Electronic Journal of Probability

Vol. 14 (2009), Paper no. 15, pages 400–430.

Journal URL
http://www.math.washington.edu/~ejpecp/

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

Bo Chen*
Department of Statistics
University of Oxford
1 South Parks Road
Oxford OX1 3TG, UK.
chen@stats.ox.ac.uk

Daniel Ford
Google Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043, USA.
dford@math.stanford.edu

Matthias Winkel
Department of Statistics
University of Oxford
1 South Parks Road
Oxford OX1 3TG, UK.
winkel@stats.ox.ac.uk
http://www.stats.ox.ac.uk/~winkel/

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.

Key words: Alpha-gamma tree, splitting rule, sampling consistency, self-similar fragmentation,

*supported by the K C Wong Education Foundation dislocation measure, continuum random tree, $\mathbb{R}$-tree, Markov branching model.

AMS 2000 Subject Classification: Primary 60J80.

Submitted to EJP on July 3, 2008, final version accepted January 20, 2009.

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 $\mathbb{T}_n$ 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 $\mathbb{T}_n$ of random trees T_n^* are determined by distributions of the delabelled tree T_n° on the space $\mathbb{T}_n^\circ$ of *unlabelled trees* and conditional label distributions, e.g. *exchangeable* labels. A sequence $(T_n^\circ, n \geq 1)$ 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 $T_{n_i}^\circ$, $1 \leq i \leq k$. The distributions of the sizes in the first branching of T_n°, $n \geq 2$, are denoted by

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

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

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 $\beta > -2$:

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

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. $q_{-2}^{\text{Aldous}}(n-1, 1) = 1$, $n \geq 2$.

For exchangeably labelled Markov branching models $(T_n, n \geq 1)$ it is convenient to set

$$p(n_1, \ldots, n_k) := \frac{m_1! \ldots m_n!}{\binom{n}{n_1, \ldots, n_k}} q((n_1, \ldots, n_k)^\downarrow), \quad n_j \geq 1, j \in [k]; k \geq 2: \; n = n_1 + \ldots + n_k, \tag{1}$$

where $(n_1, \ldots, n_k)^\downarrow$ is the decreasing rearrangement and m_r the number of rs 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 [20] that Aldous's beta-splitting models for $\beta > -2$ are the only *binary* Markov branching models for which the EPPF is of Gibbs type

$$p_{-1-\alpha}^{\text{Aldous}}(n_1, n_2) = \frac{w_{n_1} w_{n_2}}{Z_{n_1+n_2}}, \quad n_1 \geq 1, n_2 \geq 1, \qquad \text{in particular } w_n = \frac{\Gamma(n-\alpha)}{\Gamma(1-\alpha)},$$

and that the *multifurcating* Gibbs models are an *extended* Ewens-Pitman two-parameter family of random partitions, $0 \leq \alpha \leq 1$, $\theta \geq -2\alpha$, or $-\infty \leq \alpha < 0$, $\theta = -m\alpha$ for some integer $m \geq 2$,

$$p_{\alpha,\theta}^{\text{PD}^*}(n_1, \ldots, n_k) = \frac{a_k}{Z_n} \prod_{j=1}^k w_{n_j}, \quad \text{where } w_n = \frac{\Gamma(n-\alpha)}{\Gamma(1-\alpha)} \text{ and } a_k = \alpha^{k-2} \frac{\Gamma(k+\theta/\alpha)}{\Gamma(2+\theta/\alpha)}, \tag{2}$$

boundary cases by continuity (cf. p. 404), including Aldous's binary models for $\theta = -2\alpha$. Ford [12] introduced a different one-parameter *binary* model, the *alpha model* for $0 \le \alpha \le 1$, using simple sequential growth rules starting from the unique elements $T_1 \in \mathbb{T}_1$ and $T_2 \in \mathbb{T}_2$:

- (i)$^{\mathrm{F}}$ given T_n for $n \ge 2$, assign a weight $1-\alpha$ to each of the n edges adjacent to a leaf, and a weight α to each of the $n-1$ other edges;
- (ii)$^{\mathrm{F}}$ select at random with probabilities proportional to the weights assigned by step (i)$^{\mathrm{F}}$, an edge of T_n, say $a_n \to c_n$ directed away from the ROOT;
- (iii)$^{\mathrm{F}}$ to create T_{n+1} from T_n, replace $a_n \to c_n$ by three edges $a_n \to b_n$, $b_n \to c_n$ and $b_n \to n+1$ 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; 24].

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* depending on two parameters $0 \le \alpha \le 1$ and $0 \le \gamma \le \alpha$ as follows, cf. Figure 1:

- (i) given T_n for $n \ge 2$, assign a weight $1-\alpha$ to each of the n edges adjacent to a leaf, a weight γ to each of the other edges, and a weight $(k-1)\alpha - \gamma$ to each vertex of degree $k+1 \ge 3$; this distributes a total weight of $n - \alpha$;
- (ii) select at random with probabilities proportional to the weights assigned by step (i),
  - an edge of T_n, say $a_n \to c_n$ 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 $a_n \to c_n$ was selected, replace it by three edges $a_n \to b_n$, $b_n \to c_n$ and $b_n \to n+1$ 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 $v_n \to n+1$ to a new leaf labelled $n+1$.

We call this model the *alpha-gamma model*. It contains the binary alpha model for $\gamma = \alpha$. We show here that the cases $\gamma = 1-\alpha$, $1/2 \le \alpha \le 1$, and $\alpha = \gamma = 0$ form the intersection with the extended Ewens-Pitman-type two-parameter family of models (2). The growth rules for $\gamma = 1-\alpha$, when all edges have the same weight, was studied recently by Marchal [19]. It is related to the stable tree of Duquesne and Le Gall [7], see also [21] and Section 3.4 here.

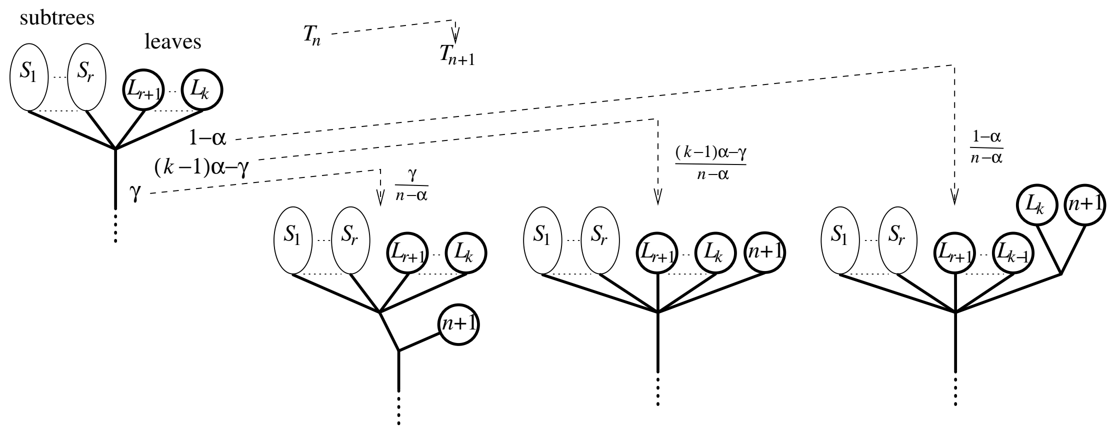


Figure 1: Sequential growth rule: displayed is one branch point of T_n with degree $k+1$, hence vertex weight $(k-1)\alpha-\gamma$, with $k-r$ leaves $L_{r+1},\dots,L_k \in [n]$ and r bigger subtrees $S_1,\dots,S_r$ attached to it; all edges also carry weights, weight $1-\alpha$ and γ are displayed here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

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

(a) *the delabelled trees T_n°, $n\ge 1$, have the Markov branching property. The splitting rules are*

$$q^{\mathrm{seq}}_{\alpha,\gamma}(n_1,\dots,n_k) \quad\propto\quad \left(\gamma+(1-\alpha-\gamma)\frac{1}{n(n-1)}\sum_{i\neq j} n_i n_j\right) q^{\mathrm{PD}^*}_{\alpha,-\alpha-\gamma}(n_1,\dots,n_k), \qquad (3)$$

in the case $0\le\alpha<1$, where $q^{\mathrm{PD}^*}_{\alpha,-\alpha-\gamma}$ is the splitting rule associated via (1) with $p^{\mathrm{PD}^*}_{\alpha,-\alpha-\gamma}$, the Ewens-Pitman-type EPPF given in (2), and LHS $\propto$ RHS means equality up to a multiplicative constant depending on n and (α,γ) that makes the LHS a probability function;

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

The normalisation constants in (2) and (3) can be expressed in terms of Gamma functions, see Section 2.4. The case $\alpha=1$ is discussed in Section 3.2.

For any function $(n_1,\dots,n_k)\mapsto q(n_1,\dots,n_k)$ that is a probability function for all fixed $n=n_1+\dots+n_k$, $n\ge 2$, we can construct a Markov branching model $(T_n^\circ, n\ge 1)$. A condition called *sampling consistency* [3] is to require that the tree $T_{n,-1}^\circ$ constructed from T_n° 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}°, for all $n\ge 2$. This is appealing for applications with incomplete observations. It was shown in [16] that all sampling consistent splitting rules admit an integral representation (c,ν) for an erosion coefficient $c\ge 0$ and a dislocation measure ν on $\mathscr{S}^\downarrow = \{s=(s_i)_{i\ge 1} : s_1\ge s_2\ge\dots\ge 0, s_1+s_2+\dots\le 1\}$ with $\nu(\{(1,0,0,\dots)\})=0$ and $\int_{\mathscr{S}^\downarrow}(1-s_1)\nu(ds)<\infty$ as in Bertoin's continuous-time fragmentation theory [4; 5; 6]. In the most relevant case for us when $c=0$ and $\nu(\{s\in\mathscr{S}^\downarrow : s_1+s_2+\dots<1\})=0$, this representation is

$$p(n_1,\dots,n_k) = \frac{1}{\widetilde{Z}_n}\int_{\mathscr{S}^\downarrow} \sum_{\substack{i_1,\dots,i_k\ge 1\\ \text{distinct}}} \prod_{j=1}^k s_{i_j}^{n_j}\,\nu(ds), \quad n_j\ge 1, j\in[k]; k\ge 2:\ n=n_1+\dots+n_k, \qquad (4)$$

where $\widetilde{Z}_n = \int_{\mathscr{S}^\downarrow}(1-\sum_{i\ge 1} s_i^n)\nu(ds)$, $n \ge 2$, are the normalisation constants. The measure ν is unique up to a multiplicative constant. In particular, it can be shown [21; 17] that for the Ewens-Pitman EPPFs $p_{\alpha,\theta}^{\mathrm{PD}^*}$ we obtain $\nu = \mathrm{PD}^*_{\alpha,\theta}(ds)$ 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

