such symmetric Dirichlet frequencies as $\kappa = -\alpha \downarrow 0$ and $M \uparrow \infty$ with $\nu M = \theta$ held fixed. In this case, the distribution of the partition Π_n is that determined by the Ewens sampling formula with parameter θ, the residual fractions W_i in the stick-breaking representation are independent and have identical Beta$(1,\theta)$ distributions, and the ranked frequencies S_i can be obtained by normalization of the jumps of a gamma process with stationary independent increments $(\gamma_\kappa, 0 \leq \kappa \leq \theta)$. Perman, Pitman and Yor [80] gave extensions of this description to the case $0 < \alpha < 1$ when the distribution of ranked frequencies can be derived from the jumps of a stable subordinator of index α. See also [84, 80, 87, 81] for further discussion and applications to the description of ranked lengths of excursion intervals of Brownian motion and Bessel processes.

In the limit case when $\kappa = -\alpha \to \infty$ and $\theta = M\kappa \to \infty$, for a fixed positive integer M, the EPPF (3.16) converges to

$$p_M(\lambda) := \frac{M(M-1)\cdots(M-k+1)}{M^n}\,, \tag{3.21}$$

corresponding to sampling from M equal frequencies

$$P_1 = P_2 = \cdots = P_M = 1/M$$

as in the classical coupon collector's problem with some fixed number M of equally frequent types of coupon. We refer to the collection of partition structures defined by (3.16) for the parameter ranges (3.15) and (3.19), as well as the limit cases (3.21), as the *extended two-parameter family*.

The partition $\mathbf{0}$ of $\mathbb{N}$ into singletons and the partition $\mathbf{1}$ of $\mathbb{N}$ into a single block both belong to the closure of the two-parameter family. As noticed by Kerov [68], a mixture of these two trivial partitions with mixing proportions t and $1-t$ also belongs to the closure, as is seen from (3.16) by letting $\alpha \to 1$ and $\theta \to -1$ in such a way that $(1-\alpha)/(\theta+1) \to t$ and $(\theta+\alpha)/(\theta+1) \to 1-t$.

3.4 The deletion property

Recall from the first paragraph of this chapter the definition of *the partition* Π' *derived by deleting the first block* B_1 *of a partition* Π *of* $\mathbb{N}$.

Definition 10. *A random partition* $\Pi = \{B_1, B_2, \ldots\}$ *of* $\mathbb{N}$ has the deletion property *if* $\mathbb{N} \setminus B_1$ *is almost surely an infinite set and if* B_1 *and* Π' *are independent, for* Π' *the partition of* $\mathbb{N}$ *derived by deleting the first block of* Π.

We will not discuss the possibility of *deleting the first block* for partitions where the complement of the first block is finite or empty. However, it follows from Kingman's theorem that if Π is an exchangeable partition of $\mathbb{N}$ then there is the almost sure equality of events

$$\{\mathbb{N} \setminus B_1 \text{ is infinite}\} = \{W_1 < 1\} = \{\Pi \neq \{\mathbb{N}\}\}$$

for W_1 the first residual limit frequency of Π. For this reason, throughout the rest of this chapter it will often be assumed that $\mathbb{P}(W_1 < 1) = 1$.

Proposition 16 (Gnedin, Haulk and Pitman [46]). *Suppose Π is an exchangeable partition of $\mathbb{N}$ with residual limits (W_i), suppose $P(W_1 < 1) = 1$ and let Π' be the partition of $\mathbb{N}$ derived by deleting the first block of Π. Let (P_i) and (P_i') [resp., (W_i) and (W_i')] be the limit frequencies [resp., residual limit frequencies] of Π and Π', respectively. Then Π' is an exchangeable partition of $\mathbb{N}$, and for $i \geq 1$, $P_i' = P_{i+1}/(1-P_1)$ and $W_i' = W_{i+1}$ almost surely.*

Proof. Exchangeability of Π' follows directly from exchangeability of Π. The assertions about the limits and residual limits are likewise straightforward to check. $\square$

The following proposition provides some alternative descriptions of the deletion property under the additional assumption of exchangeability.

Proposition 17 (Gnedin, Haulk and Pitman [46]). *Let Π be an exchangeable partition of $\mathbb{N}$ with residual limits (W_i), and suppose that $\mathbb{P}(W_1 < 1) = 1$. Let Π' be the partition derived by deleting the first block B_1 of Π. Then the following conditions are equivalent:*

(i) Π *has the deletion property:* B_1 *and* Π' *are independent.*

(ii) W_1 *is independent of* $(W_2, W_3, \ldots)$.

(iii) There is a pair of functions g and p' such that for every every integer composition λ of having at least two parts, the EPPF p factorizes as follows,

$$p(\lambda_1, \ldots, \lambda_k) = g(n_\lambda, \lambda_1) p'(\lambda_2, \ldots, \lambda_k) \tag{3.22}$$

for $n_\lambda := \lambda_1 + \ldots + \lambda_k$, for a function $g : \{(n,m) : 1 \leq m \leq n < \infty\} \mapsto [0,1]$ that is additive*, meaning*

$$g(n,m) = g(n+1,m) + g(n+1,m+1)$$

for all $1 \leq m \leq n$, such that $g(1,1) = 1$, and p' a [0,1]-valued function of compositions that is symmetric, additive, and normalized, that is, p' that is an EPPF.

Finally, if any of these conditions hold, then p' and g appearing in (iii) are uniquely determined: p' is the EPPF of Π', and $g(n,m) = \mathbb{E}[W_1^{n-1}\overline{W}_1^{n-m}]$ for $1 \leq m \leq n < \infty$.

Proof. To see that (i) implies (ii), note that W_1 is a function of B_1 and $(W_2, W_3, \ldots)$ is a function of Π' by Proposition 16. To see that (ii) implies (iii), take expectation in (3.14) and use independence of W_1 and $(W_2, W_3, \ldots)$ to factorize the expectation of the product,

$$p(\lambda) = \mathbb{E}[W_1^{\lambda_1 - 1}\overline{W}_1^{n-\lambda_1}]\, \mathbb{E}\left[\prod_{i=2}^{k} W_i^{\lambda_i - 1}\overline{W}_i^{\Lambda_i + 1}\right].$$

Then set $g(n,m) = \mathbb{E}(W_1^{m-1}\overline{W_1}^{n-m})$ and

$$p'(\lambda_2, \ldots, \lambda_k) = \mathbb{E}\left[\prod_{i=2}^{k} W_i^{\lambda_i - 1}\overline{W}_i^{\Lambda_i + 1}\right].$$

Additivity of g follows from the fact that $\mathbb{E}[W_1^{m-1}\overline{W}_1^{n-m}] = \mathbb{E}[W_1^{m-1}\overline{W}_1^{n-m}(W_1 + \overline{W}_1)]$, and p' is an EPPF by Proposition 16. To see that (iii) implies (i), without loss of generality suppose that $\Pi = \{B_1, B_2, \ldots\}$ is constructed using the EPPF p and external randomness in the form of an IID sequence of uniform random variables (U_i) as described in the proof of Proposition 13, and let (A_n) be the random variables that appear in this construction and that have the property that the equality of events

$$\{A_n = j\} = \{n \in B_j\}$$

holds almost surely. Furthermore, let Π'' be an exchangeable partition of $\mathbb{N}$ constructed in this same manner from the EPPF p' and a different sequence (V_i) of IID uniform random variables, so that Π and Π'' are independent. Now let π_n be a partition of $[n]$ with block sizes $(\lambda_1, \ldots, \lambda_k) = \lambda$ and suppose that $p(\lambda) > 0$. The equality

$$\frac{p(\lambda_1 + 1, \lambda_2, \ldots, \lambda_k)}{p(\lambda)} = \frac{g(n+1, \lambda_1 + 1)}{g(n, \lambda_1)}$$

shows that

$$\mathbb{P}(A_{n+1} = 1 \mid \Pi_n = \pi_n) = \frac{g(n+1, \lambda_1 + 1)}{g(n, \lambda)},$$

and thus, by additivity, $\mathbb{P}(A_{n+1} \neq 1 \mid \Pi_n = \pi_n) = g(n+1, \lambda_1)g(n+1, \lambda_1)^{-1}$. Therefore

$$\mathbb{P}(A_{n+1} = j \mid \Pi_n = \pi_n, A_{n+1} \neq 1) = \begin{cases} \frac{p'(\lambda_2, \ldots, \lambda_j, \ldots, \lambda_k)}{p'(\lambda_2, \ldots, \lambda_k)} & \text{if } 2 \leq j \leq k \\ \frac{p'(\lambda_2, \ldots, \lambda_k, 1)}{p'(\lambda_2, \ldots, \lambda_k)} & \text{if } j = k+1 \end{cases}, \tag{3.23}$$

and it can then be easily checked that there is the distributional equality

$$\Pi \stackrel{d}{=} \Pi_0 \stackrel{def}{=} \{B_1\} \cup \{\{G(j) : j \in B\} : B \in \Pi''\}$$

where G is the increasing bijection from $\mathbb{N} \setminus B_1$ to $\mathbb{N}$; that is, Π is equal in distribution to a partition Π_0 derived by first "breaking off" the block B_1 from $\mathbb{N}$ and then partitioning the remainder $\mathbb{N} \setminus B_1$ according Π''. Since Π'' and B_1 are independent, and Π'' equals the partition Π_0' derived from Π_0 by deleting the first block, this shows that (iii) implies (i), and also shows that Π' has EPPF p'. Since p determines the law of Π, and thereby determines the law of Π', p thus determines p' uniquely, and so by the equality $p = gp'$ determines g uniquely as well. □

Let us now restate Theorem 9 more precisely.

Theorem 11 (Gnedin, Haulk, Pitman [46]). *If Π is an exchangeable partition with residual limits (W_i) such that $\mathbb{P}(W_1 < 1) = 1$, and Π has the deletion property then either*

(i) with probability one Π is the trivial partition of $\mathbb{N}$ into singletons, or

(ii) Π *is an exchangeable partition that almost surely has exactly two blocks, or*

(iii) Π *is a member of the extended two-parameter family of Ewens Pitman partitions with EPPF* p *that either has the form specified by (3.16) for parameters in the range (3.15) or in the range* $\{\alpha < 0, \theta = M\alpha$ *for some integer* $M \geq 2\}$*, or* p *has the form (3.21) for some integer* $M \geq 2$*.*

Conversely, if any one of these three cases obtains then Π *is an exchangeable partition,* $\mathbb{P}(W_1 < 1)$ *and* Π *has the deletion property.*

The "converse" part of theorem 11 is known from previous work [82]; we will not discuss it in this chapter.

Corollary 3. *If* Π *is an exchangeable partition with residual limits* (W_i) *with* $\mathbb{P}(W_1 < 1)$ *and* W_1 *is independent of* $(W_2, W_3, \ldots)$*, then the distribution of* W_1 *determines that of* $(W_2, W_3, \ldots)$*, and the random variables* $(W_i, i \geq 1)$ *are jointly independent.*

Proof. The symmetry condition $p(r+1, s+1) = p(s+1, r+1)$ and the moment formula (3.14) in conjunction with Proposition 17 give

$$\mathbb{E}(W_1^r \overline{W}_1^{s+1})\mathbb{E}(W_2^s) = \mathbb{E}(W_1^s \overline{W}_1^{r+1})\mathbb{E}(W_2^r) \tag{3.24}$$

for non-negative integers r and s. Setting $r = 0$, this expresses moments of W_2 in terms of the moments of W_1. So the distribution of W_1 determines that of W_2. In fact, more is true: if $\mathbb{P}(0 < W_1 < 1) = 1$ then as as is shown in Lemma 12 of [82], first two moments of W_1 determine all moments of W_1 and of W_2, and one of three cases obtains:

- $W_2 = 1$ almost surely,
- $(P_1, P_2) = (\frac{1}{M}, \frac{1}{M})$ almost surely for some positive integer M
- W_1 and W_2 have nondegenerate beta distributions.

From the distribution of W_1 it can be seen whether $\mathbb{P}(0 < W_1 < 1) = 1$, and if this holds then the law of W_1 determines which of these three possibilities occurs. In the first two cases the distribution of W_1 obviously determines that of $(W_2, W_3, \ldots)$. Likewise, in the third case, W_1 has Beta$(1-\alpha, \theta+\alpha)$ distribution for some α, θ, and then the remaining residual limits have distributions determined by these parameters as a consequence of Theorem 11. If on the other hand $\mathbb{P}(0 < W_1 < 1) < 1$, then we must be in case (i) of Theorem 11 and so $(W_2, W_3, \ldots) = (1, 0, 0, \ldots)$ almost surely and so is trivially determined by the law of W_1. In any case, independence of (W_i) follows from Theorem 11

□

3.5 Proof of Theorem 11

According to Kingman's theorem we may assume without loss of generality that Π is a PB(ν) partition for some distribution ν on ∇, so that $p = p_\nu$ is the EPPF of Π. First assume that

$$p_\nu(2,2,1) = 0. \tag{3.25}$$

It follows that $\nu(\{\mathbf{s} : s_2 > 0, s_1 + s_2 < 1\}) = 0$. We claim that $\nu\{0 < s_1 < 1, s_2 = 0\} = 0$. Indeed, supposing otherwise we have by (17)

$$0 < p_\nu(2,1,1,1) = g(5,2)p'(1,1,1)$$

and

$$0 < p_\nu(2,1,1) = g(4,1)p'(2,1)$$

for some EPPF p', which shows that

$$0 < g(5,2)p'(2,1) = p_\nu(2,2,1),$$

a contradiction that proves the claim. It follows that ν concentrates on $\{s_1 = 0\} \cup \{0 < s_2 = 1 - s_1\}$. We claim furthermore that if $\nu(s_1 = 0) > 0$ then $\nu(s_1 > 0) = 0$. Supposing otherwise we have

$$0 < p_\nu(m,n) = g(m+n,m)p'(n)$$

and

$$0 < p_\nu(1^{n+1}) = g(n,1)p'(1^n)$$

for all $m, n \geq 2$ and therefore

$$0 < g(m+n,m)p'(1^n) = p_\nu(m,1^n)$$

for all $m, n \geq 2$, contradicting $\nu(\{s_2 > 0, s_1 + s_2 < 1\}) = \nu\{0 < s_1 < 1, s_2 = 0\} = 0$ and thus proving the claim. Therefore either $s_1 = 0$ ν-a.e. and we are in case (i) of Theorem 11, or $0 < s_2 = 1 - s_1$ ν-a.e., and we are in case (ii) of Theorem 11.

We may now suppose that the EPPF p of Π has the property that

$$p(2,2,1) > 0. \tag{3.26}$$

As mentioned in the proof of Corollary 3, it was shown in Lemma 12 of [82] that if the residual limits (W_i) are independent and $\mathbb{P}(0 < W_1 < 1) = 1$ then

- $W_2 = 1$ almost surely
- W_1 is a degenerate random variable, almost surely equal to $1/M$ for some positive integer M

- W_1 and W_2 have nondegenerate beta distributions.

In fact, the proof of this Lemma of [82] goes through if $\mathbb{P}(0 < W_1 < 1) = 1$ is weakened to the conjunction $\mathbb{P}(0 \leq W_1 < 1) = 1$ and $\mathbb{P}(W_1 > 0) > 0$, so that this trichotomy is holds in the present context. However, the regularity condition (3.26) reduces the situation either to the case with $M > 2$ equal frequencies with sum 1, or to the case where W_1 has a beta distribution, and hence so does W_2, by consideration of (3.24). There is nothing more to discuss in the first case, so we assume for the rest of this section that

$$\text{each of } W_1 \text{ and } W_2 \text{ has a non-degenerate beta distribution, with possibly different parameters.} \tag{3.27}$$

Recall that

$$P_1 = W_1 \text{ and } P_2 = (1 - W_1)W_2.$$

As observed in [83],

> the conditional distribution of $(P_3, P_4, \ldots)$ given P_1 and P_2 depends symmetrically on P_1 and P_2.

This can be seen from Kingman's paintbox representation, which implies that conditionally given $S_1, S_2, \ldots$, as well as P_1 and P_2, the sequence $(P_3, P_4, \ldots)$ is derived by a process of random sampling from the frequencies (S_i) with the terms P_1 and P_2 deleted. No matter what (S_i) this process depends symmetrically on P_1 and P_2, so the same is true without the extra conditioning on (S_i).

Since $P_1 + P_2$ is a symmetric function of P_1 and P_2, and $(W_3, W_4, \ldots)$ is a measurable function of $P_1 + P_2$ and $(P_3, P_4, \ldots)$,

> the conditional distribution of $W_3, W_4, \ldots$ given (P_1, P_2) depends symmetrically on P_1 and P_2.

The condition that W_1 is independent of $(W_2, W_3, W_4, \ldots)$ implies easily that

> W_1 is conditionally independent of $(W_3, W_4, \ldots)$ given W_2.

Otherwise put:

> P_1 is conditionally independent of $(W_3, W_4, \ldots)$ given $P_2/(1 - P_1)$,

hence by the symmetry discussed above

> P_2 is conditionally independent of $(W_3, W_4, \ldots)$ given $P_1/(1 - P_2)$.

