Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2013, Vol. 49, No. 3, 839–872
DOI: 10.1214/12-AIHP533

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Restricted exchangeable partitions and embedding of associated hierarchies in continuum random trees

Bo Chen and Matthias Winkel

University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK. E-mail: chen@stats.ox.ac.uk; winkel@stats.ox.ac.uk

Received 31 August 2010; revised 1 August 2012; accepted 20 October 2012

Abstract. We introduce the notion of a restricted exchangeable partition of $\mathbb{N}$. We obtain integral representations, consider associated fragmentations, embeddings into continuum random trees and convergence to such limit trees. In particular, we deduce from the general theory developed here a limit result conjectured previously for Ford's alpha model and its extension, the alpha-gamma model, where restricted exchangeability arises naturally.

Résumé. Nous introduisons la notion d'une partition restreinte échangeable de $\mathbb{N}$. Nous obtenons des représentations intégrales, nous considérons les fragmentations associées, des plongements dans des arbres aléatoires continus et la convergence vers de tels arbres limites. En particulier, nous déduisons de la théorie générale développée içi un résultat limite formulé en conjecture dans un travail précédent. Ce résultat particulier concerne les arbres alpha de Ford et leurs généralisations, les arbres alpha-gamma, deux exemples où l'échangeabilité restreinte arrive de manière naturelle.

MSC: 60G09; 60J80

Keywords: Exchangeability; Hierarchy; Coalescent; Fragmentation; Continuum random tree; Renewal theory

1. Introduction

This paper introduces the concept of restricted exchangeability, which captures a weak form of exchangeability that occurs naturally in models such as the alpha-gamma tree model of [10].

1.1. *Motivating example*: *Alpha-gamma trees as random hierarchies*

An important motivation for this paper is the study of the limiting behaviour of the alpha-gamma tree-growth model [10], which is based on a simple stochastic growth rule to build a tree T_{n+1} from a tree T_n by adding a leaf (degree-1 vertex) labelled $n+1$. Let us specify this rule in a framework of hierarchies (also called total partitions or fragmentations in the literature).

Following [20,24,29,30], we call *hierarchy on* $B \subseteq \mathbb{N}$ any subset $\mathbf{t}_B$ of the power set of B such that $B \in \mathbf{t}_B$ and $\{j\} \in \mathbf{t}_B$ for all $j \in B$, and so that for every $A, A' \in \mathbf{t}_B$, either $A \subseteq A'$ or $A' \subseteq A$ or $A \cap A' = \varnothing$. To avoid trivialities, we also require $\varnothing \in \mathbf{t}_B$. We say that a strict subset $A \in \mathbf{t}_B$ of $A' \in \mathbf{t}_B$ is *a maximal subset of A' in $\mathbf{t}_B$* if for all $A'' \in \mathbf{t}_B$ with $A \subseteq A'' \subseteq A'$ either $A = A''$ or $A'' = A'$. For finite $B \subset \mathbb{N}$ with $\#B \geq 2$, the maximal subsets $A_1, \ldots, A_k$ of B in $\mathbf{t}_B$ form a partition of B and the *restrictions* $\mathbf{t}_{A_i} = \mathbf{t}_B \cap A_i = \{A \cap A_i \colon A \in \mathbf{t}_B\}$ are hierarchies on A_i, $i \in [k] := \{1, \ldots, k\}$; a hierarchy $\mathbf{t}_B$ fully encodes a *rooted tree*, i.e. a connected acyclic graph, with vertex set $\mathbf{t}_B$ and edge relation linking each set to its maximal non-empty subsets, with *root* $\varnothing$ related to B; hierarchies $\mathbf{t}_{A_i}$ are the *subtrees* of $\mathbf{t}_B$ above the *first branchpoint* B of $\mathbf{t}_B$. See Fig. 1. We call $A \in \mathbf{t}_B$ *branchpoint* or *internal vertex* if $\#A \geq 2$. Denote by $\mathbb{T}_n$ the set of all hierarchies on $[n]$, $n \geq 1$. We say that $\mathbf{t}_n \in \mathbb{T}_n$ and $\mathbf{t}_{n+1} \in \mathbb{T}_{n+1}$ are *consistent* if $\mathbf{t}_n = \mathbf{t}_{n+1} \cap [n]$.

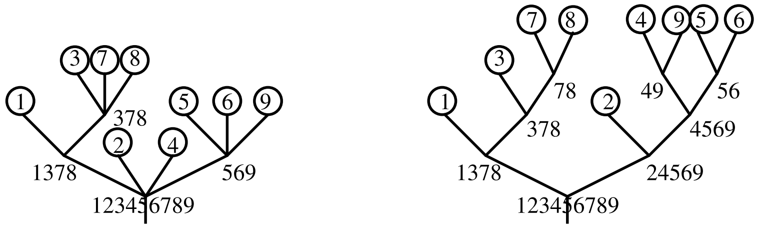


Fig. 1. Two hierarchies on $B = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ illustrated as rooted trees.

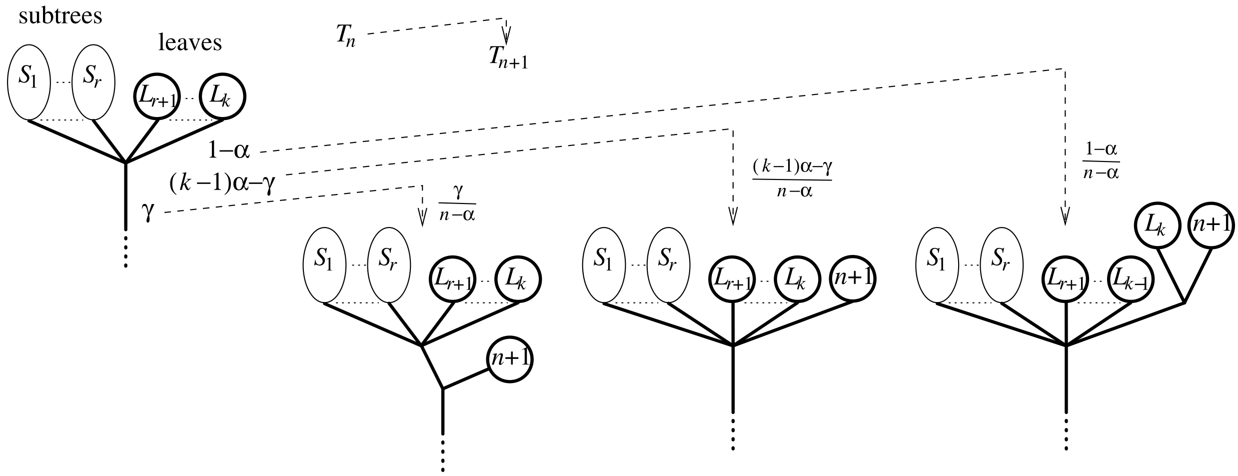


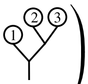
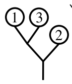
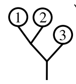
Fig. 2. Alpha-gamma growth rule: displayed is one internal vertex, B say, of T_n with degree $k+1$, hence vertex weight $(k-1)\alpha - \gamma$, with $k-r$ leaves $L_{r+1}, \dots, L_k \in [n]$ and r bigger subtrees $S_1, \dots, S_r$; all edges also carry weights, weight $1-\alpha$ and γ are displayed here for the *leaf edge* below $\{L_k\}$ and the *inner edge* below B only; the three associated possibilities for T_{n+1} are displayed.

The alpha-gamma model [10] is a consistent family $(T_n, n \ge 1)$ of random hierarchies on $[n]$, for which the conditional distributions of T_{n+1} given T_n are particularly simple. In terms of trees, passing from T_n to T_{n+1} means identifying the random place in T_n where $\{n+1\}$ connects to T_n: as illustrated in Fig. 2, for parameters $0 \le \gamma \le \alpha \le 1$ and for $n \ge 1$, vertex $\{n+1\}$ connects to

- a new vertex $\{j, n+1\}$ inserted (in the edge) below $\{j\} \in T_n$ with probability $(1-\alpha)/(n-\alpha)$;
- a new vertex $B \cup \{n+1\}$ inserted below branchpoint $B \in T_n$ with probability $\gamma/(n-\alpha)$;
- an existing branchpoint $B \in T_n$ with probability $((k-1)\alpha - \gamma)/(n-\alpha)$, where $k+1$ is the degree of vertex B in the tree T_n, or equivalently k is the number of blocks of the partition into maximal subsets $A_1, \dots, A_k$ of B in the hierarchy T_n;

now T_{n+1} is built from T_n by adding $n+1$ to all vertices on the path between $\{n+1\}$ and $\varnothing$.

A random hierarchy T_B on B is called *exchangeable* [20] if for every bijection $\beta : B \to B$, the hierarchy $\beta(T_B) = \{\{\beta(j)\colon\ j \in A\}, A \in T_B\}$ obtained by permuting labels by β is distributed like T_B. An alpha-gamma tree T_n for $n \ge 3$ is exchangeable iff $\gamma = 1-\alpha$; note for instance that

$\mathbb{P}\Big(T_3 =$  $\Big) = \mathbb{P}\Big(T_3 =$  $\Big) = \dfrac{1-\alpha}{2-\alpha}$ while $\dfrac{\gamma}{2-\alpha} = \mathbb{P}\Big(T_3 =$  $\Big)$.

However, for $\gamma \ne 1-\alpha$ there is still some exchangeability. To capture this, we introduce the partition Π_n of T_n into maximal strict subsets of $[n]$ and refer to its distribution P_n on the set $\mathcal{P}_n$ of partitions of $[n]$ as a *splitting rule*. We say that $(T_n, n \ge 1)$ is a *labelled Markov branching model* if conditionally given $\Pi_n = \{A_1, \dots, A_k\}$, the hierarchies

$T_n \cap A_i$, $i \in [k]$, are independent and distributed as $\beta_i(T_{\#A_i})$, where β_i is the unique increasing bijection from $[\#A_i]$ to A_i. Then $(P_n, n \geq 2)$ determines the distributions of T_n, $n \geq 1$. We will show in Section 6 that the alpha-gamma model is a labelled Markov branching model with splitting rules $P_n^{\alpha,\gamma}$ satisfying

$$P_n^{\alpha,\gamma}(\pi) = P_n^{\alpha,\gamma}\big(\beta(\pi)\big) \quad \text{for all bijections } \beta : [n] \to [n] \text{ with } \pi \cap \{1,2\} = \beta(\pi) \cap \{1,2\},$$

where $\beta(\pi) = \{\{\beta(j) : j \in A\}, A \in \pi\}$. Equivalently, $P_n^{\alpha,\gamma}$ satisfies $P_n^{\alpha,\gamma}(\pi) = P_n^{\alpha,\gamma}(\pi')$ if $\pi \cap \{1,2\} = \pi' \cap \{1,2\}$ and if $\pi = \{A_1, \dots, A_k\}$ and $\pi' = \{A'_1, \dots, A'_k\}$ have the same multiset of block sizes $\#A_i$, $i \in [k]$, and $\#A'_j$, $j \in [k]$.

Alpha-gamma trees $(T_n, n \geq 1)$ give rise to a random hierarchy $\mathcal{H} = \{A \subset \mathbb{N} : A \cap [n] \in T_n \text{ for all } n \geq 1\}$ on $\mathbb{N}$. We studied the limiting behaviour of T_n and identified a scaling limit in [10], but only obtained convergence in distribution. The crucial tool to strengthen to convergence in probability is restricted exchangeability, which we will use to embed $\mathcal{H}$ and more general hierarchies of (restricted exchangeable) Markov branching models into suitable limit trees.

1.2. *Restricted exchangeable partitions and integral representations*

For a partition $\pi = \{\pi_i, i \in \mathbb{N}\}$ of $B \subseteq \mathbb{N}$ with disjoint π_i, $i \in \mathbb{N}$, each non-empty $\pi_i \subseteq B$ is called a *block* of π. When π has only finitely many blocks, we often omit $\varnothing$ from π. To be definite, we arrange the blocks of π in the order of least element, i.e. $\min \pi_i < \min \pi_j$ for every $i < j$, followed by $\varnothing$ with the convention $\min \varnothing = \infty$. For finite π_i, we consider the *block size* $\#\pi_i$. We denote the set of all partitions of B by $\mathcal{P}_B$. Recall $[n] = \{1, \dots, n\}$ for $n \in \mathbb{N}$. Note that for $\Gamma \in \mathcal{P} = \mathcal{P}_{\mathbb{N}}$, the restrictions $\Gamma|_n = \Gamma \cap [n] = \{\Gamma_i \cap [n], i \in \mathbb{N}\}$ are partitions of $[n]$, $n \in \mathbb{N}$. On $\mathcal{P}$, consider the metric $d(\Gamma, \Gamma') = 2^{-\inf\{n \geq 1 : \Gamma|_n \neq \Gamma'|_n\}}$ and the associated Borel σ-algebra.

Following de Finetti and Kingman, we call a Borel measure on the space $\mathcal{P}_B$ of partitions of $B \subseteq \mathbb{N}$ *exchangeable*, if it is invariant under the natural action on $\mathcal{P}_B$ of the symmetric group on B; and a random partition is called exchangeable if its distribution is exchangeable. Then a measure μ on $\mathcal{P}$ is exchangeable if and only if the discrete measures μ_n on $\mathcal{P}_n = \mathcal{P}_{[n]}$, given by

$$\mu_n\big(\{\pi\}\big) = \mu\big(\mathcal{P}^\pi\big), \quad \pi \in \mathcal{P}_n, \text{ where } \mathcal{P}^\pi = \{\Gamma \in \mathcal{P} : \Gamma|_n = \pi\}, \tag{1}$$

are exchangeable for all $n \geq 1$. Furthermore, a measure μ_n on $\mathcal{P}_n$ is exchangeable if $\mu_n(\{\pi\}) = \mu_n(\{\pi'\})$ for all $\pi, \pi' \in \mathcal{P}_n$ with the same multiset of block sizes.

Several weaker forms of exchangeability have been studied in the literature, notably Pitman's partial exchangeability [26] and Gnedin's constrained exchangeability [13]. We introduce here a new weak form of exchangeability and discuss in Section 3.1 how these notions interact.

Definition 1. *For $\pi \in \mathcal{P}_n$, we call a measure μ on $\mathcal{P}^\pi$* exchangeable on $\mathcal{P}^\pi$ *if $\mu(\mathcal{P}^{\pi'}) = \mu(\mathcal{P}^{\pi''})$ for all $\pi', \pi'' \in \bigcup_{m \geq 1} \mathcal{P}_{n+m}$ with the same multiset of block sizes and with $\pi' \cap [n] = \pi'' \cap [n] = \pi$.*

A measure μ on $\mathcal{P}$ is called restricted exchangeable (*RE*) *if there is $\mathcal{C} \subset \mathcal{K} := \bigcup_{n \geq 1} \mathcal{P}_n$ s.th.*

- *no $\pi \in \mathcal{C}$ is the restriction of another $\pi' \in \mathcal{C}$,*
- *the measure μ is carried by $\bigcup_{\pi \in \mathcal{C}} \mathcal{P}^\pi$, i.e. $\mu(\mathcal{P} \setminus \bigcup_{\pi \in \mathcal{C}} \mathcal{P}^\pi) = 0$,*
- *and for each $\pi \in \mathcal{C}$, the restriction of μ to $\mathcal{P}^\pi$ is finite and exchangeable on $\mathcal{P}^\pi$.*

Remark 2. *A measure on $\mathcal{P}^\pi$ is exchangeable on $\mathcal{P}^\pi$ if and only if $\mu(\mathcal{P}^{\pi'}) = \mu(\mathcal{P}^{\beta(\pi')})$ for all $\pi' \in \mathcal{P}_{n+m}$ and all bijections $\beta : [n+m] \to [n+m]$ with $\pi' \cap [n] = \beta(\pi') \cap [n] = \pi$, $m \geq 1$.*

Note that the set of admissible bijections β depends on π', and while $\beta(j) = j$, $j \in [n]$, makes β admissible, there are many other admissible bijections. The point is that the specific blocks containing π_i in π' and $\beta(\pi')$ may have different sizes (*while the multisets of all block sizes of π' and $\beta(\pi')$ coincide*). *This is an important feature of our definition of restricted exchangeability. The apparently more natural but strictly weaker concept obtained by restricting the admissible bijections to the subgroup of those with $\beta(j) = j$, $j \in [n]$, is less convenient to work with, since integral representations of such measures – which we might call* weakly *RE – no longer just involve measures on decreasing sequences*, *cf. Theorem 3.*

Let $S^{\downarrow} = \{\mathbf{s} = (s_i, i \ge 1)\colon s_1 \ge s_2 \ge \cdots \ge 0, \sum_{i\ge 1} s_i \le 1\}$. For $\mathbf{s} \in S^{\downarrow}$, *Kingman's paintbox* [22] is obtained from independent random variables $(\xi_r, r \ge 1)$ with respective distributions

$$\mathbb{P}(\xi_r = i) = s_i, \quad i \ge 1, \qquad \mathbb{P}(\xi_r = -r) = s_0 := 1 - \sum_{i\ge 1} s_i,$$

as the distribution $\kappa_{\mathbf{s}}$ on $\mathcal{P}$ of the exchangeable partition $\Pi = \{\{n \ge 1\colon \xi_n = i\}, i \in \mathbb{Z}\}$, which puts any two $m, n \in \mathbb{N}$ into the same block if and only if $\xi_m = \xi_n$. By the Strong Law of Large Numbers, the vector of block sizes $\#(\Pi \cap [n])^{\downarrow}$ in decreasing order of size has *asymptotic frequencies*

$$|\Pi|^{\downarrow} = \lim_{n\to\infty} \frac{1}{n} \#\big(\Pi \cap [n]\big)^{\downarrow} = (s_i, i \ge 1) = \mathbf{s}.$$

It is well-known [1,21,22] that exchangeable measures on $\mathcal{P}$ admit integral representations $\mu = \int_{S^{\downarrow}} \kappa_{\mathbf{s}} \nu(\mathrm{d}\mathbf{s})$. To establish integral representations for RE measures here, we introduce *modified paintboxes* $\kappa_{\mathbf{s}}^{\pi}$, $\pi \in \mathcal{K} = \bigcup_{n\ge 1} \mathcal{P}_n$, by conditioning $\kappa_{\mathbf{s}}$ on the cylinder set $\mathcal{P}^{\pi} = \{\Gamma \in \mathcal{P}\colon \Gamma|_n = \pi\}$ of π in $\mathcal{P}$, but note that this conditioning is degenerate in some cases; see Section 2 for details.

Theorem 3 (Integral representation). *Let μ be a measure on $\mathcal{P}$. Then μ is RE if and only if there are a subset $\mathcal{C} \subset \mathcal{K}$ such that no $\pi \in \mathcal{C}$ is the restriction of another $\pi' \in \mathcal{C}$, and for each $\pi \in \mathcal{C}$ a finite measure ν_{π} on $S^{\downarrow}$ such that*

$$\mu = \sum_{\pi\in\mathcal{C}} \int_{S^{\downarrow}} \kappa_{\mathbf{s}}^{\pi} \nu_{\pi}(\mathrm{d}\mathbf{s}).$$

Note that a RE measure μ can be infinite, if $\mathcal{C}$ is infinite. However, as $\mathcal{C} \subset \mathcal{K}$ is countable, such infinite measures will still be σ-finite, because they are finite on $\mathcal{P}^{\pi}$, $\pi \in \mathcal{C}$.

Examples 4.

(i) *For $B \subseteq \mathbb{N}$, let $\mathbf{1}_B$ be the trivial partition of a single block B.* Dislocation measures *are measures on $\mathcal{P}$ carried by $\mathcal{P} \setminus \{\mathbf{1}_{\mathbb{N}}\}$, finite on $\mathcal{P}^{\pi}$, $\pi \in \mathcal{K} \setminus \{\mathbf{1}_{[n]}, n \ge 1\}$. We set $\mathcal{C} = \{\{[j], \{j+1\}\}, j \ge 1\}$ to naturally decompose $\mathcal{P} \setminus \{\mathbf{1}_{\mathbb{N}}\} = \bigcup_{\pi\in\mathcal{C}} \mathcal{P}^{\pi}$. Bertoin's* [6] *possibly infinite* exchangeable *dislocation measures, in the sense of* (1), *are exchangeable and finite on $\mathcal{P}^{\pi}$, $\pi \in \mathcal{C}$, so they satisfy Definition 1. See Sections 1.3 and 3.2.*

(ii) *We can associate dislocation measures with Ford's alpha model* [12] *and the alpha-gamma Markov branching model* [10], *defined in Section 1.1, so that $\mu_{\alpha,\gamma}(\mathcal{P}^{\pi}) = \lambda_n^{\alpha,\gamma} P_n^{\alpha,\gamma}(\pi)$, $\pi \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}$, for consistent rates $\lambda_n^{\alpha,\gamma}$, $n \ge 2$. These dislocation measures $\mu_{\alpha,\gamma}$ are RE, but not exchangeable, as we illustrated in terms of splitting rules $P_n^{\alpha,\gamma}$ at the end of Section 1.1. See Section 3.2 for an exploration of the relationship between splitting rules and dislocation measures in a general RE framework.*

From Theorem 3 we deduce an integral representation for restricted exchangeable dislocation measures. For simplicity we only allow as decomposition of $\mathcal{P}$ in Definition 1 the most relevant and natural $\mathcal{C} = \{\{[j], \{j+1\}\}, j \ge 1\}$.

Corollary 5. *Let κ be a RE measure with $\mathcal{C} = \{\{[j], \{j+1\}\}, j \ge 1\}$. Then for each $j \ge 1$, there are constants $c_j \ge 0$ and $k_j \ge 0$, and a measure ν_j on $S^{\downarrow}$ with*

$$\nu_j\big(\big\{(0,0,\ldots)\big\}\big) = \nu_j\big(\big\{(1,0,\ldots)\big\}\big) = 0 \quad \text{and} \quad \int_{S^{\downarrow}} \bigg(s_0 1_{\{j=1\}} + \sum_{i\ge 1} s_i^j (1 - s_i)\bigg) \nu_j(\mathrm{d}\mathbf{s}) < \infty,$$

such that, for $\varepsilon^{(j)} = \{\{j\}, \mathbb{N} \setminus \{j\}\}$ and $\omega^{[j]} = \{[j], \{j+1\}, \{j+2\}, \ldots\}$, $j \ge 1$,

$$\kappa = c_1 \delta_{\varepsilon^{(1)}} + \sum_{j\ge 1} \bigg(c_j \delta_{\varepsilon^{(j+1)}} + k_j \delta_{\omega^{[j]}} + \int_{S^{\downarrow}} \kappa_{\mathbf{s}}\big(\cdot \cap \mathcal{P}^j\big) \nu_j(\mathrm{d}\mathbf{s})\bigg), \quad \text{where } \mathcal{P}^j = \mathcal{P}^{\{[j],\{j+1\}\}}.$$

In the exchangeable case, we have $(c_j, k_j, \nu_j) = (c, 0, \nu)$, $j \ge 1$, as was shown by Bertoin [6].

1.3. *RE hierarchies and continuum random trees*

In the context of our motivating example, the alpha-gamma model, we demonstrated how consistent Markov branching trees give rise to a random hierarchy $\mathcal{H}$ of $\mathbb{N}$. Let us investigate this in the context of Bertoin's systematic studies [8] of exchangeable homogeneous and exchangeable self-similar $\mathcal{P}$-valued fragmentation processes $(F^*(t), t \geq 0)$ in continuous time, and of Haas and Miermont's [17] associated self-similar continuum random trees (CRTs).

Bertoin described exchangeable homogeneous fragmentation processes in terms of an exchangeable dislocation measure $\kappa = \sum_{j\geq 1} c\delta_{\varepsilon(j)} + \int_{\mathcal{S}^{\downarrow}} \kappa_{\mathbf{s}}\nu(\mathrm{d}\mathbf{s})$ on $\mathcal{P}$. Informally, blocks fragment independently; for each $\pi \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}$, there is a competing rate $\kappa(\mathcal{P}^{\pi})$ at which a given block $F_i^*(t)$ undergoes a split whose effect on the first n block members is a partition according to π. For an α-self-similar fragmentation process, this rate is increased (in the case $\alpha > 0$) by a factor $|F_i^*(t)|^{-\alpha}$ depending on the asymptotic frequency $|F_i^*(t)|$ of the block. The rate increase is such that singleton blocks and indeed the all-singleton state $\mathbf{0}_{\mathbb{N}}$ are obtained in finite time.

Under some regularity conditions, [17] constructed self-similar CRTs $(\mathcal{T}_{(\alpha,\nu)}, \mu)$ with characteristic pair (α, ν), i.e. random path-connected compact metric spaces $(\mathcal{T}_{(\alpha,\nu)}, \mathrm{d})$ equipped with a root $\rho \in \mathcal{T}_{(\alpha,\nu)}$ and a probability measure μ on $\mathcal{T}_{(\alpha,\nu)}$, and with the tree property that there are no cyclic paths. Self-similarity here means that conditionally given the tree up to height t above the root and given subtree masses $\mu(S_i(t)) = m_i(t)$ above height t, the subtrees $S_i(t)$, $i \geq 1$, above height t, are like independent copies of $\mathcal{T}_{(\alpha,\nu)}$, with masses rescaled by $m_i(t)$ and distances rescaled by $(m_i(t))^{\alpha}$. These CRTs can be considered as genealogical trees of Bertoin's fragmentation processes; for a μ-distributed i.i.d. sample Σ_n^*, $n \geq 1$, in $\mathcal{T}_{(\alpha,\nu)}$, we obtain an α-self-similar fragmentation process by considering the partition-valued process that has $\{n \geq 1\colon\ \Sigma_n^* \in S_i(t)\}$, $i \geq 1$, as non-singleton blocks and all other integers in singleton blocks at time t, $t \geq 0$.

To any exchangeable $\mathcal{P}$-valued fragmentation process we associate the exchangeable hierarchy $\mathcal{H}^* = \{F_i^*(t), i \geq 1, t \geq 0\}$ of all blocks ever visited, equivalently $\mathcal{H}^* = \{\mathcal{L}^*(\mathcal{T}^v)\colon\ v \in \mathcal{T}_{(\alpha,\nu)}\}$, where $\mathcal{L}^*(\mathcal{T}^v) = \{n \in \mathbb{N}\colon\ \Sigma_n^* \in \mathcal{T}^v\}$ and $\mathcal{T}^v$ is the subtree of $\mathcal{T}_{(\alpha,\nu)}$ above $v \in \mathcal{T}_{(\alpha,\nu)}$. We say that the hierarchy $\mathcal{H}^*$ is *embedded in the CRT* $\mathcal{T}_{(\alpha,\nu)}$ by the sample $\Sigma_n^* \in \mathcal{T}_{(\alpha,\nu)}$, $n \geq 1$.