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

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

$$\int_{\mathscr{S}^\downarrow} f(s)\mathrm{PD}^*_{\alpha,-2\alpha}(ds) = \int_{1/2}^1 f(x, 1-x, 0, 0, \ldots)x^{-\alpha-1}(1-x)^{-\alpha-1}dx.$$

Note that $\nu = \mathrm{PD}^*_{\alpha,\theta}$ is infinite but σ-finite with $\int_{\mathscr{S}^\downarrow}(1-s_1)\nu(ds) < \infty$ for $-2\alpha \le \theta \le -\alpha$. This is the relevant range for this paper. For $\theta > -\alpha$, the measure $\mathrm{PD}^*_{\alpha,\theta}$ just defined is a multiple of the usual Poisson-Dirichlet probability measure $\mathrm{PD}_{\alpha,\theta}$ on $\mathscr{S}^\downarrow$, so for the integral representation of $p_{\alpha,\theta}^{\mathrm{PD}^*}$ we could also take $\nu = \mathrm{PD}_{\alpha,\theta}$ in this case, and this is also an appropriate choice for the two cases $\alpha = 0$ and $m \ge 3$; the case $\alpha = 1$ is degenerate $q_{\alpha,\theta}^{\mathrm{PD}^*}(1, 1, \ldots, 1) = 1$ (for all θ) and can be associated with $\nu = \mathrm{PD}^*_{1,\theta} = \delta_{(0,0,\ldots)}$, see [20].

Theorem 2. *The alpha-gamma-splitting rules $q_{\alpha,\gamma}^{\mathrm{seq}}$ are sampling consistent. For $0 \le \alpha < 1$ and $0 \le \gamma \le \alpha$ we have no erosion ($c = 0$) and the measure in the integral representation (4) can be chosen as*

$$\nu_{\alpha,\gamma}(ds) = \left(\gamma + (1-\alpha-\gamma)\sum_{i\ne j} s_i s_j\right)\mathrm{PD}^*_{\alpha,-\alpha-\gamma}(ds). \tag{5}$$

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 $\sum_{i\ne j} s_i s_j$ to model genic 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 3.3 and 3.6 for details. This theory applies here to yield:

Corollary 3. *Let $(T_n^\circ, n \ge 1)$ be delabelled alpha-gamma trees, represented as discrete $\mathbb{R}$-trees with unit edge lengths, for some $0 < \alpha < 1$ and $0 < \gamma \le \alpha$. Then*

$$\frac{T_n^\circ}{n^\gamma} \to \mathscr{T}^{\alpha,\gamma} \qquad \textit{in distribution for the Gromov-Hausdorff topology,}$$

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

We observe that every dislocation measure ν on $\mathscr{S}^\downarrow$ gives rise to a measure ν^{sb} on the space of summable sequences under which fragment sizes are in a size-biased random order, just as the $\mathrm{GEM}_{\alpha,\theta}$ distribution can be defined as the distribution of a $\mathrm{PD}_{\alpha,\theta}$ sequence re-arranged in size-biased random order [23]. We similarly define $\mathrm{GEM}^*_{\alpha,\theta}$ from $\mathrm{PD}^*_{\alpha,\theta}$. One of the advantages of size-biased versions is that, as for $\mathrm{GEM}_{\alpha,\theta}$, we can calculate marginal distributions explicitly.

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

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

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 $\mathscr{R}_k$ of trees with k leaves labelled $[k]$, $k \geq 1$. See however [24] for developments in the binary case $\gamma = \alpha$ on how to embed $\mathscr{R}_k$, $k \geq 1$, in a CRT $\mathscr{T}^{\alpha,\alpha}$. 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 as edge lengths. Then*

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

for some discrete tree $\mathscr{R}_k$ 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

- *$W_k \sim \mathrm{beta}(k(1-\alpha) + \ell\gamma, (k-1)\alpha - \ell\gamma)$, where $\mathrm{beta}(a, b)$ is the beta distribution with density $B(a,b)^{-1}x^{a-1}(1-x)^{b-1}1_{(0,1)}(x)$;*
- *L_k with density $\dfrac{\Gamma(1+k(1-\alpha)+\ell\gamma)}{\Gamma(1+\ell+k(1-\alpha)/\gamma)}s^{\ell+k(1-\alpha)/\gamma}g_\gamma(s)$, where g_γ is the Mittag-Leffler density, the density of $\sigma_1^{-\gamma}$ for a subordinator σ with Laplace exponent λ^γ;*
- *$D_k \sim \mathrm{Dirichlet}((1-\alpha)/\gamma, \ldots, (1-\alpha)/\gamma, 1, \ldots, 1)$, where $\mathrm{Dirichlet}(a_1, \ldots, a_m)$ 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 $x_1^{a_1-1} \ldots x_{m-1}^{a_{m-1}-1}(1 - x_1 - \ldots - x_{m-1})^{a_m-1}$; 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 [18] (see Section 4.2).*

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 4.2.

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)$^{\mathrm{F}}$-(iii)$^{\mathrm{F}}$, so that in T_n^{col} each edge except those adjacent to leaves has either a blue or a red colour mark.

(iv)$^{\mathrm{col}}$ To turn T_{n+1} into a colour-marked tree T_{n+1}^{col}, keep the colours of T_n^{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 $(T_n^{\mathrm{col}}, n \geq 1)$ has been grown according to (i)$^{\mathrm{F}}$-(iii)$^{\mathrm{F}}$ and (iv)$^{\mathrm{col}}$, 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 $(\widetilde{T}_n, n \geq 1)$;

Proposition 6. *Let* $(\widetilde{T}_n, n \geq 1)$ *be a sequence of trees according to growth rules* (i)$^{\mathrm{F}}$-(iii)$^{\mathrm{F}}$,(iv)$^{\mathrm{col}}$ *and crushing rule* (cr). *Then* $(\widetilde{T}_n, n \geq 1)$ *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 $\mathscr{P}_B$ be the *set of partitions of* B into disjoint non-empty subsets called *blocks*. Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, which supports a $\mathscr{P}_B$-valued random partition Π_B for some *finite* $B \subset \mathbb{N}$. If the probability function of Π_B only depends on its block sizes, we call it *exchangeable*. Then

$$\mathbb{P}(\Pi_B = \{A_1, \ldots, A_k\}) = p(\#A_1, \ldots, \#A_k) \qquad \text{for each partition } \pi = \{A_1, \ldots, A_k\} \in \mathscr{P}_B,$$

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 Π_B. Alternatively, a random partition Π_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 $\pi \in \mathscr{P}_B$ is *finer than* $\pi' \in \mathscr{P}_B$, and write $\pi \preceq \pi'$, if any block of π is included in some block of π'. This defines a partial order $\preceq$ on $\mathscr{P}_B$. A process or a sequence with values in $\mathscr{P}_B$ 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 $\mathbf{t}$ is a collection of subsets of B with an additional member called the ROOT such that

- we have $B \in \mathbf{t}$; we call B the *common ancestor* of $\mathbf{t}$;
- we have $\{i\} \in \mathbf{t}$ for all $i \in B$; we call $\{i\}$ a *leaf* of $\mathbf{t}$;
- for all $A \in \mathbf{t}$ and $C \in \mathbf{t}$, we have either $A \cap C = \varnothing$, 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 $D \in \mathbf{t}$ 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 $C \to A$. If we equip $\mathbf{t}$ with the parent-child relation and also ROOT $\to B$, then $\mathbf{t}$ is a rooted connected acyclic graph, i.e.

a combinatorial tree. We denote the space of such trees $\mathbf{t}$ by $\mathbb{T}_B$ and also $\mathbb{T}_n = \mathbb{T}_{[n]}$. For $\mathbf{t} \in \mathbb{T}_B$ and $A \in \mathbf{t}$, the rooted subtree $\mathbf{s}_A$ of $\mathbf{t}$ with common ancestor A is given by $\mathbf{s}_A = \{\textsc{root}\} \cup \{C \in \mathbf{t} : C \subseteq A\} \in \mathbb{T}_A$. In particular, we consider the *subtrees* $\mathbf{s}_j = \mathbf{s}_{A_j}$ *of the common ancestor* B *of* $\mathbf{t}$, i.e. the subtrees whose common ancestors A_j, $j \in [k]$, are the children of B. In other words, $\mathbf{s}_1, \dots, \mathbf{s}_k$ are the rooted connected components of $\mathbf{t} \setminus \{B\}$.

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

Definition 1. Let $B \subset \mathbb{N}$ with $\#B = n$ and $\mathbf{t} \in \mathbb{T}_B$. We associate the relabelled tree

$$\mathbf{t}^\sigma = \{\textsc{root}\} \cup \{\sigma(A) : A \in \mathbf{t}\} \in \mathbb{T}_n,$$

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

$$\mathbf{t}^\circ = \{\mathbf{t}^\sigma | \sigma : B \to [n] \text{ bijection}\} \subset \mathbb{T}_n.$$

We denote by $\mathbb{T}_n^\circ = \{\mathbf{t}^\circ : \mathbf{t} \in \mathbb{T}_n\} = \{\mathbf{t}^\circ : \mathbf{t} \in \mathbb{T}_B\}$ 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 $\mathbf{t} \in \mathbb{T}_n$ and the numbers of leaves in these subtrees are invariants of the equivalence class $\mathbf{t}^\circ \subset \mathbb{T}_n$. If $\mathbf{t}^\circ \in \mathbb{T}_n^\circ$ has subtrees $\mathbf{s}_1^\circ, \dots, \mathbf{s}_k^\circ$ with $n_1 \geq \ldots \geq n_k \geq 1$ leaves, we say that $\mathbf{t}^\circ$ is formed by *joining together* $\mathbf{s}_1^\circ, \dots, \mathbf{s}_k^\circ$, denoted by $\mathbf{t}^\circ = \mathbf{s}_1^\circ * \ldots * \mathbf{s}_k^\circ$. We call the *composition* $(n_1, \dots, n_k)$ of n the *first split* of $\mathbf{t}_n^\circ$.

With this notation and terminology, a sequence of random trees $T_n^\circ \in \mathbb{T}_n^\circ$, $n \geq 1$, has the *Markov branching property* if, for all $n \geq 2$, the tree T_n° has the same distribution as $S_1^\circ * \ldots * S_{K_n}^\circ$, 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°, $j \in [k]$, are independent and distributed as $T_{n_j}^\circ$, $j \in [k]$.

