Year: 2010

A two-time-scale phenomenon for a fragmentation-coagulation process

Bertoin, J

Abstract: Consider two urns, A and B, where initially A contains a large number n of balls and B is empty. At each step, with equal probability, either we pick a ball at random in A and place it in B, or vice-versa (provided of course that A, or B, is not empty). The number of balls in B after n steps is of order n√, and this number remains essentially the same after n√ further steps. Observe that each ball in the urn B after n steps has a probability bounded away from 0 and 1 to be placed back in the urn A after n√ additional steps. So, even though the number of balls in B does not evolve significantly between n and n+n√, the precise contain of urn B does. This elementary observation is the source of an interesting two-time-scale phenomenon which we illustrate using a simple model of fragmentation-coagulation. Inspired by Pitman's construction of coalescing random forests, we consider for every n N a uniform random tree with n vertices, and at each step, depending on the outcome of an independent fair coin tossing, either we remove one edge chosen uniformly at random amongst the remaining edges, or we replace one edge chosen uniformly at random amongst the edges which have been removed previously. The process that records the sizes of the tree-components evolves by fragmentation and coagulation. It exhibits subaging in the sense that when it is observed after k steps in the regime k tn+sn√ with t>0 fixed, it seems to reach a statistical equilibrium as n→∞; but different values of t yield distinct pseudo-stationary distributions.

A TWO-TIME-SCALE PHENOMENON IN A FRAGMENTATION-COAGULATION PROCESS

JEAN BERTOIN[1]

Laboratoire de Probabilités, UPMC, 4 Place Jussieu, 75252 Paris Cedex 05, France
and

DMA, École Normale Supérieure, 45 rue d'Ulm, 75005 Paris, France.
Email: `jean.bertoin@upmc.fr`

Abstract

Consider two urns, A and B, where initially A contains a large number n of balls and B is empty. At each step, with equal probability, either we pick a ball at random in A and place it in B, or vice-versa (provided of course that A, or B, is not empty). The number of balls in B after n steps is of order $\sqrt{n}$, and this number remains essentially the same after $\sqrt{n}$ further steps. Observe that each ball in the urn B after n steps has a probability bounded away from 0 and 1 to be placed back in the urn A after $\sqrt{n}$ additional steps. So, even though the number of balls in B does not evolve significantly between n and $n + \sqrt{n}$, the precise contain of urn B does.

This elementary observation is the source of an interesting two-time-scale phenomenon which we illustrate using a simple model of fragmentation-coagulation. Inspired by Pitman's construction of coalescing random forests, we consider for every $n \in \mathbb{N}$ a uniform random tree with n vertices, and at each step, depending on the outcome of an independent fair coin tossing, either we remove one edge chosen uniformly at random amongst the remaining edges, or we replace one edge chosen uniformly at random amongst the edges which have been removed previously. The process that records the sizes of the tree-components evolves by fragmentation and coagulation. It exhibits subaging in the sense that when it is observed after k steps in the regime $k \sim tn + s\sqrt{n}$ with $t > 0$ fixed, it seems to reach a statistical equilibrium as $n \to \infty$; but different values of t yield distinct pseudo-stationary distributions.

1   Introduction

Aging refers to systems that seem to reach a statistical equilibrium in a certain regime depending on two time-scales: the age of the system and the duration of the observation. If the observation scale is much shorter that the age of the system scale, then the effect is, more precisely, called subaging. It arises in a variety of models in random media; see for instance Ben Arous and Černý [4] and references therein. The purpose of this note is to point out that subaging occurs in a process which evolves by fragmentation and coagulation, and is related to the work of Pitman [9] on coalescent random forests. It has its source in an elementary random urn dynamic which we now describe.

1.1   A two-time-scale phenomenon

Consider two urns, say A and B, and assume that at the initial time A contains n balls and B is empty. At each step $k = 1, 2, \ldots$, we flip a fair coin. If head comes up then we pick a ball in A uniformly at random and place it in B (provided of course that A is not empty, else we do nothing). Similarly, if tail comes up then we pick a ball in B uniformly at random and place it in A (provided of course that B is not empty, else we do nothing). Fix $s, t > 0$ arbitrarily. By the invariance principle, the numbers of balls in B after $\lfloor tn \rfloor$ steps and after $\lfloor tn + s\sqrt{n} \rfloor$ steps are both close to $\sqrt{n} R_t$, where R_t is a reflecting Brownian motion evaluated at time t, i.e. has the distribution of the absolute value of an $\mathcal{N}(0, t)$-variable. Precisely because B contains about $\sqrt{n}$ balls in that period, each ball in B after $\lfloor tn \rfloor$ steps has a probability bounded away from 0 and 1 to be selected and placed back in A during the next $\lfloor s\sqrt{n} \rfloor$ steps. In other words, even though the number of balls in B remains essentially unchanged between $\lfloor tn \rfloor$ and $\lfloor tn + s\sqrt{n} \rfloor$ steps, the precise contain of urn B evolves significantly.

