Scaling limits of random trees and graphs

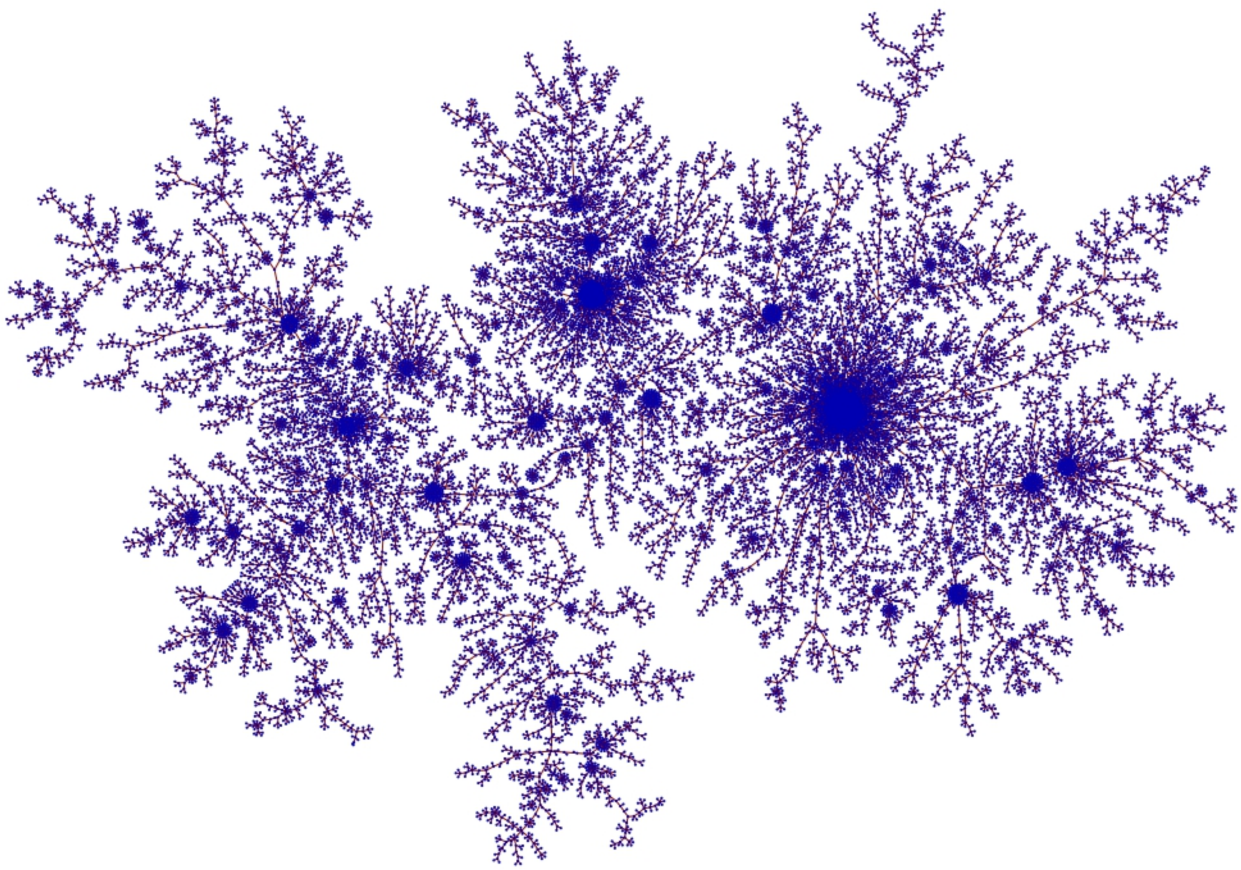

Figure 1: An image of a cool tree stolen from Igor Kortchemski's website.

Contents

Ramblings for people who stumble upon this file

The other day I woke up to a notifcation that google scholar added these notes to my profile. I guess this means people might actually end up reading these notes, so I think I should add some remarks about what these notes even are.