2.2 Colour-marked trees and the proof of Proposition 6

The growth rules $(\mathrm{i})^{\mathrm{F}}$-$(\mathrm{iii})^{\mathrm{F}}$ construct binary combinatorial trees T_n^{bin} with vertex set

$$V = \{\textsc{root}\} \cup [n] \cup \{b_1, \dots, b_{n-1}\}$$

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 $\{j \in [n] : b_i \to \ldots \to j\}$, 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 $(\mathrm{iv})^{\mathrm{col}}$ can be represented by

$$V^{\mathrm{col}} = \{\textsc{root}\} \cup [n] \cup \{(b_1, \chi_n(b_1)), \dots, (b_{n-1}, \chi_n(b_{n-1}))\}$$

for some $\chi_n(b_i) \in \{0, 1\}$, $i \in [n-1]$, where 0 represents red and 1 represents blue.

Proof of Proposition 6. We only need to check that the growth rules $(\mathrm{i})^{\mathrm{F}}$-$(\mathrm{iii})^{\mathrm{F}}$ and $(\mathrm{iv})^{\mathrm{col}}$ for $(T_n^{\mathrm{col}}, n \geq 1)$ imply that the uncoloured multifurcating trees $(\widetilde{T}_n, n \geq 1)$ obtained from $(T_n^{\mathrm{col}}, n \geq 1)$

via crushing (cr) satisfy the growth rules (i)-(iii). Let therefore $\mathbf{t}_{n+1}^{\mathrm{col}}$ be a tree with $\mathbb{P}(T_{n+1}^{\mathrm{col}} = \mathbf{t}_{n+1}^{\mathrm{col}}) > 0$. It is easily seen that there is a unique tree $\mathbf{t}_n^{\mathrm{col}}$, a unique insertion edge $a_n^{\mathrm{col}} \to c_n^{\mathrm{col}}$ in $\mathbf{t}_n^{\mathrm{col}}$ and, if any, a unique colour $\chi_{n+1}(c_n^{\mathrm{col}})$ to create $\mathbf{t}_{n+1}^{\mathrm{col}}$ from $\mathbf{t}_n^{\mathrm{col}}$. Denote the trees obtained from $\mathbf{t}_n^{\mathrm{col}}$ and $\mathbf{t}_{n+1}^{\mathrm{col}}$ via crushing (cr) by $\mathbf{t}_n$ and $\mathbf{t}_{n+1}$. If $\chi_{n+1}(c_n^{\mathrm{col}}) = 0$, denote by $k+1 \geq 3$ the degree of the branch point of $\mathbf{t}_n$ with which c_n^{col} is identified in the first step of the crushing (cr).

- If the insertion edge is a leaf edge ($c_n^{\mathrm{col}} = i$ for some $i \in [n]$), we obtain

$$\mathbb{P}(\widetilde{T}_{n+1} = \mathbf{t}_{n+1} | \widetilde{T}_n = \mathbf{t}_n, T_n^{\mathrm{col}} = \mathbf{t}_n^{\mathrm{col}}) = (1-\alpha)/(n-\alpha).$$

- If the insertion edge has colour blue ($\chi_n(c_n^{\mathrm{col}}) = 1$) and also $\chi_{n+1}(c_n^{\mathrm{col}}) = 1$, we obtain

$$\mathbb{P}(\widetilde{T}_{n+1} = \mathbf{t}_{n+1} | \widetilde{T}_n = \mathbf{t}_n, T_n^{\mathrm{col}} = \mathbf{t}_n^{\mathrm{col}}) = \alpha(1-c)/(n-\alpha).$$

- If the insertion edge has colour blue ($\chi_n(c_n^{\mathrm{col}}) = 1$), but $\chi_{n+1}(c_n^{\mathrm{col}}) = 0$, or if the insertion edge has colour red ($\chi_n(c_n^{\mathrm{col}}) = 0$, and then necessarily $\chi_{n+1}(c_n^{\mathrm{col}}) = 0$ also), we obtain

$$\mathbb{P}(\widetilde{T}_{n+1} = \mathbf{t}_{n+1} | \widetilde{T}_n = \mathbf{t}_n, T_n^{\mathrm{col}} = \mathbf{t}_n^{\mathrm{col}}) = (c\alpha + (k-2)\alpha)/(n-\alpha),$$

  because in addition to $a_n^{\mathrm{col}} \to c_n^{\mathrm{col}}$, there are $k-2$ other edges in $\mathbf{t}_n^{\mathrm{col}}$, where insertion and crushing also create $\mathbf{t}_{n+1}$.

Because these conditional probabilities do not depend on $\mathbf{t}_n^{\mathrm{col}}$ and have the form required, we conclude that $(\widetilde{T}_n, n \geq 1)$ 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 [23], which generates the Ewens-Pitman two-parameter family of exchangeable random partitions Π_∞ of $\mathbb{N}$. 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 jth table, customer $n+1$ will be placed at the jth table with probability $(n_j - \alpha)/(n+\theta)$, and at a new table with probability $(\theta + k\alpha)/(n+\theta)$. The parameters α and θ 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 (α, θ)*-seating plan*.

In the CRP $(\Pi_n, n \geq 1)$ with $\Pi_n \in \mathscr{P}_{[n]}$, 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 [23]). *For each pair of parameters (α, θ) subject to the constraints above, the Chinese restaurant with the (α, θ)-seating plan generates an exchangeable random partition Π_∞ of $\mathbb{N}$. The corresponding EPPF is*

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

boundary cases by continuity. The corresponding limiting frequencies of block sizes, in size-biased order of least elements, are $\mathrm{GEM}_{\alpha,\theta}$ *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* $\mathrm{beta}(1-\alpha, \theta + i\alpha)$ *distribution, and* $\overline{W}_i := 1 - W_i$*. The distribution of the associated ranked sequence of limiting frequencies is Poisson-Dirichlet* $\mathrm{PD}_{\alpha,\theta}$*.*

We also associate with the EPPF $p^{\mathrm{PD}}_{\alpha,\theta}$ the distribution $q^{\mathrm{PD}}_{\alpha,\theta}$ of block sizes in decreasing order via (1) and, because the Chinese restaurant EPPF is *not* the EPPF of a splitting rule leading to $k \geq 2$ block (we use notation $q^{\mathrm{PD}^*}_{\alpha,\theta}$ for the splitting rules induced by conditioning on $k \geq 2$ blocks), but can lead to a single block, we also set $q^{\mathrm{PD}}_{\alpha,\theta}(n) = p^{\mathrm{PD}}_{\alpha,\theta}(n)$.

The asymptotic properties of the number K_n of blocks of Π_n under the (α,θ)-seating plan depend on α: 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 [23])**.** *For* $0 < \alpha < 1$*,* $\theta > -\alpha$*, ,*

$$\frac{K_n}{n^\alpha} \to S \qquad \text{a.s. as } n \to \infty,$$

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_α *is the density of the Mittag-Leffler distribution with pth moment* $\Gamma(p+1)/\Gamma(p\alpha+1)$*.*

As an extension of the CRP, Pitman and Winkel in [24] 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 $\mathscr{C}_n$ of n is called *regenerative* if conditionally given that the first part of $\mathscr{C}_n$ is n_1, the remaining parts of $\mathscr{C}_n$ form a composition of $n-n_1$ with the same distribution as $\mathscr{C}_{n-n_1}$. Given any decrement matrix $(q^{\mathrm{dec}}(n,m), 1 \leq m \leq n)$, there is an associated sequence $\mathscr{C}_n$ of regenerative random compositions of n defined by specifying that $q^{\mathrm{dec}}(n,\cdot)$ is the distribution of the first part of $\mathscr{C}_n$. Thus for each composition $(n_1,\ldots,n_k)$ of n,

$$\mathbb{P}(\mathscr{C}_n = (n_1,\ldots,n_k)) = q^{\mathrm{dec}}(n,n_1)q^{\mathrm{dec}}(n-n_1,n_2)\ldots q^{\mathrm{dec}}(n_{k-1}+n_k,n_{k-1})q^{\mathrm{dec}}(n_k,n_k).$$

Lemma 9 (Proposition 6 (i) in [24])**.** *For each* (α,θ) *with* $0 < \alpha < 1$ *and* $\theta \geq 0$*, denote by* $\mathscr{C}_n$ *the composition of block sizes in the ordered Chinese restaurant partition with parameters* (α,θ)*. Then* $(\mathscr{C}_n, n \geq 1)$ *is regenerative, with decrement matrix*

$$q^{\mathrm{dec}}_{\alpha,\theta}(n,m) = \binom{n}{m}\frac{(n-m)\alpha+m\theta}{n}\frac{\Gamma(m-\alpha)\Gamma(n-m+\theta)}{\Gamma(1-\alpha)\Gamma(n+\theta)} \quad (1 \leq m \leq n). \tag{6}$$

2.4 The splitting rule of alpha-gamma trees and the proof of Proposition 1

Proposition 1 claims that the unlabelled alpha-gamma trees $(T_n^\circ, n \geq 1)$ 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° in Proposition 10. We will then establish the Markov branching property from a spinal decomposition result (Lemma 11) for T_n°.

Proposition 10. *Let T_n° 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° is*

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

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

In fact, we can express explicitly Z_n in (2) as follows (see formula (22) in [17])

$$Z_n = \frac{\Gamma(1+\theta/\alpha)}{\Gamma(1+\theta)}\left(1 - \frac{\Gamma(n-\alpha)\Gamma(1+\theta)}{\Gamma(1-\alpha)\Gamma(n+\theta)}\right)$$

in the first instance for $0 < \alpha < 1$ and $\theta > -\alpha$, and then by analytic continuation and by continuity to the full parameter range.