We now associate with any RE dislocation measure κ a *RE fragmentation process* F, in which each block fragments independently, with rates $\kappa(\mathcal{P}^{\pi})$, $\pi \in \mathcal{P}_n$, affecting the n smallest block members by partitioning according to π. We call $\mathcal{H} = \{F_i(t)\colon\ i \geq 1, t \geq 0\}$ the associated *RE hierarchy*. Alternatively (see Section 3.2), *RE splitting rules* $P_n(\pi) = \kappa(\mathcal{P}^{\pi})/\kappa(\mathcal{P} \setminus \mathcal{P}^{\mathbf{1}_{[n]}})$, $\pi \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}$, give rise to consistent *RE labelled Markov branching trees* $(T_n, n \geq 1)$ with splitting rules $(P_n, n \geq 2)$ that induce a RE hierarchy $\{A \subset \mathbb{N}\colon\ A \cap [n] \in T_n \text{ for all } n \geq 1\}$. Embedding a non-exchangeable hierarchy $\mathcal{H}$ into a CRT $\mathcal{T}$ means finding $\Sigma_n \in \mathcal{T}$, $n \geq 1$, with a non-trivial dependence structure, such that $\mathcal{H}$ is embedded in $\mathcal{T}$ by Σ_n, $n \geq 1$.

Theorem 6. *Let $\alpha > 0$, and let κ be a RE dislocation measure of the form*

$$\kappa = \sum_{j\geq 1} \int_{S^{\downarrow}} \kappa_{\mathbf{s}}\big(\cdot \cap \mathcal{P}^j\big)\nu_j(\mathrm{d}\mathbf{s}), \quad \text{with } \nu(\mathrm{d}\mathbf{s}) := \sum_{j\geq 1}\bigg(\sum_{i\geq 1} s_i^j(1 - s_i)\bigg)\nu_j(\mathrm{d}\mathbf{s}) \tag{2}$$

satisfying $\int_{S^{\downarrow}}(1 - s_1)\nu(\mathrm{d}\mathbf{s}) < \infty$ and $\nu(s_0 > 0) = 0$. Then we can construct $(\mathcal{T}_{(\alpha,\nu)}, (\Sigma_i, i \geq 1))$ such that $\mathcal{H} = \{\mathcal{L}(\mathcal{T}^v_{(\alpha,\nu)})\colon\ v \in \mathcal{T}_{(\alpha,\nu)}\}$ is a RE hierarchy with dislocation measure κ, embedded in a self-similar CRT $\mathcal{T}_{(\alpha,\nu)}$ with characteristic pair (α, ν), where $\mathcal{L}(\mathcal{T}^v_{(\alpha,\nu)}) = \{i \in \mathbb{N}\colon\ \Sigma_i \in \mathcal{T}^v_{(\alpha,\nu)}\}$.

Our proof of Theorem 6 in Section 4 gives an explicit sampling procedure for leaves $\Sigma_i \in \mathcal{T}_{(\alpha,\nu)}$, $i \geq 1$, based on the self-similarity of $\mathcal{T}_{(\alpha,\nu)}$ and recursive spinal decompositions of subtrees.

Theorem 6 partly generalises Theorem 4 of [28]. However, apart from the alpha model (the alpha-gamma model with $\gamma = \alpha$, which produces only binary trees), that theorem treats models that are not RE in the sense of Corollary 5 nor for other decompositions of $\mathcal{P}$.

It requires no extra work to also construct hierarchies associated with RE dislocation measures κ based on different decompositions of $\mathcal{P}$. However, in those more general cases, a RE measure still qualifies as a dislocation measure if and only if it is finite on $\mathcal{P}^{\{[j],\{j+1\}\}}$, $j \geq 1$, and this is necessary for hierarchies to be well-defined. Hence, the decomposition in Corollary 5 is the most natural decomposition in the context of fragmentation processes.

Exchangeable hierarchies $\mathcal{H}^*$ derived from fragmentation processes (or from Markov branching trees) have been used to construct CRTs as scaling limits [18]. We carry out a similar programme here for RE hierarchies $\mathcal{H}$, starting from a RE dislocation measure of the form identified in Corollary 5. We can delabel trees $T_n = \mathcal{H} \cap [n]$, but retain the root, to obtain rooted combinatorial trees T_n°, i.e. connected acyclic graphs with no degree-2 vertex, but some degree-1 vertices, only one of which is distinguished, as the root. We can regard T_n° as a metric space with unit distance between adjacent vertices and with adjacent vertices connected by unit length line segments. We use notation T_n°/a to scale the length of the line segments and to obtain a metric space with all connecting line segments of length $1/a$, where $a \in (0, \infty)$.

In the exchangeable case, [18] obtain CRT convergence under a regular variation condition

$$\nu(s_1 \le 1 - \epsilon) = \epsilon^{-\alpha}\ell(1/\epsilon) \quad \text{as } \epsilon \downarrow 0\text{; for some } \alpha \in (0, 1) \text{ and slowly varying } \ell \tag{3}$$

and a log-moment condition

$$\int_{S^\downarrow} \sum_{i \ge 2} s_i \left|\log(s_i)\right|^\varrho \nu(\mathrm{d}\mathbf{s}) < \infty \quad \text{for some } \varrho > 0. \tag{4}$$

Theorem 7. *If in the setting of Theorem 6, the measure ν satisfies (3) and (4), and if $\nu_j = \nu_m$ for some $m \ge 1$ and all $j \ge m$, then*

$$\frac{T_n^\circ}{n^\alpha \ell(n)\Gamma(1-\alpha)} \to \mathcal{T}_{(\alpha,\nu)} \quad \textit{in probability, in the Gromov–Hausdorff sense.}$$

Returning to the alpha-gamma model, we can now show that Theorem 7 applies to give a scaling limit in probability. The identification of ν_j, $j \ge 1$, in the parameterisation of Corollary 5 finally sheds some light on the peculiar splitting rules and ν-measures in Ford's alpha model and the alpha-gamma model [10,12,18,28]. To do this, we follow [19,24,25] and introduce Poisson–Dirichlet dislocation measures $\mathrm{PD}^*_{\alpha,\theta}(\mathrm{d}\mathbf{s})$ as σ-finite measures on $S^\downarrow$ given by

$$\mathbb{E}\left[\sigma_1^\theta; \sigma_1^{-1}\Delta\sigma_{[0,1]} \in \mathrm{d}\mathbf{s}\right], \quad \theta > -2\alpha, \alpha \in (0, 1),$$

on the interior of the parameter range, where $(\sigma_t, t \ge 0)$ is a stable subordinator with Laplace transform $\mathbb{E}[\mathrm{e}^{-\lambda\sigma_t}] = \mathrm{e}^{-t\lambda^\alpha}$ and where $\Delta\sigma_{[0,1]}$ is the decreasing rearrangements of the jumps $\Delta\sigma_t = \sigma_t - \sigma_{t-}$, $t \in [0, 1]$. For $\theta = -2\alpha$, the binary case, $\mathrm{PD}^*_{\alpha,-2\alpha}(\mathrm{d}\mathbf{s})$ is defined as the ranked beta measure on $\{(x, 1-x, 0, \dots), x \in (1/2, 1)\} \subset S^\downarrow$ with density $x^{-\alpha-1}(1-x)^{-\alpha-1}1_{(1/2,1)}(x)$; the associated Markov branching model is Aldous's [4] beta-splitting model, for $\alpha < 1$.

As the references demonstrate, Poisson–Dirichlet dislocation measures give rise to some of the nicest and best-studied parametric families of exchangeable fragmentation processes, while alpha and alpha-gamma models have as their dislocation measure what we have previously written as linear combinations of Poisson–Dirichlet measures of different parameters [10]. With the notion of restricted exchangeability, we can now obtain a stronger and more satisfactory connection.

Proposition 8. *The alpha-gamma model for $\alpha \in (0, 1)$ and $\gamma \in [0, \alpha]$ is a RE Markov branching model with dislocation measure of the form identified in Corollary 5 with $\nu_1 = (1-\alpha)\mathrm{PD}^*_{\alpha,-\alpha-\gamma}$ and $\nu_j = \gamma\mathrm{PD}^*_{\alpha,-\alpha-\gamma}$, $j \ge 2$.*

The boundary case $\alpha = 1$ degenerates [10] and leads to RE Markov branching models with

- for $\gamma = 0$ star trees corresponding to $(\nu_1, c_1, k_1) = (0, 0, 1)$ and $(\nu_2, c_2, k_2) = (0, 0, 0)$;
- for $\gamma = 1$ comb trees corresponding to $(\nu_1, c_1, k_1) = (0, 0, 0)$ and $(\nu_2, c_2, k_2) = (0, 1, 0)$;
- for $\gamma \in (0, 1)$ bushy combs corresponding to $(\nu_1, c_1, k_1) = (0, 0, 1)$, $(c_2, k_2) = (0, 0)$ and

$$\nu_2(s_2 > 0) = 0 \quad \text{and} \quad \nu_2(s_1 \in \mathrm{d}x) = \gamma x^{-2}(1-x)^{-1-\gamma}1_{(0,1)}(x)\,\mathrm{d}x.$$

1.4. *Sampling consistency and the skewed Poisson–Dirichlet model*

Proposition 8 suggests to introduce a three-parameter family of restricted exchangeable fragmentation trees that we call the *skewed Poisson–Dirichlet model*, by setting

$$\nu_1 = \lambda \mathrm{PD}^*_{\alpha,\theta}, \qquad \nu_j = (1-\lambda)\mathrm{PD}^*_{\alpha,\theta}, \quad j \geq 2,$$

for $\alpha \in [0,1]$, $\theta \geq -2\alpha$ and $\lambda \in [0,1]$. When $\lambda = (1-\alpha)/(1-\theta-2\alpha)$ and $\theta = -\alpha-\gamma$, this is the alpha-gamma model; when $\lambda = 1/2$, this is the exchangeable Poisson–Dirichlet model studied in [19,24]. We will use parameterisations by (α,θ,λ) and (α,γ,λ), where $\gamma = -\alpha-\theta$. We can apply Theorem 7 to obtain a convergence result in probability:

Corollary 9. *Let $(T_n, n \geq 1)$ be a consistent family of skewed Poisson–Dirichlet trees for parameters $0 < \alpha < 1$, $0 < \gamma = -\alpha - \theta \leq \alpha$ and $0 \leq \lambda < 1$. Then*

$$\frac{T_n^\circ}{n^\gamma} \to \mathcal{T}_{(\gamma,\nu)} \quad \textit{in probability, in the Gromov–Hausdorff sense,}$$

where $\mathcal{T}_{(\gamma,\nu)}$ is a γ-self-similar CRT associated with measure

$$\nu(\mathrm{d}\mathbf{s}) = \frac{\gamma\Gamma(1-\alpha)}{(1-\lambda)\alpha\Gamma(1-\gamma/\alpha)}\left(\lambda + (1-2\lambda)\sum_{i\geq 1} s_i^2\right)\mathrm{PD}^*_{\alpha,\theta}(\mathrm{d}\mathbf{s})$$

for $\gamma < \alpha$, while in the binary case $\gamma = \alpha$ (i.e. $\theta = -2\alpha$), we have $\nu(s_1 + s_2 < 1) = 0$ and

$$\nu(s_1 \in \mathrm{d}x) = \frac{\alpha}{(1-\lambda)\Gamma(1-\alpha)}\big((1-\lambda) + (4\lambda-2)x(1-x)\big)x^{-\alpha-1}(1-x)^{-\alpha-1}\,\mathrm{d}x.$$

Regarding the alpha model, $\alpha \in (0,1)$, $\theta = -2\alpha$, $\lambda = 1-\alpha$, this confirms in part a conjecture formulated in [28]; specifically, the setting of the conjecture was the two-parameter (α,θ)-model that contains the alpha model as a special case, and the conjecture claims almost sure convergence, while we only obtain convergence in probability here.

Another interesting feature of the skewed Poisson–Dirichlet model relates to sampling consistency. Here we say that a family of unlabelled random trees $(T_n^\circ, n \geq 1)$ is *sampling consistent* if the tree T_n° with a uniformly chosen leaf removed is distributed as T_{n-1}°. For consistent trees with exchangeable labels such as the exchangeable Poisson–Dirichlet model this is trivially so, but also and non-trivially for the alpha-gamma model that includes non-exchangeable trees [10]. Geometrically, this gives sampling consistency for two two-dimensional subsets of the three-dimensional parameter space (intersecting in the one-parameter family of stable trees [25] for $\gamma = 1-\alpha$), but somewhat surprisingly, sampling consistency does not extend any further:

Proposition 10. *The skewed Poisson–Dirichlet model is sampling consistent only for parameters that reduce it to the exchangeable Poisson–Dirichlet model or to the alpha-gamma model.*

This shows that while Theorem 6 and 7 always refer to Markov branching trees T_n° in the sense of [18], they typically do not, however, satisfy the sampling consistency property of [18], so that the theory developed in [18] does not even yield convergence in distribution for these trees, where we here establish convergence in probability.

1.5. *Structure of this paper*

In addition to proofs of main results already formulated, the content of this paper is as follows.

- Section 2 proves Theorem 3 and Corollary 5 by combining approaches of Vershik and Kerov, and of Aldous, both in the exchangeable case.
- Section 3 includes a discussion of the relationship between restricted exchangeability, partial exchangeability and constrained exchangeability, and a discussion of RE dislocation measures, RE splitting rules, RE hierarchies and RE fragmentations.

- In Section 4, we develop a new technique to sample leaves in general self-similar CRTs. We make explicit the embedding that we use to prove Theorem 6, and we obtain decomposition results along subtrees spanned by the first k sampled leaves (Corollary 22).
- In Section 5 we prove Theorem 7. Our approach is similar in spirit to [18], but with added technical difficulties. We analyse the RE embedding of Theorem 6 in detail. While in [18] consideration of a single $\Sigma^* \in \mathcal{T}_{(\alpha,\nu)}$ gives relevant estimates for all Σ_n^*, $n \geq 1$, we here need individual estimates for each Σ_n, $n \geq 1$. Methods include Gnedin's constrained paintboxes and renewal theory. We also establish almost sure convergences of rescaled subtrees of T_n spanned by k leaves, as first $n \to \infty$ in Proposition 28 and then also $k \to \infty$ in (22).
- Section 6 provides proofs for Propositions 8 and 10.
- An Appendix contains the proof of a technical lemma.

2. Integral representations, proof of Theorem 3 and Corollary 5

Our first aim is to understand exchangeability on subsets of the form $\mathcal{P}^\pi \subseteq \mathcal{P}$, for some $\pi \in \mathcal{K}$. Let us formally define modified paintboxes. For $\mathbf{s} \in S^\downarrow$ let $m \geq 0$ such that $s_m > s_{m+1} = 0$ (or $m = \infty$ if $s_i > 0$ for all $i \geq 1$), suppose $\pi \in \mathcal{K}$ has k blocks $\pi_j \neq \varnothing$, $1 \leq j \leq k$, of which ℓ with $\#\pi_j \geq 2$. For the paintbox $\kappa_\mathbf{s}$ associated with $\mathbf{s}$, we have $\kappa_\mathbf{s}(\mathcal{P}^\pi) > 0$ iff either $s_0 > 0$ and $\ell \leq m$, or $s_0 = 0$ and $k \leq m$. In these cases, set $\kappa_\mathbf{s}^\pi = \kappa_\mathbf{s}(\cdot|\mathcal{P}^\pi)$. Then $\kappa_\mathbf{s}^\pi$ is a modified paintbox:

1. Randomly assign "colours" $c(\pi) = (c(\pi_1), \dots, c(\pi_k))$ to the blocks $\pi_1, \dots, \pi_k$ using the following rule (with $Z_\mathbf{s}^\pi$ as normalisation constant)
$$\mathbb{P}\big(c(\pi) = (i_1, \dots, i_k)\big) = \frac{1}{Z_\mathbf{s}^\pi} \prod_{1 \leq j \leq k} s_{i_j}^{\#\pi_j}, \tag{5}$$
where i_j is allowed to be equal to 0 iff $\#\pi_j = 1$, and the i_j with $i_j \geq 1$ are pairwise distinct.
2. Let n be such that $\pi \in \mathcal{P}_n$. Conditionally given $c(\pi) = (i_1, \dots, i_k)$, set for $1 \leq r \leq n$ and $r \in \pi_j$,
$$\xi_r = i_j \quad \text{if } i_j \geq 1, \quad \text{and} \quad \xi_r = -\min \pi_j \quad \text{if } i_j = 0,$$
and for $r \geq n+1$, consider independent ξ_r with $\mathbb{P}(\xi_r = i) = s_i$, $i \geq 1$, $\mathbb{P}(\xi_r = -r) = s_0$. Then $\kappa_\mathbf{s}^\pi$ is the distribution of the partition $\Pi = \{\{n \geq 1\colon \xi_n = i\}, i \in \mathbb{Z}\}$, which puts any two $n, n' \in \mathbb{N}$ into the same block if and only if $\xi_n = \xi_{n'}$.

In the degenerate case when $\kappa_\mathbf{s}(\mathcal{P}^\pi) = 0$, the numerator of (5) always vanishes. Roughly speaking, we use all colours $1, \dots, m$ for the largest blocks of π. Formally, we replace 1. by 1′.:

1′. Randomly assign "colours" using the following rule (with $Z_\mathbf{s}^\pi$ as normalisation constant):
$$\mathbb{P}\big(c(\pi) = (i_1, \dots, i_k)\big) = \frac{1}{Z_\mathbf{s}^\pi} \prod_{1 \leq j \leq k : i_j \neq 0} s_{i_j}^{\#\pi_j},$$
if $\{i_1, \dots, i_k\} = \{0, \dots, m\}$, the $i_j \geq 1$ are pairwise distinct and $\sum_{j=1}^k \#\pi_j 1_{\{i_j \neq 0\}}$ is maximal.

Step 2. is applied as before to construct Π and hence $\kappa_\mathbf{s}^\pi$. Note that $\Pi_j = \pi_j$ if $i_j = 0$, while $\Pi_j \supset \pi_j$ will have limiting frequency $s_{i_j} > 0$ if $i_j \geq 1$.

Now $\kappa_\mathbf{s}^\pi(\mathcal{P} \setminus \mathcal{P}^\pi) = 0$ and, for $\pi' = (\pi'_1, \dots, \pi'_{k'}) \in \mathcal{K}^\pi := \{\pi' \in \mathcal{K}\colon \pi' \cap [n] = \pi\}$,
$$\kappa_\mathbf{s}^\pi\big(\mathcal{P}^{\pi'}\big) = \frac{1}{Z_\mathbf{s}^\pi} \sum_{(i_1,\dots,i_{k'}) \text{ admissible for } (\pi,\pi',\mathbf{s})} s_0^{\#\{J_\mathbf{s}^\pi \leq j \leq k' : i_j = 0\}} \prod_{1 \leq j \leq k' : i_j \neq 0} s_{i_j}^{\#\pi'_j 1_{\{i_j \geq 1\}}},$$
where $J_\mathbf{s}^\pi = k+1$ in the degenerate case, $J_\mathbf{s}^\pi = 1$ otherwise, and where $(i_1, \dots, i_{k'})$ is admissible for $(\pi, \pi', \mathbf{s})$ if $(i_1, \dots, i_k)$ is as in 1′. or 1. above, respectively, and if for $k+1 \leq j \leq k'$, we allow i_j equal to 0 iff $\#\pi'_j = 1$, and the i_j with $i_j \geq 1$, $1 \leq j \leq k'$, and pairwise distinct.

For $\pi = \{\{1\}\}$, this is a well-known formula for Kingman's paintbox $\kappa_\mathbf{s} = \kappa_\mathbf{s}^\pi$, with $Z_\mathbf{s}^\pi = 1$. It is easy to show that, in the general case, the modified paintboxes $\kappa_\mathbf{s}^\pi$ are exchangeable on $\mathcal{P}^\pi$.

Proposition 11. *For any $n \geq 1$ and $\pi \in \mathcal{P}_n$, the modified paintbox $\kappa_{\mathbf{s}}^{\pi}$ can be expressed in terms of any $\Gamma \in \mathcal{P}^{\pi}$ with asymptotic frequencies* $\mathbf{s}$, *provided that any blocks of Γ with zero asymptotic frequency are either subsets of $[n]$ or singletons, as*

$$\kappa_{\mathbf{s}}^{\pi}\left(\mathcal{P}^{\pi'}\right) = \lim_{r\to\infty} \frac{\#\{\pi'' \in \mathcal{K}^{\pi'}\colon \pi'' \approx \Gamma|_r\}}{\#\{\pi'' \in \mathcal{K}^{\pi}\colon \pi'' \approx \Gamma|_r\}} \quad \textit{for all } \pi' \in \mathcal{K}^{\pi} = \left\{\pi' \in \mathcal{K}\colon \pi' \cap [n] = \pi\right\},$$

where we write $\pi' \approx \pi''$ if π' and π'' have the same multiset of block sizes.

Proof. This proof is a refinement of the relevant part of the proof of Theorem 3.1 of [21], Kerov's proof of Kingman's paintbox representation of exchangeable partitions in $\mathcal{P}$, where we need to take into account the restriction to $\mathcal{P}^{\pi}$. We evaluate the right-hand side. Numerator and denominator are easily calculated, e.g. for $\pi' = (\pi'_1, \dots, \pi'_{k'}) \in \mathcal{P}^{\pi}_{n'} := \mathcal{K}^{\pi} \cap \mathcal{P}_{n'}$ as

$$\#\left\{\pi'' \in \mathcal{P}^{\pi'}_r\colon \pi'' \approx \Gamma|_r\right\} = \sum \binom{r-n'}{\#\Gamma_{i_1}|_r - \#\pi'_1, \dots, \#\Gamma_{i_{k'}}|_r - \pi'_{k'}, \#\Gamma_{\mathrm{others}}|_r} \frac{1}{\prod_{j\geq 1} p_j!},$$

where $\sum$ is over indices $(i_1, \dots, i_{k'})$ such that $\#\Gamma_{i_j}|_r - \#\pi'_j \geq 0$ for all $j \in [k']$, $\Gamma_{\mathrm{others}}|_r$ is the vector of all $\Gamma_i|_r$, $i \geq 1$, except $\Gamma_{i_1}|_r, \dots, \Gamma_{i_{k'}}|_r$, and p_j is the number of blocks of $\Gamma|_r$ with j elements, $j \geq 1$. First assume $s_0 = 1 - \sum_{i\geq 1} s_i = 0$, then the limit exists and is $Z_{\mathbf{s}}^{\pi,\pi'}/Z_{\mathbf{s}}^{\pi,\pi}$, where

$$Z_{\mathbf{s}}^{\pi,\pi'} = \lim_{r\to\infty} \frac{\sum \binom{r-n}{\#\Gamma_{i_1}|_r - \#\pi'_1, \dots, \#\Gamma_{i_k}|_r - \#\pi'_{k'}, \#\Gamma_{\mathrm{others}}|_r} r^d}{\binom{r}{\#\Gamma_1|_r, \#\Gamma_2|_r, \dots}} = \sum_{(i_1,\dots,i_{k'})\text{ admissible for }(\pi,\pi',\mathbf{s})} \prod_{j:i_j\neq 0} s_{i_j}^{\#\pi'_j},$$

with d the minimal $\sum_{j=1}^{k'} \#\pi'_j 1_{\{i_j=0\}}$, so that $d > 0$ only in the degenerate case; this power d is such that terms with higher than the minimal sum vanish as $r \to \infty$, and we identify $\kappa_{\mathbf{s}}^{\pi}(\mathcal{P}^{\pi'})$.

If $s_0 > 0$, blocks of zero limiting frequency need to be treated differently, because their union $\widetilde{\Gamma}_0$ now has a limiting frequency, and a union $\widetilde{\pi}'_0$ of blocks of π' can indeed be associated with $\widetilde{\Gamma}_0$. Specifically, we calculate a first factor as

$$\lim_{r\to\infty} \frac{\sum \binom{r-n}{\#\widetilde{\Gamma}_0 - \widetilde{\pi}'_0, \#\widetilde{\Gamma}_{i_1}|_r - \#\widetilde{\pi}_1, \dots, \#\widetilde{\Gamma}_{i_{\widetilde{k}'}}|_r - \#\widetilde{\pi}'_{\widetilde{k}'}, \#\widetilde{\Gamma}_{\mathrm{others}}|_r}}{\binom{r}{\#\widetilde{\Gamma}_0|_r, \#\widetilde{\Gamma}_1|_r, \#\widetilde{\Gamma}_2|_r, \dots}} = \sum_{(i_1,\dots,i_{\widetilde{k}'})\text{ admissible for }(\widetilde{\pi},\widetilde{\pi}',\mathbf{s})} s_0^{\#\widetilde{\pi}'_0} \prod_{j=1}^{\widetilde{k}'} s_{i_j}^{\#\widetilde{\pi}'_j},$$

but then need to also count the further partitions of the block of size $\#\widetilde{\Gamma}_0|_r$. This yields for $\widetilde{\pi}'_0 = \pi'_{j_1} \cup \dots \cup \pi'_{j_b}$ a positive limit factor if $d = \#\widetilde{\pi}'_0 - b$ is minimal, which we then calculate as

$$\lim_{r\to\infty} \frac{\sum \binom{\#\widetilde{\Gamma}_0|_r - \#\widetilde{\pi}'_0}{\#\Gamma_{i_1}|_r - \#\pi'_{j_1}, \dots, \#\Gamma_{i_b}|_r - \#\pi'_{j_b}, 1, \dots, 1} r^d}{(\#\widetilde{\Gamma}_0|_r)!} = s_0^{-d};$$

the number of available indices is asymptotically equivalent to $\#\widetilde{\Gamma}_0|_r \sim s_0 r$, so that the sum contains $\sim (\#\widetilde{\Gamma}_0|_r)^b$ terms, and this contributes to the asymptotics of the numerator. Finally we sum over the different choices of $\widetilde{\pi}'_0$ with $\#\widetilde{\pi}'_0 - b = d$ to identify $\kappa_{\mathbf{s}}^{\pi}(\mathcal{P}^{\pi'})$. □

With these representations of the modified paintboxes, we now obtain the integral representation of general measures that are exchangeable on $\mathcal{P}^{\pi}$ for some $\pi \in \mathcal{P}_n$.

