An owner's manual for scaling limits of random trees and graphs

Caelan Atamanchuk

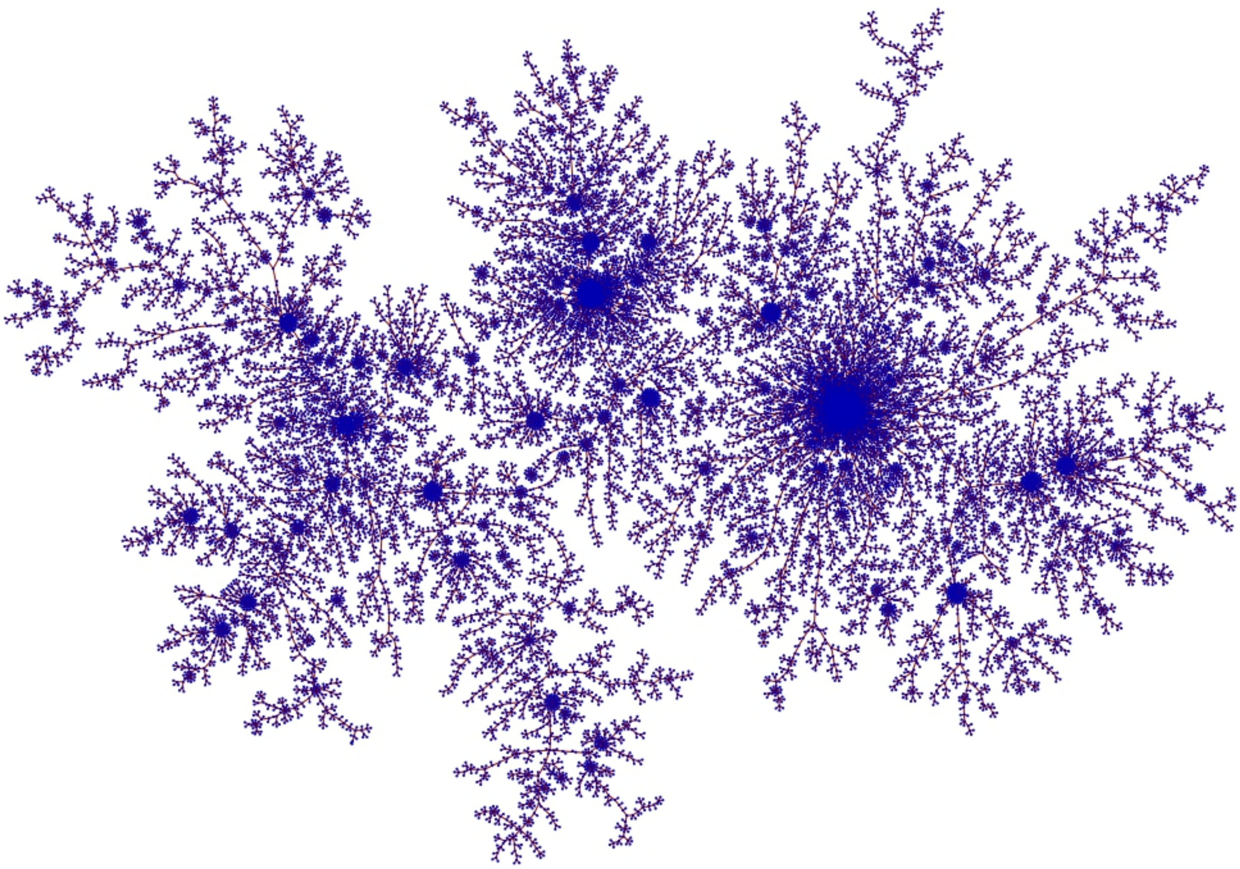

Figure 1: An image of a cool tree stolen from Igor Kortchemski's website (I'm hoping I'll find the time to make my own cool tree pictures soon).

CONTENTS

1 Random combinatorial trees

This section introduces our main object of consideration, which is random trees. We discuss two ways to encode trees with discrete functions and examine the relationships between these encodings. We then turn our attention to random trees, where the specific trees of interest are Bienaymé trees.

1.1 Encoding trees with discrete functions

Most trees we consider in these notes are *plane trees*, which are finite rooted trees with an ordering on each collection of siblings in the tree. We shall identify all plane trees as subsets of the infinite Ulam-Harris tree, which we define now. Let

$$\mathbf{U} = \bigcup_{k=0}^{\infty} \mathbb{N}^k,$$

where we take $\mathbb{N} = \{1, 2, 3, ..., \}$ and $\mathbb{N}^0 = \{\emptyset\}$. We call the elements of $\mathbf{U}$ the *vertices*. The length of the vector $u \in \mathbf{U}$, $|u|$, is called the *generation* of u. It is also called the *height* of u. If $u = (u_1, ..., u_k), v = (v_1, ..., v_m) \in \mathbf{U}$ we let $uv = u \cdot v$ denote the concatenation of the two sequences, $(u_1, ..., u_k, v_1, ..., v_m)$. The vertex $p(v) = (u_1, ..., u_{k-1})$ is called the *parent* of u and u is called the child of $p(u)$. If $w = (w_1, ..., w_k) \in \mathbf{U}$ is such that $w_i = u_i$ for all $1 \leq i \leq k-1$ and $w_k \neq u_k$, then u and w are called *siblings*. The set $\mathbf{U}$ is called the *Ulam-Harris tree* (Figure 2 highlights the tree structure), and we use it to formally define the notion of a plane tree.

Definition 1.1. *A finite subset* $\mathbf{t} \subseteq \mathbf{U}$ *is called a* plane tree *if:*

(i) $\emptyset \in \mathbf{t}$.

(ii) *If* $u \in \mathbf{t}$, *then* $p(u) \in \mathbf{t}$.

(iii) *There is a collection of non-negative integers* $(c_\mathbf{t}(u) : u \in \mathbf{t})$ *such that, for all* $j \in \mathbb{N}$ *and* $u \in \mathbf{t}$, $uj \in \mathbf{t}$ *if and only if* $1 \leq j \leq c_\mathbf{t}(u)$.

We interpret $c_\mathbf{t}(u)$ as the number of children that u has in $\mathbf{t}$. We also occasionally refer to this as the *out-degree* of u. The set of all plane trees is denoted by $\mathcal{R}$ in what follows. The set of all plane trees $\mathbf{t}$ such that $|\mathbf{t}| = n$ is denoted by $\mathcal{R}_n$. The ordering on our plane trees is the natural lexicographical ordering of the Ulam-Harris tree. We shall occasionally need to discuss the genealogical partial ordering of our trees as well, which we shall denote with $\preceq$. We write $u \preceq v$ for two vertices $u, v \in \mathbf{t}$ if v is a descendent of u, i.e., $v = uw$ for some $w \in \mathbf{U}$. The lexicographical ordering of $\mathbf{U}$ is denoted with $\leq$.

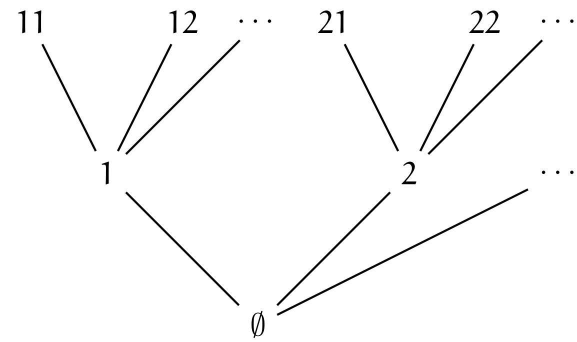

Figure 2: A depiction of the set **U** that highlights its tree structure.

The embedding of our plane trees inside the Ulam-Harris tree, and the corresponding ordering, allow for easy exploration of the tree via depth-first exploration. We first define the depth-first queue process, and then note why it is useful for characterizing plane trees.

Definition 1.2. *Let* $\mathbf{t} \in \mathcal{R}_n$ *and let* $u_1, ..., u_n$ *be the vertices written in lexicographical order. Write* $(c_1, ..., c_n) = (c_{\mathbf{t}}(u_1), ..., c_{\mathbf{t}}(u_n))$. *The sequence of integers* $(q_k)_{k=0}^n$ *with*

$$q_k = \sum_{i=1}^{k} (c_i - 1)$$

is called the depth-first queue process *of the tree* $\mathbf{t}$ *(DFQ). Any sequence* $(x_k)_{k=0}^n$ *such that*

(i) $x_0 = 0$, $x_n = -1$,

(ii) $x_k \geq 0$ *for all* $0 \leq k \leq n-1$

(iii) $x_k - x_{k-1} \geq -1$ *for all* $1 \leq k \leq n$

is called a Łukasiewicz path of length n. *We take* $\mathcal{L}$ *to denote the collection of all Łukasiewicz paths and* $\mathcal{L}_n$ *the paths of length* n. *In some places, the DFQ process of a tree is called the Łukasiewicz path of the tree.*

As the name suggests, there is an interpretation of the DFQ process of a tree $\mathbf{t} \in \mathcal{R}_n$ as the evolving size of a queue while exploring the tree. Begin with a queue $Q_0 = (\emptyset)$. Then, for $0 \leq i \leq n-1$, suppose that $Q_i = (w_1, ..., w_{q_{i+1}})$ with $q_i = |Q_i| - 1$. We pop w_1 from Q_i, query the number of children it has, and then add those children to the front of Q_i in their lexicographical order to form Q_{i+1}. The net change in the size of the queue at each step is exactly $c_i - 1$, as at each step the vertex being popped is the ith in the ordering of $\mathbf{t}$. Note that step k of the DFQ process is when we explore the vertex u_k (the kth vertex in the lexicographical order) and its children are not represented in the queue until the next step if it has any. Starting the walk at zero and not one is just a notational choice to make future convergence results a little cleaner. It removes a lot of "+1's'.'

Lemma 1.3. *The mapping* $\varphi : \mathcal{R} \to \mathcal{L}$ *given by*

$$\varphi(\mathbf{t}) = (q_0, ..., q_{|\mathbf{t}|}) \quad \forall \mathbf{t} \in \mathcal{R},$$

where $(q_0, ..., q_{|\mathbf{t}|})$ *is the DFQ process for* $\mathbf{t}$*, is a bijection.*

Proof. First, we verify that φ maps into $\mathcal{L}$, which amounts to showing (i) and (ii) in the definition as the other point is clear. The first point follows from the fact that trees on n vertices have $n-1$ edges (and hence $n-1$ children in the context of plane trees). For the second point, we note that $c_{\mathbf{t}}(u_1) + ... + c_{\mathbf{t}}(u_k) \geq k$ for $1 \leq k \leq n-1$ because $u_1, ..., u_{k+1}$ are all children of some vertex in $\{u_1, ..., u_k\}$.

Recall that two plane trees $\mathbf{t}, \mathbf{s}$ are equal if and only if they are the same subset of $\mathbf{U}$. We begin by showing that φ is injective. If $|\mathbf{t}| \neq |\mathbf{s}|$, then they do not have the same DFQ process so suppose that $|\mathbf{t}| = |\mathbf{s}| = n$ and $\mathbf{t} \neq \mathbf{s}$. Let $u^* \in \mathbf{t} \cap \mathbf{s}$ be the first vertex in the ordering that has a child in one tree and not the other. Without loss of generality, we may assume that this child is in $\mathbf{t}$, so $c_{\mathbf{t}}(u^*) > c_{\mathbf{s}}(u^*)$. If $(q_0(\mathbf{t}), ..., q_n(\mathbf{t}))$ and $(q_0(\mathbf{s}), ..., q_n(\mathbf{s}))$ are the DFQ processes of $\mathbf{t}$ and $\mathbf{s}$ respectively, the fact that u^* was chosen to be minimal implies that $q_k(\mathbf{t}) = q_k(\mathbf{s})$ for all $1 \leq k \leq i^* - 1$, where i^* is the place of u^* in the ordering. Then,

$$q_{i^*}(\mathbf{t}) = q_{i^*-1}(\mathbf{t}) + c_{\mathbf{t}}(u^*) > q_{i^*-1}(\mathbf{s}) + c_{\mathbf{s}}(u^*) = q_{i^*}(\mathbf{s}).$$

Surjectivity follows almost immediately from the fact that $q_k - q_{k-1} = c_{\mathbf{t}}(u_k) - 1$ for all $1 \leq k \leq n$. Given a Łukasiewicz path $\mathbf{q} = (q_0, ..., q_n)$ we can construct a tree that straightforwardly maps to $\mathbf{q}$. Begin with $\mathbf{t}_0 = \{\emptyset\}$. Then, inductively define $\mathbf{t}_{i+1}$ for each $0 \leq i \leq n-1$ by setting $\mathbf{t}_{i+1} = \mathbf{t}_i \cup \{x_i \cdot 1, ..., x_i \cdot (q_{i+1} - q_i + 1)\}$, where x_i is the ith element of $\mathbf{t}_i$ in lexicographical order (note that such an element exists by the assumption $q_k \geq 0$ for $0 \leq k \leq n-1$). One can check that $\varphi(\mathbf{t}_n) = (q_0, ..., q_n)$. $\square$

Another discrete function that encodes plane trees is the height function. It can be seen as a walk through the tree in lexicographical order that records the height of the current vertex.

Definition 1.4. *Let* $\mathbf{t} \in \mathcal{R}_n$ *and let* $u_0, ...u_{n-1}$ *be its vertices written in lexicographical order. The* height function *of* $\mathbf{t}$*, denoted by* $(h_{\mathbf{t}}(k))_{k=0}^{n-1}$*, is given by* $h_{\mathbf{t}}(k) = |u_{k+1}|$.

Before we get into why the height function acutally matters, let's first introduce a continuous function that is related to the height function and of great importance later on. We call this function the *contour function* of the tree. The formal definition is a little confusing, I recommend looking at the example below to make sense out of it. We informally can see the contour function as arising from a process where we trace out the tree using a pencil that never leaves the paper and draws at a single unit speed.

Definition 1.5. *Let $\mathbf{t} \in \mathcal{R}_n$ and let $u_0, ..., u_{n-1}$ be the vertices in lexicographical order. Set $u_n = \emptyset$. Let $p_0^i, p_1^i, p_2^i...$ be the interior vertices on the unique paths from u_i to u_{i+1} for each $0 \leq i \leq n-1$ in the order they would be taken if travelling from u_i to u_{i+1} in $\mathbf{t}$. We define a new sequence of vertices $v_0, ..., v_{2(n-1)}$ by inserting the p^i's between u_i and u_{i+1} for all $0 \leq i \leq n-1$ (each vertex $u \in \mathbf{t}$ appears $c_{\mathbf{t}}(u) + 1$ times in the new sequence). We define the contour function of $\mathbf{t}$, $\gamma_{\mathbf{t}} : [0, \infty) \to [0, \infty)$ by*

$$\gamma(t) = |v_{\lfloor t \rfloor}| + (t - \lfloor t \rfloor)(|v_{\lceil t \rceil}| - |v_{\lfloor t \rfloor}|)$$

for $0 \leq t \leq 2(n-1)$, and $\gamma(t) = 0$ for $t > 2(n-1)$.

