The Standard Additive Coalescent *

David Aldous and Jim Pitman

Technical Report No. 489

Department of Statistics
University of California
367 Evans Hall # 3860
Berkeley, CA 94720-3860

September 3, 1997

*Research supported in part by N.S.F. Grants DMS 96-22859 and 97-03961

Abstract

Regard an element of the set

$$\Delta := \{(x_1, x_2, \ldots) : x_1 \geq x_2 \geq \ldots \geq 0, \sum_i x_i = 1\}$$

as a fragmentation of unit mass into clusters of masses x_i. The additive coalescent of Evans and Pitman (1997) is the Δ-valued Markov process in which pairs of clusters of masses $\{x_i, x_j\}$ merge into a cluster of mass $x_i + x_j$ at rate $x_i + x_j$. They showed that a version $(\mathbf{X}^\infty(t), -\infty < t < \infty)$ of this process arises as a $n \to \infty$ weak limit of the process started at time $-\frac{1}{2}\log n$ with n clusters of mass $1/n$. We show this *standard additive coalescent* may be constructed from the continuum random tree of Aldous (1991,1993) by Poisson splitting along the skeleton of the tree. We describe the distribution of $\mathbf{X}^\infty(t)$ on Δ at a fixed time t. We show that the size of the cluster containing a given atom, as a process in t, has a simple representation in terms of the stable subordinator of index 1/2. As $t \to -\infty$, we establish a Gaussian limit for (centered and normalized) cluster sizes, and study the size of the largest cluster.

Key words and phrases. Brownian excursion, continuum random tree, random forest, splitting, stable subordinator, stochastic coalescence, stochastic fragmentation.

AMS 1991 subject classifications. 60C05, 60J25, 60J65

1 Introduction

Consider a uniform random tree $\mathcal{T}_n$ on n labeled vertices. Let

$$\mathcal{F}^n(n) := \mathcal{T}_n, \mathcal{F}^n(n-1), \mathcal{F}^n(n-2), \ldots, \mathcal{F}^n(1)$$

be the sequence of random forests obtained by deleting the edges of $\mathcal{T}_n$ one by one in uniform random order. In reversed time, this forest-valued Markov chain has transition probabilities of the following simple form (explicit in Pitman [27] and implicit in earlier work of Yao [35]). Write $\#\mathbf{t}$ for the size (number of vertices) of a tree $\mathbf{t}$.

Lemma 1 *The transition probabilities of* $(\mathcal{F}^n(m), 1 \le m \le n)$ *are as follows. Given the current forest consists of trees* $(\mathbf{t}_1, \ldots, \mathbf{t}_k)$*, pick a pair* (i,j) *with* $1 \le i < j \le k$ *with probability* $\frac{\#\mathbf{t}_i + \#\mathbf{t}_j}{n(k-1)}$*, pick uniform vertices of* $\mathbf{t}_i$ *and* $\mathbf{t}_j$*, and add an edge joining these vertices.*

Let $\bar{\mathcal{F}}^n$ be the continuous-time chain derived from $\mathcal{F}^n$ by incorporating exponential (rate $k-1$) holds between jumps, where k is the current number of trees. For a forest $\mathbf{f}$ write $\#\mathbf{f}$ for the ranked vector of sizes of the trees comprising $\mathbf{f}$ (*ranked* means: in decreasing order). Consider for each n the continuous-time process

$$\mathbf{X}^n(t) := n^{-1} \ \#\bar{\mathcal{F}}^n(\tfrac{1}{2}\log n + t), \quad t \ge -\tfrac{1}{2}\log n$$

where, by appending an infinite sequence of zeros, $\mathbf{X}^n(t)$ is regarded as a random element of the set

$$\Delta := \{(x_1, x_2, \ldots) : x_1 \ge x_2 \ge \ldots \ge 0, \sum_i x_i = 1\}$$

which we give the ℓ_1 topology. Lemma 1 implies that the process $\mathbf{X}^n$ is a (ranked) additive coalescent (Evans and Pitman [13]), that is a Δ-valued Markov process in which pairs of clusters of masses $\{x_i, x_j\}$ merge into a cluster of mass $x_i + x_j$ at rate $x_i + x_j$, and the state is re-ranked as mergers occur. The process $\mathbf{X}^n$ starts at time $-\frac{1}{2}\log n$ from the configuration $\mathbf{u}_n := (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}, 0, 0, \ldots)$ consisting of n clusters of mass $1/n$. From [13] (Proposition 18 and subsequent discussion), there is the following result:

Proposition 2 *As* $n \to \infty$

$$\mathbf{X}^n(\cdot) \xrightarrow{d} \mathbf{X}^\infty(\cdot) \tag{1}$$

in the sense of Skorokhod convergence on $D((-\infty, \infty), \Delta)$*, where the limit process* $(\mathbf{X}^\infty(t), -\infty < t < \infty)$ *is an additive coalescent.*

Call $\mathbf{X}^\infty$ the *standard additive coalescent.* The central result of this paper is the following more explicit construction of this process. In the uniform random tree $\mathcal{T}_n$, put mass $1/n$ on each vertex and let each edge have length $1/n^{1/2}$. As $n \to \infty$, a weak limit is obtained, the *Brownian continuum random tree* (concisely, the CRT) studied by Aldous [2, 3, 4]. A realization of the CRT is equipped with a *mass measure* of total mass 1 concentrated on the leaves of the tree, and a σ-finite *length measure*, such that for vertices v, w which are distance d apart, the path $[[v, w]]$ from v to w has length measure d. The *skeleton* of the CRT is the union over pairs of leaves $\{v, w\}$ of the open paths $]]v, w[[$.

In the CRT, the analog of deleting randomly-chosen edges in $\mathcal{T}_n$ is to cut the skeleton by a Poisson process of cuts with some rate λ per unit length. These cuts split the CRT into a *continuum forest*, that is a countably infinite set of smaller continuum trees. Varying λ gives a continuum-forest-valued *fragmentation process* (Section 2.2). Let

$$(\mathbf{Y}(\lambda), \lambda \geq 0) = (Y_i(\lambda), i \geq 1, \lambda \geq 0)$$

be the process of ranked masses of tree-components in the continuum forests obtained by cutting at various intensities λ. We call $\mathbf{Y}$ the *Δ-valued fragmentation process derived from the CRT.*

Theorem 3 *Let $\mathbf{X}(t) := \mathbf{Y}(e^{-t})$ where $\mathbf{Y}$ is the Δ-valued fragmentation process derived from the CRT. Then the process $(\mathbf{X}(t), -\infty < t < \infty)$ is a version of the standard additive coalescent.*

To be precise, the finite dimensional distributions of the Δ-valued process $\mathbf{X}$ defined in Theorem 3 are identical to those of $\mathbf{X}^\infty$ defined by weak-convergence in Proposition 2. We will work with a version of $\mathbf{Y}$ which has right continuous paths and left limits. Then $\mathbf{X}$ will have left continuous paths with right limits, and the process of right limits of $\mathbf{X}$ will serve as $\mathbf{X}^\infty$ in (1). We prove Theorem 3 in Section 2, after collecting background facts about the CRT.

The additive coalescent is the special case $K(x, y) = x + y$ of the general stochastic coalescent, in which clusters of masses $\{x, y\}$ merge at rate $K(x, y)$. Aldous [7] gives a lengthy survey of scientific literature related to stochastic coalescence (see our Section 6.2 for one aspect). Evans and Pitman [13] construct various coalescents with infinite numbers of clusters as strong-Markov processes with appropriate state spaces. Similarities and difference between the additive and multiplicative [6] coalescents will be listed in Section 6.3.

Theorem 3 not only brings together the several recent lines of research mentioned above, but also suggests an extensive range of new questions.

Some are answered by results stated in the remainder of this introduction and proved in the main body of the paper, and others are posed in Section 6. It turns out to be more convenient to state results in terms of $\mathbf{Y}(\lambda) = \mathbf{X}(-\log \lambda)$.

Some features of the process $(\mathbf{Y}(\lambda), 0 \leq \lambda < \infty)$ will now be described in terms a stable subordinator of index 1/2, denoted $(S_{1/2}(\lambda), 0 \leq \lambda < \infty)$, which is the increasing process with stationary independent increments such that

$$E \exp(-\theta S_{1/2}(\lambda)) = \exp(-\lambda\sqrt{2\theta}) \quad (\theta, \lambda \geq 0)$$

$$P(S_{1/2}(1) \in dx) = (2\pi)^{-1/2} x^{-3/2} e^{-\frac{1}{2x}}\, dx \quad (x > 0). \tag{2}$$

We consider first the distribution of $\mathbf{Y}(\lambda)$ on Δ for fixed λ. The following theorem is proved in Section 3.1.

Theorem 4 *Fix $0 < \lambda < \infty$. Let $J_1 \geq J_2 \geq \ldots$ be the ranked jump-sizes of $S_{1/2}(\cdot)$ over the interval $[0,1]$. Then there is the following equality of distributions on Δ:*

$$\mathbf{Y}(\lambda) \stackrel{d}{=} \left(\frac{J_1}{S_{1/2}(1)}, \frac{J_2}{S_{1/2}(1)}, \ldots \middle| S_{1/2}(1) = \frac{1}{\lambda^2} \right). \tag{3}$$

The finite-dimensional distributions of the random vector in (3) are described explicitly in Perman [25] and Pitman and Yor [31, Sec. 8.1], but they are rather complicated.