Now imagine a stochastic process governed by the contain of urn B. Suppose that the one-dimension distributions of this process can be renormalized as $n \to \infty$ in such a way that they have a non-degenerate limit, say μ_r, when there are approximately $r\sqrt{n}$ balls in B. Let n be large, $t > 0$ fixed and let s vary. The process observed after $\lfloor tn + s\sqrt{n} \rfloor$ steps then seems to be in statistical equilibrium, in the sense that its one-dimensional distributions do not change much when s increases. More precisely the almost equilibrium law can be expressed as the mixture

$$\int_0^\infty \mu_r \mathbb{P}(R_t \in \mathrm{d}r) \,.$$

However this is only a pseudo-stationarity as this almost equilibrium distribution depends on the parameter t.

The rest of this work is devoted to the rigorous analysis of this two-time-scale phenomenon in the special case when the stochastic process alluded above is a fragmentation-coagulation process induced by a natural modification of Pitman's coalescing random forests [9].

1.2   Fragmentation-coagulation of random forests

Pitman considered the following random dynamics: one first picks a tree with n labeled vertices uniformly at random, and one deletes its edges one after the other, also uniformly at random. At each step the tree-component containing the edge which is deleted splits into two smaller trees, so after $k \leq n-1$ steps, we obtain a forest with $k+1$ trees, and the process terminates after $n-1$ steps when all vertices are isolated. Pitman's motivation for considering these edge-deletion dynamics stems from its connexion with the additive coalescence. The latter governs the evolution of a particle system in continuous times in which pairs of particles coalesce with a rate proportional to the sum of their masses. More precisely, Pitman pointed out that the process of the sizes of the tree-components in the forests resulting from the edge-deletion dynamics is a Markovian fragmentation chain, and that time-reversal yields the discrete-time skeleton of the additive coalescent started from n monomers (i.e. n atoms each having a unit mass).

On the one hand, it is well-known from the work of Aldous [1] that the uniform random tree on a set of n vertices can be rescaled (specifically edges by a factor $1/\sqrt{n}$ and masses of vertices by a factor $1/n$) and then converges weakly as $n \to \infty$ towards the Continuum Random Tree (for short, CRT). On the other hand, Evans and Pitman [7] proved that the additive coalescent started from a large number of monomers possesses a scaling limit, which is known as the standard additive coalescent. Roughly speaking, Aldous and Pitman [2] put the pieces together: they showed that the properly rescaled edge-deletion procedure on finite trees converges weakly to a Poissonian logging of the CRT. The latter induces the CRT fragmentation of masses, denoted here by $(\mathbf{F}_t)_{t \geq 0}$, and in turn this yields the standard additive coalescent upon time-reversal.

We consider in this work an evolution which combines edge-removal and edge-replacement, and should be viewed as an avatar of the urn dynamic depicted in the first part. In this direction it will be convenient to use marks on edges, agreeing that a mark on an edge means that this edge has been removed, while the absence of mark means that this edge is present. For each fixed n, we first pick a tree on n vertices uniformly at random and declare that initially all edges have no mark. At each step we flip a fair coin. If head comes up then we put a mark on one edge chosen uniformly at random amongst the un-marked edges (provided of course that there still remains at least one edge without mark; else we do nothing), while if tail comes up then we erase the mark of one edge chosen uniformly at random amongst the marked edges (provided of course that there exists such an edge; else we do nothing). For every integer $k \geq 0$, we denote by $\mathcal{F}_k^{(n)}$ the random forest which results from removing the marked edges after the k-th step and by $\mathbf{X}_k^{(n)}$ the collection of the sizes of the tree-components in $\mathcal{F}_k^{(n)}$ rescaled by a factor $1/n$ and ranked in the decreasing order. In a technical jargon, $\mathbf{X}_k^{(n)}$ is a random mass-partition, that is a decreasing sequence of positive real numbers with sum 1. Plainly the addition of a mark corresponds to a fragmentation event and erasure to a coalescence; in other words the chain $(\mathbf{X}_k^{(n)})_{k \in \mathbb{N}}$ evolves by fragmentation and coagulation.

