A new family of Markov branching trees: the alpha-gamma model

Bo Chen*  Daniel Ford†  Matthias Winkel‡

July 3, 2008

Abstract

We introduce a simple tree growth process that gives rise to a new two-parameter family of discrete fragmentation trees that extends Ford’s alpha model to multifurcating trees and includes the trees obtained by uniform sampling from Duquesne and Le Gall’s stable continuum random tree. We call these new trees the alpha-gamma trees. In this paper, we obtain their splitting rules, dislocation measures both in ranked order and in sized-biased order, and we study their limiting behaviour.

AMS 2000 subject classifications: 60J80.

Keywords: Alpha-gamma tree, splitting rule, sampling consistency, self-similar fragmentation, dislocation measure, continuum random tree, $\mathbb{R}$-tree, Markov branching model

1 Introduction

Markov branching trees were introduced by Aldous [3] as a class of random binary phylogenetic models and extended to the multifurcating case in [16]. Consider the space $T_n$ of combinatorial trees without degree-2 vertices, one degree-1 vertex called the root and exactly $n$ further degree-1 vertices labelled by $[n] = \{1, \ldots, n\}$ and called the leaves; we call the other vertices branch points. Distributions on $T_n$ of random trees $T_n^*$ are determined by distributions of the delabelled tree $T_n^0$ on the space $T_n^0$ of unlabelled trees and conditional label distributions, e.g. exchangeable labels. A sequence $(T_n^0, n \geq 1)$ of unlabelled trees has the Markov branching property if for all $n \geq 2$ conditionally given that the branching adjacent to the root is into tree components whose numbers of leaves are $n_1, \ldots, n_k$, these tree components are independent copies of $T_n^0$, $1 \leq i \leq k$. The distributions of the sizes in the first branching of $T_n^0$, $n \geq 2$, are denoted by

$$q(n_1, \ldots, n_k), \quad n_1 \geq \ldots \geq n_k \geq 1, \quad k \geq 2 : \quad n_1 + \ldots + n_k = n,$$

and referred to as the splitting rule of $(T_n^0, n \geq 1)$.

Aldous [3] studied in particular a one-parameter family ($\beta \geq -2$) that interpolates between several models known in various biology and computer science contexts (e.g. $\beta = -2$ comb, $\beta = -3/2$ uniform, $\beta = 0$ Yule) and that he called the beta-splitting model, he sets for $\beta > -2$:

$$q_{\beta}^{\text{Aldous}}(n-m, m) = \frac{1}{Z_n} \binom{n}{m} B(m+1+\beta, n-m+1+\beta), \quad \text{for } 1 \leq m < n/2,$$

$$q_{\beta}^{\text{Aldous}}(n/2, n/2) = \frac{1}{2Z_n} \binom{n}{n/2} B(n/2+1+\beta, n/2+1+\beta), \quad \text{if } n \text{ even},$$

where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the Beta function and $Z_n$, $n \geq 2$, are normalisation constants; this extends to $\beta = -2$ by continuity, i.e. $q_{-2}^{\text{Aldous}}(n-1,1) = 1, n \geq 2$. 

*University of Oxford; email chen@stats.ox.ac.uk
†Google Inc.; email dford@math.stanford.edu
‡University of Oxford; email winkel@stats.ox.ac.uk
For exchangeably labelled Markov branching models \((T_n, n \geq 1)\) it is convenient to set
\[
p(n_1, \ldots, n_k) := \frac{m_1! \cdots m_r!}{(n_1, \ldots, n_k)} g((n_1, \ldots, n_k)^\dagger), \quad n_j \geq 1, j \in [k]; k \geq 2 : n = n_1 + \ldots + n_k, \tag{1}
\]
where \((n_1, \ldots, n_k)^\dagger\) is the decreasing rearrangement and \(m_r\) the number of \(r\)s of the sequence \((n_1, \ldots, n_k)\). The function \(p\) is called exchangeable partition probability function (EPPF) and gives the probability that the branching adjacent to the root splits into tree components with label sets \(\{A_1, \ldots, A_k\}\) partitioning \([n]\), with block sizes \(n_j = \#A_j\). Note that \(p\) is invariant under permutations of its arguments. It was shown in [19] that Aldous’s beta-splitting models for \(\beta > -2\) are the only binary Markov branching models for which the EPPF is of Gibbs type
\[
p_{-1-\alpha}(n_1, n_2) = \frac{w_{n_1} w_{n_2}}{Z_{n_1+n_2}}, \quad n_1 \geq 1, n_2 \geq 1, \quad \text{in particular } w_n = \frac{\Gamma(n - \alpha)}{\Gamma(1 - \alpha)},
\]
and that the multifurcating Gibbs models are an extended Ewens-Pitman two-parameter family of random partitions, \(0 \leq \alpha \leq 1, \theta \geq -2\alpha, \text{ or } -\infty \leq \alpha < 0, \theta = -m\alpha \text{ for some integer } m \geq 2,
\[
p^{PD*}_{\alpha, \theta}(n_1, \ldots, n_k) = \frac{a_k}{Z_n} \prod_{j=1}^{k} w_{n_j}, \quad \text{where } w_n = \frac{\Gamma(n - \alpha)}{\Gamma(1 - \alpha)} \text{ and } a_k = \alpha^{k-2} \frac{\Gamma(k + \theta/\alpha)}{\Gamma(2 + \theta/\alpha)}, \tag{2}
\]
boundary cases by continuity. Ford [12] introduced a different binary model, the alpha model, using simple sequential growth rules starting from the unique elements \(T_1 \in T_1 \text{ and } T_2 \in T_2:\)

(i)\(^F\) given \(T_n\) for \(n \geq 2\), assign a weight \(1 - \alpha\) to each of the \(n\) edges adjacent to a leaf, and a weight \(\alpha\) to each of the \(n - 1\) other edges;

(ii)\(^F\) select at random with probabilities proportional to the weights assigned by step (i)\(^F\), an edge of \(T_n\), say \(a_n \rightarrow c_n\) directed away from the root;

(iii)\(^F\) to create \(T_{n+1}\) from \(T_n\), replace \(a_n \rightarrow c_n\) by three edges \(a_n \rightarrow b_n, b_n \rightarrow c_n\) and \(b_n \rightarrow n + 1\) so that two new edges connect the two vertices \(a_n\) and \(c_n\) to a new branch point \(b_n\) and a further edge connects \(b_n\) to a new leaf labelled \(n + 1\).

It was shown in [12] that these trees are Markov branching trees but that the labelling is not exchangeable. The splitting rule was calculated and shown to coincide with Aldous’s beta-splitting rules if and only if \(\alpha = 0, \alpha = 1/2\) or \(\alpha = 1\), interpolating differently between Aldous’s corresponding models for \(\beta = 0, \beta = -3/2\) and \(\beta = -2\). This study was taken further in [16, 23].

In this paper, we introduce a new model by extending the simple sequential growth rules to allow multifurcation. Specifically, we also assign weights to vertices as follows, cf. Figure [11]

(i) given \(T_n\) for \(n \geq 2\), assign a weight \(1 - \alpha\) to each of the \(n\) edges adjacent to a leaf, a weight \(\gamma\) to each of the \(n - 1\) other edges, and a weight \((k - 1)\alpha - \gamma\) to each vertex of degree \(k + 1 \geq 3\);

(ii) select at random with probabilities proportional to the weights assigned by step (i),
- an edge of \(T_n\), say \(a_n \rightarrow c_n\) directed away from the root,
- or, as the case may be, a vertex of \(T_n\), say \(v_n\);

(iii) to create \(T_{n+1}\) from \(T_n\), do the following:
- if an edge \(a_n \rightarrow c_n\) was selected, replace it by three edges \(a_n \rightarrow b_n, b_n \rightarrow c_n\) and \(b_n \rightarrow n + 1\) so that two new edges connect the two vertices \(a_n\) and \(c_n\) to a new branch point \(b_n\) and a further edge connects \(b_n\) to a new leaf labelled \(n + 1\);
- if a vertex \(v_n\) was selected, add an edge \(v_n \rightarrow n + 1\) to a new leaf labelled \(n + 1\).
We call the resulting model the \textit{alpha-gamma model}. These growth rules satisfy the rules of probability for all $0 \leq \alpha \leq 1$ and $0 \leq \gamma \leq \alpha$. They contain the growth rules of the alpha model for $\gamma = \alpha$. They also contain growth rules for a model [18] [20] based on the stable tree of Duquesne and Le Gall [7], for the cases $\gamma = 1 - \alpha$, $1/2 \leq \alpha < 1$, where all edges are given the same weight; we show here that these cases $\gamma = 1 - \alpha$, $1/2 \leq \alpha \leq 1$, as well as $\alpha = \gamma = 0$ form the intersection with the extended Ewens-Pitman-type two-parameter family of models [2].

\textbf{Proposition 1.} Let $(T_n, n \geq 1)$ be alpha-gamma trees with distributions as implied by the sequential growth rules (i)-(iii) for some $0 \leq \alpha \leq 1$ and $0 \leq \gamma \leq \alpha$. Then

(a) the delabelled trees $T_n^0$, $n \geq 1$, have the Markov branching property. The splitting rules are

\begin{equation}
q_{\alpha, \gamma}^\text{seq}(n_1, \ldots, n_k) \propto \left( \gamma + (1 - \alpha - \gamma) \frac{1}{n(n-1)} \sum_{i \neq j} n_i n_j \right) q_{\alpha, \gamma}^\text{PD}((n_1, \ldots, n_k), \ldots, n_k),
\end{equation}

in the case $0 \leq \alpha < 1$, where $q_{\alpha, \gamma}^\text{PD}$ is the splitting rule associated via [7] with $p_{\alpha, \gamma}^\text{PD}$, the Ewens-Pitman-type EPPF given in [2], and LHS $\propto$ RHS means equality up to a multiplicative constant depending on $n$ and $(\alpha, \gamma)$ that makes the LHS a probability function;

(b) the labelling of $T_n$ is exchangeable for all $n \geq 1$ if and only if $\gamma = 1 - \alpha$, $1/2 \leq \alpha \leq 1$.

For any function $(n_1, \ldots, n_k) \mapsto q(n_1, \ldots, n_k)$ that is a probability function for all fixed $n = n_1 + \ldots + n_k$, $n \geq 2$, we can construct a Markov branching model $(T_n^0, n \geq 1)$. A condition called \textit{sampling consistency} [8] is to require that the tree $T_{n-1}^0$ constructed from $T_n^0$ by removal of a uniformly chosen leaf (and the adjacent branch point if its degree is reduced to 2) has the same distribution as $T_{n-1}^0$, for all $n \geq 2$. This is appealing for applications with incomplete observations. It was shown in [10] that all sampling consistent splitting rules admit an integral representation $(c, \nu)$ for an erosion coefficient $c \geq 0$ and a dislocation measure $\nu$ on $S^4 = \{ s = (s_i)_{i \geq 1} : s_1 \geq s_2 \geq \ldots \geq 0, s_1 + s_2 + \ldots \leq 1 \}$ with $\nu(\{1, 0, 0, \ldots\}) = 0$ and $\int_{S^4} (1-s_1) \nu(ds) < \infty$ as in Bertoin’s continuous-time fragmentation theory [4] [5] [6]. In the most relevant case when $c = 0$ and $\nu(\{ (s \in S^4 : s_1 + s_2 + \ldots < 1 \}) = 0$, this representation is

\begin{equation}
p(n_1, \ldots, n_k) = \frac{1}{Z_n} \int_{S^4} \sum_{\text{distinct}} \prod_{j=1}^k s_{ij}^{n_j} \nu(ds), \quad n_j \geq 1, j \in [k]; k \geq 2 : n = n_1 + \ldots + n_k,
\end{equation}
where \( Z_n = \int_{S^1} \left(1 - \sum_{i \geq 1} s_i^\alpha \right) \nu(ds) \), \( n \geq 2 \), are the normalization constants. The measure \( \nu \) is unique up to a multiplicative constant. In particular, it can be shown that for the Ewens-Pitman EPPFs \( p^{PD^*}_{\alpha,\theta} \) we obtain \( \nu = PD^*_{\alpha,\theta}(ds) \) of Poisson-Dirichlet type (hence our superscript \( PD^* \) for the Ewens-Pitman type EPPF), where for \( 0 < \alpha < 1 \) and \( \theta > -2\alpha \) we can express

\[
\int_{S^1} f(s)PD^*_{\alpha,\theta}(ds) = \mathbb{E} \left( \sigma_1^\theta f(\Delta \sigma_{[0,1]}/\sigma_1) \right),
\]

for an \( \alpha \)-stable subordinator \( \sigma \) with Laplace exponent \( -\log(\mathbb{E}(e^{-\lambda \sigma_1})) = \lambda^\alpha \) and with ranked sequence of jumps \( \Delta \sigma_{[0,1]} = (\Delta \sigma_i, t \in [0,1]) \). For \( \alpha < 1 \) and \( \theta = -2\alpha \), we have

\[
\int_{S^1} f(s)PD^*_{\alpha,\theta}(ds) = \int_{1/2}^1 f(x, 1-x, 0, 0, \ldots) x^{-\alpha-1}(1-x)^{-\alpha-1} dx.
\]

Note that \( \nu = PD^*_{\alpha,\theta} \) is infinite but \( \sigma \)-finite with \( \int_{S^1} (1 - s_1) \nu(ds) < \infty \) for \( -2\alpha \leq \theta \leq -\alpha \). This is the relevant range for this paper. For \( \theta > -\alpha \), the measure \( PD^*_{\alpha,\theta} \) just defined is a multiple of the usual Poisson-Dirichlet probability measure \( PD_{\alpha,\theta} \) on \( S^1 \), so for the integral representation of \( p^{PD^*}_{\alpha,\theta} \) we could also take \( \nu = PD_{\alpha,\theta} \) in this case, and this is also an appropriate choice for the two cases \( \alpha = 0 \) and \( m \geq 3 \); the case \( \alpha = 1 \) is degenerate \( q^{PD^*}_{0,\theta}(1,1,\ldots,1) = 1 \) (for all \( \theta \)) and can be associated with \( \nu = PD^*_{1,\theta} = \delta_{(0,0,\ldots)} \), see [19].