These notes were made for an informal course on scaling limits of random graphs at McGill in the winter 2025 semester (as it is winter 2025 right now, expect a lot of typos and half-baked ideas that were covered better in lecture than here). The intention of the course was to bring graduate students researching combinatorial probability theory up to speed with both the classical and modern work on scaling limits for random trees and grapsh. Focus was placed on introducing and proving the results from metric geometry and probability theory that pre-date the ideas of graph scaling limits and supported the emergence of it.

Much of the content from the first three sections was developed by expanding upon the excellent introduction to scaling limits provided in [LG05]. Afterwards, sections are usually dedicated to the content of a single paper that is mentioned at the beginning of the section.

Thank you to the many attendees of the course who gave me a reason to actually learn this material well enough to present it. Special thanks in particular go to my PhD supervisors Luc Devroye and Louigi Addario-Berry for helping me out with the preparation and presentation of the material.

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 $1 \le i \le k-1$ 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* $1 \le j \le c_{\mathbf{t}}(u)$.

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 $\le$.

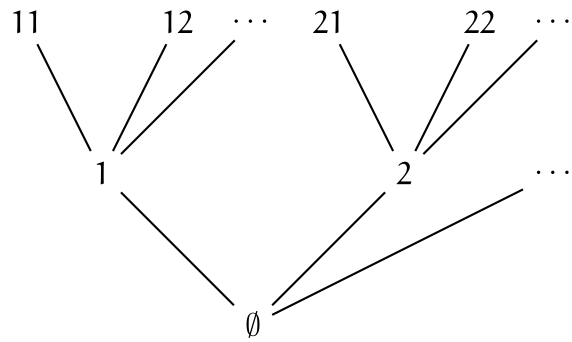

Figure 2: A depiction of the set **U** that highlights its tree structure.

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 $\varphi(\mathbf{t}_n) = (q_0, ..., q_n)$. □

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. When we deal with the contour function we often take an intuitive approach, arguing with pictures and words instead of dealing with the formal objects. This just helps us to avoid long detours with a lot of notation that end with us concluding relatively intuitive statements. Anyways, here is the definition.

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 $t > 2(n-1)$.