Suppose now that $U_1, U_2, \ldots$ is a sequence of random leaves of the CRT picked independently and uniformly at random according to the mass measure. Let $Y_i^\dagger(\lambda)$ be the mass of the tree component of the random forest that contains U_i when the cutting intensity is λ, and let $\mathbf{Y}^*(\lambda)$ be the sub-sequence of distinct masses in the sequence $(Y_i^\dagger(\lambda), i \geq 1)$. Then for each fixed $\lambda > 0$, the sequence $\mathbf{Y}^*(\lambda)$ is a *size-biased random permutation* of $\mathbf{Y}(\lambda)$. That is (see e.g. [10, 26, 28]) $Y_i^*(\lambda) = Y_{I(i)}(\lambda)$, where for $j \notin \{I(1), \ldots, I(i-1)\}$

$$P(I(i) = j | \mathbf{Y}(\lambda), I(1), \ldots, I(i-1)) = \frac{Y_j(\lambda)}{\sum_{k \notin \{I(1), \ldots, I(i-1)\}} Y_k(\lambda)}. \tag{4}$$

Theorem 4 combined with results of Perman et al. [26] yields (Section 3.2) the following simpler description of the distribution of $\mathbf{Y}^*(\lambda)$:

Corollary 5 *Let $Z_1, Z_2, \ldots$ be independent standard Gaussian variables, and let $S_m = \sum_{i=1}^m Z_i^2$. Then for each fixed $\lambda > 0$*

$$\mathbf{Y}^*(\lambda) \stackrel{d}{=} \left(\frac{\lambda^2}{\lambda^2 + S_{m-1}} - \frac{\lambda^2}{\lambda^2 + S_m}, \, m \geq 1 \right). \tag{5}$$

In particular,

$$Y_1^*(\lambda) \stackrel{d}{=} \frac{Z_1^2}{\lambda^2 + Z_1^2} \tag{6}$$

$$P(Y_1^* \leq y) = 2\Phi(\lambda y^{1/2}(1-y)^{-1/2}) - 1 \qquad (0 \leq y < 1) \tag{7}$$

where Φ is the standard normal distribution function; and $Y_1^*(\lambda)$ has density function

$$f_\lambda(y) := (2\pi)^{-1/2} \lambda y^{-1/2} (1-y)^{-3/2} \exp(-\tfrac{1}{2}\lambda^2 y/(1-y)) \qquad (0 \leq y < 1). \tag{8}$$

We note as a consequence of (8) and (4) for $i = 1$ the following formula for every non-negative Borel function g defined on $(0, 1)$:

$$E\left(\sum_i g(Y_i(\lambda)) \right) = \int_0^1 y^{-1} f_\lambda(y) g(y) \, dy. \tag{9}$$

Consider now the real-valued process $(Y_1^*(\lambda), 0 < \lambda < \infty)$, that is the mass of the tree-component containing the random leaf U_1 in the fragmentation process of the CRT. Equivalently, $X_1^*(t) := Y_1^*(e^{-t})$ is the size at time t of the cluster of the standard additive coalescent containing a point picked at random from the mass distribution. In Section 4 we show that this process admits the following simple representation:

Theorem 6

$$(Y_1^*(\lambda), 0 \leq \lambda < \infty) \stackrel{d}{=} \left(\frac{1}{1 + S_{1/2}(\lambda)}, \; 0 \leq \lambda < \infty \right).$$

Theorem 6 implies that as $\lambda \to \infty$ most of the mass is in clusters whose size is order λ^{-2}. In Section 5 we strengthen this result as follows to a Gaussian limit for the empirical measure of rescaled cluster sizes

$$F_\lambda(\cdot) := \sum_i 1(\lambda^2 Y_i(\lambda) \in \cdot).$$

Theorem 7 *Define $H_\lambda(\cdot) = \lambda^{-2} EF_\lambda(\cdot)$. As $\lambda \to \infty$,*

$$H_\lambda(\cdot) \to \chi(\cdot)$$

where χ is the measure on $(0,\infty)$ with σ-finite density $(2\pi)^{-1/2}x^{-3/2}e^{-x/2}$ and convergence is weak convergence on each (ε,∞). Define

$$G_\lambda(\cdot) = \lambda^{-1}\left(F_\lambda(\cdot) - \lambda^2\chi(\cdot)\right).$$

Then $G_\lambda(\cdot) \to G(\cdot)$, where $G(\cdot)$ is the mean-zero Gaussian random field with

$$\begin{aligned} \text{Var } G(dy) &= \chi(dy) &(10)\\ EG(dy_1)G(dy_2) &= -y_1y_2\chi(dy_1)\chi(dy_2),\ \ y_1 \neq y_2 &(11) \end{aligned}$$

where convergence is convergence in distribution of the continuous path processes $(G_\lambda(s,\infty), 0 < s < \infty)$ to $(G(s,\infty), 0 < s < \infty)$.

We interpret (10,11) in their integrated form

$$EG(s_1,\infty)G(s_2,\infty) = \int_{s_1}^\infty \int_{s_2}^\infty EG(dy_1)G(dy_2) \ + \int_{s_1 \vee s_2} \text{Var } G(dy).$$

As will be described in Section 6.2, Theorem 7 is loosely related to existing scientific literature. The $\lambda \to \infty$ asymptotics of the largest cluster size are discussed in Section 5.1, while asymptotics as $\lambda \to 0$ are considered in Section 3.3.

2 Construction of the standard additive coalescent

2.1 The CRT

Fix $k \geq 2$ and consider a tree with k leaves labeled $\{1,\ldots,k\}$ in which each internal node has degree 3 and each edge e has a positive real-valued length l_e. See Figure 1. Such a tree has $2k-3$ edges, and when edge lengths are ignored there are $\prod_{i=1}^{k-2}(2i-1)$ different possible *shapes* $\hat{t}$ for the tree. (Formally, the shape $\hat{t}$ is a combinatorial tree with k leaves labeled by $\{1,\ldots,k\}$ and $k-2$ unlabeled internal nodes.)

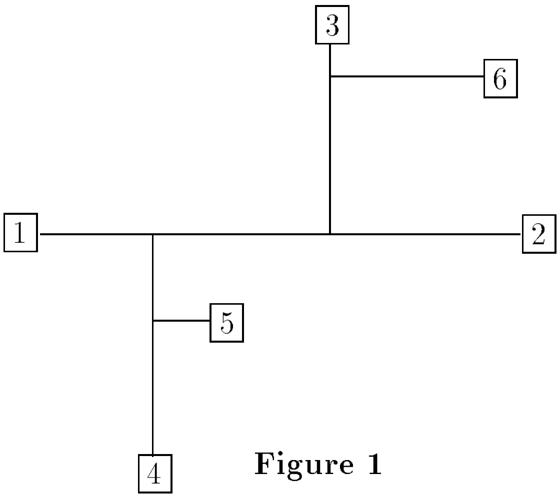

Figure 1

Consider a random such tree $\mathcal{R}(k)$ whose shape and edge-lengths (L_i) satisfy

$$P(\mathrm{shape}(\mathcal{R}(k)) = \hat{t}, L_1 \in dl_1, \ldots, L_{2k-3} \in dl_{2k-3})$$
$$= s \exp(-s^2/2)\ dl_1 \ldots dl_{2k-3} \quad \text{where } s = \sum_{i=1}^{2k-3} l_i. \tag{12}$$

Thus the shape is uniform on the set of possible shapes, the edge-lengths are independent of shape, and the edge-lengths are exchangeable, so that no labeling convention for edges need be specified. Figure 2 later shows a realization of $\mathcal{R}(50)$. Lemma 21 of [4] says that such a distribution exists (k here is $k-1$ there), and that the family $(\mathcal{R}(k), 2 \leq k < \infty)$ is *consistent* in that the subtree of $\mathcal{R}(k+1)$ spanned by leaves $\{1, \ldots, k\}$ is distributed as $\mathcal{R}(k)$. Therefore we can construct simultaneously the family $(\mathcal{R}(k), 2 \leq k < \infty)$ so that the subtree of $\mathcal{R}(k+1)$ spanned by leaves $\{1, \ldots, k\}$ is exactly $\mathcal{R}(k)$. A realization of $\mathcal{R}(k)$ can be viewed as a compact metric space, where the distance between two points is the length of the path between them. Define the realization of the CRT $\mathcal{T}$ to be the completion of the increasing union $\cup_k \mathcal{R}(k)$. Discarding a null event, results of [2, 4] include

Theorem 8 *Each realization of $\mathcal{T}$ has the following properties (i)-(iii).*
(i) $\mathcal{T}$ is compact and topologically a tree.
(ii) There is a σ-finite length measure ℓ *on $\mathcal{T}$, whose restriction to $\mathcal{R}(k) \subset \mathcal{T}$ is the natural length measure on the edges of $\mathcal{R}(k)$, and which is null outside the skeleton $\cup_k \mathcal{R}(k)$.*
(iii) There is a mass measure μ *on $\mathcal{T}$, with $\mu(\mathcal{T}) = 1$ and $\mu(\cup_k \mathcal{R}(k)) = 0$, characterized as the weak limit*

$$\mu := \lim_k \mu_k \tag{13}$$

where μ_k is the uniform probability distribution on leaves $\{1.\ldots, k\} \subset \mathcal{T}$.
(iv) [vertex-exchangeability] *Given $\mathcal{T}$, let $\{U_1, \ldots, U_k\}$ be random elements of $\mathcal{T}$ chosen independently according to $\mu(\cdot)$, and let $\hat{\mathcal{R}}(k)$ be the subtree spanned by $\{U_1, \ldots, U_k\}$, with U_i relabeled by i for $1 \leq i \leq k$. Then unconditionally $\hat{\mathcal{R}}(k) \stackrel{d}{=} \mathcal{R}(k)$.*
(v) The total length D_k of the edges of $\mathcal{R}(k)$ has distribution

$$P(D_k > d) = P(N(d^2/2) < k - 1) \tag{14}$$

where $N(\nu)$ has Poisson(ν) distribution.

How $\mathcal{T}$ arises as weak limits of random finite trees is discussed in detail in [4]. A simple aspect of such convergence is provided by the next lemma. Recall $\mathcal{T}_n$ is the uniform random tree on $\{1, \ldots, n\}$.

Lemma 9 *Assign length $1/n^{1/2}$ to each edge of $\mathcal{T}_n$. Let $\mathcal{R}(n,k)$ be the subtree of $\mathcal{T}_n$ spanned by vertices $\{1, \ldots, k\}$. Then for each fixed $k \geq 2$*

$$\mathcal{R}(n,k) \stackrel{d}{\to} \mathcal{R}(k) \text{ as } n \to \infty$$

in the sense that the joint distributions of shape and edge-lengths converge to the distribution (12).