**Theorem 2.** The alpha-gamma-splitting rules \( q_{\alpha,\gamma}^{seq} \) are sampling consistent. For \( 0 \leq \alpha < 1 \) and \( 0 \leq \gamma \leq \alpha \) the measure \( \nu \) in the integral representation can be chosen as

\[
\nu_{\alpha,\gamma}(ds) = \left( \gamma + (1 - \alpha - \gamma) \sum_{i \neq j} s_i s_j \right) PD^*_{\alpha,\theta}(ds).
\]

The case \( \alpha = 1 \) is discussed in Section 3.2. We refer to Griffiths [14] who used discounting of Poisson-Dirichlet measures by quantities involving \( \sum_{i \neq j} s_i s_j \) to model genetic selection.

In [16], Haas and Miermont’s self-similar continuum random trees (CRTs) [15] are shown to be scaling limits for a wide class of Markov branching models. See Sections 8.3 and 8.6 for details. This theory applies here to yield:

**Corollary 3.** Let \( \left( T_n^0, n \geq 1 \right) \) be delabelled alpha-gamma trees, represented as discrete \( R \)-trees with unit edge lengths, for some \( 0 < \alpha < 1 \) and \( 0 < \gamma \leq \alpha \). Then

\[
\frac{T_n^0}{n^\gamma} \to T_{\alpha,\gamma} \quad \text{in distribution for the Gromov-Hausdorff topology},
\]

where the scaling \( n^\gamma \) is applied to all edge lengths, and \( T_{\alpha,\gamma} \) is a \( \gamma \)-self-similar CRT whose dislocation measure is a multiple of \( \nu_{\alpha,\gamma} \).

We observe that every dislocation measure \( \nu \) on \( S^1 \) gives rise to a measure \( \nu^{sb} \) on the space of summable sequences under which fragment sizes are in a size-biased random order, just as the GEM distribution can be defined as the distribution of a Poisson-Dirichlet sequence re-arranged in size-biased random order [22]. We similarly define \( \text{GEM}_{\alpha,\theta}^* \) from \( PD^*_{\alpha,\theta} \). One of the advantages of size-biased versions is that, as for GEMs, we can calculate marginal distributions explicitly.

**Proposition 4.** For \( 0 < \alpha < 1 \) and \( 0 \leq \gamma < \alpha \), distributions \( \nu_{k}^{sb} \) of the first \( k \geq 1 \) marginals of the size-biased form \( \nu_{\alpha,\gamma}^{sb} \) of \( \nu_{\alpha,\gamma} \) are given, for \( x = (x_1, \ldots, x_k) \), by

\[
\nu_k^{sb}(dx) = \left( \gamma + (1 - \alpha - \gamma) \left( 1 - \sum_{i=1}^{k} x_i^2 - \frac{1-\alpha}{1+(k-1)\alpha-\gamma} \left( 1 - \sum_{i=1}^{k} x_i \right)^2 \right) \right) \text{GEM}_{\alpha,\theta}^*(dx).
\]

The other boundary values of parameters are trivial here – there are at most two non-zero parts.
We can investigate the convergence of Corollary 3 when labels are retained. Since labels are non-exchangeable, in general, it is not clear how to nicely represent a continuum tree with infinitely many labels other than by a consistent sequence $R_k$ of trees with $k$ leaves labelled $[k]$, $k \geq 1$. See however [23] for developments in the binary case $\gamma = \alpha$ on how to embed $R_k$, $k \geq 1$, in a CRT $T^\alpha$. The following theorem extends Proposition 18 of [16] to the multifurcating case.

**Theorem 5.** Let $(T_n, n \geq 1)$ be a sequence of trees resulting from the alpha-gamma-tree growth rules for some $0 < \alpha < 1$ and $0 < \gamma \leq \alpha$. Denote by $R(T_n, [k])$ the subtree of $T_n$ spanned by the root and leaves $[k]$, reduced by removing degree-2 vertices, represented as discrete $\mathbb{R}$-tree with graph distances in $T_n$ except those adjacent to leaves has either a blue or a red colour mark. Then

$$\frac{R(T_n, [k])}{n^\gamma} \to R_k \quad \text{a.s. in the sense that all edge lengths converge},$$

for some discrete tree $R_k$ with shape $T_k$ and edge lengths specified in terms of three random variables, conditionally independent given that $T_k$ has $k + \ell$ edges, as $L_k W_k^\gamma D_k$ with

- $W_k \sim \beta(k(1 - \alpha) + \ell \gamma, (k - 1)\alpha - \ell \gamma)$, where beta$(a, b)$ is the beta distribution with density $B(a, b) x^{a-1}(1-x)^{b-1} 1_{(0,1)}(x)$;
- $L_k$ with density $\frac{\Gamma(1 + k(1 - \alpha) + \ell \gamma)}{\Gamma(1 + \ell + k(1 - \alpha)/\gamma)} \ell^{+k(1-\alpha)/\gamma} g_\gamma(s)$, where $g_\gamma$ is the Mittag-Leffler density, the density of $\sigma_1^{-\gamma}$ for a subordinator $\sigma$ with Laplace exponent $\gamma$;
- $D_k \sim \text{Dirichlet}((1-\alpha)/\gamma, \ldots, (1-\alpha)/\gamma, 1, \ldots, 1)$, where Dirichlet$(a_1, \ldots, a_m)$ is the Dirichlet distribution on $\Delta_m = \{(x_1, \ldots, x_m) \in [0,1]^m : x_1 + \ldots + x_m = 1\}$ with density of the first $m - 1$ marginals proportional to $x_1^{a_1-1} \ldots x_{m-1}^{a_{m-1}-1} (1-x_1 - \ldots - x_{m-1})^{a_{m-1}-1}$; here $D_k$ contains edge length proportions, first with parameter $(1-\alpha)/\gamma$ for edges adjacent to leaves and then with parameter 1 for the other edges, each enumerated e.g. by depth first search.

In fact, $1 - W_k$ captures the total limiting leaf proportions of subtrees that are attached on the vertices of $T_k$, and we can study further how this is distributed between the branch points, see Section 1.2.

We conclude this introduction by giving an alternative description of the alpha-gamma model obtained by adding colouring rules to the alpha model growth rules (i)$^\text{F}$-(iii)$^\text{F}$, so that in $T_n^\text{col}$ each edge except those adjacent to leaves has either a blue or a red colour mark.

(iv)$^\text{col}$ To turn $T_{n+1}$ into a colour-marked tree $T_{n+1}^\text{col}$, keep the colours of $T_n^\text{col}$ and do the following:

- if an edge $a_n \to c_n$ adjacent to a leaf was selected, mark $a_n \to b_n$ blue;
- if a red edge $a_n \to c_n$ was selected, mark both $a_n \to b_n$ and $b_n \to c_n$ red;
- if a blue edge $a_n \to c_n$ was selected, mark $a_n \to b_n$ blue; mark $b_n \to c_n$ red with probability $c$ and blue with probability $1 - c$;

When $(T_n^\text{col}, n \geq 1)$ has been grown according to (i)$^\text{F}$-(iii)$^\text{F}$ and (iv)$^\text{col}$, crush all red edges, i.e. (cr) identify all vertices connected via red edges, remove all red edges and remove the remaining colour marks; denote the resulting sequence of trees by $(\tilde{T}_n, n \geq 1)$.

**Proposition 6.** Let $(\tilde{T}_n, n \geq 1)$ be a sequence of trees according to growth rules (i)$^\text{F}$-(iii)$^\text{F}$,(iv)$^\text{col}$ and crushing rule (cr). Then $(\tilde{T}_n, n \geq 1)$ is a sequence of alpha-gamma trees with $\gamma = \alpha(1 - c)$.

The structure of this paper is as follows. In Section 2 we study the discrete trees grown according to the growth rules (i)-(iii) and establish Proposition 6 and Proposition 1 as well as the sampling consistency claimed in Theorem 2. Section 3 is devoted to the limiting CRTs, we obtain the dislocation measure stated in Theorem 2 and deduce Corollary 3 and Proposition 4. In Section 4 we study the convergence of labelled trees and prove Theorem 5.
2 Sampling consistent splitting rules for the alpha-gamma trees

2.1 Notation and terminology of partitions and discrete fragmentation trees

For $B \subseteq \mathbb{N}$, let $\mathcal{P}_B$ be the set of partitions of $B$ into disjoint non-empty subsets called blocks. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which supports a $\mathcal{P}_B$-valued random partition $\Pi_B$. If the probability function of $\Pi_B$ only depends on its block sizes, we call it exchangeable. Then

$$\mathbb{P}(\Pi_B = \{A_1, \ldots, A_k\}) = p(\#A_1, \ldots, \#A_k)$$

for each partition $\pi = \{A_1, \ldots, A_k\} \in \mathcal{P}_B$, where $\#A_j$ denotes the block size, i.e. the number of elements of $A_j$. This function $p$ is called the exchangeable partition probability function (EPPF) of $\Pi_B$. Alternatively, a random partition $\Pi_B$ is exchangeable if its distribution is invariant under the natural action on partitions of $B$ by the symmetric group of permutations of $B$.

Let $B \subseteq \mathbb{N}$, we say that a partition $\pi \in \mathcal{P}_B$ is finer than $\pi' \in \mathcal{P}_B$, and write $\pi \preceq \pi'$, if any block of $\pi$ is included in some block of $\pi'$. This defines a partial order $\preceq$ on $\mathcal{P}_B$. A process or a sequence with values in $\mathcal{P}_B$ is called refining if it is decreasing for this partial order. Refining partition-valued processes are naturally related to trees. Suppose that $B$ is a finite subset of $\mathbb{N}$ and $t$ is a collection of subsets of $B$ with an additional member called the root such that

- we have $B \in t$; we call $B$ the common ancestor of $t$;
- we have $\{i\} \in t$ for all $i \in B$; we call $\{i\}$ a leaf of $t$;
- for all $A \in t$ and $C \in t$, we have either $A \cap C = \emptyset$, or $A \subseteq C$ or $C \subseteq A$.

If $A \subset C$, then $A$ is called a descendant of $C$, or $C$ an ancestor of $A$. If for all $D \in t$ with $A \subseteq D \subseteq C$ either $A = D$ or $D = C$, we call $A$ a child of $C$, or $C$ the parent of $A$ and denote $C \rightarrow A$. If we equip $t$ with the parent-child relation and also $\text{root} \rightarrow B$, then $t$ is a rooted connected acyclic graph, i.e. a combinatorial tree. We denote the space of such trees $t$ by $T_B$ and also $T_n = T_{[n]}$. For $t \in T_B$ and $A \in t$, the rooted subtree $s_A$ of $t$ with common ancestor $A$ is given by $s_A = \{\text{root}\} \cup \{C \in t : C \subseteq A\} \in T_A$. In particular, we consider the subtrees $s_j = s_{A_j}$ of the common ancestor $B$ of $t$, i.e. the subtrees whose common ancestors $A_j, j \in [k]$, are the children of $B$. In other words, $s_1, \ldots, s_k$ are the rooted connected components of $t \setminus \{B\}$.

Let $(\pi(t), t \geq 0)$ be a $\mathcal{P}_B$-valued refining process for some finite $B \subseteq \mathbb{N}$ with $\pi(0) = 1_B$ and $\pi(t) = \emptyset_B$ for some $t > 0$, where $1_B$ is the trivial partition into a single block $B$ and $\emptyset_B$ is the partition of $B$ into singletons. We define $t_\pi = \{\text{root}\} \cup \{A \in B : A \in \pi(t) \text{ for some } t \geq 0\}$ as the associated labelled fragmentation tree.

Definition 1. Let $B \subseteq \mathbb{N}$ with $\#B = n$ and $t \in T_B$. We associate the relabelled tree

$$t^\sigma = \{\text{root}\} \cup \{\sigma(A) : A \in t\} \in T_n,$$

for any bijection $\sigma : B \rightarrow [n]$, and the combinatorial tree shape of $t$ as the equivalence class

$$t^\circ = \{t^\sigma | \sigma : B \rightarrow [n] \text{ bijection} \} \subset T_n.$$

We denote by $T_n^\circ = \{t^\circ : t \in T_n\} = \{t^\circ : t \in T_B\}$ the collection of all tree shapes with $n$ leaves, which we will also refer to in their own right as unlabelled fragmentation trees.

Note that the number of subtrees of the common ancestor of $t \in T_n$ and the numbers of leaves in these subtrees are invariants of the equivalence class $t^\circ \subset T_n$. If $t^\circ \in T_n^\circ$ has subtrees $s_1^\circ, \ldots, s_k^\circ$ with $n_1 \geq \ldots \geq n_k \geq 1$ leaves, we say that $t^\circ$ is formed by joining together $s_1^\circ, \ldots, s_k^\circ$, denoted by $t^\circ = s_1^\circ \ast \ldots \ast s_k^\circ$. We call the composition $(n_1, \ldots, n_k)$ of $n$ the first split of $t_n^\circ$.