Let $X := P_2/(1-P_1)$, $Y := P_1/(1-P_2)$ and $Z := (W_3, W_4, \ldots)$. Then we have both

$$X \text{ is conditionally independent of } Z \text{ given } Y, \tag{3.28}$$

and

$$Y \text{ is conditionally independent of } Z \text{ given } X, \tag{3.29}$$

from which it follows under suitable regularity conditions (see Lemma 7 below) that

$$(X, Y) \text{ is independent of } Z, \tag{3.30}$$

meaning in the present context that

$$W_1,\ W_2 \text{ and } (W_3, W_4, \ldots) \text{ are independent.} \tag{3.31}$$

Lauritzen [74, Proposition 3.1] shows that (3.28) and (3.29) imply (3.30) under the assumption that (X, Y, Z) has a positive and continuous joint density relative to a product measure. From (3.27) and strict positivity of the beta densities on $(0, 1)$, we see that (X, Y) has a strictly positive and continuous density relative to Lebesgue measure on $(0, 1)^2$. We are not in a position to assume that (X, Y, Z) has a density relative to a product measure. However, the passage from (3.28) and (3.29) to (3.30) is justified by Lemma 7 below without need for a trivariate density. So we deduce that (3.31) holds. By Proposition 16, $(W_2, W_3, \ldots)$ is the sequence of residual fractions of an exchangeable partition Π', and W_2 has a beta density. So either $W_3 = 1$ and we are in the case (3.19) with $M = 3$, or W_3 has a beta density, and the previous argument applies to show that

$$W_1,\ W_2,\ W_3 \text{ and } (W_4, W_5, \ldots) \text{ are independent.}$$

Continue by induction to conclude the independence of $W_1, W_2, \ldots W_k$ for all k such that $p(1^k) > 0$. $\square$

Lemma 7. *Let X, Y and Z denote random variables with values in arbitrary measurable spaces, all defined on a common probability space, such that* (3.28) *and* (3.29) *hold. If the joint distribution of the pair (X, Y) has a strictly positive probability density relative to some product probability measure, then* (3.30) *holds.*

Proof. Let $p(X, Y)$ be a version of $\mathbb{P}(Z \in B \mid X, Y)$ for B a measurable set in the range of Z. By standard measure theory (e.g. Kallenberg [66, 6.8]) the first conditional independence assumption gives $\mathbb{P}(Z \in B \mid X, Y) = \mathbb{P}(Z \in B \mid X)$ a.s. so that

$$p(X, Y) = g(X) \text{ a.s. for some measurable function } g.$$

Similarly from the second conditional independence assumption,

$p(X, Y) = h(Y)$ a.s. for some measurable function h,

and we wish to conclude that

$p(X, Y) = c$ a.s. for some constant c.

To complete the argument it suffices to draw this conclusion from the above two assumptions about a jointly measurable function p, with (X, Y) the identity map on the product space of pairs $\mathcal{X} \times \mathcal{Y}$, and the two almost sure equalities holding with respect to some probability measure P on this space, with P having a strictly positive density relative to a product probability measure $\mu \otimes \nu$. Fix $u \in (0, 1)$, from the previous assumptions it follows that

$$\{p(X, Y) > u\} = \{X \in A_u\} = \{Y \in C_u\} \quad \text{a.s.} \tag{3.32}$$

for some measurable sets A_u, C_u, whence

$$\{p(X, Y) > u\} = \{X \in A_u\} \cap \{Y \in C_u\} \quad \text{a.s.,} \tag{3.33}$$

where the almost sure equalities hold both with respect to the joint distribution P of (X, Y), and with respect to a product probability measure $\mu \otimes \nu$ governing (X, Y). But under $\mu \otimes \nu$ the random variables X and Y are independent. So if $q := (\mu \otimes \nu)(p(X, Y) > u)$, then (3.32) and (3.33) imply that $q = q^2$, so $q = 0$ or $q = 1$. Thus $p(X, Y)$ is constant a.s. with respect to $\mu \otimes \nu$, hence also constant with respect to P. □

3.6 Fragmentation processes and factors

3.6.1 Fragmentation processes background

Following Bertoin [20, 23], let $\mathcal{P}$ denote the space of partitions of positive integers $\mathbb{N}$, and introduce a metric d on $\mathcal{P}$ by setting $d(\pi, \pi') = 2^{-I(\pi,\pi')}$ where $I(\pi, \pi')$ denotes the largest integer n for which the restriction of π to $[n] = \{1, 2, \ldots, n\}$ is equal to the restriction of π' to $[n]$. Say that a càdlàg $\mathcal{P}$-valued Markov process $(\Pi(t), t \geq 0)$ is a *homogeneous fragmentation process* (an HFP) if

- (refining property) $\Pi(0)$ is the trivial partition of $\mathbb{N}$ into one block, and for all pairs $0 \leq s < t$, $\Pi(t)$ is almost surely equal to or finer than $\Pi(s)$,
- (nontriviality) Π does not immediately jump to the trivial partition $\mathbf{0}$ of $\mathbb{N}$ into singletons: $\mathbb{P}(\Pi(0+) = \mathbf{0}) = 0$. Furthermore, $\mathbb{P}(\Pi(t) = \{\mathbb{N}\}) \to 0$ as $t \to \infty$.
- (exchangeability) $(\Pi(t), t \geq 0)$ is exchangeable, which is to say that the law of $(\Pi(t), t \geq 0)$ is the same as the law of $(\sigma\Pi(t), t \geq 0)$ for all permutations σ of $\mathbb{N}$, the action of a permutation being to relabel the contents of the blocks of a partition,

- (homogeneity) conditionally given $\Pi(t) = \pi$ for some partitition π with blocks (B_1, B_2, ...), the post-t process $(\Pi(t+s), s \geq 0)$ has the same law as the process whose state at time $s \geq 0$ is the partition of $\mathbb{N}$ whose blocks are those of $B_i \cap \Pi^{(i)}(s)$, where $(\Pi^{(i)}, i \geq 1)$ is a sequence of iid copies of Π.

Such processes have been the subject of much recent interest, see [20, 23] for more background. By a result of Bertoin, for all times every time $t \geq 0$ the blocks $B_1(t), B_2(t), \ldots$ of $\Pi(t)$ have limiting frequencies – that is, for every integer j the limits $\xi_j(t) := \lim_{n\to\infty} n^{-1}\#B_j(t)\cap[n]$ exist, with the convention that $\xi_j(t)$ equals zero if $\Pi(t)$ has fewer than j blocks. Another result of Bertoin shows that the law of an HFP Π is determined by a pair (ν, c) where ν is a non-negative *dislocation measure* on the Kingman simplex $\nabla := \{(s_1, s_2, \ldots) : s_1 \geq s_2 \geq \ldots \geq 0, \sum_{i\geq 1} s_i \leq 1\}$ and $c \geq 0$ is the *erosion coefficient* that describes the rate at which blocks of Π lose singletons. This result – Proposition 12 below – provides an explicit formula for the transition rates of the process $(\Pi_n(t), t \geq 0)$ that is the restriction $\Pi_n(t) := \Pi(t) \cap [n]$ of Π to $[n]$ for every positive integer n. These restrictions $(\Pi_n(t), t \geq 0)$ are Markov processes for every integer $n \geq 1$.

Theorem 12 (Bertoin [20])**.** *For every HFP Π there is a* characteristic pair (ν, c)*, where c is a nonnegative real constant and ν is a nonnegative measure on ∇ that assigns no mass to $(1, 0, 0, \ldots)$ and for which $\int_\nabla (1 - s_1)\nu(d\mathbf{s}) < \infty$, such that for every positive integer n and every partition π_n of $[n]$ having block sizes $\lambda = (\lambda_1, \ldots, \lambda_k)$ for some $k \geq 2$, the rate $K(\lambda)$ at which Π_n transitions from $[n]$ to π_n is equal to $K_{\nu,c}$ defined by*

$$K_{\nu,c}(\lambda) := \int_\nabla p_{\mathbf{s}}(\lambda)\nu(d\mathbf{s}) + c\Big(1\big(\min\{\lambda_1, \ldots, \lambda_k\} = 1\big) + 1\big(\lambda = (1, 1)\big)\Big) \tag{3.34}$$

where $p_{\mathbf{s}}$ is the function of (3.5) above, and $1(\bullet)$ is the indicator function of $\bullet$. This pair (ν, c) is uniquely determined by the distribution of Π, and the exchangeable partition rate function *(EPRF) K is a $[0, \infty)$-valued function of compositions that is* symmetric *and additive in the sense of (3.2) and (3.1) hold with K in place of p for every composition λ with at least two parts. Conversely, every symmetric, additive, $[0, \infty)$-valued function K' of compositions with at least two parts is the EPRF of some HFP Π', implying that $K' = K_{\nu,c}$ for some unique pair (ν, c) as above.*

For a proof of Proposition 12 see [20].

Remark. To clarify the role of the integrability constraint $\int_\nabla (1 - s_1)\nu(d\mathbf{s}) < \infty$, note that for $(s_1, s_2, \ldots) \in \nabla$, since $s_1 \geq s_2 \geq \ldots \geq 0$,

$$1 - s_1 \leq 1 - s_1\left(\sum_{i\geq 1} s_i\right) \leq 1 - \sum_i s_i^2 \leq 1 - s_1^2 \leq 2(1 - s_1).$$

Since $p_{\mathbf{s}}(1, 1) = 1 - \sum_i s_i^2$, the integrability constraint is equivalent to $p_\nu(1, 1) < \infty$, which by additivity is equivalent to $p_\nu(\lambda) < \infty$ for all compositions λ having at least two parts.

Note furthermore that if $\hat{\nu}$ is derived from ν by the addition of an atom at $(1, 0, 0, \ldots) \in \nabla$ then $p_\nu(\lambda) = p_{\hat{\nu}}(\lambda)$ for all compositions λ having at least two parts. In this case, p_ν and $p_{\hat{\nu}}$ may only differ on the composition (n), (and even then may not differ if they are both infinite on this composition). The imposition of the constraint $\nu((1, 0, 0, \ldots)) = 0$ ensures the uniqueness of the representation of symmetric additive K' as $K_{\nu,c}$ asserted in Proposition 12: without this constraint, there are are whole family of pairs (ν, c) that represent the given K', each differing by the value of $\nu((1, 0, 0, \ldots))$.

Finally, note that the functions $p_{\mathbf{s}}(n, 1)$ converge to monotonically zero as $n \to \infty$, pointwise in $\mathbf{s}$, and are dominated by $p_{\mathbf{s}}(1, 1)$. Thus in the present setting dominated convergence shows that the constant c can be recovered as

$$c = \lim_{n \to \infty} p(n, 1). \tag{3.35}$$

Let us illustrate Theorem 12 for an HFP Π with characteristic pair $(0, c)$. In this case, we may construct the process directly using a sequence $(Y_1, Y_2, \ldots)$ of IID exponential(c) random variables by setting

$$\Pi(t) = \{\{j \in \mathbb{N} : Y_j \geq t\}\} \cup \{\{j\} : j \in \mathbb{N}, Y_j < t\}.$$

In words, the process Π is a single block from which singletons break off at exponentially-distributed times, together with a collection of singleton blocks. It is easily seen that $(\Pi_n(t))$ has the desired transition rates, and $\Pi(t)$ is exchangeable by exchangeability of (Y_j). From the law of large numbers, almost surely after time t the unique nonsingleton block of $\Pi(t)$ has limit frequency $\exp(-ct)$.

It is likewise easy to describe the fragmentation process $(\Pi(t), t \geq 0)$ with characteristic pair $(\nu, 0)$ by describing the evolution of the *tagged fragment* $B_1(t)$, which is the block of $\Pi(t)$ containing 1, and of the blocks that break off of the tagged fragment. Following [20], construct $(B_1(t), t \geq 0)$ from a Poisson random measure on $\mathbb{R} \times \nabla$ with rate $\lambda \otimes \nu$ – let η be such a random measure. For every atom $\mathbf{a} = (t, \mathbf{s})$ of η, use external randomization to form an exchangeable partition $\pi(\mathbf{a})$ of $\mathbb{N}$ with limit frequencies $\mathbf{s}$, and do this independently for all atoms. Then set

$$B_1(r) := \text{ block of } \bigcap \pi(\mathbf{a}) \text{ that contains integer } 1 \tag{3.36}$$

where the intersection is over all atoms $(t, \mathbf{s})$ of η with $t \leq r$, and the intersection of partitions is understood to mean common refinement. To describe the blocks that break off of B_1, let $B_1(t-) = \bigcup_{s<t} B_1(s)$ denote the contents of B_1 immediately before time t. For every atom $\mathbf{a} = (t, \mathbf{s})$ of η, stipulate that the blocks of the restriction of $\Pi(t)$ to $B_1(t-)$ are the blocks of $\pi(\mathbf{a})$ restricted to $B_1(t-)$:

$$\Pi(t)\Big|_{B_1(t)} := \pi(\mathbf{a})\Big|_{B_1(t)} \qquad \text{if } \eta \text{ has atom } \mathbf{a} = (t, \mathbf{s}).$$

This completes the description of the evolution of the tagged fragment and the blocks that break off of the tagged fragment, and by exchangeability and homogeneity specifies the law of $(\Pi(t), t \geq 0)$. For more details, see [20].

In the sequel we will focus exclusively on homogeneous fragmentation processes without erosion, i.e. processes having characteristic pairs (ν, c) for which $c = 0$. Our focus on fragmentations with no erosion is not quite as narrow as it may initally seem:

Proposition 18 (Bertoin, [23])**.** *If Π^1 and Π^2 are independent HFPs with characteristic pairs (ν_1, c_1) and (ν_2, c_2) then the common refinement $\Pi^1 \cap \Pi^2$, which is the partition-of-$\mathbb{N}$-valued process whose blocks at time t are the nonempty intersections $B^1 \cap B^2$ for B^i a block of $\Pi^1(t)$ for $i \in [2]$, is an HFP with characteristic pair $(\nu_1 + \nu_2, c_1 + c_2)$.*

Proof (sketch). It is straightforward to construct Π^1 and Π^2 directly from independent Poisson random measures as in [20] and sequences $(Y^i_n, n \geq 1)$ of IID exponential(c_i) random variables, $i \in \{1, 2\}$. The assertion then follows from the superposition property of Poisson random measures and the well-known fact that the minimum of exponentially distributed random variables has exponential distribution. See [20] for more detail. □

As a consequence of this proposition, an HFP with characteristic pair (ν, c) can be regarded as the refinement of an HFP with characteristic pair $(\nu, 0)$ by a pure-erosion HFP with characteristic pair $(0, c)$. The latter process, as we have just seen, is very simple; this provides some justification for focusing attention on processes without erosion.

We now state some results of Bertoin that clarify the meaning of $\nu(\nabla) < \infty$ in the fragmentation process context.

Proposition 19 (Bertoin [20])**.** *Let Π be an HFP with characteristic pair $(\nu, 0)$, and let $T_n := \inf\{t \geq 0 : \Pi_n(t) \neq \{[n]\}\}$ denote the first time that Π_n exits state $[n]$. Then T_n is an exponential random variable with rate $\Phi(n)$ for $\Phi(n)$ defined by (3.40), and*

$$\Phi(n) := \int_{\nabla} 1 - \sum_{i \geq 1} s_i^n \nu(d\mathbf{s}) = \int_{\nabla} (1 - p_{\mathbf{s}})(n) \nu(d\mathbf{s}).$$

Furthermore, the following are equivalent:

- $\nu(\nabla) = \infty$
- $p(1) = \infty$
- $p(n) = \infty$ *for at least one integer* n
- $\lambda_n \to \infty$ *as* $n \to \infty$

Finally, letting $\xi_1(t)$ denote the asymptotic frequency of the first block $B_1(t)$ of Π at time t,

$$\xi_1(t) = \lim_{n\to\infty} \frac{1}{n} \#B_1(t) \cap [n] \qquad a.s.$$

$S_t := -\log \xi_1(t)$ is an increasing process with stationary, independent increments, i.e. a subordinator, and the sample paths of S_t have jump discontinuities in every open interval with probability one if and only if $\nu(\nabla) = \infty$.

3.6.2 Factorization

We now say what it means for a measure ν on ∇ to *admit a factor*, and thereby specify the problem addressed in this section of the chapter.

Definition 11 (Haas, Pitman, Winkel [57])**.** *For positive σ-finite measures ν and $\overline{\nu}$ on ∇, say that ν* admits $\overline{\nu}$ as a factor *if ν satisfies the regularity condition $\int_\nabla (1-s_1)\nu(d\mathbf{s}) < \infty$, if $\overline{\nu}$ is a probability measure, if $\nu((1,0,0,\ldots)) = 0$, and if there is a function $g : \{(n,m) : n > m \geq 1\} \mapsto \mathbb{R}_{\geq 0}$ for which*

$$p_\nu(\lambda_1, \ldots, \lambda_k) = g(|\lambda|, \lambda_1) p_{\overline{\nu}}(\lambda_2, \ldots, \lambda_k) \tag{3.37}$$

holds for every composition $(\lambda_1, \ldots, \lambda_k)$ with at least two parts, where $|\lambda| := \lambda_1 + \ldots + \lambda_k$. If ν admits $\overline{\nu}$ as a factor, we say that ν factorizes*.*

The following Theorem provides a characterization of measures that factorize. It has not appeared elsewhere.