Remarks. For many purposes a construction [4] of $\mathcal{T}$ from standard Brownian excursion $(B^{ex}_u, 0 \leq u \leq 1)$ is useful. In that construction, μ is the measure on $\mathcal{T}$ induced by Lebesgue measure on $[0,1]$. Elaborations of this construction, the *Brownian snake*, are used in studying superprocesses: see Le Gall [16, 18] and Dynkin-Kusnetsov [12]. Thus it is not surprising that many of our distributional expressions (e.g. Theorem 6) have a "Brownian flavor". But it is harder to interpret the length measure ℓ in the construction of $\mathcal{T}$ from a Brownian excursion, and the symmetry and self-similarity properties which the CRT inherits from the discrete random tree $\mathcal{T}_n$ tend to be obscured. While in principle one must be able to derive these properties in terms of a Brownian excursion (see e.g. Le Gall [17] for a derivation of (12)), we find it useful to take a more combinatorial approach to the CRT. Different "hidden symmetries" of Brownian excursion revealed in this way are the subject of Aldous [5].

2.2 The fragmentation process of the CRT

Any countable subset $\mathcal{C}$ of $\mathcal{T}$, viewed as a cut-set, splits $\mathcal{T}$ into a forest $\mathcal{F}$, where two elements x, y of $\mathcal{T}$ are in the same tree-component of $\mathcal{F}$ iff the unique path from x to y contains no element of $\mathcal{C}$. Now fix $0 < \lambda < \infty$ and let $\mathcal{C}_\lambda$ be a Poisson point process of mean measure $\lambda\ell(\cdot)$ on $\mathcal{T}$. That is, for each k the restriction of $\mathcal{C}_\lambda$ to $\mathcal{R}(k)$ is a Poisson point process of rate λ per unit length. Then $\mathcal{C}_\lambda$ splits $\mathcal{T}$ into a random forest $\mathcal{F}_\lambda$ (for remarks on the state space of $\mathcal{F}_\lambda$ see Section 3.5). Figure 2 shows a (genuine) simulation of $\mathcal{R}(50)$ and its cut-points with rate $e = 2.718\ldots$; the 50 leaves are the endpoints of line segments in figure 2, and the cut-points are marked $*$. Figure 3 is the same picture, with the various tree-components moved apart. The reader should imagine figures 2 and 3 as portions of the forest $\mathcal{F}_e$ obtained by using $\mathcal{C}_e$ to split $\mathcal{T}$. Of course, in the real $\mathcal{F}_e$ each tree has infinitely more smaller and smaller branches, and there are infinitely more small trees.

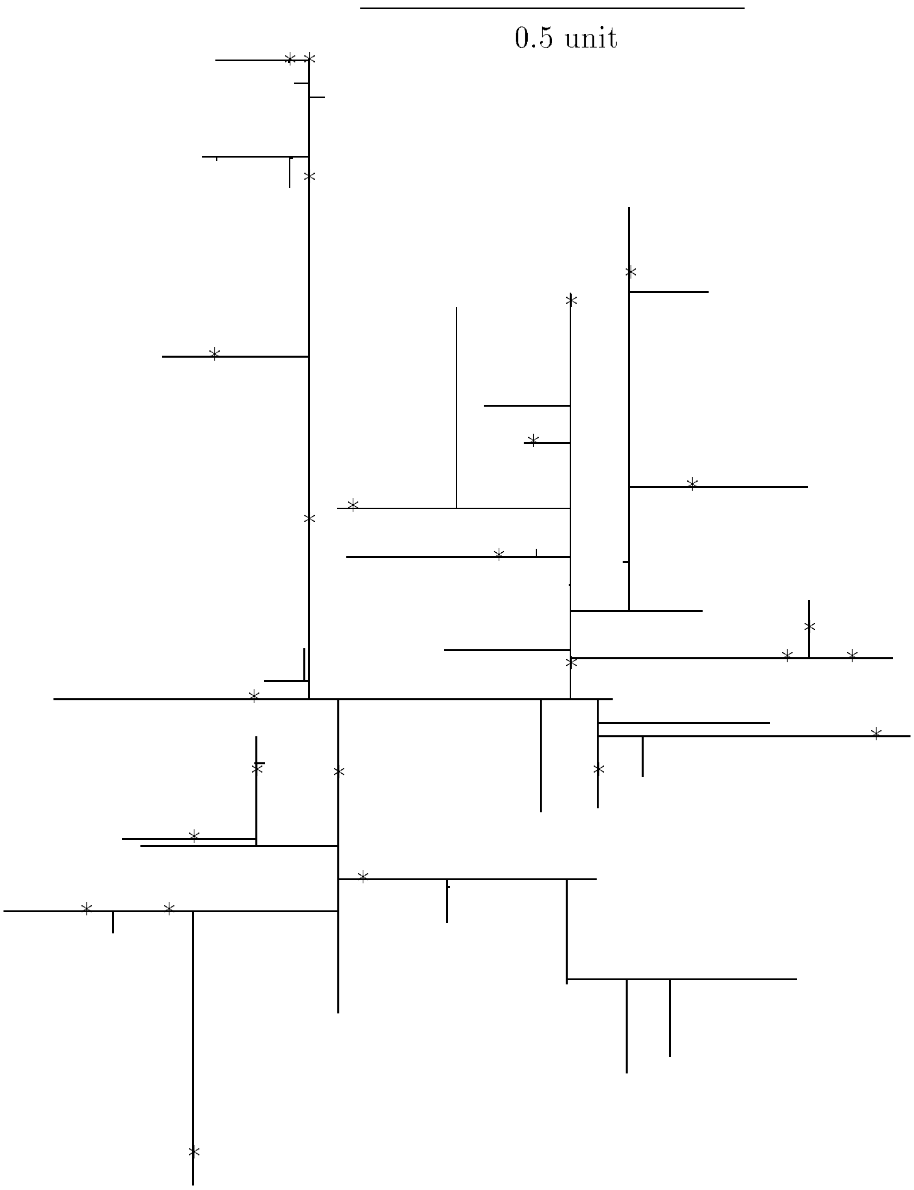

Figure 2

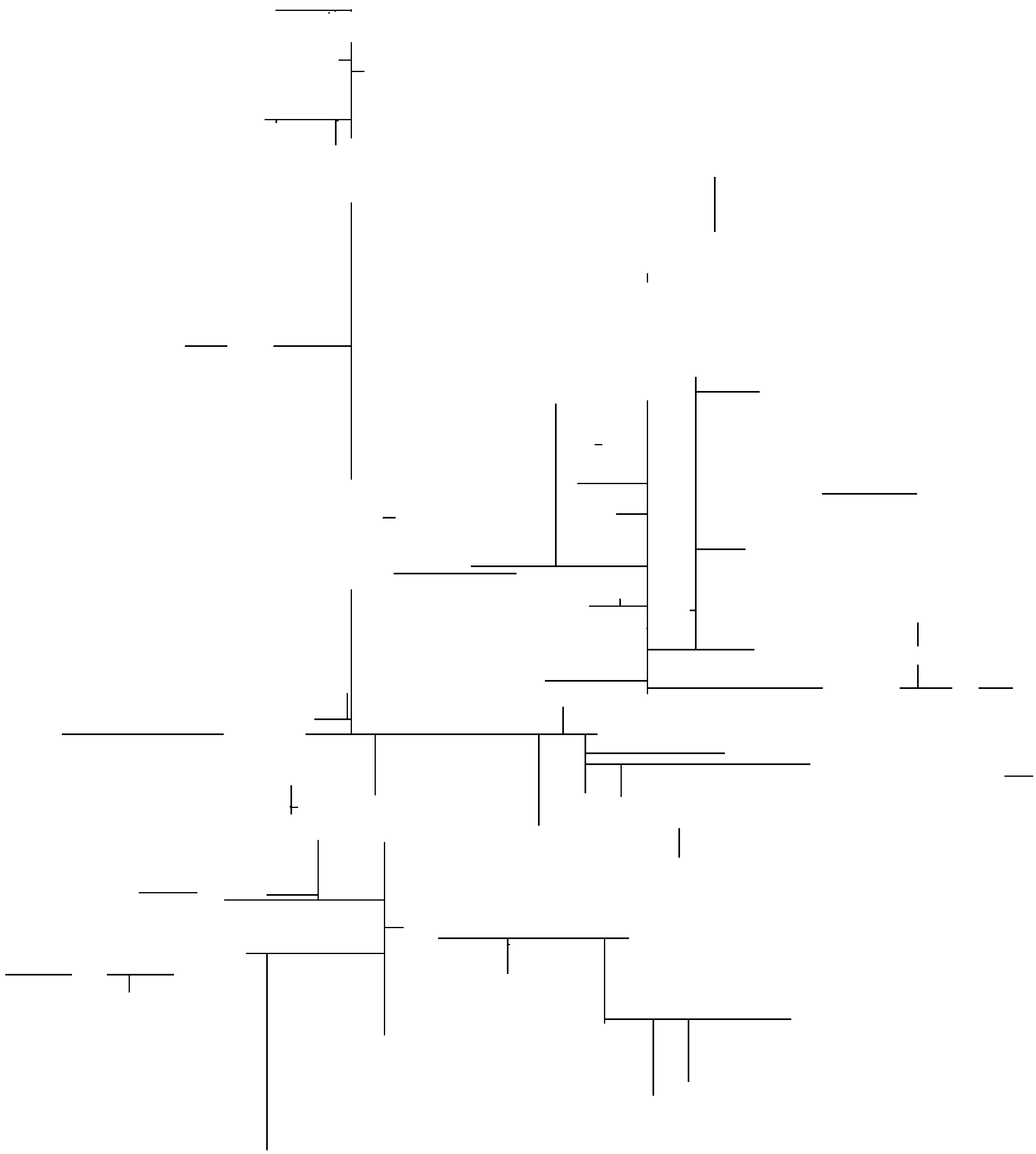

Figure 3

Write $\mathbf{Y}(\lambda) = (Y_1(\lambda), Y_2(\lambda), \ldots)$ for the ranked μ-masses of the tree-components of $\mathcal{F}_\lambda$. A priori, a countable cut-set might produce an uncountable number of tree-components each with zero μ-mass, but the following technical lemma proves this does not happen with our Poisson process of cuts.