With this notation and terminology, a sequence of random trees $T_n^\circ$ with $n \geq 1$, has the Markov branching property if, for all $n \geq 2$, the tree $T_n^\circ$ has the same distribution as $S_1^\circ \ast \ldots \ast S_{K_n}^\circ$, where $N_1 \geq \ldots \geq N_{K_n} \geq 1$ form a random composition of $n$ with $K_n \geq 2$ parts, and conditionally given $K_n = k$ and $N_j = n_j$, the trees $S_j^\circ, j \in [k]$, are independent and distributed as $T_{n_j}^\circ, j \in [k]$. 

6
2.2 Colour-marked trees and the proof of Proposition 6

The growth rules (i)F-(iii)F construct binary combinatorial trees \( T_n^{\text{bin}} \) with vertex set
\[
V = \{ \text{root} \} \cup [n] \cup \{ b_1, \ldots, b_{n-1} \}
\]
and an edge set \( E \subset V \times V \). We write \( v \to w \) if \((v,w) \in E\). In Section 2.1 we identify leaf \( i \) with the set \( \{i\} \) and vertex \( b_i \) with \( \{ j \in [n] : b_i \to \cdots \to j \} \), the edge set \( E \) then being identified by the parent-child relation. In this framework, a colour mark for an edge \( v \to b_i \) can be assigned to the vertex \( b_i \), so that a coloured binary tree as constructed in (iv)\text{col} can be represented by
\[
V^{\text{col}} = \{ \text{root} \} \cup [n] \cup \{(b_1, \chi(b_1)), \ldots, (b_{n-1}, \chi(b_{n-1}))\}
\]
for some \( \chi(b_i) \in \{0,1\} \), \( i \in [n-1] \), where 0 represents red and 1 represents blue.

**Proof of Proposition 6** We only need to check that the growth rules (i)F-(iii)F and (iv)\text{col} for \( (T_n^{\text{col}}, n \geq 1) \) imply that the uncoloured multifurcating trees \( (\tilde{T}_n, n \geq 1) \) obtained from \( (T_n^{\text{col}}, n \geq 1) \) via crushing (cr) satisfy the growth rules (i)-(iii). Let therefore \( t_{n+1}^{\text{col}} \) be a tree with \( \Pr(T_{n+1}^{\text{col}} = T_{n+1}^{\text{col}}) > 0 \). It is easily seen that there is a unique tree \( \tilde{t}_n^{\text{col}} \), a unique insertion edge \( a_n^{\text{col}} \to c_n^{\text{col}} \) in \( t_n^{\text{col}} \) and, if any, a unique colour \( \chi_n(\tilde{t}_n^{\text{col}}) \) to create \( t_{n+1}^{\text{col}} \) from \( \tilde{t}_n^{\text{col}} \). Denote the trees obtained from \( t_n^{\text{col}} \) and \( t_{n+1}^{\text{col}} \) via crushing (cr) by \( t_n \) and \( t_{n+1} \). If \( \chi_{n+1}(c_n^{\text{col}}) = 0 \), denote by \( k + 1 \geq 3 \) the degree of the branch point of \( t_n \), with which \( c_n^{\text{col}} \) is identified in the first step of the crushing (cr).

- If the insertion edge is a leaf edge \( (c_n^{\text{col}} = i \text{ for some } i \in [n]) \), we obtain
  \[
  \Pr(\tilde{T}_{n+1} = t_{n+1} | \tilde{T}_n = t_n, T_n^{\text{col}} = t_n^{\text{col}}) = (1 - \alpha)/(n - \alpha).
  \]

- If the insertion edge has colour blue \( (\chi_n(c_n^{\text{col}}) = 1) \) and also \( \chi_{n+1}(c_n^{\text{col}}) = 1 \), we obtain
  \[
  \Pr(\tilde{T}_{n+1} = t_{n+1} | \tilde{T}_n = t_n, T_n^{\text{col}} = t_n^{\text{col}}) = \alpha(1 - c)/(n - \alpha).
  \]

- If the insertion edge has colour blue \( (\chi_n(c_n^{\text{col}}) = 1) \), but \( \chi_{n+1}(c_n^{\text{col}}) = 0 \), or if the insertion edge has colour red \( (\chi_n(c_n^{\text{col}}) = 0) \), and then necessarily \( \chi_{n+1}(c_n^{\text{col}}) = 0 \) also, we obtain
  \[
  \Pr(\tilde{T}_{n+1} = t_{n+1} | \tilde{T}_n = t_n, T_n^{\text{col}} = t_n^{\text{col}}) = (\alpha + (k - 2)c)/(n - \alpha),
  \]
  because apart from \( a_n^{\text{col}} \to c_n^{\text{col}} \), there are \( k - 2 \) other edges in \( t_n^{\text{col}} \), where insertion and crushing also create \( t_{n+1} \).

Because these conditional probabilities do not depend on \( t_n^{\text{col}} \) and have the form required, we conclude that \( (\tilde{T}_n, n \geq 1) \) obeys the growth rules (i)-(iii) with \( \gamma = \alpha(1 - c) \).

2.3 The Chinese Restaurant Process

An important tool in this paper is the Chinese Restaurant Process (CRP), a partition-valued process \( (\Pi_n, n \geq 1) \) due to Dubins and Pitman, see [22], which generates the Ewens-Pitman two-parameter family of exchangeable random partitions \( \Pi_{\infty} \) of \( N \). In the restaurant framework, each block of a partition is represented by a table and each element of a block by a customer at a table. The construction rules are the following. The first customer sits at the first table and the following customers will be seated at an occupied table or a new one. Given \( n \) customers at \( k \) tables with \( n_j \geq 1 \) customers at the \( j \)th table, customer \( n + 1 \) will be placed at the \( j \)th table with probability \( (n_j - \alpha)/(n + \theta) \), and at a new table with probability \( (\theta + k\alpha)/(n + \theta) \). The parameters \( \alpha \) and \( \theta \) can be chosen as either \( \alpha < 0 \) and \( \theta = -m\alpha \) for some \( m \in \mathbb{N} \) or \( 0 \leq \alpha \leq 1 \) and \( \theta > -\alpha \). We refer to this process as the CRP with \((\alpha, \theta)\)-seating plan.

In the CRP \( (\Pi_n, n \geq 1) \) with \( \Pi_n \in \mathcal{P}[n] \), we can study the block sizes, which leads us to consider the proportion of each table relative to the total number of customers. These proportions converge to limiting frequencies as follows.
Lemma 7 (Theorem 3.2 in [22]). For each pair of parameters \((\alpha, \theta)\) subject to the constraints above, the Chinese restaurant with the \((\alpha, \theta)\)-seating plan generates an exchangeable random partition \(\Pi_\infty\) of \(\mathbb{N}\). The corresponding EPPF is

\[
p_{\alpha, \theta}^\text{PD}(n_1, \ldots, n_k) = \frac{\alpha^{k-1} \Gamma(k + \theta/\alpha) \Gamma(1 + \theta)}{\Gamma(1 + \theta/\alpha) \Gamma(n + \theta)} \prod_{i=1}^{k} \frac{\Gamma(n_i - \alpha)}{\Gamma(1 - \alpha)} , \quad n_i \geq 1, i \in [k]; k \geq 1 : \sum n_i = n,
\]

boundary cases by continuity. The corresponding limiting frequencies of block sizes, in size-biased order of least elements, are GEM\(_{\alpha, \theta}\) and can be represented as

\[
(\tilde{P}_1, \tilde{P}_2, \ldots) = (W_1, \overline{W}_1 W_2, \overline{W}_1 \overline{W}_2 W_3, \ldots)
\]

where the \(W_i\) are independent, \(W_i\) has \(\text{beta}(1 - \alpha, \theta + i\alpha)\) distribution, and \(\overline{W}_i := 1 - W_i\). The distribution of the associated ranked sequence of limiting frequencies is Poisson-Dirichlet PD\(_{\alpha, \theta}\).

We also associate with the EPPF \(p_{\alpha, \theta}^\text{PD}\) the distribution \(q_{\alpha, \theta}^\text{PD}\) of block sizes in decreasing order via \((\Pi)\) and, because the Chinese restaurant EPPF is not the EPPF of a splitting rule leading to \(k \geq 2\) block (we use notation \(q_{\alpha, \theta}^{\text{PD}, \ast}\) for the splitting rules induced by conditioning on \(k \geq 2\) blocks), but can lead to a single block, we also set \(q_{\alpha, \theta}^\text{PD}(n) = p_{\alpha, \theta}^\text{PD}(n)\).

The asymptotic properties of the number \(K_n\) of blocks of \(\Pi_n\) under the \((\alpha, \theta)\)-seating plan depend on \(\alpha\): if \(\alpha < 0\) and \(\theta = -m\alpha\) for some \(m \in \mathbb{N}\), then \(K_n = m\) for all sufficiently large \(n\) a.s.; if \(\alpha = 0\) and \(\theta > 0\), then \(\lim_{n \to \infty} K_n/\log n = \theta\) a.s. The most relevant case for us is \(\alpha > 0\).

Lemma 8 (Theorem 3.8 in [22]). For \(0 < \alpha < 1, \theta > -\alpha\),

\[
\frac{K_n}{n^\alpha} \to S \quad \text{a.s. as } n \to \infty,
\]

where \(S\) has a continuous density on \((0, \infty)\) given by

\[
\frac{d}{ds} \mathbb{P}(S \in ds) = \frac{\Gamma(\theta + 1)}{\Gamma(\theta/\alpha + 1)} s^{-\theta/\alpha} g_\alpha(s),
\]

and \(g_\alpha\) is the density of the Mittag-Leffler distribution with \(p\)th moment \(\Gamma(p + 1)/\Gamma(p\alpha + 1)\).

As an extension of the CRP, Pitman and Winkel in [23] introduced the ordered CRP. Its seating plan is as follows. The tables are ordered from left to right. Put the second table to the right of the first with probability \(\theta/(\alpha + \theta)\) and to the left with probability \(\alpha/(\alpha + \theta)\). Given \(k\) tables, put the \((k + 1)\)st table to the right of the right-most table with probability \(\theta/(k\alpha + \theta)\) and to the left of the left-most or between two adjacent tables with probability \(\alpha/(k\alpha + \theta)\) each.

A composition of \(n\) is a sequence \((n_1, \ldots, n_k)\) of positive numbers with sum \(n\). A sequence of random compositions \(C_n\) of \(n\) is called regenerative if conditionally given that the first part of \(C_n\) is \(n_1\), the remaining parts of \(C_n\) form a composition of \(n - n_1\) with the same distribution as \(C_{n-n_1}\). Given any decrement matrix \((q^{\text{dec}}(n, m), 1 \leq m \leq n)\), there is an associated sequence \(C_n\) of regenerative random compositions of \(n\) defined by specifying that \(q^{\text{dec}}(n, \cdot)\) is the distribution of the first part of \(C_n\). Thus for each composition \((n_1, \ldots, n_k)\) of \(n\),

\[
\mathbb{P}(C_n = (n_1, \ldots, n_k)) = q^{\text{dec}}(n, n_1)q^{\text{dec}}(n - n_1, n_2) \cdots q^{\text{dec}}(n_{k-1} + n_k, n_{k-1})q^{\text{dec}}(n_k, n_k).
\]

Lemma 9 (Proposition 6 (i) in [23]). For each \((\alpha, \theta)\) with \(0 < \alpha < 1\) and \(\theta \geq 0\), denote by \(C_n\) the composition of block sizes in the ordered Chinese restaurant partition with parameters \((\alpha, \theta)\). Then \((C_n, n \geq 1)\) is regenerative, with decrement matrix

\[
q^{\text{dec}}_{\alpha, \theta}(n, m) = \binom{n}{m} \frac{(n-m)\alpha + m\theta}{n} \frac{\Gamma(m-\alpha)\Gamma(n-m+\theta)}{\Gamma(1-\alpha)\Gamma(n+\theta)} (1 \leq m \leq n).
\]
2.4 The splitting rule of alpha-gamma trees and the proof of Proposition [1]

Proposition [1] claims that the unlabelled alpha-gamma trees \( (T^n, n \geq 1) \) have the Markov branching property, identifies the splitting rule and studies the exchangeability of labels. In preparation of the proof of the Markov branching property, we use CRPs to compute the probability function of the first split of \( T^n_1 \) in Proposition [10] We will then establish the Markov branching property from a spinal decomposition result (Lemma [1]) for \( T^n_0 \).

**Proposition 10.** Let \( T^n_0 \) be an unlabelled alpha-gamma tree for some \( 0 \leq \alpha < 1 \) and \( 0 \leq \gamma \leq \alpha \), then the probability function of the first split of \( T^n_0 \) is

\[
q^{seq}_{\alpha, \gamma}(n_1, \ldots, n_k) = \frac{Z_n \Gamma(1 - \alpha)}{\Gamma(n - \alpha)} \left( \gamma + (1 - \alpha - \gamma) \frac{1}{n(n - 1)} \sum_{i \neq j} n_in_j \right) q^{PD*}_{\alpha, -\alpha - \gamma}(n_1, \ldots, n_k),
\]

\( n_1 \geq \ldots \geq n_k \geq 1, k \geq 2: n_1 + \ldots + n_k = n \), where \( Z_n \) is the normalisation constant in [2].

