A note on the fragmentation of a stable tree

Philippe Marchal[1]

[1]*DMA, École Normale Supérieure, 75005 Paris, France*

We introduce a recursive algorithm generating random trees, which we identify as skeletons of a continuous, stable tree. We deduce a representation of a fragmentation process on these trees.

Keywords: stable tree, fragmentation

1　Introduction

The aim of this note is to introduce a recursive construction of a class of continuous trees, which can be identified as stable trees associated with stable processes.

Recall that stable trees of index $\alpha \in (1,2)$ can be viewed, loosely speaking, as scaling limits of large Galton-Watson trees where the probability to have n children is asymptotic to $n^{-\alpha-1}$ as $n \to \infty$. Moreover, as discrete trees can be associated with Lukasiewicz paths by the the Lukasiewicz correspondence, stable trees can be associated likewise with stable, spectrally positive Lévy processes. For a detailed account, see Duquesne–Le Gall (5).

For a discrete tree, saying that two leaves are in the same block if they share a common ancestor at height, say h, yields a partition of the set of leaves. When h increases, the partition becomes finer. The same kind of procedure can be applied to continuous trees and, under suitable conditions, this may produce a fragmentation process as defined in Bertoin (4). The case of stable trees was studied by Miermont (8), using the relations between these trees and their associated stable processes.

When $\alpha = 2$, the corresponding stable tree is Aldous' classical continuum random tree (1) and its fragmentation, first studied by Bertoin (3), is in some sense the dual of the standard additive coalescent constructed by Aldous-Pitman (2).

Our goal is to provide an alternative approach to the fragmentation of stable trees based on a recursive construction of stable trees, which can be described as follows. Choose a parameter $\alpha \in (1,2]$. Then apply the

RANDOM GROWTH ALGORITHM.

Begin with a tree T_1 with two vertices, A_0, A_1. At each step, assign to each edge a weight $\alpha - 1$ and to each node of degree $d \geq 3$ a weight $d - 1 - \alpha$. Then recursively, to construct T_{n+1} from T_n:

I. Choose at random either an edge e or a node v of degree ≥ 3, with probability proportional to the weight.

II. If an edge e has been chosen, split e into two edges e_1, e_2 and denote by V_{n+1} the vertex between e_1 and e_2. Then add a new vertex A_{n+1} and say that A_{n+1} is connected to V_{n+1}.

III. If a node v has been chosen, add a new vertex A_{n+1} and say that A_{n+1} is connected to v.

Of course, all the choices are assumed independent. Hence, at each step n, either a new leaf, adjacent to an already existing internal node, appears, or a new leaf together with an internal node is created.

In the case when $\alpha = 2$, the weight of internal nodes is always zero and therefore the algorithm generates random binary trees. Under this form, the algorithm was first introduced by Rémy (10). It is shown in (7) that using the Lukasiewicz correspondence between trees and Lukasiewicz paths, the sequence of random trees becomes a sequence of excursions of the simple random walk which converges almost surely to a Brownian excursion.

In the general case when $1 < \alpha \le 2$, we shall see that, after proper rescaling, when $n \to \infty$, the trees T_n converge, at least in some weak sense, to a continuous tree T_∞. More specifically, we can define a distance on T_∞ by normalizing the distances on T_n and T_∞ can be identified as a stable tree as in Duquesne-Le Gall (5). We can also a define a probability measure on T_∞ by normalizing the counting measure on the leaves of T_n. Moreover, we shall see that the recursive construction enables us to analyze the fragmentation of these trees, which provides an alternative approach different from Miermont's paper (8).

The next section studies the law of the trees T_n. Section 3 is devoted to the limit as $n \to \infty$. Finally, the fragmentation associated to the limiting tree is the topic of Section 4.

2   The discrete tree

2.1   Preliminaries

Let $\widetilde{I} = \{k, V_k \text{ exists}\}$ be the set of times at which an internal node is created and let

$$T_\infty = \bigcup_{n \ge 0} \{A_n\} \cup \bigcup_{n \in \widetilde{I}} \{V_n\}$$

If v is a vertex of a tree, $\deg(v)$ is the degree of v, i.e. the number of edges incident to v. A vertex v is an internal vertex if $\deg(v) > 1$ and a leaf if $\deg(v) = 1$. The set of internal vertices is denoted $I(T)$.