Theorem 13. *If a positive measure ν on ∇ factorizes, then one of the following cases obtains.*

(i) ν is atomic, with a single atom at $(0,0,\ldots)$.

(ii) ν concentrates on the set $\{\mathbf{s} : s_2 = 1 - s_1 > 0\}$, and for $n > m \geq 1$,

$$p_\nu(n,m) = g(n,m) = \int_0^1 s^m (1-s)^{n-m} \Sigma(ds),$$

for $\Sigma(ds) = \nu(s_1 \in ds) + \nu(s_2 \in ds)$, and for compositions $(\lambda_1, \ldots, \lambda_k)$ with more than two parts, $p_\nu(\lambda_1, \ldots, \lambda_k) = 0$.

(ii) ν is a member of the augmented two-parameter Poisson-Dirchlet family of measures, defined below in Section 3.7, and p_ν satisfies (3.42) below, or ν is a positive multiple of Dirac mass on the sequence with first M terms equal to $1/M$ for some integer $M \geq 3$, and remaining terms equal to zero, and there is a positive constant c such that for every composition $(\lambda_1, \ldots, \lambda_k)$,

$$p_\nu(\lambda_1, \ldots, \lambda_k) = cM(M-1)\ldots(M-k)\left(\frac{1}{M}\right)^{\lambda_1+\ldots+\lambda_k}. \tag{3.38}$$

At first sight the notion of factorization presented in Definition 11 appears unmotivated, so let us say a few words about what factorization means. First some notation: for an HFP Π let $T_n := \inf\{t : \Pi_n(t) \neq \{[n]\}\}$ denote the first time Π_n exits $[n]$. If the dislocation measure ν of Π admits $\overline{\nu}$ as a factor, it follows that for every partition $\{B_1, \ldots, B_k\}$ having a least two parts,

$$\mathbb{P}(\Pi_n(T_n) = \{B_1, \ldots, B_n\}) = \frac{g(n, \lambda_1)}{\Phi(n)} p_{\overline{\nu}}(\lambda_2, \ldots, \lambda_k), \tag{3.39}$$

for λ the sequence of block sizes of $\{B_1, \ldots, B_k\}$ and $\Phi(n)$ the total rate at which nontrivial jumps occur,

$$\Phi(n) = \sum_{\pi \neq \{[n]\}} p_\nu(\lambda) 1(\lambda = \text{block sizes of } \pi) \tag{3.40}$$

where the sum is over all partitions π of $[n]$ *except the trivial partition* $\{[n]\}$. Since $\overline{\nu}$ is a probability measure it follows from Proposition 16 that $p_{\overline{\nu}}$ is an EPPF. By summing (3.39) over all partitions $\{B_1, \ldots, B_k\}$ having at least two parts it can be seen that $\sum_{i=1}^{n-1} \binom{n-1}{i-1} g(n, i) = \Phi(n)$, where the combinatorial factor arises as the number of ways to choose a subset of $[n]$ of size i that contains 1. It follows that $\Pi_n(T_n)$ is derived as if by splitting off from $[n]$ a block $B_1(n)$ containing 1 with $\mathbb{P}(B_1(n) = B) = 1(B \text{ contains } 1) g(n, \#B)/\Phi(n)$, and then partitioning the remaining set $[n] \setminus B_1$ according to an independent PB($\overline{\nu}$)-partition, that is, a partition independent of $B_1(n)$ and having EPPF $p_{\overline{\nu}}$. This reasoning can be reversed to show that if this property holds for all $n \geq 2$ then ν admits $\overline{\nu}$ as a factor; see also Theorem 14 below.

From this discussion it is clear that if ν factorizes then the process Π has a property very much like the deletion property. The relevant distinction is that ν need not be a probability measure: $p_\nu(n) = \infty$ is possible in this context, whereas in the EPPF/deletion property context $p(1) = 1$ and $0 < p(n) \leq 1$ for all $n \geq 1$. See Proposition 19 above for discussion of what $p_\nu(n) = \infty$ signifies in terms of the associated fragmentation process.

The next theorem is a translation of a theorem from [57] that provides a condition on Π equivalent to factorization. First, some definitions: we define the *coarse spinal partition* Π^{coarse} of $\mathbb{N}' := \{2, 3, 4, \ldots\}$ as follows: put integers i and j in the same block of Π^{coarse} if and only if

$$\inf\{t \geq 0 : i \notin B_1(t)\} = \inf\{t \geq 0 : j \notin B_1(t)\},$$

where $B_1(t)$ denotes the block of $\Pi(t)$ containing 1. In other words, i and j are in the same block of Π^{coarse} if and only if i and j break off of $B_1(t)$ at the same time. Next we define the *fine spinal partition* Π^{fine} of $\mathbb{N}'$ by putting integers i and j in the same block of Π^{fine} if and only if i and j are in the same block of the coarse spinal partition and

$$i \text{ and } j \text{ are in the same block of } \Pi(T) \text{ for } T = \inf\{t > 0 : i \notin B_1\},$$

where $\Pi(T)$ in the display is the value of the fragmentation process Π evaluated at time T. In other words, i and j are in the same block of Π^{fine} if and only if i and j break off of B_1 at the same time and in the same piece.

Theorem 14 ([57], Theorem 5)**.** *If* Π *is an HFP with characteristic pair* $(\nu, 0)$*, then the following are equivalent.*

- ν *admits* $\overline{\nu}$ *as a factor*
- *The fine spinal partition is derived by shattering the blocks of the coarse spinal partitions with independent copies of PB(*$\overline{\nu}$*)-partitions.*

The proof of Theorem 13 depends on a detailed analysis of the function g appearing in (11). Let us relate this function to ν.

Proposition 20. *Suppose that* ν *admits* $\overline{\nu}$ *as a factor and that* g *is the function appearing in (11). Then for* $1 \leq m < n$*,*

$$g(n,m) = \int_{\nabla} \sum_{i \geq 1} s_i^m (1 - s_i)^{n-m} + 1(m = 1)\left(1 - \sum_{i \geq 1} s_i\right) \nu(d\mathbf{s}). \tag{3.41}$$

Proof. Let Π be a HFP with characteristic pair $(\nu, 0)$, fix a subset B of $[n]$ of size m with $1 \in B$, and compute the rate R at which Π_n exits state $[n]$ and enters partitions with first block B. On one hand,

$$R = \sum_{\pi} p_{\nu}\left(\lambda(\pi)\right) = \int_{\nabla} \sum_{\pi} p_{\mathbf{s}}\left(\lambda(\pi)\right) \nu(d\mathbf{s})$$

where the sum is over all partitions π in which the block containing 1 equals B and $\lambda(\pi)$ denotes the sequence of block sizes of π. The integrand on the right can be simplified: regarding $p_{\mathbf{s}}\left(\lambda(\pi)\right)$ as the EPPF of a PB($\delta_{\mathbf{s}}$) partition, where $\delta_{\mathbf{s}}$ is Dirac mass at $\mathbf{s}$, we have

$$p_{\mathbf{s}}\left(\lambda(\pi)\right) = \sum_{i \geq 1} s_i^m (1 - s_i)^{n-m} + 1(m = 1)\left(1 - \sum_{i \geq 1} s_i\right)$$

for every partition π in which the block containing 1 equals B. On the other hand,

$$R = g(n,m) \sum_{\pi} p_{\overline{\nu}}\left(\lambda(\pi)\right) = g(n,m),$$

the second equality holding by virtue of the fact $p_{\overline{\nu}}$ is an EPPF and therefore sums to one. □

3.7 The augmented two-parameter family of Poisson-Dirichlet measures

Say that a nonnegative sigma-finite measure ν on ∇ is a member of the *augmented Poisson-Dichlet family of measures* if $\int_\nabla (1-s_1)\nu(d\mathbf{s}) < \infty$ and

$$\frac{p_\nu(\lambda_1,\ldots,\lambda_k)}{p_\nu(1,1)} = \frac{(\theta+2\alpha)(\theta+3\alpha)\ldots(\theta+(k-1)\alpha)}{(2+\theta)(3+\theta)\ldots(n-1+\theta)}\prod_{i=1}^{k}(1-\alpha)_{\lambda_i-1} \tag{3.42}$$

for (α,θ) in the range $\{0\le\alpha<1,\theta>-2\alpha\}\cup\{\alpha<0,-\theta/\alpha\in\mathbb{N}\}$ and $(y)_j := y(y+1)\ldots(y+j-1)$. Dividing by $p_\nu(1,1)$ removes scaling: it is easily seen that if ν is in this family then so is $x\nu$ for every $x>0$. See [57, 77] for developments related to (3.42).

The right-hand-side of (3.42) is a symmetric, additive $[0,\infty)$-valued function of compositions for parameters in the indicated range, and it is easily seen that if ν is a member of the augmented Poisson-Dirichlet family then

$$\lim_{n\to\infty} p(n,1) = 0$$

Thus it follows from Proposition 12 and (3.35) that ν is determined by (3.42) up to a multiplicative constant. That is, for every $x>0$ and every pair of parameters (α,θ) in the indicated range there is exactly one positive sigma-finite measure ν on ∇ with $p_\nu(1,1)=x$ that puts no mass on $(1,0,0\ldots)$ and that verifies (3.42). On the other hand, for parameters not in the indicated range it is possible to find compositions λ for which the right-hand-side of (3.42) is negative or involves division by zero.

Conditions equivalent to finiteness of ν in the augmented Poisson-Dirichlet family are known from [57]:

$$\nu(\nabla)<\infty \Leftrightarrow \begin{cases}\alpha<0\ \theta=-M\alpha,\ M\in\{1,2,3\ldots\}\\ \alpha\ge 0\ \theta>-\alpha\end{cases} \tag{3.43}$$

Before continuing let us say a few words about another specification of part of the augmented Poisson-Dirichlet family. For $0<\alpha<1$, let η be a Poisson point process on $(0,\infty)$ with intensity $\alpha(\Gamma(1-\alpha))^{-1}x^{-\alpha-1}\mathrm{d}x$, and let the atoms of η be ranked in decreasing order $\Delta_1\ge\Delta_2\ge\ldots$. Then $T:=\sum_{i\ge1}\Delta_i$ has the distribution of the value at time 1 of an α-stable subordinator, and it is known that

$$\Gamma(1-\alpha)\lim_{j\to\infty}\left(\frac{j\Delta_j}{T}\right) = T^{-\alpha}$$

holds almost surely [81]. Then for every real θ the formula

$$\int_\nabla f(\mathbf{s})PD*(\alpha,\theta)(d\mathbf{s}) := \mathbb{E}\left[T^{-\theta}f(\Delta_1/T,\Delta_2/T,\ldots)\right],$$

which holds by fiat for every nonnegative measurable function f on ∇, defines a measure that in [57] is called the $PD^*(\alpha, \theta)$ measure. For $\theta > -2\alpha$ this measure satisfies (3.42), and is thus a member of what is here called the *augmented* Poisson-Dirichlet family. We use the term "augmented" to avoid confusion with the *extended* Ewens-Pitman family defined in Section 3.3 as the closure of the class of distributions of exchangeable partitions of $\mathbb{N}$ with EPPFs satisfying (3.16), because in [57] the term *extended* is used to describe the $PD^*(\alpha, \theta)$ measures.

3.8 Proof of Theorem 13

The case where $p(2, 2, 1) = 0$ is covered already by the proof of Theorem 11, so we will assume throughout that

$$p_\nu(2, 2, 1) > 0. \tag{3.44}$$

The proof of Theorem 13 in this case relies on the following proposition.

Proposition 21. *If ν admits $\overline{\nu}$ as a factor and g is the function for which $p_\nu = gp_{\overline{\nu}}$ as in (3.37) or (3.41), and (3.44) holds, then g uniquely determines p_ν by the formulas*

$$p_\nu(\lambda_1, \ldots, \lambda_k) = p(1^k)\frac{g(\Lambda_1, \lambda_1)}{g(k, 1)} \prod_{j=1}^{k-1} \frac{g(\Lambda_{j+1} + j, \lambda_{j+1})}{g(\Lambda_{j+1} + j, 1)} \tag{3.45}$$

which holds for all compositions $(\lambda_1, \ldots, \lambda_k)$ having $k \geq 2$ parts, where 1^k denotes a sequence of k ones and $\Lambda_j := \lambda_j + \ldots + \lambda_k$ for $1 \leq j \leq k$, and

$$p_\nu(1^k) = g(2, 1) \prod_{j=3}^{k} \left(1 - \frac{(j-1)g(j, 2)}{g(j-1, 1)}\right), \tag{3.46}$$

which holds for all $k \geq 3$, and the trivial identity $g(2, 1) = p_\nu(1, 1)$.

Proof. Fix a composition $(\lambda_1, \ldots, \lambda_k)$ having at least two parts, and for $1 \leq i \leq k$ set $\Lambda_i := \lambda_i + \ldots + \lambda_k$. From (3.44) we deduce that $\nu(\{s_1 \geq s_2 > 0\}) > 0$, so from (3.41) it can be seen that $g(n, m) > 0$ for all $n > m \geq 1$. From (3.37) it follows that

$$\frac{p_\nu(1, \ldots, \lambda_k)}{g(\Lambda_2 + 1, 1)} = p_{\overline{\nu}}(\lambda_2, \ldots, \lambda_k) = \frac{p_\nu(\lambda_1, \ldots, \lambda_k)}{g(\Lambda_2 + \lambda_1, \lambda_1)},$$

which implies

$$p_\nu(\lambda_1, \ldots, \lambda_k) = \frac{g(\Lambda_1, \lambda_1)}{g(\Lambda_2 + 1, 1)} p_\nu(1, \lambda_2, \ldots, \lambda_k).$$

One part of the composition $(\lambda_1, \ldots, \lambda_k)$ has been reduced to 1, and by using symmetry of p_ν and iterating we can likewise reduce the other parts to 1 as well, eventually obtaining

(3.45). Now, from additivity of p we have $p_\nu(1^k) = kp_\nu(2, 1^{k-1}) + p_\nu(1^{k+1})$, so if $p_\nu(1^k) \neq 0$ then we may divide to obtain

$$\frac{p_\nu(1^{k+1})}{p_\nu(1^k)} = 1 - k\frac{p_\nu(2, 1^{k-1})}{p_\nu(1^k)} = 1 - k\frac{g(k+1, 2)}{g(k, 1)}, \tag{3.47}$$

where the second equality following from (3.45). By additivity, for $k \geq 3$, if $p_\nu(1^k) > 0$ then $p_\nu(1^{k-1}) > 0$, and of course $p_\nu(1, 1) > 0$. Thus if $p_\nu(1^k) > 0$ may write $p_\nu(1^k)/p_\nu(1, 1)$ as a product of terms of the form $p_\nu(1^{j+1})/p_\nu(1^j)$ as in (3.47) and arrive at (3.46), which *a fortiori* holds for $k < k_0 := \inf\{j : p_\nu(1^j) = 0\}$ but perhaps not for all $k \geq 3$. However, if k_0 is finite then from (3.47) we see

$$0 = p_\nu(1^{k_0}) = p(1^{k_0-1})\left[1 - (k_0 - 1)\frac{g(k_0 - 1, 2)}{g(k_0, 1)}\right],$$

and it follows that (3.46) holds for all $k \geq 3$.

□

Proof of Theorem 13. First assume (3.44), which evidently implies that $\nu(\{s_2 > 0, s_1 + s_2 < 1\}) > 0$ which in turn ensures that for every integer m, either $p_\nu(2, 3, m) > 0$ or $p_\nu(2, 3, 1^m) > 0$, or both, where 1^m denotes a sequence of m ones. Accordingly we let l_m equal either 1^m or m so that $p_\nu(2, 3, l_m) > 0$. Upon using (3.45) to express $p_\nu(2, 3, l_m)$ in product form, and equating this with the analogous expression for $p_\nu(3, 2, l_m)$, and then canceling common nonzero factors, we arrive at the relation

$$\frac{g(m+5, 3)g(m+3, 2)}{g(m+3, 1)} = \frac{g(m+5, 2)g(m+4, 3)}{g(m+4, 1)} \tag{3.48}$$

which holds for every positive integer m. For $m \geq 0$, make the following definitions:

$$x_m := g(m+2, 1), \qquad y_m := \frac{x_{m+1}}{x_m}, \qquad z_m := \frac{1}{1 - y_m}. \tag{3.49}$$

Additivity of g implies that $g(m+2, 2) = x_m - x_{m+1}$, and together with (3.41) and (3.44) this shows that $y_m \in (0, 1)$ for all $m \geq 0$. Another application of additivity of g yields the further relation $g(m+4, 3) = x_m - 2x_{m+1} + x_{m+2}$. We can therefore rewrite (3.48) as follows,

$$\frac{(x_{m+1} - 2x_{m+2} + x_{m+3})(x_m - x_{m+1})}{x_{m+1}} = \frac{(x_{m+2} - x_{m+3})(x_m - 2x_{m+1} + x_{m+2})}{x_{m+2}}$$

and after multiplying both sides by x_m^{-1}, passing to y-variables, expanding and simplifying, this becomes

$$-2y_{m+1} + y_{m+1}y_{m+2} + y_m y_{m+1} = -y_m - y_{m+2} + 2y_m y_{m+2}.$$

Make the substitution $y_m = (z_m - 1)z_m^{-1}$, bring all terms to to the left hand side of the equality, and multiply both sides by $z_m z_{m+1} z_{m+2}$. The resulting expression simplifies to

$$z_{m+2} - 2z_{m+1} + z_m = 0$$

which shows that $z_m = a + bm$ for some constants a and b. Since $0 < y_m < 1$ for all m it follows that $a > 1$ and $b \geq 0$. There are now two cases to consider.

