university of groningen
faculty of science and engineering

Scaling Limits of Discrete Aggregation Trees

Master's Project Mathematics

December 2025

Student: Jeroen van Haastert

First supervisor: Dr. S. Donderwinkel

Second assessor: Dr. G. Bonnet

Abstract

In this thesis, we study the scaling limit of uniform random labeled rooted trees $\mathcal{T}_n$ to the continuum random tree (CRT), realized via the stick-breaking construction, in the Gromov–Hausdorff–Prokhorov topology. The discrete tree $\mathcal{T}_n$ is sampled using the Foata–Fuchs bijection, which can be regarded as the discrete analogue of the stick-breaking construction. We generalize existing results by introducing two families of non-uniform random labeled trees $\mathcal{T}_{n,\beta}$ and $\mathcal{T}_{n,\gamma}$ whose scaling limits are variants of the CRT constructed from Poisson point processes with intensities $t^{\beta}dt$ and $\ln^{\gamma}(t+1)dt$ respectively. In the latter case, we find a compactness threshold at $\gamma = 1$: for $\gamma > 1$, the limiting tree $\mathcal{T}_\gamma$ is compact almost surely, whereas for $\gamma \leq 1$ the tree $\mathcal{T}_\gamma$ is almost surely non-compact.

Contents

1 Introduction

The aim of the first part of this thesis is to describe the global structure of uniformly sampled labeled rooted trees on n vertices as n tends to infinity. We start with some background and basic definitions.

1.1 Background

Write $\mathbb{T}_n$ for the set of all *labeled rooted trees* on n vertices and denote the *root* by $\mathbf{r}_n$. By Cayley's formula, we have $|\mathbb{T}_n| = n^{n-1}$. We define random variable $\mathcal{T}_n$ to be the *uniform random labeled rooted tree* on n vertices. I.e. for each $T_n \in \mathbb{T}_n$, we have, $\mathbb{P}(\mathcal{T}_n = T_n) = \frac{1}{n^{n-1}}$. Note that $(\mathcal{T}_n, d_n)$ is a random metric space where d_n denotes the graph distance on $\mathcal{T}_n$.

Figure 1 contains an example of $\mathbb{T}_n$ and Figure 2 contains a sample of $\mathcal{T}_n$.

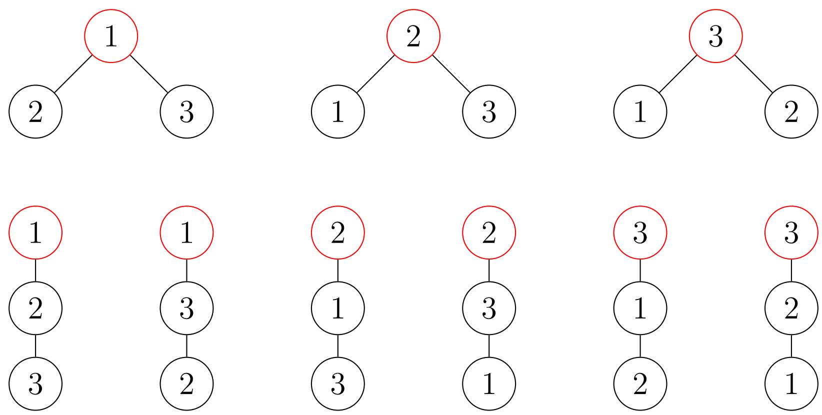


Figure 1: The set $\mathbb{T}_3$, where the red vertex denotes the root.

Write $v \in \mathcal{T}_n$, to denote a vertex in $\mathcal{T}_n$ (opposed to $v \in V(\mathcal{T}_n)$). Define $\mathrm{ht}(v) = d_n(v, \mathbf{r}_n)$ to be the *height of* v and $\mathrm{ht}(\mathcal{T}_n) = \max_{v \in \mathcal{T}_n} \mathrm{ht}(v)$ to be the *height of* $\mathcal{T}_n$. Observe that $\mathrm{ht}(v)$ and $\mathrm{ht}(\mathcal{T}_n)$ are random variables. For tree T_{60} in Figure 2, we have $\mathrm{ht}(10) = 6$ and $\mathrm{ht}(T_{60}) = 15$.

Another random variable of interest are distances in $\mathcal{T}_n$. Let $U, V \in_u \mathcal{T}_n$ be two vertices, where $\in_u$ denotes a uniformly chosen element. Then $d_n(U, V)$ is a random variable corresponding to the distance between two uniformly chosen vertices in $\mathcal{T}_n$.

The random variables $\mathrm{ht}(v)$, $\mathrm{ht}(\mathcal{T}_n)$ and $d_n(U, V)$ have one thing in common: they depend on the global structure of $\mathcal{T}_n$. This is unlike, for example, the degree distribution of $v \in_u \mathcal{T}_n$, which only depends on the neighbors of v. Coincidentally, the degree distribution of v is easy to study. Recall that Prüfer codes form a bijection between $[n]^{n-2} \times n$ and $\mathbb{T}$ (where $\times n$ is used to pick the root) such that the degree of $v \in \mathcal{T}_n$ is given by $\deg(v) = 1 + |\{S_n(i) = v\}|$ with $S_n \in_u [n]^{n-2}$. In particular, $\deg(v) \sim 1 + |\{S_n(i) = v\}|$ and it is well known that the latter converges to a $1 + \mathrm{Poi}(1)$ random variable as $n \to \infty$. This is an example of an easy to study random variable depending only on the local structure of $\mathcal{T}_n$.

It is not obvious how the asymptotic global structure of $\mathcal{T}_n$ behaves by looking at Prüfer codes, as information on the global structure is not easily recovered from the bijection. To describe the asymptotic global structure of $\mathcal{T}_n$, we turn to a different bijection.

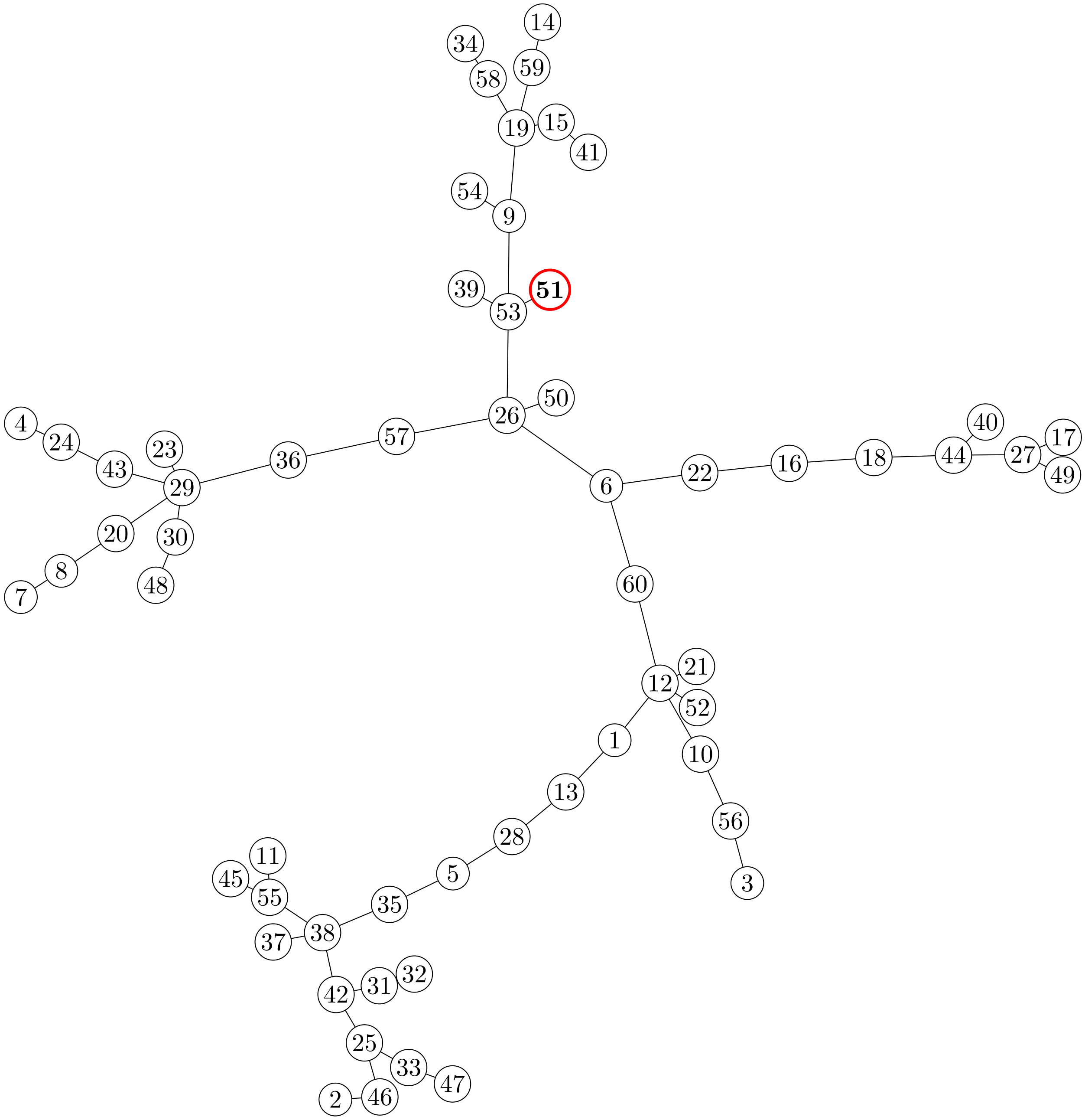
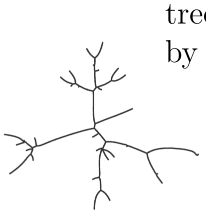


Figure 2: A sample of $\mathcal{T}_{60}$. The root is denoted in red

1.2 The Foata–Fuchs Bijection

Similar to Prüfer codes, the Foata–Fuchs bijection is a bijection between $[n]^{n-1}$ and $\mathbb{T}_n$. Whereas Prüfer codes are usually seen as bijection between $[n]^{n-2}$ and non-rooted trees, the Foata–Fuchs bijections always gives labeled rooted trees. To obtain a sequence in $[n]^{n-1}$ from a tree $T_n \in \mathbb{T}_n$, we use an exploration process: start at the root and record the path from the root to the lowest labeled leaf. Iteratively, record the path from the already explored tree to the next lowest labeled leaf until all leaves are visited. The sequence S_n is obtained by removing all leaf vertices from the recording. We formally introduce the reverse bijection.