We denote by d_n the natural distance in the tree T_n. Recall that, as a metric space, a tree T satisfies the four-point condition: for all $x, y, z, t \in T$,

$$d(x, y) + d(z, t) \le \max[d(x, z) + d(y, t), d(y, z) + d(x, t)]$$

(for more details, see for instance (6)). If T is a tree, any two points a, b define a segment $[a, b]$ by the following

$$[a, b] = \{v \in T, d(a, b) = d(a, v) + d(v, b)\}$$

If $v \in I(T)$, one can define subtrees of T cut at v as follows. Say that a, b are in the same subtree cut at v if $v \notin [a, b]$. It is easily seen that this defines an equivalence relation on $T - \{v\}$. The corresponding equivalence classes are the subtrees of T cut at v.

We call a labelled tree with n leaves a tree with n leaves where the leaves are numbered from 1 to n. So T_n is a labeled tree with n leaves.

2.2   The law of the discrete tree

To characterize of the law of T_n, set

- $p_1 = 0$

- $p_2 = (\alpha - 1)$

- for $n \geq 3$,

$$p_n = (\alpha - 1)(2 - \alpha) \ldots (n - 1 - \alpha)$$

Then

Theorem 1 *For every integer $k \geq 2$, if $\mathcal{T}$ is a labelled tree with $k + 1$ leaves,*

$$\mathbb{P}(T_k = \mathcal{T}) = c_k \prod_{v \in I(\mathcal{T})} p_{\deg(v) - 1}$$

where the c_k are constants defined by induction: $c_1 = 1$ and

$$\frac{c_k}{c_{k-1}} = k\alpha - 1$$

Proof

We first observe that after k iterations of the algorithm, the total weight given by the algorithm equals $(k + 1)\alpha - 1$. In particular, this does not depend on the shape of the tree.

We want to prove the theorem by induction on k. It is obvious for $k = 1, 2$. If it is true for $k \geq 2$, consider a labelled tree $\mathcal{T}$ with $k + 1$ leaves. Denote by L the $(k+1)$-th leaf of $\mathcal{T}$ and by V the only vertex connected to L. Also, let $\mathcal{T}_-$ be the labelled tree with k leaves obtained by removing L and removing V if V had degree 3 in $\mathcal{T}$. That is, if V had degree 3 in $\mathcal{T}$ and was connected to L, L', L", then in $\mathcal{T}_-$, we suppress L and V and L' is connected to L". By the induction hypothesis,

$$\mathbb{P}(T_{k-1} = \mathcal{T}_-) = c_{k-1} \prod_{v \in I(\mathcal{T}_-)} p_{\deg(v) - 1} \tag{1}$$

and since

$$\mathbb{P}(T_k = \mathcal{T}) = \mathbb{P}(T_{k-1} = \mathcal{T}_-)\mathbb{P}(T_k = \mathcal{T}|T_{k-1} = \mathcal{T}_-) \tag{2}$$

we are led to compute $\mathbb{P}(T_k = \mathcal{T}|T_{k-1} = \mathcal{T}_-)$. We have to distinguish between two cases. Either $\deg(V) = 3$ in $\mathcal{T}$ or $\deg(V) > 3$ in $\mathcal{T}$.

In the first case, V does not exist in $\mathcal{T}_-$ but at the k-th step, the random growth algorithm choses an edge and adds a leaf (namely, L) and an internal node (namely, V). In that case,

$$\mathbb{P}(T_k = \mathcal{T}|T_{k-1} = \mathcal{T}_-) = \frac{\alpha - 1}{k\alpha - 1} \tag{3}$$

Moreover, we also have

$$\prod_{v \in I(\mathcal{T})} p_{\deg(v)-1} = \left(\prod_{v \in I(\mathcal{T}_-)} p_{\deg(v)-1} \right) p_{\deg(V)-1}$$

$$= \left(\prod_{v \in I(\mathcal{T}_-)} p_{\deg(v)-1} \right) (\alpha - 1)$$

$$= \left(\prod_{v \in I(\mathcal{T}_-)} p_{\deg(v)-1} \right) (k\alpha - 1)\mathbb{P}(T_k = \mathcal{T}|T_{k-1} = \mathcal{T}_-)$$

and the formula of Theorem 1 is verified.