Proposition 12. *Let μ be a finite measure, exchangeable on $\mathcal{P}^{\pi}$ for some $\pi \in \mathcal{P}_n$. Then there is a finite measure ν on $S^{\downarrow}$ such that $\mu = \int_{S^{\downarrow}} \kappa_{\mathbf{s}}^{\pi} \nu(d\mathbf{s})$.*

Proof. This proof uses a combination of the martingale method due to Vershik and Kerov [31], Theorem 2, and the de Finetti method used by Aldous [1]. W.l.o.g., μ is a probability measure. Let $\Pi \sim \mu$ for an exchangeable probability measure on $\mathcal{P}^\pi$. For $n' \ge n$ and $\pi' \in \mathcal{K}^\pi \cap \mathcal{P}_{n'}$, consider the process

$$X_r = \frac{\#\{\pi'' \in \mathcal{K}^{\pi'}\colon \pi'' \approx \Pi|_r\}}{\#\{\pi'' \in \mathcal{K}^{\pi}\colon \pi'' \approx \Pi|_r\}}, \quad r \ge n',$$

in the decreasing filtration $\mathcal{F}_r$ generated by the block sizes of $\Pi|_u$, $u \ge r$. By exchangeability, X_r depends only on the block sizes B_r of $\Pi|_r$ and is hence $\mathcal{F}_r$-measurable and $\mathbb{E}[X_r|\mathcal{F}_{r+1}]$ only depends on X_{r+1}. For a multiset b of block sizes, denote by $m(b)$ (resp. $m'(b)$) the number of partitions in $\mathcal{K}^\pi$ (resp. in $\mathcal{K}^{\pi'}$) with block sizes b. By exchangeability, each of these is equally likely. For block sizes $B_{r+1} = b_{r+1}$, we denote by $m(b_r, b_{r+1})$ the number of partitions in $\mathcal{K}^{\widetilde{\pi}}$ with block sizes b_{r+1}, where $\widetilde{\pi}$ is any specific partition with block sizes b_r. Then there are $m(b_r)m(b_r, b_{r+1})$ partitions in $\mathcal{K}^\pi$ with block sizes b_{r+1} that restrict to block sizes b_r. With this notation, we have $X_r = m'(B_r)/m(B_r)$. Then

$$\mathbb{E}[X_r|B_{r+1} = b_{r+1}] = \sum_{b_r} \frac{m(b_r)m(b_r, b_{r+1})}{m(b_{r+1})} \frac{m'(b_r)}{m(b_r)} = \frac{1}{m(b_{r+1})} \sum_{b_r} m(b_r, b_{r+1})m'(b_r) = \frac{m'(b_{r+1})}{m(b_{r+1})}$$

for all admissible b_{r+1} shows that $(X_r, r \ge n')$ is a bounded martingale and hence converges a.s.

On the other hand, de Finetti's theorem yields that asymptotic frequencies exist μ-a.s. Specifically, consider a partition Π with distribution μ and, independently, a sequence U_i, $i \ge 1$, of auxiliary independent uniform random variables. Then the random variables

$$\Xi_j = U_i \quad \text{if } j \in \Pi_i,\ j \ge n+1,$$

are exchangeable. By de Finetti's theorem, they are conditionally i.i.d. and the atom sizes S_i of the random limiting distribution in random ("size-biased") order satisfy

$$S_i = \lim_{r\to\infty} \frac{\#\{j \in \{n+1, \ldots, n+r\}\colon \Xi_j = U_i\}}{r} = \lim_{r\to\infty} \frac{\#\Pi_i \cap [r]}{r}.$$

Clearly, the latter limit does not depend on the auxiliary variables $(U_i, i \ge 1)$, so asymptotic frequencies exist μ-a.s. Furthermore, μ-a.e. partition is such that blocks with zero asymptotic frequency either only involve elements of $[n]$ or are singletons. Denote by ν the distribution on $S^\downarrow$ of the asymptotic frequencies $\mathbf{S} = (S_i, i \ge 1)$ rearranged into decreasing order of Π.

This means that μ is concentrated on those partitions for which Proposition 11 yields modified paintbox representations, and we see that $X_r \to \kappa^\pi_{\mathbf{S}}(\mathcal{P}^{\pi'})$ a.s., where $\mathbf{S} \sim \nu$; but $(X_r, r \ge n')$ is a bounded martingale, so exchangeability on $\mathcal{P}^\pi$ yields

$$\int_{S^\downarrow} \kappa^\pi_{\mathbf{s}}(\mathcal{P}^{\pi'})\nu(d\mathbf{s}) = \mathbb{E}\big[\kappa^\pi_{\mathbf{S}}(\mathcal{P}^{\pi'})\big] = \mathbb{E}[X_{n'}] = \sum_{\widetilde{\pi} \in \mathcal{P}^\pi_{n'} : \widetilde{\pi} \approx \pi'} \mu(\mathcal{P}^{\widetilde{\pi}}) \frac{1}{\#\{\pi'' \in \mathcal{P}^\pi_{n'}\colon \pi'' \approx \widetilde{\pi}\}} = \mu(\mathcal{P}^{\pi'}).$$

□

This proof raises the question whether we could have done without the martingale method or without the de Finetti argument, as can be done in the exchangeable case. To avoid the de Finetti argument, we would have to generalise Proposition 11 to ensure that all Γ for which the limits in Proposition 11 exist converge to modified paintboxes, which seems more difficult given the exceptional non-singleton sets of zero limiting frequency. On the other hand, our de Finetti argument only identifies the distribution of Π restricted to $\{n+1, n+2, \ldots\}$ and gives little information about the conditional distribution of how the blocks of π attach themselves to such paintboxes. We have not found a simple and direct argument to see why the modified paintboxes describe the only way to attach π in an exchangeable way.

Now recall that Theorem 3 states that RE measures on $\mathcal{P}$ are precisely those of the form $\mu = \sum_{\pi\in\mathcal{C}} \int_{S^\downarrow} \kappa^\pi_{\mathbf{s}} \nu_\pi(d\mathbf{s})$.

Proof of Theorem 3. First consider $\mu = \sum_{\pi\in\mathcal{C}} \int_{S^\downarrow} \kappa^\pi_{\mathbf{s}} \nu_\pi(d\mathbf{s})$ with $\mathcal{C}$ such that no $\pi \in \mathcal{C}$ is a restriction of another $\pi' \in \mathcal{C}$. Since $\kappa^\pi_{\mathbf{s}}$ only charges $\mathcal{P}^\pi$, the measure μ only charges $\bigcup_{\pi\in\mathcal{C}} \mathcal{P}^\pi$. Furthermore, the restrictions of μ are finite and exchangeable on $\mathcal{P}^\pi$. Hence μ is RE.

Conversely, let μ be any RE measure on $\mathcal{P}$ with $\mathcal{C}$ such that the three bullet points of Definition 1 hold. Then the sets $\mathcal{P}^\pi$, $\pi \in \mathcal{C}$, are disjoint and the restrictions of μ to $\mathcal{P}^\pi$ are finite and exchangeable on $\mathcal{P}^\pi$. By Proposition 12, the restrictions of μ to $\mathcal{P}^\pi$ can be represented as $\int_{S^\downarrow} \kappa_{\mathbf{s}}^\pi \nu_\pi(d\mathbf{s})$. Since furthermore $\mu(\mathcal{P} \setminus \bigcup_{\pi\in\mathcal{C}} \mathcal{P}^\pi) = 0$, we have

$$\mu = \sum_{\pi\in\mathcal{C}} \mu\bigl(\cdot \cap \mathcal{P}^\pi\bigr) = \sum_{\pi\in\mathcal{C}} \int_{S^\downarrow} \kappa_{\mathbf{s}}^\pi \nu_\pi(d\mathbf{s}). \qquad \square$$

The proof of the Corollary 5 is now straightforward. Note, however, that ν_j is not ν_π for $\pi = \{[j], \{j+1\}\}$, $j \ge 1$. Instead, we set $k_j = \nu_\pi(\{(0,0,\ldots)\} \ge 0$ and $c_j = \nu_\pi(\{(1,0,\ldots)\}) \ge 0$. The corresponding modified paintboxes are $\delta_{\{[j],\{j+1\},\{j+2\},\ldots\}}$ and $\delta_{\{\{j+1\},\mathbb{N}\setminus\{j+1\}\}}$, respectively, except for $j = 1$, where it is $\frac{1}{2}(\delta_{\{\{1\},\mathbb{N}\setminus\{1\}\}} + \delta_{\{\{2\},\mathbb{N}\setminus\{2\}\}})$. We also incorporate the normalisation constants $Z_{\mathbf{s}}^\pi$ of the modified paintboxes as densities into ν_j and use restricted Kingman paintboxes $\kappa_{\mathbf{s}}(\cdot \cap \mathcal{P}^j)$ rather than normalised modified paintboxes $\kappa_{\mathbf{s}}^\pi$.

3. Basic results on restricted exchangeability and related notions

3.1. *Partially exchangeable and constrained exchangeable partitions*

Let us explore the connections between the RE partitions introduced in this paper and other generalisations of exchangeability studied in the literature, notably partial exchangeability and constrained exchangeability. *Partially exchangeable partitions* were introduced by Pitman [26]. A measure μ_n on $\mathcal{P}_n$ is partially exchangeable if $\mu_n(\pi) = \mu_n(\pi')$ for all $\pi, \pi' \in \mathcal{P}_n$ with the same *vector* of block sizes *in the order of least element*. Partially exchangeable measures are not RE, in general, nor vice versa. Specifically, $\pi = \{\{1,2\},\{3,4\}\}$ and $\pi' = \{\{1,3\},\{2,4\}\}$ have the same mass for partially exchangeable measures but not necessarily for RE measures. Vice versa, consider $\pi = \{\{1,2,3\},\{4,5\}\}$ and $\pi' = \{\{1,2\},\{3,4,5\}\}$. In fact, "the intersection" of the two concepts is exchangeability:

Proposition 13. *A measure μ_n of $\mathcal{P}_n$ is exchangeable if and only if it is both partially exchangeable and RE with $\mathcal{C} = \{\mathbf{0}_{[2]}, \mathbf{1}_{[2]}\} = \{\{\{1\},\{2\}\},\{\{1,2\}\}\}$.*

Proof. The "only if" part follows straight from the definitions. For the "if" part, suppose that $\pi, \pi' \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}$ have the same multiset of block sizes. Let $\widetilde{\pi}$ be such that, for blocks in order of least element, $\widetilde{\pi}_1 = (\pi_1 \cup \{\min \pi_2\}) \setminus \{2\}$ and $\widetilde{\pi}_2 = (\pi_2 \setminus \{\min \pi_2\}) \cup \{2\}$, $\widetilde{\pi}_j = \pi_j$, $j \ge 3$. Similarly construct $\widetilde{\pi}'$ from π'. By partial exchangeability $\mu_n(\pi) = \mu_n(\widetilde{\pi})$ and $\mu_n(\pi') = \mu_n(\widetilde{\pi}')$. But $\widetilde{\pi}', \widetilde{\pi} \in \mathcal{P}^{\{\{1\},\{2\}\}}$, so by restricted exchangeability, we have $\mu_n(\widetilde{\pi}') = \mu_n(\widetilde{\pi})$. $\square$

Constrained exchangeable partitions were introduced by Gnedin [13]. Let $\varsigma = (\varsigma_k, k \ge 1)$ be a fixed sequence of integers $\varsigma_k \ge 1$. Consider the set $\mathcal{P}^{\varsigma\text{-constr}}$ of partitions $\Gamma \in \mathcal{P}$ that are *constrained with respect to* ς in the sense that each block Γ_k contains the ς_k least elements of $\bigcup_{j\ge k} \Gamma_j$ for every $k \ge 1$ with $\Gamma_k \ne \varnothing$. A measure μ on $\mathcal{P}$ is *constrained exchangeable* if $\mu(\mathcal{P} \setminus \mathcal{P}^{\varsigma\text{-constr}}) = 0$ for some ς, and if $\mu_n(\pi) = \mu_n(\pi')$ for all $\pi, \pi' \in \{\Gamma|_n \colon \Gamma \in \mathcal{P}^{\varsigma\text{-constr}}\}$ with the same multiset of block sizes and all $n \ge 1$. For $\varsigma = (1,2,1,\ldots)$, under a constrained exchangeable measure, $\pi = \{\{1,3\},\{2,4\},\{5\}\}$ and $\pi' = \{\{1,2\},\{3,4\},\{5\}\}$ have the same mass, but not necessarily under a RE measure. Vice versa, restrictions to $\mathcal{P}^{\{[j],\{j+1\}\}}$ of a RE measure μ are constrained exchangeable if we take $\varsigma = (j,1,1,\ldots)$, but as soon as μ gives positive mass to more than one $\mathcal{P}^{\{[j],\{j+1\}\}}$, $j \ge 1$, constrained exchangeability in Gnedin's sense fails.

3.2. *RE hierarchies and fragmentation processes*

In Section 1.1, we defined hierarchies $\mathcal{H}_B$ on sets $B \subseteq \mathbb{N}$ and represented hierarchies on finite $B \subset \mathbb{N}$ as graph-theoretic trees above a root $\varnothing$, with edges between each block $A \in \mathcal{H}_B$, $\#A \ge 2$, and its maximal subsets in $\mathcal{H}_B$, which form a partition of A. For infinite $B \subseteq \mathbb{N}$, the notion of a maximal subset A of B in $\mathcal{H}_B$ is more delicate, and it is not always true that there are maximal subsets that form a partition of B.

For a hierarchy $\mathcal{H}_B$ on infinite $B \subseteq \mathbb{N}$, we say $\mathcal{H}_B$ is *closed* if for all sequences $(A_j, j \ge 1)$ in $\mathcal{H}_B$ that are increasing for the inclusion partial order, we have $\bigcup A_j \in \mathcal{H}_B$, and if for all decreasing sequences we have $\bigcap A_j \in$

$\mathcal{H}_B$. A closed hierarchy $\mathcal{H}_B$ is uniquely determined by its restrictions $\mathcal{H}_B \cap [n]$, $n \geq 1$, as $\mathcal{H}_B = \{A \subseteq B\colon\ A \cap [n] \in \mathcal{H}_B \cap [n] \text{ for all } n \geq 1\}$. For every hierarchy $\mathcal{H}_B$ there is a *closure* $\mathcal{H}_B^{\mathrm{cl}}$, the intersection of all closed hierarchies containing $\mathcal{H}_B$.

Recall from Section 1.1 definitions of labelled Markov branching models $(T_n, n \geq 1)$ with splitting rules $P_n, n \geq 2$, and that hierarchies T_n on $[n]$ are called consistent if $T_{n+1} \cap [n] = T_n$, Both consistency and the labelled Markov branching property can be viewed as properties of the distributions Q_n of T_n, $n \geq 1$. This labelled Markov branching property implies a Markov branching property [18] for rooted delabelled trees $T_n^\circ \sim Q_n^\circ$, as follows: call *size* the number of leaves (non-root degree-1 vertices), *first split* the decreasing sequence of subtree sizes for the vertex adjacent to the root; conditionally given that the first split of T_n° is $n_1 \geq \cdots \geq n_k$, the subtrees are distributed as if they were independent with respective distributions $Q_{n_i}^\circ$, $1 \leq i \leq k$. On the other hand, the associated family $(Q_n^\circ, n \geq 1)$ will not, in general, have the sampling consistency property of [18], which asserts that a tree $T_n^\circ \sim Q_n^\circ$ with a leaf picked uniformly at random removed (together with any resulting degree-2 vertex) has distribution Q_{n-1}°, for $n \geq 3$.

Let $\mathcal{P}^j = \mathcal{P}^{\{[j],\{j+1\}\}}$, $j \geq 1$. We say that a splitting rule P_n is *RE* if for all $1 \leq j \leq n-1$ and $\pi, \pi' \in \mathcal{P}_n^j := \{\pi \in \mathcal{P}_n\colon\ \pi \cap [j+1] = \{[j], \{j+1\}\}\}$ with the same multiset of block sizes, we have $P_n(\{\pi\}) = P_n(\{\pi'\})$. The alpha-gamma model of Section 1.1 is an example.

If κ is a RE dislocation measure as in Corollary 5, then

$$P_n\bigl(\{\pi\}\bigr) = \begin{cases} \kappa(\mathcal{P}^\pi)/\kappa(\mathcal{P} \setminus \mathcal{P}^{\mathbf{1}_{[n]}}), & \kappa(\mathcal{P} \setminus \mathcal{P}^{\mathbf{1}_{[n]}}) > 0, \\ \delta_{\mathbf{0}_{[n]}}(\{\pi\}), & \kappa(\mathcal{P} \setminus \mathcal{P}^{\mathbf{1}_{[n]}}) = 0 \text{ or } n = 2, \end{cases} \qquad \pi \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}, n \geq 2, \tag{6}$$

defines RE splitting rules and hence inductively a consistent Markov branching model $(Q_n, n \geq 1)$ that we also refer to as *RE*. More specifically, there is always $n_0 \in \{2, 3, \ldots\} \cup \{\infty\}$ such that the second line in (6) applies for $n < n_0$ but not for $n \geq n_0$. The second line leads to the *minimal* hierarchy $Q_n(\{\{[n], \{1\}, \ldots, \{n\}, \varnothing\}\}) = 1$ of $[n]$. We have $P_{n_0}(\{[n_0 - 1], \{n_0\}\}) = 1$ *degenerate*, while for all $n \geq n_0 + 1$, we have $P_n(\{[n-1], \{n\}\}) < 1$, *non-degenerate*, if $n_0 < \infty$.

Let us call a consistent RE Markov branching model $(Q_n, n \geq 1)$ with splitting rules $(P_n, n \geq 2)$ *regular* if there is $n_0 \geq 2$ such that Q_n is minimal for $n < n_0$, and if P_n is degenerate for $n = n_0$, non-degenerate for $n \geq n_0 + 1$.

Proposition 14. *All regular consistent labelled Markov branching models $(Q_n, n \geq 1)$ with RE splitting rules $(P_n, n \geq 2)$ are of the form (6) for some RE measure κ as in Corollary 5.*

Proof. In Pitman's [27] formalism of exchangeable partition probability functions (EPPFs)

$$p_n^j(\#\pi_1, \ldots, \#\pi_k) = P_n\bigl(\{\pi\}\bigr), \quad \pi \in \mathcal{P}_n^j = \mathcal{P}^{\{[j],\{j+1\}\}} \cap [n], j \in [n-1],$$

consistency in the RE case (extending Formula (16) of [24]) is equivalent to

$$p_n^j(n_1, \ldots, n_k) = p_{n+1}^n(n, 1) p_n^j(n_1, \ldots, n_k) + \sum_{i=1}^{k+1} p_{n+1}^j(n_1, \ldots, n_{i-1}, n_i + 1, n_{i+1}, \ldots, n_k)$$

for all $n_1, \ldots, n_k \in \mathbb{N}$, $k \geq 2$, $n = n_1 + \cdots + n_k$, $j \in [n-1]$. For $\lambda_n = 0$, $n < n_0$, any $\lambda_{n_0} \in (0, \infty)$ and $(1 - p_{n+1}^n(n,1))\lambda_{n+1} = \lambda_n$, $n \geq n_0$, we see that $\kappa(\mathcal{P}^\pi) = \lambda_n P_n(\{\pi\})$, $\pi \in \bigcup_{n \geq 2} \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}$, defines a RE measure that has the properties required. $\square$

By Kolmogorov's consistency theorem, we can consider a consistent family $(T_n, n \geq 1)$ of trees $T_n \sim Q_n$ with $T_{n+1} \cap [n] = T_n$, $n \geq 1$, and associate $\mathcal{H} = \{A \subset \mathbb{N}\colon\ A \cap [n] \in T_n \text{ for all } n \geq 2\}$ as random closed hierarchy on $\mathbb{N}$, which we call *RE* if $(Q_n, n \geq 1)$ is RE. In the regular RE case with $n_0 = 2$, we can consistently embed into continuous time the blocks of T_n, $n \geq 2$, using

- exponential holding times $\eta_{[n]}$ of rate λ_n for state $[n]$, λ_n as in the proof of Proposition 14;
- recursively and independently as blocks appear from splits, η_π at rate $\lambda_{\#\pi}$ for any $\pi \in T_n$.

With the convention that $\lambda_1 = 0$ gives infinite holding times, the collection of blocks held at any given time $t \geq 0$ forms a partition $F|_n(t)$ of $[n]$. Indeed, this construction yields consistent homogeneous fragmentation processes $(F|_n(t), t \geq 0)$ in $\mathcal{P}_n$, $n \geq 2$, that determine a $\mathcal{P}$-valued process $(F(t), t \geq 0)$, which we call a RE homogeneous fragmentation process.

We can also generalise Bertoin's [8] Poissonian construction to directly obtain RE homogeneous fragmentation processes $(F(t), t \geq 0)$ in $\mathcal{P}$ from a RE dislocation measure κ. This provides an alternative construction of the same random closed hierarchy $\mathcal{H} = \{F_i(t), i \geq 0, t \geq 0\}^{\mathrm{cl}}$, but we do not need this alternative construction and leave the details to the reader.

In the regular case with $n_0 \geq 3$, a block $[n_0 - 1]$ is never split under the exponential-rates construction above; informally $[n_0 - 1]$ is a limiting block at infinity alongside many other such blocks of size $n_0 - 1$ that still need splitting to obtain a hierarchy – they need partitioning into singletons. The simplest kind of irregular model of a RE hierarchy can be obtained here by some intermediate partitioning of these blocks of size $n_0 - 1$. It is possible, but not as natural as in the regular case with $n_0 = 2$, to incorporate such further splits in a common embedding, also when other irregularities occur with more degenerate splitting rules. For our next aim of embedding hierarchies into self-similar CRTs, such embeddings do not provide a suitable framework.

4. Embedding in self-similar CRTs, proof of Theorem 6

4.1. *Self-similar CRTs, fragmentation processes and spinal decomposition*

Aldous [2] called a pair $(\mathcal{T}, \mu)$ a *continuum tree* if $\mathcal{T}$ is an $\mathbb{R}$-tree, μ a finite measure on $\mathcal{T}$, with

1. the measure μ supported by the set $\mathrm{Lf}(\mathcal{T})$ of leaves of $\mathcal{T}$,
2. the measure μ has no atoms,
3. for every $x \in \mathcal{T} \setminus \mathrm{Lf}(\mathcal{T})$, positive mass $\mu(\mathcal{T}_x) > 0$ in the subtree $\mathcal{T}_x$ rooted at x.

We specify a root vertex $\rho \in \mathcal{T}$ and distance function d. For technical simplicity, we follow Aldous [3] and use CRTs in $\ell_1 = \ell_1(\mathbb{N})$. We endow the set of compact subsets of ℓ_1 with the Hausdorff metric, and the set of finite measures on ℓ_1 with any metric inducing the topology of weak convergence, so that the set $\mathbb{H}$ of pairs (T, μ) where T is a rooted $\mathbb{R}$-tree embedded as a subset of ℓ_1 and μ is a finite measure on T, is endowed with the product Borel σ-algebra.

A *Continuum Random Tree (CRT)* is a random variable with values in the set of continuum trees. To be specific, we call *distribution of a CRT* $(\mathcal{T}, \mu, \rho, \mathrm{d})$ the distribution on $\mathbb{H}$ of the particular random isometric embedding of $(\mathcal{T}, \mathrm{d})$ in ℓ_1 obtained from a random sample Σ_i^*, $i \geq 1$, of independent leaves with distribution $\mu/\mu(\mathcal{T})$, using $0 \in \ell_1$ as the root and the ith coordinate direction in ℓ_1 to embed the branch leading to leaf Σ_i^*, finally passing to the ℓ_1-closure and the weak limit of the $\mu(\mathcal{T})$-multiples of empirical measures of the embedded $\Sigma_1^*, \dots, \Sigma_i^*$, $i \geq 1$.

For, $\alpha \in \mathbb{R}$, $x \in [0, 1]$ and $\mathbf{s} \in S^\downarrow$, we denote by Q_x^α the distribution of the α-scaled tree $(\mathcal{T}, x\mu, \rho, x^\alpha \mathrm{d})$ and by $Q_{\mathbf{s}}^\alpha$ the distribution of a bush of independent trees with distributions $Q_{s_i}^\alpha$, $i \geq 1$, all grafted to the same root. For every $u \geq 0$, consider the bush $\mathcal{B}(u)$ obtained by grafting the connected components $\mathcal{T}_i(u)$, $i \in I$, of the open set $\{x \in \mathcal{T}\colon \mathrm{d}(x, \rho) > u\}$ to the same root. Recall that a CRT is called α-self-similar in the sense of [17], if for all $u \geq 0$ and conditionally given $(\mu(\mathcal{T}_i(u)), i \in I)^\downarrow = \mathbf{s} \neq 0$, we have $\mathcal{B}(u) \sim Q_{\mathbf{s}}^\alpha$.

For $\alpha \in \mathbb{R}$, a $\mathcal{P}$-valued process $(\Pi(t), t \geq 0)$ is an *exchangeable α-self-similar fragmentation process* if $\Pi = (\Pi(t), t \geq 0)$ is exchangeable and if given $\Pi(t) = \pi$, the partition $\Pi(t + s)$ has the same law as the random partition whose blocks are those of $\pi_i \cap \Pi^{(i)}(|\pi_i|^{-\alpha} s)$, $i \geq 1$, where $(\Pi^{(i)}, i \geq 1)$ is a sequence of i.i.d. copies of Π. The process $X = (|\Pi(t)|^\downarrow, t \geq 0)$ is an $S^\downarrow$-valued α-self-similar fragmentation. Bertoin proved in [5] that the distribution of an exchangeable $\mathcal{P}$-valued self-similar fragmentation is determined by a triple (α, c, ν), where ν is a dislocation measure on $S^\downarrow$, i.e. $\nu(s_1 = 1) = 0$ and $\int_{S^\downarrow} (1 - s_1)\nu(\mathrm{d}\mathbf{s}) < \infty$. In this paper, we take $c = 0$ and ν *conservative*, i.e. $\nu(s_0 > 0) = 0$, where $s_0 = 1 - \sum_{i \geq 1} s_i$. We call (α, ν) *characteristic pair*.

According to [17], there exists a self-similar CRT $\mathcal{T}$ associated with (α, ν), provided also that $\alpha > 0$ (and ν is infinite, but this is not essential unless it is required that the topological support of μ is $\mathcal{T}$). Specifically, $Y = ((\mu(\mathcal{T}_i(u)), i \in I_u)^\downarrow, u \geq 0)$ has the same distribution as X.

