J. Appl. Prob. **50,** 772–790 (2013)
Printed in England

A CONSTRUCTION OF A β-COALESCENT VIA THE PRUNING OF BINARY TREES

ROMAIN ABRAHAM,* *Université d'Orléans*
JEAN-FRANÇOIS DELMAS,** *École des Ponts et Chaussées*

Abstract

Considering a random binary tree with n labelled leaves, we use a pruning procedure on this tree in order to construct a $\beta(\frac{3}{2}, \frac{1}{2})$-coalescent process. We also use the continuous analogue of this construction, i.e. a pruning procedure on Aldous's continuum random tree, to construct a continuous state space process that has the same structure as the β-coalescent process up to some time change. These two constructions enable us to obtain results on the coalescent process, such as the asymptotics on the number of coalescent events or the law of the blocks involved in the last coalescent event.

Keywords: Coalescent process; binary tree; pruning; continuum random tree

2010 Mathematics Subject Classification: Primary 60J25
Secondary 60J80

1. Introduction

Let Λ be a finite measure on $[0, 1]$. A Λ-coalescent $(\Pi(t),\ t \geq 0)$ is a Markov process which takes values in the set of partitions of $\mathbb{N}^* = \{1, 2, \dots\}$ introduced in [24]. It is defined via the transition rates of its restriction $(\Pi^{[n]}(t),\ t \geq 0)$ to the first n integers: if $\Pi^{[n]}(t)$ is composed of b blocks then k $(2 \leq k \leq b)$ fixed blocks coalesce at rate

$$\lambda_{b,k} = \int_0^1 u^{k-2}(1-u)^{b-k}\Lambda(\mathrm{d}u).$$

In particular, a coalescent event arrives at rate

$$\lambda_b = \sum_{k=2}^{b} \binom{b}{k} \lambda_{b,k}.$$

As examples of Λ-coalescents, we cite Kingman's coalescent, $\Lambda(\mathrm{d}x) = \delta_0(\mathrm{d}x)$ (see [22]), the Bolthausen–Sznitman coalescent, $\Lambda(\mathrm{d}x) = \mathbf{1}_{(0,1)}(x)\,\mathrm{d}x$ (see [13]), and β-coalescents, $\Lambda(\mathrm{d}x)$ is the $\beta(2-\alpha, \alpha)$ distribution with $0 < \alpha < 2$ (see [10] and [12]) or the $\beta(2-\alpha, \alpha-1)$ distribution with $1 < \alpha < 2$ (see [16]). We refer the reader to the survey [9] for further results on coalescent processes.

The goal of this paper is to give a new representation for the $\beta(\frac{3}{2}, \frac{1}{2})$-coalescent using the pruning of random binary trees. This kind of idea has already been used in [17] where the Bolthausen–Sznitman coalescent is constructed via the cutting of a random recursive tree.

Received 29 June 2012; revision received 21 November 2012.
This work was partially supported by the 'Agence Nationale de la Recherche', under grant ANR-08-BLAN-0190.
* Postal address: Laboratoire MAPMO, CNRS, UMR 6628, Fédération Denis Poisson, FR 2964, Université d'Orléans, B.P. 6759, 45067 Orléans cedex 2, France. Email address: romain.abraham@univ-orleans.fr
** Postal address: Université Paris-Est, École des Ponts, CERMICS, 6-8 Avenue Blaise Pascal, Champs-sur-Marne, 77455 Marne La Vallée, France.

1.1. Pruning of binary trees

Now we describe the coalescent associated with the pruning of the random binary tree. We first recall the normalization constant in the beta distribution: for $a > 0$ and $b > 0$,

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_{(0,1)} x^{a-1}(1-x)^{b-1}\,\mathrm{d}x.$$

Fix an integer n, and consider a uniform, random ordered binary tree with n leaves ($2n-1$ vertices plus a root and $2n-1$ edges). Label these leaves from 1 to n uniformly at random. After an exponential time of parameter $\lambda_n = (n-1)\beta(\frac{1}{2}, n-\frac{1}{2})$, choose one of the $n-1$ inner vertices of the tree uniformly at random, coalesce all the leaves of the subtree attached at the chosen node, and remove that subtree from the original tree. Then restart the process with the resulting tree (using λ_k as the new parameter of time, where k is the new number of leaves) until all the leaves coalesce into a single leaf. In Figure 1 we give an example of such a coalescence for $n = 5$.

The above defines a process $(\Pi^{[n]}(t),\ t \geq 0)$. The main result of the paper is the following theorem.

Theorem 1. *The process* $(\Pi^{[n]}(t),\ t \geq 0)$ *defined as the coalescent associated with the pruning of the random binary tree is the restriction to* $\{1, \ldots, n\}$ *of a* $\beta(\frac{3}{2}, \frac{1}{2})$*-coalescent with coalescent measure*

$$\Lambda(\mathrm{d}u) = \sqrt{\frac{u}{1-u}}\,\mathrm{d}u. \tag{1}$$

Let us remark that the β-coalescents introduced in [12] and usually studied are $\beta(2-\alpha, \alpha)$-coalescents with $1 < \alpha < 2$. Here we have $\alpha = \frac{1}{2}$, which is not covered by the usual case. The reason is that, for $1 < \alpha < 2$, the β-coalescent comes down from ∞ ($\Pi(t)$ has almost surely (a.s.) a finite number of blocks for any $t > 0$), which is not the case for $\alpha = \frac{1}{2}$ according

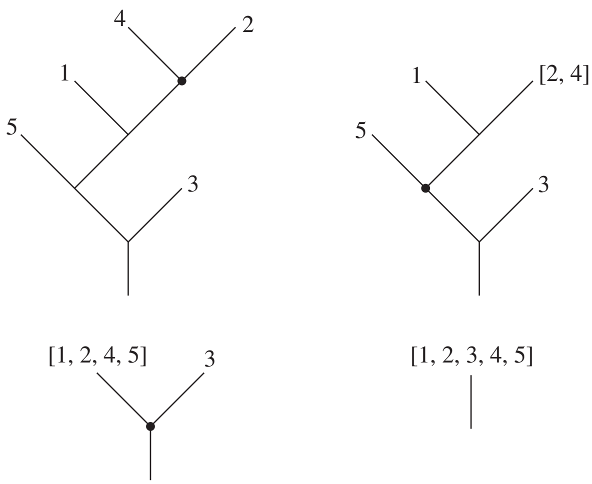

Figure 1: An example of the coalescence of a binary tree for $n = 5$. The circles represent the selected node, and the times between each picture are exponentially distributed with respective parameters λ_5, λ_4, and λ_2.

to the criterion from [9, Theorem 3.5] or [24]:

$$\int_0^1 \frac{\Lambda(\mathrm{d}u)}{u} < +\infty.$$

1.2. Pruning of Aldous's continuum random tree

A coalescent process may also be viewed as a process $(\mathcal{I}(t),\ t \geq 0)$ taking values in the interval partitions of $[0, 1]$. The length of each interval represents the mass of a block, and the process represents blocks (whose sizes sum to 1) that merge together as time passes. We can go from this interval-partition-valued process to the previous Λ-coalescent framework by the classical paintbox procedure (see, for instance, [25]): consider $(U_i)_{i\in\mathbb{N}}$, a sequence of independent and identically distributed (i.i.d.) uniform random variables on $[0, 1]$, independent of the process $\mathcal{I}$ and, for every $t \geq 0$, say that the integers i and j belong to the same block in $\Pi(t)$ if and only if U_i and U_j belong to the same interval in $\mathcal{I}(t)$.

The continuous analogue of the binary tree is Aldous's continuum random tree (CRT) which can be obtained as the limit (in an appropriate sense) of the rescaled uniform binary tree with n leaves when the number of leaves n tends to ∞. A pruning theory of such a continuum tree has been introduced in [3] (see [5] for a general theory of the pruning of Lévy trees) and will be recalled in Section 3.1. Using this pruning procedure, we are able to define an interval-partition-valued process in Section 3.2 which has the same structure as the $\beta(\frac{3}{2}, \frac{1}{2})$-coalescent except for the times (when sampling n points uniformly distributed on $(0, 1)$, the time interval between two coalescences is not exponentially distributed). However, we conjecture that an appropriate change of time would be enough so that the interval-partition-valued process is really associated with the $\beta(\frac{3}{2}, \frac{1}{2})$-coalescent via the paintbox procedure. This gives a nice interpretation of the dust (or fraction of singletons) in the $\beta(\frac{3}{2}, \frac{1}{2})$-coalescent.

1.3. Number of coalescent events and the last coalescent event

The construction using discrete trees allows us to recover in Section 4.1 the asymptotic distribution of the number of coalescent events given in [18] in a more general framework; see also [20].

Proposition 1. *Let X'_n be the number of collisions undergone by $(\Pi^{[n]}(t),\ t \geq 0)$. Then we have*

$$\frac{X'_n}{\sqrt{n}} \xrightarrow{\mathrm{D}} \sqrt{2}Z \quad \text{as } n \to \infty,$$

where Z has a Rayleigh distribution with density $x\mathrm{e}^{-x^2/2}\,\mathbf{1}_{\{x>0\}}$.