Proof. In the binary case $\gamma = \alpha$, the expression simplifies and the result follows from Ford [12], see also [16, Section 5.2]. For the remainder of the proof, let us consider the multifurcating case $\gamma < \alpha$. We start from the growth rules of the labelled alpha-gamma trees T_n. Consider the *spine* ROOT $\to v_1 \to \ldots \to v_{L_{n-1}} \to 1$ of T_n, and the *spinal subtrees* S_{ij}^{sp}, $1 \leq i \leq L_{n-1}$, $1 \leq j \leq K_{n,i}$, not containing 1 of the spinal vertices v_i, $i \in [L_{n-1}]$. By joining together the subtrees of the spinal vertex v_i we form the ith *spinal bush* $S_i^{\mathrm{sp}} = S_{i1}^{\mathrm{sp}} * \ldots * S_{iK_{n,i}}^{\mathrm{sp}}$. Suppose a bush S_i^{sp} consists of k subtrees 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 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 a subtree with n_i leaves in the first split, then the first spinal bush S_1^{sp} will have $n - n_i$ 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

$$\begin{aligned} q_{\gamma,1-\alpha}^{\mathrm{dec}}(n-1, n-n_i) &= \binom{n-1}{n-n_i}\frac{(n_i-1)\gamma + (n-n_i)(1-\alpha)}{n-1}\frac{\Gamma(n_i-\alpha)\Gamma(n-n_i-\gamma)}{\Gamma(n-\alpha)\Gamma(1-\gamma)} \\ &= \binom{n}{n-n_i}\left(\frac{n_i}{n}\gamma + \frac{n_i(n-n_i)}{n(n-1)}(1-\alpha-\gamma)\right)\frac{\Gamma(n_i-\alpha)\Gamma(n-n_i-\gamma)}{\Gamma(n-\alpha)\Gamma(1-\gamma)}. \end{aligned}$$

Next consider the probability that the first bush S_1^{sp} joins together subtrees 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 subtree with n_i leaves. The first bush has a weight of $n - n_i - \gamma$ and each subtree in it has a weight of $n_j - \alpha, j \neq i$. Consider these $k-1$ subtrees 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

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

Let m_j be the number of js 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 subtrees 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_{n_i} subtrees with n_i leaves will be

$$\frac{m_{n_i}}{m_1!\ldots m_n!}\binom{n-n_i}{n_1,\ldots,n_{i-1},n_{i+1},\ldots,n_k} p_{\alpha,-\gamma}^{\mathrm{PD}}(n_1,\ldots,n_{i-1},n_{i+1},\ldots,n_k).$$

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,

$$\begin{aligned} &\frac{m_{n_i}}{m_1!\ldots m_n!}\binom{n-n_i}{n_1,\ldots,n_{i-1},n_{i+1},\ldots,n_k} q_{\gamma,1-\alpha}^{\mathrm{dec}}(n-1,n-n_i) p_{\alpha,-\gamma}^{\mathrm{PD}}(n_1,\ldots,n_{i-1},n_{i+1},\ldots,n_k) \\ &= \quad m_{n_i}\left(\frac{n_i}{n}\gamma + \frac{n_i(n-n_i)}{n(n-1)}(1-\alpha-\gamma)\right)\frac{Z_n\Gamma(1-\alpha)}{\Gamma(n-\alpha)} q_{\alpha,-\alpha-\gamma}^{\mathrm{PD}^*}(n_1,\ldots,n_k). \end{aligned} \tag{7}$$

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 $i \in [k]$:

$$\begin{aligned} q_{\alpha,\gamma}^{\mathrm{seq}}(n_1,\ldots,n_k) &= \left(\sum_{i=1}^{k}\left(\frac{n_i}{n}\gamma + \frac{n_i(n-n_i)}{n(n-1)}(1-\alpha-\gamma)\right)\right)\frac{Z_n\Gamma(1-\alpha)}{\Gamma(n-\alpha)} q_{\alpha,-\alpha-\gamma}^{\mathrm{PD}^*}(n_1,\ldots,n_k) \\ &= \left(\gamma + (1-\alpha-\gamma)\frac{1}{n(n-1)}\sum_{i\neq j} n_i n_j\right)\frac{Z_n\Gamma(1-\alpha)}{\Gamma(n-\alpha)} q_{\alpha,-\alpha-\gamma}^{\mathrm{PD}^*}(n_1,\ldots,n_k). \end{aligned}$$

□

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}^{sp} of the spine from the ROOT to 1 and can, conversely, build T_n from S_{ij}^{sp}, for which we now introduce notation

$$T_n = \coprod_{i,j} S_{ij}^{\mathrm{sp}}.$$

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

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

- $\mathscr{C}_{n-1} = (N_1, \ldots, N_{L_{n-1}})$ *is a regenerative composition with decrement matrix* $q_{\gamma,1-\alpha}^{\mathrm{dec}}$,
- *conditionally given* $L_{n-1} = \ell$ *and* $N_i = n_i$, $i \in [\ell]$, *the sizes* $N_{i1} \geq \ldots \geq N_{iK_{n,i}} \geq 1$ *form random compositions of* n_i *with distribution* $q_{\alpha,-\gamma}^{PD}$, *independently for* $i \in [\ell]$,
- *conditionally given also* $K_{n,i} = k_i$ *and* $N_{ij} = n_{ij}$, *the trees* S_{ij}°, $j \in [k_i]$, $i \in [\ell]$, *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 $T_n^{\circ 1}$ 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}°.

The growth rules (i)-(iii) of the labelled alpha-gamma tree T_n are such that, for $0 \leq \gamma < \alpha \leq 1$

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

These selections do not depend on T_n except via $T_n^{\circ 1}$. 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 $T_n^{\circ 1}$, and also for unlabelled alpha-gamma trees such as S_{ij}°.

From these observations and the induction hypothesis, we deduce the claim for T_{n+1}°. In the multifurcating case $\gamma < \alpha$, the conditional independence of compositions $(N_{i1}, \ldots, N_{iK_{n+1,i}})$, $i \in [\ell]$, given $L_{n-1} = \ell$ and $N_i = n_i$ can be checked by explicit calculation of the conditional probability function. Similarly, the conditional independence of the trees S_{ij}° follows, because conditional probabilities such as the following factorise and do not depend on (i_0, j_0):

$$\mathbb{P}\left(S_{\bullet\bullet}^{(n+1)\circ} = \mathbf{t}_{\bullet\bullet}^{(n+1)\circ} \,\middle|\, L_n = L_{n-1} = \ell, N_{\bullet}^{(n)} = n_{\bullet}, N_{\bullet\bullet}^{(n)} = n_{\bullet\bullet}, N_{i_0 j_0}^{(n+1)} = n_{i_0 j_0} + 1 \right),$$

where $n_{\bullet} = (n_i, 1 \leq i \leq \ell)$ and $n_{\bullet\bullet} = (n_{ij}, 1 \leq i \leq \ell, 1 \leq j \leq k_i)$ etc.; superscripts ${}^{(n)}$ and ${}^{(n+1)}$ refer to the respective stage of the growth process. In the binary case $\gamma = \alpha$, the argument is simpler, because each spinal bush consists of a single tree. □

Proof of Proposition 1. (a) Firstly, the distributions of the first splits of the unlabelled alpha-gamma trees T_n° were calculated in Proposition 10, for $0 \leq \alpha < 1$ and $0 \leq \gamma \leq \alpha$.

Secondly, let $0 \leq \alpha \leq 1$ and $0 \leq \gamma \leq \alpha$. By the regenerative property of the spinal composition $\mathscr{C}_{n-1}$ and the conditional distribution of $T_n^{\circ 1}$ given $\mathscr{C}_{n-1}$ 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 $T_{n_{1j}}^{\circ}$, also independent of the remaining tree $S_{1,0} := \coprod_{i \geq 2, j} S_{ij}^{\circ}$, which, by Lemma 11, has the same distribution as T_{n-m}°.

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

(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 $\gamma \neq 1-\alpha$, let us turn to look at the distribution of T_3.

Probability: $\frac{\gamma}{2-\alpha}$ Probability: $\frac{1-\alpha}{2-\alpha}$

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

2.5 Sampling consistency and strong sampling consistency

Recall that an unlabelled Markov branching tree T_n°, $n \geq 2$ 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 $T_{n,-1}^\circ$, is distributed as T_{n-1}°. Denote by $d: \mathbb{D}_n \to \mathbb{D}_{n-1}$ the induced deletion operator on the space $\mathbb{D}_n$ of probability measures on $\mathbb{T}_n^\circ$, so that for the distribution P_n of T_n°, we define $d(P_n)$ as the distribution of $T_{n,-1}^\circ$. 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* $0 \leq \alpha \leq 1$ *and* $0 \leq \gamma \leq \alpha$ *are sampling consistent.*

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

$$\begin{aligned} q(n_1,\ldots,n_k) &= \sum_{i=1}^k \frac{(n_i+1)(m_{n_i+1}+1)}{(n+1)m_{n_i}} q((n_1,\ldots,n_i+1,\ldots,n_k)^\downarrow) \\ &\quad + \frac{m_1+1}{n+1} q(n_1,\ldots,n_k,1) + \frac{1}{n+1} q(n,1) q(n_1,\ldots,n_k) \end{aligned} \tag{8}$$

for all $n_1 \geq \ldots \geq n_k \geq 1$ with $n_1 + \ldots + n_k = n \geq 2$, where m_j is the number of n_i, $i \in [k]$, that equal j, and where q is the splitting rule of $T_n^\circ \sim P_n$. In terms of EPPFs (1), formula (8) is equivalent to

$$\left(1 - p(n,1)\right) p(n_1,\ldots,n_k) = \sum_{i=1}^k p(n_1,\ldots,n_i+1,\ldots,n_k) + p(n_1,\ldots,n_k,1). \tag{9}$$

Now according to Proposition 10, the EPPF of the alpha-gamma model with $\alpha < 1$ is

$$p_{\alpha,\gamma}^{\mathrm{seq}}(n_1,\ldots,n_k) = \frac{Z_n}{\Gamma_\alpha(n)} \left(\gamma + (1-\alpha-\gamma) \frac{1}{n(n-1)} \sum_{u \neq v} n_u n_v \right) p_{\alpha,-\alpha-\gamma}^{\mathrm{PD}^*}(n_1,\ldots,n_k), \tag{10}$$

where $\Gamma_\alpha(n) = \Gamma(n-\alpha)/\Gamma(1-\alpha)$. Therefore, we can write $p^{\mathrm{seq}}_{\alpha,\gamma}(n_1,\ldots,n_i+1,\ldots,n_k)$ using (2)

$$\begin{aligned}
&\frac{Z_{n+1}}{\Gamma_\alpha(n+1)}\left(\gamma+(1-\alpha-\gamma)\frac{1}{(n+1)n}\left(\sum_{u\neq v} n_u n_v + 2(n-n_i)\right)\right)\frac{a_k}{Z_{n+1}}\left(\prod_{j:j\neq i} w_{n_j}\right) w_{n_i+1}\\
&=\left(p^{\mathrm{seq}}_{\alpha,\gamma}(n_1,\ldots,n_k)+2(1-\alpha-\gamma)\frac{(n-1)(n-n_i)-\sum_{u\neq v} n_u n_v}{(n+1)n(n-1)}\frac{Z_n}{\Gamma_\alpha(n)}p^{\mathrm{PD}^*}_{\alpha,-\alpha-\gamma}(n_1,\ldots,n_k)\right)\\
&\qquad\times\frac{n_i-\alpha}{n-\alpha}
\end{aligned}$$

and $p^{\mathrm{seq}}_{\alpha,\gamma}(n_1,\ldots,n_k,1)$ as

$$\begin{aligned}
&\frac{Z_{n+1}}{\Gamma_\alpha(n+1)}\left(\gamma+(1-\alpha-\gamma)\frac{1}{(n+1)n}\left(\sum_{u\neq v} n_u n_v + 2n\right)\right)\frac{a_{k+1}}{Z_{n+1}}\left(\prod_{j=1}^{k} w_{n_j}\right) w_1\\
&=\left(p^{\mathrm{seq}}_{\alpha,\gamma}(n_1,\ldots,n_k)+2(1-\alpha-\gamma)\frac{(n-1)n-\sum_{u\neq v} n_u n_v}{(n+1)n(n-1)}\frac{Z_n}{\Gamma_\alpha(n)}p^{\mathrm{PD}^*}_{\alpha,-\alpha-\gamma}(n_1,\ldots,n_k)\right)\\
&\qquad\times\frac{(k-1)\alpha-\gamma}{n-\alpha}.
\end{aligned}$$