Consider $\mathbf{s} \in S^\downarrow$ and $\mathbf{s}^{(i)} \in S^\downarrow$, $i \geq 1$. We call *fragmentation of* $\mathbf{s}$ *by* $\mathbf{s}^{(\cdot)}$ the mass partition $\mathrm{Frag}(\mathbf{s}, \mathbf{s}^{(\cdot)})$ given by the decreasing rearrangement of $(s_i s_j^{(i)}, i, j \in \mathbb{N})$. Bertoin showed that the process $(X(t), t \geq 0)$ is Markovian and its

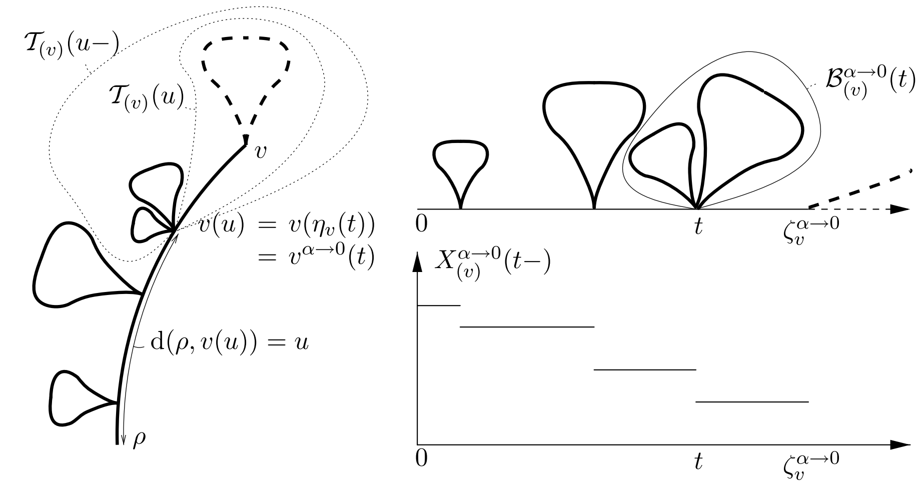


Fig. 3. The tree $\mathcal{T}$ on the left-hand side has its spine to $v \in \mathcal{T}$ exposed; for a vertex $v^{\alpha}(u)$ on the spine $[[\rho, v]]$, the subtree containing v has been indicated. The right-hand side displays rescaled subtrees and the residual mass process after passage to homogeneous time via η_v.

semigroup can be described as follows. For every $t, t' \geq 0$, the conditional distribution of $X(t+t')$ given $X(t) = \mathbf{s}$ is the law of $\mathrm{Frag}(\mathbf{s}, \mathbf{S}^{(\cdot)})$, where each $\mathbf{S}^{(i)}$ independently is distributed as $X(t' s_i^{-\alpha})$, see Proposition 3.7 of [8].

Consider an infinite block $B \subseteq \mathbb{N}$ and $\Gamma \in \mathcal{P}$. We call *fragmentation of B by Γ* the partition $\mathrm{Frag}(B, \Gamma) := \beta_B(\Gamma) \in \mathcal{P}_B$, where β_B is the unique increasing bijection from $\mathbb{N}$ to B. This is a slight variation of Bertoin's [8] notion, who uses $\Gamma \cap B$, not $\beta_B(\Gamma)$, but this is useful in Lemma 15 as it allows to recover Γ from $\mathrm{Frag}(B, \Gamma)$, and it is also instructive in the RE case.

Given a CRT $(\mathcal{T}, \mu, \rho, \mathrm{d})$ and $v \in \mathcal{T}$, we denote by $v(u)$ the point on the *spine* $[[\rho, v]]$ with $\mathrm{d}(\rho, v(u)) = u$, $0 \leq u \leq \mathrm{d}(\rho, v)$, and obtain a parameterisation $[[\rho, v]] = \{v(u), 0 \leq u \leq \mathrm{d}(\rho, v)\}$ *by distance*, cf. Fig. 3. We consider the subtree $\mathcal{T}_{(v)}(u) = \{w \in \mathcal{T}\colon \mathrm{d}(\rho, w \wedge v) > u\}$ of $\mathcal{T}$ containing v rooted at $v(u)$, and its mass $X_{(v)}(u) = \mu(\mathcal{T}_{(v)}(u))$. For $\alpha > 0$, let $\eta_v^{\alpha\to 0}$ be the *α-self-similar time change* with

$$\eta_v^{\alpha\to 0}(t) = \inf\left\{u \geq 0\colon \int_0^u \bigl(X_{(v)}(y)\bigr)^{-\alpha}\,\mathrm{d}y > t\right\}, \quad 0 \leq t < \zeta_v^{\alpha\to 0} = \int_0^{\mathrm{d}(\rho,v)} \bigl(X_{(v)}(y)\bigr)^{-\alpha}\,\mathrm{d}y. \tag{7}$$

Then $v^{\alpha\to 0}(t) = v(\eta_v^{\alpha\to 0}(t))$, $\mathcal{T}_{(v)}^{\alpha\to 0}(t) = \mathcal{T}_{(v)}(\eta_v^{\alpha\to 0}(t))$ and $X_{(v)}^{\alpha\to 0}(t) = \mu(\mathcal{T}_{(v)}^{\alpha\to 0}(t))$ are associated time-changed quantities. In particular, $[[\rho, v[[= \{v^{\alpha\to 0}(t), 0 \leq t < \zeta_v^{\alpha\to 0}\}$ is a new parameterisation of the spine, which we call parameterisation *by time*. Denote by $S^v(t) = (S_i^v(t), i \geq 1) \in \mathcal{S}^{\downarrow}$ the sequence such that $X_{(v)}^{\alpha\to 0}(t-) S^v(t)$ is the decreasing sequence of μ-masses of the connected components of $\{w \in \mathcal{T}\colon v^{\alpha\to 0}(t) \in [[\rho, w[[\}$, also $F_v(t) = X_{(v)}^{\alpha\to 0}(t)/X_{(v)}^{\alpha\to 0}(t-)$ the component of $S^v(t)$ corresponding to the subtree containing v. Moreover, we denote by

$$\left(\mathcal{B}_{(v)}^{\alpha\to 0}(t), \frac{\mu|_{\mathcal{B}_{(v)}^{\alpha\to 0}(t)}}{X_{(v)}^{\alpha\to 0}(t-)}, v^{\alpha\to 0}(t), \frac{\mathrm{d}|_{\mathcal{B}_{(v)}^{\alpha\to 0}(t)}}{(X_{(v)}^{\alpha\to 0}(t-))^{\alpha}}\right), \quad \text{where } \mathcal{B}_{(v)}^{\alpha\to 0}(t) = \mathcal{T}_{(v)}^{\alpha\to 0}(t-) \setminus \mathcal{T}_{(v)}^{\alpha\to 0}(t)$$

the associated rescaled spinal bush, of mass $1 - F_v(t)$, at time $t \geq 0$.

The following lemma is a description in the CRT framework of Bertoin's tagged particle process that is a bit richer than often stated, but follows from the same arguments.

Lemma 15. *Let $(\mathcal{T}, \mu, \rho, \mathrm{d})$ be an α-self-similar CRT with characteristic pair (α, ν) and $\Sigma^* \sim \mu$. Then $(S^{\Sigma^*}, F_{\Sigma^*})$ is a Poisson point process on $\mathcal{S}^{\downarrow} \times (0,1)$ with intensity measure $\widetilde{\nu}^*$ given by*

$$\widetilde{\nu}^*(\mathrm{d}\mathbf{s}, \mathrm{d}x) = \sum_{i\geq 1} s_i \delta_{s_i}(\mathrm{d}x) \nu(\mathrm{d}\mathbf{s}).$$

Proof. Let Y be the self-similar mass-fragmentation process associated with the CRT $(\mathcal{T}, \mu)$ and $Y^{\alpha\to 0}$ the homogeneous mass-fragmentation process obtained by applying the α-self-similar time-change to each block: $Y^{\alpha\to 0}(t) = (\mu(\mathcal{T}_i^{\alpha\to 0}(t)), i \in I_t^{\alpha\to 0})^{\downarrow}$, where $\{\mathcal{T}_i^{\alpha\to 0}(t), i \in I_t^{\alpha\to 0}\} = \{\mathcal{T}_{(v)}^{\alpha\to 0}(t)\colon v \in \mathcal{T}, \zeta_v^{\alpha\to 0} > t\}$. On an extended probability space, denote by Π a homogeneous exchangeable $\mathcal{P}$-valued fragmentation process associated with $Y^{\alpha\to 0}$. Without loss of generality, we can consider $X_{(\Sigma^*)}^{\alpha\to 0}(t) = |\Pi_1(t)|$, by exchangeability. Since $|\Pi_1(t)| > 0$ a.s., the block $\Pi_1(t)$ is infinite and there is a unique partition $\Pi^{(1)}(t)$ of $\mathbb{N}$ such that $\Pi_1(t) = \mathrm{Frag}(\Pi_1(t-), \Pi^{(1)}(t))$. Furthermore, $S^{\Sigma^*}(t) = |\Pi^{(1)}(t)|^{\downarrow}$. By Bertoin's Poissonian construction of exchangeable fragmentations, $\Pi^{(1)}$ is a (time-homogeneous) Poisson point process with intensity measure $\kappa = \int_{S^{\downarrow}} \kappa_{\mathbf{s}} \nu(\mathrm{d}\mathbf{s})$. Hence, S^{Σ^*} is a Poisson point process on $S^{\downarrow}$ with intensity measure ν.

As Σ^* is distributed according to μ, it is not hard to show that the distribution of $(S^{\Sigma^*}, F_{\Sigma^*})$ can be obtained by marking S^{Σ^*} via the size-biased marking kernel $K^*(\mathbf{s}, \cdot) = \sum_{i\geq 1} s_i \delta_{s_i}$ and so $(S^{\Sigma^*}, F_{\Sigma^*})$ is a Poisson point process with intensity $K^*(\mathbf{s}, \mathrm{d}x)\nu(\mathrm{d}\mathbf{s}) = \widetilde{\nu}^*(\mathrm{d}\mathbf{s}, \mathrm{d}x)$. □

By the stopping line argument of [19], Proposition 4, this yields the following joint description of the ordered coarse and unordered fine spinal decompositions along the spine to $\Sigma^* \sim \mu$.

Proposition 16 (Spinal decomposition [9,19]). *Let $(\mathcal{T}, \mu, \rho, \mathrm{d})$ be an α-self-similar CRT with characteristic pair (α, ν) and $\Sigma^* \sim \mu$. Then the process $(S^{\Sigma^*}, F_{\Sigma^*}, \mathcal{B}_{(\Sigma^*)}^{\alpha\to 0})$ is a Poisson point process with intensity measure*

$$\widetilde{\nu}^*_{\mathrm{bush}}(\mathrm{d}\mathbf{s}, \mathrm{d}x, \mathrm{d}T) = \sum_{i\geq 1} s_i \delta_{s_i}(\mathrm{d}x) Q^{\alpha}_{(s_1,\ldots,s_{i-1},s_{i+1},\ldots)}(\mathrm{d}T)\nu(\mathrm{d}\mathbf{s}).$$

Conversely, the isometry class of $(\mathcal{T}, \mu, \rho, \mathrm{d})$ is a measurable function of $(S^{\Sigma^*}, F_{\Sigma^*}, \mathcal{B}_{(\Sigma^*)}^{\alpha\to 0})$.

4.2. *A generic procedure to sample a leaf from a self-similar CRT*

Our aim is to generalise Lemma 15 and Proposition 16 to leaves other than the μ-sampled leaf Σ^* where we are effectively marking a Poisson point process with intensity measure ν using the size-biased marking kernel $K^*(\mathbf{s}, \cdot) = \sum_{i\geq 1} s_i \delta_{s_i}$ from $S^{\downarrow}$ to $(0, 1)$. We will now consider other marking kernels. It will be convenient to adopt an idea from Pitman's EPPF formalism and specify the probability that a *specific* part of size x is chosen with probability $P(\mathbf{s}, x)$ so that the probability of choosing a mass x is $K(\mathbf{s}, \{x\}) = m_x P(\mathbf{s}, x)$ where for $\mathbf{s} = (s_i, i \geq 1) \in S^{\downarrow}$, we let $m_x = \#\{i \geq 1\colon s_i = x\}$.

Definition 17. *A measurable function $P : S^{\downarrow} \times (0, 1) \to [0, 1]$ that fulfils the two conditions*

- *$P(\mathbf{s}, x) = 0$ if $x \notin \{s_i, i \geq 1\}$,*
- *$\sum_{i\geq 1} P(\mathbf{s}, s_i) = 1$,*

is called a selection probability function (SPF).

Example 18. *The SPF associated with a leaf chosen according to μ is $P(\mathbf{s}, s_i) = s_i$.*

We now formulate the procedure to sample a special leaf Σ based on an SPF P from an α-self-similar CRT $(\mathcal{T}, \mu, \rho, \mathrm{d})$ with dislocation measure ν, $\mathcal{T} \sim Q_1^{\alpha} = Q_1$ for short ($\alpha > 0$ fixed).

Procedure 1. *Let P be an SPF as in Definition 17 fulfilling*

$$\int_{S^{\downarrow}} \sum_{i\geq 1} (1 - s_i) P(\mathbf{s}, s_i)\nu(\mathrm{d}\mathbf{s}) < \infty. \tag{8}$$

0. *We start from $(\mathcal{T}_1, \mu_1, \rho_1, \mathrm{d}_1) := (\mathcal{T}, \mu, \rho, \mathrm{d})$ and $i = 1$ and proceed inductively.*
1. *Conditionally given $(\mathcal{T}_i, \mu_i, \rho_i, \mathrm{d}_i)$, let $\Sigma_{(i)} \sim \mu_i$.*

2. *Conditionally given* $(\mathcal{T}_i, \Sigma_{(i)})$, *we consider the parameterisation in homogeneous time of the spine* $[[\rho_i, \Sigma_{(i)}[[= \{\Sigma_{(i)}^{\alpha\to 0}(t), t \geq 0\}$ *and pick as* $\mathcal{T}_{i+1}$ *a subtree* $\mathcal{S}$ *off the spine*; *specifically, if* $\mathcal{S}$ *is a subtree rooted at the spinal vertex* $\Sigma_{(i)}^{\alpha\to 0}(t)$, *it is selected with probability*

$$\mathbb{P}(\mathcal{T}_{i+1} = \mathcal{S} | \mathcal{T}_i, \Sigma_{(i)}) = P\left(S^{\Sigma_{(i)}}(t), \frac{\mu_i(\mathcal{S})}{\mu_i(\mathcal{T}_{(\Sigma_{(i)})}^{\alpha\to 0}(t-))}\right) \prod_{t'<t} P\big(S^{\Sigma_{(i)}}(t'), F_{\Sigma_{(i)}}(t')\big).$$

3. *Let* $\tau_{(i)} = \inf\{t \geq 0\colon \mathcal{T}_{i+1} \cap \mathcal{T}_{\Sigma_{(i)}}^{\alpha\to 0}(t) = \varnothing\}$. *We turn* $\mathcal{T}_{i+1}$ *into a CRT with rescaled mass measure, root and rescaled distance function as follows*:

$$\mu_{i+1} = \frac{\mu_i|_{\mathcal{T}_{i+1}}}{\mu_i(\mathcal{T}_{i+1})}, \qquad \rho_{i+1} = \Sigma_{(i)}^{\alpha\to 0}(\tau_{(i)}), \qquad \mathrm{d}_{i+1} = \frac{\mathrm{d}_i|_{\mathcal{T}_{i+1}\times\mathcal{T}_{i+1}}}{(\mu_i(\mathcal{T}_{i+1}))^\alpha}.$$

4. *Repeat within the subtree* $(\mathcal{T}_{i+1}, \mu_{i+1}, \rho_{i+1}, \mathrm{d}_{i+1})$ *by increasing* i *by* 1 *and proceeding to* 1.
5. *As* $i \to \infty$, *we obtain a sequence* $(\Sigma_{(i)}^{\alpha\to 0}(\tau_{(i)}), i \geq 1)$ *in* $\mathcal{T}$ *that increases in the sense that* $\Sigma_{(i)}^{\alpha\to 0}(\tau_{(i)}) \in [[\rho, \Sigma_{(i+1)}^{\alpha\to 0}(\tau_{(i+1)})]]$ *and hence converges. Let* $\Sigma = \lim_{i\to\infty} \Sigma_{(i)}^{\alpha\to 0}(\tau_{(i)})$.

Note that Step 2. is well-defined as $\prod_{t'\geq 0} P(S^{\Sigma_{(i)}}(t'), F_{\Sigma_{(i)}}(t')) = 0$, by Proposition 16.

Roughly speaking, this sampling procedure is that we travel along the spine $[[\rho, \Sigma_{(1)}]]$ and keep selecting subtrees until the first time we choose a subtree not containing $\Sigma_{(1)}$ and then repeat inductively in the subtree until we reach a leaf Σ in the limit, see Fig. 4. We show in the following proposition that there is a spinal subordinator associated with Σ.

Proposition 19. *Let* Σ *be sampled according to Procedure* 1.

(i) *The process* $(S^\Sigma, F_\Sigma, \mathcal{B}_{(\Sigma)}^{\alpha\to 0})$ *is a Poisson point process with intensity measure*

$$\widetilde{\nu}_{\mathrm{bush}}^P(\mathrm{d}\mathbf{s}, \mathrm{d}x, \mathrm{d}T) = \sum_{i\geq 1} P(\mathbf{s}, s_i)\delta_{s_i}(\mathrm{d}x) Q_{(s_1,\dots,s_{i-1},s_{i+1},\dots)}(\mathrm{d}T)\nu(\mathrm{d}\mathbf{s}).$$

Specifically, $((S^\Sigma(t), F_\Sigma(t), \mathcal{B}_{(\Sigma)}^{\alpha\to 0}(t)), 0 \leq t < \tau_{(1)})$ *is a killed Poisson point process with killing rate* $\int_{\mathcal{S}^\downarrow} \sum_{i\geq 1}(1-s_i)P(\mathbf{s}, s_i)\nu(\mathrm{d}\mathbf{s})$ *and intensity measure*

$$\widetilde{\nu}_{(1),\mathrm{bush}}^P(\mathrm{d}\mathbf{s}, \mathrm{d}x, \mathrm{d}T) = \sum_{i\geq 1} s_i P(\mathbf{s}, s_i)\delta_{s_i}(\mathrm{d}x) Q_{(s_1,\dots,s_{i-1},s_{i+1},\dots)}(\mathrm{d}T)\nu(\mathrm{d}\mathbf{s}).$$

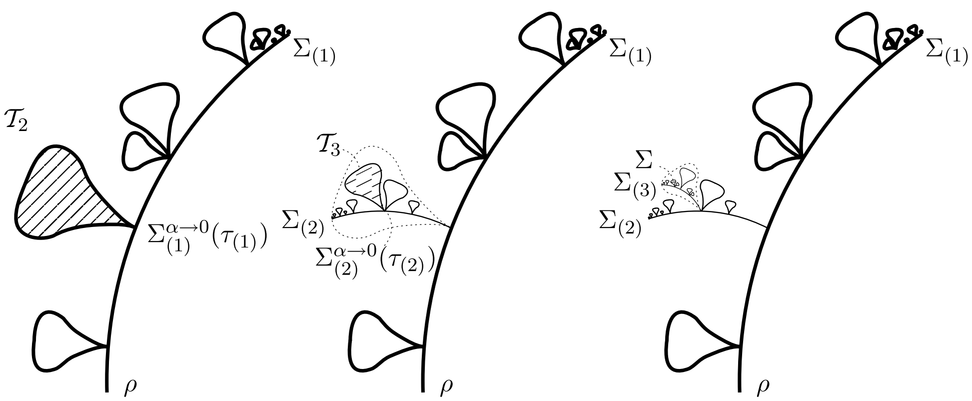


Fig. 4. In Procedure 1 we begin by sampling in $\mathcal{T}_1 = \mathcal{T}$ a leaf $\Sigma_{(1)} \sim \mu$ and pick one of the spinal subtrees as $\mathcal{T}_2$ according to SPF P. Within $\mathcal{T}_i$ for $i = 2$, rescaled, we repeat by sampling a leaf $\Sigma_{(i)} \sim \mu_i$ and pick spinal subtree $\mathcal{T}_{i+1}$ according to P. As $i \to \infty$ we indicate $\Sigma = \lim_{i\to\infty} \Sigma_{(i)}^{\alpha\to 0}(\tau_{(i)})$.

(ii) *Let* $\xi_t^\Sigma = -\log X_{(\Sigma)}^{\alpha\to 0}(t)$, $t \geq 0$. *Then* ξ^Σ *is a pure jump subordinator with Laplace exponent* Φ_Σ *and Lévy measure* Λ_Σ *given by*

$$\Phi_\Sigma(q) = \int_{\mathcal{S}^\downarrow} \sum_{i\geq 1} \left(1 - s_i^q\right) P(\mathbf{s}, s_i)\nu(\mathrm{d}\mathbf{s}) \quad \textit{and} \quad \Lambda_\Sigma = \int_{\mathcal{S}^\downarrow} \sum_{i\geq 1} P(\mathbf{s}, s_i)\delta_{-\log s_i}\nu(\mathrm{d}\mathbf{s}). \tag{9}$$

Proof. (i) This proof relies heavily on Poisson point process techniques. We use the terminology of Kingman [23]. By Proposition 16, the process $(S^{\Sigma_{(1)}}, F_{\Sigma_{(1)}}, \mathcal{B}_{(\Sigma_{(1)})}^{\alpha\to 0})$ is a Poisson point process with intensity measure $\widetilde{\nu}_{\mathrm{bush}}^*$. Step 2. of Procedure 1 can be read and analysed as follows. We mark some points of this Poisson point process with a selected subtree $\mathcal{T}_{(\Sigma_{(1)})}^{\mathrm{sel}}(t)$ using the kernel

$$K\big(\mathbf{s}, x, B'; \mathrm{d}T''\big) = P(\mathbf{s}, x)\delta_{(\{0\})}\big(\mathrm{d}T''\big) + \sum_{S \text{ connected component of } B'\setminus\{\rho'\}} P\big(\mathbf{s}, \mu'(S)\big)\delta_S\big(\mathrm{d}T''\big),$$

where B' is short for $(B', \mu', \rho', \mathrm{d}')$ and T'' is short for $(T'', \mu'', \rho'', \mathrm{d}'')$, also S for $(S, \mu'|_S, \rho', \mathrm{d}'|_{S\times S})$ and $\{0\}$ for $(\{0\}, 0, 0, 0)$. By standard marking and mapping, we get a new Poisson point process $(S^{\Sigma_{(1)}}, F_{\Sigma_{(1)}}, \mathcal{B}_{(\Sigma_{(1)})}^{\mathrm{rem}}, \mathcal{T}_{(\Sigma_{(1)})}^{\mathrm{sel}})$, where $\mathcal{B}_{(\Sigma_{(1)})}^{\mathrm{rem}}(t) = \mathcal{B}_{(\Sigma_{(1)})}^{\alpha\to 0}(t) \setminus \mathcal{T}_{(\Sigma_{(1)})}^{\mathrm{sel}}(t)$ with intensity measure

$$\sum_{i\geq 1} s_i\delta_{s_i}(\mathrm{d}x)\left(P(\mathbf{s}, s_i)Q_{\widehat{\mathbf{s}}^{(i)}}\big(\mathrm{d}B'\big)\delta_{\{0\}}\big(\mathrm{d}T''\big) + \sum_{j\neq i} P(\mathbf{s}, s_j)Q_{\widehat{\mathbf{s}}^{(i,j)}}\big(\mathrm{d}B'\big)Q_{s_j}\big(\mathrm{d}T''\big)\right)\nu(\mathrm{d}\mathbf{s}),$$

where $\widehat{\mathbf{s}}^{(i)} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots)$ is the sequence $\mathbf{s}$ with s_i removed and similarly $\widehat{\mathbf{s}}^{(i,j)}$ is the sequence $\mathbf{s}$ with s_i and s_j removed.

In Step 3., we set $\tau_{(1)} = \inf\{t \geq 0\colon \mathcal{T}_{(\Sigma_{(1)})}^{\mathrm{sel}}(t) \neq \{0\}\}$, exponentially distributed with rate

$$\int_{\mathcal{S}^\downarrow} \sum_{i\geq 1} s_i \sum_{j\neq i} P(\mathbf{s}, s_j)\nu(\mathrm{d}\mathbf{s}) = \int_{\mathcal{S}^\downarrow} \sum_{j\geq 1} (1 - s_j)P(\mathbf{s}, s_j)\nu(\mathrm{d}\mathbf{s}) < \infty,$$

note (8). Standard thinning and projecting yields that $((S^\Sigma(t), F_\Sigma(t), \mathcal{B}_{(\Sigma)}^{\alpha\to 0}(t)), 0 \leq t < \tau_{(1)}) = ((S^{\Sigma_{(1)}}(t), F_{\Sigma_{(1)}}(t), \mathcal{B}_{(\Sigma_{(1)})}^{\mathrm{rem}}(t)), 0 \leq t < \tau_{(1)})$ is an independently killed Poisson point process with intensity measure $\sum_{i\geq 1} s_i P(\mathbf{s}, s_i) \times \delta_{s_i}(\mathrm{d}x)Q_{\widehat{\mathbf{s}}^{(i)}}(\mathrm{d}B')\nu(\mathrm{d}\mathbf{s})$, as required for the second assertion. The rescaled tree $\mathcal{T}_2 = \mathcal{T}_{(\Sigma_{(1)})}^{\mathrm{sel}}(\tau_{(1)}) \sim Q_1$ is independent of this killed Poisson point process and also jointly independent of the pair formed by the bush $\mathcal{B}_{(\Sigma_{(1)})}^{\mathrm{rem}}$ and the rescaled tree $\mathcal{T}_{(\Sigma_{(1)})}^{\alpha\to 0}(\tau_{(1)})$ that has distribution Q_{s_i} for $s_i = F_{\Sigma_{(1)}}(\tau_{(1)})$, using the converse statement in Proposition 16.

In Step 4., the induction proceeds on $\mathcal{T}_i \sim Q_1$, $i \geq 2$, all independent of the past, so this Poisson point process extends indefinitely, but ignores points at $\tau_{(1)} + \cdots + \tau_{(i)}$, $i \geq 1$. These are exponentially spaced and i.i.d., hence form an independent Poisson point process. The independence and distributional properties that we noted identify the distribution of $(S^\Sigma(\tau_{(1)}), F_\Sigma(\tau_{(1)}), \mathcal{B}_{(\Sigma)}^{\alpha\to 0}(\tau_{(1)})) = (S^{\Sigma_{(1)}}(\tau_{(1)}), \mu^{\mathrm{sel}}(\mathcal{T}_{(\Sigma_{(1)})}^{\mathrm{sel}}(\tau_{(1)})), \widetilde{\mathcal{B}}_1)$, and the intensity measure

$$\sum_{i\geq 1} s_i \sum_{j\neq i} P(\mathbf{s}, s_j)\delta_{s_j}(\mathrm{d}x)Q_{(s_i, \widehat{\mathbf{s}}^{(i,j)})}\big(\mathrm{d}B'\big)\nu(\mathrm{d}\mathbf{s}) = \sum_{j\geq 1} (1 - s_j)P(\mathbf{s}, s_j)\delta_{s_j}(\mathrm{d}x)Q_{\widehat{\mathbf{s}}^{(j)}}\big(\mathrm{d}B'\big)\nu(\mathrm{d}\mathbf{s}),$$

because we define $\widetilde{\mathcal{B}}_i$ by grafting to the same root $\mathcal{B}_{(\Sigma_{(i)})}^{\mathrm{rem}}(\tau_{(i)}) \sim Q_{\hat{\mathbf{s}}^{(i,j)}}$ and the rescaled $\mathcal{T}_{(\Sigma_{(i)})}^{\alpha\to 0}(\tau_{(i)})$ has distribution Q_{s_i}. Standard superposition completes the proof of (i).