Lemma 10 $\sum_{i=1}^\infty Y_i(\lambda) = 1$ *a.s. for each* $\lambda > 0$.

Suppose now that the family of Poisson processes $(\mathcal{C}_\lambda, 0 < \lambda < \infty)$ is constructed so that for $\lambda_1 < \lambda_2$ the process $\mathcal{C}_{\lambda_1}$ is obtained by retaining each point of $\mathcal{C}_{\lambda_2}$ independently with chance λ_1/λ_2 and deleting the other points. In this way we obtain the Δ-valued fragmentation process $(\mathbf{Y}(\lambda), 0 < \lambda < \infty)$ which is the central focus of this paper. Note that by Lemma 10, $\mathbf{Y}(\lambda)$ takes values in the space Δ which is a Polish space in the topology it inherits as a subset of l_1.

Proof of Lemma 10. Write $M = \sum_i Y_i(\lambda) \leq 1$. Write A_k for the event that the leaves $\{1, \ldots, k\}$ are all in different tree-components. By vertex-exchangeability (Theorem 8 (iv))

$$P(A_k|\mathcal{T}, M) \geq (1-M)^k$$

because given $\mathcal{T}$ and M a leaf picked at random according to μ has chance $1-M$ to be in a tree-component of zero μ-mass. So $P(A_k) \geq \varepsilon^k P(M \leq 1-\varepsilon)$, and it suffices to prove

$$(P(A_k))^{1/k} \to 0 \text{ as } k \to \infty. \tag{15}$$

Now for event A_k to occur, the tree $\mathcal{R}(k)$ must contain at least $k-1$ of the cut-points $\mathcal{C}_\lambda$, and so

$$P(A_k) \leq P(N(\lambda D_k) \geq k-1)$$

where $N(\cdot)$ is a rate-1 Poisson counting process and D_k is the total edge-length of $\mathcal{R}(k)$. The distribution of D_k is given by (14), and then routine large deviation estimates establish (15). $\square$

2.3 Proof of Theorem 3

We do not know how to show that $(\mathbf{Y}(e^{-t}))$ evolves as an additive coalescent by direct calculations with continuous parameter processes. Rather, we use discrete approximation arguments. Effectively, this reproves the weak convergence result of Proposition 2 in parallel with corresponding approximations to the CRT fragmentation process. The present proof of Proposition 2 differs from the proof in Section 6.1 of [13] in that it does not involve the explicit description of the distribution of sizes of components in the discrete approximation, displayed in formula (20) of the next section. But both proofs make essential use of the existence and Feller property of the additive coalescent semigroup on Δ, which was established in [13] by a pathwise construction of the additive coalescent from an arbitrary initial state in Δ based on partition-valued processes derived from an extension of the random forest representation of Lemma 1.

Recall that $(\mathbf{X}^n(t);\ -\frac{1}{2}\log n \leq t < \infty)$ is the additive coalescent started in state $\mathbf{u}_n := (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}, 0, 0, \ldots) \in \Delta$ at time $-\frac{1}{2}\log n$. From the discussion following Lemma 1, there is the following explicit construction of $\mathbf{X}^n$ in terms of the forest-valued Markov chain $(\mathcal{F}^n(m), 1 \leq m \leq n)$:

$$\mathbf{X}^n(-\tfrac{1}{2}\log n + t) = n^{-1}\ \#\mathcal{F}^n(M^n(t)) \tag{16}$$

where

$$M^n(t) := \min\{m : \sum_{i=1}^{m} \tfrac{1}{n-i}\ \xi_i > t\}, \tag{17}$$

where the (ξ_i) are i.i.d. exponential(1), and where $\#\mathcal{F}$ is the ranked vector of sizes of tree-components of $\mathcal{F}$. Fix $k \geq 2$. Recall from Lemma 9 that $\mathcal{R}(n,k)$ denotes the subtree of $\mathcal{T}_n$ spanned by vertices $\{1,\ldots,k\}$, where each edge of $\mathcal{T}_n$ is given length $1/n^{1/2}$. Take a random sample of $m(n)$ edges of $\mathcal{T}_n$, and write $\mathcal{R}(n,k,m(n))$ for $\mathcal{R}(n,k)$ with each sampled edge marked by a cut at its midpoint. Suppose $m(n)/n^{1/2} \to \lambda > 0$. Then Lemma 9 easily extends to show that

$$\mathcal{R}(n,k,m(n)) \xrightarrow{d} \mathcal{R}(\infty,k,\lambda) \tag{18}$$

where $\mathcal{R}(\infty,k,\lambda)$ denotes the tree $\mathcal{R}(k)$ with a Poisson, rate λ per unit length, process of marked points on its edges, and the space of trees with k leaves and a finite number of marked points is given an appropriate topology. It follows that

$$\#\mathcal{R}(n,k,m(n)) \xrightarrow{d} \#\mathcal{R}(\infty,k,\lambda)$$

where in each case $\#\mathcal{R}$ denotes the decreasing vector counting numbers of vertices $\{1,\ldots,k\}$ in each of the tree-components obtained by cutting at the marks. Now by the characterization of the mass measure μ as the weak limit (13), and by definition of $\mathbf{Y}(\lambda)$,

$$k^{-1}\#\mathcal{R}(\infty,k,\lambda) \xrightarrow{d} \mathbf{Y}(\lambda) \text{ as } k \to \infty.$$

It follows that for $k_n \to \infty$ sufficiently slowly,

$$k_n^{-1}\ \#\mathcal{R}(n,k_n,m(n)) \xrightarrow{d} \mathbf{Y}(\lambda).$$

In the uniform random tree $\mathcal{T}_n$, vertices $\{1,\ldots,k_n\}$ are distributed as a simple random sample of k_n vertices, so by Lemma 11 below

$$n^{-1}\#\mathcal{F}^n(n-m(n)) = n^{-1}\#\mathcal{R}(n,n,m(n)) \xrightarrow{d} \mathbf{Y}(\lambda). \tag{19}$$

Recall (17) that $M^n(\cdot)$ is the inverse function of

$$S^n(m) := \sum_{i=1}^{m} \tfrac{1}{n-i}\ \xi_i.$$

It is easy to check that for fixed $a > 0$

$$S^n(n-an^{1/2}) - \tfrac{1}{2}\log n \xrightarrow{P}\ -\log a$$

and then to check that

$$n^{-1/2}\left(n - M^n(\tfrac{1}{2}\log n + t)\right) \xrightarrow{P} e^{-t}$$

for fixed $-\infty < t < \infty$. By inserting into (19)

$$n^{-1}\#\mathcal{F}^n(M^n(\tfrac{1}{2}\log n + t)) \xrightarrow{d} \mathbf{Y}(e^{-t})$$

and from the representation (16) this says

$$\mathbf{X}^n(t) \xrightarrow{d} \mathbf{Y}(e^{-t}), \qquad \text{for fixed } -\infty < t < \infty.$$

By the Feller property of the additive coalescent semigroup, this implies there is a version $\mathbf{X}^\infty(t)$ of the additive coalescent such that $\mathbf{X}^n(\cdot) \xrightarrow{d} \mathbf{X}^\infty(\cdot)$ in the Skorokhod sense (i.e. Proposition 2) and that $\mathbf{X}^\infty(t) \stackrel{d}{=} \mathbf{Y}(e^{-t})$ for fixed t. To complete the proof of Theorem 3, we need to identify the finite-dimensional distributions of $\mathbf{X}^\infty(t)$ with those of $\mathbf{Y}(e^{-t})$. But this just requires repeating the argument, starting with $m_i(n)/n^{1/2} \to \lambda_i$, $1 \le i \le j$ and the j-dimensional analog of (18). We omit the details. □

In the course of the proof we used the following routine consequence of the WLLN for sampling without replacement.

Lemma 11 *Let $k_n \to \infty$ with $k_n = o(n)$. For each n take a simple random sample of k_n vertices from a forest $\mathbf{f}^n$ on n vertices, and write $\mathbf{y}^n$ for the vector whose entries count the number of sampled vertices in each tree of $\mathbf{f}^n$, in decreasing order. Then, for any $\mathbf{y} \in \Delta$,*

$$k_n^{-1}\mathbf{y}^n \xrightarrow{p} \mathbf{y} \text{ iff } n^{-1}\#\mathbf{f}^n \to \mathbf{y}.$$

In the case where $\mathbf{f}^n$ is a random forest, for any random element $\mathbf{Y}$ of Δ

$$k_n^{-1}\mathbf{y}^n \xrightarrow{d} \mathbf{Y} \text{ iff } n^{-1}\#\mathbf{f}^n \xrightarrow{d} \mathbf{Y}.$$

3 The distribution of $\mathbf{Y}(\lambda)$

In this section we prove Theorem 4, extract some of its consequences, and remark on its interpretation in terms of Brownian bridge.

3.1 Proof of Theorem 4

Recall that $n^{-1}\#\mathcal{F}^n(n-k+1)$ for $1 \le k \le n$ is the random vector of ranked relative sizes of the k tree components in the random forest obtained by deleting a random sample of $k-1$ edges picked from the set of $n-1$ edges of $\mathcal{T}_n$, a uniform random tree on n vertices. It is known [27] that

$$n^{-1}\#\mathcal{F}^n(n-k+1) \stackrel{d}{=} ((J_{1,k}/\Sigma_k, J_{2,k}/\Sigma_k, \ldots, J_{k,k}/\Sigma_k) \,|\, \Sigma_k = n) \qquad (20)$$

where $J_{1,k} \ge J_{2,k} \ge \ldots \ge J_{k,k}$ are the ranked values of k independent random variables $X_1, \ldots, X_k$ with the *Borel*(1) *distribution*

$$P(X_i = m) = e^{-m}m^{m-1}/m!,\ m = 1, 2, \ldots$$

and where

$$\Sigma_k := X_1 + \cdots + X_k = J_{1,k} + \ldots + J_{k,k}.$$

By (19), if $k(n)/n^{1/2} \to \lambda > 0$ then

$$n^{-1}\#\mathcal{F}^n(n - k(n) + 1) \xrightarrow{d} \mathbf{Y}(\lambda).$$

So to show (3) we need to show that for $k = k(n)$ satisfying $n/k^2(n) \to \lambda^{-2}$ the right side of (20) converges in distribution to the right side of (3), that is

$$(J_1/S_{1/2}(1), J_2/S_{1/2}(1), \ldots \mid S_{1/2}(1) = \lambda^{-2}) \tag{21}$$

where $J_1 \geq J_2 \geq \ldots$ are the ranked jump-sizes of $S_{1/2}$ over the interval $[0,1]$. But this convergence follows from standard results regarding the asymptotic joint distribution of sums and order statistics of a sequence of independent random variables in the domain of attraction of the stable law of index 1/2. See for instance Lemma 11 of [9].