Our purpose is to investigate the asymptotic behavior in distribution of $\mathbf{X}_k^{(n)}$ as $k, n \to \infty$. This is easy as far as only one-dimensional distributions are concerned. Indeed Donsker's invariance principle implies that when n is large, the number of marks after $\lfloor tn \rfloor$ steps is about $\sqrt{n} R_t$, where $(R_t : t \geq 0)$ is a reflected Brownian motion. It then follows from results in Aldous and Pitman [2] that for every fixed $t > 0$, $\mathbf{X}_{\lfloor tn \rfloor}^{(n)}$ converges weakly as $n \to \infty$ towards $\mathbf{F}_{R_t}$, the CRT mass-fragmentation observed at the independent random time R_t.

It turns out that things are more subtle for finite-dimensional marginals. The simple rescaling of times by a factor n is too crude and the asymptotic behavior in law is better revealed in the finer regime $tn + s\sqrt{n}$. Indeed our main result says that for every fixed $t > 0$, the process $(\mathbf{X}_{\lfloor tn + s\sqrt{n} \rfloor}^{(n)})_{s \in \mathbb{R}}$ converges weakly in the sense of finite-dimensional marginals towards some non-degenerate stationary process with stationary law given by the distribution of $\mathbf{F}_{R_t}$. Thus the fragmentation-coagulation process $\mathbf{X}^{(n)}$ exhibits subaging, in the sense that this process seems to reach a statistical equilibrium in the regime $tn + s\sqrt{n}$ when n is large and t fixed.

The plan of the rest of this note is as follows. Our main result is stated and proved in Section 2 after recalling some notions on the CRT. Finally Section 3 is devoted to some comments, complements and open questions. Our approach owes much to the construction by Aldous and Pitman of the standard additive coalescent via Poissonian cuts on the skeleton of a Continuum Random Tree.

2   Main result

We start by recalling some elements on the CRT, refering the reader to Aldous [1], Evans [6] and Le Gall [8] for background, and then state our main result.

Let $\mathcal{T}$ be a Brownian CRT; in particular $\mathcal{T}$ is almost surely a compact metric space which has the structure of a real tree. Extremities of $\mathcal{T}$ are called leaves; in other words $a \in \mathcal{T}$ is a leaf means that if a lies on some path joining two points b, c in $\mathcal{T}$, then necessarily $a = b$ or $a = c$. The subset of leaves is totally disconnected; its complement is referred to as the skeleton. One endows $\mathcal{T}$ with a probability measure μ carried by the subset of leaves and with a sigma-finite length measure λ carried by the skeleton. More precisely the distance between to points in $\mathcal{T}$, say a, b, is given by $\lambda([a, b])$, where $[a, b]$ stands for the path joining a to b in $\mathcal{T}$.

We next introduce marks on the skeleton of $\mathcal{T}$ that appear and disappear randomly as time passes at some constant rate. Specifically, we fix a parameter $r > 0$ and introduce a Poisson point process on $\mathbb{R} \times (0, \infty) \times \mathcal{T}$ with intensity

$$\frac{1}{2}\mathrm{d}s \otimes \frac{1}{2r}\mathrm{e}^{-u/2r}\mathrm{d}u \otimes \mathrm{d}\lambda \,.$$

An atom (s, u, x) should be interpreted as follows: at time s a mark appears at location x on the skeleton and is erased at time $s + u$. In words, on any portion of the skeleton with length measure ℓ, marks appear with rate $\ell/2$ and the lifetime of each mark is exponentially distributed with expectation $2r$, independently of the other marks. For every $s \in \mathbb{R}$, we denote

by $\mathcal{M}_{r,s}$ the random point measure on the skeleton of $\mathcal{T}$ induced by the marks present at time s. It is immediate to check that for each fixed s, conditionally on $\mathcal{T}$, $\mathcal{M}_{r,s}$ is a Poisson point measure on the skeleton of $\mathcal{T}$ with intensity $r\lambda$.