(ii) By (i) and standard mapping, $(\Delta\xi_t^\Sigma, t \geq 0)$ is a Poisson point process with intensity measure Λ_Σ, hence $\xi_t^\Sigma = \sum_{s\leq t} \Delta\xi_s^\Sigma$ is a pure jump subordinator with Laplace exponent Φ_Σ. $\square$

4.3. *A procedure to sample a sequence of leaves from a self-similar CRT*

In this section, we formulate a special inductive procedure to sample k leaves $\Sigma_1, \dots, \Sigma_k$ from a self-similar CRT $(\mathcal{T}, \mu)$ with characteristic pair (α, ν), where

$$\nu(\mathrm{d}\mathbf{s}) = \sum_{j\geq 1}\left(\sum_{i\geq 1} s_i^j(1-s_i)\right)\nu_j(\mathrm{d}\mathbf{s})$$

for some measures ν_j, $j \geq 1$, representing a RE dislocation measure as in Corollary 5. Clearly, the measures ν_j, $j \geq 1$, are absolutely continuous with respect to ν. We denote their Radon–Nikodym derivatives by $f_j = \mathrm{d}\nu_j/\mathrm{d}\nu$, $j \geq 1$, and define selection functions

$$P_0(\mathbf{s}, s_i) = \sum_{\ell\geq 1} s_i^\ell(1-s_i)f_\ell(\mathbf{s}), \qquad P_k^{\mathrm{old}}(\mathbf{s}, s_i) = \frac{\sum_{\ell\geq k+1} s_i^\ell(1-s_i)f_\ell(\mathbf{s})}{\sum_{\ell\geq k} s_i^\ell(1-s_i)f_\ell(\mathbf{s})}, \quad k \geq 1, \quad \text{and}$$

$$\text{for } j \neq i \quad P_k^{\mathrm{new}}(\mathbf{s}, s_i, s_j) = \frac{s_i^k s_j f_k(\mathbf{s})}{\sum_{\ell\geq k} s_i^\ell(1-s_i)f_\ell(\mathbf{s})}, \quad k \geq 1.$$

Procedure 2.

(0) *To sample Σ_1 in the whole CRT $(\mathcal{T}_{1,\varnothing}, \mu_{1,\varnothing}, \rho_{1,\varnothing}, \mathrm{d}_{1,\varnothing}) = (\mathcal{T}, \mu, \rho, \mathrm{d})$ we use step $(k,\varnothing)$ for $k = 1$ and then proceed inductively.*

$(k,\varnothing)$ *Sample leaf Σ_k in $\mathcal{T}_{k,\varnothing}$ according to Procedure 1 using the SPF P_0. Then increase k by 1, set $B = [k-1]$ and $\mathcal{T}_{k,B} = \mathcal{T}$, and proceed to step (k,B).*

(k,B) *with $B \neq \varnothing$.*

1. *Given $\Sigma_i \in \mathcal{T}_{k,B}$, $i \in B$, denote by $v_{k,B}$ the branch point that separates the labels in B into several subtrees, so that $[[\rho_{k,B}, v_{k,B}]] = \bigcap_{i\in B}[[\rho_{k,B}, \Sigma_i]]$.*
2. *Conditionally given $(\mathcal{T}_{k,B}; \Sigma_i, i \in B)$, with spine $[[\rho_{k,B}, v_{k,B}[[= \{v_{k,B}^{\alpha\to 0}(t), 0 \leq t < \zeta_{v_{k,B}}^{\alpha\to 0}\}$, pick as $\mathcal{T}_{k,B'}$ either a new subtree $\mathcal{S}$ above some $v_{k,B}^{\alpha\to 0}(t)$ with probability*

$$\begin{aligned}&\mathbb{P}(\mathcal{T}_{k,B'} = \mathcal{S} | \mathcal{T}_{k,B}; \Sigma_i, i \in B)\\ &\quad = P_{\#B}^{\mathrm{new}}\left(S^{v_{k,B}}(t), F_{v_{k,B}}(t), \frac{\mu_{k,B}(\mathcal{S})}{\mu_{k,B}(\mathcal{T}_{(v_{k,B})}^{\alpha\to 0}(t-))}\right)\prod_{t'<t} P_{\#B}^{\mathrm{old}}\big(S^{v_{k,B}}(t'), F_{v_{k,B}}(t')\big),\end{aligned}$$

or, in the case $\#B \geq 2$, a new or old subtree $\mathcal{S}$ above $v_{k,B}$ with probability

$$\mathbb{P}(\mathcal{T}_{k,B'} = \mathcal{S} | \mathcal{T}_{k,B}; \Sigma_i, i \in B) = \frac{\mu_{k,B}(\mathcal{S})}{\mu_{k,B}(\mathcal{T}_{(v_{k,B})}^{\alpha\to 0}(\zeta_{v_{k,B}}-))}\prod_{t'<\zeta_{v_{k,B}}} P_{\#B}^{\mathrm{old}}\big(S^{v_{k,B}}(t'), F_{v_{k,B}}(t')\big),$$

where $B' = \{i \in B\colon\ \Sigma_i \in \mathcal{S}\}$ and new/old means without/with any Σ_i, $i \in B$.

3. *Let $\tau_{k,B} = \min\{\zeta_{v_{k,B}}^{\alpha\to 0}, \inf\{t \geq 0\colon\ \mathcal{T}_{k,B'} \cap \mathcal{T}_{v_{k,B}}^{\alpha\to 0}(t) = \varnothing\}\}$. We turn $\mathcal{T}_{k,B'}$ into a CRT with rescaled mass measure, root and rescaled distance function as follows*:

$$\mu_{k,B'} = \frac{\mu_{k,B}|_{\mathcal{T}_{k,B'}}}{\mu_{k,B}(\mathcal{T}_{k,B'})}, \qquad \rho_{k,B'} = v_{k,B}^{\alpha\to 0}(\tau_{k,B}), \qquad \mathrm{d}_{k,B'} = \frac{\mathrm{d}_{k,B}|_{\mathcal{T}_{k,B'}\times\mathcal{T}_{k,B'}}}{(\mu_{k,B}(\mathcal{T}_{k,B'}))^\alpha}.$$

4. *Repeat within the subtree $(\mathcal{T}_{k,B'}, \mu_{k,B'}, \rho_{k,B'}, \mathrm{d}_{k,B'})$ by proceeding to step (k,B').*

Note that the probabilities in Step 2. add up to 1 since $\sum_{j:j\neq i} s_i^k s_j f_k(\mathbf{s}) = s_i^k(1-s_i)f_k(\mathbf{s})$. From Proposition 19, we obtain the following by straightforward arguments.

Corollary 20. *Sample $(\Sigma_k, k \geq 1)$ following Procedure 2. Let v_k be the branch point in $\mathcal{T}$ that separates $[k]$ into different subtrees, $k \geq 1$. Then $((S^{\Sigma_k}(t), F_{\Sigma_k}(t), \mathcal{B}^{\alpha\to 0}_{(\Sigma_k)}(t)), 0 \leq t < \zeta^{\alpha\to 0}_{v_k})$ is a Poisson point process with killing rate $\lambda_k = \int_{\mathcal{S}^\downarrow} \sum_{i\geq 1} \sum_{\ell=1}^{k-1} s_i^\ell (1-s_i)\nu_\ell(\mathrm{d}\mathbf{s})$ and intensity measure*

$$\widetilde{\nu}^{(k)}_{\mathrm{bush}}(\mathrm{d}\mathbf{s}, \mathrm{d}x, \mathrm{d}T) = \sum_{i\geq 1} \delta_{s_i}(\mathrm{d}x) Q_{(s_1,\dots,s_{i-1},s_{i+1},\dots)}(\mathrm{d}T) \sum_{\ell\geq k} s_i^\ell (1-s_i)\nu_\ell(\mathrm{d}\mathbf{s}). \tag{10}$$

Note $\lambda_1 = 0$, so the Poisson point process is not killed and Corollary 20 describes the whole tree $\mathcal{T}$ jointly with Σ_1, decomposed along its spine $[[\rho, \Sigma_1[[$. For $k \geq 2$, Corollary 20 describes a spinal decomposition along $[[\rho, v_k[[$, but not the subtrees above v_k. This is done in Lemma 21.

Proof of Corollary 20. The case $k = 1$ follows straight from step $(1, \varnothing)$ of Procedure 2 and Proposition 19. We then proceed by induction in k. Assuming that the statement is true for k, step $(k+1, [k])$ 2. and standard thinning with probabilities $P^{\mathrm{old}}_{k+1}(\mathbf{s}, s_i)$ yields

$$\widetilde{\nu}^{(k+1)}_{\mathrm{bush}}(\mathrm{d}s, \mathrm{d}x, \mathrm{d}T) = \sum_{i\geq 1} P^{\mathrm{old}}_{k+1}(\mathbf{s}, s_i)\delta_{s_i}(\mathrm{d}x) Q_{(s_1,\dots,s_{i-1},s_{i+1},\dots)}(\mathrm{d}T) \sum_{\ell\geq k} s_i^\ell (1-s_i)\nu_\ell(\mathrm{d}\mathbf{s}),$$

as claimed, and an extra rate $\int_{\mathcal{S}^\downarrow} \sum_{i\geq 1}(1 - P^{\mathrm{old}}_{k+1}(\mathbf{s}, s_i)) \sum_{\ell\geq k} s_i^\ell (1-s_i)\nu_\ell(\mathrm{d}\mathbf{s})$ is added to the killing rate λ_k from the induction hypothesis. This completes the induction step. □

To identify the distribution $Q_1^{[k]}$ of $(\mathcal{T}; \Sigma_i, i \in [k])$ constructed according to Procedure 2 run up to some $k \geq 2$, we study its branching structure recursively by specifying the first branch point v_k that separates $[k]$ into several subtrees denoted by $\mathcal{T}^{[k]}_\ell$ with label partition $\Pi^{[k]}$ and a remaining bush $\mathcal{B}_{[k]}$ of unlabelled subtrees, with joint relative subtree sizes $S^{[k]} \in \mathcal{S}^\downarrow$. For $x \in (0, 1]$ and $B = \{b_1, \dots, b_k\} \subset \mathbb{N}$ with $1 \leq b_1 < \dots < b_k$, it will be convenient to denote by Q_x^B the distribution of a rescaled and relabelled version of $(\mathcal{T}; \Sigma_i, i \in [k])$, where the mass measure has been multiplied by x, the distance function by x^α, and Σ_i is renamed to Σ_{b_i}, $i \in [k]$.

Lemma 21. *The first branching of $(\mathcal{T}; \Sigma_i, i \in [k])$ separating $[k]$ and associated subtrees described in $\mathcal{C}^{\mathrm{br}}_k = (S^{[k]}, \Pi^{[k]}, \mathcal{T}^{[k]}, \mathcal{B}_{[k]})$ are independent of $\mathcal{C}^{\mathrm{pre}}_k = ((S^{v_k}(t), F_{v_k}(t), \mathcal{B}^{\alpha\to 0}_{(v_k)}(t)), 0 \leq t < \zeta_{v_k})$, with distribution given by*

$$\mathbb{P}\big(S^{[k]} \in \mathrm{d}\mathbf{s}, \Pi^{[k]} = \pi, \big(\mathcal{T}^{[k]}_1; \Sigma_i, i \in \pi_1\big) \in \mathrm{d}T_1, \dots, \big(\mathcal{T}^{[k]}_r; \Sigma_i, i \in \pi_r\big) \in \mathrm{d}T_r, \mathcal{B}_{[k]} \in \mathrm{d}B'\big)$$
$$= \frac{1}{\lambda_k}\left(\sum_{i_1,\dots,i_r \text{ distinct}} Q_{\widehat{\mathbf{s}}^{(i_1,\dots,i_r)}}\big(\mathrm{d}B'\big) \prod_{\ell=1}^r s_{i_\ell}^{\#\pi_\ell} Q^{\pi_\ell}_{s_{i_\ell}}(\mathrm{d}T_\ell)\right)\nu_m(\mathrm{d}\mathbf{s}),$$

where $\pi = (\pi_{i_1}, \dots, \pi_r) \in \mathcal{P}_k$ and $m = \min \pi_2 - 1$, also $\widehat{\mathbf{s}}^{(i_1,\dots,i_r)}$ is $\mathbf{s}$ with $s_1, \dots, s_{i_r}$ removed.

The kernel $\kappa_{\mathbf{s},\pi}(\mathrm{d}T_1, \dots, \mathrm{d}T_r, \mathrm{d}B') = \sum_{i_1,\dots,i_r \text{ distinct}} Q_{\widehat{\mathbf{s}}^{(i_1,\dots,i_r)}}(\mathrm{d}B') \prod_{\ell=1}^r s_{i_\ell}^{\#\pi_\ell} Q^{\pi_\ell}_{s_{i_\ell}}(\mathrm{d}T_\ell)$ is a fancy paintbox that equips each block under $\kappa_{\mathbf{s}}$ with a tree and embeds the labels for $\pi \in \mathcal{K}$.

Proof of Lemma 21. For $k = 1$, this is trivial since $v_1 = \Sigma_1$ is a leaf. Now suppose that the result holds for all $[j] \subseteq [k]$, and consider $k + 1$. In our use of standard Poisson point process arguments as well as in extracting from Procedure 2 as from Procedure 1, we build on the proof of Proposition 19.

For $\pi \in \mathcal{P}_{k+1} \setminus \{\mathbf{1}_{[k+1]}\}$, let $A_\pi = \{\Pi^{[k+1]} = \pi\}$ be the event that v_{k+1} splits $[k+1]$ into π. The simplest case is for $\pi = \{[k], \{k+1\}\}$. By Corollary 20, the decomposition of $\mathcal{T}$ along the spine $[[\rho, v_k[[$ is given by the Poisson point process $((S^{\Sigma_k}(t), F_{\Sigma_k}(t), \mathcal{B}^{\alpha\to 0}_{(\Sigma_k)}(t)), 0 \leq t < \zeta^{\alpha\to 0}_{v_k})$ with intensity measure (10), killed at rate $\lambda_k = \int_{\mathcal{S}^\downarrow} \sum_{i\geq 1} \sum_{\ell=1}^{k-1} s_i^\ell (1-s_i)\nu_\ell(\mathrm{d}\mathbf{s})$. By comparison with the statement of Corollary 20 for $k+1$, we see $\mathbb{P}(A_{\{[k],\{k+1\}\}}) = 1 - \lambda_k/\lambda_{k+1}$. *Conditionally given* $A_{\{[k],\{k+1\}\}}$, the distribution of $(S^{\Sigma_k}(\tau_{k+1,[k]}), F_{\Sigma_k}(\tau_{k+1,[k]}), \mathcal{B}^{\mathrm{rem}}_{(\Sigma_k)}(\tau_{k+1,[k]}),$

$\mathcal{T}^{\mathrm{sel}}_{k+1,[k]}(\tau_{k+1,[k]}))$ is

$$\frac{1}{\lambda_{k+1}-\lambda_k}\sum_{i\geq 1}\sum_{j\neq i}P^{\mathrm{new}}_k(\mathbf{s},s_i,s_j)\delta_{s_i}(\mathrm{d}x)Q_{\widehat{\mathbf{s}}^{(i,j)}}\big(\mathrm{d}B'\big)Q_{s_j}\big(\mathrm{d}T''\big)\sum_{\ell\geq k}s_i^\ell(1-s_i)\nu_\ell(\mathrm{d}\mathbf{s})$$

$$=\frac{1}{\lambda_{k+1}-\lambda_k}\sum_{i\geq 1}\sum_{j\neq i}\delta_{s_i}(\mathrm{d}x)Q_{\widehat{\mathbf{s}}^{(i,j)}}\big(\mathrm{d}B'\big)Q_{s_j}\big(\mathrm{d}T''\big)s_i^k s_j\nu_k(\mathrm{d}\mathbf{s}), \tag{11}$$

independently of the rescaled $(\mathcal{T}^{\alpha\to 0}_{(\Sigma_k)}(\tau_{k+1,[k]});\ \Sigma_i, i\in[k])$ that has $Q_1^{[k]}$ as conditional distribution given $A_{\{[k],\{k+1\}\}}$. Note also, that the sampling of Σ_{k+1} in the rescaled $\mathcal{T}^{\mathrm{sel}}_{k+1,[k]}(\tau_{k+1,[k]})$ yields conditional distribution $Q_1^{\{k+1\}}$ given $A_{\{[k],\{k+1\}\}}$, and that by standard thinning arguments these are conditionally independent of $((S^{\Sigma_{k+1}}(t), F_{\Sigma_{k+1}}(t), \mathcal{B}^{\alpha\to 0}_{(\Sigma_{k+1})}(t)), 0\leq t<\zeta^{\alpha\to 0}_{v_{k+1}})$ given $A_{\{[k],\{k+1\}\}}$. Multiplying by $\mathbb{P}(A_{\{[k],\{k+1\}\}})$, this yields the result for $\pi=\{[k],\{k+1\}\}$.

Now consider any other $\pi=\{\pi_1,\dots,\pi_r\}\in\mathcal{P}_{k+1}\setminus\{\mathbf{1}_{[k+1]}\}$ and write $m=\min\pi_2-1\in[k-1]$. Note that also $m=\min\pi_2\cap[k]-1$. By the induction hypothesis, the collections $\mathcal{C}_k^{\mathrm{pre}}$ describing the spine to the branch point separating $[k]$, and $\mathcal{C}_k^{\mathrm{br}}$ describing the branching and rescaled subtrees, are independent. We read and analyse Step 2. of Procedure 2 by marking $\mathcal{C}_k^{\mathrm{pre}}$ as we marked the Poisson point process in the proof of Proposition 19 *and* similarly and independently selecting a new or old subtree $\mathcal{S}$ above v_k with probability

$$\mathbb{P}\big(\mathcal{T}^{\mathrm{sel}}=\mathcal{S}|\mathcal{T}_{k,B};\ \Sigma_i, i\in B\big)=\frac{\mu_{k,B}(\mathcal{S})}{\mu_{k,B}(\mathcal{T}^{\alpha\to 0}_{(v_{k,B})}(t-))}.$$

Then A_π is an intersection of two independent events $A_\pi=A_k^{\mathrm{pre}}\cap A_\pi^{\mathrm{br}}$ given by

$$A_k^{\mathrm{pre}}=\big\{\mathcal{T}^{\mathrm{sel}}_{v_k}=\{0\}\text{ for all }0\leq t<\zeta_{v_k}\big\}\quad\text{and}\quad A_\pi^{\mathrm{br}}=\big\{\mathcal{L}_k\big(\mathcal{T}^{\mathrm{sel}}\big)=\pi_{(k+1)}\cap[k]\big\},$$

where $\mathcal{L}_k(\mathcal{S})=\{i\in[k]\colon\ \Sigma_i\in\mathcal{S}\}$ and $\pi_{(k+1)}$ is the block of π containing $k+1$. By construction, $(\mathcal{C}_k^{\mathrm{pre}}, A_k^{\mathrm{pre}})$ and $(\mathcal{C}_k^{\mathrm{br}}, A_\pi^{\mathrm{br}})$ are also independent and, since the random variables used to sample Σ_{k+1} in $\mathcal{T}^{\mathrm{sel}}$ are conditionally independent of $(\mathcal{C}_k^{\mathrm{pre}}, A_k^{\mathrm{pre}})$ given $\mathcal{T}^{\mathrm{sel}}$, also $\mathcal{C}_{k+1}^{\mathrm{br}}$ is independent of $(\mathcal{C}_k^{\mathrm{br}}, A_\pi^{\mathrm{br}})$, hence of $\mathcal{C}_{k+1}^{\mathrm{br}}$, since on A_π^{br}, we have $\mathcal{C}_{k+1}^{\mathrm{br}}=\mathcal{C}_k^{\mathrm{br}}$. The distribution of $\mathcal{C}_{k+1}^{\mathrm{br}}$ now follows from the conditional distribution of $\mathcal{T}^{\mathrm{sel}}$ given $\mathcal{C}_k^{\mathrm{br}}$, the recursive nature of Procedure 2 and the stability of the procedure under increasing bijections from $[j]$ to other sets $B\subset\mathbb{N}$ with $\#B=j$ that allows us to apply the induction hypothesis to obtain that the sampling of Σ_{k+1} in the rescaled $\mathcal{T}^{\mathrm{sel}}\sim Q_1^B$ yields a tree with rescaled distribution $Q_1^{B\cup\{k+1\}}$, as required. □

Inductively, Lemma 21 yields a subtree decomposition of $(\mathcal{T};\Sigma_1,\dots,\Sigma_k)$. For $\varnothing\neq B\subset[k]$, consider $[[\rho_{k,B},v_{k,B}[[=\{v_B^{\alpha\to 0}(t), 0<t<\zeta_B^{\alpha\to 0}\}$ in $\mathcal{T}_{k,B}\subset\mathcal{T}$, *branch* B, as in Procedure 2 (cf. Fig. 5, where $T_3=\{[3],\{2,3\},\{1\},\{2\},\{3\}\}$). For the rescaled $\mathcal{T}_{k,B}\sim Q_1^B$, i.e. $Q_1^{[\#B]}$ pushed forward under the increasing bijection $[\#B]\to B$, Corollary 20 gives the distribution of the analogous point process $((S^B(t), F_B(t), \mathcal{B}_B^{\alpha\to 0}(t)), 0\leq t<\zeta_B^{\alpha\to 0})$ that captures the spinal subtrees off $[[\rho_{k,B},v_{k,B}[[$. The remaining split at $v_{k,B}$ into relative sizes S^B, of which F_i^B is the size corresponding to the ith block of the split Π^B of B and $\mathcal{B}_B$ is the subbush of unlabelled subtrees of the remaining sizes in S^B, can be read from Lemma 21, as Q_1^B is just a push-forward of $Q_1^{[\#B]}$.

Corollary 22 (Subtree decomposition). *The discrete tree shapes T_k, $k\geq 1$, of the reduced trees $R(\mathcal{T};\Sigma_1,\dots,\Sigma_k):=[[\rho,\Sigma_1]]\cup\dots\cup[[\rho,\Sigma_k]]$, $k\geq 1$, are labelled Markov branching trees with*

$$\mathbb{P}\big(\Pi^{[k]}=\pi\big)=\frac{1}{\lambda_k}\int_{S^\downarrow}\kappa_{\mathbf{s}}\big(\mathcal{P}^\pi\big)\nu_m(\mathrm{d}\mathbf{s}),\quad\text{where } m=\min\pi_2-1. \tag{12}$$

Conditionally given $T_k=\mathbf{t}_k$, $\Pi^B=\pi^B=(\pi_1^B,\dots,\pi_r^B)$ with $m^B=\min\pi_2^B-1$, $B\in\mathbf{t}_k$,

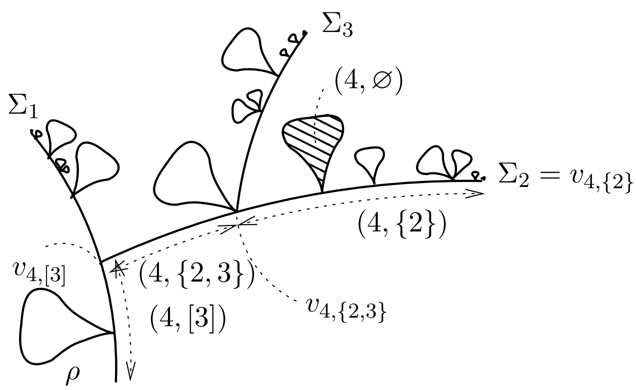


Fig. 5. Sampling Σ_4 in $(\mathcal{T}; \Sigma_i, i \in [3])$ using Procedure 2: first step is $(4, [3])$, random selection picks the old subtree $\mathcal{T}_{4,\{2,3\}}$; in step $(4, \{2, 3\})$, the old subtree $\mathcal{T}_{4,\{2\}}$ is picked; in step $(4, \{2\})$, a new subtree, shaded, is picked; step $(4, \varnothing)$ takes place in this subtree, using Procedure 1.

- *the processes* $((S^B(t), F_B(t), \mathcal{B}_B^{\alpha\to 0}(t)), 0 \le t < \zeta_B^{\alpha\to 0})$, *with distribution as in Corollary* 20,
- *and the variables* $(S^B, F_1^B, \ldots, F_r^B, \mathcal{B}_B)$, *with distribution*

$$\frac{1}{\lambda_{\#B}\mathbb{P}(\Pi^B = \pi^B)}\left(\sum_{i_1,\ldots,i_r \text{ distinct}} Q_{\widehat{\mathbf{s}}^{(i_1,\ldots,i_r)}}(\mathrm{d}B') \prod_{\ell=1}^{r} s_{i_\ell}^{\#\pi_\ell^B} \delta_{s_{i_\ell}}(\mathrm{d}x_\ell)\right)\nu_{m^B}(\mathrm{d}\mathbf{s}),$$

$B \in \mathbf{t}_k$, *are independent. The tree* $(\mathcal{T}; \Sigma_1, \ldots, \Sigma_k)$ *with* k *leaves sampled via Procedure* 2 *is a measurable function of* $(T_k; ((F_B, S^B, \mathcal{B}_B), ((S^B(t), F_B(t), \mathcal{B}_B^{\alpha\to 0}(t)), 0 \le t < \zeta_B^{\alpha\to 0})), B \in T_k)$.

Proof of Theorem 6. We will show that Procedure 2 provides an embedding for a RE hierarchy as in Corollary 5, provided that $\int_{S^\downarrow}(1 - s_1)\nu(\mathrm{d}\mathbf{s}) < \infty$ and $\nu(s_0 > 0) = c_j = k_j = 0$, $j \ge 1$.

A RE hierarchy is uniquely determined by its restrictions to $[k]$, $k \ge 1$. But the formula for κ in Corollary 5 is identical to (12), hence the hierarchy constructed via Procedure 2 is a RE hierarchy associated with $(\nu_j, j \ge 1)$ embedded in a CRT with characteristic pair (α, ν). $\square$

5. Scaling limits, proof of Theorem 7

5.1. *Asymptotics of block numbers in Gnedin's constrained partitions*

Before we describe Gnedin's framework and provide a slight extension of his asymptotic study, let us state the pth order renewal theory result that we need for this.