Definition 1.2.1. Start with a sequence $S_n = (v_1, v_2, \ldots, v_{n-1}) \in [n]^{n-1}$.

i) Set $C_0^n = 0$ and set $C_1^n < C_2^n < \cdots < C_{i-1}^n$ to be the locations of S_n that contain a repeat. Lastly, set $C_i^n = n$.
ii) Let $l_1 < l_2, \cdots < l_i$ be the elements of $[n]$ not appearing in S_n, in increasing order.
iii) For $j = 1, 2, \ldots, i$, define $P_j = (v_{C_{j-1}^n}, v_{C_{j-1}^n+1}, \ldots, v_{C_j^n-1}, l_j)$ to be the path containing the vertices between C_{j-1}^n and C_j^n and appending vertex l_j.
iv) Define tree T_n as the tree with root v_1, vertices $[n]$ and edge set given by the union of edges in the paths P_j. That is,,

$$T_n = (V, E) = \left([n], \bigcup_{j=1}^{i} E(P_j)\right).$$

Theorem 1.2.2. The Foata–Fuchs bijection is a bijection between $\mathbb{T}_n$ and $[n]^{n-1}$.

Proof. Definition 1.2.1 shows how to construct $T_n \in \mathbb{T}_n$ from $S_n \in [n]^{n-1}$. For a proof that this construction is a bijection, we refer to [1, pages 2 and 3]. □

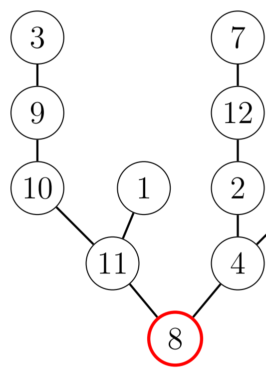


Figure 3: Tree $T_{12} \in \mathbb{T}_{12}$ with associated sequence $S_{12} = \big(8, 11, 11, 10, 9, 8, 4, 6, 4, 2, 12\big)$

Definition 1.2.3. Let $S_n \in [n]^{n-1}$ and $i \leq n-1$. The *vertex corresponding to* $S_n(i)$ is:

i) The vertex with label $S_n(i)$ if $S_n(i)$ is not a repeat.

ii) The vertex l_j if $S_n(i) = C_j^n$.

Let $F : S_n \to \mathbb{T}_n$ be the map associated with the Foata–Fuchs bijection. That is, for each $1 \leq i \leq n-1$, we set $F(S_n(i))$ to be the vertex corresponding to $S_n(i)$.

Example 1.2.4. In Figure 3, we see $F(S_{12}(4)) = 10$ and $F(S_{12}(6)) = 3$.

Remark 1.2.5. Certain global properties, like the distribution of the height of leaves in $\mathcal{T}_n$ are easily obtained through the Foata–Fuchs bijection. Indeed, let $v \in \mathcal{T}_n$ be the leaf with the smallest label and let $C_1^n \in [n]$ denote the index of the first repeat in S_n (if there are no repeats, set $C_1^n = n$). It follows that $F(S_n(C_i^n)) = v$ and $\{S_n(1), \ldots, S_n(C_i^n)\}$ forms the path from the root $\mathbf{r}_n$ to v so that $\mathrm{ht}(v) = C_1^n$. By vertex exchangeability of $\mathcal{T}_n$, we automatically obtain that for any leaf $u \in \mathcal{T}_n$, we have $\mathrm{ht}(u) \overset{d}{=} C_1^n$. In Section 1.3, we show that $C_1^n = \Theta\big(n^{\frac{1}{2}}\big)$ in probability. Thus, the height of a typical leaf in $\mathcal{T}_n$ is of the order $n^{\frac{1}{2}}$.

1.3 Tree $\mathcal{T}_n$ as Discrete Aggregation Tree

Using the Foata–Fuchs bijection, we construct $\mathcal{T}_n$ by gluing together the paths,

$$P_j = \big(S_n(C_j^n), \dots, S_n(v_{C_{j+1}^n - 1}), l_j\big).$$

This means that we may construct $\mathcal{T}_n$ iteratively by adding the paths one at a time: start with $\mathcal{T}_n^{(1)} = P_1$. Inductively, let $\mathcal{T}_n^{(k)}$ be the tree obtained from gluing the path P_k to $\mathcal{T}_n^{(k-1)}$ at the vertex with label $S_n(C_{k-1}^n)$. We illustrate this in Figure 4.

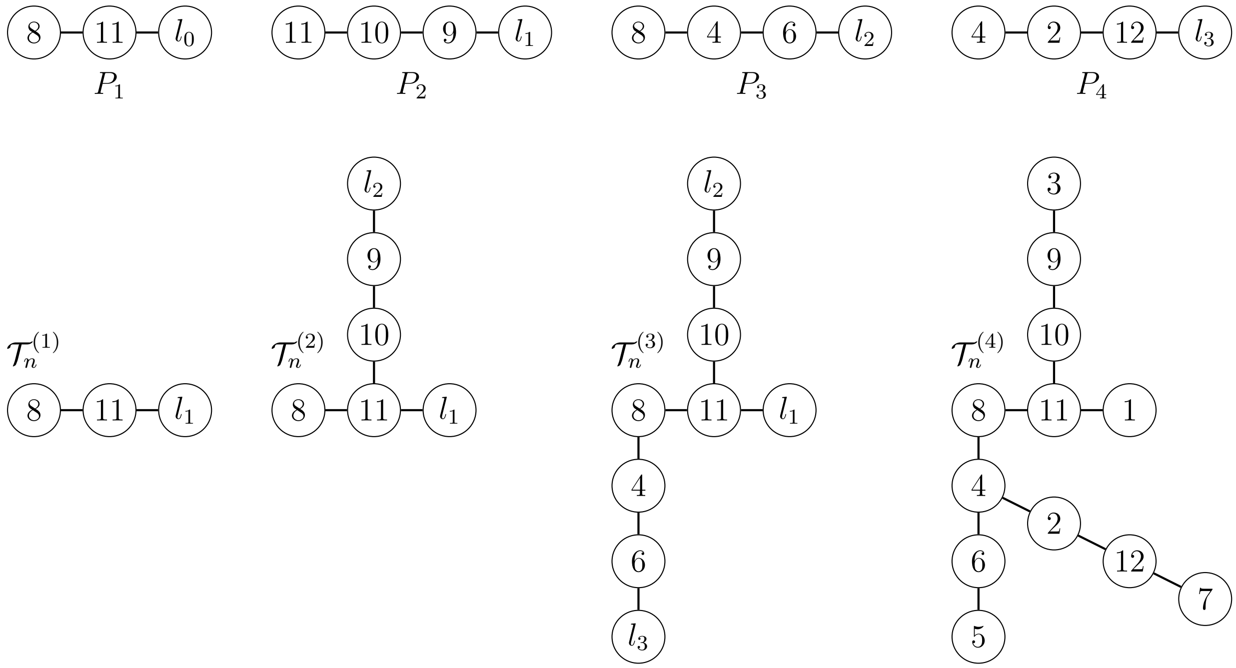


Figure 4: Trees $\mathcal{T}_{12}^{(i)}$ for $i \in \{1, \dots, 4\}$ constructed from $S_{12} = (8, 11, 11, 10, 9, 8, 4, 6, 4, 2, 12)$.

Remark 1.3.1. Observe that $\mathcal{T}_n^{(k)}$ is a random variable but inclusions $\mathcal{T}_n^{(1)} \subset \mathcal{T}_n^{(2)} \subset \cdots \subset \mathcal{T}_n$ are deterministically true. Furthermore, note that $\big(S_n(1), \dots, S_n(C_k^n - 1)\big)$ encodes the metric structure of $\mathcal{T}_n^{(k)}$ but not the labels of all vertices, as the leaves are labeled based on the missing vertices in all of S_n, not just $\big(S_n(1), \dots, S_n(C_k^n - 1)\big)$.

We call the paths P_j sticks or branches and say $\mathcal{T}_n^{(k+1)}$ is constructed from $\mathcal{T}_n^{(k)}$ by adding a stick or branch. Before trying to understanding the asymptotic global structure of $\mathcal{T}_n$, we give some intuitive reasoning what happens to the trees $\mathcal{T}_n^{(k)}$ when we take n to infinity.