Let us remark that, according to Section 5.4 of [18], Z is distributed as

$$\frac{2}{\sqrt{\pi}} \int_0^\infty \mathrm{d}t\,\mathrm{e}^{-S_t/2}$$

with $(S_t,\ t \geq 0)$ a subordinator with $\mathbb{E}[\exp(-\lambda S_t)] = \exp(-t\Phi(\lambda))$, where, for $\lambda > 0$,

$$\begin{aligned}\Phi(\lambda) &= \int_0^1 (1-(1-u)^\lambda)\frac{\Lambda(\mathrm{d}u)}{u^2}\\ &= \int_0^1 (1-(1-u)^\lambda)u^{-3/2}(1-u)^{-1/2}\,\mathrm{d}u\\ &= 2\sqrt{\pi}\frac{\Gamma(\lambda+1/2)}{\Gamma(\lambda)}.\end{aligned}$$

(See [11] for more results on exponentials of Lévy processes.)

The continuum tree construction allows us to study the last coalescent event (see [17] for a similar study of the Bolthausen–Sznitman coalescent, and [14] for stationary, continuous-state branching processes). In Section 4.2 we consider the number E_n of external branches or singletons involved in the last coalescent event as well as the number of blocks B_n involved in the last coalescent event.

Proposition 2. *Let B_n be the number of blocks, and let E_n be the number of singletons involved in the last coalescent event of $(\Pi^{[n]}(t),\ t \geq 0)$. Then we have*

$$(B_n, E_n) \xrightarrow{\mathrm{D}} (B, E) \quad \text{as } n \to \infty,$$

where $(B-E, E)$ are finite random variables with generating function Φ given for $\rho, \rho_* \in [0, 1]$ by

$$\begin{aligned}\Phi(\rho, \rho_*) &= \mathbb{E}[\rho^{B-E} \rho_*^E] \\ &= \rho\biggl(1 + \log(2) - \log\biggl(1 + \sqrt{1-\rho_*} - \frac{\rho + \rho_*}{2}\biggr)\biggr) \\ &\quad - \sqrt{\rho}\log\biggl(\frac{1 + \sqrt{1-\rho_*} + \sqrt{\rho}}{1 + \sqrt{1-\rho_*} - \sqrt{\rho}}\biggr).\end{aligned}$$

Furthermore, $B - E$ is stochastically less than (or equal to) $E + 1$.

Note that $\Phi(1, 1) = 1$, which indeed implies that B and E are finite, that is, B_n and E_n are of order 1. However, $B - E$ and E have infinite expectation, as can be checked from their generating functions given below.

The generating function of E is given by

$$\mathbb{E}[\rho_*^E] = \Phi(1, \rho_*) = 1 - 2\log\biggl(1 + \frac{\sqrt{1-\rho_*}}{2}\biggr),$$

and the generating function of $B - E$ is given by

$$\mathbb{E}[\rho^{B-E}] = \Phi(\rho, 1) = \rho(1 + \log(4) - \log(1-\rho)) - \sqrt{\rho}\log\biggl(\frac{1+\sqrt{\rho}}{1-\sqrt{\rho}}\biggr).$$

There is a nice interpretation of the distribution of $B - E - 1$ given after Proposition 4 below. The generating function of B is given by

$$\begin{aligned}\mathbb{E}[\rho^B] &= \Phi(\rho, \rho) \\ &= \rho(1 + \log(2) - \log(1 + \sqrt{1-\rho} - \rho)) - \sqrt{\rho}\log\biggl(\frac{1 + \sqrt{1-\rho} + \sqrt{\rho}}{1 + \sqrt{1-\rho} - \sqrt{\rho}}\biggr).\end{aligned}$$

Of course, we have $B \geq 2$ a.s.

Remark 1. We can compute various quantities related to E and B. We have

$$\begin{array}{lll} \mathbb{P}(E=0) = 1 - 2\log\bigl(\frac{3}{2}\bigr) \simeq 0.19, & \mathbb{P}(B-E=0) = 0, & \mathbb{P}(B=0) = 0, \\ \mathbb{P}(E=1) = \frac{1}{3}, & \mathbb{P}(B-E=1) = \log(4) - 1 \simeq 0.39, & \mathbb{P}(B=1) = 0, \\ \mathbb{P}(E=2) = \frac{1}{9}, & \mathbb{P}(B-E=2) = \frac{1}{3}, & \mathbb{P}(B=2) = \frac{5}{12}, \\ \mathbb{P}(E=3) = \frac{23}{160}. & & \end{array}$$

In particular, we have $\mathbb{P}(E > 5) \leq 25\%$, $\mathbb{P}(B > 5) \leq 32\%$, and $\mathbb{P}(B - E > 5) \leq 11\%$.

The paper is organized as follows. The proof of Theorem 1 is given in Section 2. The link with Aldous's CRT, presented in Section 3.1, is given in Section 3.2 using a pruning procedure; the reduced subtrees are presented in Section 3.3. A proof and a comment on Proposition 1 are given in Section 4.1. Proposition 2 is proved in Sections 4.2.2 and 5.

2. The $\beta(\frac{3}{2}, \frac{1}{2})$-coalescent

In order to prove Theorem 1, we first need to compute the rates $\lambda_{n,k}$ at which k given blocks among n blocks coalesce. To this end, we use the following proposition.

Proposition 3. *For the coalescent of the random binary tree, we have, for any $2 \le k \le n$,*

$$\lambda_{n,k} = \beta\big(k - \tfrac{1}{2}, n - k + \tfrac{1}{2}\big).$$

We recall the duplication formula for $a > 0$:

$$\Gamma\left(a + \frac{1}{2}\right) = \frac{\sqrt{\pi}}{2^{2a-1}} \frac{\Gamma(2a)}{\Gamma(a)}.$$

Proof of Proposition 3. Let us first remark that, by construction, since there are $n-1$ internal vertices, we have

$$\lambda_{n,n} = \frac{\lambda_n}{n-1} = \beta\left(n - \frac{1}{2}, \frac{1}{2}\right) = \sqrt{\pi}\frac{\Gamma(n-1/2)}{\Gamma(n)}.$$

It is well known that the number of ordered binary trees with n leaves is given by the Catalan numbers

$$b_n = \frac{1}{n}\binom{2n-2}{n-1} = \frac{(2n-2)!}{(n-1)!\,n!}.$$

Hence, the number of ordered binary trees with n labelled leaves is

$$C_n = n!\,b_n = \frac{(2n-2)!}{(n-1)!} = \frac{2^{2n-2}}{\sqrt{\pi}}\Gamma\left(n - \frac{1}{2}\right). \tag{2}$$

Consider a binary tree with n labelled leaves. Let $2 \le k \le n$. Fix k labels, say the first k. For these labels to coalesce at the same time, the leaves with these k labels must lie exactly in a single subtree of the initial tree. Therefore, to construct such a tree, we must choose an ordered binary tree with k leaves labelled from 1 to k (C_k possibilities), choose an ordered binary tree with $n-k+1$ leaves labelled from k to n (C_{n-k+1} possibilities), and graft the tree with k leaves at the leaf labelled k. Then, for the first k labels to coalesce, the chosen branch must be the branch that links the two subtrees (and each branch is cut at rate $\lambda_{n,n}$). Therefore, we have

$$\begin{aligned}
\lambda_{n,k} &= \frac{C_k C_{n-k+1}}{C_n}\lambda_{n,n} \\
&= \frac{2^{2k-2}}{\sqrt{\pi}}\Gamma\left(k - \frac{1}{2}\right)\frac{2^{2n-2k}}{\sqrt{\pi}}\Gamma\left(n - k + \frac{1}{2}\right)\frac{\sqrt{\pi}}{2^{2n-2}}\frac{1}{\Gamma(n-1/2)}\sqrt{\pi}\frac{\Gamma(n-1/2)}{\Gamma(n)} \\
&= \beta\big(k - \tfrac{1}{2}, n - k + \tfrac{1}{2}\big).
\end{aligned}$$

This completes the proof.

This proves that the process evolves like a Λ-coalescent with Λ given by (1) up to the time of the first merger. To complete the proof of Theorem 1, it remains to prove that, after that first merger, the resulting tree is still a uniform binary tree with uniform labelled leaves.

Let us fix $k \le n$, and let T_k be a tree with k labelled leaves, one being labelled by the block $[i_1, \dots, i_{n-k+1}]$, the others being labelled by singletons. We want to compute the probability of obtaining that tree after the first merger. From this tree, we construct a tree with n leaves by grafting, on the leaf of T_k labelled by the block $[i_1, \dots, i_{n-k+1}]$, a tree with $n-k+1$ leaves labelled by $\{i_1, \dots, i_{n-k+1}\}$. There are exactly C_{n-k+1} different trees (this corresponds to the choice of the grafted tree). Moreover, the tree T_k is obtained after the first merger if the original tree is one of those, and if the chosen internal node is the leaf labelled by the block. Hence, the probability of obtaining T_k is

$$\frac{1}{n-1}\frac{C_{n-k+1}}{C_n}.$$

We note that this probability depends only on the number k of leaves of the tree and not on the tree itself; hence, conditionally on merging $n-k+1$ leaves, the resulting tree is still uniform among all the trees with k leaves.