1. Case $b > 0$. In this case, from (3.49) we find that

$$\frac{x_{m+1}}{x_m} = \frac{a + bm - 1}{a + bm}$$

and thus that

$$\frac{g(m+2,1)}{g(2,1)} = \frac{x_m}{x_0} = \frac{(u)_m}{(v)_m}, \tag{3.50}$$

where $u := (a-1)/b$ and $v := a/b$, and $(r)_m = r(r+1)\ldots(r+k-1)$ denotes the rising factorial. Since $g(m+2,2) = g(m+1,1) - g(m+1,1)$ for $m \geq 1$ it follows from (3.50) that

$$\frac{g(m+2,2)}{g(2,1)} = \frac{(v-u)(u)_{m-1}}{(v)_m}$$

and continuing inductively using the relations $g(n+1,m+1) = g(n,m) - g(n+1,m)$ it can be seen that

$$\frac{g(n,m)}{g(2,1)} = \frac{(v-u)_{m-1}(u)_{n-m-1}}{(v)_{n-2}} \tag{3.51}$$

holds for $1 \leq m < n$. After substitution into (3.45) and cancellation we arrive at

$$p_\nu(\lambda_1, \ldots, \lambda_k) = p_\nu(1^k)\frac{(v)_{k-2}}{(v)_{n-2}} \prod_{i=1}^{k} (v-u)_{\lambda_i - 1} \tag{3.52}$$

where $n = \lambda_1 + \ldots + \lambda_k$ and 1^k denotes the composition of k into k ones.

Similarly, we find

$$1 - \frac{(j-1)g(j,2)}{g(j-1,1)} = \frac{v - 2 + (j-1)(1-v+u)}{v - 2 + j - 1}$$

for $j \geq 3$, and substitution into (3.46) yields

$$\frac{p_\nu(1^k)}{p_\nu(1,1)} = \frac{((v-2) + 2(1-v+u))\ldots(v-2+(k-1)(1-v+u))}{(v)_{k-2}}. \tag{3.53}$$

Combination of (3.52) and (3.53) yields (3.42) for $\theta = v - 2$, $\alpha = 1 - (v-u)$. Note that since $v = a/b$ we have $\theta > -2$, and because $\alpha = 1 - 1/b$ we have $\alpha < 1$.

2. If $b = 0$ then we find from (3.49) that

$$g(n+2,1) = g(2,1)w^n$$

for $w := (a-1)/a$. From additivity, $g(n+2,2) = g(n+1,1) - g(n+2,1)$ holds for all $n \geq 1$, and thus

$$\frac{g(n+2,2)}{g(2,1)} = w^{n-1}(1-w)$$

holds for $n \geq 1$. Now define a measure μ on $[0,1]$ by

$$\mu(ds) = \frac{s^2 w}{(1-w)g(2,1)} \sum_{i\geq 1} \nu(s_i \in ds), \tag{3.54}$$

so that

$$\int_0^1 (1-s)^n \mu(ds) = \frac{wg(n+2,2)}{(1-w)g(2,1)} = w^n.$$

This shows that μ is Dirac mass at $\overline{w} := 1-w$, which implies the following constraint on ν: for ν-almost every $\mathbf{s} \in \nabla$, and for every entry s_i of $\mathbf{s}$, s_i equals $\overline{w}$ or 1 or 0. And in fact, from Definition 11 it follows that $\nu((1,0,0,\ldots)) = 0$: there are no ones. We claim that $\nu(s_1 + s_2 + \ldots < 1) = 0$. Indeed, supposing otherwise, since

$$p_\nu(1^k) \geq \int_\nabla \left(1 - \sum_{i\geq 1} s_i\right)^k \nu(d\mathbf{s})$$

it follows that $p_\nu(1^k) > 0$ for all k. From (3.41) and (3.44) it can be seen that $g(n,m) > 0$ for all $1 \leq m < n < \infty$, and then from Proposition 21 one obtains $p_\nu(2^k) > 0$ for all $k \geq 2$, where 2^k denotes a sequence of k twos. Note furthermore that $p_\nu(2^k) > 0$ is implies that ν puts mass on some subset of ∇ with first k entries strictly positive. On the other hand, since $\sum_i s_i \leq 1$ for all $\mathbf{s} \in \nabla$ it follows that ν-almost-every $\mathbf{s}$ has at most $\overline{w}^{-1}$ positive entries. This contradiction proves that $\overline{w}^{-1}$ is an integer M and that ν is Dirac mass on the the sequence $(1/M, \ldots, 1/M, 0, 0, \ldots)$ consisting of M copies of $1/M$, followed by zeros. Then (3.38) can be seen directly. In light of (3.44) we see that $M \geq 3$.

□

Second Proof of Theorem 11. Suppose that Π has the deletion property and $\mathbb{P}(W_1 < 1) = 1$, where W_1 is the first residual limit frequency of Π, and let ν be the distribution of the ranked sequence of limit frequencies of Π. Then according to Proposition 17 ν factorizes. If ν is atomic, or if ν concentrates on $\{\mathbf{s} : s_2 = 1 - s_1 > 0\}$, or if ν is a pointmass at $(1/M, \ldots, 1/M, 0, 0, \ldots)$, we are done. Therefore suppose that (3.42) holds for p_ν for some

(α, θ), so that for integer compositions λ with at least two parts, $p_\nu(\lambda)$ is determined up to a multiplicative constant $p_\nu(1,1)$. This constant could be determined by revisiting the proof of Theorem 11. Instead, note that since $\mathbb{P}(W_1 < 1) = 1$ it follows that $p_\nu(n) = \mathbf{E}(W_1^{n-1}) \to 0$ as $n \to \infty$, and because ν is a probability measure,

$$1 - p(n) = \sum_\pi p(\lambda(\pi))$$

where the sum is over all partitions of $[n]$ having at least two parts and $\lambda(\pi)$ denotes the sequence of block sizes of π. Therefore the multiplicative constant is determined by the ratios (3.42) by taking a limit, and thus $p_\nu(\lambda)$ is determined uniquely by (3.42) for all compositions: if $\hat{\nu}$ is a probability measure on ∇ for which

$$\frac{p_\nu(\lambda)}{p_\nu(1,1)} = \frac{p_{\hat{\nu}}(\lambda)}{p_{\hat{\nu}}(1,1)}$$

holds for all compositions λ having at least two parts, then $\hat{\nu} = \nu$. From (3.43), either $\alpha < 0$, $\theta = -M\alpha$ for some integer $M \geq 1$ or $\alpha \in [0,1)$ and $\theta > -\alpha$.

$\square$

Chapter 4

Uniform Hierarchies

Suppose that T_n is a random rooted unoriented (i.e. nonplanar) tree, with edges endowed with lengths, having n leaves labeled by $[n] := \{1, \ldots, n\}$, and suppose that these leaf-labels are exchangeable, meaning that for every permutation $\sigma : [n] \mapsto [n]$ the trees T_n and $\sigma(T_n)$ are equal in distribution, where $\sigma(T_n)$ is the tree derived by relabeling every leaf leaf i of T_n by $\sigma(i)$. For example, T_n may be the subtree of a weighted rooted real tree (T, μ) spanned by the root of T and $(Y_1, \ldots, Y_n)$ which are exchangeable elements of T with directing measure μ; alternatively, T_n may be a Galton-Watson tree conditioned to have n leaves. The exchangeability of leaves of T_n implies a collection of distributional constraints on interleaf distances: for example, in such a tree the distance between leaves 1 and 2 is equal in distribution to the distance between leaves 2 and 3. The main theorem in this chapter is a characterization of the Brownian CRT, the proof of which uses distributional constraints imposed by leaf-exchangeability. As an application of this characterization we show that the Brownian CRT is in a limited sense the $n \to \infty$ scaling limit of *uniform random hierarchies*, that is, rooted unoriented trees on n leaves that have no internal vertices of degree two.

4.1 Introduction and statement of results

There is the following "leaves-up" method of describing a rooted tree T with no non-root vertices of degree 2 by recursively describing the subtrees T_k spanned by the first k leaves of T for $k = 1, 2, \ldots$: to specify T one first says what is the length X_1 of the path $[[\rho, 1]]$ connecting root to leaf 1, then T_1 is known to be (isometric to) a line segment of length X_1. Then, recursively for $k \geq 1$, letting $[[\rho, k+1]]$ denote the path in T from root to leaf $k+1$, one specifies the *attachment point* J_{k+1} which is the point furthest from the root in the set $[[\rho, k+1]] \cap T_k$, one specifies the distance X_{k+1} from leaf $k+1$ to the attachment point J_{k+1}, and T_{k+1} is then known to be (isometric to) the tree derived by grafting a branch of length X_{k+1} to the attachment point J_{k+1} in T_k. See [2, 29, 9] for tree models of this type. For contrast, see [75, 54] and references therein for an introduction to a large literature on "root down" constructions of trees.

The recipe above can be carried out explicitly by embedding the growing trees as subsets of ℓ_1. This construction is due to Aldous who writes that the main idea is to "always add edges orthogonally" [4]. In Aldous's sequential construction, $(0, 0, \ldots) \in \ell_1$ is regarded as root, T_1 is a tree with one branch of length X_1 that lies in the first coordinate direction, $T_1 := \{(t, 0, 0, \ldots) : 0 \leq t \leq X_1\}$, and the point $(X_1, 0, 0, \ldots)$ may be regarded as the first leaf, labeled 1. After specifying a point $J_2 \in T_1$ to be the first attachment point, one may form T_2 by adding a branch of length X_2 lying in the second coordinate direction, $T_2 := T_1 \cup \{J_2 + (0, t, 0, 0, \ldots) : 0 \leq t \leq X_2\}$, and the point $J_2 + (0, X_2, 0, 0, \ldots)$ may be regarded as a leaf labeled 2. The construction continues in this manner; see Chapter 3 for a longer description.

We call pair of sequences $((X_i), (J_i))$ of branch lengths and attachment points a *line-breaking* representation of a tree. There is a convenient *planar representation* which can encode a line-breaking representation, and therefore a tree, as follows.

Definition 12 (Aldous, Pitman [9]). *A* planar representation *of a rooted, unoriented tree T_n having n leaves is a pair $(U, V) = ((U_i, 1 \leq i \leq n), (V_i, 1 \leq i \leq n-1))$ of sequences of real numbers satisfying the inequalities $0 < U_1 < U_2 < \ldots < U_n$ and $0 < V_i < U_i$, for $1 \leq i \leq n-1$. The pair (U, V) encodes a tree as follows: set $U_0 := 0$, and define line-segments $\mathcal{X}_i$ by*

$$\mathcal{X}_i = \begin{cases} [0, U_1] & \text{if } i = 1 \\ (U_{i-1}, U_i] & \text{if } \leq i \leq n \end{cases}, \tag{4.1}$$

let $T_1 = \mathcal{X}_1$, and for $1 \leq m \leq n-1$ construct T_{m+1} recursively by gluing the open end "$(U_{i-1}$" of $\mathcal{X}_{m+1}$ onto the point V_m, which is located somewhere on some segment of T_m. For $i \in [n]$, the closed end "$U_i]$" is then the i^{th} leaf of the tree, and 0 is the root. More formally, one may set $X_1 = U_1$ and set $X_i := U_i - U_{i-1}$ for $i \geq 2$, let $(\mathbf{e}_i)$ denote the standard basis in ℓ_1, let $J_1 := (0, 0, \ldots)$ be root*, let $T_1 = \{te_{\mathbf{1}} : 0 \leq t \leq X_1\}$, and recursively for $m \geq 1$, once T_m has been defined, set*

$$B_{m+1} = \max\{k : U_k < V_m\} \qquad J_{m+1} := J_{B(m+1)} + (V_m - U_{B(m+1)})\mathbf{e}_{m+1};$$

this specifies the pair $((X_i), (J_i))$ explicitly.

In another direction, the tree T_n can be understood as a random metric space $([0, U_n], d)$ where the metric d on the interval $[0, U_n]$ depends on the sequence $(U_1, \ldots, U_n, V_2, \ldots, V_n)$. See [9] for more about this.

It is obvious that every unoriented rooted tree on $n \geq 1$ labeled leaves with no internal vertices of degree two except perhaps the root has a planar representation as above. The best example of a line-breaking representation of a random tree is provided by Aldous's Brownian continuum random tree (CRT):

Definition 13 (Line-breaking construction of Brownian CRT [2, 4])**.** *Let (X_i) be the sequence of inter-point distances of a Poisson point process on $\mathbb{R}_{\geq 0}$ of rate $t\ dt$. Let T_1 be a line segment of length X_1, with one end declared to be the root ρ and the other end declared to be a leaf and labeled 1. Then for $k \geq 1$, supposing T_k has been defined, let J_{k+1} be a point chosen from the normalized length-measure on T_k and let T_{k+1} be derived by grafting a branch of length X_{k+1} to the point J_{k+1}, labeling the new leaf – i.e. the free end of the newly-grafted branch – by $k+1$. For $n \geq 1$, the tree T_n defined above is known as the n^{th}* marginal of the Brownian CRT.

It is easily checked (see [2]) that if $((U_i),(V_i))$ is the planar representation of the n^{th} marginal of the Brownian CRT then $(U_1, V_1,), \dots, (U_{n-1}, V_{n-1})$ are the first $n-1$ points of a Poisson process of unit rate on the orthant $\mathbb{G} := \{(x, y) : 0 < y < x\}$, when points are ordered by their first coordinates. We will later make use of this fact to show that certain random trees converge to marginals of the Brownian CRT.

It is not obvious (but still true, see [4]) that the n^{th} marginal of the Brownian CRT is *leaf-exchangeable* in the following sense.

Definition 14. *If T_n is a random tree on n labeled leaves, with each leaf bearing a distinct label in the set $[n]$, say that T_n has* exchangeable leaves *or* exchangeably labeled leaves *if for every permutation σ of $[n]$, there is the distributional equality*

$$T_n \stackrel{d}{=} \sigma(T_n)$$

where $\sigma(T)$ is the tree derived by relabeling the leaves according to the permutation σ.

For a tree T_n on n leaves with planar representation (U, V) define W_i by

$$W_i = W_i(T_n) = V_i/U_i \qquad (1 \leq i \leq n-1). \tag{4.2}$$

Definition 15. *Say that a tree T_n* satisfies hypothesis H *if T_n is a rooted tree on $n \geq 1$ leaves labeled by $[n]$ with no non-root vertices of degree two,*

(a) T_n has exchangeably-labeled leaves,

(b) $\mathbb{P}(W_1 > 0) > 0$,

(c) W_1 and $(X_1, \dots, X_n)$ are independent, for W_1 defined by (4.2)

(d) X_{m+1} and $(W_1, \dots, W_{m+1})$ are conditionally independent given $(X_1, \dots, X_m)$ for $1 \leq m < n$,

(e) For all $\lambda \in \mathbb{R}$, $\mathbb{E}[\exp(\lambda X_1)] < \infty$.

Here (U, V) is the planar representation of T_n and (X_k) is the sequence of branch-lengths of T_n, $X_i = U_i - U_{i-1}$, $2 \le i \le n$, $X_1 = U_1$.

For example, the trees (T_n) appearing in the construction of the Brownian CRT evidently satisfy hypothesis H, because in that case $(W_1, \dots, W_{n-1})$ are uniform random variables independent of $(U_1, \dots, U_n)$. Part (d) is perhaps the most substantial part of hypothesis H: it is meant to encapsulate something a little stronger than the idea that the distribution of the length of the next branch to be added does not depend on where that branch is to be added.

Before stating the main theorem we must also introduce the notion of an *attachment mechanism.*

Definition 16. *If T_n is a random rooted tree with planar representation (U, V) then an* attachment mechanism $(a_1, \dots, a_{n-1})$ *is a sequence of regular conditional distributions for which*

$$\mathbb{P}(W_i \in \bullet \mid U_1, \dots, U_i, W_1, \dots, W_{i-1}) = a_i(U_1, \dots, U_i, W_1, \dots, W_{i-1}, \bullet) \qquad (4.3)$$

holds almost surely for $i = 1, \dots, n-1$, for (W_i) defined by (4.2).