3.2 Proof of Corollary 5

By the identity in distribution (3) just established, the size-biased random permutation $\mathbf{Y}^*(\lambda)$ of $\mathbf{Y}(\lambda)$ satisfies

$$\mathbf{Y}^*(\lambda) \stackrel{d}{=} (J_1^*/\Sigma, J_2^*/\Sigma, \ldots \mid \Sigma = \lambda^{-2}) \tag{22}$$

where $(J_1^*, J_2^*, \ldots)$ is a size-biased random permutation of the jump-sizes (J_i) of the stable subordinator $S_{1/2}$ over the interval $[0,1]$, and $\Sigma := \sum_i J_i^* = S_{1/2}(1)$ almost surely. The J_i are the ranked points of a Poisson random measure (PRM) on $(0,\infty)$ governed by the Lévy measure $(2\pi)^{-1/2}\ x^{-3/2}dx$ of the stable(1/2) subordinator. The joint density of the first m components of the random vector in (22) can now be read from the following lemma (Theorem 2.1 of [26]).

Lemma 12 *Let $(J_1^*, J_2^*, \ldots)$ be a size-biased random permutation of the points of a PRM on $(0,\infty)$ with intensity $\rho(x)dx$, let $\Sigma := \sum_i J_i^*$ and suppose $P(\Sigma \in dx) = f(x)dx$ where f is strictly positive and continuous on $(0,\infty)$. For $m \geq 0$ let*

$$\Sigma_m^* := \Sigma - \sum_{k=1}^{m} J_k^* = \sum_{m+1}^{\infty} J_k^*. \tag{23}$$

Then the sequence $(\Sigma_m^*, m \geq 0)$ is a Markov chain with stationary transition probabilities determined by the formula

$$P\left(J_{m+1}^*/\Sigma_m^* \in dy \,|\, \Sigma_m^* = t\right) = f^*(y|t) \qquad (0 < y < 1) \tag{24}$$

where

$$f^*(y|t) := yt\,\rho(yt)\frac{f((1-y)t)}{f(t)} \qquad (0 < y < 1). \tag{25}$$

In particular, for $\rho(x) = (2\pi)^{-1/2}x^{-3/2}$ and $f(x) = P(S_{1/2}(1) \in dx)/dx$ as displayed in (2),

$$f^*(y|\lambda^{-2}) = f_\lambda(y) \tag{26}$$

as displayed in (8). Thus (22) and (24) imply that f_λ is the density of $Y_1^*(\lambda)$, and then (6,7) are elementary reformulations of this fact. To establish the representation (5) of the distribution of the entire sequence $(Y_m^*(\lambda), m \geq 1)$, fix $\lambda > 0$ and let $R_m := 1 - \sum_{i=1}^m Y_i^*(\lambda)$. Then the assertion (5) can be rewritten as

$$(R_m, m \geq 1) \stackrel{d}{=} \left(\frac{\lambda^2}{\lambda^2 + S_m}, m \geq 1\right) \tag{27}$$

or again as

$$\left(\frac{1}{R_m} - \frac{1}{R_{m-1}}, m \geq 1\right) \stackrel{d}{=} \left(\frac{Z_m^2}{\lambda^2}, m \geq 1\right). \tag{28}$$

But from (24) and (26)

$$P(1 - R_{m+1}/R_m \in dy \,|\, R_1, \ldots, R_m) = f_{\lambda/\sqrt{R_m}}(y)dy \qquad (0 < y < 1).$$

By (6), if $1 - Y$ has density f_λ then $(1-Y)/Y \stackrel{d}{=} Z_1^2/\lambda^2$. So

$$\left(\frac{1 - R_{m+1}/R_m}{R_{m+1}/R_m}\middle| R_1, \ldots, R_m\right) \stackrel{d}{=} \left(\frac{Z_1^2 R_m}{\lambda^2}\middle| R_1, \ldots, R_m\right)$$

where Z_1 is standard Gaussian, independent of $(R_1, \ldots, R_m)$, and hence

$$\left(\frac{1}{R_{m+1}} - \frac{1}{R_m}\middle| R_1, \ldots, R_m\right) \stackrel{d}{=} \frac{Z_1^2}{\lambda^2}$$

which yields (28).

3.3 Asymptotics as $\lambda \to 0$

Theorem 4 and the calculations of the previous section yield the following description of how the fragmentation process $\mathbf{Y}(\lambda)$ gets started with λ near zero. This translates into a description of how the additive coalescent $\mathbf{X}(t)$ terminates as $t \to \infty$. See Evans and Pitman [14] for further developments.

Corollary 13 *As* $\lambda \to 0$

$$\lambda^{-2}(1 - Y_1(\lambda), Y_2(\lambda), Y_3(\lambda), \ldots) \xrightarrow{d} (S_{1/2}(1), J_1, J_2, \ldots)$$

Proof. It suffices [10] to establish the corresponding result for size-biased permutations, that is

$$\lambda^{-2}(1 - Y_1^*(\lambda), Y_2^*(\lambda), Y_3^*(\lambda), \ldots) \xrightarrow{d} (S_{1/2}(1), J_1^*, J_2^*, \ldots) \tag{29}$$

for (J_i^*) a size-biased random permutation of (J_i). For $m \geq 0$ let $R_m(\lambda) := 1 - \sum_{i=1}^m Y_i^*(\lambda)$, previously denoted R_m in (27), and as in (23) let $\Sigma_m^* := S_{1/2}(1) - \sum_{k=1}^m J_k^*$. Then (29) amounts to

$$\lambda^{-2}(R_m(\lambda), m \geq 1) \xrightarrow{d} (\Sigma_{m-1}^*, m \geq 1). \tag{30}$$

But from (27)

$$\lambda^{-2}(R_m(\lambda), m \geq 1) \stackrel{d}{=} \left(\frac{1}{\lambda^2 + S_m}, m \geq 1\right) \xrightarrow{d} \left(\frac{1}{S_m}, m \geq 1\right)$$

as $\lambda \to 0$, so it only remains to check the identity in distribution

$$(\Sigma_j^*, j \geq 0) \stackrel{d}{=} \left(\frac{1}{S_{j+1}}, j \geq 0\right). \tag{31}$$

To check this, write simply Σ instead of $\Sigma_0^* = S_{1/2}(1)$, and observe from (27) and (22) that

$$(\Sigma_m^*, m \geq 1 \,|\, \Sigma) \stackrel{d}{=} \left(\frac{1}{\Sigma^{-1} + S_m}, m \geq 1\right) \tag{32}$$

where the sequence (S_m) is independent of Σ. Now (31) follows easily, because $\Sigma^{-1} \stackrel{d}{=} Z_1^2$ by Lévy's observation that for each fixed $\lambda > 0$

$$S_{1/2}(\lambda) \stackrel{d}{=} \lambda^2 / Z_1^2 \text{ where } Z_1 \text{ is standard Gaussian.} \tag{33}$$

□

Corollary 13 is the continuous analog of the following asymptotic result for the discrete approximating scheme, which can be verified using (20). Let $\#\mathcal{F}^n(n-k+1) = (Y_1^{n,k}, Y_2^{n,k}, \ldots, Y_k^{n,k})$, say, be the random vector of ranked sizes of the k tree components in the random forest obtained by deleting a random sample of $k-1$ edges picked from the set of $n-1$ edges of $\mathcal{T}_n$. Then as n and k tend to ∞ with $k/\sqrt{n} \to 0$,

$$k^{-2}(n - Y_1^{n,k}, Y_2^{n,k}, Y_3^{n,k}, \ldots) \xrightarrow{d} (S_{1/2}(1), J_1, J_2, \ldots) \tag{34}$$

where the limit distribution is the same as in Corollary 13. A result equivalent to the convergence in distribution of the first component in (34), that is $k^{-2}(n - Y_1^{n,k}) \xrightarrow{d} S_{1/2}(1)$, was obtained by Pavlov [24].

3.4 Emergence of the majority cluster

The density $g_\lambda(y)$ of the largest cluster $Y_1(\lambda)$ relates partially to the density $f_\lambda(y)$ of $Y_1^*(\lambda)$ via the formula

$$f_\lambda(y) = y g_\lambda(y) \text{ on } \tfrac{1}{2} < y < 1 \tag{35}$$

because only the largest cluster can have size greater than 1/2. Thus

$$P(Y_1(\lambda) > \tfrac{1}{2}) = \int_{1/2}^1 g_\lambda(y)dy = \int_{1/2}^1 y^{-1} f_\lambda(y)dy$$

which combines with (8) to give an explicit integral formula for $P(Y_1(\lambda) > \frac{1}{2})$ and hence for $P(X_1(t) > \frac{1}{2})$. Here are some numerical values.

t	-1.0	-0.5	0	0.5	1.0	1.5
$P(X_1(t) > \frac{1}{2})$	0.012	0.167	0.484	0.747	0.891	0.956

3.5 Remarks on a Brownian bridge representation

Let $\mathbf{V}^{br} := (V_1^{br}, V_2^{br}, \ldots)$ be the sequence of ranked lengths of excursions of the *standard Brownian bridge* B^{br}. Write L_1^{br} for the local time of B^{br} at 0 up to time 1. Then Theorem 4 can be rewritten (see e.g. [9, 30, 31])

$$\mathbf{Y}(\lambda) \stackrel{d}{=} (\mathbf{V}^{br} \mid L_1^{br} = \lambda). \tag{36}$$

The following elaboration suggests itself. Described in [4] Theorem 13 is a deterministic mapping from a set of continuous "excursion" functions $f : [0,1] \to [0,\infty)$ satisfying $f(0) = f(1) = 0$,

$$f(u) > 0, \ 0 < u < 1 \tag{37}$$

and certain technical conditions, into a set of continuum trees. By removing requirement (37) and applying a similar mapping to each excursion of f one can define a mapping from a set of "reflecting" functions into a set of continuum forests, where the mass-measure of a tree-component equals the length of the corresponding excursion of f. The upshot of (36) is that this mapping applied to $(B^{br}|L_1^{br} = \lambda)$ yields a continuum random forest distributed like $\mathcal{F}_\lambda$. But we do not see how to obtain this result more directly in the Brownian setting and so deduce Theorem 4.