Following Aldous and Pitman [2], for every $s \in \mathbb{R}$ we use the atoms of the Poisson random measure $\mathcal{M}_{r,s}$ to decompose the set of leaves of $\mathcal{T}$. More precisely, we decide that two leaves belong to the same component if and only if $\mathcal{M}_{r,s}$ has no mass on the path that joins those leaves (note that the probability of this event is $\exp(-r\ell)$ where ℓ is the length of the path between those two leaves). The components are closed in the subset of leaves, and we denote by $\mathbf{Y}_{r,s}$ the sequence of their μ-masses ranked in the decreasing order. We stress that for every $s \in \mathbb{R}$, $\mathbf{Y}_{r,s}$ has the same distribution as $\mathbf{F}_r$, the CRT mass-fragmentation evaluated at time r, which was mentioned in the Introduction and is described in Theorem 4 of [2].

Recall from the Introduction the construction of the fragmentation-coagulation chain $\mathbf{X}_k^{(n)}$. We are now able to state

Theorem 1 *For each fixed $t > 0$, the fragmentation-coagulation process*

$$(\mathbf{X}_{\lfloor tn+s\sqrt{n}\rfloor}^{(n)})_{s\in\mathbb{R}}$$

converges weakly in the sense of finite dimensional distributions as $n \to \infty$ towards the mixed process

$$(\mathbf{Y}_{R_t,s})_{s\in\mathbb{R}},$$

where R_t denotes a random variable on $(0,\infty)$ which is independent of the preceding processes and has the distribution of a reflected Brownian motion at time t, viz.

$$\mathbb{P}(R_t \in \mathrm{d}r) = \sqrt{\frac{2}{t\pi}}\exp\left(-\frac{r^2}{2t}\right)\mathrm{d}r\,, \qquad r > 0\,.$$

The rest of this section is devoted to the proof of Theorem 1; the scheme of the argument is adapted from Aldous and Pitman [2]. We first recall the formulation of the convergence of uniform random trees towards the CTR via reduced trees.

Given $\mathcal{T}$, we sample a sequence $U_1,\ldots$ of i.i.d. random leaves according to the law μ, and for every integer $i \geq 1$, we denote by $\mathcal{R}(\infty,i)$ the subtree reduced to the first i leaves, i.e. the smallest connected subset of $\mathcal{T}$ containing $U_1,\ldots,U_i$. The reduced tree $\mathcal{R}(\infty,i)$ is a combinatorial tree (simple graph with no cycles) with leaves labeled by $1,\ldots,i$ and some unlabeled internal nodes. The paths between two adjacent internal nodes or between a leaf and an adjacent internal node are called edges. The lengths of edges are given by the length measure λ on $\mathcal{T}$, and the joint distribution of the shape and the edge-lengths is described by Lemma 21 in [1].

For every integer $n \geq 2$, we also consider a uniform random tree $\mathcal{T}_n$ on a set of n vertices, say $\{1,\ldots,n\}$, and assign length $1/\sqrt{n}$ to every edge. For every $1 \leq i \leq n$, we denote by $\mathcal{R}(n,i)$ the sub-tree reduced to the first i vertices. We agree that internal nodes with degree 2 are

discarded, so that the edge-length between two adjacent vertices in $\mathcal{R}(n,i)$ is $(1+j)/\sqrt{n}$ with j the number of internal nodes with degree 2 lying on the path connecting these vertices. It has been shown by Aldous (see (49) in [1] or Lemma 9 in [2]) that for every fixed i,

$$\mathcal{R}(n,i) \Longrightarrow \mathcal{R}(\infty,i) \quad \text{as } n \to \infty, \tag{1}$$

in the sense of weak convergence of the joint distributions of shape and edge lengths.

We then add and erase marks randomly on $\mathcal{T}_n$ as explained in the Introduction. For each integer $k \geq 0$, we denote by $\mathcal{M}_k^{(n)}$ a random point measure on $\mathcal{T}_n$ that assigns a unit mass to each marked edge after k steps, and for every $i \leq n$ by $\mathcal{M}_k^{(n,i)}$ the restriction of $\mathcal{M}_k^{(n)}$ to $\mathcal{R}(n,i)$. Similarly, we also denote by $\mathcal{M}_{r,s}^{(\infty,i)}$ the restriction of the Poisson point measure $\mathcal{M}_{r,s}$ to the reduced tree $\mathcal{R}(\infty,i)$. The key to Theorem 1 lies in the following limit theorem which can be viewed as a multi-dimensional extension of Equation (18) in [2]. Essentially it is a consequence of the law of rare events combined with the convergence of reduced trees. Recall that the reduced trees are defined by their shapes and edge-lengths, and that the set of shapes of trees with i vertices is finite. The reduced trees $\mathcal{R}(n,i)$ and $\mathcal{R}(\infty,i)$ should thus be viewed as random variables with values in some Polish space, and weak convergence of random point measures should be understood in this setting.