In the second case, V already exists in $\mathcal{T}_-$ but has degree $\deg(V) - 1$ in $\mathcal{T}_-$. Then at the k-th step, the random growth algorithm choses V and adds the leaf L, adjacent to V. Thus

$$\mathbb{P}(T_k = \mathcal{T}|T_{k-1} = \mathcal{T}_-) = \left(\frac{\deg(V) - 1 - \alpha}{k\alpha - 1} \right) \tag{4}$$

Next,

$$\prod_{v \in I(\mathcal{T})} p_{\deg(v)-1} = \left(\prod_{v \in I(\mathcal{T}_-)} p_{\deg(v)-1} \right) \left(\frac{p_{\deg(V)-1}}{p_{\deg(V)-2}} \right)$$

$$= \left(\prod_{v \in I(\mathcal{T}_-)} p_{\deg(v)-1} \right) (\deg(V) - 1 - \alpha)$$

$$= \left(\prod_{v \in I(\mathcal{T}_-)} p_{\deg(v)-1} \right) (k\alpha - 1)\mathbb{P}(T_k = \mathcal{T}|T_{k-1} = \mathcal{T}_-)$$

and again, the formula of Theorem 1 is verified.

$\square$

A consequence of Theorem 1 is the following. It is intuitively clear from the definition of the algorithm that "different parts of the tree T evolve independently". Theorem 1 enables us to give a formal proof of this idea, using the expression of the probability of a given tree as a product. For instance, suppose that $i, j \in \widetilde{I}$ and and that $V_i \notin [A_0, V_j]$, $V_j \notin [A_0, V_i]$. Then Theorem 1 shows that the subtrees above V_i and V_j are independent, conditionally on their sizes. More general results in this direction hold, although stating them formally in all generality would be a bit tedious. We shall use this remark in Section 4.

2.3   Identification with the stable tree

Given a finite labelled tree T, the number of planar embeddings of T is

$$\prod_{v \in I(T)} \deg(v) - 1$$

So let T'_n be a random random planar version of T_n. Then if $\mathcal{T}'$ is a planar tree with n leaves, Theorem 1 entails that

$$\mathbb{P}(T'_n = \mathcal{T}') = c_k \prod_{v \in I(\mathcal{T})} p'_{\deg(v)-1}$$

with the same c_k as in the statement of Theorem 1 and with

- $p'_1 = 0$

- $p'_2 = (\alpha - 1)/2$

- for $n \geq 3$,

$$p'_n = \frac{(\alpha - 1)(2 - \alpha) \ldots (n - 1 - \alpha)}{n!}$$

This is exactly the law of a random planar tree obtained by sampling n leaves from a continuous, stable tree. See (5).

3 Convergence to the continuous tree

3.1 Distances in the limiting tree

Theorem 2 *Almost surely, for every $u, v \in T_\infty$, the limit*

$$d(u, v) = \lim_{n \to \infty} \frac{d_n(u, v)}{n^{(\alpha-1)/\alpha}}$$

exists and d is a distance on T_∞ satisfying the four-point condition.

It follows from the theorem that the closure of T_∞ is almost surely a continuous tree. Of course, it would be stronger to show the convergence in the Gromov-Hausdorff tolopogy, but this would require a uniform control of the convergence for all pair of vertices. We shall not enter into these considerations here. Moreover, since the random, planar embeddings of the T_n are skeletons of a stable tree, it should be clear that, loosely speaking, a random planar embedding of T_∞ is a stable tree. Here again, we shall not enter into these details.

Proof

The proof follows from a martingale argument which is basically the same as in (7). Let $u, v \in T_\infty$ and set $a_1 = 1$,

$$a_{n+1} = \frac{a_n((n+1)\alpha - 2)}{n\alpha - 1}$$

so that $a_n \sim cn^{1-1/\alpha}$. Then $M_n(u, v) = d_{n-1}(u, v)/a_n$ is a martingale:

$$
\begin{aligned}
\mathbb{E}(M_{n+1}(u,v)|M_n(u,v)) &= \frac{(\alpha-1)d_{n-1}(u,v)}{n\alpha-1}\left(\frac{d_{n-1}(u,v)+1}{a_{n+1}}\right) \\
&\quad + \frac{n\alpha-1-(\alpha-1)d_{n-1}(u,v)}{n\alpha-1}\left(\frac{d_{n-1}(u,v)}{a_{n+1}}\right) \\
&= \frac{(\alpha-1)a_n M_n(u,v)}{n\alpha-1}\left(\frac{a_n M_n(u,v)+1}{a_{n+1}}\right) \\
&\quad + \frac{n\alpha-1-(\alpha-1)a_n M_n(u,v)}{n\alpha-1}\left(\frac{a_n M_n(u,v)}{a_{n+1}}\right) \\
&= \frac{a_n M_n(u,v)}{a_{n+1}}\left(\frac{(n+1)\alpha-2}{n\alpha-1}\right) = M_n(u,v)
\end{aligned}
$$