For example, in the Brownian CRT case, (W_i) is a family of IID random variables independent of the (U_i), in this case the left hand side of 4.3 is almost surely equal to the Lebesgue measure of $\bullet \cap [0, 1]$.

Theorem 15. *Suppose that T_n is a tree satisfying hypothesis H. Let X_1 denote be the distance in T_n from leaf 1 to root, and let (a_k) be an attachment mechanism for T_n. Then $(\mathcal{L}(X_1), (a_k))$ determines $\mathcal{L}(T_n)$.*

Here and throughout this chapter $\mathcal{L}(Z)$ denotes the law of a T_k. Put otherwise, Theorem 15 asserts that if two trees satisfying hypothesis H have the same attachment mechanism and the same first branch length, then they are equal in law. The following corollary is perhaps of greater interest than Theorem 15: it is an immediate consequence of Theorem 15 and the easily checked observation that the n^{th} marginal of the Brownian CRT satisfies hypothesis H.

Corollary 4. *Suppose that the tree T_n with planar representation (U, V) satisfies hypothesis H, and that (W_i) is defined by (4.2). If $X_1 := U_1$ has Rayleigh(1) distribution and if for all $1 \le m \le n$ the conditional distribution of W_m given $((U_1, \dots, U_m), (W_1, \dots, W_{m-1}))$ is the uniform$[0, 1]$ distribution, then then T_n has the distribution of the n^{th} marginal of the Brownian CRT.*

We illustrate the potential utility of Corollary 4 by using it to prove the following theorem.

Theorem 16. *For $n \ge 2$ let T_n have the uniform distribution on the set of unoriented trees having n leaves labeled by $\{1, \dots, n\}$, a root that is not a leaf, no non-root vertices of degree* *two, and edges of length* $2(2\log 2-1)^{1/2}n^{-1/2}$. *Let* $T_{n,k}$ *denote the subtree of* T_n *spanned by the root and the first* k *leaves. Then* $T_{n,k}$ *converges in distribution to* k^{th} *marginal of the Brownian CRT as* $n \to \infty$.

More explicitly, Theorem 16 asserts that if $(U_i(n,k), V_i(n,k))$ is the planar representation of $T_{n,k}$ above then $(U_1, \ldots, U_k)$ converge in distribution to the locations of the first k points in a Poisson process on $(0,\infty)$ of rate $t\,dt$ and that $((U_i, V_i), i = 1, \ldots, k-1)$ converge in distribution to the locations of the first $k-1$ points of a Poisson process of unit rate in the orthant $\mathbb{G} = \{(x,y) : 0 < y < x\}$ (note that there are $k-1$ V_i's but k U_i's). Trees of the type considered in Theorem 16 are called *hierarchies* [44, IV.47], and

Theorem 16 may be regarded as saying that the sequence $((T_n, \mu_n), n \geq 2)$ converges to the Brownian CRT Gromov-weakly almost surely, where for $n \geq 2$, T_n is the tree appearing in Theorem 16 and μ_n is the discrete uniform distribution on the leaves of T_n. See [53] for a discussion of this convergence. It is of interest to prove the stronger statement that uniform hierarchies converge to the Brownian CRT in the weighted Gromov-Hausdorff sense [43] – this has been done by Doug Rizzolo and Jim Pitman [85], who provide an argument relying on fragmentation-process formalism instead of the heavy use of symmetry and exchangeability found in this chapter.

4.2 Proof of Theorem 15

We begin by discussing a special case of Theorem 15 pointed out by Jim Pitman. Suppose that T_2 is a tree satisfying the conditions of Theorem 15, derived by selecting a point J_2 uniformly at random from a line segment of length X_1 and grafting to this point a line segment of length X_2. Let $W :=$(distance from root to J_2)$/X_1$. Then T_2 is a tree

- with a root of degree 1, incident to a single edge of length $X_1 W$
- and at the end of this edge is an interior vertex of degree 3 (and therefore of out-degree two)
- and the two other edges incident to this interior vertex have lengths $X_1(1-W)$ and X_2, respectively,
- and at the ends of these edges are leaves labeled 1 and 2, respectively.

Assume that T_2 has exchangeable leaves. Then the joint distribution of edge-lengths encountered on the path from root to leaf 1 is the same is as the joint distribution of edge-lengths encountered on the path from root to leaf 2; that is, $(X_1 W, X_1(1-W))$ and $(X_1 W, X_2)$ are equal in distribution. It is plausible – and Lemma 8 addresses this issue directly – that the joint distribution of $(X_1 W, X_2)$ may determine that of (X_1, X_2). Therefore from the joint distribution of (X_1, J_2) may determine the joint distribution of (X_1, X_2, J_2). Theorem 15 is proved by formalizing and generalizing this line of reasoning.

Lemma 8. *Suppose that $(W, X_1, \ldots, X_n)$ is a random element of $\mathbb{R} \times \mathbb{R}^n$ for some $n \geq 1$ and suppose furthermore that*

(i) W has the Uniform[0,1] distribution and is independent of $(X_1, \ldots, X_n)$, and

(ii) $\mathbb{E}[\exp(\lambda X_i)] < \infty$ for every $\lambda \in \mathbb{R}$ and $i \in [n]$.

Then the distribution of $(X_1, \ldots, X_n)$ is determined by the distribution of $(WX_1, X_2, \ldots, X_n)$.

Proof. The hypotheses allow to compute

$$\mathbb{E}[\frac{\lambda^n}{n!} X_1^n \exp(\lambda_2 X_2 + \ldots + \lambda_n X_n)] = \frac{(n+1)\lambda^n}{n!} \mathbb{E}[(UX_1)^n \exp(\lambda_2 X_2 + \ldots + \lambda_n X_n)]$$

(it can be seen by e.g. Hölder's inequality that these expectations are finite). By summing terms as above one may compute $\mathbb{E}[\exp(\lambda_1 X_1 + \ldots + \lambda_n X_n)]$ by dominated convergence for any $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$. These collection of expectations of this type for $\lambda \in \mathbb{R}^n$ determines the joint law of $(X_1, \ldots, X_n)$. $\square$

Definition 17 (Erasure operator)**.** *For $n \geq 2$ and t_n a rooted tree on n leaves labeled by $[n]$ with no non-root, non-leaf vertices of degree 2, let $\mathbf{e}_i(t_n)$ denote the tree derived by erasing the leaf i from t_n and also erasing the edge of t_n incident to leaf i,* but retaining the non-leaf vertex of the this edge incident to the erased leaf*. Therefore $\mathbf{e}_i(t_n)$ is a tree on $n-1$ leaves that are labeled by $[n] \setminus \{i\}$, together with a mark that distinguishes a vertex of the tree.*

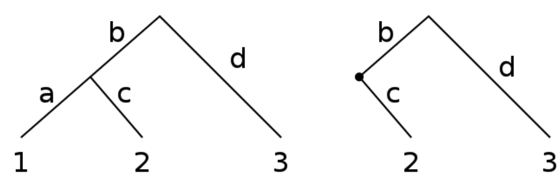

Figure 4.1: A hierarchy $\mathbf{t}_3$ and $\mathbf{e}_1(\mathbf{t_3})$. In this case, $U_1 X_1 = b$, so $(W_1 X_1, X_2, X_3)$ is measurable w.r.t. $\mathbf{e}_1(\mathbf{t}_3)$.

Proposition 22. *Suppose that $((X_k), (J_k))$ is the line-breaking representation of a tree T_n with n leaves, having planar representation (U, V) and let W be defined by (4.2). Then $(X_1 W_1, X_2, \ldots, X_n)$ is a measurable function of $\mathbf{e}_1(T_n)$.*

Proof. The proposition is obvious from pictures; see for example in Figure 4.1. $\square$

Proof of Theorem 15. We also remark at the outset that (e) of hypothesis H may be strengthened to (e') $\mathbb{E}[\exp(\lambda X_i)] < \infty$ for all $\lambda \in \mathbb{R}$ and $i \in [n]$.

Indeed, for $i \geq 2$, the distance X_i from leaf i to T_{i-1} is smaller than the distance from leaf i to root, and this latter distance is equal in distribution to X_1 by leaf-exchangeability.

We argue inductively. The base case is obviously true: $\mathcal{L}(T_1)$ is determined by $(\mathcal{L}(X_1), (a_k))$, since T_1 is a segment of length X_1 with one end labeled root and the other end a leaf labeled 1.

For the inductive step, suppose that $(\mathcal{L}(X_1), (a_k))$ determines the law of T_m, the subtree of T_n spanned by the root and the first $m < n$ vertices. Thus

$$(\mathcal{L}(X_1), (a_k)) \text{ determines } \mathcal{L}(T_m, J_{m+1}) \tag{4.4}$$

Observe that $\mathbf{e}_{m+1}(T_{m+1})$ is a measurable function of (T_m, J_{m+1}): in fact, $\mathbf{e}_{m+1}(T_{m+1})$ *is exactly equal to* the tree T_m decorated with the mark J_{m+1} indicating where the branch with leaf $m+1$ attaches to T_m. Thus

$$(\mathcal{L}(X_1), (a_k)) \;\text{ determines }\; \mathcal{L}\big(\mathbf{e}_{m+1}(T_{m+1})\big). \tag{4.5}$$

Next observe that for any permutation σ of $[m+1]$ for which $\sigma(1) = m+1$, and with erasure and permutation operators $\mathbf{e}$ and σ defined as Definitions 17 and 14, respectively,

$$\sigma^{-1}\Big(\mathbf{e}_{m+1}(T_{m+1})\Big) \stackrel{d}{=} \sigma^{-1}\Big(\mathbf{e}_{m+1}\big(\sigma(T_{m+1})\big)\Big) = \mathbf{e}_1\big(T_{m+1}\big). \tag{4.6}$$

The first equality in (4.6) is an equality in distribution which holds by exchangeability, and the second equality holds pointwise and therefore almost surely. This pointwise equality simply asserts that the tree derived from T_{m+1} by relabeling leaf 1 by $m+1$, then erasing the leaf labeled $m+1$, and then undoing the relabeling, is equal to the tree derived from T_{m+1} by just erasing the leaf labeled 1. Figure 4.2 illustrates this second equality.

From (4.5), (4.6) and Proposition 22 it follows that $(\mathcal{L}(X_1), (a_k))$ determines the joint distribution of $(X_1W_1, X_2, \ldots, X_{m+1})$. By Proposition 8 $(\mathcal{L}(X_1), (a_k))$ therefore determines the joint distribution of $(X_1, \ldots, X_{m+1})$, and thus also the conditional distribution of X_{m+1} given $(X_1, \ldots, X_m)$, which is the same, by part (d) of hypothesis H, as the conditional distribution of X_{m+1} given (T_m, J_{m+1}). From (4.4) it follows that $(\mathcal{L}(X_1), (a_k))$ determines $\mathcal{L}(T_m, J_{m+1}, X_{m+1})$; this completes the inductive step.

□

4.3 Proof of Theorem 16

Throughout this section, T_n will denote a random rooted unoriented tree having the uniform distribution on the set of trees having $n \geq 1$ leaves labeled by $\{1, \ldots, n\}$, no non-root vertices of degree two, and edges of length $c_n := 2(2\log 2 - 1)^{1/2} n^{-1/2}$. By abuse of terminology we

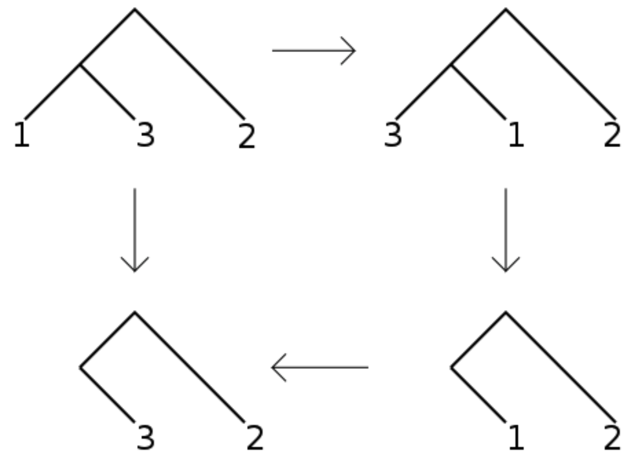

Figure 4.2: Illustration of the second equality of (4.6).

refer to trees of this type as hierarchies since trees of this type are in bijective correspondence with set hierarchies, discussed extensively in Chapter 2, and we refer to T_n as a uniform random hierarchy. There is only one hierarchy having one leaf: the trivial graph with a single vertex labeled 1 that is both leaf and root. For hierarchies having at least two leaves, we stipulate that the root is required to be a non-leaf vertex.

Proposition 23. *For $n \geq 1$ let T_n be a uniform random hierarchy on n leaves with edges of length $c_n := 2(2\log 2 - 1)^{1/2} n^{-1/2}$, and let $X_1(n)$ denote the distance from the leaf labeled 1 to the root. Then $X_1(n)$ converges in distribution as $n \to \infty$ to a random variable X_1 having Rayleigh(1) distribution,*

$$\mathbb{P}(X_1 \in dx) = x\exp(-x^2/2)1(x > 0)$$

I first heard this result from Douglas Rizzolo, who provided a proof using standard techniques of analytic combinatorics [85]. Section 4.3.1 contains a slightly different proof using such techniques.

Before proceeding we need some preliminary definitions. For $1 \leq j \leq n$ let $T_{n,k}$ denote the subtree of T_n spanned by the leaves $1, \ldots, k$, and let $(U(n,k), V(n,k))$ denote the planar representation of $T_{n,k}$. Let

$$X_i(n) = X_i(n,k) = \begin{cases} U_1(n,k) & \text{if } i = 1 \\ U_i(n,k) - U_{i-1}(n,k) & \text{if } 2 \leq i \leq k \end{cases}$$

be the branch-lengths of $T_{n,k}$ (note that branch lengths are scaled by $c_n = O(n^{-1/2})$), and let

$$W_i(n) = W_i(n,k) = \frac{V_i(n,k)}{U_i(n,k)} \qquad (1 \leq i \leq k-1). \tag{4.7}$$

Proposition 24. *For all $k \geq 1$ the family $(T_{n,k}, n \geq 1)$ is tight; i.e.,*

$$\Big((X_j(n,k), W_i(n,k), j \in [k], i \in [k-1]), n \geq 1\Big)$$

is a tight family for all $k \in \mathbb{N}$.

Proof. For $j \geq 2$, the distance from leaf j to root is greater than the distance from leaf j to the subtree $T_{n,j-1}$ spanned by the root and first $j-1$ leaves, so by exchangeability, $X_1(n)$ stochastically dominates $X_j(n)$ for $j \geq 2$. From Proposition 23 $(X_1(n), n \geq 1)$ is tight. It follows that $((X_1(n), \ldots, X_k(n)), n \geq 1)$ is tight. Since $(W_i(n), n \geq 1)$ are contained in the interval [0,1], the claim follows. □

Proposition 25. *Suppose that $(X_1, \ldots, X_k, W_1)$ is a weak subsequential limit of $((X_1(n,k), \ldots, X_k(n,k), W_1(n,k)), n \geq 1)$. Then*

(a) $(X_1, \ldots, X_k)$ and W_1 are independent

(b) W_1 has uniform distribution

(c) $\mathbb{P}(X_i = 0) = 0$ for $i \in [k]$

Proof. Consider the *spinal path* starting from the leaf labeled 1 and ending at the root in the uniform hierarchy T_n. Cutting or erasing every edge along this path shatters T_n into a sequence of subtrees called *bushes*, and the length N of this sequence is the number of edges along the path, that is, $N = X_1(n)/c_n$. Let $(B_j, j \in [N])$ denote this sequence of bushes, so that B_1 is the first bush encountered on the spinal path, and B_{L_n} is the last bush, attached to the root of T_n. Let $\{B_j, j \in [N]\}$ denote the unordered collection of these bushes, that is, the collection of trees in the sequence (B_j) with relative order forgotten. Put otherwise, $\{B_j, j \in L_n\}$ is an equivalence class of sequences of trees, two sequences $(b_j^1, j \in [l_1])$ and $(b_j^2, j \in l_2)$ being equivalent if $l_1 = l_2$ and there is a permutation σ of l_1 for which $(b_j^1, j \in [l_1]) = (b_{\sigma(j)}^2, j \in [l_2])$. The following claims are straightforward to check:

- $W_1(n,k)$ is measurable with respect to $(B_j, j \in [N])$. Moreover, if I_n equals the index of the subtree B_{I_n} containing leaf 2, then $W_1(n,k) = (N - I_n)/N$.
- Conditionally given $\{B_j, j \in [N]\}$, the distribution of $(B_j, j \in [N])$ is uniform over all $N!$ possible linear arrangements of the members of $\{B_j, j \in [N]\}$.
- $(X_1(n), \ldots, X_k(n))$ is measurable with respect to $\{B_j, j \in [N]\}$.