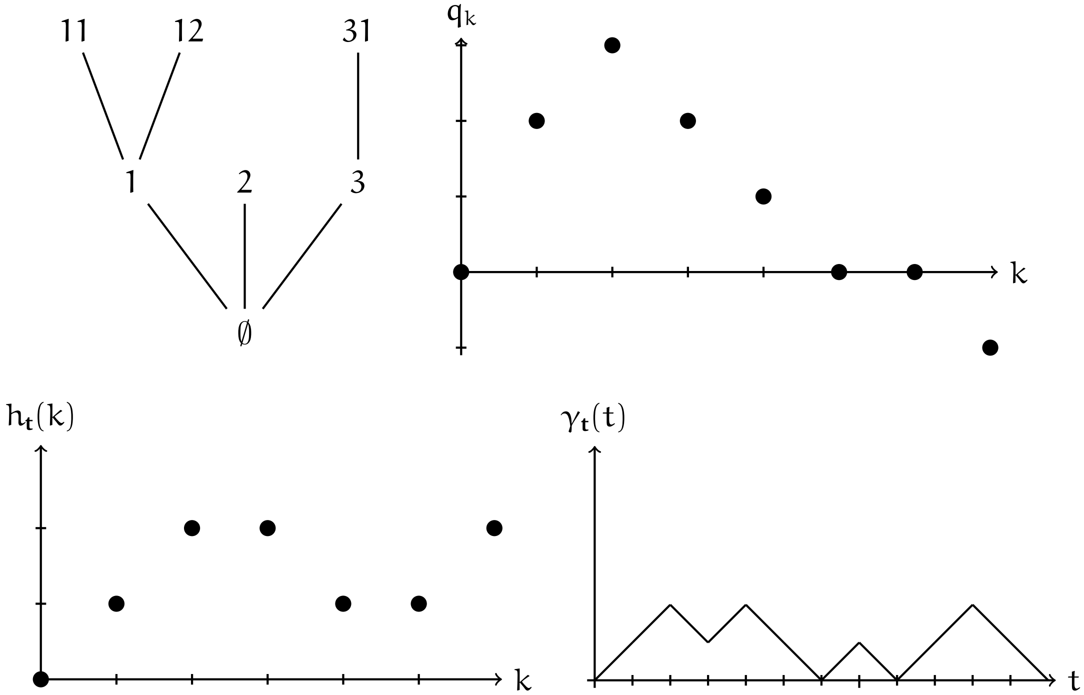

Figure 3: A tree and its many functional encodings

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

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

Let $\tau_j = \inf\{m \ge j : q_m < q_j\}$. 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 $\mathbf{t}(u_j) = \{u_m : j \le m \le \tau_j - 1\}$. (1) follows immediately from this identity. $\square$

A 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 $h_{\mathbf{t}}(k+1)-1$). I'm being a bit hand-wavy here, but the conclusion really is just that all three of the processes presented here uniquely determine our trees.

1.2 Bienaymé trees

Definition 1.7. *Let μ be a measure on $\mathbb{Z}_{\ge} = \{0, 1, 2, ...\}$ with $\sum_{k=0}^{\infty} k\mu(k) < \infty$ such that $\mu(1) \ne 1$. For all $u \in \mathbf{U}$, we associate an independent random variable $\xi_u \overset{\mathcal{L}}{=} \mu$. The subset $T = \{u = (u^1, ..., u^k) \in \mathbf{U} : u^j \le \xi_{(u^1,...,u^{j-1})}\ \forall\ 1 \le j \le k\}$ 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 $T \overset{\mathcal{L}}{=}$ Bienaymé(μ). Then,*

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

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

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

$\square$

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 $T \overset{\mathcal{L}}{=}$ 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,*

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

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

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

First, the set $T \cap \{v \in \mathbf{U} : v \leq u\}$, is measurable with respect to $\sigma(\xi_v : v < u)$. Then, for any $k \geq 0$, the event $\{U_k = u\} \cap \{|T| > k\}$, being completely determined by $T \cap \{v \in \mathbf{U} : v < u\}$, is measurable with respect to $\sigma(\xi_v : v < u)$. The set $\{U_k = u\} \cap \{|T| \leq k\}$ 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 $g_0, \ldots, g_k : \mathbb{Z}_\geq \to \mathbb{Z}_\geq$ be a collection of functions for $0 \leq k \leq |T| - 1$. Then,

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

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 $x \in (0, \tau]$, $T_n \stackrel{\mathcal{L}}{=} T_n^x$, 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 $T_n \stackrel{\mathcal{L}}{=}$ Bienaymé(n, μ).*

Proof. Let $T^* \stackrel{\mathcal{L}}{=}$ 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

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

$\square$

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 $(a_k = 1$ 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 $T \overset{\mathcal{L}}{=}$ Bienaymé(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 metric information with combinatorial trees, but erase 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 functions supported on a compact connected set. This 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 that they are in 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.

Lemma 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,and we take the following one to be our canonical definition. 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^*).

Proof. First, suppose that $d(\mathbf{T}_1, \mathbf{T}_2) < r$ for two trees $(\mathbf{T}_1, d_1)$ and $(\mathbf{T}_2, d_2)$ and let φ_1, φ_2 be isometric embeddings into a space (Z, d_Z) such that $d_H(\varphi_1 \mathbf{T}_1, \varphi_2 \mathbf{T}_2) < r$. We define a relation C by adding all pairs of vertices $(t_1, t_2) \in \mathbf{T}_1 \times \mathbf{T}_2$ such that $d_Z(\varphi_1(t_1), \varphi_2(t_2)) < r$. By the assumption at the beginning, C is a correspondance that with $\operatorname{dis}(C) < 2r$. To see this, consider two pairs of corresponding points (x_1, x_2)

and (y_1, y_2), and suppose that $d_1(x_1, y_1) \geq d_2(x_2, y_2)$. Then, a simple application of the triangle inequality gives