Lemma 23 (Gut [15], Theorem 2.3(b)). *Let* $N_t = \#\{n \ge 1\colon X_1 + \cdots + X_n \le t\}$ *be the renewal process associated with independent and identically distributed* $X_j > 0$. *Then for all* $p \in \mathbb{N}$

$$\mathbb{E}\left[\frac{N_t^p}{t^p}\right] \to \frac{1}{(\mathbb{E}[X_1])^p} \in [0, \infty), \quad \textit{as } t \to \infty.$$

Gnedin [13] introduced a *constrained paintbox* based on an $\mathbb{N}$-valued deterministic sequence $\psi = (\psi_k, k \ge 1)$ and a strictly decreasing random sequence $(G_k, k \ge 0)$ in $[0, 1]$ with $G_0 = 1$ and $\lim_{k\to\infty} G_k = 0$. Specifically, he considers a sequence $(I_n, n \ge 1)$ of independent uniform random variables on $[0, 1]$ independent of (G_k), but then associates a modified sequence $(\overline{I}_n^\psi, n \ge 1)$ that is constrained so that its lower records follow $(G_k, k \ge 1)$ with multiplicities given by $(\psi_k, k \ge 1)$:

- Set $\overline{I}_1^{\psi}=\cdots=\overline{I}_{\psi_1}^{\psi}=G_1$; inductively, consider the number K_n^{ψ} of records G_k that have been attained ψ_k times by $(\overline{I}_1^{\psi},\ldots,\overline{I}_n^{\psi})$, and the number R_n^{ψ} of times that $G_{K_n^{\psi}+1}$ has been attained by $(\overline{I}_1^{\psi},\ldots,\overline{I}_n^{\psi})$; for $n=\psi_1$, we have $K_n^{\psi}=1$ and $R_n^{\psi}=0$; this is the base case.
- Given $(\overline{I}_1^{\psi},\ldots,\overline{I}_n^{\psi})$, $K_n^{\psi}=k\geq 1$ and $R_n^{\psi}=r\in\{0,\ldots,\psi_{k+1}-1\}$, proceed as follows
  - if $I_{n+1}\in[G_k,1]$, let $\overline{I}_{n+1}^{\psi}=I_{n+1}$, $K_{n+1}^{\psi}=K_n^{\psi}$ and $R_{n+1}^{\psi}=R_n^{\psi}$;
  - if $I_{n+1}\in[0,G_k)$ and $r\leq\psi_{k+1}-2$, let $\overline{I}_{n+1}^{\psi}=G_{k+1}$, $K_{n+1}^{\psi}=K_n^{\psi}$ and $R_{n+1}^{\psi}=R_n^{\psi}+1$;
  - if $I_{n+1}\in[0,G_k)$ and $r=\psi_{k+1}-1$, let $\overline{I}_{n+1}^{\psi}=G_{k+1}$, $K_{n+1}^{\psi}=K_n^{\psi}+1$ and $R_{n+1}^{\psi}=0$.

Eventually, each G_k will appear ψ_k times as lower record in $(\overline{I}_n^{\psi},n\geq 1)$. Let $J_n^{\psi}=K_n^{\psi}+\mathbf{1}_{\{R_n^{\psi}>0\}}$ be the number of records attained by the n first terms of the sequence. Gnedin obtains the asymptotics of J_n^{ψ} when $G_k=Y_1\cdots Y_k$, where Y_k, $k\geq 1$, are i.i.d. in $(0,1)$ with $\mathbb{E}[-\log Y_1]<\infty$ and $\mathrm{Var}(-\log Y_1)<\infty$. Here we drop the requirement of finite logarithmic moments.

Lemma 24. *Let $G_k=Y_1\cdots Y_k$, where Y_k, $k\geq 1$, are i.i.d. in $(0,1)$. If $\psi=(\psi_k,k\geq 1)$ is such that $\psi_k\in\mathbb{N}$, $k\geq 1$ and*

$$\log\left(\sum_{j=1}^{k}\psi_j\right)=\mathrm{o}(k),\quad \text{as } k\to\infty,$$

then

$$\lim_{n\to\infty}\frac{J_n^{\psi}}{\log n}=\frac{1}{\mathbb{E}[-\log Y_1]}$$

in the sense that this limit vanishes when $\mathbb{E}[-\log Y_1]=\infty$. Furthermore, for every $p\geq 1$,

$$\limsup_{n\to\infty}\mathbb{E}\left[\left(\frac{J_n^{\psi}}{\log n}\right)^p\right]<\infty.$$

Proof. The case $\mathrm{Var}(-\log Y_1)<\infty$, and implicitly also $\mathbb{E}[-\log Y_1]<\infty$, has been shown in the proof of Proposition 8 of [13]. Now let $\mathbb{E}[-\log Y_1]=\infty$. Define $J_n'=\#\{k\geq 1\colon\ G_k\geq 1/n\}=\#\{k\geq 1\colon\ \sum_{i=1}^k(-\log Y_i)\leq\log n\}$. By the Renewal Theorem, see e.g. Theorem 4.1 in Chapter 3 of [11], we have $J_n'/\log n\to 0$ a.s. when $\mathbb{E}[-\log Y_1]=\infty$. Let $I_{1,n}<\cdots<I_{n,n}$ be the order statistics of $I_1,\ldots,I_n$. Define β_n by $I_{\beta_n,n}<1/n<I_{\beta_n+1,n}$. According to Gnedin's discussion, J_n' and β_n are independent, β_n is binomial$(n,1/n)$ and $J_n^{\psi}\leq J_n'+\beta_n$. By Markov's inequality, we have for all $\epsilon>0$,

$$\mathbb{P}(\beta_n>\epsilon\log n)=\mathbb{P}\big(\mathrm{e}^{2\beta_n/\epsilon}>n^2\big)\leq\frac{\mathbb{E}[\mathrm{e}^{2\beta_n/\epsilon}]}{n^2}=\frac{1}{n^2}\left(1+\frac{\mathrm{e}^{2/\epsilon}-1}{n}\right)^n.$$

Hence, $\sum_{n\geq 1}\mathbb{P}(\beta_n>\epsilon\log n)<\infty$. The Borel–Cantelli Lemma now yields $\lim_{n\to\infty}\beta_n/\log n=0$ a.s. This gives us $\limsup_{n\to\infty}J_n^{\psi}/\log n=0$ when $\mathbb{E}[-\log Y_1]=\infty$. Finally, for $p\geq 1$,

$$\mathbb{E}\left[\left(\frac{J_n^{\psi}}{\log n}\right)^p\right]\leq\mathbb{E}\left[\left(\frac{J_n'+\beta_n}{\log n}\right)^p\right]\leq 2^{p-1}\left(\mathbb{E}\left[\left(\frac{J_n'}{\log n}\right)^p\right]+\mathbb{E}\left[\left(\frac{\beta_n}{\log n}\right)^p\right]\right).$$

The first term is bounded (Lemma 23), the second tends to 0 (β_n have bounded moments). $\square$

5.2. *Special branch points and their asymptotics*

We consider the setting of Theorem 7, where $m=\inf\{n\geq 1\colon\ \nu_j=\nu_n \text{ for all } j\geq n\}<\infty$. In this setting, the selection probabilities of Section 4.3 for $k\geq m+1$ become

$$P_k^{\mathrm{old}}(\mathbf{s},s_i)=s_i\quad\text{and}\quad P_k^{\mathrm{new}}(\mathbf{s},s_i,s_j)=s_j.$$

It is now easy to see that the sampling procedure in $(\mathcal{T}, \mu)$ can be simplified in this setting so as to combine for each $k \geq m$ the steps until $\#B' < m$ into a single selection according to μ.

Procedure 3. *Use the steps of Procedure 2, but instead of steps* $(k,[k-1])$1.–2., *use the following steps for* $k \geq m$:

1′. *Given* $(\mathcal{T}; \Sigma_i \in [k-1])$, *sample* $\Sigma_k^* \sim \mu$ *independently.*
2′. *We consider the spine* $[[\rho, \Sigma_k^*[[= \{\Sigma_k^{*\alpha\to 0}(t), t \geq 0\}$ *and set*

$$\mathcal{T}_{k,B'} = \mathcal{T}_{(\Sigma_k^*)}^{\alpha\to 0}(\tau_k^*), \quad \textit{where } \tau_k^* = \inf\{t \geq 0\colon \#\mathcal{L}_{k-1}(\mathcal{T}_{(\Sigma_k^*)}^{\alpha\to 0}(t)) < m\},\ B' = \mathcal{L}_{k-1}(\mathcal{T}_{(\Sigma_k^*)}^{\alpha\to 0}(\tau_k^*)),$$

and $\mathcal{L}_{k-1}(\mathcal{S}) = \{i \in [k-1]\colon \Sigma_i \in \mathcal{S}\}$ *is the set of labels in* $\mathcal{S} \subseteq \mathcal{T}$.

Theorem 7 describes the convergence of unlabelled trees. In fact, more is true and it will be instructive to study approximations of the spines $[[\rho, \Sigma_j[[$, $j \geq 1$, in $(\mathcal{T}; \Sigma_i, i \in \mathbb{N})$ by discrete spines $\{B \in T_n\colon j \in B\}$, $n \geq j \geq 1$. In the proof of Theorem 7 we will need to control these uniformly in $j \geq 1$. In the exchangeable case, these spines can be regarded as independent uniform samples from a strongly sampling consistent regenerative interval partition [18]. In the RE case here, the analogous partitions will no longer be regenerative (except for $j = 1$, and for $j = 2$ if $m = 2$) and the sampling is not independent uniform. However, both features are still present on parts of the spine and we will cut the spines at certain *special branch points*.

Fix $j \geq 1$. A branch point $v \in [[\rho, \Sigma_j[[$ is called *special in* $(\mathcal{T}; \Sigma_i, i \in \mathbb{N})$ *for* $[[\rho, \Sigma_j[[$ if some or all of the m smallest labels $\mathcal{L}(\mathcal{T}^v)$ in the bush $\mathcal{T}^v$ above v are *not* included in the subtree $\mathcal{T}_{(\Sigma_j)}(\mathrm{d}(\rho, v))$ above v containing Σ_j. Note that a branch point v is special iff at v the m smallest labels split or j splits from the m smallest labels. In particular, a branch point that is special for $[[\rho, \Sigma_j[[$ and an element of $[[\rho, \Sigma_{j'}[[$ for some $j' < j$ may not be special for $[[\rho, \Sigma_{j'}[[$. For the analogous notion in $(\mathcal{T}; \Sigma_i, i \in [n])$, for $n \geq j$, we write

$$N_n^{(j)} = \#\{v \in [[\rho, \Sigma_j[[\colon v \text{ is a special branchpoint for } [[\rho, \Sigma_j[[\text{ in } (\mathcal{T}; \Sigma_i, i \in [n])\}$$

for the number of special branch points, and $\tau_n^{(j)} = \inf\{t \geq 0\colon \#\mathcal{L}_n(\mathcal{T}_{(\Sigma_j)}^{\alpha\to 0}(t)) < m\}$ for the time when the label set first has fewer than m elements. The significance of this time is that up to this time, all branch points that are special in $(\mathcal{T}; \Sigma_i, i \in \mathbb{N})$ will also be special in $(\mathcal{T}; \Sigma_i, i \in [n])$, but this fails afterwards. We introduce $V_n^{(j)} = \inf\{t \geq 0\colon \Sigma_n \notin \mathcal{T}_{(\Sigma_j)}^{\alpha\to 0}(t)\}$, the time when Σ_n leaves the spine $[[\rho, \Sigma_j[[$.

Proposition 25. *Let* $(\nu_j, j \geq 1)$ *and* ν *be as in Theorem 6 and* $(\mathcal{T}; \Sigma_i, i \in \mathbb{N})$ *a sampling according to Procedure 2. Suppose furthermore that there is* $m \geq 1$ *with* $\nu_j = \nu_m$, $j \geq m$, *and that* $\nu_m(s_1 \leq 1 - \epsilon) = \epsilon^{-\alpha}\ell(1/\epsilon)$, *which is equivalent to* (3) *for* ν *as in* (2). *Then,*

(i) *for all* $j \geq 1$, *we have* $N_n^{(j)}/(n^\alpha \ell(n)) \xrightarrow[n\to\infty]{a.s.} 0$;
(ii) *for every* $p \geq 1$, *we have* $\limsup_{n\to\infty} \mathbb{E}[(N_n^{(n)}/\log n)^p] < \infty$;
(iii) *for every* $p \geq 1$, *there exists a constant* C_p^{spec} *such that for all* $1 \leq j \leq n$ *and* $x > 0$

$$\mathbb{P}(N_n^{(j)} > xn^\alpha \ell(n)) < \frac{C_p^{\text{spec}}}{x^p n^{\alpha p - 1}}.$$

Proof. (i) Let us consider $N_n^{(1)}$ first. We will study the asymptotics by relating to the setting of Lemma 24. Recall that $X_{(\Sigma_1)}^{\alpha\to 0}(V_i^{(1)})$, $i \geq 2$, are the residual masses of the subtrees containing Σ_1 when Σ_i has left the spine $[[\rho, \Sigma_1[[$. Let Y_k, $k \geq 1$, be independent copies of $X_{(\Sigma_1)}^{\alpha\to 0}(\tau_m^{(1)})$, the residual mass of the subtree containing Σ_1 at the branch point separating $[m]$, and $G_k = Y_1 \cdots Y_k$.

Consider the filtration $\mathcal{F}_n^{(1)}(t) = \sigma((S^{\Sigma_1}(s), F_{\Sigma_1}(s), \mathcal{B}_{(\Sigma_1)}^{\alpha\to 0}(s), \mathcal{L}_n(\mathcal{B}_{(\Sigma_1)}^{\alpha\to 0}(s))), s \leq t)$, $t \geq 0$, of the spinal Poisson point process $(S^{\Sigma_1}, F_{\Sigma_1}, \mathcal{B}_{(\Sigma_1)}^{\alpha\to 0})$ studied in Proposition 19 augmented by label sets of spinal bushes derived from sampled leaves $\Sigma_1, \ldots, \Sigma_n$. Let $H_n^{(1)} = \#\{\tau_i^{(1)}, m \leq i \leq n\}$. Then $H_m^{(1)} = 1$ is the initial state, we will also consider

$(\tau_m^{(1)}, X_{(\Sigma_1)}^{\alpha\to 0}(\tau_m^{(1)}), \#\mathcal{L}_m(\mathcal{T}_{(\Sigma_1)}^{\alpha\to 0}(\tau_m^{(1)})))$. Now let $n \geq m+1$ and write $\overline{V}_n^{(j)} = \min\{\tau_n^{(j)}, V_n^{(j)}\}$, $n \geq 1$. Conditionally given $\mathcal{F}_{n-1}^{(1)}(\tau_{n-1}^{(1)})$, in particular $(X_{(\Sigma_1)}(\overline{V}_m^{(1)}), \dots, X_{(\Sigma_1)}(\overline{V}_{n-1}^{(1)}))$, $H_{n-1}^{(1)} = k$ and $\#\mathcal{L}_{n-1}(\mathcal{T}_{(\Sigma_1)}^{\alpha\to 0}(\tau_{n-1}^{(1)})) = \ell$, the argument to establish Procedure 3 can be used to simplify Procedure 2 slightly differently with modified steps $1'$.–$2'$. combining the steps until $\#B' < m$ or $1 \notin B'$; specifically, sample a leaf $\Sigma_n^* \sim \mu$, define $V_{n,*}^{(1)} = \inf\{t \geq 0\colon 1 \notin \mathcal{L}_{n-1}(\mathcal{T}_{(\Sigma_n^*)}^{\alpha\to 0}(t))\}$ and

- if $V_{n,*}^{(1)} \leq \tau_{n-1}^{(1)}$, set $\mathcal{T}_{n-1,B'} = \mathcal{T}_{(\Sigma_1)}^{\alpha\to 0}(V_{n,*}^{(1)})$, note $H_n^{(1)} = k$, $\tau_n^{(1)} = \tau_{n-1}^{(1)}$, $\#\mathcal{L}_n(\mathcal{T}_{(\Sigma_1)}^{\alpha\to 0}(\tau_n^{(1)})) = \ell$;
- if $V_{n,*}^{(1)} > \tau_{n-1}^{(1)}$ and $\ell < m-1$, set $\mathcal{T}_{n-1,B'} = \mathcal{T}_{(\Sigma_1)}^{\alpha\to 0}(\tau_{n-1}^{(1)})$, note $H_n^{(1)} = k$, $\tau_n^{(1)} = \tau_{n-1}^{(1)}$, $\#\mathcal{L}_n(\mathcal{T}_{(\Sigma_1)}^{\alpha\to 0}(\tau_n^{(1)})) = \ell+1$;
- if $V_{n,*}^{(1)} > \tau_{n-1}^{(1)}$ and $\ell = m-1$, then sampling of Σ_n in the rescaled subtree $\mathcal{T}_{(\Sigma_1)}^{\alpha\to 0}(\tau_{n-1}^{(1)})$ is independent of $\mathcal{F}_{n-1}^{(1)}(\tau_{n-1}^{(1)})$ and by the same procedure as Σ_m is sampled in $\mathcal{T}$, therefore

  $$X_{(\Sigma_1)}^{\alpha\to 0}\big(\overline{V}_n^{(1)}\big) \stackrel{d}{=} X_{(\Sigma_1)}^{\alpha\to 0}\big(\overline{V}_{n-1}^{(1)}\big) Y_{k+1} \stackrel{d}{=} G_{k+1},$$

  note $H_n^{(1)} = k+1$, $\tau_n^{(1)} - \tau_{n-1}^{(1)} \stackrel{d}{=} \tau_m^{(1)}$ independent of $\mathcal{F}_{n-1}^{(1)}(\tau_{n-1}^{(1)})$, and $\#\mathcal{L}_n(\mathcal{T}_{(\Sigma_1)}^{\alpha\to 0}(\tau_n^{(1)})) < m$.

Independently of $(G_k, k \geq 1)$, consider $(\Psi_k, k \geq 1) \stackrel{d}{=} (m - \#\mathcal{L}_{W_{k+1}}(\mathcal{T}_{(\Sigma_1)}^{\alpha\to 0}(\tau_{W_{k+1}}^{(1)})), k \geq 1)$, where $W_k = \inf\{n \geq 1\colon H_n^{(1)} = k\}$. As $(G_k, k \geq 1) \stackrel{d}{=} (X_{(\Sigma_1)}^{\alpha\to 0}(\overline{V}_{W_k}^{(1)}), k \geq 1)$, it is now straightforward to show that the dynamics of $H_n^{(1)}$ and J_{n-m+1}^{Ψ} are the same, hence there exists a sequence $(I_i, i \geq 1)$ of independent uniform random variables on $[0, 1]$ and an independent random sequence Ψ, each member taking values in $[m]$ such that for all $n \geq m$

$$\big(H_m^{(1)}, \dots, H_n^{(1)}\big) \stackrel{d}{=} \big(J_1^{\Psi}, \dots, J_{n-m+1}^{\Psi}\big). \tag{13}$$

Now note that $n \in \mathbb{N}$ with $W_k < n < W_{k+1}$ can only yield a new special branch point if $V_{n,*}^{(1)} > \tau_{n-1}^{(1)}$, i.e. in the middle case of the procedure above, but after at most $m-1$ such steps, the third case will apply and $H_n^{(1)}$ will increase. Therefore,

$$N_n^{(1)} \leq m\#H_n^{(1)}. \tag{14}$$

Lemma 24 ensures $H_n^{(1)}/\log n \to 1/\mathbb{E}[-\log Y_1]$, therefore $N_n^{(1)}/(n^{\alpha}\ell(n)) \to 0$ a.s. as $n \to \infty$.

The same argument, with Σ_1 replaced by $\Sigma^* \sim \mu$, yields $N_n^*/(n^{\alpha}\ell(n)) \to 0$ a.s. as $n \to \infty$. For $j \geq 2$, consider times $\chi_i^{(j)} = \inf\{t \geq 0\colon \#\mathcal{L}_j(\mathcal{T}_{(\Sigma_j)}^{\alpha\to 0}(t)) < m-i\}$, $0 \leq i < m$, when j changes rank below m in the label set, s.th. $\chi_0^{(j)} = \tau_j^{(j)}$ and $\chi_{m-1}^{(j)} = \infty$. As $\#\mathcal{L}_n(\mathcal{T}_{(\Sigma_j)}^{\alpha\to 0}(\chi_i^{(j)})) \leq n-j+m$, the number of special branch points between $\Sigma_j^{\alpha\to 0}(\chi_i^{(j)})$ and Σ_{ℓ_i}, where $\ell_i = \min \mathcal{L}_j(\mathcal{T}_{(\Sigma_j)}^{\alpha\to 0}(\chi_i^{(j)})$, will be no larger than $\widetilde{N}_{n-j+m}^{(1),i}$ where $(\widetilde{N}_k^{(1),i}, k \geq 1)$ are independent copies of $(N_k^{(1)}, k \geq 1)$. Then

$$N_n^{(j)} \leq \widetilde{N}_n^* + \sum_{i=1}^{m-1} \widetilde{N}_{n-j+m}^{(1),i}, \tag{15}$$

where $\widetilde{N}_n^*$ is the number of special branchpoints on $[[\rho, \Sigma_j^*[[$, so that $\widetilde{N}^* \stackrel{d}{=} N^*$. Hence the convergence for $N_n^{(j)}/(n^{\alpha}\ell(n))$ follows from previous cases of $N^{(1)}$ and N^*.

(ii) To study $N_n^{(n)}$, we will identify new families $(G_k, k \geq 1)$ and ψ different from the ones in (i) and again apply Lemma 24. Let $b_1^{(j)} = \Sigma_j^{\alpha\to 0}(\chi_1^{(j)})$ be the first special branch point in the spine $[[\rho, \Sigma_j]]$. By Procedure 3, $X_{(\Sigma_j)}^{\alpha\to 0}(\chi_1^{(j)}) = X_{(\Sigma_j^*)}^{\alpha\to 0}(\chi_1^{(j)})$ for all $j \geq m+1$. Also, note that $\chi_1^{(j)}$ is determined by $(\mathcal{T}; \Sigma_i, i \in [m]; \Sigma_j^*)$. As Σ_j^* is sampled according to μ in $\mathcal{T}$, we have

$$X_{(\Sigma_j)}^{\alpha\to 0}\big(\chi_1^{(j)}\big) = X_{(\Sigma_j^*)}^{\alpha\to 0}\big(\chi_1^{(j)}\big) \stackrel{d}{=} X_{(\Sigma_{m+1}^*)}^{\alpha\to 0}\big(\chi_1^{(m+1)}\big) = X_{(\Sigma_{m+1})}^{\alpha\to 0}\big(\chi_1^{(m+1)}\big). \tag{16}$$

Let Y_k, $k \geq 1$, be independent copies of $X^{\alpha\to 0}_{\Sigma_{m+1}}(\chi_1^{(m+1)})$ and consider a constrained painbox associated with $G_k = Y_1 \cdots Y_k$, $k \geq 1$, also $\psi_k = 1$, $k \geq 1$. We claim that for all $n \geq m+1$, and every $x > 0$,

$$\mathbb{P}\big(N_n^{(n)} - m + 1 > x\big) \leq \mathbb{P}\big(J^{\psi}_{n-m} > x\big). \tag{17}$$

This formula holds for $n = m+1$ as $N^{(m+1)}_{m+1} - m + 1 \leq J^{\psi}_1 = 1$. Suppose (17) holds for all $n \leq j-1$. For $n = j$, the first special branch point $b_1^{(j)}$ on the spine $[[\rho, \Sigma_j]]$ is located on the spine $[[\rho, b_1^{(1)}]]$. For $i = m+1, \dots, j-1$, let $\mathcal{T}^{\alpha\to 0}_{(\Sigma_i)}(V_i^{(1)} \wedge \chi_1^{(1)})$ be the spinal subtree of $\mathcal{T}$ containing Σ_i rooted on a branch point on the spine $[[\rho, b_1^{(1)}]]$, possibly at $b_1^{(1)}$ itself. By Procedure 3, $\Sigma_i^* \in \mathcal{T}^{\alpha\to 0}_{(\Sigma_i)}(V_i^{(1)} \wedge \chi_1^{(1)})$. We can express the number $M_1^{(j)}$ of leaves in $\{\Sigma_{m+1}, \dots, \Sigma_{j-1}\}$ belonging to the subtree containing Σ_j above branch point $b_1^{(j)}$ as

$$M_1^{(j)} = \#\big\{i \in \{m+1, \dots, j-1\}\colon\ \Sigma_i \in \mathcal{T}^{\alpha\to 0}_{(\Sigma_j)}\big(\chi_1^{(j)}\big)\big\} = \#\big\{i \in \{m+1, \dots, j-1\}\colon\ \Sigma_i^* \in \mathcal{T}^{\alpha\to 0}_{(\Sigma_j)}\big(\chi_1^{(j)}\big)\big\}.$$

As $\Sigma^*_{m+1}, \dots, \Sigma^*_{j-1}$ are sampled according to μ and $X^{\alpha\to 0}_{(\Sigma_j)}(\chi_1^{(j)}) \overset{d}{=} X^{\alpha\to 0}_{(\Sigma_{m+1})}(\chi_1^{(m+1)}) \overset{d}{=} Y_1$, by (16),

$$\begin{aligned}
\mathbb{P}\big(M_1^{(j)} = k\big) &= \mathbb{E}\left[\binom{j-m-1}{k}\big(X^{\alpha\to 0}_{(\Sigma_j)}\big(\chi_1^{(j)}\big)\big)^k\big(1 - X^{\alpha\to 0}_{(\Sigma_j)}\big(\chi_1^{(j)}\big)\big)^{j-m-k-1}\right] \\
&= \mathbb{E}\left[\binom{j-m-1}{k} Y_1^k (1-Y_1)^{j-m-k-1}\right] = \mathbb{P}\big(\overline{M}^{\psi}_{j-m} = k\big)
\end{aligned}$$

for all $0 \leq k \leq j-m-1$, where $\overline{M}^{\psi}_{j-m}$ is the number of $\overline{I}^{\psi}_1, \dots, \overline{I}^{\psi}_{j-m}$ hitting the interval $(0, G_1)$.

Let $N_j^{(j)}(\chi_1^{(j)}, \infty) = N_j^{(j)} - 1$ be the number of special branch points in $]]b_1^{(j)}, \Sigma_j]]$, and $J^{\psi}_{j-m}(0, Y_1) = J^{\psi}_{j-m} - 1$. Given $M_1^{(j)} = k$, we have $\#\mathcal{L}_j(\mathcal{T}^{\alpha\to 0}_{(\Sigma_j)}(\chi_1^{(j)})) \leq k+m \leq j-1$. Hence, applying the induction hypothesis to the rescaled $(\mathcal{T}^{\alpha\to 0}_{(\Sigma_j)}(\chi_1^{(j)}); \Sigma_i, i \in \mathcal{L}_j(\mathcal{T}^{\alpha\to 0}_{(\Sigma_j)}(\chi_1^{(j)})))$

$$\begin{aligned}
&\mathbb{P}\big(N_j^{(j)}\big(\chi_1^{(j)}, \infty\big) - m + 1 > x \,|\, M_1^{(j)} = k\big) \\
&\quad \leq \mathbb{P}\big(N^{(k+m)}_{k+m} - m + 1 > x\big) \leq \mathbb{P}\big(J^{\psi}_k > x\big) = \mathbb{P}\big(J^{\psi}_{j-m}(0, Y_1) > x \,|\, \overline{M}^{\psi}_{j-m} = k\big),
\end{aligned}$$

and then

$$\begin{aligned}
\mathbb{P}\big(N_j^{(j)} - m + 1 > x\big) &= \mathbb{E}\big[\mathbb{P}\big(N_j^{(j)}\big(\chi_1^{(j)}, \infty\big) - m + 1 > x - 1 \,|\, M_1^{(j)}\big)\big] \\
&\leq \mathbb{E}\big[\mathbb{P}\big(J^{\psi}_{j-m}(0, Y_1) > x - 1 \,|\, \overline{M}^{\psi}_{j-m}\big)\big] = \mathbb{P}\big(J^{\psi}_{j-m} > x\big).
\end{aligned}$$

Now (17) is clear and we deduce that $\mathbb{E}[(N_n^{(n)} - m + 1)^p] \leq \mathbb{E}[(J^{\psi}_{n-m})^p]$ for every $p \geq 1$. The result in (ii) now follows from Lemma 24.