Lemma 1 *Fix $t > 0$ and an integer i. The process of random point measures on the reduced trees*

$$(\mathcal{M}_{\lfloor tn+s\sqrt{n}\rfloor}^{(n,i)}, \mathcal{R}(n,i))_{s\in\mathbb{R}}$$

converges weakly in the sense of finite dimensional distributions as $n \to \infty$ towards the mixed process of point measures on the reduced CRT

$$(\mathcal{M}_{R_t,s}^{(\infty,i)}, \mathcal{R}(\infty,i))_{s\in\mathbb{R}},$$

where R_t denotes a random variable on $(0,\infty)$ which is independent of the preceding processes and has the distribution of a reflected Brownian motion at time t, viz.

$$\mathbb{P}(R_t \in \mathrm{d}r) = \sqrt{\frac{2}{t\pi}} \exp\left(-\frac{r^2}{2t}\right) \mathrm{d}r, \qquad r > 0.$$

Proof: We first deal with the one-dimensional convergence in the statement, rephrasing (and slightly developing) the argument for Equation (18) in [2].

By Skorohod's representation, we may assume that the convergence (1) for the reduced trees holds almost surely and not merely in distribution. Thus with a probability close to 1 when n is large, the shape of $\mathcal{R}(n,i)$ coincides with that of $\mathcal{R}(\infty,i)$, and the edge lengths of $\mathcal{R}(n,i)$ and of $\mathcal{R}(\infty,i)$ are close. We denote by $N_k^{(n)}$ the total number of marks on $\mathcal{T}_n$ after k steps and consider a sequence $(r_n)_{n\in\mathbb{N}}$ of integers with $r_n \sim r\sqrt{n}$ for some $r > 0$. We first work for each n conditionally on the event that $N_{\lfloor tn+s\sqrt{n}\rfloor}^{(n)} = r_n$.

Recall that $\mathcal{T}_n$ has $n-1$ edges, each of length $1/\sqrt{n}$, so the number of edges in a segment of $\mathcal{T}_n$ is $\sqrt{n}$ times the length of that segment. We stress that when n is large, the number of edges in the reduced tree $\mathcal{R}(n,i)$ is of order $\sqrt{n} = o(n)$ and the number of marked edges in $\mathcal{R}(n,i)$ of order $O(1) = o(r_n)$. This is important to justify the claims of asymptotic independence which will be made below.

As $(n-1)^{-1}r_n\sqrt{n} \sim r$, it follows from the law of rare events that when n is large, the number of marked edges after $\lfloor tn+s\sqrt{n}\rfloor$ steps on a segment in $\mathcal{T}_n$ is approximately Poisson distributed with parameter given by r times the length of that segment, and further to disjoint segments correspond asymptotically independent Poisson variables. This entails that the conditional distribution of $\mathcal{M}^{(n,i)}_{\lfloor tn+s\sqrt{n}\rfloor}$ converges weakly as $n \to \infty$ towards a Poisson random measure on $\mathcal{R}(\infty,i)$ with intensity $r\lambda$, i.e.

$$\mathcal{L}\left(\mathcal{M}^{(n,i)}_{\lfloor tn+s\sqrt{n}\rfloor} \mid N^{(n)}_{\lfloor tn+s\sqrt{n}\rfloor} = r_n\right) \implies \mathcal{M}^{(\infty,i)}_{r,s}, \tag{2}$$

where the notation $\mathcal{L}(Z \mid \Lambda)$ refers to the conditional law of the variable Z given the event Λ.