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

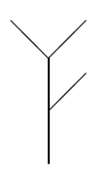
$$\left(1-\frac{1}{n-\alpha}\left(\gamma+\frac{2}{n+1}(1-\alpha-\gamma)\right)\right)p^{\mathrm{seq}}_{\alpha,\gamma}(n_1,\ldots,n_k).$$

Notice that the factor is indeed $p^{\mathrm{seq}}_{\alpha,\gamma}(n,1)$. 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. □

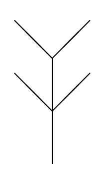
Moreover, sampling consistency can be enhanced to *strong sampling consistency* [16] by requiring that $(T^\circ_{n-1},T^\circ_n)$ has the same distribution as $(T^\circ_{n,-1},T^\circ_n)$.

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.



$\mathbf{t}^\circ_3$



$\mathbf{t}^\circ_4$

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

$$\mathbb{P}(T^\circ_4=\mathbf{t}^\circ_4)=q^{\mathrm{seq}}_{\alpha,\gamma}(2,1,1)q^{\mathrm{seq}}_{\alpha,\gamma}(1,1)=(\alpha-\gamma)(5-5\alpha+\gamma)/((2-\alpha)(3-\alpha)).$$

Then we delete one of the two leaves at the first branch point of $\mathbf{t}^\circ_4$ to get $\mathbf{t}^\circ_3$. Therefore

$$\mathbb{P}((T^\circ_{4,-1},T^\circ_4)=(\mathbf{t}^\circ_3,\mathbf{t}^\circ_4))=\frac{1}{2}\mathbb{P}(T^\circ_4=\mathbf{t}^\circ_4)=\frac{(\alpha-\gamma)(5-5\alpha+\gamma)}{2(2-\alpha)(3-\alpha)}.$$

On the other hand, if $T_3^\circ = \mathbf{t}_3^\circ$, we have to add the new leaf to the first branch point to get $\mathbf{t}_4^\circ$. Thus

$$\mathbb{P}((T_3^\circ, T_4^\circ) = (\mathbf{t}_3^\circ, \mathbf{t}_4^\circ)) = \frac{\alpha-\gamma}{3-\alpha}\mathbb{P}(T_3^\circ = \mathbf{t}_3^\circ) = \frac{(\alpha-\gamma)(2-2\alpha+\gamma)}{(2-\alpha)(3-\alpha)}.$$

It is easy to check that $\mathbb{P}((T_{4,-1}^\circ, T_4^\circ) = (\mathbf{t}_3^\circ, \mathbf{t}_4^\circ)) \neq \mathbb{P}((T_3^\circ, T_4^\circ) = (\mathbf{t}_3^\circ, \mathbf{t}_4^\circ))$ 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. In the binary case $\gamma = \alpha$, the expression simplifies and the result follows from Ford [12], see also [16, Section 5.2].

In the multifurcating case $\gamma < \alpha$, we first make some rearrangement for the coefficient of the sampling consistent splitting rules of alpha-gamma trees identified in Proposition 10:

$$\begin{aligned}
&\gamma + (1-\alpha-\gamma)\frac{1}{n(n-1)}\sum_{i\neq j} n_i n_j \\
&\quad = \frac{(n+1-\alpha-\gamma)(n-\alpha-\gamma)}{n(n-1)}\left(\gamma + (1-\alpha-\gamma)\left(\sum_{i\neq j} A_{ij} + 2\sum_{i=1}^{k} B_i + C\right)\right),
\end{aligned}$$

where

$$\begin{aligned}
A_{ij} &= \frac{(n_i-\alpha)(n_j-\alpha)}{(n+1-\alpha-\gamma)(n-\alpha-\gamma)}, \\
B_i &= \frac{(n_i-\alpha)((k-1)\alpha-\gamma)}{(n+1-\alpha-\gamma)(n-\alpha-\gamma)}, \\
C &= \frac{((k-1)\alpha-\gamma)(k\alpha-\gamma)}{(n+1-\alpha-\gamma)(n-\alpha-\gamma)}.
\end{aligned}$$

Notice that $B_i p_{\alpha,-\alpha-\gamma}^{\mathrm{PD}^*}(n_1,\ldots,n_k)$ simplifies to

$$\begin{aligned}
&\frac{(n_i-\alpha)((k-1)\alpha-\gamma)}{(n+1-\alpha-\gamma)(n-\alpha-\gamma)}\frac{\alpha^{k-2}\Gamma(k-1-\gamma/\alpha)}{Z_n\Gamma(1-\gamma/\alpha)}\Gamma_\alpha(n_1)\ldots\Gamma_\alpha(n_k) \\
&\quad = \frac{Z_{n+2}}{Z_n(n+1-\alpha-\gamma)(n-\alpha-\gamma)}\frac{\alpha^{k-1}\Gamma(k-\gamma/\alpha)}{Z_{n+2}\Gamma(1-\gamma/\alpha)}\Gamma_\alpha(n_1)\ldots\Gamma_\alpha(n_i+1)\ldots\Gamma_\alpha(n_k) \\
&\quad = \frac{\widetilde{Z}_{n+2}}{\widetilde{Z}_n}p_{\alpha,-\alpha-\gamma}^{\mathrm{PD}^*}(n_1,\ldots,n_i+1,\ldots,n_k,1),
\end{aligned}$$

where $\Gamma_\alpha(n)=\Gamma(n-\alpha)/\Gamma(1-\alpha)$ and $\widetilde{Z}_n=Z_n\alpha\Gamma(1-\gamma/\alpha)/\Gamma(n-\alpha-\gamma)$ is the normalisation constant in (4) for $\nu=\mathrm{PD}^*_{\alpha,-\gamma-\alpha}$. The latter can be seen from [17, Formula (17)], which yields

$$\widetilde{Z}_n=\sum_{\{A_1,\ldots,A_k\}\in\mathscr{P}_{[n]}\setminus\{[n]\}}\frac{\alpha^{k-1}\Gamma(k-1-\gamma/\alpha)}{\Gamma(n-\alpha-\gamma)}\prod_{i=1}^{k}\frac{\Gamma(\#A_i-\alpha)}{\Gamma(1-\alpha)},$$

whereas Z_n is the normalisation constant in (2) and hence satisfies

$$Z_n=\sum_{\{A_1,\ldots,A_k\}\in\mathscr{P}_{[n]}\setminus\{[n]\}}\frac{\alpha^{k-2}\Gamma(k-1-\gamma/\alpha)}{\Gamma(1-\gamma/\alpha)}\prod_{i=1}^{k}\frac{\Gamma(\#A_i-\alpha)}{\Gamma(1-\alpha)}.$$

According to (4),

$$p^{\mathrm{PD}^*}_{\alpha,-\alpha-\gamma}(n_1,\ldots,n_k)=\frac{1}{\widetilde{Z}_n}\int_{\mathscr{S}^\downarrow}\sum_{\substack{i_1,\ldots,i_k\geq 1\\ \text{distinct}}}\prod_{l=1}^{k}s_{i_l}^{n_l}\,\mathrm{PD}^*_{\alpha,-\alpha-\gamma}(ds).$$

Thus,

$$\sum_{i=1}^{k}B_i p^{\mathrm{PD}^*}_{\alpha,-\alpha-\gamma}(n_1,\ldots,n_k) \quad=\quad \frac{1}{\widetilde{Z}_n}\int_{\mathscr{S}^\downarrow}\sum_{\substack{i_1,\ldots,i_k\geq 1\\ \text{distinct}}}\prod_{l=1}^{k}s_{i_l}^{n_l}\left(\sum_{u\in\{i_1,\ldots,i_k\},v\notin\{i_1,\ldots,i_k\}}s_u s_v\right)\mathrm{PD}^*_{\alpha,-\alpha-\gamma}(ds)$$

Similarly,

$$\begin{aligned}\sum_{i\neq j}A_{ij}p^{\mathrm{PD}^*}_{\alpha,-\alpha-\gamma}(n_1,\ldots,n_k) &= \frac{1}{\widetilde{Z}_n}\int_{\mathscr{S}^\downarrow}\sum_{\substack{i_1,\ldots,i_k\geq 1\\ \text{distinct}}}\prod_{l=1}^{k}s_{i_l}^{n_l}\left(\sum_{u,v\in\{i_1,\ldots,i_k\}:u\neq v}s_u s_v\right)\mathrm{PD}^*_{\alpha,-\alpha-\gamma}(ds)\\ Cp^{\mathrm{PD}^*}_{\alpha,-\alpha-\gamma}(n_1,\ldots,n_k) &= \frac{1}{\widetilde{Z}_n}\int_{\mathscr{S}^\downarrow}\sum_{\substack{i_1,\ldots,i_k\geq 1\\ \text{distinct}}}\prod_{l=1}^{k}s_{i_l}^{n_l}\left(\sum_{u,v\notin\{i_1,\ldots,i_k\}:u\neq v}s_u s_v\right)\mathrm{PD}^*_{\alpha,-\alpha-\gamma}(ds),\end{aligned}$$