**Proof.** We start from the growth rules of the labelled alpha-gamma trees \( T_n \). Consider the spine root \( v_1 \to v_2 \to \ldots \to v_{L_{n-1}} \to 1 \) of \( T_n \), and the spinal sub-trees \( S^\text{sp}_{ij} \), \( 1 \leq i \leq L_{n-1}, 1 \leq j \leq K_{n,i} \), not containing \( 1 \) of the spinal vertices \( v_i, i \in [L_{n-1}] \). By joining together the sub-trees of the spinal vertex \( v_i \) we form the \( i \)th spinal bush \( S^\text{sp}_i = S^\text{sp}_{i1} \ast \ldots \ast S^\text{sp}_{iK_{n,i}} \). Suppose a bush \( S^\text{sp}_i \) consists of \( k \) sub-trees with \( m \) leaves in total, then its weight will be \( m - k\alpha - \gamma + k\alpha = m - \gamma \) according to growth rule (i) – recall that the total weight of the tree \( T^n_1 \) is \( n - \alpha \).

Now we consider each bush as a table, each leaf \( n = 2, 3, \ldots \) as a customer, \( 2 \) being the first customer. Adding a new leaf to a bush or to an edge on the spine corresponds to adding a new customer to an existing or to a new table. The weights are such that we construct an ordered Chinese restaurant partition of \( \mathbb{N} \setminus \{1\} \) with parameters \((\gamma, 1 - \alpha)\).

Suppose that the first split of \( T_n \) is into tree components with numbers of leaves \( n_1 \geq \ldots \geq n_k \geq 1 \). Now suppose further that leaf 1 is in the sub-tree with \( n_i \) leaves in the first split, then the first spinal bush \( S^\text{sp}_i \) will have \( n_i \) leaves. Notice that this event is equivalent to that of \( n - n_i \) customers sitting at the first table with a total of \( n - 1 \) customers present, in the terminology of the ordered CRP. According to Lemma [9] the probability of this is

\[
q^{dec}_{\gamma, 1 - \alpha}(n - 1, n - n_i) = \binom{n - 1}{n - n_i} \frac{(n - 1)\gamma + (n - n_i)(1 - \alpha)}{n - 1} \frac{\Gamma(n - 1)\Gamma(n - n_i)}{\Gamma(n - \alpha)\Gamma(1 - \gamma)} = \frac{n}{n - n_i} \frac{\gamma + n_i(n - n_i)}{n(n - 1)} \frac{\Gamma(n - 1)\Gamma(n - n_i)}{\Gamma(n - \alpha)\Gamma(1 - \gamma)}.
\]

Next consider the probability that the first bush \( S^\text{sp}_i \) joins together sub-trees with \( n_1 \geq \ldots \geq n_{i-1} \geq n_{i+1} \geq \ldots n_k \geq 1 \) leaves conditional on the event that leaf 1 is in a sub-tree with \( n_i \) leaves. The first bush has a weight of \( n - n_i - \gamma \) and each sub-tree in it has a weight of \( n_j - \alpha, j \neq i \). Consider these \( k - 1 \) sub-trees as tables and the leaves in the first bush as customers. According to the growth procedure, they form a second (unordered, this time) Chinese restaurant partition with parameters \((\alpha, -\gamma)\), whose EPPF is

\[
p^{PD}_{\alpha, -\gamma}(n_1, \ldots, n_{i-1}, n_{i+1}, \ldots, n_k) = \frac{\alpha^{k-2}\Gamma(k - 1 - \gamma/\alpha)\Gamma(1 - \gamma)}{\Gamma(1 - \gamma/\alpha)\Gamma(n - n_i - \gamma)} \prod_{j \in [k] \setminus \{i\}} \frac{\Gamma(n_j - \gamma)}{\Gamma(1 - \alpha)}.
\]

Let \( m_j \) be the number of \( j \)s in the sequence of \( (n_1, \ldots, n_k) \). Based on the exchangeability of the second Chinese restaurant partition, the probability that the first bush consists of sub-trees with \( n_1 \geq \ldots \geq n_{i-1} \geq n_{i+1} \geq \ldots \geq n_k \geq 1 \) leaves conditional on the event that leaf 1 is in one of the \( m_i \) sub-trees with \( n_i \) leaves will be

\[
\frac{m_{n_i}}{m_1! \ldots m_n!} \binom{n - n_i}{n_1, \ldots, n_{i-1}, n_{i+1}, \ldots, n_k} p^{PD}_{\alpha, -\gamma}(n_1, \ldots, n_{i-1}, n_{i+1}, \ldots, n_k).
\]
Thus the joint probability that the first split is \((n_1, \ldots, n_k)\) and that leaf 1 is in a subtree with \(n_i\) leaves is,

\[
\frac{m_{n_1}}{m_1! \ldots m_n!} \left( \frac{n-n_i}{n_1, \ldots, n_i-1, n_{i+1}, \ldots, n_k} \right) q_{\gamma, -\alpha}^{\text{dec}}(n-1, n-n_i) p_{\gamma, -\alpha}^{\text{PD}}(n_1, \ldots, n_i, \ldots, n_k) \]

\[
= m_{n_i} \left( \frac{n_i}{n} + \frac{n_i(n-n_i)}{n(n-1)} (1 - \alpha - \gamma) \right) \frac{Z_n \Gamma(1-\alpha)}{\Gamma(n-\alpha)} q_{\alpha, -\alpha-\gamma}^{\text{PD}^*}(n_1, \ldots, n_k). \tag{7}
\]

Hence the splitting rule will be the sum of \(7\) for all different \(n_i\) (not \(i\)) in \((n_1, \ldots, n_k)\), but they contain factors \(m_{n_i}\), so we can write it as sum over \(i \in [k]\):

\[
q_{\alpha, \gamma}^{\text{seq}}(n_1, \ldots, n_k) = \left( \sum_{i=1}^{k} \left( \frac{n_i}{n} + \frac{n_i(n-n_i)}{n(n-1)} (1 - \alpha - \gamma) \right) \right) \frac{Z_n \Gamma(1-\alpha)}{\Gamma(n-\alpha)} q_{\alpha, -\alpha-\gamma}^{\text{PD}^*}(n_1, \ldots, n_k)
\]

\[
= \left( \gamma + (1 - \alpha - \gamma) \frac{1}{n(n-1)} \sum_{i \neq j} n_i n_j \right) \frac{Z_n \Gamma(1-\alpha)}{\Gamma(n-\alpha)} q_{\alpha, -\alpha-\gamma}^{\text{PD}^*}(n_1, \ldots, n_k).
\]

We can use the nested Chinese restaurants described in the proof to study the subtrees of the spine of \(T_n\). We have decomposed \(T_n\) into the subtrees \(S_{ij}^{\text{sp}}\) of the spine from the root to 1 and can, conversely, build \(T_n\) from \(S_{ij}^{\text{sp}}\), for which we now introduce notation

\[
T_n = \coprod_{i,j} S_{ij}^{\text{sp}}.
\]

We will also write \(\coprod_{i,j} S_{ij}^{\circ}\) when we join together unlabelled trees \(S_{ij}^{\circ}\) along a spine. The following unlabelled version of a spinal decomposition theorem will entail the Markov branching property.

**Lemma 11** (Spinal decomposition). Let \((T_n^{\text{ol}}, n \geq 1)\) be alpha-gamma trees, delabelled apart from label 1. For all \(n \geq 2\), the tree \(T_n^{\text{ol}}\) has the same distribution as \(\coprod_{i,j} S_{ij}^{\circ}\), where

- \(C_{n-1} = (N_1, \ldots, N_{L_{n-1}})\) is a regenerative composition with decrement matrix \(q_{\gamma, 1-\alpha}^{\text{dec}}\),
- conditionally given \(L_{n-1} = \ell\) and \(N_i = n_i, i \in [\ell]\), the sizes \(N_{i1} \geq \ldots \geq N_i K_{n,i} \geq 1\) form random compositions of \(n_i\) with distribution \(q_{\alpha, -\gamma}^{\text{PD}}\), independently for \(i \in [\ell]\),
- conditionally given also \(K_{n,i} = k_i\) and \(N_{ij} = n_{ij}\), the trees \(S_{ij}^{\circ}\), \(j \in [k_i], i \in [\ell]\), are independent and distributed as \(T_{n_{ij}}^{\circ}\).

**Proof.** For an induction on \(n\), note that the claim is true for \(n = 2\), since \(T_n^{\text{ol}}\) and \(\coprod_{i,j} S_{ij}^{\circ}\) are deterministic for \(n = 2\). Suppose then that the claim is true for some \(n \geq 2\) and consider \(T_{n+1}^{\text{ol}}\).

The growth rules (i)-(iii) of the labelled alpha-gamma tree \(T_n\) are such that

- leaf \(n+1\) is inserted into a new bush or any of the bushes \(S_i^{\text{sp}}\) selected according to the rules of the ordered CRP with \((\gamma, 1-\alpha)\)-seating plan,
- further into a new subtree or any of the subtrees \(S_{ij}^{\text{sp}}\) of the selected bush \(S_i^{\text{sp}}\) according to the rules of a CRP with \((\alpha, -\gamma)\)-seating plan,
- and further within the subtree \(S_{ij}^{\text{sp}}\) according to the weights assigned by (i) and growth rules (ii)-(iii).

These selections do not depend on \(T_n\) except via \(T_n^{\text{ol}}\). In fact, since labels do not feature in the growth rules (i)-(iii), they are easily seen to induce growth rules for partially labelled alpha-gamma trees \(T_n^{\text{ol}}\), and also for unlabelled alpha-gamma trees such as \(S_{ij}^{\circ}\).

From these observations and the induction hypothesis, we deduce the claim for \(T_{n+1}^{\text{ol}}\). 

\[\square\]
Proof of Proposition 2. (a) Firstly, the distributions of the first splits of the unlabelled alpha-gamma trees \( T_n^0 \) were calculated in Proposition 10 for 0 ≤ \( \alpha \) < 1 and 0 ≤ \( \gamma \) ≤ \( \alpha \).

Secondly, let 0 ≤ \( \alpha \) ≤ 1 and 0 ≤ \( \gamma \) ≤ \( \alpha \). By the regenerative property of the spinal composition \( C_{n-1} \) and the conditional distribution of \( T_n^1 \) given \( C_{n-1} \) identified in Lemma 11, we obtain that given \( N_1 = m, K_{n,1} = k_1 \) and \( N_{1j} = n_{1j}, j \in [k_1] \), the subtrees \( S_{1j}, j \in [k_1] \), are independent alpha-gamma trees distributed as \( T_{n1j}^0 \), also independent of the remaining tree \( S_{1,0} := \prod_{i \geq 2} S_{ij}^0 \), which, by Lemma 11, has the same distribution as \( T_{n-m}^0 \).

This is equivalent to saying that conditionally given that the first split is into subtrees with \( n_1 ≥ \ldots ≥ n_i ≥ \ldots ≥ n_k ≥ 1 \) leaves and that leaf 1 is in a subtree with \( n_i \) leaves, the delabelled subtrees \( S_1^0, \ldots, S_k^0 \) of the common ancestor are independent and distributed as \( T_{n1j}^0 \) respectively, \( j \in [k] \). Since this conditional distribution does not depend on \( i \), we have established the Markov branching property of \( T_n^0 \).

(b) Notice that if \( \gamma = 1 - \alpha \), the alpha-gamma model is the model related to stable trees, the labelling of which is known to be exchangeable, see Section 3.4.

On the other hand, if \( \gamma ≠ 1 - \alpha \), let us turn to look at the distribution of \( T_3 \).

![Diagram](image)

We can see the probabilities of the two labelled tree in the above picture are different although they have the same unlabelled tree. So if \( \gamma ≠ 1 - \alpha \), \( T_n \) is not exchangeable. □

### 2.5 Sampling consistency and strong sampling consistency

Recall that an unlabelled Markov branching tree \( T_n^0, n ≥ 2 \) has the property of sampling consistency, if when we select a leaf uniformly and delete it (together with the adjacent branch point if its degree is reduced to 2), then the new tree, denoted by \( T_{n-1}^0 \), is distributed as \( T_{n-1}^0 \). Denote by \( d: \mathbb{D}_n → \mathbb{D}_{n-1} \) the induced deletion operator on the space \( \mathbb{D}_n \) of probability measures on \( T_n^0 \), so that for the distribution \( P_n \) of \( T_n^0 \), we define \( d(P_n) \) as the distribution of \( T_{n-1}^0 \). Sampling consistency is equivalent to \( d(P_n) = P_{n-1} \). This property is also called deletion stability in [12].

**Proposition 12.** The unlabelled alpha-gamma trees for 0 ≤ \( \alpha \) ≤ 1 and 0 ≤ \( \gamma \) ≤ \( \alpha \) are sampling consistent.