We start with $\mathcal{T}_n^{(1)}$. Recall that $\mathcal{T}_n^{(1)}$ is the path from $\mathbf{r}_n$ to the lowest labeled leaf, which is the vertex corresponding to $S_n(C_1^n)$. Thus, $\mathcal{T}_n^{(1)}$ is a line graph of random number of vertices C_1^n. To compute the distribution of C_1^n, we observe that $C_1^n > x$ happens precisely when $S_n(1), \dots, S_n(x)$ are all unique. Note that $S_n(i) \in_u [n]$, and for $i < C_1^n$, we have $\big(S_n(1), \dots, S_n(i-1)\big)$ contains $i-1$ distinct elements. Thus,

$$\mathbb{P}\big(C_1^n > x\big) = \prod_{i=1}^{\lfloor x \rfloor} \left(1 - \frac{i-1}{n}\right) \xrightarrow[n \to \infty]{} 1.$$

We hence see that as $n \to \infty$, the height of a typical leaf in $\mathcal{T}_n$ goes to infinity. However, how fast does the height of a typical leaf go to infinity? For this, let $g(n) : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be some function to be determined. Then, by the reasoning above, we see,

$$\begin{aligned}
\mathbb{P}(C_1^n > g(n)x) = \prod_{i=1}^{\lfloor g(n)x \rfloor} \left(1 - \frac{i-1}{n}\right) &= \exp\left(\sum_{i=1}^{\lfloor g(n)x \rfloor} \ln\left(1 - \frac{i-1}{n}\right)\right), \\
&= \exp\left(-\frac{1}{n} \sum_{i=1}^{\lfloor g(n)x \rfloor} i - 1\right) + o(1), \\
&= \exp\left(-\frac{(g(n)x)^2}{2n}\right) + o(1).
\end{aligned}$$

Here, we used first order expansion $\ln(1-x) = -x + O(x^2)$. For a rigorous justification on the error bounds, we refer to the proof of Lemma 3.1.3. From the above analysis we see,

$$\mathbb{P}(C_1^n > g(n)x) \xrightarrow[n\to\infty]{} \begin{cases} 1 & \text{if } g(n) = o(n^{\frac{1}{2}}), \\ 0 & \text{if } g(n) = \omega(n^{\frac{1}{2}}), \\ \exp\left(-\frac{x^2}{2}\right) & \text{if } g(n) = n^{\frac{1}{2}}. \end{cases}$$

In particular, $n^{-\frac{1}{2}} C_1^n \xrightarrow[n\to\infty]{d} C_1$ where $C_1 \sim \text{Rayleigh}(1)$. Thus, the typical height of a leaf is on the order $n^{\frac{1}{2}}$ and for large n, the rescaled partial tree $\left(\mathcal{T}_n^{(1)}, n^{-\frac{1}{2}} d_n\right)$ can be viewed as 'approximating' a line segment of random length C_1 with vertices at equal distance $n^{-\frac{1}{2}}$. We illustrate this in Figure 5.

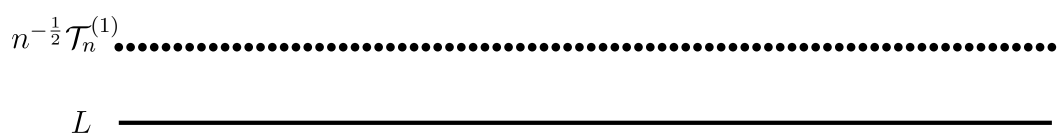


Figure 5: $n^{-\frac{1}{2}}\mathcal{T}_n^{(1)}$ approximates line segment L of random length C_1.

This gives intuition in the natural limit object of $\left(\mathcal{T}_n^{(1)}, n^{-\frac{1}{2}} d_n\right)$, also simply denoted $n^{-\frac{1}{2}}\mathcal{T}_n^{(1)}$: a line segment with random length $C_1 \sim \text{Rayleigh}(1)$. The correct topology for this convergence is the Gromov–Hausdorff topology, which is a metric on the space of compact metric spaces up to isometries. A comprehensive overview of the Gromov–Hausdorff distance can be found in Section 2.3.1.

We continue with understanding $\mathcal{T}_n^{(2)}$. Recall that $\mathcal{T}_n^{(2)}$ is created by gluing path P_2 to the vertex $S_n(C_1^n) \in \mathcal{T}_n^{(1)}$. Let B_1^n be the value in $\{1, \ldots, C_1^n - 1\}$ for which $S_n(B_1^n) = S_n(C_1^n)$. Note $B_1^n \in_u \{S_n(1), \ldots, S_n(C_1^n - 1)\}$, thus the second branch is glued to a uniform point in $\mathcal{T}_n^{(1)}$. For the length of the second branch, similar reasoning shows that $n^{-\frac{1}{2}} C_2^n \xrightarrow[n\to\infty]{d} C_2$ where C_2 is some random variable dependent on C_1 to still be determined.

This gives a heuristic for the limiting object of $\mathcal{T}_n^{(2)}$: let L_1 be a line segment of random length C_1 as above and L_2 a line segment of random length $C_2 - C_1$, for C_2 to be determined.

Attach L_2 to a uniformly random point on L_1. Let $\mathcal{T}^{(2)}$ be the resulting random metric space (with induced path metric). This metric space can be seen as limiting object for $\mathcal{T}_n^{(2)}$.

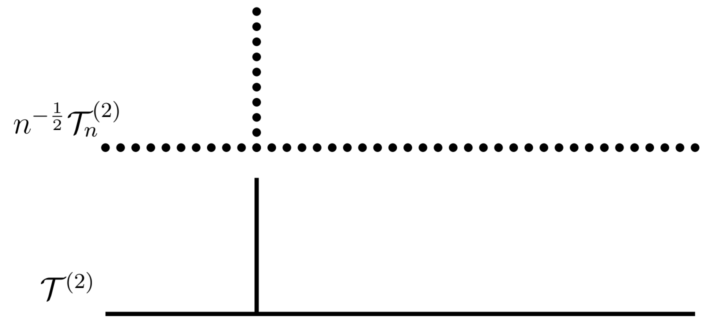


Figure 6: $n^{-\frac{1}{2}}\mathcal{T}_n^{(2)}$ approximates $\mathcal{T}^{(2)}$

We iterate this argument: construct $\mathcal{T}_n^{(k+1)}$ by attaching stick P_{k+1} to $\mathcal{T}_n^{(k)}$ at vertex $S_n(C_k)$. First, let $C_1^n < C_2^n < \cdots < C_k^n$ denote the indices of the first k repeats in S_n. We aim to find a heuristic for limiting distribution for $n^{-\frac{1}{2}}\big(C_1^n, \ldots, C_k^n\big)$. For this, we have,

$$\mathbb{P}\big(C_k^n > C_{k-1}^n + xn^{\frac{1}{2}} \mid C_{k-1}^n = \lfloor sn^{\frac{1}{2}} \rfloor\big) = \prod_{i=1}^{x} \left(1 - \frac{|\{S_n(1), \ldots, S_n(C_{k-1}^n)\}| + i}{n}\right),$$
$$\approx \prod_{i=1}^{\lfloor xn^{\frac{1}{2}} \rfloor} \left(1 - \frac{sn^{\frac{1}{2}} + i}{n}\right) \xrightarrow[n\to\infty]{} e^{-sx-\frac{x^2}{2}}.$$

where the approximation comes from $|\{S_n(1), \ldots, S_n(C_{k-1}^n)\}| = \lfloor sn^{\frac{1}{2}} \rfloor - (k-1) \approx sn^{\frac{1}{2}}$ since the fixed k repeats are negligible on scale $n^{\frac{1}{2}}$. The limit is independent of k: that is, for large n the distribution of the point $n^{-\frac{1}{2}}C_k^n$ depends only on the location of the point $n^{-\frac{1}{2}}C_{k-1}^n$, but is independent of the number of prior points and their locations. This memoryless property hints at the fact that $n^{-\frac{1}{2}}\big(C_1^n, C_2^n, \ldots\big)$ approaches an inhomogeneous Poisson point process, (inhomogeneous since the distribution of $n^{-\frac{1}{2}}C_k^n$ depends on the location of $n^{-\frac{1}{2}}C_{k-1}^n$). In Section 3.1, this heuristic is worked out rigorously. We refer to Section 2.2 for a formal introduction to the Poisson point process.

We saw a heuristic for the limiting object for the scaled repeat points $n^{-\frac{1}{2}}C_k^n$. This determines the distribution of the length of the branches of $\mathcal{T}_n$. Next, we understand how these branches are attached to the partial trees $\mathcal{T}_n^{(k)}$. For this, define $B_k^n = \min\{l : S_n(l) = S_n(C_k)\}$ to be the first time $S_n(C_k^n)$ appeared in S_n. This is the index corresponding to the attachment point of stick P_{k+1} in $\mathcal{T}_n^{(k)}$. Note B_k^n is almost uniform over $\{1, \ldots, C_k^n - 1\}$. Indeed, B_k^n is uniform if $S_n(1), \ldots, S_n(C_k^n - 1)$ contains no repeats. In our setting, $S_n(1), \ldots, S_n(C_k^n - 1)$ contains exactly $k-1$ repeats. However, since C_k^n is on the scale $n^{\frac{1}{2}}$, the proportion of repeats becomes negligible, which explains $n^{-\frac{1}{2}}B_1^n \xrightarrow[n\to\infty]{} B_i$ with $B_i \sim \mathrm{Unif}([0, C_i])$. Section 3.1 contains a formal argument.

Informally, $n^{-\frac{1}{2}}\mathcal{T}_n^{(k)}$ consists of k branches of vertices at distance $n^{-\frac{1}{2}}$ such that the length of branch i converges to $C_i - C_{i-1}$ for $C_1 < C_2 < \ldots$ the ordered points of some Poisson point process and the branches are glued to a roughly uniform point on the already existing tree. As limit, define $\mathcal{T}^{(k)}$ to be the random metric space obtained from inductively gluing together line segments of length $C_i - C_{i-1}$ at a uniformly chosen point in the already constructed space, together with the induced path metric. See Figure 7 for an illustration.