4 The subordinator representation

In this section we prove Theorem 6. The key ingredient is the formula (40) for the splitting rate of the CRT.

4.1 The splitting rate of the CRT

Each point v of the skeleton $\cup_k \mathcal{R}(k)$ of a realization $\mathcal{T}$ of the CRT (except for the countable number of branch points) specifies a bipartition $\mathcal{T} \setminus \{v\} = B_1(v) \cup B_2(v)$ where the two components are written in random order. Thus the unit mass of $\mathcal{T}$ is split into masses $\mu(B_1(v))$ and $\mu(B_2(v)) = 1 - \mu(B_1(v))$. Choosing v according to the σ-finite length measure ℓ on $\cup_k \mathcal{R}(k)$ gives a σ-finite measure on $(0,1)$

$$s(\cdot|\mathcal{T}) := \ell\{v : \ \mu(B_1(v)) \in \cdot\}.$$

Now define unconditionally

$$s(\cdot) = Es(\cdot|\mathcal{T}).$$

Similarly, if $B^*(v)$ denotes the component of the split containing a leaf of $\mathcal{T}$ picked at random according to μ, that is a μ-mass-biased choice from $\{B_1(v), B_2(v)\}$, then we can define

$$s^*(\cdot) = E\ell\{v : \ \mu(B^*(v)) \in \cdot\}. \tag{38}$$

Lemma 14

$$s(dx) = (8\pi)^{-1/2} x^{-3/2} (1-x)^{-3/2} dx, \ 0 < x < 1. \tag{39}$$

$$s^*(dx) = (2\pi)^{-1/2} x^{-1/2} (1-x)^{-3/2} dx, \ 0 < x < 1. \tag{40}$$

Proof. We will prove (40), and then (39) follows from the relationship $s^*(dx) = 2x \ s(dx)$.

In the random tree $\mathcal{T}_n$, let $B_n^*(e)$ be the component containing vertex 1 when edge e is cut, and let B_n^* be the random component obtained by taking e to be a uniform random edge. An elementary counting argument based on Cayley's formula (there are n^{n-2} trees on vertices $\{1, 2, \ldots, n\}$) shows

$$P(\#B_n^* = a) = \frac{\binom{n-1}{a-1} a^{a-2} b^{b-2} \ ab}{n^{n-2} \ (n-1)}, \quad \text{where } b = n - a.$$

Using Stirling's formula we obtain

$$P(\#B_n^* = a) \sim n^{-3/2} (2\pi)^{-1/2} y^{-1/2} (1-y)^{-3/2} \text{ as } n \to \infty, \ a/n \to y. \tag{41}$$

Write ℓ_n for the measure assigning weight $n^{-1/2}$ to each edge, and set

$$s_n^*(\cdot | \mathcal{T}_n) = \ell_n \{ e : n^{-1} \# B_n^*(e) \in \cdot \}$$

$$s_n^*(\cdot) = E s_n^*(\cdot | \mathcal{T}_n).$$

Then (41) implies

$$s_n^* \text{ converges vaguely to } s^{**}$$

where s^{**} denotes the measure with density given by the formula (40). Lemma 9 implies that the joint distribution of $\mathcal{R}(n,k)$ and the midpoint of an edge chosen with measure ℓ_n restricted to $\mathcal{R}(n,k)$ converges vaguely (as $n \to \infty$) to the joint distribution of $\mathcal{R}(k)$ and a point chosen with measure ℓ restricted to $\mathcal{R}(k)$. Applying this with $k(n) \to \infty$ slowly, and then using Lemma 11 and (13), we see that the vague limit s^{**} is indeed the splitting rate s^* defined by (38).

4.2 Transition rates

Lemma 15 $(Y_1^*(\lambda), \lambda \geq 0)$ *is a decreasing process on* $(0,1]$ *with conditional jump rate density*

$$q^*(x,y) = (2\pi)^{-1/2} x^{3/2} y^{-1/2} (x-y)^{-3/2} \qquad (0 < y < x) \tag{42}$$

by which we mean that the process jumps from x *into* $[y, y+dy]$ *at rate* $q^*(x,y)dy$.

Proof. When $x = 1$ (that is, when $\lambda = 0$) the jump rate has density $s^*(\cdot)$ at (40), which coincides with the formula for $q^*(1,\cdot)$. The general case will be derived by scaling.

We first discuss scaling of the CRT $\mathcal{T}$. For $0 < c < \infty$ write $\mathcal{T}_c$ for the rescaled tree obtained from $\mathcal{T}$ by replacing the mass measure $\mu(\cdot)$ and the length measure $\ell(\cdot)$ by $\mu_c(\cdot) = c\mu(\cdot)$ and $\ell_c(\cdot) = c^{1/2}\ell(\cdot)$. To motivate the definition of $\mathcal{T}_c$, consider the uniform random tree on $j(n)$ vertices. When we assign mass $1/j(n)$ to each vertex and length $1/\sqrt{j(n)}$ to each edge, then (e.g. in the sense of Lemma 9) the random tree converges in distribution to $\mathcal{T}$. If instead we assign mass $1/n$ to each vertex and length $1/\sqrt{n}$ to each edge, where $j(n)/n \to c$, then the random tree converges in distribution to $\mathcal{T}_c$.

By the scaling construction, the instantaneous splitting rate $s_c^*(\cdot)$ of $\mathcal{T}_c$ analogous to (38) when $\mathcal{T}_c$ is split by a ℓ_c-distributed point is

$$\begin{aligned} s_c^*(dy) &= c^{1/2} s^*(dy'), \ y = cy' \\ &= c^{-1/2} q^*(1, y/c) \\ &= q^*(c,y). \end{aligned} \tag{43}$$

Regard $Y_1^*(\lambda)$ as the μ-mass of the tree-component $\mathcal{T}^*(\lambda)$ of the continuum forest $\mathcal{F}_\lambda$ containing a fixed μ-random atom. By (43), to show that $(Y_1^*(\lambda))$ is Markov with transition rates q^* it is enough to show:

$$\text{conditional on } (Y_1^*(\lambda'), 0 \leq \lambda' \leq \lambda) \text{ with } Y_1^*(\lambda) = c$$

$$\text{we have } \mathcal{T}^*(\lambda) \stackrel{d}{=} \mathcal{T}_c. \tag{44}$$

Now as the discrete analog, let $\mathcal{T}^{*n}(m)$ be the tree-component containing vertex 1 in the forest-valued process $(\mathcal{F}^n(m), n \geq m \geq 1)$ in the beginning of the Introduction. Conditional on $(\mathcal{F}^n(m'), n \geq m' \geq m)$ with $\#\mathcal{T}^{*n} = j$, and conditional on the vertex-set of $\mathcal{T}^{*n}(m)$ being A, it is clear that the random tree $\mathcal{T}^{*n}(m)$ is uniform on all trees with vertex-set A. By considering the limit (19) $n^{-1}\#\mathcal{F}^n(n-m(n)) \xrightarrow{d} \mathbf{Y}(\lambda)$ for $m(n)/n^{1/2} \to \lambda$, and taking $j(n)/n \to c$, these uniform conditioned distributions of rescaled $\mathcal{T}^{*n}(n-m(n))$ converge to $\mathcal{T}_c$, establishing (44).

Proof of Theorem 6. We prove the equivalent assertion that

$$S(\lambda) := \frac{1}{Y_1^*(\lambda)} - 1$$

is the stable subordinator of index 1/2. We first show that the jump rate density q for $S(\lambda)$ is

$$q(s,t) = (2\pi)^{-1/2}(t-s)^{-3/2} \qquad (0 < s < t) \tag{45}$$

which is the jump rate density for the subordinator. This is just a calculation. In terms of the rate $q^*(x,y)$ for $Y_1^*(\lambda)$,

$$q(s,t) = q^*(x,y)\left|\frac{dy}{dt}\right|$$

where

$$x = (1+s)^{-1},\ y = (1+t)^{-1},\ \left|\frac{dy}{dt}\right| = (1+t)^{-2}.$$

So using Lemma 15

$$q(s,t) = (2\pi)^{-1/2}(1+s)^{-3/2}(1+t)^{1/2}\left(\tfrac{1}{1+s} - \tfrac{1}{1+t}\right)^{-3/2}(1+t)^{-2}$$

which reduces to (45).

Thus the jumps of the processes $S(\lambda)$ and $S_{1/2}(\lambda)$ form the same Poisson point process. Lemma 15 does not preclude the possibility that $Y_1^*(\lambda)$ and hence $S(\lambda)$ might have an additional (monotone) drift term. But this is not possible because (33) and (6) imply the identity of marginal distributions $S(\lambda) \stackrel{d}{=} S_{1/2}(\lambda)$ for each λ.

5 Asymptotics as $\lambda \to \infty$

We start by proving Theorem 7, and then briefly describe the behavior of the largest cluster.