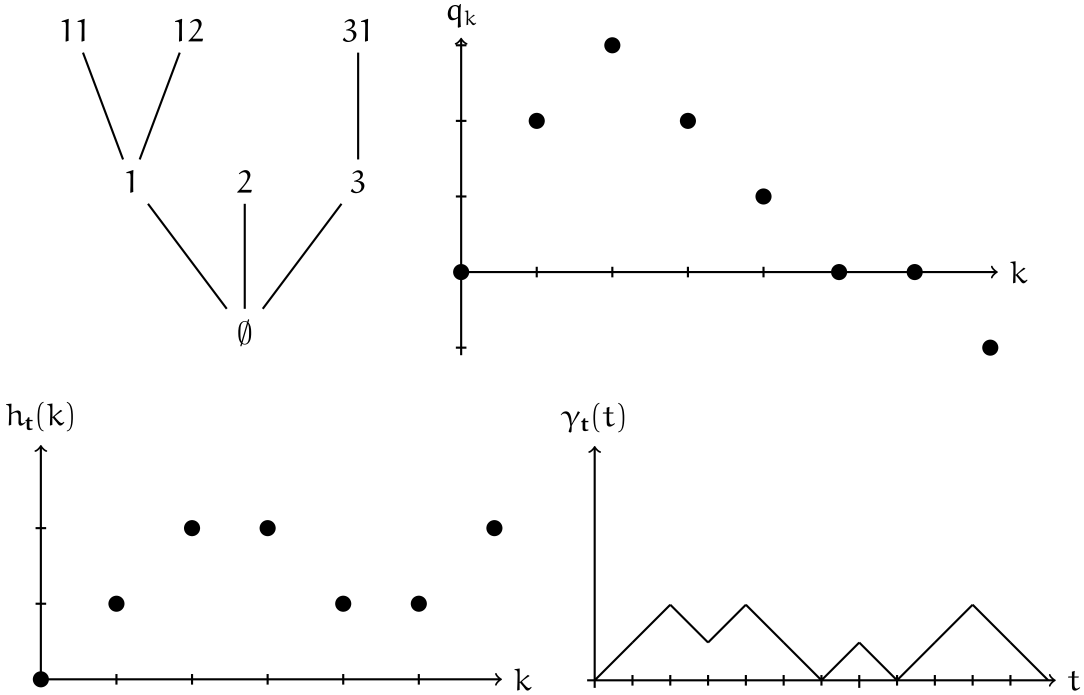

Figure 3: A tree and its many functional encodings

There is a simple way to convert between the height function and the DFQ process of a tree. This relationship will allow us to describe the height function in terms of sums of i.i.d. random variables when discussing Bienaymé trees later.

Theorem 1.6. *Let $\mathbf{t} \in \mathbb{R}_n$ have DFQ process $(q_0, ..., q_n)$. Then, for all $0 \leq k \leq n-1$,*

$$h_{\mathbf{t}}(k) = \left| \left\{ 1 \leq j \leq k-1 : q_j = \inf_{j \leq m \leq k} q_m \right\} \right|.$$

Proof sketch. It is clear that $h_t(k) = |\{0 \leq j \leq k-1 : u_j \preceq u_k\}|$, so we only need to show that

$$u_j \preceq u_k \iff q_j = \inf_{j \leq m \leq k} q_m.$$

It can be observed immediately from the definition that, if $\mathbf{t}(u_j)$ is the subtree of $\mathbf{t}$ rooted at u_j, then $u_j \preceq u_k$ if and only if $u_k \in \mathbf{t}(u_j)$, so we can instead show

$$u_k \in \mathbf{t}(u_j) \iff q_j = \inf_{j \leq m \leq k} q_m. \tag{1}$$

Let $\tau_j = \inf\{m \geq j : q_m < q_j\}$. At step j of the DFQ process we add u_j's children to the queue and remove u_j. The process only leaves the subtree $\mathbf{t}(u_j)$ all of the children of u_j have been removed (along with any children they have). This is exactly τ_j. In particular, we have that $\mathbf{t}(u_j) = \{u_m : j \leq m \leq \tau_j - 1\}$. (1) follows immediately from this identity. □

An immediate corollary of Theorem 1.6 is that the height function of a tree uniquely determines it. By taking the end point of all length one intervals on which the contour function is increasing, we can recover the height process of a tree. Moreover, from the height function we can recover the tree and from the tree we can get the contour function. Hence, the contour function uniquely determines the tree as well. Of course, one can prove this fact directly via the "pencil and paper" analogy. One can also prove the height function encodes its tree directly by observing that, if one knows the u_k and $h_{\mathbf{t}}(k+1)$, then there is only one possible vertex that could be u_{k+1} (it is a child of the ancestor of u_k that is at height $h_{\mathbf{t}}(k+1) - 1$).

1.2 Bienaymé trees

Definition 1.7. *Let μ be a measure on $\mathbb{Z}_{\geq} = \{0, 1, 2, ...\}$ with $\sum_{k=0}^{\infty} k\mu(k) < \infty$ such that $\mu(1) \neq 1$. For all $u \in \mathbf{U}$, we associate an independent random variable $\xi_u \overset{\mathcal{L}}{=} \mu$. The subset $T = \{u = (u^1, ..., u^k) \in \mathbf{U} : u^j \leq \xi_{(u^1,...,u^{j-1})}\ \forall\ 1 \leq j \leq k\}$ is called a* Bienaymé tree *with offspring distribution μ. We often write $T \overset{\mathcal{L}}{=}$ Bienaymé(μ). Collections of many i.i.d. Bienaymé trees are sometimes called Bienaymé forests. We call a Bienaymé tree critical if $\sum_{k=0}^{\infty} k\mu(k) = 1$, subcritical if $\sum_{k=0}^{\infty} k\mu(k) < 1$, and supercritical otherwise.*

These trees are ubiquitous in probability theory and combinatorics, having been studied as far back as the 1800's. Those familiar with the classic Galton-Watson martingale process may notice that these two structures are essentially the same. It is mostly straightforward to prove from the definition that Bienymé trees are plane trees except for the criteria that T must be finite. This fact is a corollary of a result known by many as the fundamental theorem of Bienaymé trees. See [ANN04] for a proof.

Theorem 1.8. *Let $T \overset{\mathcal{L}}{=}$ Bienaymé(μ) for some μ matching the above criteria. If T is sub-critical or critical, then $|T| < \infty$ almost surely. In particular, T is a plane tree. Otherwise, $\mathbf{P}(|T| = \infty) > 0$.*

The independence in the variables $(\xi_u : u \in \mathbf{U})$ has some nice consequences concerning the distribution of T over the set $\mathcal{R}$.

Lemma 1.9. *Let* $\mathbf{t} \in \mathcal{R}$ *and let* $\mathsf{T} \overset{\mathcal{L}}{=}$ *Bienaymé*(μ)*. Then,*

$$\mathbf{P}(\mathsf{T} = \mathbf{t}) = \prod_{u \in \mathbf{t}} \mu(c_{\mathbf{t}}(u)).$$

Proof. Since T is a plane tree almost surely, $\{\mathsf{T} = \mathbf{t}\} = \cap_{u \in \mathbf{t}} \{\xi_u = c_{\mathbf{t}}(u)\}$. Using the independence of the ξ's we get,

$$\mathbf{P}(\mathsf{T} = \mathbf{t}) = \mathbf{P}\left(\bigcap_{u \in \mathbf{t}} \{\xi_u = c_{\mathbf{t}}(u)\}\right) = \prod_{u \in \mathbf{t}} \mu(c_{\mathbf{t}}(u)).$$

$\square$

With the standard pleasantries out of the way, we can turn our attention to the most important property of Bienaymé trees from the perspective of scaling limits. The DFQ process of these trees is distributed like a simple random walk, and their sizes are exactly distributed like the first time that the simple random walk hits -1. At first glance, knowing the definition of the DFQ process, one might think that this statement is trivially true by the definition of Bienaymé trees. However, the presence of the stopping time in the expression below makes the claim not immediate as it could (in theory) disturb the natural independence between the number of children each vertex has.

Theorem 1.10. *Let* $\mathsf{T} \overset{\mathcal{L}}{=}$ *Bienaymé*(μ)*, and let its DFQ process be denoted by* Q*. Let* $(S_k : k \geq 0)$ *be a simple random walk with step sizes distributed like* ν*, where for all* $k \geq -1$*,* $\nu(k) = \mu(k+1)$*. Then,*

$$Q \overset{\mathcal{L}}{=} (S_0, ..., S_\tau),$$

where $\tau = \inf\{n \geq 1 : S_n = -1\}$*. In particular* $|\mathsf{T}| \overset{\mathcal{L}}{=} \tau$*.*

Proof. It suffices to just check that the vector $(c_{\mathbf{t}}(U_0), ..., c_{\mathbf{t}}(U_{|\mathsf{T}|-1}))$ is distributed like a collection of i.i.d. μ-distributed random variables, where $(U_0, ..., U_{|\mathsf{T}|-1})$ is the vertices of T written in lexicographical order. To be able to remove the random indexing, we want $\{U_k = u\}$ for $0 \leq k \leq |\mathsf{T}| - 1$ and $u \in \mathcal{U}$ to be measurable with respect to only the vertices below u in the lexicographical order.

First, the set $\mathsf{T} \cap \{v \in \mathbf{U} : v \leq u\}$, is measurable with respect to $\sigma(\xi_v : v < u)$. Then, for any $k \geq 0$, the event $\{U_k = u\} \cap \{|\mathsf{T}| > k\}$, being completely determined by $\mathsf{T} \cap \{v \in \mathbf{U} : v < u\}$, is measurable with respect to $\sigma(\xi_v : v < u)$. The set $\{U_k = u\} \cap \{|\mathsf{T}| \leq k\}$ is also measurable with respect to $\sigma(\xi_v : v < u)$ for the same reason. Combining the two facts we get that $\{U_k = u\}$ is measurable with respect to $\sigma(\xi_v : v < u)$.

Now, from here we can proceed via a standard induction. Let $g_0, ..., g_k : \mathbb{Z}_{\geq} \to \mathbb{Z}_{\geq}$ be a collection of functions for $0 \leq k \leq |\mathsf{T}| - 1$. Then,

$$\mathbf{E}\left[g_1(\xi_{U_0}) \cdots g_k(\xi_{U_k})\right]$$

$$
\begin{aligned}
&= \sum_{u_0<\ldots<u_k} \mathbf{E}\left[\mathbf{1}_{\{U_0=u_0,\ldots,U_k=u_k\}} g_1(\xi_{u_1})\cdots g_k(\xi_{u_k})\right] \\
&= \sum_{u_0<\ldots<u_k} \mathbf{E}\left[\mathbf{1}_{\{U_0=u_0,\ldots,U_k=u_k\}} g_1(\xi_{u_1})\cdots g_{k-1}(\xi_{u_{k-1}})\right] \mathbf{E}[g_k(\xi_{u_k})] \\
&= \sum_{u_0<\ldots<u_{k-1}} \mathbf{E}\left[\mathbf{1}_{\{U_0=u_0,\ldots,U_{k-1}=u_{k-1}\}} g_1(\xi_{u_1})\cdots g_{k-1}(\xi_{u_{k-1}})\right] \mathbf{E}[g_k(\xi_{u_k})] \\
&= \mathbf{E}\left[g_1(\xi_{U_0})\cdots g_k(\xi_{U_{k-1}})\right] \mathbf{E}[g_k(\xi_{u_0})],
\end{aligned}
$$

where in the first equality we used the measurability we just proved and in the second we use the independence of child distribution for fixed indices. The sum is only over vertices in generation at most k. Applying induction completes the proof of the independence, and as noted at the start completes the proof as a whole. □

1.3 Bienaymé tree conditioned to have a fixed size

Bienaymé trees are interesting structures in their standard form. However, their ability to generalize so many canonical random tree models is what has kept them an ongoing topic of discussion for so many years since their origins in the study of family trees. The way we observe this generalizing property is by sampling Bienaymé trees conditioned on their size being some parameter $n \in \mathbb{N}$. We write $T \stackrel{\mathcal{L}}{=} \text{Bienaymé}(n, \mu)$ for a random plane tree T if, for all $\mathbf{t} \in \mathcal{R}_n$,

$$\mathbf{P}(T = \mathbf{t}) = \mathbf{P}(T' = \mathbf{t} \mid |T'| = n),$$

where $T' \stackrel{\mathcal{L}}{=} \text{Bienaymé}(\mu)$. For the rest of this subsection, we are going to cover a variety of random tree models, and explain how they fit into the category of conditioned critical Bienaymé trees. First, however, we need to explain why this is something that we should be able to do.

Definition 1.11. *Let M be a multiset of plane trees. We define the weight of a tree in $\mathbf{t} \in \mathbf{U}$, $\Omega(\mathbf{t})$, to be the number of occurrences of $\mathbf{t}$ in M. Then, we call*

$$z_n = \sum_{\mathbf{t} \in M : |\mathbf{t}| = n} \Omega(\mathbf{t})$$

the partition function of M. For each $n \geq 1$, let T_n be a random tree with distribution,

$$\mathbf{P}(T_n = \mathbf{t}) = \frac{\Omega(\mathbf{t})}{z_n}.$$

For each $\mathbf{t} \in \mathcal{U}$, let $(m_k(\mathbf{t}))_{k=0}^{\infty}$ be the number of vertices with k children for $k \geq 0$. If there exists a sequence $(a_k)_{k=1}^{\infty}$ of integers such that

$$\Omega(\mathbf{t}) = \prod_{k=0}^{\infty} a_k^{m_k(\mathbf{t})},$$

then we call the random trees $(T_n)_{n=1}^{\infty}$ a simply generated family of random trees.

In many cases, simply generated trees can be described as Bienaymé trees conditioned on their size. Let $(T_n)_{n=1}^{\infty}$ be a family of simply generated tree, and let μ^x be a measure defined by $\mu^x(k) = a_k x^k / f(x)$ for all $k \geq 0$ and some $x > 0$. We define T_n^x for all $n \geq 1$ to be a Bienaymé(n, μ^x).

Lemma 1.12. *Let $f(x) = \sum_{k=0}^{\infty} a_k x^k$ and suppose that there is some $x^* > 0$ such that $1 \leq f(x^*) < \infty$. Then, there exists some $\tau > 0$ such that $f(\tau) = \tau f'(\tau)$.*

We shall skip the proof as it not particularly instructive and generating functions are not the topic of interest.

Theorem 1.13. *Let $f(x) = \sum_{k=0}^{\infty} a_k x^k$ and suppose that there is some $x^* > 0$ such that $1 \leq f(x^*) < \infty$. Let $\tau > 0$ such that $f(\tau) = \tau f'(\tau)$ (exists from the above lemma). Then, for all $x \in (0, \tau]$, $T_n \overset{\mathcal{L}}{=} T_n^x$, where both $(T_n)_{n=1}^{\infty}$ and $(T_n^x)_{n=1}^{\infty}$ are defined above. In particular, there is a critical child distribution μ such that $T_n \overset{\mathcal{L}}{=}$ Bienaymé(n, μ).*

Proof. Let $T^* \overset{\mathcal{L}}{=}$ Bienaymé(μ^t). By Lemma 1.9,

$$\begin{aligned}
\mathbf{P}(T^* = \mathbf{t}) &= \prod_{k=0}^{\infty} (\mu^x(k))^{m_k(\mathbf{t})} \\
&= \prod_{k=0}^{\infty} \left(\frac{a_k x^k}{f(x)} \right)^{m_k(\mathbf{t})} \\
&= \left(\prod_{k=0}^{\infty} a_k^{m_k(\mathbf{t})} \right) (f(x))^{-n} \left(x^{\sum_{k=0}^{\infty} k m_k(\mathbf{t})} \right) \\
&= \Omega(\mathbf{t}) (f(x))^{-n} \left(x^{\sum_{k=0}^{\infty} k m_k(\mathbf{t})} \right).
\end{aligned}$$

Then,

$$\mathbf{P}(|T^*| = n) = \sum_{\mathbf{t}:|\mathbf{t}|=n} \Omega(\mathbf{t}) (f(x))^{-n} \left(x^{\sum_{k=0}^{\infty} k m_k(\mathbf{t})} \right) = z_n (f(x))^{-n} \left(x^{\sum_{k=0}^{\infty} k m_k(\mathbf{t})} \right).$$

Hence,

$$\mathbf{P}(T_n^x = \mathbf{t}) = \frac{\Omega(\mathbf{t})}{z_n}.$$

The second statement follows the above lemma and the fact that the mean of the child distribution μ^x is

$$\sum_{k=0}^{\infty} \frac{k a_k x^k}{f(x)} = \frac{x f'(x)}{f(x)}.$$

□

What is the takeaway of this theorem? Our claim at the beginning of this section was that we could view many canonical random tree models as Bienaymé trees conditioned on their size. This theorem just asserts that we only need to be able to view them as simply generated trees, which is a much nicer family for this purpose. It is fairly easy to find a weight function that results in the correct distribution for many families of random trees. Let us finish things off by giving some examples. Verifying the claims is not too hard and I don't even know if I'll cover this material, so I'm just going to write the coefficients that give the desired tree for each example.