3. Links with the pruning of Aldous's CRT

3.1. Aldous's CRT

In [6], Aldous introduced a CRT which can be obtained as the scaling limit of critical Galton–Watson trees when the length of the branches tends to 0. This tree can also be seen as a compact (with respect to the Gromov–Hausdorff topology) real tree. Indeed, a real tree is a metric space $(\mathcal{T}, d)$ satisfying the following two properties for every $x, y \in \mathcal{T}$.

- (Unique geodesic) There is a unique isometric map $f_{x,y}$ from $[0, d(x, y)]$ into $\mathcal{T}$ such that $f_{x,y}(0) = x$ and $f_{x,y}(d(x, y)) = y$.
- (No loop) If φ is a continuous injective map from $[0, 1]$ into $\mathcal{T}$ such that $\varphi(0) = x$ and $\varphi(1) = y$, then
$$\varphi([0, 1]) = f_{x,y}([0, d(x, y)]).$$

A rooted real tree is a real tree with a distinguished vertex, denoted by $\varnothing$ and called the root.

It is well known that every compact real tree can be coded by a continuous function (this coding is described below); Aldous's tree is just the real tree coded by a normalized Brownian excursion e. We refer the reader to [23] for more details on real trees and their coding by continuous functions.

Remark 2. Aldous's CRT is in fact coded by $2e$. We omit here the factor 2 for convenience, but some constants may vary between this paper and Aldous's results. Our setting corresponds to the branching mechanism $\psi(\lambda) = 2\lambda^2$ in [15].

Let e be a normalized Brownian excursion on $[0, 1]$. For $s, t \in [0, 1]$, we set

$$d(s, t) = e(s) + e(t) - 2 \inf_{u \in [s\wedge t, s\vee t]} e(u).$$

We then define the equivalence relation $s \sim t$ if and only if $d(s, t) = 0$ and the tree $\mathcal{T}$ is the quotient space $\mathcal{T} = [0, 1]/\sim$. We denote by p the canonical projection from $[0, 1]$ to $\mathcal{T}$. The distance d induces a distance on $\mathcal{T}$, and we keep the notation d for this distance. The metric space $(\mathcal{T}, d)$ is then a real tree. The metric space $(\mathcal{T}, d)$ can be seen as a rooted tree by choosing $\varnothing = p(0)$ as the root.

For $x, y \in \mathcal{T}$, we denote by $[\![x, y]\!]$ the range of the unique injective continuous path between x and y in $\mathcal{T}$. We also define a length measure $\ell(\mathrm{d}x)$ on the skeleton of $\mathcal{T}$ (i.e. nonleave vertices) by

$$\ell([\![x, y]\!]) = d(x, y).$$

Finally, for $x, y \in \mathcal{T}$, we denote by $a(x, y)$ their last common ancestor, i.e. the unique point in $\mathcal{T}$ such that

$$[\![\varnothing, x]\!] \cap [\![\varnothing, y]\!] = [\![\varnothing, a(x, y)]\!].$$

For simplicity, we write $a(s, t)$ instead of $a(p(s), p(t))$ for $s, t \in [0, 1]$.

3.2. The interval-partition-valued process

As in [2] and [3], we throw points on the CRT 'uniformly' on the skeleton of the CRT, and add more and more points as time goes by. More precisely, we consider a Poisson point measure $M(\mathrm{d}\theta, \mathrm{d}x)$ on $[0, +\infty) \times \mathcal{T}$ with intensity $4\,\mathrm{d}\theta\,\ell(\mathrm{d}x)$.

For $\theta > 0$, we define an equivalence relation '$\sim_\theta$' on $[0, 1]$ by

$$s \sim_\theta t \quad \Longleftrightarrow \quad s = t \quad \text{or} \quad M([0, \theta] \times [\![\varnothing, a(s, t)]\!]) > 0.$$

We set $(I_k^\theta, k \in \mathcal{K}_\theta)$ to be the equivalence classes associated with '$\sim_\theta$', nonreduced to a singleton. Let us remark that each I_k^θ is an interval.

Equivalently, we define

$$\mathcal{T}_\theta = \{x \in \mathcal{T},\ M([0, \theta] \times [\![\varnothing, x]\!]) = 0\},$$

which is the set of vertices that have no marks on their lineage. The tree $\mathcal{T}_\theta$ is called the pruned tree; it corresponds to the whole dust of the coalescent process. Then consider the set $(\mathcal{T}_k^\theta, k \in \mathcal{K}_\theta)$ of the connected components of $\mathcal{T} \setminus \mathcal{T}_\theta$ which are the subtrees that are grafted on the leaves of $\mathcal{T}_\theta$ to get $\mathcal{T}$ (see Figure 2). Then I_k^θ is just the set of $s \in [0, 1]$ such that $p(s) \in \mathcal{T}_k^\theta$.

By the definition of the mark process, for $\theta' > \theta$, we have

$$s \sim_\theta t \quad \Longrightarrow \quad s \sim_{\theta'} t,$$

and, consequently,

$$\text{for all } k \in \mathcal{K}_\theta, \text{ there exists } k' \in \mathcal{K}_{\theta'} \text{ such that } I_k^\theta \subset I_{k'}^{\theta'}.$$

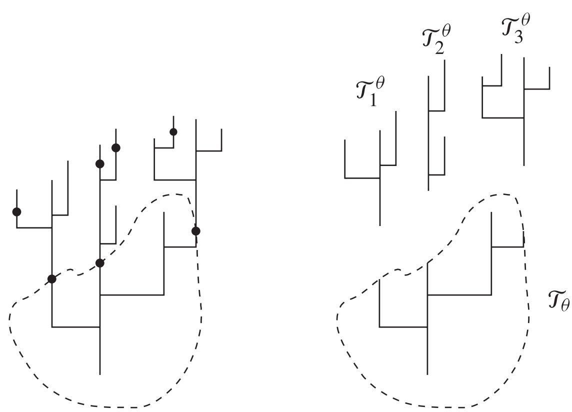

FIGURE 2: *Left*: Aldous's CRT with the marks. *Right*: the subtrees constructed from the marks.

Therefore, the process $\mathcal{I} = ((I_k^\theta,\ k \in \mathcal{K}_\theta),\ \theta \ge 0)$ can be viewed as a process where several blocks coalesce together (with part of the dust) into a single larger block. In the framework of trees, when θ increases, the number of marks also increases and, when a mark appears on $\mathcal{T}_\theta$, some subtrees above $\mathcal{T}_\theta$ coalesce with part of $\mathcal{T}_\theta$.

Let us remark that, as announced in the introduction, this process always has dust which corresponds to individuals that have no marks on their lineage, i.e. that belong to $\mathcal{T}_\theta$. The dust has Lebesgue measure σ_θ:

$$\sigma_\theta = \int_0^1 \mathbf{1}_{\{M([0,\theta]\times[\![\varnothing, p(s)]\!])=0\}}\ \mathrm{d}s = \int_0^1 \mathbf{1}_{\{p(s)\in\mathcal{T}_\theta\}}\ \mathrm{d}s.$$

We recall the distribution of $(\sigma_\theta,\ \theta \ge 0)$ from [8] on the size process of a tagged fragment for a self-similar fragmentation; see also Proposition 9.1 or Corollary 9.2 (but with $\beta = 2$) of [1]. The distribution of $(\sigma_\theta,\ \theta \ge 0)$ under the normalized Brownian excursion measure is given by $(1/(1+4\tau_\theta),\ \theta \ge 0)$, where $(\tau_\theta,\ \theta \ge 0)$ is a stable subordinator with index $\frac{1}{2}$ with no drift, no killing, and Lévy measure $(2\pi x^3)^{-1/2}\,\mathrm{d}x$ on $(0,\infty)$: for $\lambda \ge 0$, $\mathbb{E}[\mathrm{e}^{-\lambda\tau_\theta}] = \mathrm{e}^{-\theta\sqrt{2\lambda}}$.

3.3. The reduced tree with *n* leaves

We apply the paintbox procedure to this process. Consider n independent (and independent of the process) random variables, uniformly distributed on $(0, 1)$. This corresponds to choosing n leaves uniformly on Aldous's CRT. Then the law of the reduced tree, $\mathcal{T}^n$, containing these n leaves is given in [15, Section 3.3] or [7, Section 4.3]. The shape of the reduced tree is a binary tree with uniform probability on all ordered binary trees with n leaves. As the tree is binary, it is composed of $2n-1$ branches with lengths $(h_1, \dots, h_{2n-1})$ and distribution

$$2^{n+1}\frac{(2n-1)!}{(n-1)!} s_n \mathrm{e}^{-2s_n^2}\,\mathbf{1}_{\{h_1>0,\dots,h_{2n-1}>0\}}\ \mathrm{d}h_1\cdots\mathrm{d}h_{2n-1},$$

where $s_n = \sum_{k=1}^{2n-1} h_k$ is the total length of the reduced tree.