**Proof.** The sampling consistency formula (14) in [16] states that \( d(P_n) = P_{n-1} \) is equivalent to

\[
q(n_1, \ldots, n_k) = \sum_{i=1}^{k} \left( \frac{(n_i + 1)(m_{n_i} + 1)}{nm_{n_i}} q((n_1, \ldots, n_i + 1, \ldots, n_k)^1) + \frac{m_1 + 1}{n} q(n_1, \ldots, n_k, 1) + \frac{1}{n} q(n-1,1)q(n_1, \ldots, n_k) \right)
\]

(8)

for all \( n_1 ≥ \ldots ≥ n_k ≥ 1 \) with \( n_1 + \ldots + n_k = n - 1 \), where \( m_j \) is the number of \( n_i, i \in [k] \), that equal \( j \), and where \( q \) is the splitting rule of \( T_n^0 \sim P_n \). In terms of EPPFs [11], formula (8) is equivalent to

\[
(1 - p(n-1,1)) p(n_1, \ldots, n_k) = \sum_{i=1}^{k} p((n_1, \ldots, n_i + 1, \ldots, n_k)^1) + p(n_1, \ldots, n_k, 1).
\]

(9)
Now according to Proposition 1 the EPPF of the alpha-gamma model with $\alpha < 1$ is

$$p_{\alpha, \gamma}^\text{seq}(n_1, \ldots, n_k) = \frac{Z_n}{\Gamma_\alpha(n)} \left( \gamma + (1 - \alpha - \gamma) \frac{1}{(n-1)(n-2)} \sum_{u \neq v} n_u n_v \right) p_{\alpha, -\alpha - \gamma}^\text{PD^*}(n_1, \ldots, n_k),$$

where $\Gamma_\alpha(n) = \Gamma(n - \alpha)/\Gamma(1 - \alpha)$. Therefore, $p_{\alpha, \gamma}^\text{seq}(n_1, \ldots, n_i + 1, \ldots, n_k)$ can be written as

$$\left( p_{\alpha, \gamma}^\text{seq}(n_1, \ldots, n_k) + 2(1 - \alpha - \gamma) \frac{(n-2)(n-1-n_i)}{n(n-1)(n-2)} \frac{Z_n}{\Gamma_\alpha(n)} p_{\alpha, -\alpha - \gamma}^\text{PD^*}(n_1, \ldots, n_k) \right) \times \frac{n_i - \alpha}{n - 1 - \alpha}$$

and $p_{\alpha, \gamma}^\text{seq}(n_1, \ldots, n_k, 1)$ as

$$\left( p_{\alpha, \gamma}^\text{seq}(n_1, \ldots, n_k) + 2(1 - \alpha - \gamma) \frac{(n-1)(n-2) - \sum_{u \neq v} n_u n_v}{n(n-1)(n-2)} \frac{Z_n}{\Gamma_\alpha(n)} p_{\alpha, -\alpha - \gamma}^\text{PD^*}(n_1, \ldots, n_k) \right) \times \frac{(k-1)\alpha - \gamma}{n - 1 - \alpha}.$$

Sum over the above formulas, then the right-hand side of (9) is

$$\left( 1 - \frac{1}{n - 1 - \alpha} \left( \gamma + \frac{2}{n} (1 - \alpha - \gamma) \right) \right) p_{\alpha, \gamma}^\text{seq}(n_1, \ldots, n_k).$$

Notice that $p_{\alpha, \gamma}^\text{seq}(n-1, 1) = (\gamma + 2(1 - \alpha - \gamma)/n) / (n-1 - \alpha)$. Hence, the splitting rules of the alpha-gamma model satisfy (9), which implies sampling consistency for $\alpha < 1$. The case $\alpha = 1$ is postponed to Section 3.2.

Moreover, sampling consistency can be enhanced to strong sampling consistency by requiring that $(T_{n-1}^\circ, T_n^\circ)$ has the same distribution as $(T_{n-1}^\circ, T_n^\circ)$.

Proposition 13. The alpha-gamma model is strongly sampling consistent if and only if $\gamma = 1 - \alpha$.

Proof. For $\gamma = 1 - \alpha$, the model is known to be strongly sampling consistent, cf. Section 3.4.

If $\gamma \neq 1 - \alpha$, consider the above two deterministic unlabelled trees.

$\mathbb{P}(T_4^\circ = t_3^\circ) = \delta_{a, \gamma}^\text{seq}(2, 1, 1) \delta_{a, \gamma}^\text{seq}(1, 1) = (\alpha - \gamma)(5 - 5\alpha + \gamma)/(2 - \alpha)(3 - \alpha)).$

Then we delete one of the two leaves at the first branch point of $t_4^\circ$ to get $t_3^\circ$. Therefore

$$\mathbb{P}((T_{n-1}^\circ, T_n^\circ) = (t_3^\circ, t_4^\circ)) = \frac{1}{2} \mathbb{P}(T_4^\circ = t_3^\circ) = \frac{(\alpha - \gamma)(5 - 5\alpha + \gamma)}{2(2 - \alpha)(3 - \alpha)}.$$

On the other hand, if $T_3^\circ = t_3^\circ$, we have to add the new leaf to the first branch point to get $t_4^\circ$. Thus

$$\mathbb{P}((T_3^\circ, T_4^\circ) = (t_3^\circ, t_4^\circ)) = \frac{\alpha - \gamma}{3 - \alpha} \mathbb{P}(T_3^\circ = t_3^\circ) = \frac{(\alpha - \gamma)(2 - 2\alpha + \gamma)}{(2 - \alpha)(3 - \alpha)}.$$

It is easy to check that $\mathbb{P}((T_{n-1}^\circ, T_n^\circ) = (t_3^\circ, t_4^\circ)) \neq \mathbb{P}((T_3^\circ, T_4^\circ) = (t_3^\circ, t_4^\circ))$ if $\gamma \neq 1 - \alpha$, which means that the alpha-gamma model is then not strongly sampling consistent.
3 Dislocation measures and asymptotics of alpha-gamma trees

3.1 Dislocation measures associated with the alpha-gamma-splitting rules

Theorem 2 claims that the alpha-gamma trees are sampling consistent, which we proved in Section 2.5 and identifies the integral representation of the splitting rule in terms of a dislocation measure, which we will now establish.

Proof of Theorem 2. Firstly, we make some rearrangement for the coefficient of the sampling consistent splitting rules of alpha-gamma trees identified in Proposition 10:

\[
\frac{1}{n(n - 1)} \sum_{i \neq j} n_i n_j = \frac{(n + 1 - \alpha - \gamma)(n - \alpha - \gamma)}{n(n - 1)} \left( \gamma + (1 - \alpha - \gamma) \left( \sum_{i \neq j} A_{ij} + 2 \sum_{i=1}^{k} B_i + C \right) \right),
\]

where

\[
A_{ij} = \frac{(n_i - \alpha)(n_j - \alpha)}{(n + 1 - \alpha - \gamma)(n - \alpha - \gamma)},
\]

\[
B_i = \frac{(n_i - \alpha)((k - 1)\alpha - \gamma)}{(n + 1 - \alpha - \gamma)(n - \alpha - \gamma)},
\]

\[
C = \frac{((k - 1)\alpha - \gamma)(k\alpha - \gamma)}{(n + 1 - \alpha - \gamma)(n - \alpha - \gamma)}.
\]

Notice that \( B_{ikn^*} \) simplifies to

\[
\frac{(n_i - \alpha)((k - 1)\alpha - \gamma)}{(n + 1 - \alpha - \gamma)(n - \alpha - \gamma)} \frac{\alpha^{k-1} \Gamma(k - 1 - \gamma/\alpha)}{Z_n \Gamma(1 - \gamma/\alpha)} \Gamma_{\alpha}(n_1) \ldots \Gamma_{\alpha}(n_k)
\]

\[
= \frac{Z_{n+2}}{Z_n(n + 1 - \alpha - \gamma)(n - \alpha - \gamma)} \frac{\alpha^{k-1} \Gamma(k - 1 - \gamma/\alpha)}{Z_{n+2} \Gamma(1 - \gamma/\alpha)} \Gamma_{\alpha}(n_1) \ldots \Gamma_{\alpha}(n_i + 1) \ldots \Gamma_{\alpha}(n_k)
\]

\[
= \frac{Z_{n+2}}{Z_n} p_{\alpha,\alpha-\gamma}(n_1, \ldots, n_i + 1, \ldots, n_k, 1),
\]

where \( \Gamma_{\alpha}(n) = \Gamma(n - \alpha)/\Gamma(1 - \alpha) \) and \( Z_n = Z_0 \alpha \Gamma(1 - \gamma/\alpha)/\Gamma(n - \alpha - \gamma) \) is the normalisation constant in (4) for \( \nu = PD_{\alpha,\alpha-\gamma}^* \), as can be read from [17 Formula (17)]. According to (4),

\[
p_{\alpha,\alpha-\gamma}^*(n_1, \ldots, n_k) = \frac{1}{Z_n} \int_{\mathbb{S}^1} \sum_{i_1, \ldots, i_k \geq 1 \text{ distinct}} \prod_{l=1}^{k} s_{i_l}^{n_l} PD_{\alpha,\alpha-\gamma}^*(ds).
\]

Thus,

\[
B_{ikn^*} = \frac{1}{Z_n} \int_{\mathbb{S}^1} \sum_{i_1, \ldots, i_k \geq 1 \text{ distinct}} \prod_{l=1}^{k} s_{i_l}^{n_l} \left( \sum_{u \in \{i_1, \ldots, i_k\}, v \notin \{i_1, \ldots, i_k\}} s_u s_v \right) PD_{\alpha,\alpha-\gamma}^*(ds).
\]

Similarly,

\[
A_{ij} = \frac{1}{Z_n} \int_{\mathbb{S}^1} \sum_{i_1, \ldots, i_k \geq 1 \text{ distinct}} \prod_{l=1}^{k} s_{i_l}^{n_l} \left( \sum_{u, v \in \{i_1, \ldots, i_k\}, u \neq v} s_u s_v \right) PD_{\alpha,\alpha-\gamma}^*(ds)
\]

\[
C_{ij} = \frac{1}{Z_n} \int_{\mathbb{S}^1} \sum_{i_1, \ldots, i_k \geq 1 \text{ distinct}} \prod_{l=1}^{k} s_{i_l}^{n_l} \left( \sum_{u, v \in \{i_1, \ldots, i_k\}, u \neq v} s_u s_v \right) PD_{\alpha,\alpha-\gamma}^*(ds).
\]
Hence, the EPPF $p_{\alpha,\gamma}^\text{seq}(n_1, \ldots, n_k)$ of the sampling consistent splitting rule takes the following form:

\[
\frac{(n+1-\alpha-\gamma)(n-\alpha-\gamma)}{n(n-1)\Gamma_\alpha(n)}Z_\alpha \left( \gamma + (1-\alpha-\gamma) \left( \sum_{i \neq j} A_{ij} + 2 \sum_{i=1}^k B_i + C \right) \right)^{p_{\alpha,\gamma}^\text{PD}^* (n_1, \ldots, n_k)}
\]

\[
= \frac{1}{Y_n} \int_{S_1^{\alpha,\gamma}} \prod_{i, j \geq 1} s_{ij}^{n_i} \left( \gamma + (1-\alpha-\gamma) \sum_{i \neq j} s_1 s_j \right) PD_{\alpha,\gamma-\gamma}^*(ds), \tag{11}
\]

where $Y_n = n(n-1)\Gamma_\alpha(n)\alpha \Gamma(1-\gamma/\alpha)/\Gamma(n+2-\alpha-\gamma)$ is the normalization constant. Hence, we have $\nu_{\alpha,\gamma}(ds) = \left( \gamma + (1-\alpha-\gamma) \sum_{i \neq j} s_1 s_j \right) PD_{\alpha,\gamma-\gamma}^*(ds)$.

3.2 The alpha-gamma model when $\alpha = 1$, spine with bushes of singleton-trees

Within the discussion of the alpha-gamma model so far, we restricted to $0 \leq \alpha < 1$. In fact, we can still get some interesting results when $\alpha = 1$. The weight of each leaf edge is $1 - \alpha$ in the growth procedure of the alpha-gamma model. If $\alpha = 1$, the weight of each leaf edge becomes zero, which means that the new leaf can only be inserted to internal edges or branch points. Starting from the two leaf tree, leaf 3 must be inserted into the root edge or the branch point. Similarly, any new leaf must be inserted into the spine leading from the root to the common ancestor of leaf 1 and leaf 2. Hence, the shape of the tree is just a spine with some bushes of one-leaf subtrees rooted on it. Moreover, the first split of an $n$-leaf tree will be $(n-k+1, 1, \ldots, 1)$ for some $2 \leq k \leq n-1$. The cases $\gamma = 0$ and $\gamma = 1$ lead to degenerate trees with, respectively, all leaves connected to a single branch point and all leaves connected to a spine of binary branch points (comb).

**Proposition 14.** Consider the alpha-gamma model with $\alpha = 1$ and $0 < \gamma < 1$.

(a) The model is sampling consistent with splitting rules

\[
q_{\alpha,\gamma}^\text{seq}(n_1, \ldots, n_k)
\]

\[
= \begin{cases} 
\gamma \Gamma_\gamma(k-1)/(k-1)!, & \text{if } 2 \leq k \leq n-1 \text{ and } (n_1, \ldots, n_k) = (n-k+1, 1, \ldots, 1); \\
\Gamma_\gamma(n-1)/(n-2)!, & \text{if } k = n \text{ and } (n_1, \ldots, n_k) = (1, \ldots, 1); \\
0, & \text{otherwise},
\end{cases}
\tag{12}
\]

where $n_1 \geq \ldots \geq n_k \geq 1$ and $n_1 + \ldots + n_k = n$.

(b) The dislocation measure associated with the splitting rules can be expressed as follows

\[
\int_{S_1} f(s_1, 0, \ldots) \nu_{\alpha,\gamma}(ds) = \int_0^1 f(s_1, 0, \ldots) \left( \gamma(1-s_1)^{-1-\gamma}ds_1 + \delta_0(ds_1) \right). \tag{13}
\]

In particular, it does not satisfy $\nu(\{s \in S_1 : s_1 + s_2 + \ldots < 1\}) = 0$.