Furthermore, $M_n(u,v)$ is bounded in L^2:

$$
\begin{aligned}
\mathbb{E} \ & ((M_{n+1}(u,v)-M_n(u,v))^2|M_n(u,v)) \\
=\ & \frac{(\alpha-1)d_{n-1}(u,v)}{n\alpha-1}\left(\frac{d_{n-1}(u,v)+1}{a_{n+1}}-\frac{d_{n-1}(u,v)}{a_n}\right)^2 \\
& + \frac{n\alpha-1-(\alpha-1)d_{n-1}(u,v)}{n\alpha-1}\left(\frac{d_{n-1}(u,v)}{a_{n+1}}-\frac{d_{n-1}(u,v)}{a_n}\right)^2 \\
=\ & \frac{(\alpha-1)a_n M_n(u,v)}{n\alpha-1}\left(\frac{a_n M_n(u,v)+1}{a_{n+1}}-M_n(u,v)\right)^2 \\
& + \frac{n\alpha-1-(\alpha-1)a_n M_n(u,v)}{n\alpha-1}\left(\frac{a_n M_n(u,v)}{a_{n+1}}-M_n(u,v)\right)^2 \\
=\ & \frac{(\alpha-1)a_n[M_n(u,v)+(a_n-a_{n+1})M_n^2(u,v)]}{(n\alpha-1)a_{n+1}^2} \\
\leq\ & C\frac{M_n(u,v)}{n^{2-1/\alpha}}
\end{aligned}
$$

Thus $M_n(u,v)$ converges almost surely to some limit $d(u,v)$ Since we have a countable number of vertices, almost surely, for every pair (u,v), $d(u,v)$ exists. By passage to the limit, it is clear that for all $u,v,w \in V$, $d(u,u)=0$, $d(u,v)=d(v,u)$ and $d(u,v)+d(v,w) \leq d(u,w)$. Similarly, the four-point condition is satisfied.

Let us show that if $u \neq v$, then almost surely, $d(u,v) > 0$. Let p_n be the probability, for two vertices u,v such that $d_n(u,v)=1$, that $d(u,v)=0$. Clearly, p_n increases in n and thus tends to some p_∞. We want to show that $p_\infty = 0$.

It is easily seen that almost surely, there exists a minimal $n' > n$ such that $d_{n'}(u,v)=2$. Let w be the vertex between u and v in $T(n')$. Then on the event that $d(u,v)=0$, one must have $d(u,w)=d(w,v)=0$, and thus $p_n \leq \sup_{n'>n} p_{n'}^2$. Taking limits, $p_\infty \leq p_\infty^2$ and therefore, p_∞ is either 0 or 1. Since M_n is bounded in L^2, $\mathbb{E}(M_\infty)=\mathbb{E}(M_n)>0$, which implies that $p_\infty \neq 1$. Hence $p_\infty = 0$.

Notice that the L^2 estimate entails that for every n,

$$\mathbb{E}[(M_n(u,v) - d(u,v))^2] \leq \frac{C'|M_n(u,v)|}{n^{1-1/\alpha}}$$

and the maximal inequality implies

$$\mathbb{P}[\sup_{N \geq n} |M_n(u,v) - M_N(u,v)| \leq x] \leq \frac{C'|M_n(u,v)|}{x^2 n^{1-1/\alpha}} \tag{5}$$

□

3.2 Probability measure on the continuous tree

If v is an internal node in T_∞, define $T(v, A_n)$ as the subtree of T_∞ cut at v and containing A_n. Similarly, for $k \geq 1$, let $T_k(v, A_n)$ be the subtree of T_k cut at v and containing A_n. Then we have:

Proposition 1 *Let $m, n \geq 1$. Then conditionally on $m \in \widetilde{I}$, the limit*

$$\lim_{N \to \infty} \frac{|T_N(V_m, A_n)|}{N}$$

exists almost surely, where $|T|$ denotes the number of leaves of T.

We denote this limit $\widetilde{m}(V_m, A_n)$.