Since the reduced tree is binary and all its edges are identically distributed, the first mark that appears on the reduced tree is uniformly distributed among all the edges and we deduce that this continuous coalescent procedure has the same structure as a $\beta(\frac{3}{2}, \frac{1}{2})$-coalescent process. However, this process is not *stricto sensu* a coalescent process as the time at which n leaves chosen at random undergo a coalescence is not exponentially distributed, but is distributed according to an exponential random variable of (random) parameter H_n, with H_n equal to four times the total length of the internal branches. Thus, H_n is distributed as $4\sum_{k=1}^{n-1} h_k$. Note that the random variables $(h_1, \dots, h_{2n-1})$ are exchangeable and $\mathbb{E}[h_1] = 2^{-3/2}\Gamma(n-\frac{1}{2})/\Gamma(n)$. In particular, we have

$$\mathbb{E}[H_n] = 4(n-1)\mathbb{E}[h_1] = \sqrt{2}(n-1)\frac{\Gamma(n-1/2)}{\Gamma(n)} = \sqrt{\frac{2}{\pi}}\lambda_n.$$

So $\mathbb{E}[H_n]$ corresponds (up to a scaling constant) to the rate of the $\beta(\frac{3}{2}, \frac{1}{2})$-coalescent starting with n individuals.

Remark 3. We conjecture that there exists a random time change $(A_t,\ t \ge 0)$ such that the process $((I_k^{A_t},\ k \in \mathcal{K}_{A_t}),\ t \ge 0)$ is the interval-partition-valued process associated with the β-coalescent. However, we were not able to exhibit such a time change.

4. Applications

4.1. Number of coalescent events

Proposition 1 is just a consequence of Theorem 6.2 of [21] on the number of cuts used to isolate the root in a Galton–Watson tree with a given number of leaves. We must just remark that a binary tree with n leaves has $2n-1$ edges and that a Galton–Watson tree with binary branching conditioned to have n leaves is uniformly distributed among the binary trees with n leaves.

Remark 4. According to [2], we also have the following equality in distribution:

$$Z \overset{\mathrm{D}}{=} \Theta \quad \text{with} \quad \Theta = \int_0^\infty \sigma_\theta \,\mathrm{d}\theta.$$

Note that if a pruning mark appears twice or more on the same internal branch, only one will be taken into account as a coalescent event, and that the pruning marks which appear on the external branch will not be taken into account as a coalescent event. Let X_n be the number of pruning events of the reduced tree with n leaves, and let X'_n be the number of coalescent events. We deduce that X_n is stochastically larger than (or equal to) X'_n. But, the almost-sure convergence which appears in [2] (see also [19] for the fluctuations) gives

$$\lim_{n\to+\infty} \frac{X_n}{\sqrt{n}} = \sqrt{2}\Theta \quad \text{a.s.}$$

This implies that the number of marked external branches, say W_n, is of order $o(n)$. We shall see in Section 4.2.2 that it is in fact of order 1.

4.2. Last coalescent event

4.2.1. *The CRT framework.* We refer the reader to [15] for the construction of Lévy trees and their main properties. Let $\mathcal{T}$ be a continuum Lévy tree associated with the branching mechanism $\psi(\lambda) = \alpha\lambda^2$, i.e. coded as in Section 3.1 by a positive excursion e of $\sqrt{2\alpha}B$, where B is a standard Brownian motion. We denote by $\mathbb{N}$ the 'law' of this tree when the coded function e is distributed according to the Itô measure (hence, $\mathbb{N}$ is an infinite measure and is not really a distribution) and by $\mathbb{N}^{(r)}$ the same law when e is distributed as a normalized excursion of length r. We denote by $\boldsymbol{m}$ the mass measure on the tree, i.e. the image of the Lebesgue measure on $[0, +\infty)$ by the canonical projection p.

Conditionally on $\mathcal{T}$, let M be defined as in Section 3.2 with the intensity $2\alpha \,\mathrm{d}\theta\,\ell(\mathrm{d}s)$ instead of $4\,\mathrm{d}\theta\,\ell(\mathrm{d}s)$. (We introduce the parameter α in order to make the references to [1] easier.) Consider the pruning of the tree $\mathcal{T}$ at time $\theta > 0$:

$$\mathcal{T}_\theta = \{s \in \mathcal{T};\, M([0,\theta] \times [\![\varnothing, s]\!]) = 0\}.$$

Also, set $\sigma_\theta = \boldsymbol{m}(\mathcal{T}_\theta)$. This notation is consistent with the definition of σ_θ in Section 3.2.

Using Theorem 1.1 of [5] (see also Proposition 5.4 of [1]), we find that, under $\mathbb{N}$, the pruned tree $\mathcal{T}_\theta$ is distributed as a Lévy tree associated with the branching mechanism ψ_θ defined by

$$\psi_\theta(u) = \psi(u+\theta) - \psi(\theta).$$

Moreover, using Lemma 3.8 of [1], we have the following Girsanov formula that links the law of $\mathcal{T}_\theta$ with that of $\mathcal{T}$: for every nonnegative measurable functional F on the space of trees,

$$\mathbb{N}[F(\mathcal{T}_\theta)] = \mathbb{N}[F(\mathcal{T})\mathrm{e}^{-\alpha\theta^2\sigma}]. \tag{3}$$

Let n be a positive integer. We consider, under $\mathbb{N}$ (or $\mathbb{N}^{(r)}$), conditionally given the tree $\mathcal{T}$, n i.i.d. leaves $x_1, \dots, x_n$, uniformly chosen on the set of leaves, i.e. sampled with the probability $\boldsymbol{m}(\mathrm{d}x)/\sigma$. For $\theta > 0$, let $(\mathcal{T}^j,\ j \in J)$ be the connected components of $\mathcal{T} \setminus \mathcal{T}_\theta$. We write $Y_\theta^0(n) = \sum_{\ell=1}^n \mathbf{1}_{\{x_\ell \in \mathcal{T}_\theta\}}$ to denote the number of chosen leaves on the subtree $\mathcal{T}_\theta$, and, for $k \geq 1$, $Y_\theta^k(n) = \operatorname{card}\{j \in J;\ \sum_{\ell=1}^n \mathbf{1}_{\{x_\ell \in \mathcal{T}^j\}} = k\}$ to denote the number of subtrees with exactly k chosen leaves. In particular, we have

$$Y_\theta^0(n) + \sum_{k \geq 1} k Y_\theta^k(n) = n. \tag{4}$$

We set $N_\theta(n) = \sum_{k \geq 0} Y_\theta^k(n)$ to be the number of chosen leaves on $\mathcal{T}_\theta$ plus the number of subtrees with chosen leaves. For convenience, we shall consider

$$Y_\theta(n) = \sum_{k \geq 2} Y_\theta^k(n) = N_\theta(n) - Y_\theta^0(n) - Y_\theta^1(n).$$

Let $\mathcal{T}^n$ be the reduced tree of the chosen leaves, that is, the smallest connected component of $\mathcal{T}$ containing the root $\varnothing$ and $\{x_i,\ 1 \leq i \leq n\}$. Let $\mathcal{T}_\theta^n$ be the reduced tree $\mathcal{T}^n$ pruned at time $\theta > 0$:

$$\mathcal{T}_\theta^n = \{s \in \mathcal{T}^n;\ M([0,\theta] \times [\![0,s]\!]) = 0\}.$$

Note that N_θ is the number of leaves of $\mathcal{T}_\theta^n$ (with the convention that the root is not a leaf). Define the last pruning event as

$$L_n = \inf\{\theta > 0;\ N_\theta(n) = 1\}.$$

We define

$$U_n = N_{L_n-}(n), \qquad V_n = Y_{L_n-}^0(n) + Y_{L_n-}^1(n), \quad \text{and} \quad W_n = Y_{L_n-}^1(n).$$

We can interpret U_n as the number of leaves of the pruned reduced tree, $V_n - W_n$ as the number of chosen leaves of the pruned reduced tree, and W_n as the number of subtrees with only one chosen leaf just before the last pruning event.

4.2.2. *Proof of Proposition 2.* Let B_n be the number of blocks, and let E_n be the number of singletons involved in the last coalescent event of $(\Pi^{[n]}(t),\ t \geq 0)$. Using the link with the pruning of CRT from the previous section, (B_n, E_n) is distributed as (U_n, V_n) under $\mathbb{N}^{(1)}$.

Following Remark 4, we can interpret V_n as the sum of $V_n - W_n$ (number of leaves of $\mathcal{T}^n$ with no mark before the last pruning event) and the number W_n of leaves of $\mathcal{T}^n$ with no mark on their ancestral lineage until the mark corresponding to the last coalescent pruning, but for the external branch, where there is at least one mark.

Before giving the asymptotic distribution of (U_n, V_n), we need to introduce some notation. For $a \geq 0$ and $b \geq 0$ such that $a + b > 0$, we define $\Delta_0(a,b)$ as

$$\Delta_0(a,b) = \begin{cases} \dfrac{1}{\sqrt{|1+b-2a|}} \log\left(\dfrac{1+\sqrt{b}+\sqrt{|1+b-2a|}}{1+\sqrt{b}-\sqrt{|1+b-2a|}}\right) & \text{if } 1+b-2a \neq 0, \\ \dfrac{2}{1+\sqrt{b}} & \text{if } 1+b-2a = 0. \end{cases} \tag{5}$$

It is easy to check that the function Δ_0 is continuous in (a,b) and that

$$\lim_{(a,b)\to(0,0)} \Delta_0(a,b) + \log(\sqrt{b}+a) = \log(2).$$

We set, for $a \geq 0$ and $b \geq 0$ (with the convention that $I(0,0) = 1$),

$$I(a,b) = 1 + \log(2) - \log(\sqrt{b} + a) - \Delta_0(a,b). \tag{6}$$

Note that the function I is continuous on $[0, +\infty)^2$ (and in particular at $(0,0)$).