**Proof.** (a) We start from the growth procedure of the alpha-gamma model when $\alpha = 1$. Consider a first split into $(n-k+1, 1, \ldots, 1)$ for some labelled $n$-leaf tree. Suppose its first branch point is created when the leaf $l$ is inserted to the root edge for $l \geq 3$. At this time the first split is $(l-1, 1)$ with a probability $\gamma/(l-2)$ as $\alpha = 1$. In the following insertion, leaves $l+1, \ldots, n$ have been added either to the first branch point or to the subtree with $l-1$ leaves at this time. Hence the probability that the first split of this tree is $(n-k+1, 1, \ldots, 1)$ is

\[
\frac{(n-k-1)!}{(n-2)!} \gamma \Gamma_\gamma(k-1),
\]
which does not depend on \( l \). Notice that the growth rules imply that if the first split is \((n - k + 1,1,\ldots,1)\) with \( k \leq n - 1 \), then leaves 1 and 2 will be located in the subtree with \( n - k + 1 \) leaves. There are \( \binom{n-2}{n-k-1} \) labelled trees with the above first split. Therefore,

\[
q_{1,\gamma}^{\text{seq}}(n - k + 1, 1, \ldots, 1) = \left( \frac{n - 2}{n - k - 1} \right) \frac{(n - k - 1)!}{(n - 2)!} \gamma \Gamma_{\gamma}(k - 1) = \gamma \Gamma_{\gamma}(k - 1)/(k - 1)!.
\]

On the other hand, there is only one \( n \)-leaf labelled tree with a first split \((1,\ldots,1)\) and in this case, all leaves have been added to the only branch point. Hence

\[
q_{1,\gamma}^{\text{seq}}(1,\ldots,1) = \Gamma_{\gamma}(n - 1)/(n - 2)!.
\]

For sampling consistency, we check criterion \( [8] \), which reduces to the two formulas

\[
(1 - q_{1,\gamma}^{\text{seq}}(n - 1, 1))q_{1,\gamma}^{\text{seq}}(n - k + 1, 1, \ldots, 1) = \frac{k}{n} q_{1,\gamma}^{\text{seq}}(n - k, 1, \ldots, 1) + \frac{n - k + 1}{n} q_{1,\gamma}^{\text{seq}}(n - k + 1, \ldots, 1)
\]

\[
(1 - q_{1,\gamma}^{\text{seq}}(n - 1, 1))q_{1,\gamma}^{\text{seq}}(1,\ldots,1) = \frac{2}{n} q_{1,\gamma}^{\text{seq}}(2,1,\ldots,1) + q_{1,\gamma}^{\text{seq}}(1,\ldots,1).
\]

(b) According to \( [12] \),

\[
q_{1,\gamma}^{\text{seq}}(n - k + 1, 1, \ldots, 1)
= \left( \frac{n}{n - k + 1} \right) \frac{\Gamma_{\gamma}(n + 1)}{n!} \gamma \beta(n - k + 2, k - 1 - \gamma)
= \frac{1}{Y_{n}} \left( \frac{n}{n - k + 1} \right) \int_{0}^{1} s_{1}^{n-k+1}(1 - s_{1})^{k-1} (\gamma(1 - s_{1})^{-1-\gamma}ds_{1})
= \frac{1}{Y_{n}} \left( \frac{n}{n - k + 1} \right) \int_{0}^{1} s_{1}^{n-k+1}(1 - s_{1})^{k-1} (\gamma(1 - s_{1})^{-1-\gamma}ds_{1} + \delta_{0}(ds_{1})) ,
\]

where \( Y_{n} = n!/\Gamma_{\gamma}(n + 1) \). Similarly,

\[
q_{1,\gamma}^{\text{seq}}(1,\ldots,1) = \frac{1}{Y_{n}} \int_{0}^{1} (n - 1 - s_{1})^{n-1}s_{1} + (1 - s_{1})^{n} (\gamma(1 - s_{1})^{-1-\gamma}ds_{1} + \delta_{0}(ds_{1})) .
\]

Formulas \( [14] \) and \( [15] \) are of the form of \( [16] \) Formula (2)], which generalises \( [4] \) to the case where \( \nu \) does not necessarily satisfy \( \nu(\{s \in S_{1}: s_{1} + s_{2} + \ldots < 1\}) = 0 \), hence \( \nu_{1,\gamma} \) is identified.

### 3.3 Continuum random trees and self-similar trees

Let \( B \subset \mathbb{N} \) finite. A labelled tree with edge lengths is a pair \( \vartheta = (t, \eta) \), where \( t \in T_{B} \) is a labelled tree, \( \eta = (\eta_{A}, A \in t \setminus \{ \text{ROOT} \}) \) is a collection of marks, and every edge \( C \to A \) of \( t \) is associated with mark \( \eta_{A} \in (0, \infty) \) at vertex \( A \), which we interpret as the edge length of \( C \to A \). Let \( \Theta_{B} \) be the set of such trees \( (t, \eta) \) with \( t \in T_{B} \).

We now introduce continuum trees, following the construction by Evans et al. in \( [9] \). A complete separable metric space \((\tau, d)\) is called an \( \mathbb{R} \)-tree, if it satisfies the following two conditions:

1. for all \( x, y \in \tau \), there is an isometry \( \varphi_{x,y} : [0,d(x,y)] \to \tau \) such that \( \varphi_{x,y}(0) = x \) and \( \varphi_{x,y}(d(x,y)) = y \).
2. for every injective path \( c : [0,1] \to \tau \) with \( c(0) = x \) and \( c(1) = y \), one has \( c([0,1]) = \varphi_{x,y}([0,d(x,y)]) \).

We will consider rooted \( \mathbb{R} \)-trees \((\tau, d, \rho)\), where \( \rho \in \tau \) is a distinguished element, the root. We think of the root as the lowest element of the tree.
We denote the range of \( \varphi_{x,y} \) by \([x, y]\) and call the quantity \( d(\rho, x) \) the **height** of \( x \). We say that \( x \) is an ancestor of \( y \) whenever \( x \in [\rho, y] \). We let \( x \land y \) be the unique element in \( \tau \) such that \([\rho, x] \cap [\rho, y] = [\rho, x \land y] \), and call it the **highest common ancestor** of \( x \) and \( y \) in \( \tau \). Denoted by \((\tau_x, d_{\tau_x}, x)\) the set of \( y \in \tau \) such that \( x \) is an ancestor of \( y \), which is an \( \mathbb{R} \)-tree rooted at \( x \) that we call the **fringe subtree** of \( \tau \) above \( x \).

Two rooted \( \mathbb{R} \)-trees \((\tau, d, \rho), (\tau', d', \rho')\) are called **equivalent** if there is a bijective isometry between the two metric spaces that maps the root of one to the root of the other. We also denote by \( \Theta \) the set of equivalence classes of compact rooted \( \mathbb{R} \)-trees. We define the **Gromov-Hausdorff distance** between two rooted \( \mathbb{R} \)-trees (or their equivalence classes) as
\[
d_{\text{GH}}(\tau, \tau') = \inf \{d_H(\tilde{\tau}, \tilde{\tau}')\}
\]
where the infimum is over all metric spaces \( E \) and isometric embeddings \( \tilde{\tau} \subset E \) of \( \tau \) and \( \tilde{\tau}' \subset E \) of \( \tau' \) with common root \( \tilde{\rho} \in E \); the Hausdorff distance on compact subsets of \( E \) is denoted by \( d_H \). Evans et al. [2] showed that \((\Theta, d_{\text{GH}})\) is a complete separable metric space.

We call an element \( x \in \tau \), \( x \neq \rho \), in a rooted \( \mathbb{R} \)-tree \( \tau \), a **leaf** if its removal does not disconnect \( \tau \), and let \( \mathcal{L}(\tau) \) be the set of leaves of \( \tau \). On the other hand, we call an element of \( \tau \) a **branch point**, if it has the form \( x \land y \) where \( x \) is neither an ancestor of \( y \) nor vice-versa. Equivalently, we can define branch points as points disconnecting \( \tau \) into three or more connected components when removed. We let \( B(\tau) \) be the set of branch points of \( \tau \).

A **weighted \( \mathbb{R} \)-tree** \((\tau, \mu)\) is called a **continuum tree** [1], if \( \mu \) is a probability measure on \( \tau \) and
1. \( \mu \) is supported by the set \( \mathcal{L}(\tau) \),
2. \( \mu \) has no atom,
3. for every \( x \in \tau \setminus \mathcal{L}(\tau) \), \( \mu(\tau_x) > 0 \).

A **continuum random tree** (CRT) is a random variable whose values are continuum trees, defined on some probability space \((\Omega, \mathcal{A}, \mathbb{P})\). Several methods to formalize this have been developed [2][10][13]. For technical simplicity, we use the method of Aldous [2]. Let the space \( \ell_1 = \ell_1(\mathbb{N}) \) be the base space for defining CRTs. We endow the set of compact subsets of \( \ell_1 \) with the Hausdorff metric, and the set of probability measures on \( \ell_1 \) with any metric inducing the topology of weak convergence, so that the set of pairs \((T, \mu)\) where \( T \) is a rooted \( \mathbb{R} \)-tree embedded as a subset of \( \ell_1 \) and \( \mu \) is a measure on \( T \), is endowed with the product \( \sigma \)-algebra.

An exchangeable \( \mathcal{P}_n \)-valued fragmentation process \((\Pi(t), t \geq 0)\) is called **self-similar** with index \( a \in \mathbb{R} \) if given \( \Pi(t) = \pi = \{\pi_i, i \geq 1\} \) with asymptotic frequencies \( |\pi_i| = \lim_{n \to \infty} n^{-1} \lfloor n \rfloor \cap \pi_j \), the random variable \( \Pi(t + s) \) has the same law as the random partition whose blocks are those of \( \pi_i \cap \Pi(i)(\lfloor n \rfloor s), i \geq 1, \) where \( \Pi(i), i \geq 1 \) is a sequence of i.i.d. copies of \((\Pi(t), t \geq 0)\).

The process \((\Pi(t)^{\frac{1}{a}}, t \geq 0)\) is an \( S_1 \)-valued self-similar fragmentation process. Bertoin [5] proved that the distribution of a \( \mathcal{P}_n \)-valued self-similar fragmentation process is determined by a triple \((a, c, \nu)\), where \( a \in \mathbb{R}, c \geq 0 \) and \( \nu \) is a dislocation measure on \( S^1 \). For this article, we are only interested in the case \( c = 0 \) and when \( \nu(s_1 + s_2 + \ldots < 1) = 0 \). We call \((a, \nu)\) the characteristic pair. When \( a = 0 \), the process \((\Pi(t), t \geq 0)\) is also called homogeneous fragmentation process.

A CRT \((T, \mu)\) is a **self-similar** CRT with index \( a = -\gamma < 0 \) if for every \( t \geq 0 \), given \((\mu(T_t^i), i \geq 1)\) where \( T_t^i, i \geq 1 \) is the ranked order of connected components of the open set \( \{x \in \tau : d(x, \rho(\tau)) > t\} \), the continuum random trees
\[
\left( \frac{\mu(T_t^1)^{-\gamma}T_t^1}{\mu(T_t^1)} \right), \left( \frac{\mu(T_t^2)^{-\gamma}T_t^2}{\mu(T_t^2)} \right), \ldots
\]
are i.i.d copies of \((T, \mu)\), where \( \mu(T_t^i)^{-\gamma}T_t^i \) is the tree that has the same set of points as \( T_t^i \), but whose distance function is divided by \( \mu(T_t^i)^{-\gamma} \). Haas and Miermont in [15] have shown that there exists a self-similar continuum random tree \( T_{(\gamma, \nu)} \) characterized by such a pair \((\gamma, \nu)\), which can be constructed from a self-similar fragmentation process with characteristic pair \((\gamma, \nu)\).
3.4 The alpha-gamma model when \( \gamma = 1 - \alpha \), sampling from the stable CRT

Let \( (T, \rho, \mu) \) be the stable tree of Duquesne and Le Gall \([7]\). The distribution on \( \Theta \) of any CRT is determined by its so-called finite-dimensional marginals: the distributions of \( \mathcal{R}_k \), \( k \geq 1 \), the subtrees \( \mathcal{R}_k \subset T \) defined as the discrete trees with edge lengths spanned by \( \rho, U_1, \ldots, U_k \), where given \( (T, \mu) \), the sequence \( U_i \in T, i \geq 1 \), of leaves is sampled independently from \( \mu \). See also \([21, 8, 16, 17, 18]\) for various approaches to stable trees. Let us denote the discrete tree without edge lengths associated with \( \mathcal{R}_k \) by \( T_k \) and note the Markov branching structure.

**Lemma 15** (Corollary 22 in \([16]\)). Let \( 1/\alpha \in (1, 2) \). The trees \( T_n, n \geq 1 \), sampled from the \((1/\alpha)\)-stable CRT are Markov branching trees, whose splitting rule has EPPF

\[
p_{1/\alpha}^{\text{stable}}(n_1, \ldots, n_k) = \frac{\alpha^{k-2}(k-1/\alpha)\Gamma(2-\alpha)}{(2-1/\alpha)\Gamma(n-\alpha)} \prod_{j=1}^{k} \frac{\Gamma(n_j - \alpha)\Gamma(1-\alpha)}{\Gamma(1)}
\]

for any \( k \geq 2, n_1 \geq 1, \ldots, n_k \geq 1, n = n_1, \ldots, n_k \).