(i) If we set $a_0 = 1,\ a_1 = 2,\ a_2 = 1$, then T_n is a uniform rooted binary tree on n vertices.

(ii) If we set $(a_0 = 1,\ a_2 = 1)$, then T_n is a uniform full binary tree on n vertices.

(iii) If we set $(a_0 = 1,\ a_k = 1)$, then T_n is a uniform rooted k-ary tree on n vertices.

(iv) If we set $(a_k = 1 \text{ for all } k \geq 0)$, then T_n is a uniform rooted plane tree on n vertices.

There is one last case that needs to be separated out on its own as we can deal directly with the Bieanymé tree instead of the simply generated tree. The tree of interest is the uniform random labelled tree on n vertices. Let $T \overset{\mathcal{L}}{=} \text{Bienaymé}(\text{Poi}(1))$. Erase the planar ordering and root, and then give T a uniformly chosen labelling from $\{1, ..., |T|\}$. Then, for a labelled rooted tree $\mathbf{t}$,

$$\mathbf{P}(T = \mathbf{t}) = \frac{e^{-|\mathbf{t}|}}{|\mathbf{t}|!},$$

implying that $\mathbf{P}(T = \mathbf{t} \mid |T| = n)$ is a uniform labelled tree on n vertices (the identity is not trivial, but can be verified without too much sweat by permuting vertices with the same degree).

2 Real trees and the Brownian CRT

We introduce a second notion of a tree in this section, specifically that of a real tree. These are connected metric spaces that share a lot of global metric information with combinatorial trees, but erase some of the meaning of things like vertices and adjacency. We discuss how the space of all real trees can be made into a complete separable metric space, setting ourselves up the groundwork for how one can make sense out of scaling limits for trees. We also cover the encoding of real trees via continuous compactly supported functions. This crucially sets up a bridge between the combinatorial and the continuum via the contour function.

2.1 The space of rooted real trees

As was done with combinatorial trees, we shall begin our exploration of real trees by setting them up as formal structures. Naturally, the starting place is the definition.

Definition 2.1. *A compact metric space* $(\mathbf{T}, d)$ *is called a real tree if, for all* $x, y \in \mathbf{T}$*:*

(i) *there is a unique is isometric embedding* $f_{xy} : [0, d(x, y)] \to \mathbf{T}$ *such that* $f_{xy}(0) = x$ *and* $f_{xy}(d(x, y)) = y$*;*

(ii) *if* $g : [0, 1] \to \mathbf{T}$ *is a continuous injective map with* $g(0) = x$ *and* $g(1) = y$*, then* $g([0, 1]) = f([0, d(x, y)])$.

Despite no longer feeling like vertices in the sense of a combinatorial tree, we shall still call elements of $\mathbf{T}$ its *vertices*. The real trees we discuss in these notes shall be rooted, meaning that each $\mathbf{T}$ has some distinguished vertex $\rho \in \mathbf{T}$. Its role shall mostly be as a constraint for the equivalence of two trees, though its existence also allows to discuss things like height. Real trees are not considered planar, but some results we prove later about how much branching can occur in a real tree imply that we could define an ordering analogous to the sibling ordering that defines plane trees. We need some more notation to go along with our new definition.

(i) The range of the isometric embedding f_{xy} for any $x, y \in \mathbf{T}$ shall be denoted by $[x, y]$. The sets $(x, y]$, $[x, y)$, (x, y), $[x, x]$, $(x, x]$, $[x, x)$, (x, x) are all defined analogously.

(ii) The distance $d(\rho, x)$ for $x \in \mathbf{T}$ is called the *height* of x. The segment $[\rho, x]$ is called the ancestral line of x.

(iii) We define the *genealogical partial ordering* on $\mathbf{T}$, written as $\preceq$, by $x \preceq y$ if $x \in [\rho, y]$.

(iv) The *degree* of a vertex $x \in \mathbf{T}$ is the cardinality of the set of components in the metric space $(\mathbf{T} \setminus \{x\}, d)$. We say that y and z are in the same component of $\mathbf{T}\setminus\{x\}$ if they are connected in $\mathbf{T}\setminus\{x\}$ in the topological sense. Vertices of degree one are called *leaves*.

(v) For $x, y \in \mathbf{T}$, we call the unique $z \in \mathbf{T}$ such that $[\rho, x] \cap [\rho, y] = [\rho, z]$ the *least common ancestor* of x and y. We denote this vertex by $x \wedge y$.

(vi) We call two real trees $\mathbf{T}_1$ and $\mathbf{T}_2$ *equivalent* if there is a root preserving isometry $f : \mathbf{T}_1 \to \mathbf{T}_2$. The set $\mathbb{T}$ will denote the space of all equivalence classes of real trees. We often conflate a tree with its equivalence class.

Item (v) above contained the claim that there exists such an element. Since it gives us a chance to get acquainted with the definition of a real tree, let's prove this claim.

Theorem 2.2. *For every pair $x, y \in \mathbf{T}$, there exists a unique vertex $z \in \mathbf{T}$ such that $[\rho, x] \cap [\rho, y] = [\rho, z]$.*

Proof. Let $a = \sup\{b \in [0, d(\rho, x)] : f_{\rho x}(b) \in [\rho, y]\}$, and let $z = f_{\rho x}(a)$. By the closeness of the sets $[\rho, x]$ and $[\rho, y]$, we know that $z \in [\rho, x] \cap [\rho, y]$, implying that $[\rho, z] \subseteq [\rho, x] \cap [\rho, y]$. On the other hand, if $z' \in [\rho, x] \cap [\rho, y]$, then $f_{\rho x}^{-1}(z') \in \{b \in [0, d(\rho, x)] : f_{\rho x}(b) \in [\rho, y]\}$, and so $f_{\rho x}^{-1}(z') \leq a$. Using the fact that $f_{\rho x}$ is an isometric embedding we can see that $d(\rho, z) = a$ and that $f|_{[0,a]}$ is the unique isometric embedding of $[0, d(\rho, z)]$ into $\mathbf{T}$. Hence, $z' \in [\rho, z]$ and $[\rho, x] \cap [\rho, y] \subseteq [\rho, z]$. Uniqueness is straightforward. If $[\rho, x] = [\rho, y]$ for any $x, y \in \mathbf{T}$, then $x \preceq y$ and $y \preceq x$. In particular $x = y$. □

There are many equivalent notions of real trees. Almost all of them use (i) (which is called the unique geodesic condition), but (ii) (the no-loop property) could be restated in any number of ways [Jan23]. Item (i) also is the property that asserts connectedness. There is one common equivalent description that does not use (i) and we shall record it because it is fun. Rather than pretend that I can say anything about the proof, I shall simply state it and bask in its glory ([Jan23] discusses this equivalent definition as well if you would like to learn about it).

Theorem 2.3. *A compact rooted metric space (X, d) is a real tree if and only if it is path-connected and satisfies the four-point condition :*

$$d(x_1, x_2) + d(x_3, x_4) \leq \max\{d(x_1, x_3) + d(x_2, x_4), d(x_1, x_4) + d(x_2, x_3)\},$$

for all $x_1, x_2, x_3, x_4 \in X$.

Ok, moving on. With the goal of convergence theorems in mind, we would like to have a notion of distance between two real trees. In most cases, our particular choice of distance function is the Gromov-Hausdorff distance. There are multiple equivalent definitions of this distance, however, for now at least we take the following one to be ours. For $(\mathbf{T}_1, d_1)$ and $(\mathbf{T}_2, d_2)$ real trees, we call $C \subseteq \mathbf{T}_1 \times \mathbf{T}_2$ a (root-preserving) correspondence between $\mathbf{T}_1$ and $\mathbf{T}_2$ if:

(i) $\forall x_1 \in \mathbf{T}_1\ \exists x_2 \in \mathbf{T}_2$ such that $(x_1, x_2) \in C$,

(ii) $\forall x_2 \in \mathbf{T}_2\ \exists x_1 \in \mathbf{T}_1$ such that $(x_1, x_2) \in C$, and

(iii) $(\rho_1, \rho_2) \in C$, where ρ_1 and ρ_2 are the roots of the trees $\mathbf{T}_1$ and $\mathbf{T}_2$ respectively.

The space of all correspondences between $\mathbf{T}_1$ and $\mathbf{T}_2$ is denoted by $\mathcal{C}(\mathbf{T}_1, \mathbf{T}_2)$. Then, we define the Gromov-Hausdorff distance between $(\mathbf{T}_1, d_1)$ and $(\mathbf{T}_2, d_2)$ as

$$d_{GH}(\mathbf{T}_1, \mathbf{T}_2) = \frac{1}{2} \inf_{C \in \mathcal{C}(\mathbf{T}_1, \mathbf{T}_2)} \operatorname{dis}(C),$$

where

$$\operatorname{dis}(C) = \sup \left\{ |d_1(x_1, y_1) - d_2(x_2, y_2)| : (x_1, x_2),\ (y_1, y_2) \in C \right\}.$$

There is a slightly more intuitive definition of the GH distance in terms of the Hausdorff distance of isometric embeddings of $\mathbf{T}_1$ and $\mathbf{T}_2$ into a mutual space. This definition will be of use later down the line, and for this sake we introduce it now.

Definition 2.4. *The Hausdorff distance* d_H *between two compact sets* K_1, K_2 *of a metric space* (X, d) *is defined by*

$$\inf\{\epsilon > 0 : K_1 \subseteq K_2^\epsilon, K_2 \subseteq K_1^\epsilon\},$$

where $S^\epsilon = \{x \in X : d(x, S) \leq \epsilon\}$.

Lemma 2.5. *For two real trees* $(\mathbf{T}_1, d_1)$ *and* $\mathbf{T}_2, d_2)$ *with roots* ρ_1 *and* ρ_2 *we define a metric*

$$d(\mathbf{T}_1, \mathbf{T}_2) = \inf_{\varphi_1, \varphi_2} \left(d_H(\varphi(\mathbf{T}_1), \varphi(\mathbf{T}_2)) \vee d^*(\varphi_1(\rho_1), \varphi_2(\rho_2)) \right),$$

where the infimum is taken over all isometric embeddings of $\mathbf{T}_1$ *and* $\mathbf{T}_2$ *and choices of destination* (X^*, d^*).

Before moving on to the discussion of functional encodings, we shall record that $(\mathbb{T}, d_{GH})$ is a good metric space to work with.

Theorem 2.6. $(\mathbb{T}, d_{GH})$ *is a complete separable metric space.*

This proof is also quite Gromovian, and so I'll skip the proof for now too. Hopefully there is not too much protest about this.

2.2 Encoding real trees with functions

The point of this subsection is to argue why we can replace the study of real trees with the study of certain types of continuous functions. As noted in the summary of this section, this offers a bridge between the real trees of this section, and the plane trees of the previous section. To prove the results we need some further properties of real trees, which we prove before getting ahead of ourselves.

For a real tree $(\mathbf{T}, d)$ we shall call the set $\mathbf{T} \setminus L(\mathbf{T})$, where $L(\mathbf{T}) = \{x \in \mathbf{T} : \deg_{\mathbf{T}}(x) = 1\}$, the *skeleton* of the tree $\mathbf{T}$ and we shall write $\mathrm{skel}(\mathbf{T}) = \mathbf{T} \setminus L(\mathbf{T})$. The set of vertices with degree at least three is called the *branching set* of $\mathbf{T}$ and we denote it by $\mathrm{br}(\mathbf{T})$. For the next lemma, we recall the standard fact from analysis that all compact metric spaces are separable.

Lemma 2.7. *Let $(\mathbf{T}, d)$ be a real tree. Then,*

(i) $\mathrm{skel}(\mathbf{T}) = \cup_{n=1}^{\infty}[\rho, x_n)$ *for any dense sequence $(x_n)_{n=1}^{\infty}$ in $\mathbf{T}$.*

(ii) $\mathrm{br}(\mathbf{T})$ *is countable.*

(iii) *For any $x \in \mathbf{T}$, $\deg(x) \in \mathbb{N} \cup \{|\mathbb{N}|\}$. Moreover, if $(C_n)_{n=1}^{\infty}$ are the components in $(\mathbf{T} \setminus \{x\}, d)$, then $\mathrm{diam}(C_n) \to 0$ as $n \to \infty$.*

Proof. (i) Suppose that there was a half-open geodesic $[\rho, x)$ and a leaf y such that $y \in [\rho, x)$. Then, this would imply that deleting y separates x from the rest of the tree by the unique geodesic property. Hence, $\mathbf{T} \setminus \{y\}$ has two components, contradicting the fact that it has degree 1. Thus $\cup_{n=1}^{\infty}[\rho, x_n) \subseteq \mathrm{skel}(\mathbf{T})$.

Now, let $x \in \mathrm{skel}(\mathbf{T})$. There are at least two components in the set $\mathbf{T} \setminus \{x\}$. Let $\mathbf{T}'$ be one component not containing the root and let $z \in \mathbf{T}'$. By density, there exists some $n^* \geq 1$ be chosen such that $x_n^* \in (x, z)$ (we can see that (x, z) is not empty by the fact that $f_{\rho z}$ is an isometry). Then, this clearly implies that $x \in [\rho, x_{n^*})$ since $f_{\rho z}|_{[0, f_{\rho z}^{-1}(x_{n^*})]}$ is the unique isometry embedding of $[0, d(\rho, z)]$ into $\mathbf{T}$.

(ii) Suppose that $\mathrm{br}(T)$ is uncountable. Then, there is some set $[\rho, z)$ such that $\mathrm{br}(T) \cap [\rho, z)$ is uncountable by (i). Let $f : \mathbb{R} \to \mathrm{br}(T) \cap [\rho, z]$ be injective. In particular, for each $x \in \mathbb{R}$, we can associate a component $C(x)$ of $\mathbf{T} \setminus \{f(x)\}$ that does not intersect $[\rho, z]$. Let $v_x \in C(x)$ and $v_y \in C(y)$. Then, since $f(x) \neq f(y)$, we know that the unique geodesic from v_x to v_y must go through $[\rho, z]$ as the path $[v_x, f(x)] \cdot [f(x), f(y)] \cdot [f(y), v_y]$ is a continuous injective map (here $\cdot$ is seen as adjoining the two paths). In particular, this implies that $v_y \notin C(x)$ and $v_x \notin C(y)$. Thus $C(x) \cap C(y) = \emptyset$. Altogether, we get that there are uncountable many components of $\mathbf{T} \setminus [\rho, x]$, which means that $\mathbf{T}$ cannot be separable (any dense set of $\mathbf{T}$ would require an element in each component).

(iii) If there was a vertex of infinite degree, then at least one of the sets

$$S_k = \left\{ C \text{ a component of } (\mathbf{T} \setminus \{x\}, d) : \exists y \in C \text{ s.t. } d_{\mathbf{T}}(x, y) \geq \frac{1}{k} \right\}$$

would need to be uncountable. This contradicts the compactness of $\mathbf{T}$. If $\mathrm{diam}(C_n) \not\to 0$ as $n \to \infty$, then at least one of the sets

$$S_k = \left\{ C \text{ a component of } (\mathbf{T} \setminus \{x\}, d) : \exists y \in C \text{ s.t. } d_{\mathbf{T}}(x, y) \geq \frac{1}{k} \right\}$$

would be countable, also contradicting compactness. $\square$

The properties from Lemma 2.7 give us the ability to embed our trees into a continuum version of the Ulam-Harris tree. This allows for easy construction of valid functional encodings. What is a "valid" encoding? Let's getn into that now.