We next present the main lines of the argument for extending (2) to multi-dimensional convergence by analyzing the evolution of the random point measures as s increases. It is readily checked that with probability one

$$N^{(n)}_{\lfloor tn+(s+s')\sqrt{n}\rfloor} \sim r_n \sim r\sqrt{n}$$

uniformly for $s' \geq 0$ in an arbitrary bounded interval. Thus for every $k = nt + O(\sqrt{n})$, each atom of $\mathcal{M}^{(n,i)}_k$ has a probability close to $1/(2r_n) \sim 1/(2r\sqrt{n})$ to be erased at the next step, where the factor $1/2$ accounts for the probability that head turns up when the fair coin is flipped. The probability that a given atom of $\mathcal{M}^{(n,i)}_{\lfloor tn+s\sqrt{n}\rfloor}$ has not been erased after $\lfloor s'\sqrt{n}\rfloor$ further steps is close to

$$(1 - 1/(2r_n))^{\lfloor s'\sqrt{n}\rfloor} \sim \exp(-s'/(2r))$$

when n is large; in other words if one unit of time corresponds to $\sqrt{n}$ steps, each atom of $\mathcal{M}^{(n,i)}_{\lfloor tn+s\sqrt{n}\rfloor}$ is removed after a time which is approximately exponentially distributed with mean $2r$. A similar argument shows that asymptotically, each atom of $\mathcal{M}^{(n,i)}_{\lfloor tn+s\sqrt{n}\rfloor}$ is removed or not after $\lfloor s'\sqrt{n}\rfloor$ more steps independently on the other atoms.

On the other hand, at each step, a mark appears on an un-marked edge with probability close to $1/(2(n - r_n)) \sim 1/(2n)$. Recalling that any given mark is also erased at each step with probability close to $1/(2r\sqrt{n})$, and neglecting the event of multiple appearances and erasures of a mark whose probability is of lower order, we deduce that the probability that an edge

with no mark after $\lfloor tn + s\sqrt{n} \rfloor$ steps be marked after $\lfloor s'\sqrt{n} \rfloor$ further steps is

$$
\sum_{j=1}^{\lfloor s'\sqrt{n} \rfloor} \frac{1}{2n} \left(1 - \frac{1}{2r\sqrt{n}}\right)^j \;=\; \frac{r}{\sqrt{n}} \left(\left(1 - \frac{1}{2r\sqrt{n}}\right) - \left(1 - \frac{1}{2r\sqrt{n}}\right)^{\lfloor s'\sqrt{n} \rfloor + 1} \right)
$$
$$
\sim \; \frac{r}{\sqrt{n}} \left(1 - \exp\left(-\frac{s'}{2r}\right) \right).
$$

It then follows from the law of rare events that for any given segment of $\mathcal{R}(n,i)$ with length ℓ (i.e. with $\ell\sqrt{n}$ edges), the number of marked edges after $\lfloor tn + (s + s')\sqrt{n} \rfloor$ steps which were un-marked after $\lfloor tn + s\sqrt{n} \rfloor$ steps is approximately Poisson with parameter

$$
r\ell \left(1 - \exp\left(-\frac{s'}{2r}\right) \right).
$$

Further, one checks readily that the evolutions of marks on a given finite sequence of disjoint segments are asymptotically independent.

Putting the pieces together, this shows that when n is large, the distribution of $\mathcal{M}^{(n,i)}_{\lfloor tn+(s+s')\sqrt{n} \rfloor}$ given $\mathcal{M}^{(n,i)}_{\lfloor tn+s\sqrt{n} \rfloor}$ is close to that of a measure obtained from $\mathcal{M}^{(n,i)}_{\lfloor tn+s\sqrt{n} \rfloor}$ by removing each atom with probability $1 - \exp(-s'/(2r))$ independently one of the others (i.e. by thinning), and further adding an independent Poisson measure on $\mathcal{R}(n,i)$ with intensity $r\left(1 - \exp\left(-s'/2r\right)\right)\lambda$. Comparing with the evolution of the random point measure $\mathcal{M}^{(\infty,i)}_{r,s}$ when s increases, we see that (2) can be extended as follows: for every $s' \geq 0$ we have

$$
\mathcal{L}\left(\left(\mathcal{M}^{(n,i)}_{\lfloor tn+s\sqrt{n} \rfloor}, \mathcal{M}^{(n,i)}_{\lfloor tn+(s+s')\sqrt{n} \rfloor} \right) \mid N^{(n)}_{\lfloor tn+s\sqrt{n} \rfloor} = r_n \right) \implies \left(\mathcal{M}^{(\infty,i)}_{r,s}, \mathcal{M}^{(\infty,i)}_{r,s+s'} \right).
$$

An iteration based on the obvious Markov property of the processes of the radom measures involved (in the case of $(\mathcal{M}^{(\infty,i)}_{r,s})_{s \in \mathbb{R}}$, this Markov property follows from the absence of memory of the exponential variables that specify the lifetime of the marks) enables us to conclude that

$$
\mathcal{L}\left(\left(\mathcal{M}^{(n,i)}_{\lfloor tn+s'\sqrt{n} \rfloor} \right)_{s' \geq s} \mid N^{(n)}_{\lfloor tn+s\sqrt{n} \rfloor} = r_n \right) \implies \left(\mathcal{M}^{(\infty,i)}_{r,s'} \right)_{s' \geq s}, \tag{3}
$$

in the sense of finite dimensional distributions.