$$\begin{aligned}
&d_1(x_1, y_1) - d_2(x_2, y_2)\\
=&d_Z(\varphi_1 x_1, \varphi_1 y_1) - d_Z(\varphi_2 x_2, \varphi_2 y_2)\\
\leq&d_Z(\varphi_1 x_1, \varphi_2 x_2) + d_Z(\varphi_2 x_2, \varphi_1 y_1) - d_Z(\varphi_2 x_2, \varphi_2 y_2)\\
\leq&d_Z(\varphi_1 x_1, \varphi_2 x_2) + d_Z(\varphi_2 x_2, \varphi_1 y_2) + d_Z(\varphi_2 y_2, \varphi_1 y_1) - d_Z(\varphi_2 x_2, \varphi_2 y_2)\\
=&d_Z(\varphi_1 x_1, \varphi_2 x_2) + d_Z(\varphi_2 y_2, \varphi_1 y_1),
\end{aligned}$$

which is strictly below $2r$ by definition. Hence, we can conclude that $d_{GH} \leq d$. Now suppose that $\operatorname{dis}(C) = 2r$ for some correspondance C. Then, in the disjoint union of $\mathbf{T}_1$ and $\mathbf{T}_2$ (mark all the points in $\mathbf{T}_1$ with a zero and in $\mathbf{T}_2$ with a one and then take the union) we define a pseudometric

$$d^*(t_1, t_2) = \begin{cases} \inf_{(t_1', t_2') \in C} \Big(d_1(t_1, t_1') + d_2(t_2, t_2') + r \Big), \text{ if } t_1 \in \mathbf{T}_1, t_2 \in \mathbf{T}_2 \\ d_1(t_1, t_2), \text{ if } t_1, t_2 \in \mathbf{T}_1 \\ d_2(t_1, t_2), \text{ if } t_1, t_2 \in \mathbf{T}_2 \end{cases}.$$

Note that $d^*(t_1, t_2) = r$ when the two vertices correspond with each other. In particular, since every vertex has a partner in the correspondance (and the roots correspond), we have that $d_H(\mathbf{T}_1, \mathbf{T}_2) \leq r$. There are some issues with the fact that d^* is only a pseudometric, but simply modding out by the standad distance zero equivalence relation finishes the job. □

An important remark to make is that there was nothing special about the fact that our compact metric spaces of choice were trees in any of the proof of any of those definitions. One can extend the notion of Gromov-Hausdorff distance that we just provided to the set of all isometry classes of compact metric spaces. We will often make reference to this larger space containing $\mathbb{T}$ when working with real trees and especially when working with real graphs. We denote it by $\mathbb{K}$. The last thing to cover about Gromov-Hausdorff space before moving on to functional encodings is the question of completeness.

Theorem 2.6. *Both* $(\mathbb{K}, d_{GH})$ *and* $(\mathbb{T}, d_{GH})$ *are complete separable metric spaces.*

Proof sketch. Separability of $\mathbb{K}$ is not too hard to show with the correspondance definition of the Gromov-Hausdorff distance. Since our metric spaces are compact, we can find finite ϵ-covers of them for all $\epsilon > 0$. This implies that the set of finite metric spaces is dense in $\mathbb{K}$. If we take all finite metric spaces that have only rational distances, then we get a countable dense set. We can do a very similar thing for $\mathbb{T}$ by considering all real trees that branch out only finitely many times

I didn't quite have time to type up a full argument for this proof following what was done in class. Hopefully I can fill this in later when I have time.

□

Due mostly to time constraints we have not ventured very deep into the theory of Gromov-Hausdorff space, only presenting the results that are needed. I would just like to remark that this is not due to lack of relevance or because the connetions end with what has been discussed here. Deep knowledge of the theory of convergence for metric spaces and the surrounding material has and will continue to be important to developing the theory of graph scaling limits. I recommend taking a look at [Bur01] to learn more about the topic, it was my main source of deeper information about Gromov-Hausdorff convergence when preparing these notes. I also stole a couple ideas from [Pet06].

2.2 ENCODING REAL TREES WITH FUNCTIONS

In this subsection, we 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. First we set up our candidates for the encodings.