Let $f \in \{g : [0,\infty) \to [0,\infty) : \text{supp}(f) \text{ compact and connected}, g(0) = 0\} := C_c^+[0,\infty)$. We shall construct a real tree from the function. Define, for all $s, t \geq 0$,

$$m_f(s,t) = \inf_{\min(s,t) \leq r \leq \max(s,t)} f(r),$$

and $d_f(s,t) = f(s) + f(t) - 2m_f(s,t)$. Then, d_f is a metric on the set of equivalence classes $[0,\infty)/R_f$, where $R_f = \{(s,t) \in [0,\infty)\times[0,\infty) : d_f(s,t) = 0\}$. Our final theorem of this subsection asserts that the collection of all metric spaces $([0,\infty)/R_f, d_f)$ for functions $f \in C_c^+[0,\infty)$ is a rich enough set to fill our needs. For a function $f \in C_c^+[0\infty)$, we let $(\mathbf{T}_f, df_f)$ denote the space $([0,\infty)/R_f, d_f)$ with root $\rho = [0]_{R_f}$, the equivalence class of 0 under R_f. We would like to show the following:

(i) For any $f \in C_c^+[0,\infty)$, the pair $(\mathbf{T}_f, d_f)$ is a real tree.

(ii) For any two real trees $(\mathbf{T}_f, d_f)$ and $(\mathbf{T}_g, d_g)$, $d_{GH}(\mathbf{T}_f, \mathbf{T}_g) \leq 2\|f - g\|_\infty$.

(iii) For every real tree $(\mathbf{T}, d)$, there exists a function $f \in C_c^+[0,\infty)$ such that $(\mathbf{T}, d) = (\mathbf{T}_f, d_f)$.

One way to show (i) is to observe that any metric spaces of the form $(\mathbf{T}_f, d_f)$ satisfy the four-point condition, which implies they are all real trees via Theorem 2.3. We shall take a more elementary approach that relies most on basic analysis techniques. To prove (i) and (ii) we first prove the results for semi-linear functions (defined below) and then invoke the completeness of $(\mathbb{T}, d_{GH})$ to extend to all functions in $C_c^+[0,\infty)$. Then, (iii) will follow from an approximation of real trees with trees that are "equivalent" to a combinatorial tree, in the sense that they are described by semi-linear functions. A different approach to prove the same theorem that argues directly with the isometry definition is covered in [LG05].

Let $f \in C_c^+[0,\infty)$. We say that f is a semi-linear if there is $\epsilon, \Delta > 0$ such that for any $n \geq 0$ $f(x) = f(n\epsilon) + \Delta(x - n\epsilon)$ or $f(x) = f(n\epsilon) - \Delta(x - n\epsilon)$ for $x \in [n\epsilon, (n+1)\epsilon]$. We shall label the set of semi-linear functions in $C_c^+[0\infty)$ with C_L.

Lemma 2.8. *For any $f \in C_L$, the pair $(\mathbf{T}_f, d_f)$ is a real tree. This proof needs some love.*

Proof. Let $f \in C_L$. We can construct a metric space isometric to $\mathbf{T}_f, d_f$ by grafting together multiple intervals of length Δ. In fact, we shall construct the particular tree via a modified DFQ construction, similar to the one used to explain the contour function.

We shall see ourselves as drawing the tree with pencil and paper. The "up" movements shall correspond to the drawing of new branches of length Δ and at the end of these intervals our "pencil" will be at the leaf which completes this interval. "Down" movements correspond to tracing our pencil back down the drawn paths towards the root, with the pencil now at the unique interior vertex that is distance Δ closer to the root than where it was before. For the sake of formalism, we complete the construction inside a continuum version of the Ulam-Harris tree,

$$\mathbf{U} = (\mathbb{N}^0 \times \{0\}) \cup \bigcup_{n=1}^{\infty} (\mathbb{N}^n \times (0, \Delta]),$$

where the final coordinate in $(u^1, ..., u^n, x)$ is meant to indicate how far along the length Δ line from $(u_1, ..., u^{n-1})$ to $(u^1, ..., u^n)$ the point is. Nothing about the structure of the tree has changed, just that we now view each edge as existing in tree and having length Δ. We shall order the vertices $(\mathbb{N}^0 \times \{0\}) \cup (\cup_{n=1}^{\infty} \mathbb{N}^n \times \{1\})$ with the lexicographical order from the ordinary Ulam-Harris tree. For any plane tree $\mathbf{t}$, the subset of the continuum Ulam-Harris tree $\mathbf{T}_\Delta = \mathbf{t} \setminus \emptyset \times (0, 1] \cup \{\emptyset\}$ can be shown to be a real tree when equipped with the metric

$$d_{\mathbf{T}_\Delta}((u, x), (v, y)) = \begin{cases} x - y + \Delta \cdot d_{\mathbf{t}}(u^-, y^-), & \text{if } y \preceq_{\mathbf{t}} x \\ x - y + \Delta \cdot d_{\mathbf{t}}(u^-, y^-), & \text{if } x \preceq_{\mathbf{t}} y \\ x + y + \Delta \cdot d_{\mathbf{t}}(u^-, y^-), & \text{if } x \perp y \end{cases},$$

where $\perp$ is indicating incomparability in the genealogical order, $u^- = (u^1, ..., u^{n-1})$ for $u = (u^1, ..., u^n) \in \mathbf{t} \setminus \emptyset$, and $d_{\mathbf{t}}$ is the ordinary graph distance metric in $\mathbf{t}$. The metric $d_{\mathbf{T}_\Delta}$ is just an extension of $d_{\mathbf{t}}$ to include these points that are between two vertices scaled by Δ. It is quite straightforward to show that $(\mathbf{T}_\Delta, d_{\mathbf{T}_\Delta})$ is a real tree as it is embedded in the Ulam-Harris tree (the proof is really a sanity check about the definition of a real tree because it would be quite a failure if such a tree failed to meet the definition).

Consider the function $\Gamma(t) = \Delta^{-1} f(t\epsilon)$ for $0 \leq t \leq \text{length}(\text{supp}(f))\epsilon^{-1} := \ell\epsilon^{-1}$. Notice that this is the contour function for a unique plane tree $\mathbf{t}^f$ (see remark after Theorem 1.6). We define T^f_Δ to be the continuum version of the plane tree $\mathbf{t}^f$ as defined above. We can construct an isometry $\varphi : \mathbf{T}^f_\Delta \to \mathbf{T}_f$. For this, we recall the depth-first exploration of $\mathbf{t}^f$ that we defined when introducing the contour function $v_0, ..., v_{2(n-1)}$. Since T^f_Δ contains all of the in-between points between two vertices we can define a new T^f_Δ-valued version of the contour function $\gamma : [0, \ell] \to T^f_\Delta$:

$$\gamma(t) = \sum_{k=1}^{2(n-1)} f_{v_{k-1} v_k}(t - \epsilon(k-1)) \mathbf{1}_{\{\epsilon(k-1) \leq t \leq \epsilon k\}},$$

where $f_{v_{k-1} v_k}$ is the unique isometry from $[0, \Delta] \to \mathbf{T}^f_\Delta$ starting at v_{k-1} and ending at v_k. Now, define $\varphi(v) = [\inf\{0 \leq t \leq \ell : \gamma(t) = v\}]_{R_f}$. It is a bit clunky, but not too hard, to verify that φ is in fact an isometry. □

Lemma 2.9. *Let* $f, g \in C_L$. *Then,* $d_{GH}(\mathbf{T}_f, \mathbf{T}_g) \leq 2\|f - g\|_\infty$.

Proof. Let $C = \{([x]_{R_f}, [y]_{R_g}) : x = f(t),\ y = g(t) \ \text{ for some } t \geq 0\}$. It can be observed easily that this is a root-preserving correspondence. Let $(x_1, y_1), (x_2, y_2) \in C$ (we are supressing the $[\cdot]_{R_f}$ now for clarity). Then, for some $s, t \geq 0$ we have,

$$|d_f(x_1, x_2) - d_g(y_1, y_2)| \leq |f(s) - g(s)| + |f(t) - g(t)| + 2|m_f(s,t) - m_g(s,t)|.$$

Without loss of generality we can assume that $m_f(s,t) \geq m_g(s,t)$. By the continuity of the two functions and the fact that $[s \wedge t, s \vee t]$ is closed there is some $p \geq 0$ such that $m_g(s,t) = g(p)$. Then,

$$2|m_f(s,t) - m_g(s,t)| \leq 2(f(p) - g(p)) \leq 2\|f - g\|_\infty.$$

Altogether, we get that

$$d_{GH}(\mathbf{T}_f, \mathbf{T}_g) \leq \frac{1}{2}\,\mathrm{dis}(C) \leq 2\|f - g\|_\infty.$$

□

Lemma 2.10. C_L *is dense in* $C_c^+[0,\infty)$ *under the norm* $\|\cdot\|_\infty$.

Proof. It suffices to show the result for Lipschitz functions in $C_c^+[0,\infty)$ as they are dense in the set $C_c^+[0,\infty)$. Let $f \in C_c^+[0,\infty)$ be C-Lipschitz. Let $\Delta_n = C$ and $\epsilon_n = (S - I)n^{-1}$, where $S = \sup \mathrm{supp}(f)$ and $I = \inf \mathrm{supp}(f)$. Define recursively

$$P_n(j) = \begin{cases} +1, & \text{if } f(j\epsilon + I) \geq f_n(j\epsilon + I) \\ -1, & \text{otherwise} \end{cases}.$$

Finally, we set

$$f_n(t) = \sum_{j=0}^{(n-1)} P_n(j)\Delta_n\big((t - j\epsilon)_+ \vee \epsilon\big) - \sum_{j=0}^{f_n(S)(\Delta_n\epsilon_n)^{-1}} \Delta_n\big((t - S) - j\epsilon) \vee \epsilon\big).$$

The second sum exists only to make sure that the function is in $C_c^+[0,\infty)$ as promised, it disappears in the limit. We claim that $\|f - f_n\|_\infty \leq 2\Delta_n\epsilon_n$. We can proceed via induction. Suppose that $\sup_{x\in[I,k\epsilon+I]} |f_n(x) - f(x)| \leq 2\Delta_n\epsilon_n$ for some $0 \leq k < n-1$. Then, in particular $|f_n(k\epsilon + I) - f(k\epsilon + I)| \leq 2\Delta_n\epsilon_n$. There are two cases to consider. case 1: $f(k\epsilon + I) \geq f_n(k\epsilon + I)$. In this case the function f_n increases on the next interval. Since $|f(t) - f(k\epsilon + I)| \leq C(t - k\epsilon - I)$, we have that

$$\sup_{t\in[k\epsilon+I,(k+1)\epsilon+I]} (f(t) - f_n(t)) \leq f(k\epsilon+I) + C(t - k\epsilon - I) - f_n(k\epsilon+I) - C(t - k\epsilon - I) \leq 2\Delta_n\epsilon_n,$$

and

$$\sup_{t\in[k\epsilon+I,(k+1)\epsilon+I]} (f_n(t) - f(t)) \leq f(k\epsilon + I) + \Delta_n\epsilon_n - f_n(k\epsilon + I) - (-\Delta_n\epsilon_n) \leq 2\Delta_n\epsilon_n.$$

In particular, we have using the assumption that $\sup_{x\in[I,(k+1)\epsilon+I]} |f_n(x) - f(x)| \leq 2\Delta_n\epsilon_n$. case 2: $f(k\epsilon + I) < f_n(k\epsilon + I)$. This case goes almost identically to the first case so we shall omit this. We note that this induction actually extends to include times above S without changing anything as the second sum defining $f_n(t)$ is only empty when $f_n(S) > 0 = f(S)$. Thus, the proof is done as $\Delta_n\epsilon_n \to 0$ as $n \to \infty$. □

Theorem 2.11. *The claims stated at the beginning of the section hold.*

(i) *For any* $f \in C_c^+[0, \infty)$*, the pair* $(\mathbf{T}_f, d_f)$ *is a real tree.*

(ii) *For any two real trees* $(\mathbf{T}_f, d_f)$ *and* $(\mathbf{T}_g, d_g)$*,* $d_{GH}(\mathbf{T}_f, \mathbf{T}_g) \leq 2\|f - g\|_\infty$.

(iii) *For every real tree* $(\mathbf{T}, d)$*, there exists a function* $f \in C_c^+[0, \infty)$ *such that* $(\mathbf{T}, d) = (\mathbf{T}_f, d_f)$.

Proof. (i) and (ii) are immediate from the work done in the section up to now. Come back to this. Honestly, it seems like you need to be careful and use a lot of the properties from Lemma 2.7. Also, note that $\mathbf{T} \setminus \mathrm{br}(\mathbf{T})$ has countably many components. □

3 Scaling limits of random walks and Bienaymé trees

We finally prove some scaling limits in this section. We begin with building up the theory of scaling limits for random functions, explaining the topological backing behind it and proving Donsker's Theorem. Using the theorem and results from the previous two sections, we prove scaling limits for the height function of both conditioned and un-conditioned critical Bienaymé trees. As a corollary, we obtain a scaling limit in the Gromov-Hausdorff topology for critical conditioned trees to a random real tree called the Brownian CRT. It is defined to be a real tree that is encoded by a unit length Brownian excursion.

3.1 Random functions in $C[0,1]$ and Donsker's Theorem

I borrowed a lot of the material in this subsection from [Bil13]. In order to discuss scaling limits, we require some results connecting random walks and Brownian motion. We also desire some good tools to explore the convergence of random functions with our functional encodings of real trees in mind. Our setup in this section is a sequence of i.i.d. random variables $(\xi_n)_{n\geq 1}$ with mean 0 and variance 1. Let $S_k = \sum_{i=1}^k \xi_i$. The sequence of random functions that we consider is $(W_n)_{n\geq 1}$, where $W_n : [0,1] \to \mathbb{R}$ is such that

$$W_n(t) = \frac{S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)\xi_{\lceil nt \rceil}}{\sqrt{n}}. \tag{2}$$

Donsker's Theorem essentially asserts that the functions $W_n(t)$ converge towards Brownian motion on the interval $[0,1]$.

Theorem 3.1 (Donsker's Theorem)**.**

$$\Big(W_n(t) : t \in [0,1]\Big) \xrightarrow{\mathcal{L}} \Big(B(t) : t \in [0,1]\Big),$$

as $n \to \infty$ *in the space* $(C[0,1], \|\cdot\|_\infty)$*, where* $(B(t) : t \geq 0)$ *is standard one dimensional Brownian motion that starts with* $B(0) = 0$*.*

While we can intuitively view this theorem as being a sort of generalization of the central limit theorem (the sequence $(W_n(1)/\sqrt{n})_{n\geq 1}$ is exactly the sequence $(S_n/\sqrt{n})_{n\geq 1}$), we need to recall some topological tools to be able to complete the proof. This increased difficulty is due to the fact that the claimed convergence is in the space $C[0,1]$ rather than $\mathbb{R}$. Specifically, we desire an equivalence between convergence in distribution and convergence of finite dimensional marginals for continuous functions.

3.1.1 Convergence of measures on $C[0,1]$

Let us begin by dragging some old dusty theorems out from our attic.

Definition 3.2. *Let (X,τ) be a Hausdorff space and let $\mathcal{P}$ be the space of all probability measures on X equipped with the Borel sigma-algebra. A set $S \subseteq \mathcal{P}$ is called* tight *if for all $\epsilon > 0$ there is a compact set $K(\epsilon)$ such that $\sup_{\mu\in S}\mu(X \setminus K(\epsilon)) < \epsilon$.*

Theorem 3.3 (Prokhorov's Theorem)**.** *Let (X,d) be a separable metric space and let $\mathcal{P}$ be the set of all probability measures on X with the Borel sigma-algebra. Then, $S \subseteq \mathcal{P}$ is tight if and only if it is pre-compact.*

An almost direct consequence of Prokhorov's Theorem is worth recording.