Thus, conditionally given $\{B_j, j \in [N]\}$, the distribution of $W_1(n,k)$ is discrete uniform on $\{0, 1/N, \ldots, (N-1)/N\}$ where $N = X_1(n)/c_n$ and, conditionally given $X_1(n)$, $W_1(n,k)$ independent of $(X_2(n,k), \ldots, X_n(n,k))$. Since $\mathbb{P}(N_n < K) \to 0$ as $n \to \infty$ for all fixed K as a consequence of Proposition 23, Parts (a) and (b) follow.

Part (c) is obvious for $i = 1$ from Proposition 23. For the general case, let $(i \wedge j)_n$ denote the vertex of T_n where the path from root to leaf i first diverges from the path from root to leaf j. By leaf-exchangeability it follows that the distance from leaf 1 to $(1 \wedge 2)_n$ is equal in distribution to the distance from leaf 2 to $(1 \wedge 2)_n$. The former distance is $(1 - W_1(n,k))X_1(n,k)$ and the latter distance is $X_2(n,k)$, and thus $X_2 \stackrel{d}{=} X_1 W_1$, proving (c) for $i = 2$.

For the general case, fix $\delta > 0$ and consider the set $\mathcal{S}_n^i$ defined as

$$\mathcal{S}_n^i = \{\text{hierarchies } \mathbf{t}_n \text{ on } n \text{ leaves } : \{T_n = \mathbf{t}_n\} \subseteq \{X_i(n,k) < \delta\}\},$$

i.e. $\mathcal{S}_n^i$ is the set of hierarchies on n leaves for which the i^{th} branch has length less than δ. Now, if $\mathbf{t}_n \in \mathcal{S}_n^i$ has its i^{th} branch attached to its j^{th} branch, then any permutation of $[n]$ that sends j to 1 and i to 2 sends $\mathbf{t}_n$ to a tree in $\mathcal{S}_n^2$. Let Σ denote a uniform random permutation of $[k]$ independent of T_n, let $T'_{n,k} := \Sigma(T_{n,k})$, and let $(X'_1(n,k), \ldots, X'_k(n,k))$ be the first k branch lengths of $T'_{n,k}$. Since T_n is uniformly distributed, $T_{n,k}$ is leaf exchangeable, so $X_2(n,k) \stackrel{d}{=} X'_2(n,k)$, and thus

$$\frac{1}{k(k-1)} \mathbb{P}(X_i(n,k) < \delta) \leq \mathbb{P}(X'_2(n,k) < \delta).$$

Now send $n \to \infty$, follow with $\delta \to 0$, and apply the $i = 2$ case of (c) to obtain the general case.

□

Proposition 26. *For $1 \leq k \leq n$ let $B_{n,k}$ be the probability that $T_{n,k}$ is binary. Then for each $k \geq 1$, $\mathbb{P}(B_{n,k}) \to 1$ as $n \to \infty$.*

Proof. Consider the spinal paths from leaf i to root for $i \in \{1, 2\}$. Let $\mathcal{S}$ be the set of vertices of $T_{n,2}$ that are found on both of these spinal paths, and let $(1 \wedge 2)_n$ denote the member of $\mathcal{S}$ that is furthest from the root. Thus $(1 \wedge 2)_n$ is the vertex in $T_{n,2}$ where the path from root to leaf 1 diverges from the path from root to leaf 2.

Let

$$\mathcal{R} = \bigcup_{i=1}^{2} \{\text{vertices } v \in T_n : v \text{ is on the spinal path from leaf } i \text{ to root.}\}$$

Cutting all of the edges of $T_{n,2}$ that lie between vertices of $\mathcal{R}$ shatters $T_{n,2}$ into bushes, and each vertex that lies on one of these paths becomes the root of exactly one bush. After cutting edges, 1 and 2 form singleton bushes. The bush rooted at $(1 \wedge 2)_n$ is special: it may be a singleton, consisting solely of $(1 \wedge 2)_n$, as in Figure 4.3. It is easily seen that when edges between vertices of $\mathcal{R}$ are cut, except for $(1 \wedge 2)_n$ and the first two leaves, every vertex in $\mathcal{R}$

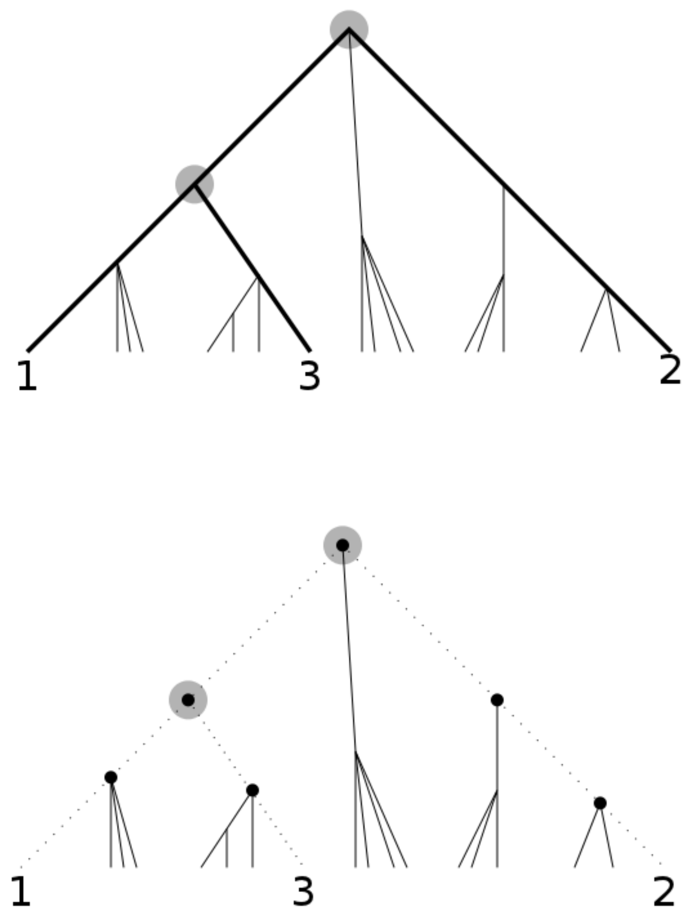

Figure 4.3: The two highlighted vertices are members of the set $\{v : v \text{ is the vertex at which the path from root to leaf } i \text{ diverges from the path from root to leaf } j \text{ for some distinct } i, j \in [3]\}$. When edges along the first three spinal paths are erased, one of these highlighed vertices becomes a singleton bush, the other becomes a nonsingleton bush, and all other vertices along the first three spinal paths except the leaves 1,2,3 become nonsingleton bushes.

must become the root of a nonsingleton bush. Letting Y_n denote the number of nonsingleton bushes, we have thus

$$Y_n \in \{(X_1(n)+X_2(n))/c_n, (X_1(n)+X_2(n))/c_n - 1\} \qquad a.s.$$

The nonsingleton bushes in $T_{n,2}$ are exchangeable. More explicitly, if b_1 is a nonsingleton bush rooted at $r_1 \in \mathcal{R}$ and b_2 a nonsingleton vertex rooted at a different vertex $r_2 \in \mathcal{R}$, one may prune off b_1 and b_2 and regraft b_2 to r_1 and b_1 to r_2 to produce a new hierarchy. Since T_n is uniformly distributed this operation preserves probability. Since the leaf 3 lies in a nonsingleton bush, it follows that

$$\mathbb{P}(B^c_{n,3} \mid X_1(n), X_2(n)) \leq \frac{c_n}{X_1(n)+X_2(n)-c_n}$$

and then by Proposition 23 it follows that

$$\limsup_{n\to\infty} \mathbb{P}(B^c_{n,3}) = 0. \tag{4.8}$$

Now for distinct integers $i, j \in [k]$ let $(i \wedge j)_n$ denote the vertex of $T_{n,k}$ at which the path from root to leaf i first diverges from the path from root to leaf j. Then

$$\mathbb{P}(B^c_{n,k}) \leq \sum_{i,j,k \text{ distinct}} \mathbb{P}\left((i\wedge j)_n = (j \wedge k)_n\right).$$

Since T_n is uniformly distributed, $T_{n,k}$ is leaf-exchangeable, so

$$\sum_{i,j,k \text{ distinct}} \mathbb{P}\left((i\wedge j)_n \leq (j \wedge k)_n\right) \leq k^3 \mathbb{P}((1\wedge 2)_n = (2\wedge 3)_n) = k^3 \mathbb{P}(B^c_{n,3})$$

and the claim follows from (4.8). □

Proposition 27. *Suppose that* $((X_1,\dots,X_k),(W_1,\dots,W_{k-1})$ *is a weak subsequential limit of* $((X_1(n,k),\dots,X_k(n,k)),(W_1(n,k),\dots,W_{k-1}(n,k)))$*. Then* $(W_1,\dots,W_{k-1})$ *is a sequence of independent and identically-distributed uniform[0,1] random variables independent of* $(X_1, \dots, X_k)$*.*

Proof. Fix open intervals $(r_1,s_1),\dots,(r_k,s_k)$ in $(0,\infty)$ with $r_i > 0$ for $i = 1,\dots,n$, and suppose that n is large enough that $\mathbb{P}(X_1(n,k) > r_1,\dots,X_k(n,k) > r_k) > 0$, for example $n > 2(r_1+\ldots+r_k)/c_n$ suffices. Also fix intervals $(a_1,b_1),\dots,(a_{k-1},b_{k-1})$ contained in $[0,1]$. Our first goal is to approximate the probability

$$\mathbb{P}\left(W_i(n,k) \in (a_i,b_i) \text{ for } i \in [k-1], X_1(n,k) = x_1,\dots,X_k(n,k) = x_k\right) \tag{4.9}$$

for sequences $(x_1,\dots,x_k)$ in $\prod_{i=1}^k (r_i,s_i)$ with $\mathbb{P}(X_1(n,k) = x_1,\dots,X_k(n,k) = x_k) > 0$. Fixing such a sequence, we will compute a lower bound for (4.9) by means of the following stochastic algorithm for generating hierarchies $\mathbf{t}_n$ on n leaves having first k branch lengths equal to $x_1,\dots,x_k$.

$\star$-Algorithm: Part 1

1. Let $\mathbf{t}_1$ be a branch of length x_1 with leaf 1 at the end of the branch. Put a grid of width c_n along this first branch so that the branch has x_1/c_n grid points, not counting the leaf 1 as a grid point.

2. Next, select one of these grid points of $\mathbf{t}_1$ uniformly at random, mark this grid point with a $\star$, and attach to this grid point a branch of length x_2. Put a leaf labeled 2 at the end of this second branch and as before add a grid of width c_n to the second branch. As before the leaf 2 is not to be regarded as a point in the grid. Call the resulting tree $\mathbf{t}_2$.

3. Next, select a grid point from $\mathbf{t}_2$ uniformly at random from among the $x_1/c_n+x_2/c_n-1$ grid points in $\mathbf{t}_2$ that are *not* marked with a $\star$. Mark this new grid point with another $\star$ and attach to it a branch of length x_3. Put a leaf labeled 3 at the end of this branch and likewise put a grid along this new branch. This produces a tree $\mathbf{t}_3$ with 3 leaves and $(x_1+x_2+x_3)/c_n$ internal vertices, two of which are marked with stars.

4. Proceed in this manner, selecting new grid points uniformly at random, marking each selected grid point with a star and adding a branch to the selected grid point, subject to the constraint that once a point is marked with a $\star$ it can never be chosen again. This procedure produces a binary tree $\mathbf{t}_k$ having $(x_1+\ldots+x_k)/c_n-(k-1)$ unstarred grid points and $k-1$ starred grid points and k leaves labeled by $\{1,\ldots,k\}$. We will regard these grid points as ordered: the first x_1/c_n-1 grid points lie on the first branch, etc., the first starred grid point is the first grid point to receive a star, etc.

The second part of the algorithm adds bushes to $\mathbf{t}_k$ to produce a hierarchy. Before specifying the second part of the algorithm let us say a few more words about bushes.

Let $\mathbf{t}_n$ be a hierarchy with n leaves and let

$$\mathcal{R}_k := \bigcup_{i=1}^{k}\{\text{ vertices } v \text{ of } \mathbf{t}_n : v \text{ is on path from leaf } i \text{ to root}\}$$

denote the collection of vertices of $\mathbf{t}_n$ that lie in the first k spinal paths from leaves $i \in [k]$ to root. Remove every edge of $\mathbf{t}_n$ that lies between vertices of $\mathcal{R}_k$ to form a new graph G'_n, and remove from G'_n the leaves $1,\ldots,k$, which form singleton components, to form another graph G_n. The connected components of this graph are *bushes*. The root of a bush b in G_n is the unique vertex of b that belongs to $\mathcal{R}_k$. It is easily seen that if b bush on $j \geq 2$ leaves, then either b has no internal vertices of degree two except perhaps the root, or the root ρ of b has degree 1 and the graph H derived by rerooting b at the neighbor of ρ and then removing ρ, has no internal vertices of degree two except perhaps the new root. In other words, bushes look very much like hierarchies. If b is a singleton bush– that is, a single vertex in $\mathcal{R}_k$ – it follows that b must be the vertex at which the path in $\mathbf{t}_n$ from root to leaf i diverges from the path from root to leaf j. There are at most $k-1$ such singleton bushes since there are at most $k-1$ exceptional vertices where spinal paths diverge.

Now let $\mathcal{C}(n,k)$ be the set of sequences $\mathbf{s}$ of length $L = (x_1 + \ldots + x_k)/c_n$, where each entry in $\mathbf{s}$ is a bush, and for each $i \in \{1, \ldots, i \leq L-(k-1)\}$ the i^{th} entry s_i of $\mathbf{s}$ is a nonsingleton bush (so that any or all of the final $(k-1)$ entries of $\mathbf{s}$ may be, but are not required to be, singleton bushes), such that all bushes have disjoint leaf-sets, and such that the union of leaves in the bushes of $\mathbf{s}$ is the set $\{k+1, k+2, \ldots, n\}$. We can now describe the second part of the $\star-$algorithm.

$\star-$Algorithm, part 2

5. Once $\mathbf{t}_k$ has been produced, select uniformly at random a sequence $\mathbf{s} = (s_1, s_2, \ldots)$ from $\mathcal{C}_{n,k}$. Sequentially attach the bushes $s_1, s_2, \ldots$ to the grid-points of $\mathbf{t}_k$ by identifying grid-points with roots of bushes, saving the last $k-1$ entries of $\mathbf{s}$ to be attached at the starred grid points of $\mathbf{t}_k$. This produces a hierarchy $\mathbf{t}_n$.

It can then be seen the hierarchy $\mathbf{t}_n$ produced by the $\star$-algorithm is uniformly distributed over the set $\mathcal{S}$,

$\mathcal{S} :=$ { hierarchies τ_n on n leaves : the subtree $\tau_{n,k}$ spanned by the root and first k leaves of τ is *binary* and has first k edge-lengths equal to $x_1, \ldots, x_n$ }.

Indeed, regarded in another light, the $\star$-algorithm simply provides an explicit parametrization or coordinatization of $\mathcal{S}$: speaking a little loosely, every element of $\mathcal{S}$ corresponds to

- a sequence $(j_1, \ldots, j_{k-1})$ of real numbers, each j_i taking values in a grid, subject to a *ties* constraint that ensures binary-ness of $\tau_{n,k}$: this sequence parameterizes $\tau_{n,k}$,
- an element of $\mathcal{C}_{n,k}$ that parametrizes "the complement, $\tau_n \setminus \tau_{n,k}$".

The uniform measure on $\mathcal{S}$ is the pushforward of uniform measure on the space of parameters $\{(j_1, \ldots, j_{k-1})\} \times \mathcal{C}_{n,k}$ via this parametrization, and the $\star$-algorithm produces hierarchies by selecting an element of the parameter space uniformly at random.