(iii) Formula (15) implies that for every $p \geq 1$ and $x > 0$ and $z_n = x n^{\alpha} \ell(n)$

$$\begin{aligned}
\mathbb{P}\big(N_n^{(j)} > z_n\big) &\leq \mathbb{P}\big(\widetilde{N}_n^* > z_n/m\big) + \sum_{i=1}^{m-1} \mathbb{P}\big(\widetilde{N}^{(1),i}_{n-j+m} > z_n/m\big) \\
&\leq \frac{\mathbb{E}[(N_n^*)^p]}{z_n^p/m^p} + (m-1)\frac{\mathbb{E}[(N^{(1)}_{n-j+m})^p]}{z_n^p/m^p} \leq \frac{C_p(\log n)^p}{z_n^p}.
\end{aligned}$$

The last line is obtained by Markov's inequality. Formula (14) together with Lemma 24 gives the upper bounds. The result in (iii) follows. □

Procedure 3 and the notion of special branch points are also useful to show that the sampling uses the whole CRT $(\mathcal{T}, \mu)$ and does not leave any subtrees of positive mass unlabelled. One way of making this precise is to say that the reduced trees converge to the CRT:

Proposition 26. *In the setting of Procedure 3, we have*

$$R(\mathcal{T}; \Sigma_i, i \in [k]) \to \mathcal{T} \quad \text{a.s. in the Gromov–Hausdorff sense as } k \to \infty.$$

Proof. Let $\epsilon > 0$. Consider $[[\rho, \Sigma_1[[$ and the associated spinal mass partition [19]. Here we denote by $\nu_\epsilon^{\mathrm{sp}}$ the distribution on $S^\downarrow$ of the masses of spinal subtrees that are greater than ϵ. Let $\sigma_\epsilon^{(1)} = \inf\{t \geq 0\colon\ \mu(\mathcal{T}_{(\Sigma_1)}^{\alpha\to 0}(t)) < \epsilon\}$. Note that $W_1 := \inf\{n \geq 1\colon\ \tau_n^{(1)} \geq \sigma_\epsilon^{(1)}\} < \infty$ a.s., by the previous proof. By Procedure 3, leaves Σ_n^* and Σ_n are in the same subtree of $[[\rho, \Sigma_1^{\alpha\to 0}(\sigma_\epsilon)]]$ for each $n > W_1$, in particular each subtree of mass greater than ϵ is selected with an asymptotic frequency greater than ϵ. Inductively, we use Corollary 20 and leaves selected according to Procedure 3 to further split according to scaled $\nu_\epsilon^{\mathrm{sp}}$ each subtree of mass greater than ϵ.

After a finite number of steps, all subtrees have mass less than ϵ, e.g. because a homogeneous mass fragmentation process $(F_t, t \geq 0)$ in $S^\downarrow$ with finite dislocation measure $\nu_\epsilon^{\mathrm{sp}}$ satisfies $F_t \to 0$ as $t \to \infty$, see e.g. Eq. (4) of [7], and so only has finitely many splits before $|F_1(t)| < \epsilon$. □

Using arguments of [28], Corollary 23, we can also show joint a.s. convergence in the Gromov–Prohorov sense of weighted trees $(R(\mathcal{T}; \Sigma_i, i \in [n]), n^{-1} \sum_{i=1}^n \delta_{\Sigma_i}) \to (\mathcal{T}, \mu)$.

5.3. *Convergence of reduced trees and large deviation estimates for spines*

By Corollary 22, reduced trees $R(\mathcal{T}; \Sigma_i, i \in [k])$ of self-similar CRTs with labelled leaves sampled according to Procedure 3, can be assigned subtree masses on edges (parts of spines) in terms of Poisson point processes and associated spinal subordinators, and away from existing leaves, sampling of new leaves is according to subtree masses. To study the asymptotics of the number of spinal branchpoints, we will need the following refinement of results in [14,18].

Lemma 27. *Let $\xi = (\xi_t, t \geq 0)$ be a pure jump subordinator with Lévy measure Λ satisfying $\Lambda([x, \infty)) = x^{-\alpha}\ell_\Lambda(1/x)$, $x \downarrow 0$, for some $\alpha \in (0, 1)$. Let (ϵ, τ, τ') be any random variables on $[0, \infty)^2 \times [0, \infty]$ with $\tau \leq \tau'$. Let $(V_i, i \geq 1)$ be any random variables conditionally independent given (ξ, ϵ, τ) with*

$$\mathbb{P}(V_i \leq \tau | \xi, \epsilon, \tau) = 1 - e^{-\epsilon} \quad \text{and} \quad \mathbb{P}(V_i > \tau + v | \xi, \epsilon, \tau) = e^{-\epsilon - \xi_v}, \quad v \geq 0,$$

and $K_n(\epsilon, \tau, \tau') = \#\{V_i\colon\ 1 \leq i \leq n, \tau < V_i \leq \tau'\}$. Then

$$\lim_{n\to\infty} \frac{K_n(\epsilon, \tau, \tau')}{n^\alpha \ell_\Lambda(n)\Gamma(1-\alpha)} = \int_0^{\tau'-\tau} \exp\bigl(-\alpha(\epsilon + \xi_v)\bigr)\,\mathrm{d}v \quad \text{a.s. as } n \to \infty.$$

If furthermore $\Lambda([xy, \infty)) \leq C_\Lambda y^{-\varrho}\Lambda([x, \infty))$ for all $y \geq 1$ and $0 < x \leq 1$, and some $\varrho > 0$, then there is a constant C_p for all $p > 1/\alpha$, such that for all $x \geq 1$, $n \geq 1$ and all (ϵ, τ, τ') as above, but with the additional property that $\tau' = \tau + \tau''$ for a stopping time τ'' for a filtration in which ξ is a subordinator,

$$\mathbb{P}\biggl(\frac{K_n(\epsilon, \tau, \tau')}{n^\alpha \ell_\Lambda(n)\Gamma(1-\alpha)} > (1+x)Y\bigl(\epsilon, \tau, \tau'\bigr)\biggr) \leq \frac{C_p}{x^p n^{\alpha p - 1}}, \tag{18}$$

where $Y(\epsilon, \tau, \tau') = 1 + (1 + A_\alpha)C_\Lambda \sum_{j=0}^{[\tau'-\tau]} \exp(-\varrho(\epsilon + \xi_j))$ with $A_\alpha = 2\sum_{j\geq 1} \frac{(j+1)^{\sqrt{\alpha}}}{j(j+1)}$.

This lemma is an extension of Lemmas 8 and 12 of [18], which we recover as the special case $\tau = \epsilon = 0$ and/or $\tau' = \infty$. The proof is also essentially the same, but since this result is more general, we reproduce the proof rewritten in the present generality in the Appendix.

Proposition 28. *Let $\nu_1, \dots, \nu_m$ be conservative with $\nu(s_1 \le 1-\epsilon) = \epsilon^{-\alpha}\ell(1/\epsilon)$, where ν is as in Theorem 6 with $\nu_j = \nu_m$, $j \ge m$. Let $R(\mathcal{T}, \Sigma_1, \dots, \Sigma_n)$ be an $\mathbb{R}$-tree sampled from an α-self-similar CRT $(\mathcal{T}, \mu)$ with dislocation measure ν by Procedure 3. Let $(T_n)_{n\ge 1}$ be the associated labelled discrete RE Markov branching trees with unit edge lengths. Then*

$$\frac{R(T_n, [k])}{n^\alpha \ell(n) \Gamma(1-\alpha)} \xrightarrow[n\to\infty]{a.s.} R(\mathcal{T}, \Sigma_1, \dots, \Sigma_k) \quad \textit{in the sense that all edge lengths converge.}$$

In particular, the delabelled trees $(R(T_n, [k]))^\circ$, $n \ge k$, converge in the Gromov–Hausdorff sense.

Proof. Consider $k = 1$ and denote by $D_n^{(1)}$ the length of $R(T_n, \{1\})$.

If $\nu_1 = \dots = \nu_{m-1} = 0$, then $\Sigma_1, \dots, \Sigma_m$ are always in the same subtree in $\mathcal{T}$, then $\tau_1^{(1)} = \dots = \tau_m^{(1)} = \infty$. Conditionally on the subordinator ξ^{Σ_1} associated with leaf Σ_1, cf. Proposition 19(ii), the leaves $\Sigma_{m+1}, \dots, \Sigma_n$ are sampled according to μ along the spine $[[\rho, \Sigma_1[[$. Using Proposition 19(ii), we see that the hypotheses of the first part of Lemma 27 are satisfied, and the convergence result then follows. Specifically, it is easy to see that by (3), as $x \downarrow 0$,

$$\Lambda_1\bigl([x,\infty)\bigr) \sim \int_{\{1/2 < s_1 \le \mathrm{e}^{-x}\}} P_0(\mathbf{s}, s_1)\nu(\mathrm{d}\mathbf{s}) \sim \nu\bigl(s_1 \le \mathrm{e}^{-x}\bigr) \sim x^\alpha \ell(1/x),$$

since by (8), $\int_{\{1/2<s_1\le \mathrm{e}^{-x}\}}(1-P_0(\mathbf{s}, s_1))\nu(\mathrm{d}\mathbf{s}) \le 2\int_{\mathcal{S}^\downarrow} \sum_{i\ge 2}(1-s_i)P_0(\mathbf{s}, s_i)\nu(\mathrm{d}\mathbf{s}) < \infty$.

Now suppose that at least one of $\nu_1, \dots, \nu_{m-1}$ is non-zero. By Procedure 3, each Σ_i is either placed in the same subtree of $[[\rho, \Sigma_1[[$ as $\Sigma_i^* \sim \mu$ or contributes a special branch point. Now

$$D_n^{(1)} = \#\bigl\{V_i^{(1)}, 1 \le i \le n\bigr\} \le 1 + N_n^{(1)} + \#\bigl\{V_{i,*}^{(1)}, 2 \le i \le n\bigr\}, \tag{19}$$

with $V_{i,*}^{(1)} = \inf\{t \ge 0\colon 1 \notin \mathcal{L}_{n-1}(\mathcal{T}_{(\Sigma_n^*)}^{\alpha\to 0}(t))\}$, where Lemma 27 yields the asymptotics of $K_{n-1}^{(1)}(0,0,\infty) = \#\{V_{i,*}^{(1)}, 2 \le i \le n\}$. Together with the asymptotics of $N_n^{(1)}$ obtained in Proposition 25, this yields

$$\limsup_{n\to\infty} \frac{D_n^{(1)}}{n^\alpha \ell(n)\Gamma(1-\alpha)} \le \int_0^\infty \exp\bigl(-\alpha \xi_t^{\Sigma_1}\bigr)\,\mathrm{d}t \quad \text{a.s.} \tag{20}$$

On the other hand, no special branch points are created for $n \ge l+1 \ge m+2$ below $\tau_l^{(1)}$, so

$$D_n^{(1)} \ge \#\bigl\{V_i^{(1)}\colon 0 < V_i^{(1)} \le \tau_l^{(1)}, l+1 \le i \le n\bigr\} = \#\bigl\{V_{i,*}^{(1)}\colon 0 < V_{i,*}^{(1)} \le \tau_l^{(1)}, l+1 \le i \le n\bigr\}.$$

At least one of $\nu_j \ne 0$, $j \le m-1$, so $\tau_m^{(1)} < \infty$. By the proof of Proposition 25, $\tau_l^{(1)} \to \infty$, so

$$\liminf_{n\to\infty} \frac{D_n^{(1)}}{n^\alpha \ell(n)\Gamma(1-\alpha)} \ge \sup_{l\ge m+1} \liminf_{n\to\infty} \frac{\#\{V_{i,*}^{(1)}\colon V_{i,*}^{(1)} \le \tau_l^{(1)}, l+1 \le i \le n\}}{n^\alpha \ell(n)\Gamma(1-\alpha)} = \int_0^\infty \exp\bigl(-\alpha\xi_t^{\Sigma_1}\bigr)\,\mathrm{d}t.$$

Combining this with (20), the convergence for $D_n^{(1)}$ follows and establishes the result for $k = 1$.

Next, consider $k \ge 2$ assuming the result for $1, \dots, k-1$. For the branch point v_k adjacent to ρ in $R(\mathcal{T}, \Sigma_1, \dots, \Sigma_k)$, set $D^{[k]} = \mathrm{d}(\rho, v_k)$, with time ζ_{v_k} given by

$$D^{[k]} = \int_0^{\zeta_{v_k}} \exp\bigl(-\alpha\xi_t^{\Sigma_1}\bigr)\,\mathrm{d}t.$$

Let $D_n^{[k]}$ be the height of the branch point adjacent to the root in $R(T_n, [k])$, then $D_n^{[k]} - 1$ is the number of distinct branch points of $R(\mathcal{T}, \Sigma_1, \dots, \Sigma_n)$ belonging to $[[\rho, v_k[[$, i.e.

$$D_n^{[k]} = 1 + \#\bigl\{V_i^{(1)}\colon 0 < V_i^{(1)} < \zeta_{v_k}, k+1 \le i \le n\bigr\}.$$

If $2 \le k \le m$, then $1 \le D_m^{[k]} \le m-1$ and, by the same argument as for $k=1$,

$$K_{n-m}^{(k)}(0,0,\zeta_{v_k}) = \#\big\{V_{i,*}^{(1)}\colon\ V_{i,*}^{(1)} < \zeta_{v_k}, m+1 \le i \le n\big\} \le D_n^{[k]} \le m + K_{n-m}^{(k)}(0,0,\zeta_{v_k}). \tag{21}$$

If $k \ge m+1$, then $D_n^{[k]} = 1 + \#\{V_{i,*}^{(1)}\colon\ V_{i,*}^{(1)} < \zeta_{v_k}, k+1 \le i \le n\}$. In all cases, by Lemma 27

$$\frac{D_n^{[k]}}{n^\alpha \ell(n)\Gamma(1-\alpha)} \xrightarrow[n\to\infty]{\text{a.s.}} \int_0^{\zeta_{v_k}} \exp\big(-\alpha \xi_s^{\Sigma_1}\big)\,\mathrm{d}s = D^{[k]}.$$

So the renormalized length of the root edge of $R(T_n,[k])$ converges as required.

Now argue conditionally given that $[k]$ is first separated into $\Pi^{[k]} = (\pi_1,\dots,\pi_r)$. For all $n \ge k+1$ and $1 \le j \le r$, denote by $B_j(n) = \mathcal{L}_n(\mathcal{T}_j^{[k]}) \supset \pi_j$ the jth block of the partition at v_k in $(\mathcal{T};\,\Sigma_i, i \in [n])$, and by $T_{n,j}^{[k]}$ the corresponding subtree of T_n. By Lemma 21, Procedure 3 and the Strong Law of Large Numbers,

$$\frac{\#B_j(n)}{n} \xrightarrow[n\to\infty]{\text{a.s.}} \mu\big(\mathcal{T}_j^{[k]}\big), \quad 1 \le j \le r,$$

and the Induction Hypothesis yields convergence of the remaining edge lengths, for $1 \le j \le r$

$$\begin{aligned}\frac{R(T_{n,j}^{[k]},\pi_j)}{n^\alpha \ell(n)\Gamma(1-\alpha)} \quad &= \quad \frac{(\#B_j(n))^\alpha \ell(\#B_j(n))}{n^\alpha \ell(n)} \frac{R(T_{n,j}^{[k]},\pi_j)}{(\#B_j(n))^\alpha \ell(\#B_j(n))\Gamma(1-\alpha)} \\ &\xrightarrow[n\to\infty]{\text{a.s.}} \big(\mu\big(\mathcal{T}_j^{[k]}\big)\big)^\alpha R\big(\mathcal{T}_j^{[k]};\,\Sigma_i, i \in \pi_j\big),\end{aligned}$$

in the sense that all edge lengths converge, which implies Gromov–Hausdorff convergence. □

While the arguments of the analogous but much more specific Proposition 22 of [28], do not apply here in cases where the densities $f_k = \mathrm{d}\nu_k/\mathrm{d}\nu$ are degenerate, we can now deduce from our Proposition 26 that in the setting of Proposition 28 here, delabelled trees converge a.s. when taking double limits

$$\lim_{k\to\infty}\lim_{n\to\infty} \frac{(R(T_n,[k]))^\circ}{n^\alpha \ell(n)\Gamma(1-\alpha)} = \mathcal{T} \quad \text{in the Gromov–Hausdorff sense a.s.} \tag{22}$$

Theorem 7, instead of restricting to $[k]$, then letting $n\to\infty$ and then $k\to\infty$, considers $n\to\infty$ directly, at the cost of weakening the mode of convergence to convergence in probability. To prepare the proof of Theorem 7, we study the spines $[[\rho,\Sigma_j[[$, $j \ge 1$.

We denote by Λ_1 and Λ^* the Lévy measures of the subordinators ξ^{Σ_1} and ξ^{Σ^*} generated, respectively, by the first sampled leaf Σ_1 and by a leaf Σ^* sampled according to μ. For $k \ge 1$ and $n \ge k$, denote by $D_n^{(k)}$ the length of $R(T_n,\{k\})$.

Lemma 29. *For all $p \ge 0$, there is a constant $C_p' > 0$ such that for all $k \ge 1$, $n \ge k$ and $x \ge 1$*

$$\mathbb{P}\big(D_n^{(k)} > 2(1+x)(2+Z_k)\max\big\{\overline{\Lambda}_1\big(n^{-1}\big), \overline{\Lambda}^*\big(n^{-1}\big)\big\}\big) \le \frac{C_p'}{x^p n^{\alpha p-1}},$$

where $Z_k = m + (1+A_\alpha)\max\{C_{\Lambda_1}, C_{\Lambda^*}\}(m + \sum_{i=0}^\infty (X_{(\Sigma_k)}^{\alpha\to 0}(i))^\varrho)$ has all moments finite.

Proof. For $k=1$, we use (19) to write $D_n^{(1)} \le 2(D_n^{(1)}-1) \le 2N_n^{(1)} + 2K_{n-1}^{(1)}(0,0,\infty)$ and deduce from Proposition 25 and Lemma 27 that for all $p \ge 0$ and all $n \ge 1$, $x \ge 1$,

$$\begin{aligned}&\mathbb{P}\big(D_n^{(1)} > 2(1+x)(2+Z_1)\overline{\Lambda}_1\big(n^{-1}\big)\big) \\ &\quad \le \mathbb{P}\big(N_n^{(1)} > (1+x)2\overline{\Lambda}_1\big(n^{-1}\big)\big) + \mathbb{P}\big(K_{n-1}^{(1)}(0,0,\infty) > (1+x)Z_1\overline{\Lambda}_1\big(n^{-1}\big)\big) \le \frac{C_p^{\text{spec}} + C_p^{(1)}}{x^p n^{\alpha p-1}}.\end{aligned}$$

Next, consider $2 \le k \le m$. Recall that we denote by $\mathcal{L}_k(\mathcal{S}) = \{i \in [k]\colon\ \Sigma_i \in \mathcal{S}\}$ the set of labels in a subtree $\mathcal{S} \subseteq \mathcal{T}$. We set $\gamma_k^{(k)} = 0$ and split the spine $[[\rho, \Sigma_k[[$ at times $\gamma_j^{(k)} = \inf\{t \ge 0\colon\ \#\mathcal{L}_k(\mathcal{T}_{(\Sigma_k)}^{\alpha\to 0}(t)) \le j\}$ for $k-1 \ge j \ge 1$, some of which may coincide. Repeated application of Corollary 20, Lemma 21 and arguing as for (19) yields that

$$D_n^{(k)} \le 2\big(D_n^{(k)} - 1\big) \le 2N_n^{(k)} + 2K_{n-k}^{(k,1)}\big(\xi^{\Sigma_k}\big(\gamma_1^{(k)}\big), \gamma_1^{(k)}, \infty\big) + \sum_{j=2}^{k} 2K_{n-k}^{(k,j)}\big(\xi^{\Sigma_k}\big(\gamma_j^{(k)}\big), \gamma_j^{(k)}, \gamma_{j-1}^{(k)}\big),$$

where $K_{n-k}^{(k,j)}(\xi^{\Sigma_k}(\gamma_j^{(k)}), \gamma_j^{(k)}, \gamma_{j-1}^{(k)})$ is as in Lemma 27, but here associated with the subordinator $\xi^{(k,j)} = \xi^{\Sigma_{k,j}}(\gamma_j^{(k)} + \cdot) - \xi^{\Sigma_{k,j}}(\gamma_j^{(k)})$ that has Lévy measure $\Lambda^{(k,j)} = \Lambda_1$ and with random variables $V_i^{(k,j)} = \inf\{t \ge 0\colon\ \Sigma_{k+i}^* \notin \mathcal{T}_{(\Sigma_{k,j})}^{\alpha\to 0}(t)\}$, $i \ge 1$, where $\Sigma_{k,j} = \Sigma_\ell$ if $\ell = \min \mathcal{L}_k(\mathcal{T}_{(\Sigma_k)}^{\alpha\to 0}(\gamma_j^{(k)}))$.

Let $Z^{(k)}(\gamma_j^{(k)}, \gamma_{j-1}^{(k)}) = 1 + (1 + A_\alpha) C_{\Lambda_1} \sum_{i=[\gamma_j^{(k)}]}^{[\gamma_{j-1}^{(k)}]} \exp(-\varrho \xi_i^{\Sigma_k})$, noting $\sum_{j=1}^{k} Z^{(k)}(\gamma_j^{(k)}, \gamma_{j-1}^{(k)}) < Z_k$. Then

$$\begin{aligned}
&\mathbb{P}\big(D_n^{(k)} > 2(1+x)(2+Z_k)\overline{\Lambda}_1\big(n^{-1}\big)\big) \\
&\quad \le \mathbb{P}\big(N_n^{(k)} > 2(1+x)\overline{\Lambda}_1\big(n^{-1}\big)\big) + \sum_{j=1}^{k} \mathbb{P}\big(K_{n-k}^{(k,j)}\big(\xi^{\Sigma_k}\big(\gamma_j^{(k)}\big), \gamma_j^{(k)}, \gamma_{j-1}^{(k)}\big) > (1+x) Z^{(k)}\big(\gamma_j^{(k)}, \gamma_{j-1}^{(k)}\big)\Lambda_1\big(n^{-1}\big)\big)
\end{aligned}$$

and we can conclude again by Proposition 25 and Lemma 27 with constant $C_p^{\mathrm{spec}} + kC_p^{(1)}$.

Now consider $k \ge m+1$. We set $\gamma_{m+1}^{(k)} = 0$ and $\gamma_0^{(k)} = \infty$. We split $[[\rho, \Sigma_k[[$ at times $\gamma_m^{(k)} = \inf\{t \ge 0\colon\ \Sigma_k^* \notin \mathcal{T}_{(\Sigma_k)}^{\alpha\to 0}(t)\}$ and $\gamma_j^{(k)} = \inf\{t \ge \gamma_m^{(k)}\colon\ \#\mathcal{L}_k(\mathcal{T}_{(\Sigma_k)}^{\alpha\to 0}(t)) \le j\}$ for $m-1 \ge j \ge 1$. Note that, by Procedure 3, $\#\mathcal{L}_k(\mathcal{T}_{(\Sigma_k)}^{\alpha\to 0}(\gamma_m^{(k)})) \le m$. Again

$$D_n^{(k)} \le 2\big(D_n^{(k)} - 1\big) \le 2N_n^{(k)} + \sum_{j=1}^{k} 2K_{n-m}^{(k,j)}\big(\varepsilon_j^{(k)}, \gamma_j^{(k)}, \gamma_{j-1}^{(k)}\big) + 2K_{n-m}^{(k,*)}\big(0, 0, \gamma_m^{(k)}\big),$$

where $\varepsilon_j^{(k)} = \xi^{\Sigma_k}(\gamma_j^{(k)})$, other notation as for $k \le m$, and $K_{n-m}^{(k,*)}(0,0,\gamma_m^{(k)})$ is as in Lemma 27, here based on the subordinator $\xi^{\Sigma_k^*}$ with Lévy measure Λ^*, and $V_i^{(k,*)} = \inf\{t \ge 0\colon\ \Sigma_i^* \notin \mathcal{T}_{(\Sigma_k^*)}^{\alpha\to 0}(t)\}$, the time when Σ_k^* and Σ_i^* are first in different subtrees. We get

$$\begin{aligned}
&\mathbb{P}\big(D_n^{(k)} > 2(1+x)(2+Z_k)\max\big\{\overline{\Lambda}_1\big(n^{-1}\big), \overline{\Lambda}^*\big(n^{-1}\big)\big\}\big) \\
&\quad \le \mathbb{P}\big(N_n^{(k)} > 2(1+x)\overline{\Lambda}_1\big(n^{-1}\big)\big) \\
&\qquad + \sum_{j=1}^{k} \mathbb{P}\big(K_{n-m}^{(k,j)}\big(\xi^{\Sigma_k}\big(\gamma_j^{(k)}\big), \gamma_j^{(k)}, \gamma_{j-1}^{(k)}\big) > (1+x) Z^{(k)}\big(\gamma_j^{(k)}, \gamma_{j-1}^{(k)}\big)\Lambda_1\big(n^{-1}\big)\big) \\
&\qquad + \mathbb{P}\big(K_{n-m}^{(k,*)}\big(0, 0, \gamma_m^{(k)}\big) > (1+x) Z^{(k)}\big(0, \gamma_m^{(k)}\big)\overline{\Lambda}^*\big(n^{-1}\big)\big)
\end{aligned}$$

and conclude again by Proposition 25 and Lemma 27 with constant $C_p' = C_p^{\mathrm{spec}} + mC_p^{(1)} + C_p^*$.

Let $H_{\mathcal{T}}^{\varrho}$ be the height of the ϱ-self-similar CRT $(\mathcal{T}^\varrho, \mu^\varrho)$ obtained from $(\mathcal{T}, \mu)$ by ϱ-self-similar time-change. By Proposition 14 of [16], the height $H_{\mathcal{T}}^{\varrho}$ has exponential moments and so does Z_k:

$$\begin{aligned}
\sup_{k\ge 1} Z_k &\le m + (1 + A_\alpha)\max\{C_{\Lambda_1}, C_{\Lambda^*}\}\left(m + \sup_{k \ge 1}\int_0^\infty \big(X_{(\Sigma_k)}^{\alpha\to 0}(t)\big)^\varrho \,\mathrm{d}t\right) \\
&\le m + (1 + A_\alpha)\max\{C_{\Lambda_1}, C_{\Lambda^*}\}\big(m + H_{\mathcal{T}}^{\varrho}\big).
\end{aligned}$$

□

5.4. *Proof of Theorem 7*

The previous sections contain the new developments that we need to apply the techniques developed in [18] for the exchangeable case in the higher generality of Theorem 7. We only briefly retrace this argument here so as to identify the places where a result in the previous sections here replaces a more specific result of [18].