Corollary 3.4. *Let $(\mu_n)_{n=1}^{\infty}$, μ be probability measures on $(C[0,1], \|\cdot\|_\infty)$. If the finite-dimensional marginals of $(\mu_n)_{n=1}^{\infty}$ converge in distribution to the finite-dimensional marginals of μ, and if $(\mu_n)_{n=1}^{\infty}$ is tight, then $\mu_n \xrightarrow{\mathcal{L}} \mu$ as $n \to \infty$.*

Proof. Recall that, for probability measures μ and ν on $[0,1]$, $\mu = \nu$ if and only if $\mu\pi_{t_1,\dots,t_k} = \nu\pi_{t_1,\dots,t_k}$ for $0 \le t_1 \le \dots \le t_k \le 1$, where $\pi_{t_1,\dots,t_k}$ is the projection onto the coordinates $t_1,\dots,t_k$ (this can be observed by a standard $\pi - \lambda$ system proof).

Let $(\mu_{n_k})_{k=1}^{\infty})$ be a subsequence of $(\mu_n)_{n=1}^{\infty})$. By pre-compactness, this sequence has a convergent subsequence, tending to some limit μ^*. By the finite-dimensional marginals convergence and the fact from the previous paragraph, it holds that $\mu^* = \mu$. Hence, every subsequence of $(\mu_n)_{n=1}^{\infty}$ has a further subsequence that converges to μ. It is well known that this implies that $\mu_n \xrightarrow{\mathcal{L}} \mu$ as $n \to \infty$. □

Theorem 3.5 (Arzelà-Ascoli Theorem)**.** *A set $S \subseteq C[0,1]$ is pre-compact if and only if $\sup_{f\in S}|f(0)| < \infty$ and $\lim_{\delta\to 0}\sup_{f\in S} w_f(\delta) = 0$, where $w_f(\delta) = \sup_{|s-t|<\delta}|f(s) - f(t)|$ for all $0 < \delta < 1$.*

The function w_x is called the modulus of continuity. For the rest of this subsection is essentially to prove that sums of i.i.d. random variables are tight. A couple technical lemmas about tightness are needed for this. The first is a translation of tightness in $C[0,1]$ to a pair of conditions that mirror the pre-compactness definition given by the Arzelà-Ascoli Theorem.

Lemma 3.6. *A sequence of measures $(\mu_n)_{n=1}^{\infty}$ on $(C[0,1], \|\cdot\|_\infty)$ is tight if and only if the following two conditions hold:*

(i) *for all $\epsilon > 0$ there is $N, t \ge 0$ such that $\mu(\{x : |x(0)| > t\}) \le \epsilon$ for all $n \ge N$,*

(ii) *for all $\epsilon > 0$, $\lim_{\delta\to 0}\limsup_{n\to\infty}\mu_n(\{x : w_x(\delta) \ge \epsilon\})$.*

Proof. Suppose that the sequence is tight. Choose some $K_\epsilon \subseteq C[0,1]$ and $t \ge 0$ such that $\mu_n(K_\epsilon) \ge 1-\epsilon$, $K_\epsilon \subseteq \{x : |x(0)| \ge t\}$ and $K_\epsilon \subseteq \{x : w_x(\delta) \le \epsilon\}$ for all $n \ge 1$ and $\delta > 0$ chosen sufficiently small. The set is guaranteed to exist by the

Arzelà-Ascoli Theorem. Then, it quickly follows that $\mu_n(\{x : |x(0)| \geq t\}) \leq \epsilon$ and $\lim_{\delta\to 0} \sup_{n\geq 1} \mu_n(\{x : w_x(\delta) \geq \epsilon\})$ by the continuity of the function w_x.

For the reverse direction, we may instead show (ii)': for all $\eta, \epsilon > 0$ that $\mu_n(\{x : w_x(\delta) \geq \epsilon\}) \leq 1 - \eta$ for all n above some chosen $N \geq 0$.

Suppose that (i) and (ii)' hold for $N \geq 0$. We claim that each of the individual measures $\mu_1, ..., \mu_N$ are tight.

Since $C[0, 1]$ is separable there is, for each $k \geq 0$, a collection of balls of radius k, $A_1, ..., A^{(k)}_{n_k}$ such that $\mu_1(\cup_{i=1}^{n_k} A_i^{(k)}) \geq 1 - \epsilon 2^{-k}$. The closure K of the set $\cap_{k=1}^{\infty} \cup_{i=1}^{n_k} A_i^{(k)}$ has measure $\mu_1(K) \geq 1 - \epsilon$ and is totally bounded. By the completeness of $C[0, 1]$ we can conclude that K is compact.

Returning back to the proof, a simple application of the union bound proves that the collection $\mu_1, ..., \mu_N$ is tight. This implies that the inequalities from (i) and (ii)' hold for this collection too. In particular, this allows us to assume that $N = 1$ in (i) and (ii)'. Choose some $t \geq 0$ such that $\mu_n(\{x : |x(0)| \leq t\}) \geq 1 - \epsilon$ for all $n \geq 1$ and choose δ_k such that $\mu_n(\{x : w_x(\delta) < k^{-1}\}) \geq 1 - \epsilon 2^{-k}$ for all $n \geq 1$. Then, if we set K to be the closure of

$$(\{x : |x(0)| \leq t\}) \cap \bigcap_{k=1}^{\infty} \left\{x : w_x(\delta) < k^{-1}\right\},$$

we have that $\mu_n(K) \geq 1 - 2\epsilon$ for all $n \geq 1$. By the Arzelà-Ascoli Theorem K is compact. $\square$

In order to do probabilistic computations cleanly we need to be able to work with a nicer form of the modulus of continuity than is provided via its definition. Our final lemma covers this for us.

Lemma 3.7. *Suppose that* $0 = t_0 \leq ... \leq t_k = 1$ *is such that* $\min_{1\leq i\leq k}(t_i - t_{i-1}) \geq \delta$. *Then, for any* $x \in C[0, 1]$,

$$w_x(\delta) \leq 3 \max_{1\leq i\leq k} \sup_{t_{i-1}\leq t\leq t_i} |x(t) - x(t_{i-1})|,$$

and

$$\mu(\{x : w_x(\delta) \geq 3\epsilon\}) \leq \sum_{i=1}^{k} \mu\left(\left\{x : \sup_{t_{i-1}\leq t\leq t_i} |x(t) - x(t_{i-1})| \geq \epsilon\right\}\right)$$

for any measure μ *on* $C[0, 1]$.

Proof. The first inequality is a simple triangle inequality argument. Let

$$M = \max_{1\leq i\leq k} \sup_{t_{i-1}\leq t\leq t_i} |x(t) - x(t_{i-1})|.$$

If $|s - t| \leq \delta$, then they are either in adjacent intervals or the same interval. Suppose that $s, t \in [t_{i-1}, t_i]$ for some chosen i. Then,

$$|x(s) - x(t)| \leq |x(s) - x(t_{i-1})| + |x(t) - x(t_{i-1})| \leq 2M.$$

Suppose that $s \in [t_{i-1}, t_i]$ and $t \in [t_i, t_{i+1}]$ for some chosen i. Then,

$$|x(s) - x(t)| \leq |x(s) - x(t_{i-1})| + |x(t_{i-1}) - x(t_i)| + |x(t) - x(t_i)| \leq 3M.$$

The second inequality follows from a union bound. □

3.1.2 Back to Donsker's Theorem

Equipped with Corollary 3.4, proving Donsker's Theorem is as easy as verifying the convergence for finite-dimensional marginals and the tightness condition.

Lemma 3.8. *Suppose that $(W_n)_{n=1}^{\infty}$ is defined as in (2). If*

$$\lim_{x\to\infty} \limsup_{n\to\infty} x^2 \mathbf{P}\left(\max_{1\leq k\leq n} |S_k| \geq x\sqrt{n}\right) = 0,$$

then the sequence $(W_n)_{n=1}^{\infty}$ is tight.

Proof. We proceed by showing the Arzelà-Ascoli conditions hold in Lemma 3.6. Condition (i) is immediate as $W_n(0) = 0$ for all $n \geq 1$, so we only need to verify the condition on the modulus of continuity for an arbitrary $\epsilon > 0$,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbf{P}(w_x(W_n, \delta) \geq \epsilon) = 0.$$

Let $= m_0 \leq ... \leq m_k = n$, and consider times $t_i = \frac{m_i}{n}$. Applying Lemma 3.7 we get that

$$\mathbf{P}(w(W_n, \delta) \geq 3\epsilon) \leq \sum_{i=1}^{k} \mathbf{P}\left(\sup_{t_{i-1}\leq t\leq t_i} |W_n(t) - W_n(t_{i-1})| \geq \epsilon\right)$$

whenever $\delta \leq \frac{m_i - m_{i-1}}{n}$ for all $1 \leq i \leq k$. The chosen times are important because, by definition, $W_n(t_i) = S_{m_i}/\sqrt{n}$. Thus,

$$\sup_{t_{i-1}\leq t\leq t_i} |W_n(t) - W_n(t_{i-1})| = \frac{1}{\sqrt{n}} \max_{m_{i-1}\leq j\leq m_i} |S_j - S_{m_{i-1}}|,$$

and

$$\begin{aligned}\mathbf{P}(w(W_n, \delta) \geq 3\epsilon) &\leq \sum_{i=1}^{k} \mathbf{P}\left(\frac{1}{\sqrt{n}} \max_{m_{i-1}\leq j\leq m_i} |S_j - S_{m_{i-1}}| \geq \epsilon\right) \\ &= \sum_{i=1}^{k} \mathbf{P}\left(\max_{0\leq j\leq m_i - m_{i-1}} |S_j| \geq \sqrt{n}\epsilon\right)\end{aligned}$$

for appropriately chosen $(m_i)_{i=1}^{k}$ to suit the conditions on δ (the second equality is a consequence of the ξ_n's being i.i.d.). This bound leaves us with a much more familiar expression to deal with. First, we need to finalize our choices of parameters though.

Let $m = \lceil n\delta \rceil$, let $k = \lceil \delta^{-1} \rceil$, and let $m_i = 2im$ for each $0 \leq i \leq k$. Then, $m_i - m_{i-1} = m$ for all i and $(m_i - m_{i-1})/n \to 2\delta > \delta$ as $n \to \infty$.

With these chosen parameters the above expression becomes

$$\begin{aligned}\mathbf{P}(w(W_n, \delta) \geq 3\epsilon) &\leq \delta^{-1}\mathbf{P}\left(\max_{0\leq j\leq 2m} |S_j| \geq \epsilon\sqrt{\frac{m}{\delta}}\right) \\ &= 2 \cdot (2\delta)^{-1}\mathbf{P}\left(\max_{0\leq j\leq 2m} |S_j| \geq \epsilon\frac{1}{\sqrt{2\delta}}\sqrt{2m}\right) \\ &= \frac{2}{\epsilon^2}x^2\mathbf{P}\left(\max_{0\leq j\leq 2m} |S_j| \geq x\sqrt{2m}\right),\end{aligned}$$

where we set $x = \epsilon(2\delta)^{-1/2}$. Note that, as $\delta \to 0$, $x \to \infty$. From here, applying the assumption is enough to yield condition (ii) in Lemma 3.6, which proves tightness. □

We are now ready to prove Donsker's Theorem, but first let us quickly recall the properties that characterize Brownian motion.

Definition 3.9. *One dimensional* Brownian motion *is a real-valued stochastic process* $(B(t) : t \geq 0)$ *that satisfies the following properties:*

(i) $B(0) = 0$.

(ii) *If* $t_0 < t_1 < ... < t_n$, *then* $B(t_0), B(t_1) - B(t_0), ..., B(t_n) - B(t_{n-1})$ *are independent.*

(iii) *If* $s < t$, *then* $B(s+t) - B(s) \stackrel{\mathcal{L}}{=} N(0, t-s)$.

These properties need to be shown for the limit of the finite-dimensional marginals of $(W_n)_{n=1}^{\infty}$ to complete the proof. If we show that, for any collection of times $0 = t_0 \leq ... \leq t_k$ for some $k \geq 0$,

$$(W_n(t_1) - W_n(t_0), ..., W_n(t_k) - W_n(t_{k-1})) \xrightarrow{\mathcal{L}} (X_1, ..., X_k),$$

where the X_i's are independent with $X_i \stackrel{\mathcal{L}}{=} N(0, t_i - t_{i-1})$, then we are done.

Theorem (Donsker's Theorem). *Let* $(\xi_n)_{n\geq 1}$ *be a sequence of i.i.d. random variables with mean* 0 *and variance* 1. *Let* $S_k = \sum_{i=1}^{k} \xi_i$. *Define random functions* $(W_n)_{n\geq 1}$ *where*

$$W_n(t) = \frac{S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)\xi_{\lceil nt \rceil}}{\sqrt{n}}.$$

Then,

$$\Big(W_n(t) : t \in [0,1]\Big) \xrightarrow{\mathcal{L}} \Big(B(t) : t \in [0,1]\Big),$$

as $n \to \infty$ *in the space* $(C[0,1], \|\cdot\|_\infty)$, *where* $(B(t) : t \geq 0)$ *is standard one dimensional Brownian motion that starts with* $B(0) = 0$.

Proof. Let $t \geq s \geq 0$. $W_n(s) = S_{\lfloor ns \rfloor}/\sqrt{n} + X_n$ and $W_n(t) - W_n(s) = (S_{\lfloor nt \rfloor} - S_{\lfloor sn \rfloor})/\sqrt{n} + Y_n$, where X_n and Y_n are random variables that tend to 0 almost surely as $n \to \infty$. Basic properties of random walks assert that $S_{\lfloor ns \rfloor}$ and $(S_{\lfloor nt \rfloor} - S_{\lfloor sn \rfloor})$ are independent. By the central limits theorem and the continuous mapping theorem, we get that $W_n(s) \xrightarrow{\mathcal{L}} X$ and $W_n(t) - W_n(s) \xrightarrow{\mathcal{L}} Y$, where $X \stackrel{\mathcal{L}}{=} N(0, s)$ and $Y \stackrel{\mathcal{L}}{=} N(0, t-s)$ are independent. The general case is similar, and so we can move on to tightness. By Etemadi's inequality (see remark below if you are unfamiliar),

$$x^2 \mathbf{P}\left(\max_{0 \leq k \leq n} |S_k| \geq x\sqrt{n}\right) \leq 3x^2 \max_{0 \leq k \leq n} \mathbf{P}\left(|S_k| \geq x\sqrt{n}/3\right).$$

Let $k^*(x)$ be a constant depending only on x, chosen such that $\mathbf{P}(|S_k| \geq x\sqrt{k}/3) \leq \mathbf{P}(N(0,1) \geq x/3) + o(x^{-3})$ for all $k^* \leq k$. Then, by Markov's inequality,

$$3x^2 \max_{k^*(x) \leq k \leq n} \mathbf{P}\left(|S_k| \geq x\sqrt{n}/3\right) \leq \frac{3^4 \mathbf{E}|N(0,1)|}{x} = o_x(1)$$

for any $n \geq 1$. In particular,

$$3x^2 \limsup_{n \to \infty} \max_{k^*(x) \leq k \leq n} \mathbf{P}\left(|S_k| \geq x\sqrt{n}/3\right) = o_x(1)$$

Then, for $1 \leq k < k^*$ Chebyshev's inequality gives

$$3x^2 \limsup_{n \to \infty} \max_{0 \leq k < k^*} \mathbf{P}\left(|S_k| \geq x\sqrt{n}/3\right) \leq \limsup_{n \to \infty} \frac{3^3 k^*}{n} = 0$$

for any x. Altogether, this proves tightness by Lemma 3.8. □

Remark. Since I had never seen it before, I will present Etemadi's inequality (a pretty tidy tool to have in your kit in my opinion). Let $(\xi_n)_{n=1}^{\infty}$ be a sequence of i.i.d. random variables, let $(S_n)_{n=0}^{\infty}$ be the partial sum of the first n ξ's, and let $t \geq 0$. Then, Etemadi's inequality states that

$$\mathbf{P}\left(\max_{1 \leq k \leq n} |S_k| \geq 3t\right) \leq 3 \max_{1 \leq k \leq n} \mathbf{P}(|S_k| \geq t).$$

With it, you can prove a weaker form of Kolmogorov's maximal inequality (one still strong enough to prove the strong law of large numbers though).