We recognize \( p_{1/\alpha}^{\text{stable}} = p_{\alpha,-1}^{\text{PD}} \) in \([2]\), and by Proposition [1] we have \( p_{\alpha,-1}^{\text{PD}} = p_{\alpha,1-\alpha}^{\text{seq}} \). This observation yields the following corollary:

**Corollary 16.** The alpha-gamma trees with \( \gamma = 1 - \alpha \) are strongly sampling consistent and exchangeable.

**Proof.** These properties follow from the representation by sampling from the stable CRT, particularly the exchangeability of the sequence \( U_i, i \geq 1 \). Specifically, since \( U_i, i \geq 1 \), are conditionally independent and distributed given \( (T, \mu) \), they are exchangeable. If we denote by \( \mathcal{L}_{n,-1} \) the random set of leaves \( \mathcal{L}_n = \{U_1, \ldots, U_n\} \) with a uniformly chosen member removed, then \( \mathcal{L}_{n,-1} \) has the same conditional distribution as \( \mathcal{L}_{n-1} \). Hence the pairs of (unlabelled) tree shapes spanned by \( \rho \) and these sets of leaves have the same distribution — this is strong sampling consistency as defined before Proposition [13]. \( \square \)

3.5 Dislocation measures in size-biased order

In actual calculations, we may find that the splitting rules in Proposition [1] are quite difficult and the corresponding dislocation measure \( \nu \) is always inexplicit, which leads us to transform \( \nu \) to a more explicit form. The method proposed here is to change the space \( S^1 \) into the space \([0,1]^{\mathbb{N}}\) and to rearrange the elements \( s \in S^1 \) under \( \nu \) into the size-biased random order that places \( s_{i_1} \) first with probability \( s_{i_1} \) (its size) and, successively, the remaining ones with probabilities \( s_{i_j}/(1-s_{i_1} - \ldots - s_{i_{j-1}}) \) proportional to their sizes \( s_{i_j} \) into the following positions, \( j \geq 2 \).

**Definition 2.** We call a measure \( \nu^{sb} \) on the space \([0,1]^{\mathbb{N}}\) the size-biased dislocation measure associated with dislocation measure \( \nu \), if for any subset \( A_1 \times A_2 \times \ldots \times A_k \times [0,1]^{\mathbb{N}} \) of \([0,1]^{\mathbb{N}}\),

\[
\nu^{sb}(A_1 \times A_2 \times \ldots \times A_k \times [0,1]^{\mathbb{N}}) = \sum_{i_1 \ldots i_k \geq 1 \text{ distinct}} \int_{\{s \in S^1 : s_1 \in A_1, \ldots, s_k \in A_k\}} \frac{s_{i_1} \ldots s_{i_k}}{\prod_{j=1}^{k-1}(1 - \sum_{l=1}^{j} s_{i_l})} \nu(ds)
\]

(16)

for any \( k \in \mathbb{N} \), where \( \nu \) is a dislocation measure on \( S^1 \) satisfying \( \nu(s \in S^1 : s_1 + s_2 + \ldots < 1) = 0 \). We also denote by \( \nu_k^{sb}(A_1 \times A_2 \times \ldots \times A_k) = \nu^{sb}(A_1 \times A_2 \times \ldots \times A_k \times [0,1]^{\mathbb{N}}) \) the distribution of the first \( k \) marginals.

The sum in (16) is over all possible rank sequences \((i_1, \ldots, i_k)\) to determine the first \( k \) entries of the size-biased vector. The integral in (16) is over the decreasing sequences that have the \( j \)th entry of the re-ordered vector fall into \( A_j \), \( j \in [k] \). Notice that the support of such a size-biased dislocation measure \( \nu^{sb} \) is a subset of \( S^{sb} := \{s \in [0,1]^{\mathbb{N}} : \sum_{i=1}^{\infty} s_i = 1\} \). If we denote by \( s^j \) the sequence \( s \in S^{sb} \) rearranged into ranked order, taking (16) into formula (1), we obtain
Proposition 17. The EPPF associated with a dislocation measure \( \nu \) can be represented as:

\[
p(n_1, \ldots, n_k) = \frac{1}{Z_n} \int_{[0,1]^k} x_1^{n_1-1} \cdots x_k^{n_k-1} \prod_{j=1}^{k-1} (1 - \sum_{l=1}^{j} x_l) \nu^{sb}(dx),
\]

where \( \nu^{sb} \) is the size-biased dislocation measure associated with \( \nu \), where \( n_1 \geq \ldots \geq n_k \geq 1, k \geq 2, n = n_1 + \ldots + n_k \) and \( x = (x_1, \ldots, x_k) \).

Now turn to see the case of Poisson-Dirichlet measures \( \text{PD}^{*}_{\alpha, \theta} \) to then study \( \nu^{sb}_{\alpha, \gamma} \).

Lemma 18. If we define \( \text{GEM}^{*}_{\alpha, \theta} \) as the size-biased dislocation measure associated with \( \text{PD}^{*}_{\alpha, \theta} \) for \( 0 < \alpha < 1 \) and \( \theta > -2\alpha \), then the first \( k \) marginals have joint density

\[
\text{GEM}^{*}_{\alpha, \theta}(x_1, \ldots, x_k) = \frac{\alpha \Gamma(2 + \theta/\alpha)}{\Gamma(1 - \alpha) \Gamma(\theta + \alpha + 1)} \left( 1 - \sum_{i=1}^{k} x_i \right)^{\theta + k \alpha} \prod_{j=1}^{k} x_j^{-\beta},
\]

where \( B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} \, dx \) is the beta function.

This is a simple \( \sigma \)-finite extension of the GEM distribution and (17) can be derived analogously to Lemma 7. Applying Proposition 17, we can get an explicit form of the size-biased dislocation measure associated with the alpha-gamma model.

Proof of Proposition 17. We start our proof from the dislocation measure associated with the alpha-gamma model. According to (5) and (16), the first \( k \) marginals of \( \nu^{sb}_{\alpha, \gamma} \) are given by

\[
\nu^{sb}_{\alpha, \gamma}(A_1 \times \ldots \times A_k) = \sum_{i_1, \ldots, i_k \geq 1 } \int_{\{s \in S^k: s_i \in A_j, j \in \{k\}\}} \frac{s_{i_1} \cdots s_{i_k}}{\prod_{j=1}^{k-1} (1 - \sum_{l=1}^{j} s_i)} \left( \gamma + (1 - \alpha - \gamma) \sum_{i \neq j} s_i s_j \right) \text{PD}^{*}_{\alpha, \alpha - \gamma}(ds)
\]

where

\[
D = \sum_{i_1, \ldots, i_k \geq 1 } \int_{\{s \in S^k: s_i \in A_j, j \in \{k\} \}} \frac{s_{i_1} \cdots s_{i_k}}{\prod_{j=1}^{k-1} (1 - \sum_{l=1}^{j} s_i)} \text{PD}^{*}_{\alpha, \alpha - \gamma}(ds)
\]

\[
E = \sum_{i_1, \ldots, i_k \geq 1 } \int_{\{s \in S^k: s_i \in A_j, j \in \{k\} \}} (1 - \sum_{u=1}^{k} s_{i_u}^2) \frac{s_{i_1} \cdots s_{i_k}}{\prod_{j=1}^{k-1} (1 - \sum_{l=1}^{j} s_i)} \text{PD}^{*}_{\alpha, \alpha - \gamma}(ds)
\]

\[
F = \sum_{i_1, \ldots, i_k \geq 1 } \int_{\{s \in S^k: s_i \in A_j, j \in \{k\} \}} \left( \sum_{u \neq \{i_1, \ldots, i_k\}} s_{i_u}^2 \right) \frac{s_{i_1} \cdots s_{i_k}}{\prod_{j=1}^{k-1} (1 - \sum_{l=1}^{j} s_i)} \text{PD}^{*}_{\alpha, \alpha - \gamma}(ds)
\]

\[
= \sum_{i_1, \ldots, i_k \geq 1 } \int_{\{s \in S^k: s_i \in A_j, j \in \{k\} \}} \frac{s_{i_{k+1}}^2}{\prod_{j=1}^{k-1} (1 - \sum_{l=1}^{j} s_i)} \text{PD}^{*}_{\alpha, \alpha - \gamma}(ds)
\]

\[
= \int_{A_1 \times \ldots \times A_k \times [0,1]} \frac{x_{k+1}}{1 - \sum_{i=1}^{k} x_i} \text{GEM}^{*}_{\alpha, \alpha - \gamma}(d(x_1, \ldots, x_{k+1})).
\]
Applying (17) to $F$ (and setting $\theta = -\alpha - \gamma$), then integrating out $x_{k+1}$, we get:

$$F = \int_{A_1 \times \ldots \times A_k} \frac{1 - \alpha}{1 + (k-1)\alpha - \gamma} \left( 1 - \sum_{i=1}^{k} x_i \right)^2 \text{GEM}^*_{\alpha, -\alpha - \gamma}(dx).$$

Summing over $D, E, F$, we obtain the formula stated in Proposition 4.

As the model related to stable trees is a special case of the alpha-gamma model when $\gamma = 1 - \alpha$, the sized-biased dislocation measure for it is

$$\nu_{\alpha, 1-\alpha}^{sb}(ds) = \gamma \text{GEM}^*_{\alpha, -1}(ds).$$

For general $(\alpha, \gamma)$, the explicit form of the dislocation measure in size-biased order, specifically the density $g_{\alpha, \gamma}$ of the first marginal of $\nu_{\alpha, \gamma}^{sb}$, yields immediately the tagged particle [4] Lévy measure associated with a fragmentation process with alpha-gamma dislocation measure.

**Corollary 19.** Let $(\Pi_{\alpha, \gamma}(t), t \geq 0)$ be an exchangeable homogeneous $\mathcal{P}_\mathbb{N}$-valued fragmentation process with dislocation measure $\nu_{\alpha, \gamma}$. Then, for the size $|\Pi_{(i)}(t)|$ of the block containing $i \geq 1$, the process $\xi(i) = -\log |\Pi_{(i)}(t)|$, $t \geq 0$, is a pure-jump subordinator with Lévy measure

$$\Lambda_{\alpha, \gamma}(dx) = e^{-x} g_{\alpha, \gamma}(e^{-x}) dx = \frac{\alpha \Gamma(1 - \gamma/\alpha)}{\Gamma(1 - \alpha) \Gamma(1 - \gamma)} \left( 1 - e^{-x} \right)^{1-\gamma} \left( e^{-x} \right)^{1-\alpha} \times \left( \gamma + (1 - \alpha - \gamma) \left( 2e^{-x}(1 - e^{-x}) + \frac{\alpha - \gamma}{1 - \gamma}(1 - e^{-x})^2 \right) \right) dx.$$

### 3.6 Convergence of alpha-gamma trees to self-similar CRTs

In this subsection, we will prove that the delabelled alpha-gamma trees $T_n$, represented as $\mathbb{R}$-trees with unit edge lengths and suitably rescaled converge to CRTs as $n$ tends to infinity.

**Lemma 20.** If $(\tilde{T}_n^0)_{n \geq 1}$ are strongly sampling consistent discrete fragmentation trees associated with dislocation measure $\nu_{\alpha, -\alpha - \gamma}$, then

$$\frac{\tilde{T}_n^0}{n^{\gamma}} \rightarrow T^{\alpha, \gamma}$$

in the Gromov-Hausdorff sense, in probability as $n \rightarrow \infty$.

**Proof.** Theorem 2 in [16] says that a strongly sampling consistent family of discrete fragmentation trees $(T_n^0)_{n \geq 1}$ converges in probability to a CRT

$$\frac{\tilde{T}_n^0}{n^{\gamma} \ell(n)^{1-\gamma/\nu}} \rightarrow T_{(\gamma, \nu)}$$

for the Gromov-Hausdorff metric if the dislocation measure $\nu$ satisfies following two conditions:

$$\nu(s_1 \leq 1 - \varepsilon) = \varepsilon^{-\gamma} \ell(1/\varepsilon); \quad (18)$$

$$\int_{S^1} \sum_{i \geq 2} s_i \ln s_i |\nu(ds) < \infty, \quad (19)$$

where $\rho$ is some positive real number, $\gamma_{\nu} \in (0, 1)$, and $x \mapsto \ell(x)$ is slowly varying as $x \rightarrow \infty$.

By virtue of (19) in [16], we know that (18) is equivalent to

$$\Lambda((x, \infty)) = x^{-\gamma} \ell(1/x), \quad \text{as } x \downarrow 0,$$
where $\Lambda$ is the Lévy measure of the tagged particle subordinator as in Corollary \[19\]. So, the dislocation measure $\nu_{\alpha,\gamma}$ satisfies (18) with $\ell(x) \rightarrow \gamma \alpha \Gamma(1-\gamma/\alpha) / \Gamma(1-\alpha) \Gamma(2-\gamma)$ and $\gamma_{\alpha,\gamma} = \gamma$. Notice that

$$\int_{S^{+}} \sum_{i \geq 2} s_i |\ln s_i|^p \nu_{\alpha,\gamma}(ds) \leq \int_{0}^{\infty} x^p \Lambda_{\alpha,\gamma}(dx).$$

As $x \to \infty$, $\Lambda_{\alpha,\gamma}$ decays exponentially, so $\nu_{\alpha,\gamma}$ satisfies condition (19). This completes the proof. \hfill \Box

**Proof of Corollary \[3\]**. The splitting rules of $T_n^\circ$ are the same as those of $\tilde{T}_n^\circ$, which leads to the identity in distribution for the whole trees. The preceding lemma yields convergence in distribution for $T_n^\circ$. \hfill \Box

### 4 Limiting results for labelled alpha-gamma trees

In this section we suppose $0 < \alpha < 1$ and $0 < \gamma \leq \alpha$. In the boundary case $\gamma = 0$ trees grow logarithmically and do not possess non-degenerate scaling limits; for $\alpha = 1$ the study in Section \[2\] can be refined to give results analogous to the ones below, but with degenerate tree shapes.