We shall prove the next result in Section 5.

Proposition 4. *Under $\mathbb{N}^{(1)}$, as n goes to ∞, (U_n, V_n, V_n) converges, in distribution, to a finite random variable (U, V, W). Furthermore, the distribution of (U, V, W) is characterized by the following generating function: for $\rho, \rho_0, \rho_1 \in [0,1]$,*

$$\Psi(\rho, \rho_0, \rho_1) = \mathbb{E}[\rho^{U-V} \rho_0^{V-W} \rho_1^{W}] = \rho I\left(1 - \frac{\rho + \rho_1}{2}, 1 - \rho_0\right).$$

Note that the random variable $U - V - 1$ is distributed as W. Since $W \leq V$, this implies that $U - V - 1$ is stochastically smaller than V. This last remark and Proposition 4 readily imply Proposition 2.

The generating function of $V - W$ is given, for $\rho_0 \in [0,1]$ by $\mathbb{E}[\rho_0^{V-W}] = I(0, 1 - \rho_0)$, that is,

$$\mathbb{E}[\rho_0^{V-W}] = 1 + \log(2) - \log(\sqrt{1-\rho_0}) - \frac{1}{\sqrt{2-\rho_0}} \log\left(\frac{1 + \sqrt{1-\rho_0} + \sqrt{2-\rho_0}}{1 + \sqrt{1-\rho_0} - \sqrt{2-\rho_0}}\right).$$

The generating function of W is given for $\rho_1 \in [0,1]$ by

$$\mathbb{E}[\rho_1^{W}] = I\left(\frac{1-\rho_1}{2}, 0\right) = 1 + \log(4) - \log(1-\rho_1) - \frac{1}{\sqrt{\rho_1}} \log\left(\frac{1+\sqrt{\rho_1}}{1-\sqrt{\rho_1}}\right).$$

5. Proof of Proposition 4

In order to compute the generating function that appears in Proposition 4, it is easier to work first under $\mathbb{N}$ and then condition on σ to have the result under $\mathbb{N}^{(1)}$ and to consider a Poissonian number of chosen leaves. Let $\lambda > 0$. Under $\mathbb{N}$ or $\mathbb{N}^{(r)}$, conditionally given the tree $\mathcal{T}$, we consider a Poisson point measure $\mathcal{N} = \sum_{i \in I} \delta_{x_i}$ on $\mathcal{T}$ with intensity $\lambda \boldsymbol{m}$. We denote by $\tilde{N} = \mathcal{N}(\mathcal{T})$ the number of chosen leaves. The law of the total mass $\sigma = \boldsymbol{m}(\mathcal{T})$ of $\mathcal{T}$ under $\mathbb{N}$ is given by the following Laplace transform:

$$\mathbb{N}[1 - \mathrm{e}^{-\kappa\sigma}] = \psi^{-1}(\kappa) = \sqrt{\frac{\kappa}{\alpha}}. \tag{7}$$

Conditionally on σ, the random variable $\tilde{N}$ is Poisson with parameter $\lambda\sigma$. Therefore, by first conditioning on σ, we get, for $k \geq 1$,

$$\mathbb{N}[\tilde{N} = k] = \frac{\lambda^k}{k!} \mathbb{N}[\sigma^k \mathrm{e}^{-\lambda\sigma}] = \frac{1}{2\sqrt{\pi}} \sqrt{\frac{\lambda}{\alpha}} \frac{\Gamma(k - 1/2)}{\Gamma(k+1)}. \tag{8}$$

Let $\theta > 0$. It follows from the special Markov property, Theorem 5.6 of [1] (see also [4]), that, under $\mathbb{N}$, conditionally on σ_θ, the random variables $(Y_\theta^k(\tilde{N}), k \geq 0)$ are independent, $Y_\theta^0(\tilde{N})$ is Poisson with parameter $\lambda\sigma_\theta$, and, for $k \geq 1$, $Y_\theta^k(\tilde{N})$ is Poisson with parameter $2\alpha\theta\sigma_\theta \mathbb{N}[\tilde{N} = k]$.

For $a \in [0,1]$, we set

$$f_\theta(a) = \mathbb{N}[a^{\tilde{N}} \rho^{Y_\theta(\tilde{N})} \rho_0^{Y_\theta^0(\tilde{N})} \rho_1^{Y_\theta^1(\tilde{N})} \mathbf{1}_{\{\tilde{N} > 0\}}].$$

Lemma 1. *We have*

$$f_\theta(a) = \theta + \sqrt{\frac{\lambda}{\alpha}} - \sqrt{\delta_0 + 2\delta_1\sqrt{1-a} - \delta_2 a}, \tag{9}$$

with

$$\delta_0 = \theta^2 + \frac{\lambda}{\alpha} + 2\theta\sqrt{\frac{\lambda}{\alpha}}(1-\rho), \qquad \delta_1 = \theta\rho\sqrt{\frac{\lambda}{\alpha}}, \qquad \delta_2 = \frac{\lambda}{\alpha}\rho_0 - \theta\sqrt{\frac{\lambda}{\alpha}}(\rho - \rho_1).$$

Note that

$$\delta_0 - \delta_2 = \theta^2 + \frac{\lambda}{\alpha}(1-\rho_0) + \theta\sqrt{\frac{\lambda}{\alpha}}(2-\rho-\rho_1) \geq \theta^2 > 0. \tag{10}$$

Consequently, the right-hand side of (9) is well defined.

Proof of Lemma 1. We set $\mu = -\log(\rho)$, $\mu_i = -\log(\rho_i)$ for $i \in \{0, 1\}$, and $\kappa = -\log(a)$. We have

$$\begin{aligned} f_\theta(a) &= \mathbb{N}[\mathrm{e}^{-\mu_0 Y_\theta^0(\tilde{N}) - \mu_1 Y_\theta^1(\tilde{N}) - \mu Y_\theta(\tilde{N}) - \kappa\tilde{N}} - \mathbf{1}_{\{\tilde{N}=0\}}] \\ &= \mathbb{N}\bigg[\mathbb{N}[\mathrm{e}^{-(\mu_0+\kappa)Y_\theta^0(\tilde{N})} \mid \sigma_\theta]\tilde{\mathbb{N}}[\mathrm{e}^{-(\mu_1+\kappa)Y_\theta^1(\tilde{N})} \mid \sigma_\theta]\prod_{k\geq 2}\mathbb{N}[\mathrm{e}^{-(\mu+k\kappa)Y_\theta^k(\tilde{N})} \mid \sigma_\theta]\bigg] \\ &\quad - \mathbb{N}[\tilde{N} = 0], \end{aligned}$$

using (4) and the independence of the variables $Y_\theta^k(\tilde{N})$ conditionally given σ_θ. Now, as the variables $Y_\theta^k(\tilde{N})$ are, conditionally given σ_θ, Poisson variables, and thanks to (8), we have

$$f_\theta(a) = \mathbb{N}[\mathrm{e}^{-\gamma\sigma_\theta} - \mathrm{e}^{-\lambda\sigma}],$$

with

$$\begin{aligned} \gamma &= \lambda(1 - \mathrm{e}^{-(\kappa+\mu_0)}) + 2\alpha\theta\lambda\mathbb{N}[\sigma\mathrm{e}^{-\lambda\sigma}](1 - \mathrm{e}^{-(\kappa+\mu_1)}) \\ &\quad + 2\alpha\theta\sum_{k\geq 2}\frac{\lambda^k}{k!}\mathbb{N}[\sigma^k\mathrm{e}^{-\lambda\sigma}](1 - \mathrm{e}^{-(k\kappa+\mu)}) \\ &= \lambda(1 - \mathrm{e}^{-(\kappa+\mu_0)}) + 2\alpha\theta\lambda\mathbb{N}[\sigma\mathrm{e}^{-\lambda\sigma}](1 - \mathrm{e}^{-(\kappa+\mu_1)}) \\ &\quad + 2\alpha\theta(\mathbb{N}[1 - \mathrm{e}^{-\lambda\sigma}](1 - \mathrm{e}^{-\mu}) - \lambda\mathbb{N}[\sigma\mathrm{e}^{-\lambda\sigma}](1 - \mathrm{e}^{-(\mu+\kappa)}) \\ &\qquad + \mathrm{e}^{-\mu}\mathbb{N}[1 - \mathrm{e}^{-\lambda(1-\mathrm{e}^{-\kappa})\sigma}]). \end{aligned}$$

We now use (7) to obtain

$$\begin{aligned} \gamma &= \lambda(1 - \mathrm{e}^{-(\kappa+\mu_0)}) + \theta\sqrt{\lambda\alpha}\mathrm{e}^{-\kappa}(\mathrm{e}^{-\mu} - \mathrm{e}^{-\mu_1}) + 2\theta\sqrt{\lambda\alpha}(1 - \mathrm{e}^{-\mu}) \\ &\quad + 2\theta\mathrm{e}^{-\mu}\sqrt{\lambda\alpha}\sqrt{1 - \mathrm{e}^{-\kappa}} \\ &= 2\theta\sqrt{\lambda\alpha}(1 - \mathrm{e}^{-\mu}) + \lambda + 2\theta\mathrm{e}^{-\mu}\sqrt{\lambda\alpha}\sqrt{1-a} - a(\lambda\mathrm{e}^{-\mu_0} - \theta\sqrt{\lambda\alpha}(\mathrm{e}^{-\mu} - \mathrm{e}^{-\mu_1})). \end{aligned}$$