Recall from Section 3.1 that the J_i in Theorem 4 are the ranked points of a PRM on $(0,\infty)$ whose intensity measure is the Lévy measure $\rho(dx) := (2\pi)^{-1/2}\, x^{-3/2} dx$ of the stable(1/2) subordinator. The image of $\rho(dx)$ via the map $h(x) = \lambda^4 x$ is $\lambda^2 \rho(dh)$. Let $H_i(\lambda^2) := \lambda^4 J_i$ to see that Theorem 4 can be rewritten as

$$\lambda^2 \mathbf{Y}(\lambda) \stackrel{d}{=} (H_1(\lambda^2), H_2(\lambda^2), \ldots \mid \sum_i H_i(\lambda^2) = \lambda^2) \tag{46}$$

where $H_i(\lambda^2)$ is the ith largest point in a PRM on $(0,\infty)$ with intensity $\lambda^2 \rho(dh)$. Moreover, by a well known change of measure formula for Poisson processes [22], formula (46) holds also when $H_i(\lambda^2)$ is the ith largest point in a PRM on $(0,\infty)$ with intensity $\lambda^2 e^{-bh}\rho(dh)$ for any fixed $b \geq 0$. Choosing $b = 1/2$, this intensity measure is $\lambda^2 e^{-h/2}\rho(dh) = \lambda^2 \chi(dh)$ for χ as in Theorem 7. For this choice of b we find that

$$E(\sum_i H_i(\lambda)) = \lambda^2 \int_0^\infty h\, \chi(dh) = \lambda^2$$

and

$$\mathrm{Var}(\sum_i H_i(\lambda)) = \lambda^2 \int_0^\infty h^2\, \chi(dh) = \lambda^2.$$

As $\lambda \to \infty$, the central limit theorem implies that the asymptotic distribution of $(\sum_i H_i(\lambda^2) - \lambda^2)/\lambda$ is standard normal, and the local limit theorem gives convergence of densities. It follows that the asymptotic behavior of the sequence $\lambda^2 \mathbf{Y}(\lambda)$ for large λ can be read from classical (cf. [11, 21]) conditioned limit theorems for independent random variables.

Theorem 7 concerns the empirical measure

$$F_\lambda(\cdot) = \sum_i 1(\lambda^2 Y_i(\lambda) \in \cdot).$$

Consider

$$F^*_\lambda(\cdot) = \sum_i 1(H_i(\lambda^2) \in \cdot)$$

so F^*_λ is a PRM with intensity measure $\lambda^2\chi$. For a measure μ on $(0,\infty)$ with $\int_0^\infty x\mu(dx) < \infty$ let $\bar\mu$ be the finite measure $\bar\mu(dx) := x\mu(dx)$, so $\bar\mu(0,h] := \int_{(0,h]} x\mu(dx)$. By the invariance principle, there is the convergence in distribution of processes on $D[0,\infty)$

$$\frac{1}{\lambda}(\bar F^*_\lambda(0,h] - \lambda^2\bar\chi(0,h], h \geq 0) \xrightarrow{d} B\left(\int_0^h x^2\chi(dx),\ h \geq 0\right) \tag{47}$$

where $(B(t), t \geq 0)$ is a standard Brownian motion. We assert that convergence still holds after conditioning that the $h = \infty$ values of the processes equal 0, i.e. we assert that

$$\frac{1}{\lambda}(\bar F_\lambda(0,h] - \lambda^2\bar\chi(0,h], h \geq 0) \xrightarrow{d} B^{br}\left(\int_0^h x^2\chi(dx),\ h \geq 0\right) := \bar G[0,h] \tag{48}$$

where $(B^{br}(t), 0 \leq t \leq 1)$ is a standard Brownian bridge. Though we cannot find a precise reference for such a result, the methods are standard. Check the hypotheses of the local CLT in the setting (47); use the conclusion of the local CLT to establish convergence of finite dimensional distributions in (48); check tightness by using a sufficient condition such as [1]. We omit details.

From the familiar variances and covariances of Brownian bridge

$$\begin{aligned} \text{Var } B^{br}(dh) &= dh \\ EB^{br}(dh_1)B^{br}(dh_2) &= -dh_1dh_2 \end{aligned}$$

we get

$$\begin{aligned} \text{Var}\bar G(dh) &= h^2\chi(dh) \\ E\bar G(dh_1)\bar G(dh_2) &= -h_1^2h_2^2\chi(dh_1)\chi(dh_2). \end{aligned}$$

Now (48) is equivalent to the convergence $G_\lambda(\cdot) \xrightarrow{d} G(\cdot)$ asserted in Theorem 7, where G is defined via its density

$$\bar G(dx) = xG(dx).$$

Clearly G has covariances specified at (11).

5.1 The largest cluster

It is not hard to show that for large λ the conditioning in (46) has a negligible effect on the joint distribution of the first few terms. This leads to the result that, as $\lambda \to \infty$,

$$P(\lambda^2 Y_1(\lambda) \le h) = \exp\left(-\lambda^2 \int_h^\infty \chi(dx)\right) + o(1), \text{ uniformly in } h. \tag{49}$$

Routine calculations, (or comparison with classical i.i.d. extreme value theory [15, p.158, Ex. 24]), then show

$$\frac{\lambda^2}{2} Y_1(\lambda) - a(\lambda) \xrightarrow{d} \xi, \quad \text{where } P(\xi \le x) = \exp(-e^{-x}) \tag{50}$$

where the centering constant $a(\lambda)$ is the solution of

$$\frac{\lambda^2}{2\sqrt{\pi}} \int_{a(\lambda)}^\infty y^{-3/2} e^{-y}\ dy = 1$$

which is

$$a(\lambda) = \log \frac{\lambda^2}{2\sqrt{\pi}} - \frac{3}{2} \log\log \frac{\lambda^2}{2\sqrt{\pi}} + o(1).$$

From (50) we deduce easily that

$$\lim_{\lambda\to\infty} \frac{Y_1(\lambda)}{\lambda^{-2}\log\lambda} = 4$$

in the sense of convergence in probability, and in fact this can be sharpened to give a.s. convergence. Rather than give the details of (49,50), let us give a simple alternate proof of an upper bound.

Lemma 16

$$\limsup_{\lambda\to\infty} \frac{Y_1(\lambda)}{\lambda^{-2}\log\lambda} \le 4 \ a.s.$$

Proof. By size-biasing, $P(Y_1^*(\lambda) \ge y) \ge y P(Y_1(\lambda) \ge y)$, and so

$$\begin{aligned} P(Y_1(\lambda) \ge y) &\le y^{-1} P\left(\frac{Z^2}{\lambda^2 + Z^2} \ge y\right) \\ &\le y^{-1} \Phi(\lambda y^{1/2}) \\ &\le (2\pi)^{-1/2} \lambda^{-1} y^{-3/2} \exp(-\lambda^2 y/2). \end{aligned}$$

So for fixed $\varepsilon > 0$ we have

$$P(Y_1(\lambda) \geq (4+3\varepsilon)\lambda^{-2} \log \lambda) = O(\lambda^{-\varepsilon}) \text{ as } \lambda \to \infty$$

and a routine Borel-Cantelli argument through $\lambda_j = (1+\delta)^j$ leads to the lemma.

6 Final Remarks

6.1 Further distributional properties

Recall that $Y_i^\dagger(\lambda)$ is the mass of the tree component of the random forest that contains U_i when the cutting intensity is λ, for independent U_i picked according to the mass distribution on leaves of the CRT. Write $\Pi_n(\lambda)$ for the partition of $[n] := \{1, \ldots, n\}$ generated by the values of $Y_i^\dagger(\lambda), i \in [n]$. One can derive a formula for the joint distribution of $Y_i^\dagger(\lambda), i \in [n]$, extending the $n = 1$ case (8), and it is then possible by integration to obtain a formula for the distribution of $\Pi_n(\lambda)$. These results, which refine those of Pitman [29], and related descriptions of the processes $((Y_i(\lambda), i \in [n]), 0 \leq \lambda < \infty)$, and $(\Pi_n(\lambda), 0 \leq \lambda < \infty)$, will be treated in a subsequent paper.

6.2 The additive Marcus-Lushnikov process

The scientific literature [7] on mean-field models for stochastic coalescence of mass focuses on the *Marcus-Lushnikov process*, which is the continuous-time finite-state Markov chain in which n unit-mass atoms merge into clusters according to the rule

> for each pair of clusters, of masses $\{x_i, x_j\}$ say, they merge into a cluster of mass $x_i + x_j$ at rate $K(x_i, x_j)/n$

where $K(x, y)$ is a specified rate kernel. The particular case of the *additive* kernel $K(x, y) = x + y$ is one of several tractable kernels, and formulae for the time-t distribution have been given by Lushnikov [23] and Hendriks et al. [20]. The additive Marcus-Lushnikov process may be constructed as $\#\bar{\mathcal{F}}^n(t)$, in the notation of the Introduction. The qualitative import of Theorem 3 is that the transition from clusters being all of size $o(n)$ to a single cluster containing mass $n - o(n)$ occurs over the *critical* time-interval $\frac{1}{2} \log n \pm O(1)$. (This fact isn't stated clearly in the scientific literature, though Tanaka and Nakazawa [33] make related assertions.) Quantitatively, the sizes of the largest clusters at time $\frac{1}{2}\log n + t$ are asymptotically like the (random) sizes $n\mathbf{X}(t)$. So our distributional results for the standard additive coalescent are immediately interpretable as $n \to \infty$ limits in the additive Marcus-Lushnikov process over the critical time-interval.