Proof Let $k = \max(m, n)$. For every $l \geq k$, let w_l be the weight, given by the random growth algorithm, of the subtree $T_l(V_m, A_n)$. From the definition the random growth algorithm, the sequence w_l follows the scheme of a generalized Polya urn. That is, $w_{l+1} = w_l + \alpha$ with probability $w_l/[(l+1)\alpha+1]$ and $w_{l+1} = w_l$ with probability $1 - w_l/[(l+1)\alpha+1]$. This entails the almost sure convergence of w_l/l as $l \to \infty$ and the convergence stated in the proposition follows.

□

4 Fragmentation

Let us briefly describe the fragmentation process we want to study. When we supress from a tree an internal node, we disconnect the tree and obtain different subtrees. In the case of T_∞, we have seen that these subtrees have a mass and we would like to know the distribution of the respective masses when we have removed all internal nodes "sufficiently closed to" A_0. For a more precise statement, see Section 4.4. The question of fragmentation of stable trees was already solved in (8). We want to show that our recursive construction provides an alternative approach.

4.1 Partitions and the Chinese restaurant

A partition of mass is a family of nonnegative numbers $(a_1 \geq a_2 \geq \ldots)$ such that $\sum a_n \leq 1$. Of course, if $(a_1, a_2, \ldots)$ is a family of nonnegative numbers such that $\sum a_n \leq 1$, we can turn it into a partition of mass by reordering the terms in decreasing order.

A way to construct random partitions of mass is Pitman's Chinese restaurant process (9). We present here a slight modification thereof. We call a Chinese restaurant with parameters (β, θ, n) the following process. At time 0, there is a first table with n customers and a second table with 1 customer. Then sequentially, other customers arrive. At time $N \geq 1$, the new customer chooses either to sit at a table with already k customers, with probability $(k - \beta)/(N + n + \theta)$, or to sit at a new table, with probability $(t(N)\beta + \theta)/(N + n + \theta)$, $t(N)$ being the number of occupied tables at time N.

For $k \geq 1$ let $C_k(N)$ denote the number of customers seated at the k-th table at time N. Then it is clear that $C_k(N)$ follows a generalized urn scheme and therefore, the family $(C_k(N)/N, k \geq 1)$ converges almost surely, as $N \to \infty$, to a partition of mass (this is the same argument as, for instance, (9), Chapter 2.2). We denote the distribution of this partition of mass $R(\beta, \theta, n)$.

4.2   Indices of internal nodes and fragmentation

For every $n \in \widetilde{I}$, we say that the *index* of V_n is k if the number of leaves of $T_n - T_n(V_n, A_0)$ is $k + 1$. By convention, say that A_0 has index 1. Also, if $n \in \widetilde{I}$, we define $J(n)$ as a set of integers such that for every i, there is exactly one $j \in J(n)$ such that A_j, A_i are in the same subtree of T_∞ cut at V_n. Then we have

Proposition 2 *For every n, conditionally on $n \in \widetilde{I}$, putting $M = 1 - \widetilde{m}(V_n, A_0)$, the partition of mass*

$$Part(V_n) := (\widetilde{m}(V_n, A_i)/M, i \in J(n))$$

has the same law as the partition of mass $R(1/\alpha, -1, k)$

Proof   This follows readily from the construction by the random growth algorithm. Indeed, the leaves $A_k \notin T(V_n, A_0)$ can be identified as customers and the subtrees cut at V_n can be identified as tables and this way, the algorithm exactly reproduces the Chinese restaurant scheme described above.

□

We also have:

Proposition 3 *Let $n \neq m$. Then conditionally on $n, m \in \widetilde{I}$ and on the indices of V_n, V_m, the partitions of masses $Part(V_n)$, $Part(V_m)$ are independent*

Proof   Let $G(n) = \{i, A_i \notin T(V_n, A_0)\}$. Define the partition $part(V_n)$ as follows. Say that $i, j \in G(n)$ are in the same block of $part(V_n)$ if A_i, A_j are in the same subtree of T_∞ cut at V_n.

Then the partitions $part(V_n), part(V_m)$ are determined by the trees $T(V_n, A_i)$, $i \neq 0$ and $T(V_m, A_i)$, $i \neq 0$. First, if $V_n \notin [A_0, V_m]$ and $V_m \notin [A_0, V_n]$, then conditionally on $G(n), G(m)$ and on the indices of V_n, V_m, these trees are independent, as noticed after the proof of Theorem 1.