Remark. An entirely equivalent argument with A replacing 1 proves Donsker's Theorem on all compact sets of $[0, \infty)$, and hence proves that the result holds in the space $C[0, \infty)$ under the topology of uniform convergence on compact sets.

3.2 Skorokhod space

Just drop theorems later no proofs.

3.3 Convergence of the height process for Bienaymé forests

We now have everything we need to start exploring relationships between random combinatorial trees and random real trees. In this subsection, we start by showing that the height process of a critical Bienaymé tree converges to a Brownian excursion (see remark below). Let $(T_n)_{n=1}^{\infty}$ be a sequence of independent Bienaymé(μ) distributed random variables for some critical offspring distribution μ. Throughout the rest of this section we assume that all child distributions are critical. Let $S_i = |T_1| + ... + |T_i|$ for all $i \geq 1$. We define the height process of the forest by setting $H_k = h_{T_i}(k - X_{i-1})$ for all $X_{i-1} \leq k < X_i$ (recall that the height process of a tree $\mathbf{t}$ is defined on $0, ..., |\mathbf{t}| - 1$). Since the height process visits zero only once, the height process encodes the whole forest.

Before getting to the main theorem let's pause to address why the height function is the one we need to analyze. Our end goal is to prove the convergence of Bienaymé trees (specifically conditioned ones) to the Brownian CRT. To do this with Theorem 2.11, we need to show that the contour function of the tree converges to a Brownian excursion in distribution. We study the height function instead of the contour function is that the height function enjoys a nice connection with the DFQ process, which is distributed like a simple random walk for Bienaymé trees. Extending the result to include convergence of contour functions does not take much extra work. Of course, the desire to instead study the height function is what leads us to explore the convergence in Skorokhod space rather than $(C[0, 1], \|\cdot\|_\infty)$.

Remark. A Brownian excursion is, informally, a Brownian motion that is conditioned to be non-negative and takes the value 0 at time 1. This event of course has probability zero of occurring so we should be more careful than this. There are many legal ways to generate such stochastic processes, but one simple one goes as follows: Let $\tau_1, \tau_2 > 0$ be such that $B(\tau_1) = B(\tau_2) = 0$, $B(t) \geq 0$ for all $\tau_1 < t < \tau_2$ and $\tau_2 - \tau_1 \geq 1$ for some Brownian motion $(B(t) : t \geq 0)$. These times exist almost surely as Brownian motion is recurrent with expected return time to zero being unbounded. Then, set $e(t) = B((\tau_2 - \tau_1)t + \tau_1)/\sqrt{\tau_2 - \tau_1}$ for each $0 \leq t \leq 1$. This gives us a stochastic process with the correct characteristics.

Much of the work on combinatorial trees from Section 1 can be summarized with the following lemma.

Lemma 3.10. *For all* $n \geq 0$, $H_n = |\{0 \leq k \leq n-1 : S_k = \inf_{k \leq j \leq n} S_j\}|$, *where* $(S_n)_{n=0}^{\infty}$ *is a simple random walk with jump distribution* ν *defined by* $\nu(k) = \mu(k+1)$ *for all* $k \geq -1$.

Proof. Note that for $X_{i-1} \leq k < X_i$, the indices in $\{0 \leq k \leq n-1 : S_k = \inf_{k \leq j \leq n} S_j\}$ must be at least X_{i-1}. This is because each new tree is marked by a new global minimum in the random walk $(S_n)_{n=0}^{\infty}$. In particular, the kth tree ends where the random walk first visits the state $-k$. Thus, H_n coincides with $h_{\mathbf{t}_i}$ for $X_{i-1} \leq k < X_i$. From here, applying Theorems 1.6 and 1.10 complete the proof. $\square$

Here is the main theorem.

Theorem 3.11. $(H_{\lfloor nt \rfloor}/\sqrt{n} : t \geq 0) \xrightarrow{\mathcal{L}} (2Z(t)/\sigma : t \geq 0)$ *as* $n \to \infty$*, where* σ^2 *is the variance of* μ*, and* $(Z(t) : t \geq 0)$ *is a reflected Brownian motion. The convergence occurs in* $D[0, \infty)$.

Remark. Reflected Brownian motion is $B(t) - \inf_{0 \leq s \leq t} B(s)$ for each $t \geq 0$, where $(B(t) : t \geq 0)$ is standard one dimensional Brownian motion. It has been study as far back as Lévy, and it is known to be distributed as $|B(t)|$.

Much of the heavy lifting in the proof of Theorem 3.11 is done by a couple of technical lemmas about random walks and a nice concentration inequality for the height process. We separate these pieces into their own pieces and then quickly explain why this completes the proof at the end. There exists proofs for the statement in full generality [Ald93], but they are not fully optimized to be able to present in a reasonable amount of time. For this proof we make one simplifying assumption that allows for the proving of the aforementioned concentration inequality we need. We assume that there is some $t > 0$ such that $\sum_{k \geq 0} \exp(tk)\mu(k) < \infty$, i.e., we assume that the moment generating function exists on some interval in the postive reals.

A few new pieces of notation need to be introduced before continuing. For the random walk defined in Lemma 3.10, $M_n := \sup_{0 \leq k \leq n} S_k$ and $I_n := \inf_{0 \leq k \leq n} S_k$. For all $n \geq 0$, we define the time reversed random walk starting from n by $\hat{S}^n_k := S_n - S_{n-k}$ for all $0 \leq k \leq n$. The duality principle for random walks asserts that $(\hat{S}^n_k : 0 \leq k \leq n) \stackrel{\mathcal{L}}{=} (S_k : 0 \leq k \leq n)$. For any sequence $x = (x_n)_{n=0}^m$ (m can be ∞), we define

$$\Phi_n(x) = \left| \left\{ 1 \leq k \leq n : x_k = \sup_{0 \leq j \leq k} x_j \right\} \right|.$$

Note that we do not count $k = 0$ in the size of the set. We can rewrite our expression for H_n in terms of our new notation.

Lemma 3.12. $H_n = \Phi_n(\hat{S}^n)$ *for all* $n \geq 0$.

Proof. Indeed,

$$\begin{aligned} S_k = \inf_{k \leq j \leq n} S_j &\iff S_n - S_k = S_n - \inf_{k \leq j \leq n} S_j \\ &\iff \hat{S}^n_k = \sup_{k \leq j \leq n} \left(S_n - S_{n-(n-j)}\right) \\ &\iff \hat{S}^n_k = \sup_{0 \leq j \leq n-k} \hat{S}^n_j. \end{aligned}$$

Thus, the cardinalities defining both functions (using the definition from Lemma 3.10) are the same. $\square$

Lemma 3.13. *Let $(\tau_n)_{n=0}^{\infty}$ be a sequence of stopping times defined inductively by setting $\tau_0 = 0$ and $\tau_j = \inf\{n > \tau_{j-1} : S_n = M_n\}$ for all $j > 0$. The sequence random variables $(S_{\tau_j} - S_{\tau_{j-1}})_{j=1}^{\infty}$ are i.i.d. with distribution given by*

$$\mathbf{P}(S_{\tau_1} - S_{\tau_0} = k) = \nu[k, \infty) = \mu[k+1, \infty)$$

for all $k \geq 0$.

Proof. The independence property is and immediate consequence of the Markov property. Let $R = \inf\{n \geq 1 : S_n = 0\}$ and let $k \in \mathbb{Z}$. Let $(\sigma_n)_{n=0}^{\infty}$ be the times at which the random walk is at either the state 0 or state k. The sequence $(S_{\sigma_n})_{n=0}^{\infty}$ is a symmetric Markov chain on the state space $\{0, k\}$. In particular, $\mathbf{E}T_{0,0} = 2$ as the stationary distribution is uniform. Hence, we expect to visit k once before returning to 0. Altogether, this shows that

$$\mathbf{E}\left[\sum_{n=0}^{R-1} \mathbf{1}_{\{S_n = k\}}\right] = 1. \tag{3}$$

Now, note that $\tau_1 \leq R$. If $S_1 > 0$, then $\tau_1 = 1$, and $R > 1$. If $S_1 < 0$, then τ_1 is the first time that the random walk is ≥ 0, which contains the event that the random walk returns to the origin. Moreover, since negative jumps of the walk are at most -1, the portion of the random walk on (τ_1, R) is all positive integers and the portion on $(1, \tau_1)$ is all negative integers. In particular, if $k \leq 0$, then by (3),

$$\mathbf{E}\left[\sum_{n=0}^{\tau_1 - 1} f(S_n)\right] = \sum_{i=0}^{\infty} g(-i)\mathbf{E}\left[\sum_{n=0}^{R-1} \mathbf{1}_{\{S_n = -i\}}\right] = \sum_{i=0}^{\infty} f(-i) \tag{4}$$

for any function $f : \mathbb{Z} \to \mathbb{Z}_{\geq}$. Continuing,

$$\begin{aligned}
\mathbf{E}[f(S_{\tau_1})] &= \sum_{n=0}^{\infty} \mathbf{E}\left[f(S_{n+1})\mathbf{1}_{\{n < \tau_1\} \cap \{S_{n+1} \geq 0\}}\right] \\
&= \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \mathbf{E}\left[f(S_n + j)\nu(j)\mathbf{1}_{\{n < \tau_1\} \cap \{S_n + j \geq 0\}}\right] \\
&= \sum_{j=0}^{\infty} \nu(j)\mathbf{E}\left[\sum_{n=0}^{\infty} f(S_n + j)\mathbf{1}_{\{n < \tau_1\} \cap \{S_n + j \geq 0\}}\right] \\
&= \sum_{j=0}^{\infty} \nu(j)\mathbf{E}\left[\sum_{n=0}^{\tau_1 - 1} f(S_n + j)\mathbf{1}_{\{S_n + j \geq 0\}}\right] \\
&= \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \nu(j) f(j - i)\mathbf{1}_{\{j \geq i\}} && \text{(by (4))}
\end{aligned}$$

$$= \sum_{m=0}^{\infty} \sum_{\ell=m}^{\infty} f(m)\nu(\ell).$$

From here, just take f to be the indicator that $S_{\tau_1} = k$. □

With this, we can prove a key part of the proof of Theorem 3.11.

Lemma 3.14.

$$\frac{H_n}{S_n - I_n} \xrightarrow{\mathbb{P}} \frac{2}{\sigma^2}$$

as $n \to \infty$.

Proof. Let the sequence $(\tau_n)_{n=0}^{\infty}$ be defined as above. As ν has mean zero,

$$\mathbf{E}[S_{\tau_1}] = \sum_{k=0}^{\infty} k\nu[k, \infty) = \sum_{j=0}^{\infty} \frac{j(j+1)}{2}\nu(j) = \sigma^2/2.$$

Moreover,

$$M_n = \sum_{k:\tau_k \le n} (S_{\tau_k} - S_{\tau_{k-1}}) = \sum_{k=1}^{\Phi_n(S)} (S_{\tau_k} - S_{\tau_{k-1}}).$$

By Lemma 3.13 and the law of large numbers, $M_n/\Phi_n(S) \xrightarrow{\text{a.s.}} \sigma^2/2$ as $n \to \infty$ ($\Phi_n(S) \to \infty$ almost surely as $n \to \infty$ by null recurrence). Using Lemma 3.12 and the duality principle, we have that $(M_n, \Phi_n(S)) \stackrel{\mathcal{L}}{=} (S_n - I_n, H_n)$ for all $n \ge 0$. Hence,

$$\frac{S_n - I_n}{H_n} \xrightarrow{\mathbb{P}} \frac{\sigma^2}{2}$$

as $n \to \infty$. □

Now we turn our attention to the issue of concentration. In the proof of Theorem 3.11 we use a stronger result than just the law of large numbers convergence from the previous proof. Given the previous two results, the proof is not too different from that for most standard concentration inequalities in probabilistic combinatorics. A full proof can be found in [LG05], we shall just record the result and move on.

Lemma 3.15. *For any* $\epsilon \in (0, 1/4)$ *there exists a* $\delta > 0$ *and an* $N \ge 1$ *such that for all* $n \ge N$ *and all* $0 \le j \le n$,

$$\mathbf{P}\left(\left|M_j - \frac{\sigma^2}{2}K_j\right| \ge n^{1/4+\epsilon}\right) \le e^{-n^{\delta}}.$$

We are now ready to prove Theorem 3.11.

Theorem. $(H_{\lfloor nt \rfloor}/\sqrt{n} : t \ge 0) \xrightarrow{\mathcal{L}} (2Z(t)/\sigma : t \ge 0)$ *as* $n \to \infty$, *where* σ^2 *is the variance of* μ, *and* $(Z(t) : t \ge 0)$ *is a reflected Brownian motion. The convergence occurs in* $D[0, \infty)$ *with its associated metric.*

Proof. Most of the tough computations were done in the above lemmas. We just need to carefully go through and check that all of the convergences line up in the right way.

Step 1: (The function $\varphi : D[0,A] \to D[0,A]$ defined by $\varphi(f)(t) = \sup_{0\le s\le t} f(s)$ is continuous with respect to the Skorokhod topology) Suppose that x, y are such that $d(x,y) < \delta$ and without loss of generality assume that there is no dilation (of course, we could just redefine y to be λy). Let $t \in [0,A]$ and suppose without loss of generality that $\sup_{0\le s\le t} x(s) \ge \sup_{0\le s\le t} y(s)$. Let $(s_k)_{k=1}^{\infty}$ be such that $x(s_k) \to \sup_{0\le s\le t} x(s)$. We have for large k that $\delta \le y(s_k) \le x(s_k)$. By compactness, we may take some subsequence $(s_{k_m})_{m=1}^{\infty}$ such that $y(s_{k_m}) \to \alpha^*$ for some α^*. Then, it must hold that $\sup_{0\le s\le t} x(s) - \delta \le \alpha^* \le \sup_{0\le s\le t} y(s) \le \sup_{0\le s\le t} x(s)$. Since t was chosen arbitrarily the result follows.

Step 1, Donsker's Theorem, and the continuous mapping theorem combine to give that

$$\left(\frac{1}{\sqrt{n}}(S_{\lfloor nt\rfloor} - I_{\lfloor nt\rfloor}) : t \ge 0\right) \xrightarrow{\mathcal{L}} \left(\sigma(B(t) - \inf_{0\le s\le t} B(s)) : t \ge 0\right)$$

as $n \to \infty$ in $D[0,\infty)$ (recall that convergence in $D[0,\infty)$ is equivalent to convergence in $D[0,A]$ for all values of A).