Lemma 30 (Lemma 10 and Corollary 11 of [18]). *Let $H_n = \max_{1\le k\le n} D_n^{(k)}$ be the height of T_n. Then there is a constant $C_{p,a}$ for all $a>0$, $p\ge 2/\alpha$, such that for all $x\ge 1$ and $n\ge 1$*

$$\mathbb{P}\left(\frac{H_n}{n^\alpha \ell(n)} > ax\right) \le \frac{C_{p,a}}{x^p}.$$

The proof is based on Lemma 29 here replacing Lemma 12 of [18], and $\overline{\Lambda}_1(n^{-1}) \sim \overline{\Lambda}^*(n^{-1}) \sim n^\alpha \ell(n)$.

Lemma 31 (Proposition 9 of [18]). *Under the hypotheses of Theorem 7, let for $n\ge k$*

$$\Delta(n,k) := \max_{1\le i\le n} \mathrm{d}_n\big(\{i\}, R\big(T_n,[k]\big)\big),$$

d_n *being the metric associated with T_n. Then for each $\eta>0$,*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\frac{\Delta(n,k)}{n^\alpha \ell(n)} > \eta\right) = 0.$$

The proof is based on the following three replacements. First, Proposition 26 or (22) here replace "Clearly, $\lambda^k_{\max} := \max_{j\ge 1}\lambda^k_j \to 0$ a.s." on page 1819 of [18]. Second, Corollary 22 here replaces the reference made in [18] to Lemma 3.14 of their reference [10]. Third, Lemma 30 here replaces Corollary 11 of [18].

Proof of Theorem 7. This proof is now based on the following three replacements. First, (22) here replaces the reference made in [18] to their reference [29]. Second, Lemma 31 here replaces Proposition 9 of [18]. Third, Proposition 28 here replaces Proposition 7 of [18]. □

6. Skewed PD model; proofs of Propositions 8 and 10

Recall that Proposition 8 asserts that the alpha-gamma model for $\alpha\in[0,1)$ and $\gamma\in[0,\alpha]$ is a RE Markov branching model with dislocation measures of the form identified in Corollary 5 with $\nu_1 = (1-\alpha)\mathrm{PD}^*_{\alpha,-\alpha-\gamma}$ and $\nu_j = \gamma\mathrm{PD}^*_{\alpha,-\alpha-\gamma}$, $j\ge 2$.

Proof of Proposition 8. We focus on the multifurcating case $\alpha\in(0,1)$ and $\gamma\in[0,\alpha)$, the binary case being easier. We claim that the distribution of the partition Π_n of $T_n \sim Q_n^{\alpha,\gamma}$ at $[n]$ is given by

$$\mathbb{P}(\Pi_n = \pi) = \begin{cases} p_n^1(n_1,\dots,n_k) = (1-\alpha)\frac{\Gamma(2-\alpha)}{\Gamma(n+1-\alpha)}\frac{\alpha^{k-2}\Gamma(k-1-\gamma/\alpha)}{\Gamma(1-\gamma/\alpha)}\prod_{i=1}^k \frac{\Gamma(n_i-\alpha)}{\Gamma(1-\alpha)}, & \pi\in\mathcal{K}^{\mathbf{0}}{}_{[2]}\cap\mathcal{P}_n,\\ p_n^2(n_1,\dots,n_k) = \gamma\frac{\Gamma(2-\alpha)}{\Gamma(n+1-\alpha)}\frac{\alpha^{k-2}\Gamma(k-1-\gamma/\alpha)}{\Gamma(1-\gamma/\alpha)}\prod_{i=1}^k \frac{\Gamma(n_i-\alpha)}{\Gamma(1-\alpha)}, & \pi\in\mathcal{K}^{\mathbf{1}}{}_{[2]}\cap\mathcal{P}_n,\end{cases}$$

and that $(Q_n^{\alpha,\gamma}, n\ge 2)$ has the labelled Markov branching property

$$\mathbb{P}\big(\Pi_n=\pi, S_1^n=\mathbf{s}_1,\dots,S_k^n=\mathbf{s}_k\big) = p_n^j(\#\pi_1,\dots,\#\pi_k)\prod_{i=1}^k Q_{\pi_i}^{\alpha,\gamma}\big(\{\mathbf{s}_i\}\big), \quad \pi\in\mathcal{P}_n^j,$$

where S_i^n is the ith subtree of T_n above the first branchpoint, and $Q_{\pi_i}^{\alpha,\gamma}$ is the push-forward of $Q_{\#\pi_i}^{\alpha,\gamma}$ under the natural bijection on the set of hierarchies induced by the increasing bijection from $[\#\pi_i]$ to π_i.

We show this by induction on n. Specifically, for $n=2$, this is trivial, for $n=3$ we have e.g.

$$\mathbb{P}\big(\Pi_3=\big\{\{1,3\},\{2\}\big\}\big)=\mathbb{P}\big(\Pi_3=\big\{\{1\},\{2,3\}\big\}\big)=\frac{1-\alpha}{2-\alpha},\qquad \mathbb{P}\big(\Pi_3=\big\{\{1,2\},\{3\}\big\}\big)=\frac{\gamma}{2-\alpha}.$$

If the claim holds for n, we can apply the growth rules and the induction hypothesis to see

$$\mathbb{P}\big(\Pi_{n+1}=\big\{[n],\{n+1\}\big\},S_1^{n+1}=\mathbf{s}_1,S_2^{n+1}=\big\{\{n+1\}\big\}\big)=\frac{\gamma}{n-\alpha}Q_n\big(\{\mathbf{s}_1\}\big)Q_{\{n+1\}}\big(\{n+1\}\big),$$

and for $\pi=(\pi_1,\dots,\pi_k)\in\mathcal{P}_n^j$, $j=1,2$, and hierarchies $\mathbf{s}_i$ of π_i, $i\neq i'$, and $\mathbf{s}_{i'}$ of $\pi_{i'}\cup\{n+1\}$,

$$\begin{aligned}&\mathbb{P}\big(\Pi_{n+1}=\big(\pi_1,\dots,\pi_k,\{n+1\}\big),S_1^{n+1}=\mathbf{s}_1,\dots,S_k^{n+1}=\mathbf{s}_k,S_{k+1}^{n+1}=\big\{\{n+1\}\big\}\big)\\&\quad=\frac{(k-1)\alpha-\gamma}{n-\alpha}p_n^j(\#\pi_1,\dots,\#\pi_k)Q_{\{n+1\}}\big(\big\{\{n+1\}\big\}\big)\prod_{i=1}^k Q_{\pi_i}\big(\{\mathbf{s}_i\}\big),\\&\mathbb{P}\big(\Pi_{n+1}=\big(\pi_1,\dots,\pi_{i'}\cup\{n+1\},\dots,\pi_k\big),S_1^{n+1}=\mathbf{s}_1,\dots,S_k^{n+1}=\mathbf{s}_k\big)\\&\quad=\frac{n_{i'}-\alpha}{n-\alpha}p_n^j(\#\pi_1,\dots,\#\pi_k)Q_{\pi_{i'}\cup\{n+1\}}\big(\{\mathbf{s}_{i'}\}\big)\prod_{i\neq i'}Q_{\pi_i}\big(\{\mathbf{s}_i\}\big),\end{aligned}$$

as conditionally given that the insertion of $n+1$ is in subtree $S_{i'}^n$, it is just as an insertion of $\#\pi_{i'}+1$ into $T_{\#\pi_{i'}}$, pushed forward from $[\#\pi_{i'}+1]$ to $\pi_{i'}\cup\{n+1\}$. The result follows. □

Recall that Proposition 10 asserts that the skewed Poisson–Dirichlet model is sampling consistent only for parameters that reduce it to the exchangeable Poisson–Dirichlet model or to the alpha-gamma model.

Proof of Proposition 10. By Corollary 5, the skewed Poisson–Dirichlet model has dislocation measure

$$\kappa=\int_{S^\downarrow}\big(\lambda\kappa_{\mathbf{s}}\big(\cdot\cap\mathcal{P}^{\mathbf{0}_{[2]}}\big)+(1-\lambda)\kappa_{\mathbf{s}}\big(\cdot\cap\mathcal{P}^{\mathbf{1}_{[2]}}\big)\big)\mathrm{PD}^*_{\alpha,\theta}(d\mathbf{s}).$$

From this, we can calculate splitting rules. Specifically, we can calculate the distribution of the ranked sequence $S_n=(\#\Pi_{n,1},\dots,\#\Pi_{n,K_n})^\downarrow$ of block sizes of $\Pi_n=(\Pi_{n,1},\dots,\Pi_{n,K_n})$ by summing (6) over partitions of equal ranked sequence of block sizes and obtain

$$\begin{aligned}&\mathbb{P}\big(S_2=(1,1)\big)=1,\qquad \mathbb{P}\big(S_3=(1,1,1)\big)=\frac{\lambda(2\alpha+\theta)}{D_3},\qquad \mathbb{P}\big(S_3=(2,1)\big)=\frac{(1+\lambda)(1-\alpha)}{D_3}\\&\mathbb{P}\big(S_4=(1,1,1,1)\big)=\frac{\lambda(3\alpha+\theta)(2\alpha+\theta)}{D_4},\qquad \mathbb{P}\big(S_4=(2,1,1)\big)=\frac{(1+4\lambda)(2\alpha+\theta)(1-\alpha)}{D_4}\\&\mathbb{P}\big(S_4=(2,2)\big)=\frac{(1+\lambda)(1-\alpha)^2}{D_4},\qquad \mathbb{P}\big(S_4=(3,1)\big)=\frac{2(1-\alpha)(2-\alpha)}{D_4},\end{aligned}$$

where D_3 and D_4 are normalisation constants of the form $a_3\lambda+b_3$ and $a_4\lambda+b_4$. Using the criterion of [18], sampling consistency requires, in particular, that

$$\mathbb{P}\big(S_3=(1,1,1)\big)=\mathbb{P}\big(S_4=(1,1,1,1)\big)+\frac{1}{2}\mathbb{P}\big(S_4=(2,1,1)\big)+\frac{1}{4}\mathbb{P}\big(S_4=(3,1)\big)\mathbb{P}\big(S_3=(1,1,1)\big),$$

which upon multiplication by D_3D_4 is a quadratic equation in λ. Common coefficients of all terms include $(1-\alpha)$ and $(\theta+2\alpha)$. For $\alpha<1$ and $\theta>-2\alpha$, the quadratic equation has the two solutions $\lambda=1/2$ and $\lambda=(1-\alpha)/(1-\theta-2\alpha)$ corresponding, respectively, to the Poisson–Dirichlet and alpha-gamma models, so no other models can be sampling consistent.

The exchangeable Poisson–Dirichlet model is trivially sampling consistent. The alpha-gamma model was shown in [10] to be sampling consistent. In the excluded case $\alpha = 1$ models for all θ collapse to the same deterministic model where all leaves are connected directly to a single branch point [24]. For the binary case $\theta = -2\alpha$, which we also had to exclude for our argument here, we need to consider S_5. This gives similar quadratic equations, but also leads to the required conclusion that only the alpha model $\lambda = 1-\alpha$ and the beta-splitting model $\lambda = 1/2$ are sampling consistent. We leave details to the reader. □

Appendix: Proof of Lemma 27

The first part of Lemma 27 is a straightforward consequence of [14], see also Lemma 8 of [18]. The second part generalises Lemma 12 of [18]. In the following, we indicate the most relevant changes that needed for our higher generality.

Let $N_y(t_1,t_2)$ denote the number of jumps of ξ of size at least y in the time interval $[t_1,t_2]$, $\widetilde{N}_y^{\epsilon,\tau}(t_1,t_2)$ denote the number of jumps of $\exp(-\epsilon)(1-\exp(-\xi))$ of size at least y in the same time interval.

Step 1. *Large deviations for* $\widetilde{N}_y^{\epsilon,\tau}(0,\tau'')$

Lemma 32. *For all* $x > 0$ *and* $0 < y \le 1$,

$$\mathbb{P}\left(\widetilde{N}_y^{\epsilon,\tau}\big(0,\tau''\big) > (1+x)C_\Lambda \sum_{i=0}^{[\tau'']} \exp\big(-\varrho(\epsilon+\xi_i)\big)\overline{\Lambda}(y)\right) \le \exp\big(-a_x\overline{\Lambda}(y)\big),$$

where $a_x := (1+x)\ln(1+x) - x > 0$.

Proof. We adapt the proof of Lemma 36 of [18]. Let $\mathcal{F}_t^{\epsilon,\tau}$ denote the σ-field generated by $(\epsilon,\tau,\tau''\wedge t)$ and ξ until time t, and $\mathcal{F}_\infty^{\epsilon,\tau}$ the one generated by (ϵ,τ,τ'') and ξ, and observe that

$$\widetilde{N}_y^{\epsilon,\tau}\big(0,\tau''\big) \le \sum_{i=0}^{[\tau'']} \widetilde{N}_y^{\epsilon,\tau}(i,i+1) \le \sum_{i=0}^{[\tau'']} N_{y\exp(\epsilon+\xi_i)}(i,i+1).$$

Conditional on $\mathcal{F}_i^{\epsilon,\tau}$, $N_{y\exp(\epsilon+\xi_i)}(i,i+1)$ is a Poisson random variable with mean $\overline{\Lambda}(y\exp(\epsilon+\xi_i))$. The remainder of the proof of Lemma 36 of [18], now applies to give

$$\mathbb{P}\left(\sum_{i=0}^{[\tau'']\wedge n} N_{y\exp(\epsilon+\xi_i)}(i,i+1) \ge (1+x)C_\Lambda \sum_{i=0}^{[\tau'']\wedge n} \exp\big(-\varrho(\epsilon+\xi_i)\big)\overline{\Lambda}(y)\right) \le \exp\big(-a_x\overline{\Lambda}(y)\big),$$

and we can let $n \to \infty$ and apply Fatou's lemma to complete the proof. □

Step 2. *Large deviations for* $\mathbb{E}[K_n(\epsilon,\tau,\tau')|\mathcal{F}_{\tau''}^{\epsilon,\tau}]$

Lemma 33. *Let* $B_\alpha := \sum_{k\ge 1}\exp(-4^{-1}a_1k^{\alpha/2})$ *with* $a_1 = 2\ln 2 - 1$. *Then for all* $x \ge 1$ *and all integers* n *large enough*,

$$\mathbb{P}\big(\mathbb{E}\big[K_n\big(\epsilon,\tau,\tau'\big)|\mathcal{F}_{\tau''}^{\epsilon,\tau}\big] > (1+x)\big(Y\big(\epsilon,\tau,\tau'\big)-1\big)\overline{\Lambda}\big(n^{-1}\big)\big) \le (1+B_\alpha)\exp\big(-4^{-1}a_1x\overline{\Lambda}\big(n^{-1}\big)\big).$$

Proof. We adapt the proof of Lemma 14 of [18]. According to formula (4) of [14],

$$\mathbb{E}\big[K_n\big(\epsilon,\tau,\tau'\big)|\mathcal{F}_{\tau''}^{\epsilon,\tau}\big] = n\int_0^1 (1-y)^{n-1}\widetilde{N}_y^{\epsilon,\tau}\big(0,\tau''\big)\,\mathrm{d}y \le \widetilde{N}_{1/n}^{\epsilon,\tau}\big(0,\tau''\big) + n\int_0^{1/n} \widetilde{N}_y^{\epsilon,\tau}\big(0,\tau''\big)\,\mathrm{d}y.$$

Hence, setting $S := C_\Lambda \sum_{i=0}^{[\tau'']} \exp(-\varrho(\epsilon + \xi_i))$,

$$\begin{aligned}
&\mathbb{P}\big(\mathbb{E}\big[K_n\big(\epsilon,\tau,\tau'\big)|\mathcal{F}^{\epsilon,\tau}_{\tau''}\big] > (1+x)(1+A_\alpha)S\overline{\Lambda}\big(n^{-1}\big)\big)\\
&\quad\le \mathbb{P}\big(\widetilde{N}^{\epsilon,\tau}_{1/n}\big(0,\tau''\big) > (1+x)S\overline{\Lambda}\big(n^{-1}\big)\big) + \mathbb{P}\bigg(n\int_0^{1/n}\widetilde{N}^{\epsilon,\tau}_y\big(0,\tau''\big)\,\mathrm{d}y > (1+x)A_\alpha S\overline{\Lambda}\big(n^{-1}\big)\bigg).
\end{aligned}$$

The first probability in the RHS is smaller than $\exp(-a_x\overline{\Lambda}(n^{-1}))$ by Lemma 32. To bound the second probability, we use $n\int_{1/(k+1)n}^{1/kn}\widetilde{N}^{\epsilon,\tau}_y(0,\tau'')\,\mathrm{d}y \le \widetilde{N}^{\epsilon,\tau}_{1/(n(k+1))}(0,\tau'')\frac{1}{k(k+1)}$, which gives

$$\begin{aligned}
&\mathbb{P}\bigg(n\int_0^{1/n}\widetilde{N}^{\epsilon,\tau}_y\big(0,\tau''\big)\,\mathrm{d}y > A_\alpha(1+x)S\overline{\Lambda}\big(n^{-1}\big)\bigg)\\
&\quad\le \sum_{k\ge 1}\mathbb{P}\big(\widetilde{N}^{\epsilon,\tau}_{1/(n(k+1))}\big(0,\tau''\big) > 2(k+1)^{\sqrt{\alpha}}(1+x)S\overline{\Lambda}\big(n^{-1}\big)\big),
\end{aligned}$$

and we proceed as in Lemma 14 of [18], to see that this is bounded by $\exp(-4^{-1}a_1x\overline{\Lambda}(n^{-1}))B_\alpha$ for all $x\ge 1$ and n large enough. □

Step 3. *Proof of inequality* (18)

We adapt the proof of formula (28) of [18]. To start with, fix $x \ge 1$, $n \in \mathbb{N}$, and note that

$$\begin{aligned}
&\mathbb{P}\big(K_n\big(\epsilon,\tau,\tau'\big) > (1+x)Y\big(\epsilon,\tau,\tau'\big)\overline{\Lambda}\big(n^{-1}\big)\big)\\
&\quad\le \mathbb{P}\big(\mathbb{E}\big[K_n\big(\epsilon,\tau,\tau'\big)|\mathcal{F}^{\epsilon,\tau}_{\tau''}\big] > (1+x)\big(Y\big(\epsilon,\tau,\tau'\big)-1\big)\overline{\Lambda}\big(n^{-1}\big)\big)\\
&\qquad + \mathbb{P}\big(K_n\big(\epsilon,\tau,\tau'\big) - \mathbb{E}\big[K_n\big(\epsilon,\tau,\tau'\big)|\mathcal{F}^{\epsilon,\tau}_{\tau''}\big] > (1+x)\overline{\Lambda}\big(n^{-1}\big)\big).
\end{aligned}\tag{23}$$

Lemma 33 gives an upper bound for the first probability provided n is large enough. To get an upper bound for the second probability, we proceed as for formula (28) of [18], to find that for all $m \ge 1$, there exists some deterministic constant B_m depending only on m such that

$$\begin{aligned}
&\mathbb{P}\big(K_n\big(\epsilon,\tau,\tau'\big) - \mathbb{E}\big[K_n\big(\epsilon,\tau,\tau'\big)|\mathcal{F}^{\epsilon,\tau}_{\tau''}\big] > (1+x)\overline{\Lambda}\big(n^{-1}\big)|\mathcal{F}^{\epsilon,\tau}_{\tau''}\big)\\
&\quad\le 2^{m-1}B_m\frac{\mathbb{E}[(K_n(\epsilon,\tau,\tau'))^m|\mathcal{F}^{\epsilon,\tau}_{\tau''}] + ((1+x)\overline{\Lambda}(n^{-1}))^m}{((1+x)\overline{\Lambda}(n^{-1}))^{2m}}.
\end{aligned}$$

We then take expectations on both sides of the resulting inequality. Theorem 6.3 of [14] ensures that $\mathbb{E}[(K_n(\epsilon,\tau,\tau'))^m|\epsilon,\tau] \le \mathbb{E}[(K_n(0,0,\infty)^m] \sim (\overline{\Lambda}(n^{-1}))^m$, up to a constant. Hence, we have

$$\mathbb{P}\big(K_n\big(\epsilon,\tau,\tau'\big) - \mathbb{E}\big[K_n\big(\epsilon,\tau,\tau'\big)|\mathcal{F}^{\epsilon,\tau}_{\tau''}\big] > (1+x)\overline{\Lambda}\big(n^{-1}\big)\big) \le B_{m,\Lambda}\big((1+x)\overline{\Lambda}\big(n^{-1}\big)\big)^{-m},\tag{24}$$

where $B_{m,\Lambda}$ depends only on m and Λ. The proof of (18) now follows the proof of formula (28) of [18].

This completes the proof of Lemma 27. □

Acknowledgments

We would like to thank Sasha Gnedin for drawing Kerov's work to our attention and for further stimulating discussions. Thanks are also due to Gerold Alsmeyer for directing us to Allan Gut's work [15]. The first author would like to acknowledge support from the K. C. Wong Education Foundation. We would like to thank an anonymous referee for carefully reading the paper and for making suggestions that led to an improvement of the presentation.

References

[1] D. Aldous. Exchangeability and related topics. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1983)* 1–198. *Lecture Notes in Math.* **1117**. Springer, Berlin, 1985. MR0883646

[2] D. Aldous. The continuum random tree. I. *Ann. Probab.* **19**(1) (1991) 1–28. MR1085326

[3] D. Aldous. The continuum random tree. III. *Ann. Probab.* **21**(1) (1993) 248–289. MR1207226

[4] D. Aldous. Probability distributions on cladograms. In *Random Discrete Structures (Minneapolis, MN, 1993)* 1–18. *IMA Vol. Math. Appl.* **76**. Springer, New York, 1996. MR1395604

[5] J. Bertoin. *Lévy Processes*. *Cambridge Tracts in Mathematics* **121**. Cambridge Univ. Press, Cambridge, 1996. MR1406564

[6] J. Bertoin. Homogeneous fragmentation processes. *Probab. Theory Related Fields* **121**(3) (2001) 301–318. MR1867425

[7] J. Bertoin. The asymptotic behavior of fragmentation processes. *J. Euro. Math. Soc.* **5** (2003) 395–416. MR2017852

[8] J. Bertoin. *Random Fragmentation and Coagulation Processes*. *Cambridge Studies in Advanced Mathematics* **102**. Cambridge Univ. Press, Cambridge, 2006. MR2253162

[9] J. Bertoin and A. Rouault. Discretization methods for homogeneous fragmentations. *J. London Math. Soc. (2)* **72**(1) (2005) 91–109. MR2145730

[10] B. Chen, D. Ford and M. Winkel. A new family of Markov branching trees: The alpha-gamma model. *Electron. J. Probab.* **14**(15) (2009) 400–430 (electronic). MR2480547

[11] R. Durrett. *Probability: Theory and Examples*, 2nd edition. Duxbury Press, Belmont, CA, 1996. MR1609153

[12] D. J. Ford. Probabilities on cladograms: Introduction to the alpha model. Preprint, 2005. Available at arXiv:math/0511246v1.

[13] A. Gnedin. Constrained exchangeable partitions. In *Fourth Colloquium on Mathematics and Computer Science, Vol. AG* 391–398. Discrete Mathematics and Theoretical Computer Science, Nancy, 2006. MR2509650

[14] A. Gnedin, J. Pitman and M. Yor. Asymptotic laws for compositions derived from transformed subordinators. *Ann. Probab.* **34**(2) (2006) 468–492. MR2223948

[15] A. Gut. On the moments and limit distributions of some first passage times. *Ann. Probab.* **2** (1974) 277–308. MR0394857

[16] B. Haas. Loss of mass in deterministic and random fragmentations. *Stochastic Process. Appl.* **106**(2) (2003) 245–277. MR1989629

[17] B. Haas and G. Miermont. The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electron. J. Probab.* **9**(4) (2004) 57–97 (electronic). MR2041829

[18] B. Haas, G. Miermont, J. Pitman and M. Winkel. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *Ann. Probab.* **36**(5) (2008) 1790–1837. MR2440924

[19] B. Haas, J. Pitman and M. Winkel. Spinal partitions and invariance under re-rooting of continuum random trees. *Ann. Probab.* **37**(4) (2009) 1381–1411. MR2546748

[20] C. Haulk and J. Pitman. A representation of exchangeable hierarchies by sampling from real trees. Preprint, 2011. Available at arXiv:1101.5619v1.

[21] S. V. Kerov. Combinatorial examples in the theory of AF-algebras. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **172**(Differentsialnaya Geom. Gruppy Li i Mekh. Vol. 10) (1989) 55–67, 169–170. MR1015698

[22] J. F. C. Kingman. The representation of partition structures. *J. London Math. Soc. (2)* **18**(2) (1978) 374–380. MR0509954

[23] J. F. C. Kingman. *Poisson Processes*. *Oxford Studies in Probability* **3**. Oxford Univ. Press, New York, 1993. MR1207584

[24] P. McCullagh, J. Pitman and M. Winkel. Gibbs fragmentation trees. *Bernoulli* **14**(4) (2008) 988–1002. MR2543583

[25] G. Miermont. Self-similar fragmentations derived from the stable tree. I. Splitting at heights. *Probab. Theory Related Fields* **127**(3) (2003) 423–454. MR2018924

[26] J. Pitman. Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102**(2) (1995) 145–158. MR1337249

[27] J. Pitman. *Combinatorial Stochastic Processes*. *Lecture Notes in Mathematics* **1875**. Springer, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. MR2245368

[28] J. Pitman and M. Winkel. Regenerative tree growth: Binary self-similar continuum random trees and Poisson–Dirichlet compositions. *Ann. Probab.* **37**(5) (2009) 1999–2041. MR2561439

[29] E. Schroeder. Vier combinatorische Probleme. *Z. f. Math. Phys.* **15** (1870) 361–376.

[30] R. P. Stanley. *Enumerative Combinatorics, Vol. 2*. *Cambridge Studies in Advanced Mathematics* **62**. Cambridge Univ. Press, Cambridge, 1999. With a foreword by Gian-Carlo Rota and Appendix 1 by Sergey Fomin. MR1676282

[31] A. M. Vershik and S. V. Kerov. Asymptotic theory of the characters of a symmetric group. *Funktsional. Anal. i Prilozhen.* **15**(4) (1981) 15–27, 96. MR0639197