Hence, the EPPF $p^{\mathrm{seq}}_{\alpha,\gamma}(n_1,\ldots,n_k)$ of the sampling consistent splitting rule takes the following form:

$$\begin{aligned}&\frac{(n+1-\alpha-\gamma)(n-\alpha-\gamma)Z_n}{n(n-1)\Gamma_\alpha(n)}\left(\gamma+(1-\alpha-\gamma)\left(\sum_{i\neq j}A_{ij}+2\sum_{i=1}^{k}B_i+C\right)\right)p^{\mathrm{PD}^*}_{\alpha,\gamma}(n_1,\ldots,n_k)\\ &=\frac{1}{Y_n}\int_{\mathscr{S}^\downarrow}\sum_{\substack{i_1,\ldots,i_k\geq 1\\ \text{distinct}}}\prod_{l=1}^{k}s_{i_l}^{n_l}\left(\gamma+(1-\alpha-\gamma)\sum_{i\neq j}s_i s_j\right)\mathrm{PD}^*_{\alpha,-\alpha-\gamma}(ds),\end{aligned}\tag{11}$$

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

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 $0 \le \alpha < 1$. 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 into k parts $(n-k+1, 1, \ldots, 1)$ for some $2 \le k \le n$. 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*

$$q^{\mathrm{seq}}_{1,\gamma}(n_1, \ldots, n_k)$$
$$= \begin{cases} \gamma\Gamma_\gamma(k-1)/(k-1)!, & \text{if } 2 \le k \le n-1 \text{ and } (n_1, \ldots, n_k) = (n-k+1, 1, \ldots, 1); \\ \Gamma_\gamma(n-1)/(n-2)!, & \text{if } k = n \text{ and } (n_1, \ldots, n_k) = (1, \ldots, 1); \\ 0, & \text{otherwise}, \end{cases} \tag{12}$$

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

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

$$\int_{\mathscr{S}^\downarrow} f(s_1, s_2, \ldots)\nu_{1,\gamma}(ds) = \int_0^1 f(s_1, 0, \ldots)\left(\gamma(1-s_1)^{-1-\gamma}ds_1 + \delta_0(ds_1)\right). \tag{13}$$

In particular, it does not *satisfy $\nu(\{s \in \mathscr{S}^\downarrow : s_1 + s_2 + \ldots < 1\}) = 0$. The erosion coefficient c vanishes.*

The presence of the Dirac measure δ_0 in the dislocation measure means that the associated fragmentation process exhibits dislocation events that split a fragment of positive mass into infinitesimal fragments of zero mass, often referred to as *dust* in the fragmentation literature. Dust is also produced by the other part of $\nu_{1,\gamma}$, where also a fraction s_1 of the fragment of positive mass is retained.

Proof. (a) We start from the growth procedure of the alpha-gamma model when $\alpha = 1$. Consider a first split into k parts $(n-k+1, 1, \ldots, 1)$ for some labelled n-leaf tree for some $2 \le k \le n$. Suppose $k \le n-1$ and that the branch point adjacent to the root is created when leaf l is inserted to the root edge, where $l \ge 3$. This insertion happens with probability $\gamma/(l-2)$, as $\alpha = 1$. At stage l, the first split is $(l-1, 1)$. In the following insertions, leaves $l+1, \ldots, n$ have to be added either to the first branch point or to the subtree with $l-1$ leaves at stage l. 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 of $[n]$ is $(n-k+1,1,\dots,1)$ with $k\le n-1$, 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,

$$q^{\mathrm{seq}}_{1,\gamma}(n-k+1,1,\dots,1)=\binom{n-2}{n-k-1}\frac{(n-k-1)!}{(n-2)!}\gamma\Gamma_\gamma(k-1)=\gamma\Gamma_\gamma(k-1)/(k-1)!.$$

On the other hand, for $k=n$, there is only one n-leaf labelled tree with the corresponding first split $(1,\dots,1)$ and in this case, all leaves have to be added to the only branch point. Hence

$$q^{\mathrm{seq}}_{1,\gamma}(1,\dots,1)=\Gamma_\gamma(n-1)/(n-2)!.$$

For sampling consistency, we check criterion (8), which reduces to the two formulas for $2\le k\le n-1$ and $k=n$, respectively,

$$\begin{aligned}\left(1-\frac{1}{n+1}q^{\mathrm{seq}}_{1,\gamma}(n,1)\right)q^{\mathrm{seq}}_{1,\gamma}(n-k+1,1,\dots,1) &= \frac{n-k+2}{n+1}q^{\mathrm{seq}}_{1,\gamma}(n-k+2,1,\dots,1)\\ &\quad+\frac{k}{n+1}q^{\mathrm{seq}}_{1,\gamma}(n-k+1,1,\dots,1)\\ \left(1-\frac{1}{n+1}q^{\mathrm{seq}}_{1,\gamma}(n,1)\right)q^{\mathrm{seq}}_{1,\gamma}(1,\dots,1) &= \frac{2}{n+1}q^{\mathrm{seq}}_{1,\gamma}(2,1,\dots,1)+q^{\mathrm{seq}}_{1,\gamma}(1,\dots,1),\end{aligned}$$

where the right-hand term on the left-hand side is a split of n (into k parts), all others are splits of $n+1$.

(b) According to (12), we have for $2\le k\le n-1$

$$\begin{aligned}&q^{\mathrm{seq}}_{1,\gamma}(n-k+1,1,\dots,1)\\ &=\binom{n}{n-k+1}\frac{\Gamma_\gamma(n+1)}{n!}\gamma B(n-k+2,k-1-\gamma)\\ &=\frac{1}{Y_n}\binom{n}{n-k+1}\int_0^1 s_1^{n-k+1}(1-s_1)^{k-1}\left(\gamma(1-s_1)^{-1-\gamma}ds_1\right)\\ &=\frac{1}{Y_n}\binom{n}{n-k+1}\int_0^1 s_1^{n-k+1}(1-s_1)^{k-1}\left(\gamma(1-s_1)^{-1-\gamma}ds_1+\delta_0(ds_1)\right),\end{aligned}\tag{14}$$

where $Y_n=n!/\Gamma_\gamma(n+1)$. Similarly, for $k=n$,

$$q^{\mathrm{seq}}_{1,\gamma}(1,\dots,1)=\frac{1}{Y_n}\int_0^1\left(n(1-s_1)^{n-1}s_1+(1-s_1)^n\right)\left((\gamma(1-s_1)^{-1-\gamma}ds_1+\delta_0(ds_1)\right).\tag{15}$$

Formulas (14) and (15) are of the form of [16, Formula (2)], which generalises (4) to the case where ν does not necessarily satisfy $\nu(\{s\in\mathscr{S}^\downarrow : s_1+s_2+\ldots<1\})=0$, hence $\nu_{1,\gamma}$ is identified. $\square$

3.3 Continuum random trees and self-similar trees

Let $B\subset\mathbb{N}$ finite. A *labelled tree with edge lengths* is a pair $\vartheta=(\mathbf{t},\eta)$, where $\mathbf{t}\in\mathbb{T}_B$ is a labelled tree, $\eta=(\eta_A,A\in\mathbf{t}\setminus\{\textsc{root}\})$ is a collection of marks, and every edge $C\to A$ of $\mathbf{t}$ is associated with mark

$\eta_A \in (0,\infty)$ at vertex A, which we interpret as the *edge length* of $C \to A$. Let Θ_B be the set of such trees $(\mathbf{t},\eta)$ with $\mathbf{t} \in \mathbb{T}_B$.

We now introduce continuum trees, following the construction by Evans et al. in [9]. A complete separable metric space (τ, 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 (τ, d, ρ), 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 $x \wedge y$ be the unique element in τ such that $[[\rho,x]] \cap [[\rho,y]] = [[\rho, x\wedge y]]$, and call it the *highest common ancestor* of x and y in τ. 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 τ 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 Θ 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

$$d_{\mathrm{GH}}(\tau,\tau') = \inf\{d_{\mathrm{H}}(\widetilde{\tau},\widetilde{\tau}')\}$$

where the infimum is over all metric spaces E and isometric embeddings $\widetilde{\tau} \subset E$ of τ and $\widetilde{\tau}' \subset E$ of τ' with common root $\widetilde{\rho} \in E$; the Hausdorff distance on compact subsets of E is denoted by d_{H}. Evans et al. [9] showed that $(\Theta, d_{\mathrm{GH}})$ is a complete separable metric space.

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

A *weighted $\mathbb{R}$-tree* (τ,μ) is called a *continuum tree* [1], if μ is a probability measure on τ and

1. μ is supported by the set $\mathscr{L}(\tau)$,
2. μ has no atom,
3. for every $x \in \tau \backslash \mathscr{L}(\tau)$, $\mu(\tau_x) > 0$.

A *continuum random tree (CRT)* is a random variable whose values are continuum trees, defined on some probability space $(\Omega, \mathscr{A}, \mathbb{P})$. 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 ℓ_1 with the Hausdorff metric, and the set of probability measures on ℓ_1 with any metric inducing the topology of weak convergence, so that the set of pairs (T,μ) where T is a rooted $\mathbb{R}$-tree embedded as a subset of ℓ_1 and μ is a measure on T, is endowed with the product σ-algebra.

An exchangeable $\mathscr{P}_{\mathbb{N}}$*-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 $|\pi_i| = \lim_{n\to\infty} n^{-1}\#[n] \cap \pi_j$, the random variable $\Pi(t+s)$ has the same law as the random partition whose blocks are those of $\pi_i \cap \Pi^{(i)}(|\pi_i|^a s), i \geq 1$, where $(\Pi^{(i)}, i \geq 1)$ is a sequence of i.i.d. copies of $(\Pi(t), t \geq 0)$. The process $(|\Pi(t)|^{\downarrow}, t \geq 0)$ is an $S^{\downarrow}$*-valued self-similar fragmentation process*. Bertoin [5] proved that the distribution of a $\mathscr{P}_{\mathbb{N}}$-valued self-similar fragmentation process is determined by a triple (a, c, ν), where $a \in \mathbb{R}$, $c \geq 0$ and ν is a dislocation measure on $S^{\downarrow}$. 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, ν) the characteristic pair. When $a = 0$, the process $(\Pi(t), t \geq 0)$ is also called *homogeneous fragmentation process*.