Let $f \in \{g : [0,\infty) \to [0,\infty) : \operatorname{supp}(f) \text{ compact and connected}, g(0) = 0\} := C_c^+[0,\infty)$. 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 $R_f = \{(s,t) \in [0,\infty) \times [0,\infty) : d_f(s,t) = 0\}$. Essentially, our main 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 tree related 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. It is relatively straightforward to show that $\mathbf{T}_f$ is in fact a compact metric space using uniform continuity of continuous functions over compact intervals, however we need to still show that they are real trees. In particulat, we would like the following to be true:

(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 $(\mathbf{T}_g, d_g)$, $d_{GH}(\mathbf{T}_f, \mathbf{T}_g) = \Theta(\|f - g\|_\infty)$.

(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 almost-linear functions (defined below) and then invoke the completeness of $(\mathbb{T}, d_{GH})$ to extend to all functions in $C_c^+[0,\infty)$. A different approach to prove the same results that argues directly with the definition of a real tree is covered in [LG05]. The third point is actually not relevant in these notes and so we won't prove it. However, for the sake of completeing the analogy with the results from the previous section we think it is worthy to mention that (iii) is also true. An excellent constructive proof can be found in [Duq06].

Let $f \in C_c^+[0,\infty)$. We say that f is a almost-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 almost-linear functions in $C_c^+[0\infty)$ with C_L. We begin by asserting that almost linear-function produce real trees. We can conclude this fact by observing that the metric spaces produced by almost-linear functions are essentially equivalent to combinatorial plane trees.

Lemma 2.7. *Let $f \in C_L$. Then, the function $\gamma : [0,\infty) \to [0,\infty)$ given by $\gamma(t) = \Delta^{-1} f(\epsilon t)$ is the contour function for some plane tree $\mathbf{t}_f$. Moreover, $(\mathbf{T}_f, d_f)$ is isometric to the real tree version of $\mathbf{t}_f$ with edge lengths Δ.*

We shall skip past proving Lemma 2.7 or producing a formal construction of the real tree version of $\mathbf{t}_f$ with edge lengths Δ, favouring an appeal to intuition (see figure below). The idea is essentially that, as we sketch out the contour function with our pencil and paper, we can graft on intervals of length Δ every time that we begin an up interval for the function. One other thing worth observing is that ϵ actually plays no role in the structure of $(\mathbf{T}_f, d_f)$. This is not an issue and makes sense for what we want our functional encodings to be. We can see straight from the definition of $\mathbf{T}_f$ that, if we define $g(x) = f(\alpha x)$ for any $\alpha > 0$, the mapping $x \mapsto \alpha x$ induces an isometry $\mathbf{T}_f \to \mathbf{T}_g$.

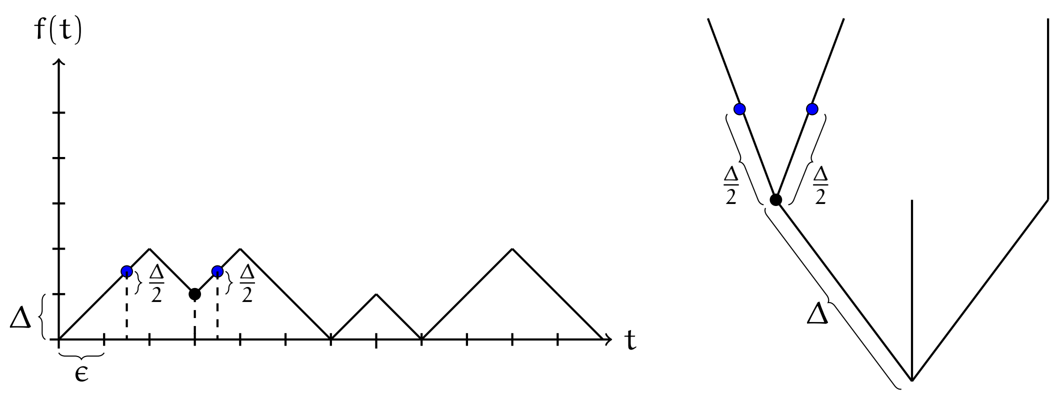

Figure 4: An almost-linear function and its corresponding real tree. Two points on the graph of the function are highlighted in blue, along with their corresponding vertices in the real tree to highlight how the distance d_f matches the natural extension of graph distance we get from sketching out the contour function. The greatest common ancestor of the points/vertices is black in both drawings.