Using the special Markov property of [1, Theorem 5.6], we find that, conditionally given σ_θ, σ is distributed as $\sigma_\theta + \sum\sigma_i$, where the σ_i are the atoms of a Poisson point measure of intensity $2\alpha\theta\sigma_\theta\mathbb{N}[\mathrm{d}\sigma]$. This yields

$$f_\theta(a) = \mathbb{N}[\mathrm{e}^{-\gamma\sigma_\theta} - \mathrm{e}^{-\sigma_\theta(\lambda+2\theta\sqrt{\lambda\alpha})}].$$

To conclude, we use Girsanov's formula (3) to obtain

$$f_\theta(a) = \mathbb{N}[\mathrm{e}^{-(\gamma+\alpha\theta^2)\sigma} - \mathrm{e}^{-\sigma(\alpha\theta^2+\lambda+2\theta\sqrt{\lambda\alpha})}] = \theta + \sqrt{\frac{\lambda}{\alpha}} - \sqrt{\theta^2 + \frac{\gamma}{\alpha}}.$$

This gives the result.

Let us set $A_n = \mathbb{N}^{(1)}[\mathrm{e}^{-\mu(U_n-V_n)-\mu_0(V_n-W_n)-\mu_1 W_n}]$. To prove Proposition 4, it is enough to prove that

$$\lim_{n\to+\infty} A_n = \rho I\left(1 - \frac{\rho+\rho_1}{2}, 1-\rho_0\right).$$

As the CRT is coded by a Brownian excursion, it enjoys a scaling property, namely the law of $r\mathcal{T}$ under $\mathbb{N}^{(r^2)}$ is that of $\mathcal{T}$ under $\mathbb{N}^{(1)}$ (where $r\mathcal{T}$ means that we multiply the distance on the tree by a factor r). Consequently, the mark process (defined as a Poisson point measure with intensity proportional to the length measure) also satisfies a scaling property. It is then easy to deduce that the law of (U_n, V_n, W_n) does not depend on σ. So, we have

$$A_n = \mathbb{N}^{(r)}[\mathrm{e}^{-\mu(U_n-V_n)-\mu_0(V_n-W_n)-\mu_1 W_n}]$$

for every positive r. We set

$$\mathcal{A}_n = A_n \mathbb{N}[\sigma^n \mathrm{e}^{-\lambda\sigma}]. \tag{11}$$

By conditioning on σ, we obtain

$$\mathcal{A}_n = \mathbb{N}[\sigma^n \mathrm{e}^{-\lambda\sigma}\mathrm{e}^{-\mu(U_n-V_n)-\mu_0(V_n-W_n)-\mu_1 W_n}].$$

For $n \geq 1$, we set

$$F_\theta(n,r) = \mathbb{N}^{(r)}[\rho^{Y_\theta(n)}\rho_0^{Y_\theta^0(n)}\rho_1^{Y_\theta^1(n)}] \quad \text{and} \quad H_n(\theta) = \mathbb{N}[F_\theta(n,\sigma)\sigma^n\mathrm{e}^{-\lambda\sigma}]. \tag{12}$$

Recall that

$$\mathbb{N}[\sigma^n \mathrm{e}^{-\lambda\sigma}] = \lambda^{-n} n!\,\mathbb{N}[\tilde{N} = n] = \frac{1}{2\sqrt{\alpha\pi}\lambda^{n-1/2}}\Gamma\left(n - \frac{1}{2}\right). \tag{13}$$

Therefore, we have

$$f_\theta(a) = \mathbb{N}[a^{\tilde{N}} F_\theta(\tilde{N},\sigma)\,\mathbf{1}_{\{\tilde{N}>0\}}],$$

and, thanks to (13), for $n \geq 1$,

$$f_\theta^{(n)}(0) = n!\,\mathbb{N}[F_\theta(n,\sigma)\,\mathbf{1}_{\{\tilde{N}=n\}}] = \lambda^n H_n(\theta).$$

We use the description in [15, Section 3.3] of the reduced tree spanned by n leaves under the σ-finite measure $\mathbb{N}$: it is a uniform binary tree with n leaves and with edge lengths i.i.d. and 'distributed' as $\alpha\,\mathrm{d}h$. We denote by $[\![\varnothing, x_1]\!]$ the edge of the reduced tree attached to the root, and by H its length. The time L_n at which the last coalescent event occurs is just the first time θ at which a mark appears on $[\![\varnothing, x_1]\!]$ and, therefore, it is, conditionally given $H = h$, exponentially distributed with parameter $2\alpha h$. Moreover, if we denote by $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ the two subtrees attached to x_1, and by $\sigma^{(1)}$ and $\sigma^{(2)}$ their respective total masses, we have

$$\sigma = \sigma^{(1)} + \sigma^{(2)} + \sum_{i\in I'} \sigma_i,$$

where the σ_i are the total masses of the subtrees attached on the edge $[\![\varnothing, x_1]\!]$. The random measure $\sum_{i\in I'}\sigma_i$ is independent of $\sigma^{(1)}$ and $\sigma^{(2)}$ and is, conditionally on $\{H = h\}$, distributed as a Poisson point measure with intensity $2\alpha h\mathbb{N}[\mathrm{d}\sigma]$. Eventually, the two reduced subtrees attached to x_1 are independent and distributed as uniform binary trees with k and $n-k$ leaves $(0 < k < n)$, respectively. Recall that $2\alpha h\mathbb{N}[1-\mathrm{e}^{-\lambda\sigma}] = 2\sqrt{\alpha\lambda}h$. We deduce from this description that

$$\begin{aligned}\mathcal{A}_n &= \int_0^\infty \alpha \mathrm{e}^{-2\sqrt{\alpha\lambda}h}\,\mathrm{d}h \int_0^\infty 2\alpha h \mathrm{e}^{-2\alpha h\theta}\,\mathrm{d}\theta \sum_{k=1}^{n-1}\binom{n}{k} H_k(\theta)H_{n-k}(\theta)\\ &= \lambda^{-n}\int_0^\infty \frac{\mathrm{d}\theta}{(\theta+\sqrt{\lambda/\alpha})^2}G_n(\theta), \qquad (14)\end{aligned}$$

with

$$G_n(\theta) = \frac{\lambda^n}{2}\sum_{k=1}^{n-1}\binom{n}{k}H_k(\theta)H_{n-k}(\theta) = \frac{1}{2}\sum_{k=1}^{n-1}\binom{n}{k} f_\theta^{(k)}(0) f_\theta^{(n-k)}(0). \qquad (15)$$

Let $c_0 = \theta + \sqrt{\lambda/\alpha}$ and $g = c_0 - f_\theta$, so that

$$g(a) = \sqrt{\delta_0 + 2\delta_1\sqrt{1-a} - \delta_2 a}.$$

If h is a function, we write $\partial^n_{a=0}h(a)$ for $h^{(n)}(0)$. Then, using the formula for the nth derivative of a product of functions, we have, for $n \geq 2$,

$$G_n(\theta) = \frac{1}{2}\sum_{k=1}^{n-1}\binom{n}{k} g^{(k)}(0)g^{(n-k)}(0) = \frac{1}{2}\partial^n_{a=0}g^2(a) - g(0)g^{(n)}(0).$$

That is, since $\partial^n_{a=0}\sqrt{1-a} = -\Gamma(n-\frac{1}{2})/2\sqrt{\pi}$,

$$G_n(\theta) = \delta_1\partial^n_{a=0}\sqrt{1-a} - \sqrt{\delta_0+2\delta_1}\,g^{(n)}(0) = -\delta_1\frac{\Gamma(n-1/2)}{2\sqrt{\pi}} - \sqrt{\delta_0+2\delta_1}\,g^{(n)}(0). \qquad (16)$$

The next lemma gives an equivalent expression for $g^{(n)}(0)$.