A CRT $(\mathscr{T}, \mu)$ is a *self-similar CRT* with index $a = -\gamma < 0$ if for every $t \geq 0$, given $(\mu(\mathscr{T}_t^i), i \geq 1))$ where $\mathscr{T}_t^i, i \geq 1$ is the ranked order of connected components of the open set $\{x \in \tau : d(x, \rho(\tau)) > t\}$, the continuum random trees

$$\left(\mu(\mathscr{T}_t^1)^{-\gamma}\mathscr{T}_t^1, \frac{\mu(\cdot \cap \mathscr{T}_t^1)}{\mu(\mathscr{T}_t^1)}\right), \left(\mu(\mathscr{T}_t^2)^{-\gamma}\mathscr{T}_t^2, \frac{\mu(\cdot \cap \mathscr{T}_t^2)}{\mu(\mathscr{T}_t^2)}\right), \ldots$$

are i.i.d copies of $(\mathscr{T}, \mu)$, where $\mu(\mathscr{T}_t^i)^{-\gamma}\mathscr{T}_t^i$ is the tree that has the same set of points as $\mathscr{T}_t^i$, but whose distance function is divided by $\mu(\mathscr{T}_t^i)^{\gamma}$. Haas and Miermont in [15] have shown that there exists a self-similar continuum random tree $\mathscr{T}_{(\gamma,\nu)}$ characterized by such a pair (γ, ν), which can be constructed from a self-similar fragmentation process with characteristic pair (γ, ν).

3.4 The alpha-gamma model when $\gamma = 1 - \alpha$, sampling from the stable CRT

Let $(\mathscr{T}, \rho, \mu)$ be the stable tree of Duquesne and Le Gall [7]. The distribution on Θ of any CRT is determined by its so-called finite-dimensional marginals: the distributions of $\mathscr{R}_k$, $k \geq 1$, the subtrees $\mathscr{R}_k \subset \mathscr{T}$ defined as the discrete trees with edge lengths spanned by $\rho, U_1, \ldots, U_k$, where given $(\mathscr{T}, \mu)$, the sequence $U_i \in \mathscr{T}$, $i \geq 1$, of leaves is sampled independently from μ. See also [22; 8; 16; 17; 19] for various approaches to stable trees. Let us denote the discrete tree without edge lengths associated with $\mathscr{R}_k$ 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*

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

for any $k \geq 2$, $n_1 \geq 1, \ldots, n_k \geq 1$, $n = n_1 + \ldots + n_k$.

We recognise $p_{1/\alpha}^{\text{stable}} = p_{\alpha,-1}^{\text{PD}^*}$ in (2), and by Proposition 1, we have $p_{\alpha,-1}^{\text{PD}^*} = p_{\alpha,1-\alpha}^{\text{seq}}$. The full distribution of $\mathscr{R}_n$, $n \geq 1$, is displayed in Theorem 5, which in the stable case was first obtained by [7, Theorem 3.3.3]. Furthermore, it can be shown that the trees $(T_k, k \geq 1)$ obtained by sampling from the stable CRT follow the alpha-gamma growth rules for $\gamma = 1 - \alpha$, see e.g. Marchal [19]. 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 identically distributed given $(\mathscr{T}, \mu)$, they are exchangeable. If we denote by $\mathscr{L}_{n,-1}$ the random set of leaves $\mathscr{L}_n = \{U_1, \ldots, U_n\}$ with a uniformly chosen member removed, then $(\mathscr{L}_{n,-1}, \mathscr{L}_n)$ has the same conditional distribution as $(\mathscr{L}_{n-1}, \mathscr{L}_n)$. Hence the pairs of (unlabelled) tree shapes spanned by ρ 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 ν is always inexplicit, which leads us to transform ν to a more explicit form. For simplicity, let us assume that ν satisfies $\nu(\{s \in \mathscr{S}^\downarrow : s_1 + s_2 + \ldots < 1\}) = 0$. The method proposed here is to change the space $\mathscr{S}^\downarrow$ into the space $[0,1]^{\mathbb{N}}$ and to rearrange the elements $s \in \mathscr{S}^\downarrow$ under ν 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 ν^{sb} on the space $[0,1]^{\mathbb{N}}$ the size-biased dislocation measure associated with dislocation measure ν, if for any subset $A_1 \times A_2 \times \ldots \times A_k \times [0,1]^{\mathbb{N}}$ of $[0,1]^{\mathbb{N}}$,

$$\nu^{\mathrm{sb}}(A_1 \times A_2 \times \ldots \times A_k \times [0,1]^{\mathbb{N}}) = \sum_{\substack{i_1,\ldots,i_k \geq 1 \\ \text{distinct}}} \int_{\{s \in \mathscr{S}^\downarrow : s_{i_1} \in A_1, \ldots, s_{i_k} \in A_k\}} \frac{s_{i_1} \ldots s_{i_k}}{\prod_{j=1}^{k-1}(1-\sum_{l=1}^{j} s_{i_l})} \nu(ds) \qquad (16)$$

for any $k \in \mathbb{N}$, where ν is a dislocation measure on $\mathscr{S}^\downarrow$ satisfying $\nu(s \in \mathscr{S}^\downarrow : s_1 + s_2 + \ldots < 1) = 0$. We also denote by $\nu_k^{\mathrm{sb}}(A_1 \times A_2 \times \ldots \times A_k) = \nu^{\mathrm{sb}}(A_1 \times A_2 \times \ldots \times A_k \times [0,1]^{\mathbb{N}})$ 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 jth entry of the re-ordered vector fall into A_j, $j \in [k]$. Notice that the support of such a size-biased dislocation measure ν^{sb} is a subset of $\mathscr{S}^{\mathrm{sb}} := \{s \in [0,1]^{\mathbb{N}} : \sum_{i=1}^{\infty} s_i = 1\}$. If we denote by $s^\downarrow$ the sequence $s \in \mathscr{S}^{\mathrm{sb}}$ rearranged into ranked order, taking (16) into formula (4), we obtain

Proposition 17. *The EPPF associated with a dislocation measure ν can be represented as:*

$$p(n_1, \ldots, n_k) = \frac{1}{\widetilde{Z}_n} \int_{[0,1]^k} x_1^{n_1-1} \ldots x_k^{n_k-1} \prod_{j=1}^{k-1}(1-\sum_{l=1}^{j} x_l) \nu_k^{\mathrm{sb}}(dx),$$

where ν^{sb} is the size-biased dislocation measure associated with ν, 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 $\mathrm{PD}^*_{\alpha,\theta}$ to then study $\nu^{\mathrm{sb}}_{\alpha,\gamma}$.

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

$$\mathrm{gem}^*_{\alpha,\theta}(x_1,\ldots,x_k)=\frac{\alpha\Gamma(2+\theta/\alpha)}{\Gamma(1-\alpha)\Gamma(\theta+\alpha+1)\prod_{j=2}^k B(1-\alpha,\theta+j\alpha)}\frac{(1-\sum_{i=1}^k x_i)^{\theta+k\alpha}\prod_{j=1}^k x_j^{-\alpha}}{\prod_{j=1}^k(1-\sum_{i=1}^j x_i)}, \tag{17}$$

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

This is a simple σ-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 4. We start our proof from the dislocation measure associated with the alpha-gamma model. According to (5) and (16), the first k marginals of $\nu^{\mathrm{sb}}_{\alpha,\gamma}$ are given by

$$\begin{aligned}
&\nu_k^{\mathrm{sb}}(A_1\times\ldots\times A_k)\\
&=\sum_{\substack{i_1,\ldots,i_k\geq 1\\ \text{distinct}}}\int_{\{s\in\mathscr{S}^\downarrow: s_{i_j}\in A_j, j\in[k]\}}\frac{s_{i_1}\ldots s_{i_k}}{\prod_{j=1}^{k-1}(1-\sum_{l=1}^j s_{i_l})}\left(\gamma+(1-\alpha-\gamma)\sum_{i\neq j}s_is_j\right)\mathrm{PD}^*_{\alpha,-\alpha-\gamma}(ds)\\
&=\gamma D+(1-\alpha-\gamma)(E-F),
\end{aligned}$$

where

$$\begin{aligned}
D &= \sum_{\substack{i_1,\ldots,i_k\geq 1\\ \text{distinct}}}\int_{\{s\in\mathscr{S}^\downarrow:\ s_{i_1}\in A_1,\ldots,s_{i_k}\in A_k\}}\frac{s_{i_1}\ldots s_{i_k}}{\prod_{j=1}^{k-1}(1-\sum_{l=1}^j s_{i_l})}\mathrm{PD}^*_{\alpha,-\alpha-\gamma}(ds)\\
&= \mathrm{GEM}^*_{\alpha,-\alpha-\gamma}(A_1\times\ldots\times A_k),\\
E &= \sum_{\substack{i_1,\ldots,i_k\geq 1\\ \text{distinct}}}\int_{\{s\in\mathscr{S}^\downarrow:\ s_{i_1}\in A_1,\ldots,s_{i_k}\in A_k\}}\left(1-\sum_{u=1}^k s_{i_u}^2\right)\frac{s_{i_1}\ldots s_{i_k}}{\prod_{j=1}^{k-1}(1-\sum_{l=1}^j s_{i_l})}\mathrm{PD}^*_{\alpha,-\alpha-\gamma}(ds)\\
&= \int_{A_1\times\ldots\times A_k}\left(1-\sum_{i=1}^k x_i^2\right)\mathrm{GEM}^*_{\alpha,-\alpha-\gamma}(dx)\\
F &= \sum_{\substack{i_1,\ldots,i_k\geq 1\\ \text{distinct}}}\int_{\{s\in\mathscr{S}^\downarrow:\ s_{i_1}\in A_1,\ldots,s_{i_k}\in A_k\}}\left(\sum_{v\notin\{i_1,\ldots,i_k\}}s_v^2\right)\frac{s_{i_1}\ldots s_{i_k}}{\prod_{j=1}^{k-1}(1-\sum_{l=1}^j s_{i_l})}\mathrm{PD}^*_{\alpha,-\alpha-\gamma}(ds)\\
&= \sum_{\substack{i_1,\ldots,i_{k+1}\geq 1\\ \text{distinct}}}\int_{\{s\in\mathscr{S}^\downarrow:\ s_{i_1}\in A_1,\ldots,s_{i_k}\in A_k\}}\frac{s_{i_{k+1}}^2}{1-\sum_{l=1}^k s_{i_l}}\frac{s_{i_1}\ldots s_{i_{k+1}}}{\prod_{j=1}^{k}(1-\sum_{l=1}^j s_{i_l})}\mathrm{PD}^*_{\alpha,-\alpha-\gamma}(ds)\\
&= \int_{A_1\times\ldots\times A_k\times[0,1]}\frac{x_{k+1}}{1-\sum_{i=1}^k x_i}\mathrm{GEM}^*_{\alpha,-\alpha-\gamma}(d(x_1,\ldots,x_{k+1})).
\end{aligned}$$

Applying (17) to F (and setting $\theta=-\alpha-\gamma$), then integrating out x_{k+1}, we get:

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

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

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

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

For general (α,γ), the explicit form of the dislocation measure in size-biased order, specifically the density $g_{\alpha,\gamma}$ of the first marginal of $\nu^{\mathrm{sb}}_{\alpha,\gamma}$, 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 $\mathscr{P}_{\mathbb{N}}$-valued fragmentation process with dislocation measure $\nu_{\alpha,\gamma}$ for some $0 < \alpha < 1$ and $0 \leq \gamma < \alpha$. Then, for the size $|\Pi^{\alpha,\gamma}_{(i)}(t)|$ of the block containing $i \geq 1$, the process $\xi_{(i)}(t) = -\log |\Pi^{\alpha,\gamma}_{(i)}(t)|$, $t \geq 0$, is a pure-jump subordinator with Lévy measure*

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

A similar result holds for the binary case $\gamma = \alpha$, see [24, Equation (10), also Section 4.2].