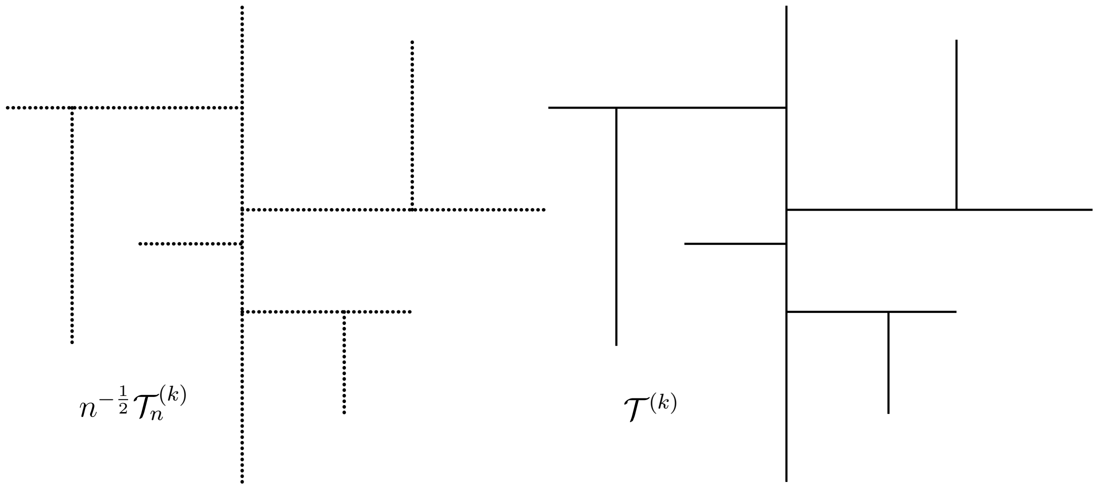


Figure 7: A sample of $n^{-\frac{1}{2}}\mathcal{T}_n^{(k)}$ and a sample of $\mathcal{T}^{(k)}$.

Remark 1.3.2. It should be noted that we draw samples of $n^{-\frac{1}{2}}\mathcal{T}_n^{(k)}$ and $\mathcal{T}^{(k)}$ that suggestively look alike. This way of representing the trees is not misleading: since the branch length and attachment points converge in distribution, we may work on a probability spaces where this convergence happens almost surely. See Section 3.2 for details.

Lastly, we define measures on $\mathcal{T}_n$ and $\mathcal{T}_n^{(k)}$.

Definition 1.3.3. Let ν_n be the uniform probability measure defined on the vertices of $\mathcal{T}_n$. That is, $\nu_n(A) = \frac{|A|}{n}$ for all $A \subset \mathcal{T}_n$. Similarly, $\nu_n^{(k)}$ denotes the uniform probability measure on $\mathcal{T}_n^{(k)}$ so that $\nu_n^{(k)}(A) = \frac{|A|}{C_k^n - 1} = \frac{|A|}{|\mathcal{T}_n^{(k)}|}$ for all $A \subset \mathcal{T}_n^{(k)}$.

1.4 The Continuum Random Tree

We informally defined $\mathcal{T}^{(k)}$ as gluing together line segment of random length. We formalize this construction, leading to the definition of the Continuum Random Tree (CRT). The CRT can be viewed as the scaling limit of $\mathcal{T}_n$ and was first introduced by Aldous in [2]. It should be noted that the CRT has various definitions, we will be using the stick-breaking definition.

Definition 1.4.1. Let ℓ^1 be the space of all absolute real sequences. That is,

$$\ell^1 = \left\{ (x_1, x_2, \dots) : x_i \in \mathbb{R} \text{ and } \sum_{i=1}^{\infty} |x_i| < \infty \right\}.$$

The space ℓ^1 comes with basis vectors $z_i = (0, \dots, 0, 1, 0, 0, \dots)$: $i-1$ zeros followed by a one.

Let η be a Poisson point process of intensity tdt on $\mathbb{R}_{\geq 0}$ and let $0 < C_1 < C_2 < \dots$ denote its ordered points. To each C_i, associate $B_i \sim \text{Unif}([0, C_i))$. Define $\rho(t) : \mathbb{R}_{\geq 0} \to \ell^1$ piecewise on the intervals $(C_{i-1}, C_i]$ by,

$$\rho(t) = \begin{cases} (t, 0, \dots, 0) & \text{for } x \in [0, C_1], \\ \rho(B_{i-1}) + tz_i & \text{for } x \in (C_{i-1}, C_i] \text{ with } i \geq 2. \end{cases}$$

Define $\mathcal{T}^{(k)} = \rho([0, C_k])$ and $\mathcal{T}(t) = \rho([0, t])$. Observe that $\mathcal{T}^{(k)}$ and $\mathcal{T}(t)$ are compact for each $k \in \mathbb{N}$ and $t \in \mathbb{R}$. Lastly, we define,

$$\mathcal{T} = \overline{\bigcup_{k=1}^{\infty} \mathcal{T}^{(k)}},$$

to be the closure of the union of partial trees $\mathcal{T}^{(k)}$. Observe that $\mathcal{T}^{(1)} \subset \mathcal{T}^{(2)} \subset \cdots \subset \mathcal{T}$ is deterministically true and $(\mathcal{T}, d)$ is a random metric space, where d denotes distance in ℓ^1.

Remark 1.4.2. The intervals $[C_{i-1}, C_i]$ are called branches and $\mathcal{T}^{(k)}$ is constructed by attaching a branch of length $C_k - C_{k-1}$ to a uniform point on the already constructed tree. Hence $\mathcal{T}^{(k)}$ is a formalization of the description of the limit objects of $n^{-\frac{1}{2}}\mathcal{T}_n^{(k)}$ as described in Section 1.3.

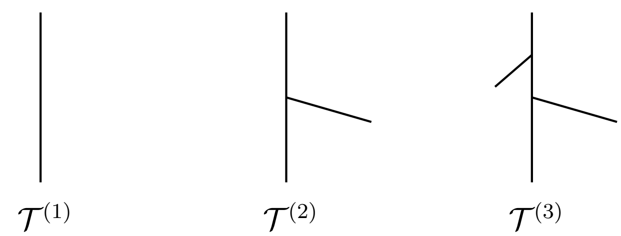


Figure 8: Sample of tree $\mathcal{T}^{(i)}$ for $i = 1, 2, 3$ as subset of ℓ^1.

Remark 1.4.3. For $k \geq 4$, we draw $(\mathcal{T}^{(k)}, d)$ in $\mathbb{R}^2$. Given that d is the ℓ^1 metric, and branches in $\mathcal{T}^{(k)}$ are orthogonal in ℓ^1, we see that distances on $\mathcal{T}^{(k)}$ should be interpreted as distances along the branches in $\mathcal{T}^{(k)}$ and not Euclidean distance in $\mathbb{R}^2$.

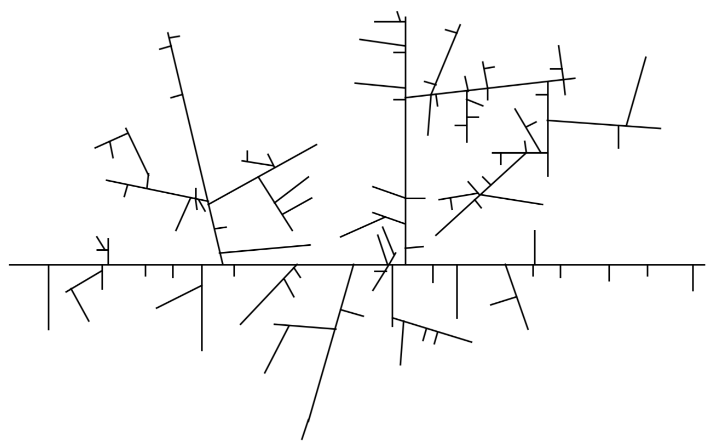
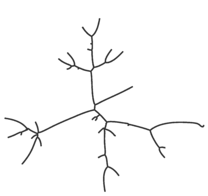


Figure 9: A sample of $\mathcal{T}^{(110)}$.



Besides $\mathcal{T}$ being the scaling limit of $\mathcal{T}_n$ as random metric spaces, we will also see that we can pick points in $\mathcal{T}$ uniformly in a way that resembles picking uniform vertices in $\mathcal{T}_n$. To this end, we introduce measures on $\mathcal{T}^{(k)}$.

Definition 1.4.4. Let $\mu^{(k)} = \rho^*\overline{\lambda}$ where $\overline{\lambda}$ is the normalized Lebesgue measure on $[0, C_k]$. Here, ρ^* denotes the push forward by ρ.

Remark 1.4.5. Observe that $\mu^{(k)}$ is a random measure such that $\mu^{(k)}(A)$ denotes the proportion of A in $\mathcal{T}^{(k)}$ in much the same way that $\nu_n^{(k)}(B)$ denotes the proportion of B in $\mathcal{T}_n^{(k)}$. In particular, $\mu^{(k)}$ and $\nu_n^{(k)}$ are similar in the following sense: use a probability space where the branch lengths of $n^{-\frac{1}{2}}\mathcal{T}_n^{(k)}$ and $\mathcal{T}^{(k)}$ converge almost surely (see Remark 1.3.2), and isometrically embed $n^{-\frac{1}{2}}\mathcal{T}_n^{(k)}$ and $\mathcal{T}^{(k)}$ in a common space Z via embeddings ϕ and ψ respectively. Let $A \subset Z$. Then, $\phi^*\mu^{(k)}(A)$ and $\psi^*\nu_n^{(k)}(A)$ are roughly identical as illustrated in Figure 10.