#### 4.1 The scaling limits of reduced alpha-gamma trees

For $\tau$ a rooted $\mathbb{R}$-tree and $x_1, \ldots, x_n \in \tau$, we call $R(\tau, x_1, \ldots, x_n) = \bigcup_{i=1}^{n}[[\rho, x_i]]$ the reduced subtree associated with $\tau, x_1, \ldots, x_n$, where $\rho$ is the root of $\tau$.

As a fragmentation CRT, the limiting CRT ($T^\circ\alpha,\gamma, \mu$) is naturally equipped with a mass measure $\mu$ and contains subtrees $\tilde{R}_k, k \geq 1$ spanned by $k$ leaves chosen independently according to $\mu$. Denote the discrete tree without edge lengths by $\tilde{T}_n$ — it has *exchangeable* leaf labels. Then $\tilde{R}_n$ is the almost sure scaling limit of the reduced trees $R(\tilde{T}_n, [k])$, by Proposition 7 in [16].

On the other hand, if we denote by $T_n$ the (non-exchangeably) labelled trees obtained via the alpha-gamma growth rules, the above result will not apply, but, similarly to the result for the alpha model shown in Proposition 18 in [16], we can still establish a.s. convergence of the reduced subtrees in the alpha-gamma model as stated in Theorem [5] and the convergence result can be strengthened as follows.

**Proposition 21.** In the setting of Theorem [3]

$$(n^{-\gamma} R(T_n, [k]), n^{-1} W_{n,k}) \rightarrow (R_k, W_k) \quad \text{a.s. as } n \to \infty,$$

in the sense of Gromov-Hausdorff convergence, where $W_{n,k}$ is the total number of leaves in subtrees of $T_n \setminus R(T_n, [k])$ that are linked to the present branch points of $R(T_n, [k])$.

**Proof of Theorem [2] and Proposition [21]**. Actually, the labelled discrete tree $R(T_n, [k])$ with edge lengths removed is $T_k$ for all $n$. Thus, it suffices to prove the convergence of its total length and of its edge length proportions.

Let us consider a first urn model, cf. [11], where at level $n$ the urn contains a black ball for each leaf in a subtree that is directly connected to a branch point of $R(T_n, [k])$, and a white ball for each leaf in one of the remaining subtrees connected to the edges of $R(T_n, [k])$. Suppose that the balls are labelled like the leaves they represent. If the urn then contains $W_{n,k} = m$ white balls and $n - k - m$ black balls, the induced partition of $\{k+1, \ldots, n\}$ has probability function

$$p(m, n - k - m) = \frac{\Gamma(n - m - \alpha - w) \Gamma(w + m) \Gamma(k - \alpha)}{\Gamma(k - \alpha - w) \Gamma(w) \Gamma(n - \alpha)} = \frac{B(n - m - \alpha - w, w + m)}{B(k - \alpha - w, w)},$$

where $w = k(1 - \alpha) + \ell \gamma$ is the total weight on the $k$ leaf edges and $\ell$ other edges of $T_k$. As $n \to \infty$, the urn is such that $W_{n,k}/n \rightarrow W_k$ a.s., where $W_k \sim \text{beta}((k - 1)\alpha - l\gamma, k(1 - \alpha) + l\gamma)$. 

20
We will partition the white balls further. Extending the notions of spine, spinal subtrees and spinal bushes from Proposition \(10\) \((k = 1)\), we call, for \(k \geq 2\), skeleton the tree \(S(T_n, [k])\) of \(T_n\) spanned by the root and leaves \([k]\) including the degree-2 vertices, for each such degree-2 vertex \(v \in S(T_n, [k])\), we consider the skeleton subtrees \(S_{v,j}^{sk}\) that we join together into a skeletal bush \(S_{v,j}^{sk}\). Note that the total length \(L_k^{(n)}\) of the skeleton \(S(T_n, [k])\) will increase by 1 if leaf \(n + 1\) in \(T_{n+1}\) is added to any of the edges of \(S(T_n, [k])\); also, \(L_k^{(n)}\) is equal to the number of skeletal bushes (denoted by \(\overline{K}_n\)) plus the original total length of \(k + \ell\) of \(T_k\). Hence, as \(n \to \infty\)

\[
\frac{L_k^{(n)}}{n^\gamma} \sim \frac{\overline{K}_n}{W_{n,k}^\gamma} \left(\frac{W_{n,k}}{n}\right)^\gamma \sim \frac{\overline{K}_n}{W_{n,k}^\gamma} W_k^\gamma.
\]

(20)

The partition of leaves (associated with white balls), where each skeletal bushes gives rise to a block, follows the dynamics of a Chinese Restaurant Process with \((\gamma, w)\)-seating plan: given that the number of white balls in the first urn is \(m\) and that there are \(K_m := \overline{K}_n\) skeletal bushes on the edges of \(S(T_n, [k])\) with \(n_i\) leaves on the \(i\)th bush, the next leaf associated with a white ball will be inserted into any particular bush with \(n_i\) leaves with probability proportional to \(n_i - \gamma\) and will create a new bush with probability proportional to \(w + K_m^\gamma\). Hence, the EPPF of this partition of the white balls is

\[
p_{\gamma,w}(n_1, \ldots, n_{K_m}) = \frac{\gamma^{K_m-1} \Gamma(K_m + w/\gamma) \Gamma(1 + w) \Gamma(m + w)}{\Gamma(1 + w/\gamma) \Gamma(m + w)} \prod_{i=1}^{K_m} \Gamma(n_i).
\]

Applying Lemma 8 in connection with (20), we get the probability density of \(L_k/W_k^\gamma\) as specified.

Finally, we set up another urn model that is updated whenever a new skeletal bush is created. This model records the edge lengths of \(R(T_n, [k])\). The alpha-gamma growth rules assign weights \(1 - \alpha + (n_i - 1)\gamma\) to leaf edges of \(R(T_n, [k])\) and weights \(n_i\gamma\) to other edges of length \(n_i\), and each new skeletal bush makes one of the weights increase by \(\gamma\). Hence, the conditional probability that the length of each edge is \((n_1, \ldots, n_{k+\ell})\) at stage \(n\) is that

\[
\frac{\prod_{i=1}^{k} \Gamma_1 - \alpha(n_i) \prod_{i=k+1}^{k+\ell} \Gamma(n_i)}{\Gamma_{k\alpha + \ell\gamma}(n - k)}.
\]

Then \(D_k^{(n)}\) converge a.s. to the Dirichlet limit as specified. Moreover, \(L_k^{(n)} D_k^{(n)} \to L_k D_k\) a.s., and it is easily seen that this implies convergence in the Gromov-Hausdorff sense.

The above argument actually gives us the conditional distribution of \(L_k/W_k^\gamma\) given \(T_k\) and \(W_k\), which does not depend on \(W_k\). Similarly, the conditional distribution of \(D_k\) given \(T_k\), \(W_k\) and \(L_k\) does not depend on \(W_k\) and \(L_k\). Hence, the conditional independence of \(W_k\), \(L_k/W_k^\gamma\) and \(D_k\) given \(T_k\) follows.

4.2 Further limiting results

Alpha-gamma trees not only have edge weights but also vertex weights, and the latter are in correspondence with the vertex degrees. We can get a result on the limiting ratio between the degree of each vertex and the total number of leaves.

**Proposition 22.** Let \((c_1 + 1, \ldots, c_\ell + 1)\) be the degree of each vertex in \(T_k\), listed by depth first search. The ratio between the degrees in \(T_n\) of these vertices and \(n^\alpha\) will converge to

\[
C_k = (C_k,1, \ldots, C_k,\ell) = \overline{W}_k M_k D_k',
\]

where \(D_k' \sim \text{Dirichlet}((c_1 - 1 - \gamma/\alpha, \ldots, c_\ell - 1 - \gamma/\alpha))\) and \(M_k\) are conditionally independent of \(W_k\) given \(T_k\), where \(\overline{W}_k = 1 - W_k\), and \(M_k\) has density

\[
\frac{\Gamma(\overline{w} + 1)}{\Gamma(\overline{w}/\alpha + 1)} s^{\overline{w}/\alpha} g_\alpha(s), \quad s \in (0, \infty),
\]
\( w = (k - 1)\alpha - \ell\gamma \) is total branch point weight in \( T_k \) and \( g_\alpha(s) \) is the Mittag-Leffler density.

**Proof.** Recall the first urn model in the preceding proof which assigns colour black to leaves attached in subtrees of branch points of \( T_k \). We will partition the black balls further. The partition of leaves (associated with black balls), where each subtree \( S_{vk} \) of a branch point \( v \in R(T_n, [k]) \) gives rise to a block, follows the dynamics of a Chinese Restaurant Process with \((\alpha, w)\)-seating plan. Hence, the total degree \( C_k^{\text{tot}}(n)/W_{n,k} \to M_k \) a.s., where \( C_k^{\text{tot}}(n) \) is the sum of degrees in \( T_n \) of the branch points of \( T_k \), and \( W_{n,k} = n - k - W_{n,k} \) is the total number of leaves of \( T_n \) that are in subtrees directly connected to the branch points of \( T_k \).

Similarly to the discussion of edge length proportions, we now see that the sequence of degree proportions will converge a.s. to the Dirichlet limit as specified. Since \( 1 - W_k \) is the a.s. limiting proportion of leaves in subtrees connected to the vertices of \( T_k \).

Given an alpha-gamma tree \( T_n \), if we decompose along the spine that connects the root to leaf 1, we will find the leaf numbers of subtrees connected to the spine is a Chinese restaurant partition of \( \{2, \ldots, n\} \) with parameters \((\alpha, 1 - \alpha)\). Applying Lemma \( 7 \) we get following result.

**Proposition 23.** Let \( (T_n, n \geq 1) \) be alpha-gamma trees. Denote by \((P_1, P_2, \ldots)\) the limiting frequencies of the leaf numbers of each subtree of the spine connecting the root to leaf 1 in the order of appearance. These can be represented as

\[(P_1, P_2, \ldots) = (W_1, W_1 W_2, W_1 W_2 W_3, \ldots)\]

where the \( W_i \) are independent, \( W_i \) has beta\((1 - \alpha, 1 + (i - 1)\alpha)\) distribution, and \( \overline{W}_i = 1 - W_i \).

Observe that this result does not depend on \( \gamma \). This observation also follows from Proposition \( 6 \) because colouring (iv)\textsuperscript{col} and crushing (cr) do not affect the partition of leaf labels according to subtrees of the spine.

**References**

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

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

[3] D. Aldous. Probability distributions on cladograms. In Random discrete structures (Minneapolis, MN, 1993), volume 76 of IMA Vol. Math. Appl., pages 1–18. Springer, New York, 1996.

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

[5] J. Bertoin. Self-similar fragmentations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(3):319–340, 2002.

[6] J. Bertoin. *Random fragmentation and coagulation processes*, volume 102 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2006.

[7] T. Duquesne and J.-F. Le Gall. Random trees, Lévy processes and spatial branching processes. *Astérisque*, (281):vi+147, 2002.

[8] T. Duquesne and J.-F. Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields*, 131(4):553–603, 2005.
[9] S. N. Evans, J. Pitman, and A. Winter. Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Related Fields*, 134(1):81–126, 2006.

[10] S. N. Evans and A. Winter. Subtree prune and regraft: a reversible real tree-valued Markov process. *Ann. Probab.*, 34(3):918–961, 2006.

[11] W. Feller. *An introduction to probability theory and its applications. Vol. I*. Third edition. John Wiley & Sons Inc., New York, 1968.

[12] D. J. Ford. Probabilities on cladograms: introduction to the alpha model. 2005. Preprint, arXiv:math.PR/0511246.

[13] A. Greven, P. Pfaffelhuber, and A. Winter. Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Preprint, arXiv:math.PR/0609801v2*, 2006.

[14] R. C. Griffiths. Allele frequencies with genic selection. *J. Math. Biol.*, 17(1):1–10, 1983.

[15] B. Haas and G. Miermont. The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electron. J. Probab.*, 9:no. 4, 57–97 (electronic), 2004.

[16] B. Haas, G. Miermont, J. Pitman, and M. Winkel. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *Preprint, arXiv:math.PR/0604350*, 2006, to appear in Annals of Probability.

[17] B. Haas, J. Pitman, and M. Winkel. Spinal partitions and invariance under re-rooting of continuum random trees. *Preprint, arXiv:0705.3602*, 2007.

[18] P. Marchal. A note on the fragmentation of a stable tree. *Work in progress*, 2008.

[19] P. McCullagh, J. Pitman, and M. Winkel. Gibbs fragmentation trees. *Preprint, arXiv:0704.0945*, 2007, to appear in Bernoulli.

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

[21] G. Miermont. Self-similar fragmentations derived from the stable tree. II. Splitting at nodes. *Probab. Theory Related Fields*, 131(3):341–375, 2005.

[22] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002.

[23] J. Pitman and M. Winkel. Regenerative tree growth: binary self-similar continuum random trees and Poisson-Dirichlet compositions. *Preprint, arXiv:0803.3098*, 2008.