Otherwise, suppose that $V_n \in [A_0, V_m]$. Then the partitions $part(V_n)$ and $part(V_m)$ are determind by the trees $T(V_m, A_i), i \in J(m)$. Now for every N, conditionally on the number of their leaves, on the indices of V_n, V_m and on $G(m) \cap \{1, 2, \ldots N\}$, these trees are independent, again according to the remark after the proof of Theorem 1. This proves the proposition.

□

Proposition 4 *Conditionally on the indices of $V_n, n \in \widetilde{I}$, the partitions of masses $(Part(V_n), n \in \widetilde{I})$ are independent.*

Proof The proof is the same as for Proposition 3.

$\square$

So we know the law of the fragmentation at a node conditionally on the index of this node and we have shown the independence of the fragmentations at different nodes, conditionally on their indices. It remains to study the distribution of the indices of the nodes.

4.3   Another length parametrization

The distribution of the indices of internal nodes can be described as follows. First consider the segment $[A_0, A_1]$. Along this segment, there is a natural order: say that if $V_n, V_m \in [A_0, A_1]$, V_m comes before V_m if $V_m \in [A_0, V_n]$.

Let $U(k)$ be the set of nodes of index $\leq k$ in the segment $[A_0, A_1]$. Since there is an order on $[A_0, A_1]$, we can define consecutive vertices of the set $U(k)$. When k increases, the sets $U(k)$ are nested sets. We then have:

Proposition 5 *The number of vertices in $U(k+1)$ between two consecutive vertices in $U(k)$ is a geometric random variable with mean $(\alpha - 1)/k\alpha$. Moreover these random variables are independent.*

Proof Suppose that V_m, V_n are consecutive vertices in $U(k)$, with V_m coming before V_n. First, remark that necessarily, by definition of the index of a node, $n \geq m$ and that there is no vertex in the segment $[V_m, V_n]$ of the tree T_n.

Let N be the first time $\geq n$ at which the random growth algorithm inserts a leaf in the subtree $T(V_m, A_n)$. If the insertion is not along the edge (V_m, V_n), which happens with probability

$$1 - \frac{\alpha - 1}{(k+1)\alpha - 1}$$

then there can be no vertex of index $k+1$ in the segment $[V_m, V_n]$ of T_∞. If the algorithm inserts a vertex along the edge (V_m, V_n), which happens with probability

$$\frac{\alpha - 1}{(k+1)\alpha - 1}$$

then V_N has index $k + 1$. Moreover, in that case there can be no vertex of index $k + 1$ in the segment $[V_m, V_N]$ of T_∞ but there can be a vertex of index $k + 1$ in the segment $[V_N, V_n]$ of T_∞, and this occurs with the same probability. Therefore, the number $N_{k+1}(V_m, V_n)$ of vertices of index $k+1$ in the segment $[V_m, V_n]$ is geometric with mean $(\alpha - 1)/k\alpha$.

The proof of independence is tedious. To simplify, we shall only examine the case $k = 1$. Suppose that V_m, V_n are, in that order, consecutive vertices in $U(1)$. Let N be the first time at which the algorithm inserts a vertex along the segment $[V_n, A_1]$. Then $N_2(V_m, V_n)$ is entirely determined at time M, while $N_2(V_n, V_N)$ only depends on what occurs after time N. Therefore, $N_2(V_m, V_n)$ and $N_2(V_n, V_N)$ are independent and V_m, V_n, V_N are consecutive vertices in $U(1)$. The same argument can be used for i consecutive vertices in $U(1)$. $\square$

With the same arguments, one can show that if V_m, V_n are consecutive vertices $\in U(k)$, if $l \geq k$ and if $N_l(V_m, V_n)$ is the number of vertices in the intersection $U(l) \cap [V_m, V_n]$, then $N_l(V_m, V_n)$ is a geometric random variable with mean

$$\left[\prod_{i=k}^{l-1} \left(1 + \frac{\alpha - 1}{i\alpha} \right) \right] - 1$$

One can also show that

$$\frac{N_l(V_m, V_n)}{\prod_{i \leq l} \left(1 + \frac{\alpha-1}{i\alpha} \right)}$$

converges, as $l \to \infty$, to an exponential random variable $E(V_m, V_n)$. Moreover, again by the same arguments, the random variables $E(V_m, V_n)$ and $E(V_l, V_p)$ are independent if $[V_m, V_n] \cap [V_l, V_p] = \emptyset$. Of course, the same results can be proved for vertices $\notin [A_0, A_1]$. This leads to