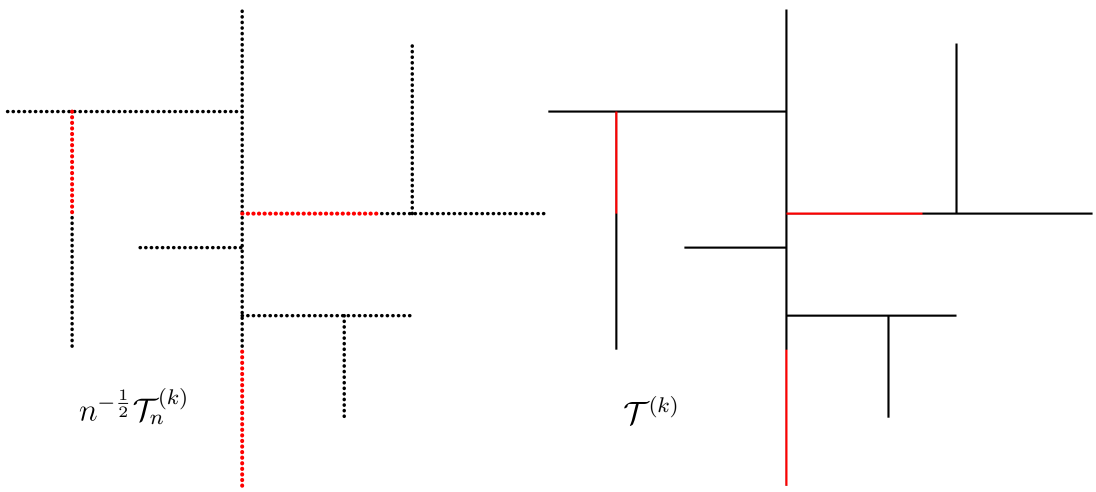


Figure 10: the proportion of red vertices in $n^{-\frac{1}{2}}\mathcal{T}_n^{(k)}$ and $\mathcal{T}^{(k)}$ is roughly equal.

1.5 Global Convergence and Proof Strategy

We have seen a heuristic that the length of the first k branches in $n^{-\frac{1}{2}}\mathcal{T}_n^{(k)}$ converge in distribution to the length of the first k sticks of $\mathcal{T}^{(k)}$ and in both cases, attachment is roughly uniform on the constructed tree. Furthermore, both $\mu^{(k)}(A)$ and $\nu_n^{(k)}(B)$ are similar in the sense of Remark 1.4.5. Hence, the following result is not surprising.

Theorem 1.5.1. For $k \in \mathbb{N}$, we have,

$$\left(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}}d_n, \nu_n^{(k)}\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}^{(k)}, d, \mu^{(k)}\right),$$

in the Gromov–Hausdorff–Prokhorov topology.

The Gromov–Hausdorff–Prokhorov (GHP) distance is a metric on the space of compact measure metric spaces, (up to isometries). We refer to Section 2.3.3 for details. Section 3 is dedicated to a formal proof of the theorem. Theorem 1.5.1 is known as convergence of the finite dimensional distribution. We aim to upgrade this to,

Theorem 1.5.2. There exists a probability measure μ on $\mathcal{T}$ such that,

$$\left(\mathcal{T}_n, n^{-\frac{1}{2}} d_n, \nu_n\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}, d, \mu\right),$$

in the GHP topology.

To upgrade Theorem 1.5.1 to Theorem 1.5.2, we follow the reasoning used in [2]. For this, we introduce the following standard result, which is obtained through the Portmanteau lemma.

Lemma 1.5.3. Suppose that $X_{n,k}, X_n, X_k$ and X are random variables living in the same metric space and for all $\epsilon > 0$, we have,

$$i)\ X_{n,k} \xrightarrow[n\to\infty]{d} X_k,\ ii)\ \lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(d(X_{n,k}, X_k) > \epsilon) = 0 \text{ and } iii)\ \lim_{k\to\infty} \mathbb{P}(d(X_k, X) > \epsilon) = 0,$$

then we have $X_n \xrightarrow[n\to\infty]{d} X$.

This is a standard result. We refer to Section 2.4 for details. To use this result, we set,

$$X = (\mathcal{T}, d, \mu),\ X_n = \left(\mathcal{T}_n, n^{-\frac{1}{2}} d_n, \nu_n\right),\ X_k = \left(\mathcal{T}^{(k)}, d, \mu^{(k)}\right) \text{ and } X_{n,k} = \left(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}} d_n, \nu_n^{(k)}\right).$$

Theorem 1.5.1 implies statement $i)$ of Lemma 1.5.3 is satisfied. By rewriting statement $ii)$ and statement $iii)$, we obtain that it suffices to show,

Corollary 1.5.4. We have shown $\left(\mathcal{T}_n, n^{-\frac{1}{2}} d_n, \nu_n\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}, d, \mu\right)$ in the GHP topology, if we can show that for all $\epsilon > 0$, and some probability measure μ defined on $\mathcal{T}$, we have,

$$i)\ \lim_{t\to\infty} \mathbb{P}\left(d_H\left(\mathcal{T}(t), \mathcal{T}\right) > \epsilon\right) = 0, \qquad ii)\ \lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(d_H\left(\mathcal{T}_n(tn^{\frac{1}{2}}), \mathcal{T}_n\right) > \epsilon\right) = 0,$$

$$iii)\ \lim_{k\to\infty} \mathbb{P}\left(d_P\left(\mu^{(k)}, \mu\right) > \epsilon\right)) = 0, \qquad iv)\ \lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(d_P\left(\nu_n^{(k)}, \nu_n\right) > \epsilon\right) = 0.$$

The first and third statement combine into $iii)$ of Lemma 1.5.3 and the second and fourth statement combine into $ii)$ of Lemma 1.5.3. Together with Theorem 1.5.1, this shows all criteria of 1.5.3 are satisfied. We refer to Section 2.4 for the details.

This gives a clear guideline for the proof of Theorem 1.5.2. We state the results we have to show, and in which sections the corresponding proofs can be found.

i) We show $n^{-\frac{1}{2}}\left(C_1^n, \ldots, C_k^n, B_1^n, \ldots, B_k^n\right) \xrightarrow[n\to\infty]{d} \left(C_1, \ldots, C_k, B_1, \ldots, B_k\right)$. Section 3.1

ii) We show $\left(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}} d_n, \nu_n^{(k)}\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}^{(k)}, d, \mu^{(k)}\right)$. Section 3.2

iii) For all $\epsilon > 0$, we have $\mathbb{P}\left(d_H\left(\mathcal{T}(t), \mathcal{T}\right) > \epsilon\right) \to 0$, as $t \to \infty$. Section 4.1

iv) For all $\epsilon > 0$, we have $\limsup_{n\to\infty} \mathbb{P}\left(d_H\left(\mathcal{T}_n(tn^{\frac{1}{2}}), \mathcal{T}_n\right) > \epsilon\right) \to 0$, as $t \to \infty$. Section 4.2

v) For all $\epsilon > 0$, we have $\mathbb{P}\left(d_P\left(\mu^{(k)}, \mu\right) > \epsilon\right)) \to 0$, as $k \to \infty$. Section 5.1

vi) For all $\epsilon > 0$, we have $\limsup_{n\to\infty} \mathbb{P}\left(d_P\left(\nu_n^{(k)}, \nu_n\right) > \epsilon\right) \to 0$ as $k \to \infty$. Section 5.2

Remark 1.5.5. Convergence $\left(\mathcal{T}_n, n^{-\frac{1}{2}}d_n, \nu_n\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}, d, \mu\right)$ in the GHP topology allows us to deduce many asymptotic global properties of $\mathcal{T}_n$ from $\mathcal{T}$. An immediate example is that almost surely $\mathrm{ht}(\mathcal{T}_n) = O(n^{\frac{1}{2}})$, which follows from $\mathcal{T}$ being compact with probability 1 (see Remark 4.0.1). Another example includes convergence of mass in ϵ-balls. The short survey [14] contains some examples on showing which properties can be passed to the limit in GHP convergence. A last example is convergence of distances of uniformly chosen points in $n^{-\frac{1}{2}}\mathcal{T}_n$ and $\mathcal{T}_n$, which is covered in [9].

1.6 Generalizations

Recall Definition 1.4.1, in which $\mathcal{T}$ was constructed by gluing sticks of lengths determined by a Poisson point process of intensity tdt. We ask 2 questions: what if we consider Poisson point processes of different intensities, is the corresponding tree $\mathcal{T}$ still compact almost surely? If so, can we find random variable $\mathcal{T}_n$ taking values in $\mathbb{T}_n$ such that $\left(\mathcal{T}_n, g(n)d_n, \nu_n\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}, d, \mu\right)$ in the GHP topology? Here $g(n)$ is some scaling function and ν_n is the uniform measure on the vertices of $\mathcal{T}_n$ and μ is some probability measure on $\mathcal{T}$.

In chapter 6, we construct random tree $\mathcal{T}_{n,f}$ non uniformly by defining,

$$S_n(i) \sim \begin{cases} S_n^f(j) \text{ with } j \in_u \{1, \dots, i-1\}, & \text{with probability } f\left(\frac{i}{n}\right), \\ \mathrm{Unif}\big([n] \setminus \{S_n(1), \dots, S_n(i-1)\}\big), & \text{with probability } 1 - f\left(\frac{i}{n}\right), \end{cases}$$

for some function $f : [0,1] \to [0,1]$ and letting $\mathcal{T}_{n,f}$ be the tree obtained by applying the Foata–Fuchs bijection to S_n^f. See Section 6 for intuition behind this exact choice of S_n. We have the following result.

Theorem 1.6.1. Let $f(x) = x^\beta$ for $\beta > 0$. Then, there exists a probability measure μ on $\mathcal{T}_\beta$ such that,

$$\left(\mathcal{T}_{n,f}, n^{-\frac{\beta}{\beta+1}}d_n, \nu_n\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}_\beta, d, \mu\right),$$

in the GHP topology. Here, $\mathcal{T}_\beta$ denotes the tree obtained from the stick-breaking construction with Poisson point process of intensity $t^\beta dt$.