Step 2: (Turning $S - I$ into H) Recall from the proof of Lemma 3.14 that $(S_n - I_n, H_n) \stackrel{\mathcal{L}}{=} (M_n, \Phi_n(S))$. Thus, Lemma 3.15 implies that for all $0 \le j \le n$ for n large that

$$\mathbf{P}\left(\left|S_j - I_j - \frac{\sigma^2}{2}H_j\right| > n^{3/8}\right) \le e^{-n^{\epsilon'}}$$

for some $\epsilon' > 0$. An application of the union bound gives

$$\mathbf{P}\left(\sup_{0\le j\le n}\left|S_j - I_j - \frac{\sigma^2}{2}H_j\right| > n^{3/8}\right) \le ne^{-n^{\epsilon'}}.$$

We can easily extend the event to the continuous height function on the interval $[0,A]$,

$$\mathbf{P}\left(\sup_{0\le t\le A}\left|S_{\lfloor nt\rfloor} - I_{\lfloor nt\rfloor} - \frac{\sigma^2}{2}H_{\lfloor nt\rfloor}\right| > (An)^{3/8}\right) \le Ane^{-(An)^{\epsilon'}}.$$

Summing and applying the Borel-Cantelli lemma we get that

$$\sup_{0\le t\le A}\left|\frac{S_{\lfloor nt\rfloor} - I_{\lfloor nt\rfloor}}{\sqrt{n}} - \frac{H_{\lfloor nt\rfloor}}{\sqrt{n}}\right| \xrightarrow{\text{a.s.}} 0$$

as $n \to \infty$. Combining this with the conclusion after step 1 yields the final result. $\square$

3.4 Convergence of the contour process

Towards the goal of proving convergence in the Gromov-Hausdorff topology, we would also like to say something about the convergence of contour functions for trees (and forests). Luckily this follows quite easily from the convergence for the height process. In this subsection, we give a contour function analogue for Theorem 3.11.

If we want to make a contour process for a sequence of independent Bienaymé(μ) trees $(T_n)_{n=1}^{\infty}$ then we need to deal with the fact that the contour function for the tree $\{\emptyset\}$ is trivial. Recall that the contour function $\gamma_{\mathbf{t}}$ for a tree $\mathbf{t}$ is defined on the interval $[0, 2(|\mathbf{t}|-1)]$. We define a new contour function $\gamma'_{\mathbf{t}}$ by $\gamma'_{\mathbf{t}}(t) = \gamma(t)\mathbf{1}_{\{t\in[0,2(|\mathbf{t}|-1)]\}}$. We define the contour process $(\Gamma(t) : t \geq 0)$ by concatenating the functions $(\gamma'_{T_n})_{n=1}^{\infty}$. For all $n \geq 0$ define $J_n = 2n - H_n + I_n$, where we recall that $I_n = \sup_{0\leq k\leq n} S_k$.

Lemma 3.16. *Let $(T_n)_{n=1}^{\infty}$ be a sequence of independent Bienaymé(μ) trees with $(U_n)_{n=0}^{\infty}$ being the vertices written in order (the ordering is the one obtained from making the root of T_{n+1} larger than every vertex of T_n for all $n \geq 1$). Then, over the interval $[J_n, J_{n+1}]$ the contour process goes from U_n to U_{n+1}. Moreover,*

$$\sup_{t\in[J_n,J_{n+1}]} |\Gamma(t) - H_n| \leq |H_{n+1} - H_n| + 1.$$

Proof. There are three possible cases (proof by look at Figure 4):

(i) u_{n+1} is a child of u_n;

(ii) u_{n+1} is a child of ancestor of u_n;

(iii) u_{n+1} is the root of the next tree in the sequence.

It is pretty simple to verify both the first and the second statements by induction by splitting them into these cases. □

Theorem 3.17. *If $(\Gamma(t) : t \geq 0)$ is the contour process for a sequence of Bienaymé(μ) random forests, then*

$$\left(\frac{1}{\sqrt{n}}\Gamma(2nt) : t \geq 0\right) \xrightarrow{\mathcal{L}} \left(\frac{2}{\sigma}B(t) : t \geq 0\right)$$

in $D[0,\infty)$ as $n \to \infty$, where $(B(t) : t \geq 0)$ is standard Brownian motion.

Proof. Let $A > 0$. Let $\varphi : [0,\infty) \to \mathbb{N}$ be a random function defined by $\varphi(t) = \sum_{n=0}^{\infty} n\mathbf{1}_{\{t\in[J_n,J_{n+1})\}}$. By Lemma 3.16 and Theorem 3.11,

$$\sup_{t\leq A}\left|\frac{1}{\sqrt{n}}\Gamma(2nt) - \frac{1}{\sqrt{n}}H_{\varphi(2nt)}\right| \leq \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sup_{t\leq A}|H_{\lfloor nt\rfloor+1} - H_{\lfloor nt\rfloor}| \xrightarrow{\mathbb{P}} 0$$

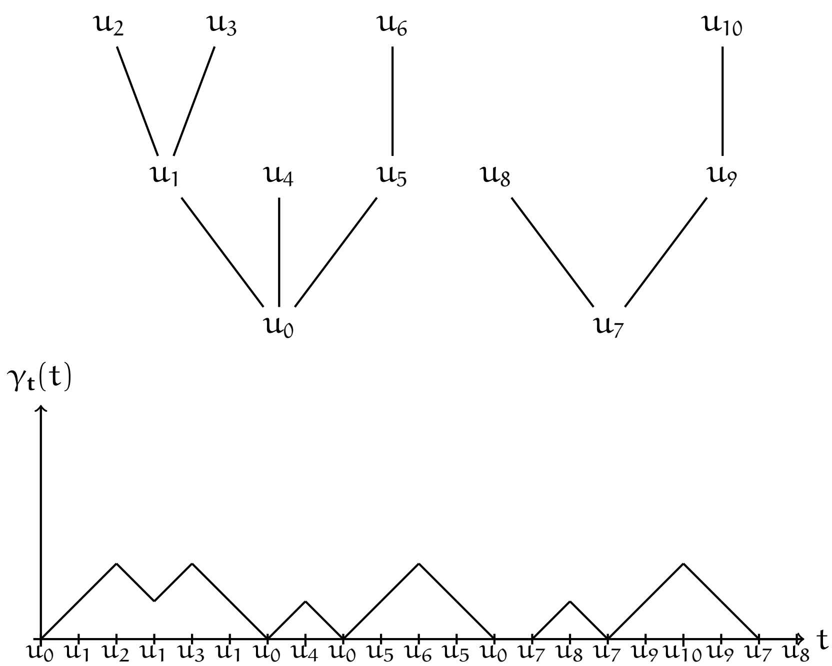

Figure 4: The first two trees in a realization of a Bienaymé forest along with the contour process $\Gamma(t)$ for the first two trees.

as $n \to \infty$. From the definition of the intervals $(J_n)_{n=0}^{\infty}$ and the continuity of the supremum/infimum from the proof of Theorem 3.11,

$$\frac{1}{n} \sup_{0 \le t \le A} |\varphi(2nt) - nt| \le \frac{1}{n} \sup_{0 \le k \le 2An} H_k + \frac{1}{n}|I_{2An}| + \frac{2}{n} \xrightarrow{\mathbb{P}} 0$$

as $n \to \infty$. Combining the two inequalities and applying Theorem 3.11 one more time we arrive at the final result. □

3.5 Aldous' Theorem

We are ready to turn our attention to combinatorial trees again and prove our first scaling limit theorem for random trees. Specifically, we identify a universal limit for conditioned Bienaymeé trees. The universal limit is known as the Brownian continuum random tree.

Definition 3.18. *Let $(e(t) : t \in [0,1])$ be a Brownian excursion. Extend the function to $[0,\infty)$ by defining $e(t) = 0$ for $t > 1$. The random metric space $\mathbf{T}_e$ is called the Brownian continuum random tree (CRT).*

We shall learn about the CRT as we continue to develop the theory of scaling limits (specifically, Section 4 offers a lot of insight into the structure of the tree), though for the moment it's, main importance is that it is the limit in the following theorem.

Theorem 3.19 (Aldous' Theorem). *Let $\mathbf{T}_n \overset{\mathcal{L}}{=}$ Bienaymé(n, μ) be a non-trivial critical Bienaymé tree considered as a real tree with edge lengths $(2n)^{-1/2}$ (take the tree encoded by the contour function and then scale the edge lengths). If μ has finite variance, then $\mathbf{T}_n \xrightarrow{\mathcal{L}} \mathbf{T}_e$ as $n \to \infty$ in the space $(\mathbb{T}, d_{GH})$.*

As one can imagine from the work done above, the convergence is essentially a corollary of a functional convergence result for the height/contour functions.

Theorem 3.20. *Let $T_n \overset{\mathcal{L}}{=}$ Bienaymé(n, μ) be a non-trivial critical Bienaymé tree, and let σ^2 be the variance of μ. Let $(H_k^{(n)})_{k=1}^n$ be the height process for T_n for each $n \geq 1$. Then,*

$$\left(\frac{1}{\sqrt{n}} H^{(n)}_{\lfloor nt \rfloor} : 0 \leq t \leq 1\right) \xrightarrow{\mathcal{L}} \left(\frac{2}{\sigma} e(t) : 0 \leq t \leq 1\right)$$

as $n \to \infty$ in $D[0,1]$. $(e(t) : 0 \leq t \leq 1)$ is a normalized length 1 Brownian excursion.

The proof builds on the work done on Theorem 3.11, though some additional effort is needed to address the fact that the trees have a fixed size. This change removes the independence between each jump in the walk and breaks the ability to apply Donsker's Theorem. Because of this, we need a version of Donsker's Theorem for discrete excursions.

Lemma 3.21. *Let $(\xi_k)_{k=1}^\infty$ be a sequence of i.i.d. random variables with mean 0 and variance 1 and let $S_k = \sum_{i=1}^k \xi_i$ for all $k \geq 0$. Let $\tau = \inf\{k \geq 1 : S_k \leq 0\}$. Let $(S_k^*)_{k=0}^\infty$ be distributed like S_k under $\mathbf{P}(\cdot | \tau = n)$, i.e., $\mathbf{P}(S_k^* = j) = \mathbf{P}(S_n = j | \tau = n)$ for all $k \geq 0$.*

$$\left(\frac{1}{\sqrt{n}} S^*_{\lfloor nt \rfloor} : 0 \leq t \leq 1\right) \xrightarrow{\mathcal{L}} (e(t) : 0 \leq t \leq 1)$$

as $n \to \infty$ in $D[0,1]$.

The proof of this lemma follows a similar structure to the proof of the original, and was developed over many papers in the 1970's [Bel72, Kai75, Kai76]. If I have time later I might try to fill this proof in, but for now I'm going to skip past it.

Proof of Theorem 3.19. We shall deal only with the convergence of the height process for the trees $(T_n)_{n=1}^\infty$, noting that converting the result to be about the contour function follows the exact same structure as the conversion of Theorem 3.11 provided in Theorem 3.17. Let $T \overset{\mathcal{L}}{=}$ Bienaymé(μ) be an unconditioned tree and let $(S_n)_{n=0}^\infty$ and $(H_n)_{n=0}^{|T|-1}$ be its corresponding DFQP and height process. From the local limit theorem for simple random walks we have that

$$\lim_{n\to\infty} \sup_x \left| \sqrt{2\pi n}\sigma \mathbf{P}(S_n = x) - e^{-\frac{x^2}{2n\sigma^2}} \right|.$$

Then, using the cycle lemma for simple random walks,

$$\mathbf{P}(|T| = n) = \mathbf{P}(S_0 \geq 0, ..., S_{n-1} \geq 0, S_n = -1) = \frac{1}{n}\mathbf{P}(S_n = -1) \sim \frac{1}{\sigma\sqrt{2\pi n^3}}.$$

In proving Theorem 3.11, we proved that

$$\mathbf{P}\left(\sup_{0\leq t\leq 1}\left|\frac{H_{\lfloor nt\rfloor}}{\sqrt{n}} - \frac{2(S_{\lfloor nt\rfloor} - I_{\lfloor nt\rfloor})}{\sigma^2\sqrt{n}}\right| > n^{-1/8}\right) \leq ne^{-n^{\epsilon}}$$

for some $\epsilon > 0$. As $\mathbf{P}(|T| = n)$ is polynomial in n we can condition without changing much,

$$\mathbf{P}\left(\sup_{0\leq t\leq 1}\left|\frac{H_{\lfloor nt\rfloor}}{\sqrt{n}} - \frac{2(S_{\lfloor nt\rfloor} - I_{\lfloor nt\rfloor})}{\sigma^2\sqrt{n}}\right| > n^{-1/8} \;\middle|\; |T| = n\right) \leq O\left(n^{5/2}e^{-n^{\epsilon}}\right).$$

Recalling the continuity of the supremum and infimum with respect to the Skorokhod topology and applying Lemma 3.21 we get that $(H^{(n)}_{\lfloor nt\rfloor} : 0 \leq t \leq 1) \xrightarrow{\mathcal{L}} (e(t) : 0 \leq t \leq 1)$ in the space $D[0,1]$, where we are defining $(H^{(n)}_n)_{n=0}^{n-1}$ to be the height process for T_n. □

A remark on the height of the Brownian CRT

A nice corollary of Aldous' Theorem is that the height of critical Bienaymé trees scaled by $1/\sqrt{n}$ converges to the height of the Brownian CRT. This naturally leads one to wonder what the height of the Brownian CRT is. Recall that the root of $\mathbf{T}_e$ is the equivalence class $[0]_{R_e}$. Since $e(0) = 0$, this implies that $\mathrm{ht}(\mathbf{T}_e) = \sup_{0\leq t\leq 1} e(t)$. That is a pretty clean description, but studying $\sup_{0\leq t\leq 1} e(t)$ is far from an easy job. For example, the diameter (which is closely related to the height), has a probability density given by

$$f(y) = \frac{\sqrt{2\pi}}{3}\sum_{n\geq 1}\left(\frac{64}{y^2}\left(4b_{n,y}^4 - 36b_{n,y}^3 + 75b_{n,y}^2 - 30b_{n,y}\right) + \frac{16}{y^2}\left(2b_{n,y}^3 - 5b_{n,y}^2\right)\right)e^{-b_{n,y}},$$

where $b_{n,y} = \frac{8\pi^2 n^2}{y^2}$ [Sze06, Wan15]. It can be proven either by using the Brownian CRT's close relationship with combinatorial trees or via direct analysis of the the supremeum of Brownian excursions.

4 The line-breaking construction of the CRT

In this section we cover an alternative perspective from which one can prove Aldous' Theorem in the specific case that the tree is distributed uniformly over all trees of size n. Rather than using functional encodings, this proof is algorithmic, providing a more tangible way to both generate and study the Brownian CRT. The result is stronger than Aldous' Theorem, as the convergence is with respect to a generalization of the Gromov-Hausdorff distance from Section 2. Most of the material from this section was initially explored in [Ald90, Ald91].

4.1 The Gromov-Hausdorff-Prokhorov distance

In this section we discuss a stronger form of convergence than the Gromov-Hausdorff distance we considered before. Specifically, it is a generalization of the distance to the space of measured metric spaces. Essentially, it just meant to encode both weak convergence of measure and metric space. It will appear quite daunting, but just keep in mind its reason for existence. First let's cover the Prokhorov metric (are you tired of new metrics yet?).

Definition 4.1. *For two measures μ and ν defined on a metric space (X, d), we define the Prokhorov distance between them to be*

$$d_P = \inf\{\epsilon > 0 : \mu(A) \leq \nu(A^\epsilon) + \epsilon, \nu(A) \leq \mu(A^\epsilon) + \epsilon \;\forall A \in \mathcal{B}(X)\},$$

where $\mathcal{B}(X)$ are the Borel-measurable sets and $S^\epsilon = \{x \in X : d(x, S) \leq \epsilon\}$.

The next lemma explains why we introduced another metric. Using the Portmanteau theorem it is not too hard to prove (see e.g., [Bil13]).

Lemma 4.2. *Let $M(X)$ be the space of finite measures on some complete separable metric space (X, d). Let $(\mu_n)_{n=1}^\infty$ be a sequence in $M(X)$ and let $\mu \in M(X)$. Then $\mu_n \xrightarrow{\mathcal{L}} \mu$ if and only if $d_P(\mu_n, \mu) \to 0$ as $n \to \infty$, i.e., d_P is a metrization of the topology of weak convergence.*

Now we can formally introduce the idea of Gromov-Hausdorff-Prokhorov convergence (GHP). The Prokhorov distance is needed in order to have a well define GHP metric that encapsulates what we want it to.