Now let us define events E_n and F_n by

$$\begin{cases} E_n := \{X_1(n,k) = x_1, \ldots, X_k(n,k) = x_k\} \\ F_n = \{W_i(n,k) \in (a_i, b_i) \text{ for } i \in [k-1]\} \end{cases} \tag{4.10}$$

Note that because T_n is uniformly distributed, conditionally given E_n and the event $B_{n,k}$ that $T_{n,k}$ is binary, T_n is uniformly distributed over the set $\mathcal{S}$ above. Then

$$P(E_n, F_n) \geq \mathbb{P}(F_n \mid E_n, B_{n,k})\mathbb{P}(E_n, B_{n,k}) \geq \mathbb{P}(F_n \mid E_n, B_{n,k})(\mathbb{P}(E_n) - \mathbb{P}(B_{n,k}^c))$$

and we may compute $\mathbb{P}(F_n \mid E, B_{n,k})$ by summing over attachment points for "next branches" in part one of $\star$-algorithm. We find

$$\mathbb{P}(E_n F_n) \geq (\mathbb{P}(E_n) - \mathbb{P}(B_{n,k})) \prod_{i=1}^{k-1} \left(\frac{(b_i - a_i)(x_1 + \ldots + x_i)}{c_n} - 2 - (i-1) \right) \frac{c_n}{x_1 + \ldots + x_i}; \tag{4.11}$$

where the 2 comes from possible "edge effects" when counting the number of discrete grid-points in a real interval, and the $(i-1)$ comes from possibly having to avoid previous attachment points. In a more manageable form,

$$\mathbb{P}(E_nF_n) \geq \left(\mathbb{P}(E_n) - \mathbb{P}(B_{n,k}^c)\right)\left(\prod_{i=1}^{k-1}(b_i - a_i)\right)(1 - D\frac{c_n}{x_1}) \tag{4.12}$$

where D is a constant not depending on $((a_i, b_i), i \geq 1)$ or on $x_2, \dots x_k$. Likewise,

$$\mathbb{P}(E_nF_n) \leq \mathbb{P}(B_{n,k}^c) + \mathbb{P}(F_n \mid E_n, B_{n,k})\mathbb{P}(E_n)$$

and thus

$$\mathbb{P}(E_nF_n) \leq P(B_{n,k}^c) + \mathbb{P}(E_n)\prod_{i=1}^{k-1}\left(\frac{(b_i - a_i)(x_1 + \dots + x_i)}{c_n} + 2\right)\frac{c_n}{x_1 + \dots + x_i}, \tag{4.13}$$

or more simply

$$\mathbb{P}(E_nF_n) \leq \mathbb{P}(B_{n,k}^c) + \mathbb{P}(E_n)\left(\prod_{i=1}^{k-1}(b_i - a_i)\right)(1 + D\frac{c_n}{x_1}) \tag{4.14}$$

for another constant D that does not depend on $x_2, \dots, x_k$ or $((a_i, b_i), i \in [k-1])$. Now set

$$\hat{E}_n := \{X_1(n,k) \in (r_1, s_1), \dots, X_k(n,k) \in (r_k, s_k)\}$$

and sum (4.12) and (4.14) over the set the set $\{(x_1, \dots, x_n) : x_i \in (r_i, s_i), \text{ for } i \in [k]\}$ to obtain

$$\left(\mathbb{P}(\hat{E}_n) - \mathbb{P}(B_{n,k}^c)\right)\left(\prod_{i=1}^{k-1}(b_i - a_i)\right)(1 - D\frac{c_n}{r_1}) \leq \mathbb{P}(\hat{E}_n, F_n) \tag{4.15}$$

and

$$\mathbb{P}(\hat{E}_n, F_n) \leq \mathbb{P}(B_{n,k}^c) + \mathbb{P}(\hat{E}_n)\left(\prod_{i=1}^{k-1}(b_i - a_i)\right)(1 + D\frac{c_n}{r_1}) \tag{4.16}$$

Send $n \to \infty$, noting that $c_n \downarrow 0$ and $\mathbb{P}(B_{n,k}) \to 1$ by Proposition 26, to conclude that

$$\mathbb{P}(\hat{E}, F) = \mathbb{P}(\hat{E})\prod_{i=1}^{k-1}(b_i - a_i) \tag{4.17}$$

for

$$\begin{cases} \hat{E} = \{X_1 \in (r_1, s_1), \dots, X_k \in (r_k, s_k)\} \\ F = \{W_1 \in (a_1, b_1), \dots, W_{k-1} \in (a_{k-1}, b_{k-1}). \end{cases}$$

At the outset we required $r_i > 0$ for $i \in [k]$. However, in view of Part (c) of Proposition 25, the desired conclusion still follows from (4.17). $\square$

Proof of Theorem 16. Suppose that $((X_1,\dots,X_k),(W_1,\dots,W_{k-1}))$ is a weak subsequential limit of $((X_i(n,k),W_j(n,k),i\in[k],j\in[k-1]),n\geq 1)$. Then $((X_i(n,k),V_j(n,k),i\in[k],j\in[k-1]),n\geq 1)$ converges in law along the same subsequence as well to the planar representation of a tree T_n. The limit satisfies hypothesis H, and X_1 has Rayleigh distribution by Proposition 23. In view of 24, the conclusion follows from Corollary 4.

□

The following proposition is not needed to prove any of the results mentioned in the beginning of this chapter. Nonetheless it may be of some interest, because speaking loosely it can be said to rely principally on symmetry or uniformity as does the rest of this section. The proposition also follows from stronger results derived using different methods [85].

If $(\tau_n, n\geq 1)$ is a family of random trees with τ_n having n leaves, then (τ_n) is said to be *leaf-tight* if for all $\epsilon>0$,

$$\limsup_{n\to\infty}\mathbb{P}(\min\{\text{ distance between leaf 1 and leaf } j \text{ in } \tau_n : 2\leq j\leq n\}>\epsilon)=0. \tag{4.18}$$

Proposition 28. *The family (T_n) of uniform random hierarchies is leaf tight.*

Proof. Cutting the edges on the spinal path from 1 to root shatters T_n into a collection of subtrees called bushes. Each bush b has a distinguished vertex r_b called the root of b which is the unique vertex of b found on this spinal path – say that b is *rooted at* r_b to indicate this relationship. If b is bush then cutting the edges of b that are adjacent to r_b shatters b into a collection of subtrees, each nominally rooted at the vertex previously adjacent to r_b. These subtrees are easily seen to be a *almost-hierarchies*, i.e. graphs that satisfy the postulates of a hierarchy except that leaves need not be consecutively ordered. This defect of leaf-labeling can be remedied by relabeling the leaves of such a graph h' by the unique increasing bijection sending the set $S(h')$ of leaves of h to $\{1,\dots,\#S\}$ to produce a graph h that is a hierarchy in the strict sense. A tedious straightforward counting argument establishes that the entire collection of hierarchies $\mathcal{H}_n$ derived by first shattering T_n into bushes, then shattering these bushes into almost hierarchies, and then relabeling these almost-hierarchies to produce hierarchies, is a collection of independent hierarchies, each uniformly distributed, conditionally given the collection of sizes of leaf sets, ie. conditionally given the multiset $\{\#h : h\in\mathcal{H}_n\}$.

Fix $n>1$ and $\epsilon>0$. Then by Proposition 23 are then $O_P(\epsilon\sqrt{n})$ vertices on the spinal path from leaf 1 to root within distance ϵ of the leaf 1. As above let $\mathcal{H}_n$ denote the collection of hierarchies derived from the the bushes rooted at these $O_P(\epsilon\sqrt{n})$ vertices by the edge-cutting recipe described above. For a hierarchy $h\in\mathcal{H}_n$ let $N(h)$ denote the number of leaves of h, and let $d(h)$ denote the distance from r_b to the first leaf of h. Conditionally given $N(h)=m$, h is uniformly distributed and $d(h)$ is equal in distribution to $X_1(m)c_n/c_m$ for sequences (c_n) and $(X_1(n))$ defined as in Proposition 23. We claim that there are integers (m_0,n_0) for which

$$\inf\{\mathbb{P}(X_1(m)c_n/c_m\leq\epsilon) : m\geq m_0 \text{ or } n\geq n_0\}>0; \tag{4.19}$$

Indeed, the existence of m_0 for which

$$\inf\{\mathbb{P}(X_1(m)c_n/c_m \leq \epsilon) : m \geq m_0\} > 0;$$

is obvious from Proposition 23 and then any n_0 for which $c_{n_0} < (1/2)\epsilon m_0^{-1}$ can be shown to work for (4.19), for then $X_1(m)c_{n_0}/c_m < \epsilon$ almost surely for all $m < m_0$.

Now set $M_n :=$ min{distance between leaf 1 and leaf j in T_n: $2 \leq j \leq n$}. Observe that

$$0 \leq M_n \leq \epsilon + c_n + \min\{d(h) : h \in \mathcal{H}_n\}. \tag{4.20}$$

In view of (4.19), for $n > n_0$, for every hierarchy h appearing in (4.20) there is some $\delta > 0$ not depending on n for which $\mathbb{P}(d(h) \leq \epsilon) > \delta$. By the previously-asserted conditional independence of the family $(h : h \in \mathcal{H}_n)$ given the sizes $(N(h) : h \in \mathcal{H}_n)$, and the easily checked fact, following from Proposition 23, that the number of hierarchies in $\mathcal{H}_n$ grows without bound almost surely, it follows that $\mathbb{P}(M_n \leq 2\epsilon) \to 1$ as $n \to \infty$. Therefore (T_n) is leaf-tight.

□

4.3.1 Proof of Proposition 23

This section closely follows ideas from [44, pp 128, 280, 472-4, 479], which contains a number of facts about hierarchies. The main result (4.30) was first obtained in [85] by methods similar to those employed here.

Let $\mathcal{L}_n$ denote the class of hierarchies having exactly n leaves, let $L_n := \#\mathcal{L}_n$, and let $L(z)$ denote the exponential generating function for the class $\mathcal{L} := \bigcup_{n\geq 1} \mathcal{L}_n$ of hierarchies,

$$L(z) := \sum_{n=1}^{\infty} \frac{L_n}{n!} z^n. \tag{4.21}$$

By convention, $L_0 = 0$.

Proposition 29. *For $n \geq 1$, $L_n < 3^n n!$. Furthermore, $L(z)$ defines a complex-analytic function on the disc $\{z \in \mathbb{C} : |z| < 1/3\}$.*

Proof. For a hierarchy t_n having $n \geq 2$ leaves write $t_n\Big|_{[n-1]}$ to denote the hierarchy derived from t_n by erasing the leaf labeled n and the edge adjacent to this leaf, and then erasing any non-root degree-two vertices formed by this erasure. Then t_n is derived as if by attaching a branch to t_{n-1} in one of three ways:

- selecting an internal vertex of t_{n-1} and attaching an extra branch to this vertex
- selecting an edge of t_{n-1} and attaching an extra branch to this edge, thereby creating a new internal vertex

- creating a new root, attaching the new branch to this new root and attaching the old root of t_{n-1} to this new root as well.

It follows that with H_n denoting the set of hierarchies on n leaves, and t_{n-1} any hierarchy on $n-1$ leaves,

$$\#\{t \in H : t\Big|[n-1] = t_{n-1}\} = e(t_{n-1}) + v(t_{n-1}) + 1 \tag{4.22}$$

where $e(t_{n-1})$ and $v(t_{n-1})$ denote the number of edges and number of vertices of t_{n-1}, respectively. Since $e(t_{n-1}) + v(t_{n-1}) + 1 \leq 3n$, as can be seen by induction, the claim $L_n \leq 3^n n!$ follows by induction. Analyticity of $L(z)$ follows by comparison with a geometric series. □

Let $\overline{\mathcal{L}}_n$ denote the class of *pointed hierarchies* on n leaves,

$$\overline{\mathcal{L}}_n := \{(h,k) : h \text{ a hierarchy with } n \text{ leaves and } k \in [n]\},$$

and let $\overline{\mathcal{L}} := \bigcup_{n\geq 1} \overline{\mathcal{L}}_n$. Obviously, $\overline{L}_n := \#\overline{\mathcal{L}}_n = nL_n$. Less formally, a pointed hierarchy is a hierarchy together with a distinguished leaf. A pointed hierarchy (h,i) has a distinguished spinal path, which is the path in h from leaf i to the root. Erasing the edges in this path shatters h into a collection of subtrees called *bushes* that inherit a linear order from their position on the path. Conversely, a vertex labeled i for some $i \in [n]$ followed by a linearly ordered collection of bushes with distinct labels in $[n] \setminus \{k\}$ can be stitched together in the obvious manner to produce a pointed hierarchy.

Let us say a little more about bushes. Let (h,i) be a pointed hierarchy, and let v be a vertex different from leaf k in the spinal path in h from leaf k to root. When edges in this spinal path are erased, v will belong to a unique bush b. We define the *leaves of* b to be those vertices of b that are leaves in h. The degree of v in b is one if and only if v has two children in h, and in this case it is easily checked that the graph derived from b by

- declaring the unique neighbor of v in b to be root

- and then erasing v and the edge from v to its neighbor

is a hierarchy. Also, if the degree of v in the bush b is two or more, it is easily checked that the graph derived from b by declaring v to be root is a hierarchy. Furthermore, if b has exactly one leaf then the degree of v in b must be 1. Letting $\mathcal{B}_n$ denote the class of bushes on n leaves with leaves labeled by $[n]$ and $B_n := \#\mathcal{B}_n$, this discussion establishes that $B_1 = 1$ and

$$B_n = 2L_n \quad (n \geq 2).$$

Therefore

$$B(z) := \sum_{n=1}^{\infty} B_n \frac{z^n}{n!} = 2L(z) - z. \tag{4.23}$$

Proposition 30. *If $L_{n,k,1}$ denotes the number of hierarchies on n leaves where vertex 1 is at distance kc_n from the root then*

$$L_{n,k,1} = (n-1)![z^{n-1}]\,(B(z))^k \tag{4.24}$$

(recall that all edges in a hierarchy on n leaves have length c_n, so that a distance equal to kc_n in T_n corresponds to "graph distance" k).

Proof. From the discussion above relating pointed hierarchies and sequences of bushes, $L_{n,k,1}$ equals

$$\sum_{(n_1,\ldots,n_k)\vdash n-1} \binom{n-1}{n_1,\ldots,n_k} \prod_{i=1}^{k} B_{n_i} = (n-1)! \sum_{(n_1,\ldots,n_k)\vdash n-1} \prod_{i=1}^{k} \frac{B_{n_i}}{n_i!}$$

where the sums are over all sequences $(n_1,\ldots,n_k)$ of positive integers with sum $n-1$. The right-hand-side of the equality above is evidently $(n-1)![z^{n-1}]\,(B(z))^k$, since $B_0 = 0$. □

Proposition 31. *There are the following equalities of formal power series:*

(a) $L(z) = \exp(L(x)) - L(x) - 1 + x$

(b) $B(x) = G(x, B(x))$ *for* $G(z,w) := 2(\exp(z/2 + w/2) - w/2 - 1)$.

Proof. Let h be a hierarchy having at least two vertices, and $k \geq 2$ be the degree of the root of h. Erasing the root of h and all edges attached to the root shatters h into a forest of k subtrees. For every subtree h' derived in this manner, declare the root of h' to be the vertex of h' that was adjacent to the root in h. Each of these subtrees h' has all the properties of a hierarchy except that the union of leaf labels in h' may not be a consecutive set of integers $\{1,\ldots,m\}$. Conversely any collection of k such subtrees can be glued together in the obvious manner to form hierarchy on n leaves in which the root has degree k, provided that the leaves of the subtrees are have distinct labels and jointly constitute the set $[n]$. From standard combinatorial facts it then follows that $n![z^n]\frac{1}{k!}(L(z))^k$ is the number of hierarchies on n leaves in which the root has degree k for $k \geq 2$. There is only one hierarchy on one leaf, and it is the only hierarchy in which the root has degree 1. It follows that

$$L(z) = z + \sum_{k=2}^{\infty} \frac{(L(z))^k}{k!},$$

which proves (a). Using (4.23) in conjunction with (a) yields (b):

$$\begin{aligned}
B(x) &= 2L(x) - x \\
&= 2(\exp(L(x)) - L(x) - 1 + x) - x \\
&= 2(\exp((2L(x) - x)/2 + x/2) - (2L(x) - x)/2 - 1 - x/2 + x) - x \\
&= 2(\exp(B(x)/2 + x/2) - B(x)/2 - 1) \\
&= G(x, B(x))
\end{aligned}$$

□

Definition 18. *Let* $y(z) = \sum_{n\geq 0} y_n z^n$ *be a function complex-analytic at 0, with* $y_0 = 0$ *and* $y_n \geq 0$ *for all* $n \geq 0$. *Then* y *is said to* belong to the smooth implicit-function schema defined by G *if*

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

for z *in a neighborhood of zero, where* $G(z, w) = \sum_{m,n\geq 0} g_{m,n} z^m w^n$ *is a bivariate function analytic in a domain* $|z| < R$, $|w| < S$, *for some* $R, S > 0$, *additionally satisfying the following conditions:*

- *The coefficients of* G *satisfy* $g_{m,n} \geq 0$ *for all* $m, n \geq 0$; $g_{0,0} = 0$, $g_{0,1} \neq 1$, *and* $g_{m,n} > 0$ *for some* $n \geq 2$, $m \geq 0$.
- *There exist two numbers* r, s *such that* $0 < r < R$ *and* $0 < s < S$, *satisfying the system of equations*

  $$G(r, s) = s \quad G_w(r, s) = 1$$

  which is called the characteristic system.

Here and henceforth, subscripts denote differentiation, so e.g. $G_{ww}(w_0, z_0)$ is the second derivative of G with respect to its first argument, evaluated at the point (w_0, z_0). For the next theorem, recall that a generating function or power series $\sum_n y_n z^n$ is called *aperiodic* if there exist three indices i, j, k such that $y_i y_j y_k \neq 0$ and $\gcd(i, j, k) = 1$.