In Section 7, we construct $\mathcal{T}_{n,f}$ by applying the Foata–Fuchs bijection to,

$$S_n \equiv S_n^f(i) \sim \begin{cases} S_n(j) \text{ where } j \in_u \{1, \dots, i-1\}, & \text{with probability } f(i,n), \\ \mathrm{Unif}\big([n] \setminus \{S_n(1), \dots, S_n(i-1)\}\big), & \text{with probability } 1 - f(i,n), \end{cases}$$

where $f(i,n) \equiv f_n^\gamma(i) = \ln^\gamma(in^{-\frac{1}{2}} + 1)n^{-\frac{1}{2}}$. We have the following result,

Theorem 1.6.2. For $\gamma > 1$, we have convergence in the GHP-topology,

$$\left(\mathcal{T}_{n,f}, n^{-\frac{1}{2}}d_n, \nu_n\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}_\gamma, d, \mu\right),$$

where $\mathcal{T}$ is the tree obtained from the stick-breaking construction with a Poisson point process with intensity $\ln^\gamma(t+1)dt$. If $\gamma \leq 1$, then $\mathcal{T}_\gamma$ is not compact almost surely.

In particular, at $\gamma = 1$ we find a threshold where $\mathcal{T}_\gamma$ fails to be compact.

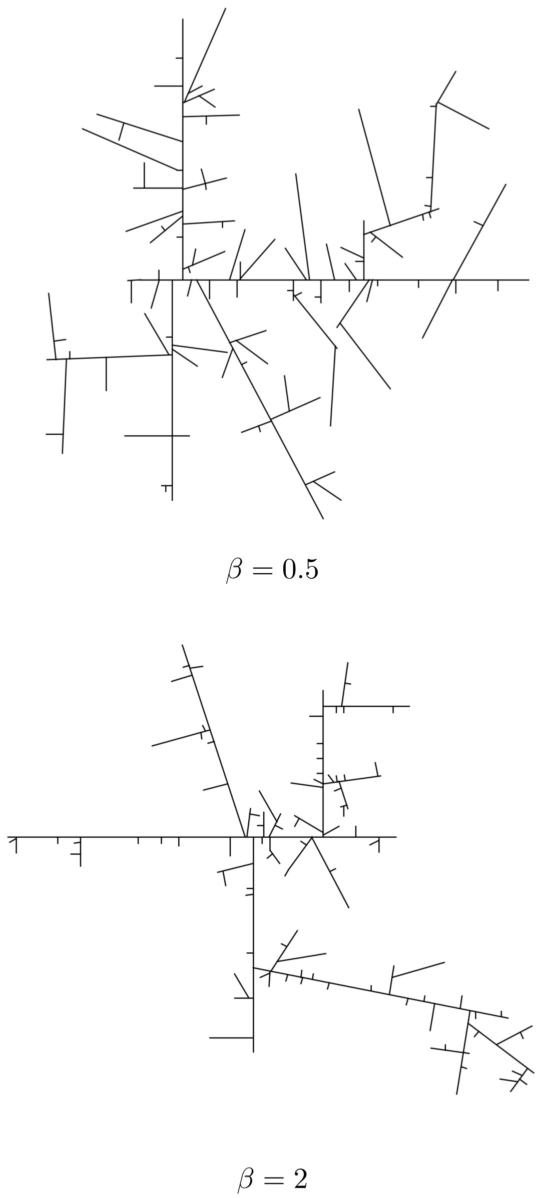
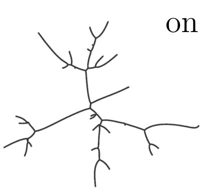


Figure 11: four trees constructed from a Poisson point process with intensity $t^\beta dt$. The trees are constructed from the interval $\left[0, (100(\beta+1))^{\frac{1}{\beta+1}}\right]$ so that the expected number of branches of each tree equals 100.

Remark 1.6.3. One can sample trees $\mathcal{T}_\beta$ in a coupled manner: start with a Poisson point process of intensity 1 on $\mathbb{R}^2$. Let $T_\beta = \{(x,y) \in \mathbb{R}^2 : 0 \leq x, 0 \leq y \leq x^\beta\}$. It follows that $\eta_\beta = \{(X,Y) \in \eta : (X,Y) \in T_\beta\}$ can be written as $\eta = \sum_{i=1}^{\infty} \delta_{(C_i,B_i)}$ where $0 < C_1 < C_2 < \dots$ equal (in distribution) the k first ordered points of a Poisson point process of intensity $t^\beta dt$ and $B_i \sim \mathrm{Unif}([0, C_i])$. Thus, all $\mathcal{T}_\beta$ can be constructed from a single PPP such that $\mathcal{T}_\beta$ has the correct law for each β. On the bottom left of each page, we have drawn $\mathcal{T}_\beta$ sampled in this manner. The intensity of the tree drawn on page number i is approximately $1 + i/1000$. The images can be turned into a video by using the thesis as a flipbook. Note that you start on the last page, so that you watch the tree change from intensity 1.1 to 1.



2 Theoretical Framework

This section aims to provide a rigorous background to the Poisson point process and the Gromov–Hausdorff–Prokhorov topology as well as a formalization of the proof outline discussed in Section 1.5. We start with the introduction of basic measure theory.

2.1 Measure Theory

The Poisson point process is defined through measure theory. Hence, we introduce some basic measure theory first.

Definition 2.1.1. A *measurable space* is a set X together with set of subsets $\mathcal{F}$, where $\mathcal{F}$ is a σ-algebra, i.e. it satisfies,

i) $X \in \mathcal{F}$,

ii) if $A \in \mathcal{F}$ then $X \setminus A \in \mathcal{F}$,

iii) if $(A_n)_{n\geq 1} \subseteq \mathcal{F}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Definition 2.1.2. Let $\mathcal{R}$ be a family of sets. The σ-algebra *generated by* $\mathcal{R}$, denoted $\sigma(\mathcal{R})$, is the smallest σ-algebra containing $\mathcal{R}$. Observe that $\sigma(\mathcal{R})$ always exists, since the power set of X trivially is a σ-algebra containing $\mathcal{R}$.

Example 2.1.3. The Borel σ-algebra on metric space (M, d), denoted $\mathcal{B}(M)$, is defined to be the σ-algebra generated by the open sets of M.

Definition 2.1.4. Let $(X, \mathcal{F}_X)$ and $(Y, \mathcal{F}_Y)$ be two measurable spaces. A function $f : X \to Y$ is called *measurable* if $f^{-1}(B) \in \mathcal{F}_X$ for all $B \in \mathcal{F}_Y$.

Definition 2.1.5. Let $(X, \mathcal{F})$ be a measurable space. A *measure* $\mu : \mathcal{F} \to [0, \infty]$ is a function for which,

i) $\mu(\emptyset) = 0$,

ii) If $(A_n)_{n\geq 1} \subseteq \mathcal{F}$ are pairwise disjoint, then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.

The triple $(X, \mathcal{F}, \mu)$ is a *measure space*. If $\mu(X) < \infty$, μ is a *finite measure*. If $\mu(X) = 1$, then μ is a *probability measure* in which case $(X, \mathcal{F}, \mu)$ is called a *probability space*.

Definition 2.1.6. A measure μ on $(X, \mathcal{F})$ is,

i) *σ-finite* if there exists $(A_n)_{n\geq 1} \subseteq \mathcal{F}$ so that $X = \bigcup_{n=1}^{\infty} A_n$ and $\mu(A_n) < \infty$ for all n.

ii) s-*finite* if μ is a countable sum of finite measures. In other words, we can find $(\mu_i)_{i\geq 1}$ so that $\mu(A) = \sum_{i=1}^{\infty} \mu_i(A)$ for all $A \in \mathcal{F}$ and $\mu_i(X) < \infty$ for all i.

Observe that every σ-finite measure is automatically s-finite. The converse does not hold.

In a measurable space $(X, \mathcal{F})$, the set of subsets $\mathcal{F}$ is often large and impractical to work with directly. However, due to the structural properties of σ-algebras it usually suffices to work on a smaller generating subset of $\mathcal{F}$. Carathéodory's extension theorem and Dynkin's $\pi - \lambda$ lemma are two results using this line of reasoning. We introduce these results below.

Theorem 2.1.7. Let $\mathcal{R}$ be a collection of sets for which,

$$A, B \in \mathcal{R} \implies A \cup B \in \mathcal{R} \text{ and } A \backslash B \in \mathcal{R}.$$

Let $\mu_0 : \mathcal{R} \to [0, \infty]$ satisfy $\mu_0(\emptyset) = 0$ and $\mu_0\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$ whenever A_i are pairwise disjoint and $\bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$. Then there exists a measure on $\sigma(\mathcal{R})$ which agrees with μ_0 on $\mathcal{R}$. Furthermore, the extension is unique whenever $\mu_0(A_i) < \infty$ for all i.

Proof. This is Carathéodory's extension theorem. See [11, Theorem 1.7.3] for a proof. □

Lemma 2.1.8. Let $P, L \subset 2^X$, for X a set. If P is closed under intersection, and L satisfies,

i) $\emptyset \in L$,

ii) $A, B \in L$ with $A \subset B \implies B \backslash A \in L$,

iii) $A_1, A_2, \ldots$ disjoint implies $\bigcup_{i=1}^{n} A_i \in L$.

Then $P \subset L$, implies $\sigma(P) \subset L$. Family P is known as a π-system, L as a Dynkin-system.

Proof. This is Dynkin's $\pi - \lambda$ Lemma, a proof can be found in [4, Theorem 3.2]. □

Corollary 2.1.9. Let X be a set and $P \subset 2^X$ be closed under finite intersections. If μ and μ' agree on P and are σ-finite, then μ and μ' agree on $\sigma(P)$.

Proof. We first assume μ and μ' are finite and define $L = \{A \subset X : \mu(A) = \mu'(A)\}$ to be the sets of which μ, μ' agree. One may check L satisfies $i) - iii)$ of Lemma 2.1.8 and $P \subset L$. We conclude $\sigma(P) \subset L$ and hence μ, μ' agree on $\sigma(P)$.