Lemma 2. *We have*

$$\lim_{n\to+\infty} -\frac{g^{(n)}(0)}{\Gamma(n-1/2)} = \frac{\delta_1}{2\sqrt{\pi}}\frac{1}{\sqrt{\delta_0-\delta_2}}.$$

Proof. Using the fact that the density, which corresponds to the density of the $\frac{1}{2}$-stable subordinator with no drift,

$$h(x) = \frac{\delta_1 r}{\sqrt{\pi}}\frac{1}{x^{3/2}}\mathrm{e}^{-\delta_1^2 r^2/x}\,\mathbf{1}_{\{x>0\}}$$

has Laplace transform

$$\int_0^{+\infty}\mathrm{e}^{-\lambda x}h(x)\,\mathrm{d}x = \mathrm{e}^{-2\delta_1 r\sqrt{\lambda}},$$

we can write

$$\begin{aligned}g(a) &= \frac{1}{2\sqrt{\pi}}\int_0^\infty \frac{\mathrm{d}r}{r^{3/2}}(1-\mathrm{e}^{-(\delta_0+2\delta_1\sqrt{1-a}-\delta_2 a)r})\\ &= \frac{1}{2\pi}\int_0^\infty\frac{\mathrm{d}r}{r^{3/2}}\int_0^\infty \delta_1 r\frac{\mathrm{d}x}{x^{3/2}}\mathrm{e}^{-\delta_1^2r^2/x}(1-\mathrm{e}^{-\delta_0 r - x + a(\delta_2 r + x)}).\end{aligned}$$

We deduce, with $y = r/x$, that, for $n \geq 1$,

$$
\begin{aligned}
-g^{(n)}(0) &= \frac{1}{2\pi}\int_0^\infty \frac{\mathrm{d}r}{r^{3/2}}\int_0^\infty \delta_1 r \frac{\mathrm{d}x}{x^{3/2}}(\delta_2 r + x)^n \mathrm{e}^{-\delta_0 r - x - \delta_1^2 r^2/x} \\
&= \frac{\delta_1}{2\pi}\int_0^\infty r^{n-1}\,\mathrm{d}r \int_0^\infty \frac{\mathrm{d}y}{\sqrt{y}}\Big(\frac{1+\delta_2 y}{y}\Big)^n \mathrm{e}^{-r(\delta_0 + 1/y + \delta_1^2 y)} \\
&= \frac{\delta_1 \Gamma(n)}{2\pi}\int_0^\infty \frac{\mathrm{d}y}{\sqrt{y}}\Big(\frac{1+\delta_2 y}{y}\Big)^n \frac{1}{(\delta_0 + 1/y + \delta_1^2 y)^n} \\
&= \frac{\delta_1 \Gamma(n)}{2\pi}\int_0^\infty \frac{\mathrm{d}y}{\sqrt{y}}\varphi(y)^n,
\end{aligned}
$$

with

$$
\varphi(y) = \frac{1+\delta_2 y}{1+\delta_0 y + \delta_1^2 y^2}.
$$

To obtain an equivalent expression for $g^{(n)}(0)$, we use Laplace's method. Note that $\varphi(0) = 1$ and $\varphi'(0) = \delta_2 - \delta_0 < 0$ thanks to (10). Note that $\delta_0 > 0$ and $\delta_1 > 0$.

The only root to $\varphi(y) = 1$ is 0. Let us assume that there is a root to $\varphi(y) = -1$. This implies that $1 + \delta_0 y + \delta_1^2 y^2 + 1 + \delta_2 y = 0$, that is,

$$
(\delta_0 + \delta_2) y + \delta_1^2 y^2 = -2.
$$

But we have

$$
\delta_0 + \delta_2 \geq \theta^2 + \frac{\lambda}{\alpha} - \theta\sqrt{\lambda}\alpha > 0.
$$

Therefore, $|\varphi(y)| \in (-1, 1)$ for $y > 0$ since $\lim_{y\to+\infty}\varphi(y) = 0$.

The discussion above proves that

there exists $\delta_0 > 0$ for all $\delta \in (0, \delta_0)$, and, for all $x \geq \delta$, $|\varphi(x)| \leq \varphi(\delta) < 1$.

Moreover, we have

$$
\lim_{x\downarrow 0} \frac{\log(\varphi(x))}{(\delta_2 - \delta_0)x} = 1.
$$

Therefore, we can apply the standard Laplace method and obtain the following asymptotic in n:

$$
\begin{aligned}
-g^{(n)}(0) &\sim \frac{\delta_1 \Gamma(n)}{2\pi}\int_0^\infty \frac{\mathrm{d}y}{\sqrt{y}}\mathrm{e}^{-n(\delta_0 - \delta_2)y} \\
&= \frac{\delta_1 \Gamma(n)}{2\pi}\frac{\Gamma(1/2)}{\sqrt{n(\delta_0 - \delta_2)}} \\
&\sim \frac{\delta_1 \Gamma(n - 1/2)}{2\sqrt{\pi}}\frac{1}{\sqrt{\delta_0 - \delta_2}}.
\end{aligned}
$$

Here $a_n \sim b_n$ means that $\lim_{n\to+\infty} a_n/b_n = 1$.

Recall the notation of Proposition 4:

$$
\Psi(\rho, \rho_0, \rho_1) = \rho I\Big(1 - \frac{\rho + \rho_1}{2}, 1 - \rho_0\Big).
$$

Lemma 3. *We have, for $\rho, \rho_0, \rho_1 \in [0, 1]$,*

$$\lim_{n\to+\infty} \mathbb{N}^{(1)}[\rho^{U_n - V_n} \rho_0^{V_n - W_n} \rho_1^{W_n}] = \Psi(\rho, \rho_0, \rho_1).$$

Note that $\Psi(1, 1, 1) = I(0, 0) = 1$. This implies that (U_n, V_n, W_n) converges in distribution, as n goes to ∞, to an almost-sure finite random variable (U, V, W) and that the generating function of $(U - V, V - W, W)$ is given by Ψ.

Proof of Lemma 3. On the one hand, we deduce from (16) and Lemma 2 that

$$\lim_{n\to+\infty} \frac{G_n(\theta)}{\Gamma(n - 1/2)} = -\frac{\delta_1}{2\sqrt{\pi}} + \sqrt{\delta_0 + 2\delta_1}\frac{\delta_1}{2\sqrt{\pi}}\frac{1}{\sqrt{\delta_0 - \delta_2}} = \frac{\delta_1}{2\sqrt{\pi}}\left(\frac{\sqrt{\delta_0 + 2\delta_1}}{\sqrt{\delta_0 - \delta_2}} - 1\right).$$

On the other hand, we deduce from (12) (by considering $\rho = \rho_0 = \rho_1 = 1$) and (13) that

$$H_n(\theta) \le \mathbb{N}[\sigma^n \mathrm{e}^{-\lambda\sigma}] = \frac{1}{2\sqrt{\alpha\pi}\lambda^{n-1/2}}\Gamma\left(n - \frac{1}{2}\right). \tag{17}$$

Recall (2). By decomposing an ordered binary tree with n labelled leaves into two ordered binary subtrees attached to the closest node of the root, we obtain

$$\sum_{k=1}^{n-1}\binom{n}{k} C_k C_{n-k} = C_n.$$

This readily implies that

$$\sum_{k=1}^{n-1}\binom{n}{k}\Gamma\left(k - \frac{1}{2}\right)\Gamma\left(n - k - \frac{1}{2}\right) = 4\sqrt{\pi}\Gamma\left(n - \frac{1}{2}\right).$$

We deduce from the first equality of (15) and (17) that

$$G_n(\theta) = \frac{\lambda^n}{2}\sum_{k=1}^{n-1}\binom{n}{k} H_k(\theta) H_{n-k}(\theta) \le \frac{\lambda}{2\alpha\sqrt{\pi}}\Gamma\left(n - \frac{1}{2}\right).$$

This implies that

$$\frac{G_n(\theta)}{\Gamma(n - 1/2)} \le \frac{\lambda}{2\alpha\sqrt{\pi}}.$$

By dominated convergence, we deduce from (14) that

$$\begin{aligned}
\lim_{n\to+\infty} &\frac{\lambda^n \mathcal{A}_n}{\Gamma(n - 1/2)} \\
&= \int_0^\infty \frac{\mathrm{d}\theta}{(\theta + \sqrt{\lambda/\alpha})^2} \lim_{n\to+\infty} \frac{G_n(\theta)}{\Gamma(n - 1/2)} \\
&= \frac{\sqrt{\lambda}}{2\sqrt{\alpha\pi}}\rho\int_0^\infty \frac{\theta\,\mathrm{d}\theta}{(\theta + \sqrt{\lambda/\alpha})^2}\left(\frac{\theta + \sqrt{\lambda/\alpha}}{\sqrt{\theta^2 + \theta\sqrt{\lambda\alpha^{-1}}(2 - \rho - \rho_1) + \lambda\alpha^{-1}(1 - \rho_0)}} - 1\right) \\
&= \frac{\sqrt{\lambda}}{2\sqrt{\alpha\pi}}\rho\int_0^\infty \frac{\theta\,\mathrm{d}\theta}{(\theta + 1)^2}\left(\frac{\theta + 1}{\sqrt{\theta^2 + \theta(2 - \rho - \rho_1) + (1 - \rho_0)}} - 1\right).
\end{aligned}$$

We deduce from (11) that

$$\begin{aligned}\lim_{n\to+\infty} A_n &= \frac{2\sqrt{\alpha\pi}}{\sqrt{\lambda}} \lim_{n\to+\infty} \frac{\lambda^n \mathcal{A}_n}{\Gamma(n-1/2)} \\ &= \rho \int_0^\infty \frac{\theta\,\mathrm{d}\theta}{(\theta+1)^2}\bigg(\frac{\theta+1}{\sqrt{\theta^2+\theta(2-\rho-\rho_1)+(1-\rho_0)}} - 1\bigg).\end{aligned}$$

Then we use Lemma 5 below to conclude.

Before stating and proving Lemma 5, we first give a preliminary result.