All that we need now is to get rid of the conditioning in (3), which is straightforward. Indeed Donsker's invariance principle shows that there is the weak convergence

$$
\frac{1}{\sqrt{n}} N^{(n)}_{\lfloor tn+s\sqrt{n} \rfloor} \implies R_t,
$$

where R_t is a random variable distributed as in the statement. On the other hand, it is easy seen from the construction of the random point measures $\mathcal{M}_{r,s}$ that the finite-dimensional distributions of the process $\left(\mathcal{M}^{(\infty,i)}_{r,s'} \right)_{s' \geq s}$ depend continuously on the parameter r. We derive

from above that

$$\left(\mathcal{M}^{(n,i)}_{\lfloor tn+s'\sqrt{n}\rfloor}\right)_{s'\geq s} \implies \left(\mathcal{M}^{(\infty,i)}_{R_t,s'}\right)_{s'\geq s}$$

in the sense of finite dimensional distributions, which is our claim. $\qquad\square$

We now explain how Theorem 1 follows from Lemma 1, focussing on one-dimensional distributions as the multi-dimensional case is similar but with heavier notation.

For every integers $n, k \geq 1$, we denote by $\Pi^{(n)}_k$ the random partition of $\{1, \dots, n\}$ induced by the marks on edges of $\mathcal{T}_n$ after k steps, that is the blocks of $\Pi^{(n)}_k$ are characterized by the property that there is no mark on the paths connecting two vertices in the same block. By definition, $\mathbf{X}^{(n)}_k$ is the sequence of the sizes of the blocks of $\Pi^{(n)}_k$ ranked in the decreasing order and rescaled by a factor $1/n$.

Similarly, for every $s \in \mathbb{R}$, we denote by $\Pi^{(\infty)}_{r,s}$ the random partition of $\mathbb{N}$ such that two integers j, j' belong to the same block of $\Pi^{(\infty)}_{r,s}$ if and only if there is no atom of the independent mixed-Poisson random measure $\mathcal{M}_{r,s}$ on the path in $\mathcal{T}$ from the leaf U_j to the leaf $U_{j'}$. Plainly the random partition $\Pi^{(\infty)}_{r,s}$ is exchangeable and the asymptotic frequencies of its blocks are given by $\mathbf{Y}_{r,s}$, i.e. the sequence of the μ-masses of the connected components of the CRT $\mathcal{T}$ cut at the atoms of $\mathcal{M}_{r,s}$.

For every integer $i \leq n$, we also denote by $\Pi^{(n,i)}_k$ (respectively by $\Pi^{(\infty,i)}_{r,s}$) the restriction of $\Pi^{(n)}_k$ (respectively of $\Pi^{(\infty)}_{r,s}$) to the first i vertices. Plainly these restricted partition only depend on the reduced tree $\mathcal{R}(n,i)$ and $\mathcal{R}(\infty,i)$, and the marks on their edges after k steps and the atoms of the random measure $\mathcal{M}^{(\infty,i)}_{r,s}$, respectively. Lemma 1 implies that for every i, in the obvious notation, when $n \to \infty$ there is the weak convergence

$$\Pi^{(n,i)}_{\lfloor tn+s\sqrt{n}\rfloor} \implies \Pi^{(\infty,i)}_{R_t,s}. \tag{4}$$

Repeating the argument of Aldous and Pitman for proving Theorem 3 in [2] enables us to conclude from (4) and the preceding observations that

$$\mathbf{X}^{(n)}_{\lfloor tn+s\sqrt{n}\rfloor} \implies \mathbf{Y}_{R_t,s},$$

which is the one-dimensional version of Theorem 1. The multidimensional case is similar, using the full strength of Lemma 1.