Definition 4.3. *Let (X, d_X, μ) and (Y, d_Y, ν) be compact measured metric spaces with roots ρ_X and ρ_Y, where $\mu \in M(X)$ and $\nu \in M(Y)$. We define the Gromov-Hausdorff-Prokhorov distance between X and Y as*

$$d_{GHP}(X, Y) = \inf_{\varphi_X, \varphi_Y} \left(\max\{d_H(\varphi_X(X), \varphi_Y(Y)), d^*(\varphi_X(\rho_X), \varphi_Y(\rho_Y)), d_P(\varphi_X\mu, \varphi_Y\nu)\}\right),$$

where the infimum is taken over all isometric embeddings φ_X and φ_Y into a metric space (X^*, d^*).

The reason that convergence in the GHP topology is preferred in many cases is quite simple - measures on metric spaces can tell us information about the space. We are going to try to keep rigorous discussion of the GHP topology to a minimum and refer the reader to [ADH13] if they are interested in the finer details. We'll record below that the metric induces good metric spaces as was done with the GH metric. There is an equivalent definition of GHP distance that induces the same topology and uses the notion of correspondances instead of isometries that one can find in [ABBGM17]. For the next theorem, we note that two measured compact metric spaces (X, d, μ) and (Y, d, ν) are called *equivalent* if there is an isometry $\varphi : X \to Y$ such that $\varphi\mu = \nu$.

Theorem 4.4. *If $\mathbb{K}_M$ is the space of equivalence classes of measure compact metric spaces, then $(\mathbb{K}_M, d_{GHP})$ is a complete separable metric space as well. $\mathbb{T}_M$, the space of equivalence classes of measured real trees, is a closed set in $\mathbb{K}_M$*

4.2 Growing random trees (the set up)

Consider the following combinatorial tree growth algorithm for a fixed $n \geq 2$ and vertex set $V_n$labelled in $[n]$ and edge set E_n:

(i) For $2 \leq i \leq n$, let $V_i = \min\{U_i, i-1\}$, where $U_i \stackrel{\mathcal{L}}{=} \mathrm{Unif}(\{1, ..., n\})$. Add the edge $\{i, V_i\}$ to E_n.

(ii) Remove the labels from V_n and then give them a new label distributed uniformly over all bijections $V_n \to [n]$.

(iii) output $T_n = (V_n, E_n)$.

As the technique deviates a bit from those in this course, we won't prove the following theorem that can be found in [Ald90]. The proof relies on connections between uniform spanning trees of graphs and random walks on them.

Theorem 4.5. *Let T_n be the output from the above algorithm. Then, T_n is distributed like a uniform lablled tree on n vertices.*

Now, consider the following real tree growth "algorithm" that is done within the space ℓ_1 of absolutely summable sequences with the metric $d(x, y) = \sum_{n=1}^{\infty} |x_n - y_n|$ (the quotations are there because it is an infinite algorithm). Let $C_0 = L_0 = 0$, define $(C_n)_{n=1}^{\infty}$ to be the arrival times of a non-homogenous poisson process with rate $r(t) = t$, and define $(L_n)_{n=1}^{\infty}$ by $L_n = C_n \cdot \xi_n$, where ξ_n is a Unif$[0, 1]$ random variable. Then define $f(0) = 0$ and

$$f(x) = \sum_{n=1}^{\infty} (f(L_{i-1}) + (x - C_{i-1})z_i)\mathbf{1}_{\{C_{i-1} < x \leq C_i\}},$$

where z_i is the basis vector in ℓ_1 in the canonical orthonormal basis for the space that has the ith entry non-zero. Then, we set $\mathcal{T}_x = f([0,x])$ for all $x \geq 0$, $\mathcal{T}_\infty = f([0,\infty))$, and $\mathcal{T} = \overline{\mathcal{T}_\infty}$ (the closure). These two procedures are the main stars of this section. Let's present what we prove.

Theorem 4.6. *Let $(\mu_t : t \geq 0)$ be defined by setting $\mu_t = f\lambda_{[0,t]}$ for all $t \geq 0$ ($\lambda_{[0,t]}$ is the Lebesgue measure on $[0,t]$ and f is the function from the continuum algorithm). Then, almost surely, there is some measure $\mu_\mathcal{T}$ such that $\mu_t \to \mu_\mathcal{T}$ as $t \to \infty$ pointwise. and $\mu_\mathcal{T}(\mathcal{T} \setminus \mathcal{T}_\infty) = 1$.*

Theorem 4.6 is not just a fun result. It plays a large role in the proof the following theorem.

Theorem 4.7. *Let $(T_n)_{n=2}^\infty$ be defined as in the algorithm (viewed as a real tree), let $\mathcal{T}$ be defined as the continuum algorithm, and let μ_n be the uniform measure on vertices in T_n (viewed as a measure on the real tree version of T_n). Then, $(T_n, \mu_n) \xrightarrow{\mathcal{L}} (\mathcal{T}, \mu_\mathcal{T})$ as $n \to \infty$ with respect to the GHP topology. In particular, $\mathcal{T}$ is distributed like the Brownian CRT.*

Corollary 4.8. *The Brownian CRT has only vertices of degree one, two, and three almost surely. Its mass measure is supported almost surely on the set of leaves.*

The proof strongly utilizes the nice properties of Poisson procceses to poke at properties of the tree $\mathcal{T}$. In the previous section, continuous functions were our window into the the world of real trees. Here, Poisson processes can be seen as playing a similar role. We pay special attention to their usage in the proof. We are also quite careful when it comes time to discuss the measures on our trees as this is a new concept as well. Other details that feel more familiar we are a little more casual with.

4.3 Non-homogeneous Poisson processes

Crucial to the continuum algorithm is a non-homogeneous Poisson process on $[0,\infty)]$. It is worthwhile to review some basic facts about them. Fill in this section with cool material

4.4 The fractal dimension of the tree $\mathcal{T}$

The first thing to cover is the question of compactness for the tree $\mathcal{T}$. Our approach towards proving this also happens to proves that the Minkowski dimension of the tree is 2. For the rest of the section we shall concern ourselves mostly with the function

$$D(s,t) = \inf_{0 \leq r \leq s} \|f(t) - f(r)\|$$

whre $0 < s < t$. $D(s,t)$ is the distane between the point $f(t)$ and the tree $\mathcal{T}_s$.

Lemma 4.9. *For $0 < s < t$, $D(s,t)$ is stochastically dominated by an $\mathrm{Exp}(s)$ random variable. Moreover, there is a random function $(X(s) : s > 0)$ such that $D(s,t) \xrightarrow{\mathcal{L}} X(s)$ as $t \to \infty$ and $sX(s) \xrightarrow{\mathcal{L}} \mathrm{Exp}(1)$ as $s \to \infty$.*

Proof. Aldous does not prove either of these two facts fully in his original paper. I have been struggling to fill in the gaps. □

4.5 Convergence with respect to the GH topology

As $\mathcal{T}$ is defined via a limit as $t \to \infty$, and the uniform tree has a paramter $n \to \infty$ we need a technical lemma that takes control of the double limit.

Lemma 4.10. *Let (X, d) be a metric space. Let $(X_n(t) : t \geq 0)_{n=1}^{\infty}$ be a sequence of random functions $[0, \infty) \to X$ and let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables taking values in X. Suppose, for each fixed $t \geq 0$ there exists some random variable $Y(t)$ such that $X_n(t) \xrightarrow{\mathcal{L}} Y(t)$ as $n \to \infty$,and further that there is some random variable Y such that $Y(t) \xrightarrow{\mathcal{L}} Y$ as $t \to \infty$. If*

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbf{P}(d(X_n(t), X_n) \geq \epsilon) = 0$$

for all $\epsilon > 0$, then $X_n \xrightarrow{\mathcal{L}} Y$ as $n \to \infty$.

This result is also quite quick to prove using the Portmanteau lemma [Bil13] (sorry for skipping so much in this section, there is just a lot to cover). There are some more technical lemmas, but first let's set up the structures for the actual proof. In particular we need a "continuum" version of a uniform random tree, which is where the discrete algorithm comes in. Let $(U_i)_{i=2}^{n}$ be defined as in the discrete algorithm. We introduce a sequence of random variables $(J_i)_{i=2}^{n}$ with $J_1 = 1$ defined recursively as

$$J_i = (J_{i-1} + 1)\mathbf{1}_{\{U_i < i-1\}} + J_{i-1}\mathbf{1}_{\{U_i \geq i-1\}}.$$

Then, we define a new sequence $(V_i^*)_{i=2}^{n}$ with $V_1 = 0$ in ℓ_1 recursively as

$$V_i^* = V_{\min(U_i, i-1)}^* + z_{J_i}.$$

We define a geometric realization of the uniform random tree for all $t \geq 0$, denoted by $T_n(t)$ to be the smallest connected set in ℓ_1 containing the set

$$\left\{V_1^*, \ldots, V_{\lfloor t \rfloor \wedge n}^*\right\}.$$

We define $T_n := T_n(n)$. We define similarly to before, $D_n(i,j) = \min_{1 \leq m \leq i} \|V_j^* - V_m^*\|$. A result analogous tothe lemma about $D(s,t)$ can be proved about D_n. I similarly have no idea how one does this.

Lemma 4.11. *There exists some $K \geq 0$ and $C > 0$ such that for all $k \geq K$,*

$$\mathbf{P}\left(\max_{1 \leq m \leq n} D_n(\lfloor \sqrt{n} e^k \rfloor, m) \geq 6ke^{-k}\sqrt{n}\right) \leq Ce^{-k}.$$

Some work is needed to make the proof discrete rather than continuous, but we leave out the details and move forward to the result we have been aiming to prove in this subsection.

Theorem 4.12. $\frac{1}{\sqrt{n}} T_n \xrightarrow{\mathcal{L}} \mathcal{T}$ *as $n \to \infty$ with respect to the* GH *topology.*

Proof. According to Lemma 4.10, it suffices to show that $T_n(t) \xrightarrow{\mathcal{L}} \mathcal{T}_t$ as $n \to \infty$ for all $t \geq 0$, that $\mathcal{T}_t \xrightarrow{\mathcal{L}} \mathcal{T}$ as $t \to \infty$, and that

$$\lim_{t \to \infty} \limsup_{n \to \infty} \mathbf{P}(d(X_n(t), X_n) \geq \epsilon) = 0$$

for all $\epsilon > 0$. By the compactness of $\mathcal{T}$, the partial trees $\mathcal{T}_t$ converge to the tree $\mathcal{T}$ almost surely with respect to the Hausdorff distance. To see this, note that $\mathcal{T}_t$ is a subset of $\mathcal{T}$ and take some finite covering of $\mathcal{T}$ with ϵ-balls. Then, for some $t^* \geq 0$ the tree $\mathcal{T}_{t^*}$ contains the centre of each ball. Lemma 4.11 implies the third point. It is not immediate, but playing around with the two expressions a little bit reveals the truth quite quickly. The point that requires some real work to prove is the first.

For all $n \geq 1$ we construct a sequence $(C_j^n, B_j^n)_{j=1}^n$, where $C_j^n = x_j$ for $(x_j)_{j=1}^n$ an enumeration of the set $\{i \geq 0 : U_i < i - 1\}$ and $B_j^n = U_{x_j}$. As T_n and $\mathcal{T}$ are simply just deterministic functions of $(C_j^n, B_j^n)_{j=1}^n$ and $(C_j, B_j)_{j=1}^\infty$, to prove the desired convergence for point one, we only need to show that

$$\frac{1}{\sqrt{n}}(C_j^n, B_j^n)_{j=1}^n \xrightarrow{\mathcal{L}} (C_j, B_j)_{j=1}^\infty$$

as $n \to \infty$ in the product topology (for concerns of placing the two sequences in the same space, fill the rest of the first sequence with zeros after slot n). Since B_j^n can be generated by a function of C_j^n and an independent Unif$[0, 1]$ random variable ξ_j, it suffices by the continuous mapping theorme to only prove the result for the sequences containing only the C_j^n's and C_j's. Moreover, from the independence structure of the inter-arrival times of a Poisson process, the result will follow from just proving the first case and then applying induction (the general case is almost identical). After all of these reductions we just need to do a simple computation,

$$\begin{aligned} \mathbf{P}(C_1^n > \lfloor \alpha\sqrt{n} \rfloor) &= \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{\lfloor \alpha\sqrt{n} \rfloor}{n}\right) \\ &= \exp\left(\sum_{i=1}^{\lfloor \alpha\sqrt{n} \rfloor} \log\left(1 - \frac{i}{n}\right)\right) \end{aligned}$$

$$
\begin{aligned}
&= \exp\left(-\sum_{i=1}^{\lfloor\alpha\sqrt{n}\rfloor}\left(\frac{i}{n}+O\left(\frac{i^2}{n^2}\right)\right)\right)\\
&= \exp\left(-\frac{(\lfloor\alpha\sqrt{n})(\lfloor\alpha\sqrt{n}\rfloor-1)}{2n}+o(1)\right)\\
&= \exp\left(-\frac{\alpha^2}{2}+o(1)\right).
\end{aligned}
$$

Recalling the characterizations of non-homogeneous Poisson processes that we saw before, we can conclude the first point and conclude the whole proof. Fill in the computation for convergence. What about independent increments property? □

4.6 Measuring the tree $\mathcal{T}$

4.7 Completing Theorem 4.7

5 Scaling limits of components in critical random graphs

Utilising the ideas from the previous section about scaling limits for random trees, we explore how one can derive meaning scaling limits for very sparse random graphs. Specifically, we prove that the components in an Erdös-Rényi graph with parameter $p = n^{-1} + \lambda n^{-4/3}$ converge in Gromov-Hausdorff topology to a random graph that can be described as a random real tree encoded by an excursion that is "decorated" with a finite number of extra edges. This requires some background on the $G(n, p)$ that we begin with.

References

[ABBGM17] Louigi Addario-Berry, Nicolas Broutin, Christina Goldschmidt, and Grégory Miermont. The scaling limit of the minimum spanning tree of the complete graph. 2017.

[ADH13] Romain Abraham, Jean-François Delmas, and Patrick Hoscheit. A note on the gromov-hausdorff-prokhorov distance between (locally) compact metric measure spaces. 2013.

[Ald90] David J Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM Journal on Discrete Mathematics*, 3(4):450–465, 1990.

[Ald91] David Aldous. The Continuum Random Tree. I. *The Annals of Probability*, 19(1):1 – 28, 1991.

[Ald93] David Aldous. The continuum random tree iii. *The annals of probability*, pages 248–289, 1993.

[ANN04] Krishna B Athreya, Peter E Ney, and PE Ney. *Branching processes*. Courier Corporation, 2004.

[Bel72] Barry Belkin. An invariance principle for conditioned recurrent random walk attracted to a stable law. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 21(1):45–64, 1972.

[Bil13] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[Jan23] Svante Janson. Real trees, 2023.

[Kai75] WD Kaigh. A conditional local limit theorem for recurrent random walk. *The Annals of Probability*, pages 883–888, 1975.

[Kai76] William D Kaigh. An invariance principle for random walk conditioned by a late return to zero. *The Annals of Probability*, pages 115–121, 1976.

[LG05] Jean-François Le Gall. Random trees and applications. 2005.

[Sze06] George Szekeres. Distribution of labelled trees by diameter. In *Combinatorial Mathematics X: Proceedings of the Conference held in Adelaide, Australia, August 23–27, 1982*, pages 392–397. Springer, 2006.

[Wan15] Minmin Wang. Height and diameter of brownian tree. 2015.

Caelan Atamanchuk
email: caelan.atamanchuk@gmail.com
Department of Mathematics and Statistics
McGill University