Lemma 2.7 covers point (i) for the case of almost-linear functions. What is left to do is to argue that we can approximate all of the metric spaces for $C_c^+[0,\infty)$ via those generated by functions in C_L.

Lemma 2.8. 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 $S = \sup\operatorname{supp}(f)$ and $I = \inf\operatorname{supp}(f)$. 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 $n \to \infty$. □

Combining the previous lemmas we can conclude what we wanted to show.

Theorem 2.9. *The two claims stated at the beginning of the section hold.*

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

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

Proof. (i) can be proven using the correspondance definition of the Gromov-Hausdorff distance (we have not yet shown that the metric spaces are trees, but recall that we can define the GH-distance for any two compact metric spaces). Let

$$C = \big\{([x]_{R_f}, [y]_{R_g}) : \exists t \geq 0 \text{ such that } t \in [x]_{R_f},\ t \in [y]_{R_y}\big\}.$$

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, there exists $s, t \geq 0$ such that

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

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

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

We can easily prove (ii) using (i), the density of C_L in $C_c^+[0, \infty)$, and the fact that $\mathbb{T}$ is closed in $\mathbb{K}$. $\square$

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 $\pi_{t_1,\dots,t_k}\mu = \pi_{t_1,\dots,t_k}\nu$ 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 $n \to \infty$. $\square$

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 our purposes, we need a translation of tightness into more criteria that are more easily verified by computations. We can begin by deriving 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 $\epsilon > 0$, $\lim_{\delta \to 0} \limsup_{n \to \infty} \mu_n(\{x : w_x(\delta) \ge \epsilon\}) = 0$.*

Proof. Suppose that the sequence is tight. Choose some $K \subseteq C[0,1]$ and $t \ge 0$ such that $\mu_n(K) \ge 1 - \eta$. Then, by compactness, $K \subseteq \{x : |x(0)| \le t\}$ and $K \subseteq \{x : w_x(\delta) \le \epsilon\}$ for all $n \ge 1$ and $\delta > 0$ chosen sufficiently small by the Arzelà-Ascoli Theorem. It quickly follows that $\mu_n(\{x : |x(0)| \geq t\}) \leq \epsilon$ and $\lim_{\delta\to 0}\sup_{n\geq 1}\mu_n(\{x : w_x(\delta) \geq \epsilon\}) = 0$ by choosing K appropriately.

For the reverse direction, we may instead show the result under the assumption (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, we can find for each $k \geq 0$ a collection of balls of radius k, $A_1, ..., A^{(k)}_{n_k}$ such that $\mu_1(\cup_{i=1}^{n_k} A^{(k)}_i) \geq 1-\epsilon 2^{-k}$. The closure K of the set $\cap_{k=1}^{\infty}\cup_{i=1}^{n_k} A^{(k)}_i$ 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. □

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. Afterwards, we are left with criteria for weak convergence that are much more easily verified.

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. $\square$

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. It is enough to 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 $X_i \stackrel{\mathcal{L}}{=} N(0, t_i - t_{i-1})$. This, along with tightness, is enough to complete the proof.

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.

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.

[Ath69] Krishna B Athreya. On a characteristic property of pólya's urn. *Studia Sci. Math. Hungar*, 4:31–35, 1969.

[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.

[Bur01] D Burago. A course in metric geometry. *American Mathematical Society*, 2001.

[Duq06] Thomas Duquesne. The coding of compact real trees by real valued functions. *arXiv preprint math/0604106*, 2006.

[Jan23] Svante Janson. Real trees, 2023.

[JK77] Norman Lloyd Johnson and Samuel Kotz. Urn models and their application; an approach to modern discrete probability theory. 1977.

[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.

[Pem07] Robin Pemantle. A survey of random processes with reinforcement. 2007.

[Pet06] Peter Peterson. *Riemannian Geometry*. Springer-Verlarg, 2006.

[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