Theorem 3 *(i) Let v be an internal node in T_∞. Let $N_n(v)$ be the number of internal nodes of index $\leq n$ in the segment $[A_0, v]$. Then*

$$\frac{N_n(v)}{\prod_{i \leq n} \left(1 + \frac{\alpha-1}{i\alpha} \right)}$$

converges almost surely, as $n \to \infty$, to a positive real number $L(v)$.
   (ii) For all $n, k \geq 1$, let

$$\mathcal{N}(n) = \{t \geq 0, L(V_i) = t \text{ for some } i \text{ such that } V_i \text{ has index } \leq n \text{ and } V_i \in [A_0, A_k]\}$$

Then $\mathcal{N}(n)$ is a Poisson point process on $\mathbb{R}_+$ with intensity

$$\prod_{i=1}^{n-1} \left(1 + \frac{\alpha - 1}{i\alpha} \right) dx$$

4.4   Fragmentation of the tree

Let $t \geq 0$ and define the partition of integers F_t' as follows. Say that $i, j \geq 1$ are in the same block of F_t' if, for every $V_m \in [A_i, A_j]$, $L(V_m) > t$. Since these blocks are unions of subtrees, which have asymptotic frequencies, it is easily seen that the blocks of F_t also have asymptotic frequencies and induce a partition of mass F_t. When t increases, these partitions "split" into finer partitions. More precisely, if, for some $L(V_n) = t$, then the block containing A_n at time t_- splits at time t into blocks given by the subtrees cut at V_n. One can describe this phenomeneon using the general framework of fragmentation introduced by Bertoin (4). Then we can state:

Theorem 4 *The process $(F_t, t \geq 0)$ is a homogeneous fragmentation process with dislocation measure*

$$\sum_{n \geq 1} p_n(\alpha) R(1/\alpha, -1, n)$$

where $p_1(\alpha) = 1$ and for $n \geq 2$,

$$p_n(\alpha) = \prod_{i=1}^{n-1} \left(1 + \frac{\alpha - 1}{i\alpha} \right) - \prod_{i=1}^{n-2} \left(1 + \frac{\alpha - 1}{i\alpha} \right)$$

As already mentioned, this result appears in (8), where the dislocation measure is expressed in a different way. As observed therein, loosely speaking, this dislocation measure corresponds to a Chinese restaurant with parameters $(1/\alpha, -1)$ except that, strictly speaking, this has no meaning since $1/\alpha < 1$. The decomposition of the dislocation measure as a "mixture" of dislocation measures of the form $R(\alpha, -1, n)$ gives a precise sense to this informal idea.

Theorem 4 can be proved using the description of the dislocation depending on the index in Proposition 2, the independence of these dislocations stated in Proposition 3 and the distribution of the indices of internal nodes given by Theorem 3.

References

[1] Aldous, David. The continuum random tree. III. *Ann. Probab.* **21** (1993), no. 1, 248–289.

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

[3] Bertoin, Jean. A fragmentation process connected to Brownian motion. *Probab. Theory Related Fields* **117** (2000), no. 2, 289–301.

[4] Bertoin, Jean. *Random fragmentation and coagulation processes.* Cambridge Studies in Advanced Mathematics, **102**. Cambridge University Press, Cambridge, 2006.

[5] Duquesne, Thomas; Le Gall, Jean-François. Random trees, Lévy processes and spatial branching processes. *Astérisque* No. 281 (2002).

[6] Evans, Steven N. *Probability and real trees.* Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. Lecture Notes in Mathematics, 1920. Springer, Berlin, 2008.

[7] Marchal, Philippe. Constructing a sequence of random walks strongly converging to Brownian motion. Discrete random walks (Paris, 2003), 181–190 *Discrete Math. Theor. Comput. Sci.* 2003.

[8] Miermont, Grégory. Self-similar fragmentations derived from the stable tree. I. Splitting at heights. *Probab. Theory Related Fields***127** (2003), no. 3, 423–454.

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

[10] Rémy, Jean-Luc. Un procédé itératif de dénombrement d'arbres binaires et son application à leur génération aléatoire. *RAIRO Inform. ThÃ©or.* **19** (1985), no. 2, 179–195.