In case μ, μ' are only σ-finite, let $(A_n)_{n \geq 1}$ be a measurable partition of X such that $A_n \in \sigma(P)$ and $\mu(A_n), \mu'(A_n) < \infty$ for all n. This is guaranteed possible since μ, μ' are σ-finite. Define π-system $P_n = \{B \cap A_n \text{ for all } B \in P\}$ and define $\mu_n(E) = \mu(E \cap A_n)$ and $\mu_n'(E) = \mu_n'(E \cap A_n)$. By the finite measure case, we know μ_n and μ_n' agree on $\sigma(P_n)$ for all n. Let $B \in \sigma(P)$. Then,

$$\mu(B) = \mu\Big(\bigcup_{n=1}^{\infty} (B \cap A_n)\Big) = \sum_{n=1}^{\infty} \mu\big(B \cap A_n\big) = \sum_{n=1}^{\infty} \mu'\big(B \cap A_n\big) = \mu'\Big(\bigcup_{n=1}^{\infty} (B \cap A_n)\Big) = \mu'(B),$$

which concludes the proof. □

We use Carathéodory's Extension Theorem to define product measure spaces. Below, we give the construction for finite products. We restrict our attention to the product of two spaces, but any finite number products can be reached inductively. A similar, yet more technical construction can be applied to achieve similar results for (countably) infinite product spaces.

Definition 2.1.10. Let $(X_1, \mathcal{F}_1), (X_2, \mathcal{F}_2)$ be two measurable spaces. We define their *product measurable space* as $(X_1, \mathcal{F}_1) \times (X_2, \mathcal{F}_2) = (X_1 \times X_2, \mathcal{F}_1 \otimes \mathcal{F}_2) = (X_1 \times X_2, \sigma(\mathcal{F}_1 \times \mathcal{F}_2))$.

Theorem 2.1.11. For two finite measure spaces $(X_1, \mathcal{F}_1, \mu_1), (X_2, \mathcal{F}_2, \mu_2)$, there exists a unique measure μ on the product measurable space $(X_1 \times X_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ for which,

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2), \qquad \text{for all } A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2.$$

Proof. This follows directly from Carathéodory's Extension Theorem. Define $\mathcal{R} = F_1 \times F_2$ and $\mu_0 : \mathcal{R} \to [0,\infty]$ as $\mu_0(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$. Then Carathéodory's Extension Theorem gives a unique extension μ to $\sigma(\mathcal{R}) = \mathcal{F}_1 \otimes \mathcal{F}_2$ such that $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for all $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$. Uniqueness follows as μ_0 is σ-finite on $\mathcal{R}$. □

Definition 2.1.12. Let $(X_1, \mathcal{F}_1, \mu_1), (X_2, \mathcal{F}_2, \mu_2)$ be two finite measure spaces. The *product measure space* is defined as,

$$(X_1, \mathcal{F}_1, \mu_1) \times (X_2, \mathcal{F}_2, \mu_2) = (X_1 \times X_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu),$$

where μ is the unique measure so that $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for all $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$.

This definition can be extended to countable products of probability measure spaces.

Definition 2.1.13. Let $(X_i, \mathcal{F}_i)_{i=1}^{\infty}$ be a sequence of probability spaces. Define,

$$X = \prod_{i=1}^{\infty} X_i, \quad \text{and} \quad \mathcal{F} = \bigotimes_{i=1}^{\infty} \mathcal{F}_i = \sigma\left(\left\{\prod_{i=1}^{\infty} A_i : A_i \in \mathcal{F}_i \text{ and } A_i \neq X_i \text{ finitely often}\right\}\right).$$

Then $(X, \mathcal{F})$ denotes the *product measurable space* of $(X_i, \mathcal{F}_i)_{i=1}^{\infty}$.

Remark 2.1.14. Note that the above definition for $\mathcal{F}$ is chosen precisely such that the projection maps become measurable. In other words, this choice of σ-algebra is the smallest such that $\pi_j^{-1}(A) \in \mathcal{F}$ for all $A \in X_j$. This is often taken as equivalent definition.

Theorem 2.1.15. Let $(X_i, \mathcal{F}_i, \mu_i)_{i=1}^{\infty}$ be probability measure spaces with $(X, \mathcal{F})$ their product space. We can define a measure μ on X such that for all finite $J \subset \{1, 2, \dots\}$, we have,

$$\mu\left(\prod_{j\in J} A_j \prod_{j\notin J} X_i\right) = \prod_{j\in J} \mu_j(A_j), \qquad \text{where } A_j \in \mathcal{F}_j \text{ for all } j \in J,$$

The triple $(X, \mathcal{F}, \mu)$ is called the *product measure space* of $(X_i, \mathcal{F}_i, \mu_i)_{i=1}^{\infty}$.

Proof. Similar to the proof of the finite case, we define $\mathcal{R}$ and $\mu_0 : \mathcal{R} \to [0, 1]$ as,

$$\mathcal{R} = \left\{\prod_{i=1}^{\infty} A_i : A_i \in \mathcal{F}_i \text{ and } A_i \neq X_i \text{ finitely often}\right\} \text{ and } \mu_0\left(\prod_{j\in J} A_j \prod_{j\notin J} X_i\right) = \prod_{j\in J} \mu_j(A_j),$$

To apply Carathéodory's Extension Theorem, one needs to check that $\mathcal{R}$ is closed under finite unions and intersections, μ_0 is well defined, and $\mu_0\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu_0(A_i)$ for pairwise disjoint A_i with $\bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$. Unlike the finite case, these statements are no longer trivial. We outsource the details to [15, Theorem 4]. □

2.2 The Poisson Point Process

This subsection aims to give a formal description of the Poisson point process (PPP). We start with its definition and proceed by proving the existence of the PPP. The theory in this subsection is largely based on work by Günter Last and Mathew Penrose [12]. Throughout this section, we write $\mathbb{N}_0 = \mathbb{N} \cup 0$ and $\overline{\mathbb{N}}_0 = \mathbb{N}_0 \cup \infty$.

2.2.1 Definition of the PPP

Definition 2.2.1. Let $(X, \mathcal{F})$ be a measurable space. We define N to be the set of countable sums of all nonnegative integer valued measures. That is,

$$N \equiv N(X) = \left\{ \sum_{i=1}^{\infty} \mu_i : \mu_i(A) \in \mathbb{N}_0 \text{ for all } A \in \mathcal{F} \right\},$$

where we write N instead of $N(X)$ for brevity. We turn this set into a measurable space with the following σ-algebra,

$$\mathcal{N} = \sigma\big(\{\{\mu \in N : \mu(B) = k\} \text{ for all } B \in \mathcal{F} \text{ and } k \in \mathbb{N}_0\}\big).$$

Example 2.2.2. Let $(X, \mathcal{F})$ be a measurable space and let $x \in X$. Then, the Dirac measure defined as $\delta_x(A) = \mathbb{1}_{\{x \in A\}}$ is an element of N. For $x_1, x_2, \dots \in X$, the measure $\sum_{i=1}^{\infty} \delta_{x_i}$ is also in N. One may think that any measure in N can be written as $\sum_{i=1}^{k} \delta_{x_i}$ for some $k \in \overline{\mathbb{N}}_0$. This however is not the case.

Indeed, take $X = [0, 1]$ and $\mathcal{F} = \{A \subset [0, 1] : \#A < \infty \text{ or } \#A^c < \infty\}$. Set $\mu(A) = \mathbb{1}_{\{\#A = \infty\}}$ for $A \in \mathcal{F}$. One may check that $(X, \mathcal{F}, \mu)$ is a measure space. Clearly $\mu \in \{0, 1\}$ so $\mu \in N$. However, $\mu(\{x\}) = 0$ for all $x \in [0, 1]$ and thus $\mu \neq \delta_x$ for any x. This problem only came about from the poor choice of σ-algebra on $[0, 1]$ as the σ-algebra was too coarse.

This problem disappears if we work on a finer σ-algebra such as $([0, 1], \mathcal{B}([0, 1]))$. Suppose $\mu \in N$ and $\mu(A) = 1$ for some $A \in \mathcal{B}([0, 1])$. By a standard halving argument, we can find measurable closed intervals $I_i \in A$ such that $\bigcap_{i=1}^{\infty} I_i = \{x\}$ and $\mu(I_i) = 1$ for all i. By continuity from above for measures, we conclude $\mu(\{x\}) = 1$.

Remark 2.2.3. Throughout this thesis, we will be working with random elements in N, where the measures in N are defined on $(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0})$. In this setting, every measure $\mu \in N$ is locally finite and thus repeating the above argument at all $A \in \mathcal{B}(\mathbb{R}_{\geq 0})$ where $\mu(A) > 1$, we see μ can be written as sums of Dirac measures. However, the theory presented in the current section also works on a more general level.

We continue with the definition of a point process, informally a point process on $(X, \mathcal{F})$ is a random variable taking values in the space N. The formal definition follows below.

Definition 2.2.4. Let $(X, \mathcal{F})$ be a measurable space. A *point process* on X is a measure valued random variable η defined on probability space $(\Omega, \mathcal{A}, \mathbb{P})$, taking values in the space N.

Using shorthand notation, $\{\eta(B) = k\} = \{\omega \in \Omega : \eta(\omega)(B) = k\}$, we see η is a point process precisely when $\{\eta(B) = k\} \in \mathcal{A}$ for all $B \in \mathcal{F}$ and $k = \overline{\mathbb{N}}_0$.

Example 2.2.5. Let $(X, \mathcal{F})$ be a measurable space and let Y be a random variable defined on $(\Omega, \mathcal{A}, \mathbb{P})$ taking values in X. Then $\eta = \delta_Y$ is a point process as for $B \in \mathcal{F}$ and $k \in \overline{\mathbb{N}}_0$,

$$\{\eta(B) = k\} = \begin{cases} \{\omega : Y(\omega) \notin B\} & k = 0 \\ \{\omega : Y(\omega) \in B\} & k = 1 \\ \emptyset & k > 1. \end{cases}$$

And $\{\omega : Y(\omega) \notin B\}, \{\omega : Y(\omega) \in B\}, \emptyset \in \mathcal{A}$. Also note $\mathbb{P}(\eta(B) = 1) = \mathbb{P}(Y \in B)$.