Definition 19 ([44] p 389)**.** *For real numbers* $R > 1$ *and* $\phi \in (0, \pi/2]$, *let* $\Delta(\phi, R)$ *be the following subset of the complex plane:*

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

A domain is a Δ*-domain at* $\zeta \in \mathbb{C}$ *if it is the image under* $z \mapsto z\zeta$ *of* $\Delta(R, \phi)$ *for some* R *and* ϕ *as above.*

Theorem 17 ([44], Theorem VII.3)**.** *Let* $y(z)$ *belong to the* implicit function schema *defined by* $F(z, w)$ *with* (r, s) *the positive solution of the characteristic system. Then* $y(z)$ *converges at* $z = r$, *where it has a square-root singularity, meaning*

$$y(z) = s - \gamma\sqrt{1 - z/r} + O(1 - z/r), \tag{4.25}$$

as $z \to r$ *within a* Δ*-domain, where* $\gamma := \sqrt{\frac{2\,z_0 F_z(z_0,w_0)}{F_{ww}(z_0,w_0)}}$. *Additionally, if* $y(z)$ *is aperiodic, then*

$$[z^n]y(z) \sim \frac{\gamma}{2\sqrt{\pi n^3}} r^{-n} \left(1 + O(n^{-1})\right) \qquad \text{as } n \to \infty. \tag{4.26}$$

A proof of Theorem 17 may be found in [44][Theorem VII.3].

Corollary 5. *With $r := 2\log 2 - 1$ and $\gamma = 2(2\log 2 - 1)^{1/2}$, and B defined by (4.23)*

$$B(z) = 1 - \gamma(1 - z/r)^{1/2} + R(z) \tag{4.27}$$

for a function $R(z)$ analytic in a Δ domain at z that is $O(1 - z/r)$ as $z \to r$, in this Δ-domain. Furthermore,

$$[z^n]B(z) = \frac{1}{\sqrt{\pi n^3}}(2\log 2 - 1)^{-n+1/2}\left(1 + O(n^{-1})\right) \tag{4.28}$$

as $n \to \infty$.

Proof. From Proposition 29 and (4.23) function $B(z)$ is analytic in a neighborhood of zero; furthermore B is obviously aperiodic. Proposition 31 (b) shows that B belongs to the smooth implicit function schema defined by $G(z, w) = 2(\exp(z/2 + w/2) - w/2 - 1)$. It is easily checked that the solution of the characteristic system $G(r, s) = s,\ G_w(r, s) = 1$ is $s = 1$, $r = 2\log 2 - 1$. The claim then follows from Theorem 17. □

Theorem 18 ([44], Theorem IX.16)**.** *If $y(z)$ satisfies the hypotheses of Theorem 17 then for any compact subset $F \subseteq (0, \infty)$, any sequence $x = x_n$ contained in F, and and $k := [xn^{1/2}]$, the coefficient of z^n in $y(z)^k$ admits the following asymptotic estimate:*

$$[z^n]y(z)^k \sim s^k r^{-n}\frac{1}{n}\frac{x\gamma/s}{\sqrt{\pi}}\exp(-\frac{1}{2}x^2\gamma^2/s^2), \tag{4.29}$$

for r, s, γ the constants appearing in Theorem 17 and $[u]$ the nearest integer to $u \in \mathbb{R}$.

This is the "$\lambda = 1/2$" instance of Theorem IX.16 of [44]. See also [14].

Proposition 32. *If T_n is is a uniform random hierarchy and $D_1(n)$ denotes the number of vertices on the path from the leaf labeled 1 to the root, then*

$$\sqrt{n}\mathbb{P}(D_1(n) = [xn^{1/2}]) \to c\,x\exp\left(-\frac{1}{2}cx^2\right) \tag{4.30}$$

as $n \to \infty$, where $c = 4(2\log(2) - 1)$ and $[u]$ denotes the nearest integer to $u \in \mathbb{R}$.

Proof of Proposition 23. According to (4.23), $L_n = \frac{1}{2}B_n$ for $n \geq 2$. Making use of Proposition 30, we have

$$\sqrt{n}\mathbb{P}(X_1(n) = [xk]) = \frac{L_{n,k,1}}{L_n} = \frac{\sqrt{n}(n-1)![z^{n-1}]\left(B(z)\right)^k}{\frac{1}{2}n![z^n]B(z)}. \tag{4.31}$$

The claim then follows from (4.28) and (4.29). □

Proof of Proposition 23. All branches in T_n have length $2(2\log 2 - 1)^{-1/2}n^{1/2}$ so the result follows from Proposition 32 and Scheffé's Lemma [88, p 57] after convolving T_n with an independent Uniform$[0, c_n]$ random variable. □

Bibliography

[1] Ryan Adams, Zoubin Ghahramani, and Michael Jordan. Tree-structured stick breaking for hierarchical data. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 19–27. 2010.

[2] David Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.

[3] 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.

[4] David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[5] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25(2):812–854, 1997.

[6] David Aldous. More uses of exchangeability: Representations of complex random structures. In N.H. Bingham and C.M. Goldie, editors, *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*, pages 35–63. Cambridge University Press, 2010.

[7] David Aldous, Maxim Krikun, and Lea Popovic. Stochastic models for phylogenetic trees on higher-order taxa. *J. Math. Biol.*, 56(4):525–557, 2008.

[8] David Aldous and Jim Pitman. A family of random trees with random edge lengths. *Random Structures Algorithms*, 15(2):176–195, 1999.

[9] David Aldous and Jim Pitman. Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. *Probab. Th. Rel. Fields*, 118:455–482, 2000.

[10] David J. Aldous. Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.*, 11(4):581–598, 1981.

[11] David J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985.

[12] David J. Aldous. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Statist. Sci.*, 16(1):23–34, 2001.

[13] Tim Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.*, 5:80–145, 2008.

[14] Cyril Banderier, Philippe Flajolet, Gilles Schaeffer, and Michèle Soria. Random maps, coalescing saddles, singularity analysis, and Airy phenomena. *Random Structures Algorithms*, 19(3-4):194–246, 2001. Analysis of algorithms (Krynica Morska, 2000).

[15] Federico Bassetti and Eugenio Regazzini. The unsung de Finetti's first paper about exchangeability. *Rend. Mat. Appl. (7)*, 28(1):1–17, 2008.

[16] Julien Berestycki. Ranked fragmentations. *ESAIM Probab. Statist.*, 6:157–175 (electronic), 2002.

[17] Julien Berestycki, Nathanaël Berestycki, and Jason Schweinsberg. Small-time behavior of beta coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44(2):214–238, 2008.

[18] Jean Bertoin. A fragmentation process connected to Brownian motion. *Probab. Theory Related Fields*, 117(2):289–301, 2000.

[19] Jean Bertoin. Eternal additive coalescents and certain bridges with exchangeable increments. *Ann. Probab.*, 29(1):344–360, 2001.

[20] Jean Bertoin. Homogeneous fragmentation processes. *Probab. Theory Related Fields*, 121(3):301–318, 2001.

[21] Jean Bertoin. Self-similar fragmentations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(3):319–340, 2002.

[22] Jean Bertoin. Random covering of an interval and a variation of Kingman's coalescent. *Random Structures Algorithms*, 25(3):277–292, 2004.

[23] Jean Bertoin. *Random fragmentation and coagulation processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

[24] Jean Bertoin. Homogeneous multitype fragmentations. In *In and out of equilibrium. 2*, volume 60 of *Progr. Probab.*, pages 161–183. Birkhäuser, Basel, 2008.

[25] Jean Bertoin and Jean-François Le Gall. Stochastic flows associated to coalescent processes. *Probab. Theory Related Fields*, 126(2):261–288, 2003.

[26] Jean Bertoin and Alain Rouault. Discretization methods for homogeneous fragmentations. *J. London Math. Soc. (2)*, 72(1):91–109, 2005.

[27] D. M. Blei, T. L. Griffiths, and M. I. Jordan. The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. *Journal of the ACM*, 57:1–30, 2010.

[28] C. Blundell, Y. W. Teh, and K. A. Heller. Bayesian rose trees. In *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*, 2010.

[29] Michael Camarri and Jim Pitman. Limit distributions and random trees derived from the birthday problem with unequal probabilities. *Electron. J. Probab.*, 5:no. 2, 18 pp. (electronic), 2000.

[30] Bruno de Finetti. La prévision : ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré*, 7(1):1–68, 1937.

[31] P. Diaconis and D. Freedman. de Finetti's theorem for Markov chains. *Ann. Probab.*, 8(1):115–130, 1980.

[32] P. Diaconis and D. Freedman. Finite exchangeable sequences. *Ann. Probab.*, 8(4):745–764, 1980.

[33] Persi Diaconis and Svante Janson. Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)*, 28(1):33–61, 2008.

[34] Peter Donnelly. The heaps process, libraries, and size-biased permutations. *J. Appl. Probab.*, 28(2):321–335, 1991.

[35] Peter Donnelly, Steven N. Evans, Klaus Fleischmann, Thomas G. Kurtz, and Xiaowen Zhou. Continuum-sites stepping-stone models, coalescing exchangeable partitions and random trees. *Ann. Probab.*, 28(3):1063–1110, 2000.

[36] Peter Donnelly and Paul Joyce. Consistent ordered sampling distributions: characterization and convergence. *Adv. in Appl. Probab.*, 23(2):229–258, 1991.

[37] Peter Donnelly and Thomas G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.*, 27(1):166–205, 1999.

[38] David Draper, James Hodges, Colin Mallows, and Daryl Pregibon. Exchangeability and data analysis. *J R Stat Soc Ser A*, 156(1):9–37, 1993.

[39] Thomas Duquesne and Jean-François Le Gall. Random trees, Lévy processes and spatial branching processes. *Astérisque*, (281):vi+147, 2002.

[40] Richard Durrett. *Probability: theory and examples.* Thomson, Belmont, CA, third edition, 2005.

[41] Steven N. Evans. Kingman's coalescent as a random metric space. In *Stochastic models (Ottawa, ON, 1998)*, volume 26 of *CMS Conf. Proc.*, pages 105–114. Amer. Math. Soc., Providence, RI, 2000.

[42] Steven N. Evans. *Probability and real trees*, volume 1920 of *Lecture Notes in Mathematics.* Springer, Berlin, 2008. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005.

[43] Steven N. Evans and Anita Winter. Subtree prune and regraft: a reversible real tree-valued Markov process. *Ann. Probab.*, 34(3):918–961, 2006.

[44] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics.* Cambridge University Press, Cambridge, 2009.

[45] David Freedman and Persi Diaconis. de Finetti's theorem for symmetric location families. *Ann. Statist.*, 10(1):184–189, 1982.

[46] Alexander Gnedin, Chris Haulk, and Jim Pitman. Characterizations of exchangeable partitions and random discrete distributions by deletion properties. In N.H. Bingham and C.M. Goldie, editors, *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*, pages 264–298. Cambridge University Press, 2010.

[47] Alexander Gnedin, Alex Iksanov, and Martin Möhle. On asymptotics of exchangeable coalescents with multiple collisions. *J. Appl. Probab.*, 45(4):1186–1195, 2008.

[48] Alexander Gnedin and Grigori Olshanski. A q-analogue of de Finetti's theorem. *Electron. J. Combin.*, 16(1):Research Paper 78, 16, 2009.

[49] Alexander V. Gnedin. On a class of exchangeable sequences. *Statist. Probab. Lett.*, 25(4):351–355, 1995.

[50] Alexander V. Gnedin. The representation of composition structures. *Ann. Probab.*, 25(3):1437–1450, 1997.

[51] Alexander V. Gnedin. On convergence and extensions of size-biased permutations. *J. Appl. Probab.*, 35(3):642–650, 1998.

[52] Christina Goldschmidt and James B. Martin. Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron. J. Probab.*, 10:no. 21, 718–745 (electronic), 2005.

[53] Andreas Greven, Peter Pfaffelhuber, and Anita Winter. Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probab. Theory Related Fields*, 145(1-2):285–322, 2009.

[54] B. Haas and G. Miermont. Scaling limits of Markov branching trees, with applications to Galton-Watson and random unordered trees. *ArXiv e-prints*, March 2010.

[55] 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.

[56] 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.

[57] Bénédicte Haas, Jim Pitman, and Matthias Winkel. Spinal partitions and invariance under re-rooting of continuum random trees. *Ann. Probab.*, 37(4):1381–1411, 2009.

[58] K. A. Heller and Z. Ghahramani. Bayesian hierarchical clustering. In Luc De Raedt and Stefan Wrobel, editors, *Proceedings of the 22nd International Conference on Machine Learning*. ACM Press, 2005.

[59] Edwin Hewitt and Leonard J. Savage. Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.*, 80:470–501, 1955.

[60] Ulrich Hirth. Exchangeable random ordered trees by positive definite functions. *J. Theoret. Probab.*, 16(2):339–344, 2003.

[61] Ulrich Hirth and Paul Ressel. Random partitions by semigroup methods. *Semigroup Forum*, 59(1):126–140, 1999.

[62] Ulrich Hirth and Paul Ressel. Exchangeable random orders and almost uniform distributions. *J. Theoret. Probab.*, 13(3):609–634, 2000.

[63] Saul Jacka and Jon Warren. Random orderings of the integers and card shuffling. *Stochastic Process. Appl.*, 117(6):708–719, 2007.

[64] Svante Janson. Poset limits and exchangeable random posets. *Institut Mittag-Leffler preprint 02*, 2009.

[65] Olav Kallenberg. Exchangeable random measures in the plane. *J. Theoret. Probab.*, 3(1):81–136, 1990.

[66] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[67] Olav Kallenberg. *Probabilistic symmetries and invariance principles*. Probability and its Applications (New York). Springer, New York, 2005.

[68] S. Kerov. Coherent random allocations, and the Ewens-Pitman formula. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 325(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 12):127–145, 246, 2005.

[69] J. F. C. Kingman. Random partitions in population genetics. *Proc. Roy. Soc. London Ser. A*, 361(1704):1–20, 1978.

[70] J. F. C. Kingman. The representation of partition structures. *J. London Math. Soc. (2)*, 18(2):374–380, 1978.

[71] J. F. C. Kingman. The coalescent. *Stochastic Process. Appl.*, 13(3):235–248, 1982.

[72] J. F. C. Kingman. Exchangeability and the evolution of large populations. In *Exchangeability in probability and statistics (Rome, 1981)*, pages 97–112. North-Holland, Amsterdam, 1982.

[73] J. F. C. Kingman, S. J. Taylor, A. G. Hawkes, A. M. Walker, David Roxbee Cox, A. F. M. Smith, B. M. Hill, P. J. Burville, and T. Leonard. Random discrete distribution. *J. Roy. Statist. Soc. Ser. B*, 37:1–22, 1975. With a discussion by S. J. Taylor, A. G. Hawkes, A. M. Walker, D. R. Cox, A. F. M. Smith, B. M. Hill, P. J. Burville, T. Leonard and a reply by the author.

[74] Steffen L. Lauritzen. *Graphical models*, volume 17 of *Oxford Statistical Science Series*. The Clarendon Press Oxford University Press, New York, 1996. Oxford Science Publications.

[75] Peter McCullagh, Jim Pitman, and Matthias Winkel. Gibbs fragmentation trees. *Bernoulli*, 14(4):988–1002, 2008.

[76] E.W. Meeds, D.A. Ross, R.S. Zemel, and S.T. Roweis. Learning stick-figure models using nonparametric bayesian priors over trees. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 –8, june 2008.

[77] Grégory Miermont. Self-similar fragmentations derived from the stable tree. I. Splitting at heights. *Probab. Theory Related Fields*, 127(3):423–454, 2003.

[78] Martin Möhle and Serik Sagitov. A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.*, 29(4):1547–1562, 2001.

[79] Robin Pemantle. Automorphism invariant measures on trees. *Ann. Probab.*, 20(3):1549–1566, 1992.

[80] Mihael Perman, Jim Pitman, and Marc Yor. Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields*, 92(1):21–39, 1992.

[81] 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, With a foreword by Jean Picard.

[82] Jim Pitman. Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields*, 102(2):145–158, 1995.

[83] Jim Pitman. Random discrete distributions invariant under size-biased permutation. *Adv. in Appl. Probab.*, 28(2):525–539, 1996.

[84] Jim Pitman. Poisson-Kingman partitions. In *Statistics and science: a Festschrift for Terry Speed*, volume 40 of *IMS Lecture Notes Monogr. Ser.*, pages 1–34. Inst. Math. Statist., Beachwood, OH, 2003.

[85] Jim Pitman and Douglas Rizzolo. Schröder's problems and scaling limits of random trees. *Preprint*, 2011.

[86] Jim Pitman and Matthias Winkel. Regenerative tree growth: binary self-similar continuum random trees and Poisson-Dirichlet compositions. *Ann. Probab.*, 37(5):1999–2041, 2009.

[87] Jim Pitman and Marc Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25(2):855–900, 1997.

[88] David Pollard. *A user's guide to measure theoretic probability*, volume 8 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2002.

[89] Paul Ressel. Subdiagonal and almost uniform distributions. *Electron. Comm. Probab.*, 7:97–101 (electronic), 2002.

[90] Paul Ressel. Exchangeability and semigroups. *Rend. Mat. Appl. (7)*, 28(1):63–81, 2008.

[91] Serik Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, 36(4):1116–1125, 1999.

[92] Mattias Winkel and Bo Chen. Restricted exchangeable partitions and embedding of associated hierarchies in continuum random trees. *Preprint on arxiv.org : arXiv:0911.5647v1*, 2009.