One could alternatively consider the *initial* time-period (where $t = O(1)$ as $n \to \infty$) and the *intermediate* time-period (where $t \to \infty$ and $\frac{1}{2}\log n - t \to \infty$). On the initial period, the distribution of the cluster at time t containing a given atom tends to the Borel$(1 - e^{-t})$ distribution, where the Borel(μ) distribution is the total progeny in the Galton-Watson branching process with Poisson(μ) offspring:

$$B_\mu(i) = \frac{(\mu i)^{i-1}}{i!} e^{-\mu i}, \;\; i = 1, 2, \ldots$$

This may readily be deduced from the construction of $\bar{\mathcal{F}}^n(t)$, and leads to an asymptotic result for the number $N_n(i,t)$ of size-i clusters at time t: as $n \to \infty$

$$n^{-1} E N_n(i,t) \to n(i,t) := i^{-1} B_{1-e^{-t}}(i), i = 1, 2, \ldots. \tag{51}$$

The scientific literature (surveyed in [7]) historically started by analytic derivation of the formula for $n(i,t)$ as the solution of the deterministic *Smoluchowsky coagulation equations*. This derivation goes back at least to Golovin [19]; see [34] sec. A1 or [32] for recent treatments and further references. The explicit interpretation of $n(i,t)$ as a limit (51) of the (stochastic) additive Marcus-Lushnikov process was studied in most detail by van Dongen [34]. Equation (3.3) of that paper presents a simple formula, analogous to (11), for $\mathrm{cov}(N_n(i,t), N_n(j,t))$. However, its derivation ([34] sec. 2) is via "Ω-expansion of the master equation ...... up to [second-order] terms ...... the higher terms are neglected", which we interpret as tantamount to an *assumption* that $(N_n(i,t); i \geq 1)$ is Gaussian. The upshot is that, while it seems intuitively clear that in the initial and intermediate time-periods the limit result (51) may be refined to a Gaussian limit result, this has not been rigorously proved.

6.3 Comparisons with the standard multiplicative coalescent, and entrance boundary

We list similarities and differences between the standard additive coalescent $(\mathbf{X}(t))$ introduced here and the standard *multiplicative* coalescent, $(\mathbf{Z}(t))$ say, studied in Aldous [6].

- The natural time parameter set for both process is $(-\infty, \infty)$.
- Analogous to Proposition 2, $(\mathbf{Z}(t))$ also arises as a weak limit of a simple discrete process, the process of component sizes in the random graphs $\mathcal{G}(n, \lambda/n)$.
- $\mathbf{Z}(t)$ takes values in l_2 rather than l_1; its total mass is infinite.
- The distribution of $\mathbf{Z}(t)$ for fixed t can be described in terms of excursion lengths in a Brownian-type process; cf. the Brownian excursion construction of the CRT.
- But the distribution of $\mathbf{Z}(t)$ does not seem to permit such explicit formulas as those implied by Theorem 4.

Looking at results about the multiplicative coalescent suggests two specific conjectures about the additive coalescent, as follows. The first (cf. [6] Proposition 4) concerns the non-uniform analog of Proposition 2.

Conjecture 1 *Let* $\mathbf{v}^n \in \Delta$ *satisfy*
(i) $s^n := \sum_i (v_i^n)^2 \to 0$
(ii) $v_1^n / s^n \to 0$.
Let $(\mathbf{X}^n(t), \frac{1}{2}\log s^n \leq t < \infty)$ *be the additive coalescent started at time* $\frac{1}{2}\log s^n$ *from configuration* $\mathbf{v}^n$. *Then* $\mathbf{X}^n(\cdot) \xrightarrow{d} \mathbf{X}^\infty(\cdot)$.

The second concerns the *entrance boundary*, that is the set of extreme distributions for an additive coalescent $(\widetilde{\mathbf{X}}(t), -\infty < t < \infty)$. One might guess that the standard additive coalescent was essentially the only such process, but we know this is false. Let

$$\Delta_2 := \{\mathbf{x} = (x_i) : x_1 \geq x_2 \geq \ldots \geq 0, \sum_i x_i^2 < \infty\}$$

.

Conjecture 2 (cf. [8]). *For each $-\infty < \tau < \infty$ and each $\mathbf{x} \in \Delta_2$ there exists a unique distribution for an additive coalescent $(\widetilde{\mathbf{X}}(t), -\infty < t < \infty)$ such that*
(i) $e^{-2(t-\tau)} \sum_i (\widetilde{X}_i(t))^2 \to 1$ *a.s. as* $t \to -\infty$
(ii) $e^{-2(t-\tau)} \widetilde{X}_i(t) \to x_i, \ i \geq 1$ *a.s. as* $t \to -\infty$.

This set of distributions, together with the distribution of the constant process $(1, 0, 0, \ldots)$*, is exactly the set of extreme distributions for an additive coalescent paramaterized by* $(-\infty, \infty)$.

One can show that the standard additive coalescent is the case where $\tau = 0$ and $\mathbf{x} = \mathbf{0}$ (the latter is a consequence of Lemma 16). In general τ is just a time-shift parameter. Heuristically, we expect the process $(\widetilde{\mathbf{X}}(t))$ with $\mathbf{x} = (x_1, \ldots, x_k)$ to arise as a weak limit (analogous to Proposition 2) $\mathbf{X}^n(\cdot) \xrightarrow{d} \widetilde{\mathbf{X}}(\cdot)$ where $\mathbf{X}^n$ is the additive coalescent starting at a suitable time from initial configuration $(x_1 n^{-1/2}, \ldots, x_k n^{-1/2}, n^{-1}, \ldots, n^{-1}, 0, 0, \ldots)$.

References

[1] D.J. Aldous. Stopping times and tightness II. *Ann. Probab.*, 17:586–595, 1989.

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

[3] D.J. Aldous. The continuum random tree II: an overview. In M.T. Barlow and N.H. Bingham, editors, *Stochastic Analysis*, pages 23–70. Cambridge University Press, 1991.

[4] D.J. Aldous. The continuum random tree III. *Ann. Probab.*, 21:248–289, 1993.

[5] D.J. Aldous. Recursive self-similarity for random trees, random triangulations and Brownian excursion. *Ann. Probab.*, 22:527–545, 1994.

[6] D.J. Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25:812–854, 1997.

[7] D.J. Aldous. Deterministic and stochastic models for coalescence: a review of the mean-field theory for probabilists. Unpublished. Available via homepage http://www.stat.berkeley.edu/users/aldous, 1997.

[8] D.J. Aldous and V. Limic. The entrance boundary of the multiplicative coalescent. In Preparation, 1997.

[9] D.J. Aldous and J. Pitman. Brownian bridge asymptotics for random mappings. *Random Structures Algorithms*, 5:487–512, 1994.

[10] P. Donnelly and P. Joyce. Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex. *Stochastic Process. Appl.*, 31:89–103, 1989.

[11] M. Dwass and S. Karlin. Conditioned limit theorems. *Ann. Math. Stat.*, 34:1147–1167, 1963.

[12] E.B. Dynkin and S.E. Kuznetsov. Markov snakes and superprocesses. *Probab. Th. Rel. Fields*, 103:433–473, 1995.

[13] S.N. Evans and J. Pitman. Construction of Markovian coalescents. Technical Report 465, Dept. Statistics, U.C. Berkeley, 1996.

[14] S.N. Evans and J. Pitman. Stationary Markov processes related to stable Ornstein-Uhlenbeck processes and the additive coalescent. Technical Report 488, Dept. Statistics, U.C. Berkeley, 1997.

[15] J. Galambos. *The Asymptotic Theory of Extreme Order Statistics.* Robert E. Krieger Pub. Co., Malabar, Florida, 1987.

[16] J.-F. Le Gall. Brownian excursions, trees and measure-valued branching processes. *Ann. Probab.*, 19:1399–1439, 1991.

[17] J.-F. Le Gall. The uniform random tree in a Brownian excursion. *Probab. Th. Rel. Fields*, 96:369–384, 1993.

[18] J.-F. Le Gall. The Brownian snake and solutions of $\Delta u = u^2$ in a domain. *Probab. Th. Rel. Fields*, 102:393–432, 1995.

[19] A.M. Golovin. The solution of the coagulating equation for cloud droplets in a rising air current. *Izv. Geophys. Ser.*, 5:482–487, 1963.

[20] E.M. Hendriks, J.L. Spouge, M. Eibl, and M. Shreckenberg. Exact solutions for random coagulation processes. *Z. Phys. B - Condensed Matter*, 58:219–227, 1985.

[21] F. Knight. A pair of conditional normal convergence theorems. *Theory Probab. Appl.*, 11:191–210, 1966.

[22] U. Küchler and S. L. Lauritzen. Exponential families, extreme point models and minimal space-time invariant functions for stochastic processes with stationary and independent increments. *Scandinavian Journal of Statistics*, 16:237 – 261, 1989.

[23] A.A. Lushnikov. Coagulation in finite systems. *J. Colloid and Interface Science*, 65:276–285, 1978.

[24] Yu.V. Pavlov. A case of the limit distribution of the maximal volume on a tree in a random forest. *Mathematical Notes of the Acad. Sci. USSR*, 25:387–392, 1979.

[25] M. Perman. Order statistics for jumps of normalized subordinators. *Stoch. Proc. Appl.*, 46:267–281, 1993.

[26] M. Perman, J. Pitman, and M. Yor. Size-biased sampling of Poisson point processes and excursions. *Probab. Th. Rel. Fields*, 92:21–39, 1992.

[27] J. Pitman. Coalescent random forests. Technical Report 457, Dept. Statistics, U.C. Berkeley, 1996.

[28] J. Pitman. Random discrete distributions invariant under size-biased permutation. *Adv. Appl. Prob.*, 28:525–539, 1996.

[29] J. Pitman. Partition structures derived from Brownian motion and stable subordinators. *Bernoulli*, 3:1–18, 1997.

[30] J. Pitman and M. Yor. Arcsine laws and interval partitions derived from a stable subordinator. *Proc. London Math. Soc. (3)*, 65:326–356, 1992.

[31] J. Pitman and M. Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25:855–900, 1997.

[32] R.K. Sheth and J. Pitman. Coagulation and branching process models of gravitational clustering. To appear in *Mon. Not. R. Astron. Soc.*, 1997.

[33] H. Tanaka and K. Nakazawa. Validity of the stochastic coagulation equation and runaway growth of protoplanets. *Icarus*, 107:404–412, 1994.

[34] P.G.J. van Dongen. Fluctuations in coagulating systems II. *J. Statist. Phys.*, 49:927–975, 1987.

[35] A. Yao. On the average behavior of set merging algorithms. In *Proc. 8th ACM Symp. Theory of Computing*, pages 192–195, 1976.