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 fragmentation CRTs $\mathscr{T}^{\alpha,\gamma}$ as n tends to infinity, where $\mathscr{T}^{\alpha,\gamma}$ is a γ-selfsimilar fragmentation CRT whose dislocation measure is a multiple of $\nu_{\alpha,\gamma}$, as in Corollary 3, cf. Section 3.3.

Lemma 20. *If $(\widetilde{T}^\circ_n)_{n\geq 1}$ are strongly sampling consistent discrete fragmentation trees in the sense that $(T^\circ_{n-1}, T^\circ_n)$ has the same distribution as $(T^\circ_{n,-1}, T^\circ_n)$ for all $n \geq 2$, cf. Section 2.5, associated with dislocation measure $\nu_{\alpha,\gamma}$ for some $0 < \alpha < 1$ and $0 < \gamma \leq \alpha$, then*

$$\frac{\widetilde{T}^\circ_n}{n^\gamma} \to \mathscr{T}^{\alpha,\gamma}$$

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

Proof. For $\gamma = \alpha$ this is [16, Corollary 17]. For $\gamma < \alpha$, we apply Theorem 2 in [16], which says that a strongly sampling consistent family of discrete fragmentation trees $(\widetilde{T}^\circ_n)_{n\geq 1}$ converges in probability to a CRT

$$\frac{\widetilde{T}^\circ_n}{n^{\gamma_\nu}\ell(n)\Gamma(1-\gamma_\nu)} \to \mathscr{T}_{(\gamma_\nu,\nu)}$$

for the Gromov-Hausdorff metric if the dislocation measure ν satisfies following two conditions:

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

$$\int_{\mathscr{S}^{\downarrow}} \sum_{i\geq 2} s_i |\ln s_i|^{\rho} \nu(ds) < \infty, \tag{19}$$

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

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

$$\Lambda([x,\infty)) = x^{-\gamma_\nu}\ell^*(1/x), \qquad \text{as } x \downarrow 0,$$

where Λ is the Lévy measure of the tagged particle subordinator as in Corollary 19. Specifically, the slowly varying functions ℓ and ℓ^* are asymptotically equivalent since

$$\Lambda([x,\infty)) = \int_{\mathscr{S}^{\downarrow}} (1-s_1)\nu(ds) + \nu(s_1 \leq e^{-x}) = \int_{\mathscr{S}^{\downarrow}} (1-s_1)\nu(ds) + (1-e^{-x})^{-\gamma_\nu}\ell\left(\frac{1}{1-e^{-x}}\right)$$

implies that

$$\frac{\ell^*(1/x)}{\ell(1/x)} = \frac{\Lambda([x,\infty)}{x^{-\gamma_\nu}\ell(1/x)} \sim \left(\frac{1-e^{-x}}{x}\right)^{-\gamma_\nu} \frac{\ell(x+\frac{1-x+xe^{-x}}{1-e^{-x}})}{\ell(x)} \to 1.$$

So, the dislocation measure $\nu_{\alpha,\gamma}$ satisfies (18) with $\ell(x) \to \alpha\Gamma(1-\gamma/\alpha)/\Gamma(1-\alpha)\Gamma(1-\gamma)$ and $\gamma_{\nu_{\alpha,\gamma}} = \gamma$. Notice that

$$\int_{\mathscr{S}^{\downarrow}} \sum_{i\geq 2} s_i |\ln s_i|^{\rho} \nu_{\alpha,\gamma}(ds) \leq \int_0^\infty x^\rho \Lambda_{\alpha,\gamma}(dx).$$

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

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

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 3.2 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 τ 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 ρ is the root of τ.

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

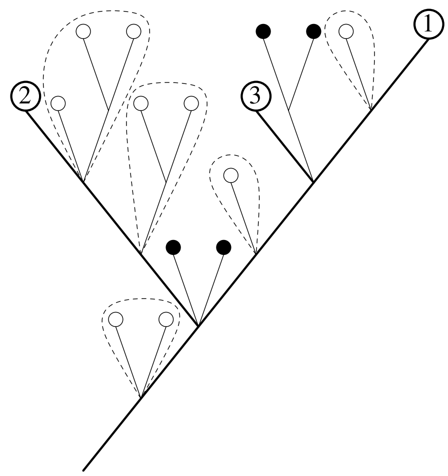


Figure 2: We display an example of $S(T_{16},[3])$, seven skeletal subtrees in five skeletal bushes (within the dashed lines, white leaves) and further subtrees in the branch points of $S(T_{16},[3])$ (with black leaves).

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 5*

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

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

Proof of Theorem 5 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 $W_{n,k}=m$ black balls and $n-k-m$ white 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 ℓ other edges of T_k. As $n\to\infty$, the urn is such that $W_{n,k}/n\to W_k$ a.s., where $W_k\sim \text{beta}((k-1)\alpha-l\gamma, k(1-\alpha)+l\gamma)$.

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 skeletal subtrees S^{sk}_{vj} that we join together into a *skeletal bush* S^{sk}_v, cf. Figure 2. Note that the total length $L^{(n)}_k$ 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 $k+\ell$ of T_k. Hence, as $n\to\infty$

$$\frac{L_k^{(n)}}{n^\gamma} \sim \frac{\overline{K}_n}{W_{n,k}^\gamma}\left(\frac{W_{n,k}}{n}\right)^\gamma \sim \frac{\overline{K}_n}{W_{n,k}^\gamma} W_k^\gamma. \tag{20}$$

The partition of leaves (associated with white balls), where each skeletal bush gives rise to a block, follows the dynamics of a Chinese Restaurant Process with (γ,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 ith 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

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

Applying Lemma 8 in connection with (20), we get the probability density of L_k/W_k^γ 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 γ. Hence, the conditional probability that the length of each edge is $(n_1,\dots,n_{k+l})$ at stage n is that

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

Then $D_k^{(n)}$ converge a.s. to the Dirichlet limit as specified. Moreover, $L_k^{(n)}D_k^{(n)} \to L_kD_k$ 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^γ given T_k and W_k, which does not depend on W_k. Similarly, the conditional distribution of D_k given 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^γ 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. To be specific, it is useful to enumerate all vertices in a unique way, e.g. in the order they are visited by depth first search [18], where beginning from the root each subtree is visited recursively, in the order of least labels.

Proposition 22. *Let $(c_1+1,\dots,c_\ell+1)$ 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^α will converge to*

$$C_k = (C_{k,1},\dots,C_{k,\ell}) = \overline{W}_k^\alpha M_k D_k', \qquad \textit{where } D_k' \sim \mathrm{Dirichlet}(c_1-1-\gamma/\alpha,\dots,c_\ell-1-\gamma/\alpha),$$

M_k and W_k are conditionally independent given T_k, where $\overline{W}_k = 1 - W_k$, and M_k has density

$$\frac{\Gamma(\overline{w}+1)}{\Gamma(\overline{w}/\alpha+1)} s^{\overline{w}/\alpha} g_\alpha(s), \qquad s \in (0,\infty),$$

$\overline{w} = (k-1)\alpha - \ell\gamma$ 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_{vj}^{sk} of a branch point $v \in R(T_n,[k])$ gives rise to a block, follows the dynamics of a Chinese Restaurant Process with $(\alpha,\overline{w})$-seating plan. Hence, the total degree $C_k^{\mathrm{tot}}(n)/\overline{W}_{n,k}^{\alpha} \to M_k$ a.s., where $C_k^{\mathrm{tot}}(n)$ is the sum of degrees in T_n of the branch points of T_k, and $\overline{W}_{n,k} = n - k - W_{n,k}$ 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*

$$(P_1, P_2, \ldots) = (W_1, \overline{W}_1 W_2, \overline{W}_1 \overline{W}_2 W_3, \ldots)$$

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 γ. This observation also follows from Proposition 6, because colouring (iv)$^{\mathrm{col}}$ and crushing (cr) do not affect the partition of leaf labels according to subtrees of the spine.

Acknowledgement

We would like to thank the referee for various suggestions that led to an improved presentation.

References

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

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

[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. MR1395604

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

[5] J. Bertoin. Self-similar fragmentations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(3):319–340, 2002. MR1899456

[6] J. Bertoin. *Random fragmentation and coagulation processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006. MR2253162

[7] T. Duquesne and J.-F. Le Gall. Random trees, Lévy processes and spatial branching processes. *Astérisque*, (281):vi+147, 2002. MR1954248

[8] T. Duquesne and J.-F. Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields*, 131(4):553–603, 2005. MR2147221

[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. MR2221786

[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. MR2243874

[11] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968. MR0228020

[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). *Probability Theory and Related Fields – Online First*, DOI 10.1007/s00440-008-0169-3, 2008.

[14] R. C. Griffiths. Allele frequencies with genic selection. *J. Math. Biol.*, 17(1):1–10, 1983. MR0707220

[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. MR2041829

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

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

[18] D. E. Knuth. *The art of computer programming. Vol. 1: Fundamental algorithms*. Second printing. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont, 1969. MR0286317

[19] P. Marchal. A note on the fragmentation of a stable tree. In *Fifth Colloquium on Mathematics and Computer Science*, volume AI, pages 489–500. Discrete Mathematics and Theoretical Computer Science, 2008.

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

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

[22] G. Miermont. Self-similar fragmentations derived from the stable tree. II. Splitting at nodes. *Probab. Theory Related Fields*, 131(3):341–375, 2005. MR2123249

[23] 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. MR2245368

[24] J. Pitman and M. Winkel. Regenerative tree growth: binary self-similar continuum random trees and Poisson-Dirichlet compositions. *Preprint, arXiv:0803.3098,* 2008*, to appear in Annals of Probability*.