3   A comment, a complement, and an open question

The stationary limiting process $(\mathbf{Y}_{R_t,s})_{s\in\mathbb{R}}$ which appears in Theorem 1 is expressed as a mixture. The mixing variable R_t may be thought of as the effective age of the system as it represents the intensity of cuts along the skeleton of the CRT. In this direction, we note that the variables R_t are stochastically increasing with t. We also mention that R_t can be recovered from a sample of $\mathbf{Y}_{R_t,s}$. Indeed, it follows easily from Theorem 4 in Aldous and Pitman [2] and the law of large numbers for Poisson processes that if $F_t(i)$ denotes the i-th largest term

of $\mathbf{F}_t$, the CRT fragmentation of masses observed at time t, then with probability one

$$F_t(i) \sim \frac{2}{\pi t^2 i^2} \qquad \text{as } i \to \infty \,.$$

Equivalently

$$Y_{r,s}(i) \sim \frac{2}{\pi r^2 i^2} \qquad \text{as } i \to \infty \,,$$

where $Y_{r,s}(i)$ denotes the i-th largest term of $\mathbf{Y}_{r,s}$ and we conclude that

$$R_t = \lim_{i \to \infty} \sqrt{\frac{2}{\pi Y_{R_t,s}(i) i^2}} \qquad \text{a.s.}$$

We used uniform random trees $\mathcal{T}_n$ merely to stick to Pitman's original framework [9]. Nonetheless the same results hold if we replace $\mathcal{T}_n$ by any other sequence of discrete random tree which converges to the Brownian CRT after rescaling edge lengths by a factor $1/\sqrt{n}$ and masses of vertices by a factor $1/n$ (for instance critical Galton-Watson trees with finite variance and conditioned to have total size n). In the same vein, the results of this note can be extended to certain sequences of so-called birthday trees. Indeed, Camarri and Pitman [5] have established the weak convergence of suitably rescaled birthday trees towards certain Inhomogeneous Continuum Random Trees. On the other hand, dynamics of edge-deletion for birthday trees bears the same connection to the additive coalescence as uniform random trees, except that the initial distribution of masses is inhomogeneous. The asymptotic behavior of the latter has been characterized by Aldous and Pitman [3], in the study of the entrance boundary of the additive coalescence. We thus have all the ingredients needed to apply the arguments of the present work to this more general setting. Of course, the limiting processes will then have different distributions.

Our aim in this work was to point at the phenomenon of subaging in a fragmentation-coagulation process. The model that we used for this purpose is easy to deal with although somewhat artificial. There are other discrete models for the evolution of random forests which may be more natural, but are also much harder to investigate. Here is an example, which inspired by the subtree prune and regraft algorithm; see Chapter 9 of Evans [6]. We now work with rooted forests on n vertices, that is each tree has one distinguished vertex that serves as the root. At each step we flip a fair coin. With probability $1/2$ we delete an edge chosen uniformly at random in this forest. This disconnects the tree containing that edge into two rooted subtrees. With probability $1/2$, we create a new edge between a vertex chosen uniformly at random and the root of a tree chosen uniformly at random amongst the trees to which the chosen vertex does not belong. Our result suggests that a similar subaging phenomenon might occur at the same scale as in the present study. Proving or disproving this property would be interesting, but does not seem easy.

References

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

[2] D. J. Aldous and J. Pitman (1998). The standard additive coalescent. *Ann. Probab.* **26**, 1703-1726.

[3] D. J. Aldous and J. Pitman (2000). Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. *Probab. Theory Related Fields* **118**, 455-482.

[4] G. Ben Arous and J. Černý (2006). Dynamics of trap models. In: A. Bovier et al. (Eds) *Lecture notes of Les Houches summer school 2005*. Mathematical Statistical Physics vol. **LXXXIII**, pp. 331-394, Elsevier.

[5] M. Camarri and J. Pitman (2000). Limit distributions and random trees derived from the birthday problem with unequal probabilities. *Electronic J. Probab.* **5**, 1-18.

[6] S. N. Evans (2008). *Probability and real trees*. École d'Été de Probabilités de Saint-Flour XXXV-2005. Lecture Notes in Mathematics **1920**, Springer.

[7] S. N. Evans and J. Pitman (1998). Construction of Markovian coalescents. *Ann. Inst. H. Poincaré, Probabilités et Statistiques* **34**, 339-383.

[8] J.-F. Le Gall (2005). Random trees and applications. *Probability Surveys* **2**, 245-311 (electronic).

[9] J. Pitman (1999). Coalescent random forests. *J. Comb. Theory A.* **85**, 165-193.