Lemma 4. *For $a \geq 0$, $b \geq 0$, and $a+b>0$, we have $\Delta(a,b) = \Delta_0(a,b)$.*

Proof. We first assume that $1+b-2a>0$. From (18) we have

$$\begin{aligned}-\partial_a \Delta(a,b) &= \int_0^\infty \frac{\theta\,\mathrm{d}\theta}{(\theta+1)(\theta^2+2a\theta+b)^{3/2}} \\ &= -2\partial_b I(a,b) \\ &= -\frac{\Delta(a,b)}{1+b-2a} + \frac{1}{1+b-2a}\frac{\sqrt{b}+1}{\sqrt{b}+a}.\end{aligned}$$

Then, by computing the derivative $\partial_a \Delta_0$ we deduce that $\Delta(a,b) = \Delta_0(a,b) + h_b(a)$ for a function h_b solving

$$-h_b'(a) = -\frac{h_b(a)}{1+b-2a},$$

that is, for some function c_+,

$$h_b(a) = \frac{c_+(b)}{\sqrt{1+b-2a}}.$$

Similarly, we have, for $1+b-2a<0$,

$$\Delta(a,b) = \Delta_0(a,b) + \frac{c_-(b)}{\sqrt{|1+b-2a|}}$$

for some function c_-. Note that Δ and Δ_0 are, by definitions (18) below and (5), continuous on $(0,+\infty)^2$. By letting a go to $(1+b)/2$ we deduce that $c_+ = c_- = 0$. This proves the result.

Recall the definition of I given in (6). We set, for $a \geq 0$ and $b \geq 0$,

$$J(a,b) = \int_0^\infty \frac{\mathrm{d}\theta}{\theta+1}\bigg(\frac{\theta}{\sqrt{\theta^2+2a\theta+b}} - \frac{\theta}{\theta+1}\bigg).$$

Lemma 5. *For $a \geq 0$ and $b \geq 0$, we have $J(a,b) = I(a,b)$.*

Proof. We first note that

$$\int_0^\infty \frac{\theta\,\mathrm{d}\theta}{(\theta^2+2a\theta+b)^{3/2}} = \frac{1}{\sqrt{b}+a} \quad\text{and}\quad \int_0^\infty \frac{\mathrm{d}\theta}{(\theta^2+2a\theta+b)^{3/2}} = \frac{1}{\sqrt{b}(\sqrt{b}+a)}.$$

For $a \geq 0$, $b \geq 0$, and $a+b>0$, we set

$$\Delta(a,b) = \int_0^\infty \frac{\mathrm{d}\theta}{(\theta+1)\sqrt{\theta^2+2a\theta+b}}. \tag{18}$$

We have

$$-\partial_b J(a,b) = \frac{1}{2}\int_0^\infty \frac{\theta\,\mathrm{d}\theta}{(\theta+1)(\theta^2+2a\theta+b)^{3/2}}$$
$$= \frac{1}{2(1+b-2a)}\int_0^\infty \frac{\mathrm{d}\theta}{\sqrt{\theta^2+2a\theta+b}}\left(\frac{\theta+b}{\theta^2+2a\theta+b}-\frac{1}{\theta+1}\right)$$
$$= -\frac{\Delta(a,b)}{2(1+b-2a)} + \frac{1}{2(1+b-2a)}\left(\frac{1}{\sqrt{b}+a}+\frac{b}{\sqrt{b}(\sqrt{b}+a)}\right)$$
$$= -\frac{\Delta(a,b)}{2(1+b-2a)} + \frac{1}{2(1+b-2a)}\frac{\sqrt{b}+1}{\sqrt{b}+a}. \tag{19}$$

We also have

$$-\partial_a J(a,b) = \int_0^\infty \frac{\theta^2\,\mathrm{d}\theta}{(\theta+1)(\theta^2+2a\theta+b)^{3/2}}$$
$$= \frac{1}{1+b-2a}\int_0^\infty \frac{\mathrm{d}\theta}{\sqrt{\theta^2+2a\theta+b}}\left(\frac{(b-2a)\theta-b}{\theta^2+2a\theta+b}+\frac{1}{\theta+1}\right)$$
$$= \frac{\Delta(a,b)}{1+b-2a} + \frac{1}{1+b-2a}\left(\frac{b-2a}{\sqrt{b}+a}-\frac{b}{\sqrt{b}(\sqrt{b}+a)}\right)$$
$$= \frac{\Delta(a,b)}{1+b-2a} + \frac{1}{1+b-2a}\frac{b-\sqrt{b}-2a}{\sqrt{b}+a}. \tag{20}$$

After computing $\partial_a\Delta(a,b)$, we deduce from (20) that, for $a+b>0$,

$$J(a,b) = -\Delta(a,b) - \log(\sqrt{b}+a) + g(b)$$

for some function g. Then computing $\partial_b J(a,b)$, it follows using (19) that $g(b)$ is a constant c. Eventually, on the one hand, taking $a=\sqrt{b}=1$, we obtain

$$J(1,1) = \int_0^\infty \frac{\mathrm{d}\theta}{\theta+1}\left(\frac{\theta}{\theta+1}-\frac{\theta}{\theta+1}\right) = 0.$$

On the other hand, we have

$$J(1,1) = -\Delta(1,1) - \log(2) + c = -1-\log(2)+c.$$

This gives $c = 1+\log(2)$. We obtain

$$J(a,b) = -\Delta(a,b) - \log(\sqrt{b}+a) + 1 + \log(2).$$

That is, $I=J$ for $a+b>0$. Then, we use the continuity of I and J to obtain $I=J$.

References

[1] Abraham, R. and Delmas, J.-F. (2012). A continuum tree-valued Markov process. *Ann. Prob.* **40,** 1167–1211.
[2] Abraham, R. and Delmas, J.-F. (2013). Record process on the continuum random tree. *Latin Amer. J. Prob. Math. Statist.* **10,** 225–251.
[3] Abraham, R. and Serlet, L. (2002). Poisson snake and fragmentation. *Electron. J. Prob.* **7,** 1–15.
[4] Abraham, R., Delmas, J.-F. and He, H. (2013). Pruning of CRT sub-trees. Submitted. Available at http://arxiv.org/abs/1212.2765v1.

[5] Abraham, R., Delmas, J.-F. and Voisin, G. (2010). Pruning a Lévy continuum random tree. *Electron. J. Prob.* **15,** 1429–1473.
[6] Aldous, D. (1991). The continuum random tree. I. *Ann. Prob.* **19,** 1–28.
[7] Aldous, D. (1993). The continuum random tree. III. *Ann. Prob.* **21,** 248–289.
[8] Aldous, D. and Pitman, J. (1998). The standard additive coalescent. *Ann. Prob.* **26,** 1703–1726.
[9] Berestycki, N. (2009). *Recent Progress in Coalescent Theory* (Ensaios Matemáticos **16**). Sociedade Brasileira de Matemática, Rio de Janeiro.
[10] Berestycki, J., Berestycki, N. and Schweinsberg, J. (2007). Beta-coalescents and continuous stable random trees. *Ann. Prob.* **35,** 1835–1887.
[11] Bertoin, J. and Yor, M. (2005). Exponential functionals of Lévy processes. *Prob. Surveys* **2,** 191–212.
[12] Birkner, M. *et al.* (2005). Alpha-stable branching and beta-coalescents. *Electron. J. Prob.* **10,** 303–325.
[13] Bolthausen, E. and Sznitman, A.-S. (1998). On Ruelle's probability cascades and an abstract cavity method. *Commun. Math. Phys.* **197,** 247–276.
[14] Delmas, J.-F. and Chen, Y.-T. (2012). Smaller population size at the MRCA time for stationary branching processes. *Ann. Prob.* **40,** 2034–2068.
[15] Duquesne, T. and Le Gall, J.-F. (2002). *Random Trees, Lévy Processes and Spatial Branching Processes* (Astérisque **281**). Societe Mathematique de France.
[16] Foucard, C. and Hénard, O. (2013). Stable continuous branching processes with immigration and Beta-Fleming-Viot processes with immigration. *Electron. J. Prob.* **18,** 1–21.
[17] Goldschmidt, C. and Martin, J. (2005). Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron. J. Prob.* **10,** 718–745.
[18] Gnedin, A., Iksanov, A. and Marynych, A. (2011). On lambda-coalescents with dust component. *J. Appl. Prob.* **48,** 1133–1151.
[19] Hoscheit, P. (2013). Fluctuations for the number of records on random binary trees. Submitted. Available at http://arxiv.org/abs/1212.5434v1.
[20] Iksanov, A. and Möhle, M. (2008). On the number of jumps of random walks with a barrier. *Adv. Appl. Prob.* **40,** 206–228.
[21] Janson, S. (2006). Random cutting and records in deterministic and random trees. *Random Structures Algorithms* **29,** 139–179.
[22] Kingman, J. F. C. (1982). The coalescent. *Stoch. Process. Appl.* **13,** 235–248.
[23] Le Gall, J.-F. (2006). Random real trees. *Ann. Fac. Sci. Toulouse Math.* **15,** 35–62.
[24] Pitman, J. (1999). Coalescents with multiple collisions. *Ann. Prob.* **27,** 1870–1902.
[25] Pitman, J. (2006). *Combinatorial Stochastic Processes* (Lectures Notes Math. **1875**). Springer, Berlin.