Example 2.2.6. Let $(X, \mathcal{F})$ be a measurable space and μ be a probability measure on X. Let $Y_1, \ldots, Y_m$ be i.i.d. random variables with distribution μ. Then, $\eta = \delta_{Y_1} + \cdots + \delta_{Y_m}$ is a point process, since $\{\eta(B) = k\}$ is measurable. Indeed, $\mathbb{P}(\eta(B) = k) = \binom{m}{k}\mu(B)^k(1 - \mu(B))^{m-k}$.

Next, we introduce the Poisson point process, a special type of point process.

Definition 2.2.7. Let $(X, \mathcal{F}, \mu)$ be a measure space with a s-finite measure μ. A point process η on X is a *Poisson point process* with intensity measure μ if,

i) For all $A \in \mathcal{F}$ with $\mu(A) < \infty$, we have $\eta(A) \sim \text{Poi}(\mu(A))$.

ii) For mutually disjoint sets $A_1, \ldots, A_n \in \mathcal{F}$, $\eta(A_1), \ldots, \eta(A_n)$ are independent.

On $\mathbb{R}^d$, we make the distinction between homogeneous and inhomogenous measures. If $\mu = \gamma\lambda$ with λ the Lebesgue measure and $\gamma > 0$, we say η is a *homogeneous* PPP of intensity γ. If $\mu(A) = \int_A f(t)dt$, we say η is an *inhomogeneous* PPP of intensity $f(t)dt$.

Observe that the definition of a PPP only tells us how counts in finite disjoint sets are distributed based on averaging an intensity measure μ over sets. It is not obvious that this uniquely determines a PPP. In principle, there could be a PPP with a different law that produces the same average intensities. Below, we show that this is not the case. The proof relies on the fact that the σ-algebra $\mathcal{N}$ is generated by the events $\{\eta : \eta(A) = k\}$ hence $\mathcal{N}$ consists of events of the form $\{\eta : \eta(A_1) = k_1, \ldots, \mu(A_n) = k_n\}$, which is precisely determined by the intensity measure of a PPP.

Lemma 2.2.8. Let η and η' be two point processes on $(X, \mathcal{F})$. We have,

$$\eta \overset{(d)}{=} \eta' \iff (\eta(A_1), \ldots, \eta(A_k)) \overset{(d)}{=} (\eta'(A_i), \ldots, \eta'(A_k)) \text{ for all } k \in \mathbb{N}_0 \text{ and } A_i \in \mathcal{F}$$

Proof. Suppose η and η' are two point processes with $\eta \overset{(d)}{=} \eta'$. Then $\mathbb{P}(\eta \in A) = \mathbb{P}(\eta' \in A)$ for all $A \in \mathcal{N}$. Note that $\{\mu : \mu(A_1) = k_1, \ldots, \mu(A_n) = k_n\} \in \mathcal{N}$ for all $n \in \mathbb{Z}_{\geq 0}, A_i \in \mathcal{F}$. Thus, $\mathbb{P}(\eta(A_1) = k_1, \ldots, \eta(A_n) = k_n) = \mathbb{P}(\eta'(A_1) = k_1, \ldots, \eta'(A_n = k_n))$ follows immediately.

Assume $(\eta(A_1), \ldots, \eta(A_k)) \overset{(d)}{=} (\eta'(A_i), \ldots, \eta'(A_k))$ for all $k \in \mathbb{Z}_{\geq 0}$ and $A_i \in \mathcal{F}$. We define,

$$P = \{\{\mu : \mu(A_1) = k_1, \ldots, \mu(A_n) = k_n\} \text{ for } n \in \mathbb{N}, k_1, \ldots, k_n \in \mathbb{N}, A_i \in \mathcal{F}\} \subset \mathcal{N}.$$

Then P is closed under intersection and $\sigma(P) = \mathcal{N}$. Furthermore, $\mathbb{P}(\eta \in A) = \mathbb{P}(\eta' \in A)$ for all $A \in P$. Hence by Corollary 2.1.9, η and η' agree on $\sigma(P) = \mathcal{N}$. We conclude $\eta \overset{(d)}{=} \eta'$. ☐

Corollary 2.2.9. Let η and η' be two PPP's with the same intensity measure. Then $\eta \overset{(d)}{=} \eta'$.

This finishes the formal introduction of the Poisson point process and shows that, if such a measure valued random variable exists, is uniquely determined by its intensity measure. The following sections proves the existence of PPP's of arbitrary intensity measure.

2.2.2 Existence of the PPP

Before rigorously proving the existence of the PPP, we give an outline of the proof strategy, by first informally constructing a PPP of finite intensity measure μ and then upgrading this to arbitrary s-finite measures μ.

Let $(X, \mathcal{F}, \mu)$ be a probability space and $\gamma > 0$. To construct a PPP of intensity $\gamma\mu$ on X, we sample $X_1, \ldots, X_N \sim \mu$ i.i.d. where $N \sim \text{Poi}(\gamma)$. We claim $\eta = \sum_{i=1}^{N} \delta_{X_i}$ is a PPP of intensity measure $\gamma\mu$. To see this, for $A \subset X$, we have $\eta(A) \sim \text{Binom}(N, \mu(A)) \sim \text{Poi}(\gamma\mu(A))$. Given that $X_1, \ldots, X_N$ are independent, and the Poisson thinning property, we may expect mutually disjoint sets to have independent counts, making η a PPP of intensity measure $\gamma\mu$ on X. This construction allows for PPP's of arbitrary intensity measure μ as long as $\mu(X) < \infty$.

If $\mu(X) = \infty$, we will use that μ is s-finite. Indeed, suppose $\mu = \sum_{i=1}^{\infty} \mu_i$ with $\mu_i(X) < \infty$. Using the approach above, we may construct PPP's η_i with intensity measure μ_i. It turns out that $\eta = \sum_{i=1}^{\infty} \eta_i$ is a PPP with intensity measure μ. This last claim is known as the superposition principle. In the remainder of this section, we formalize the reasoning in the first paragraph, introduce the superposition principle and show how this ensures the existence of PPP of arbitrary s-finite intensity measures.

Lemma 2.2.10. Let $(X, \mathcal{F}, \mu)$ be a probability space and fix $\gamma > 0$. Suppose that $N \sim \text{Poi}(\gamma)$ and $X_1, \ldots, X_N \sim \mu$ i.i.d. Then $\eta = \sum_{i=1}^{N} \delta_{X_i}$ is a PPP with intensity $\gamma\mu$.

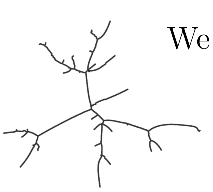
Proof. Let $B_1, \ldots, B_m \in \mathcal{F}$ be pairwise disjoint sets, and define $B_{m+1} = X \setminus \bigcup_{i=1}^{m} B_i$ so that $B_1, \ldots, B_{m+1}$ forms a partition of X. Let $k_1, \ldots, k_m \in N_{\geq 0}$ and set $k = k_1 + \ldots + k_m$. Conditional on $N = n$, $(\eta(B_1), \ldots, \eta(B_{m+1}))$ follows a multinomial distribution. We obtain,

$$\begin{aligned}
\mathbb{P}(\eta(B_1) = k_1, \ldots, \eta(B_m) = k_m) &= \sum_{n=k}^{\infty} \mathbb{P}(N = n)\mathbb{P}(\eta(B_1) = k_1, \ldots, \eta(B_{m+1}) = n - k) \\
&= \sum_{n=k}^{\infty} e^{-\gamma} \frac{\gamma^n}{n!} \frac{n!}{k_1! \ldots k_m! (n-k)!} \mu(B_{m+1})^{n-k} \prod_{i=1}^{m} \mu(B_i)^{k_i}, \\
&= e^{-\gamma} \prod_{i=1}^{m} \frac{(\gamma\mu(B_i))^{k_i}}{k_i!} \sum_{n=k}^{\infty} \gamma^{n-k} \frac{\mu(B_{m+1})^{n-k}}{(n-k)!}, \\
&= e^{-\gamma(1-\mu(B_{m+1}))} \prod_{i=1}^{m} \frac{(\gamma\mu(B_i))^{k_i}}{k_i!}.
\end{aligned}$$

By setting $m = 1$ in the above expression, we get, $\mathbb{P}(\eta(B_1) = k_1) = \frac{(\gamma\mu(B_1))^{k_1}}{k_1!} e^{-\gamma\mu(B_1)}$ and hence $\eta(B_1) \sim \text{Poi}\big(\gamma\eta(B_1)\big)$. By symmetry, we have $\eta(B_i) \sim \text{Poi}(\gamma\mu(B_i))$ for $1 \leq i \leq m$. Independence of $\eta(B_i)$ follows from the fact that the product of marginals equals the joint pmf as seen below,

$$\begin{aligned}
\prod_{i=1}^{m} \mathbb{P}(\eta(B_i) = k_i) = \prod_{i=1}^{m} \frac{(\gamma\mu(B_i))^{k_i}}{k_i!} e^{-\gamma\mu(B_i)} &= e^{\gamma(\mu(B_{m+1})-1)} \prod_{i=1}^{m} \frac{(\gamma\mu(B_i))^{k_i}}{k_i!}, \\
&= \mathbb{P}(\eta(B_1) = k_1, \ldots, \eta(B_m) = k_m).
\end{aligned}$$

We conclude that η is a PPP with intensity $\gamma\mu$. □



This shows that for measurable space $(X, \mathcal{F})$ and probability measure μ, we can find a PPP of intensity $\gamma\mu$. Since $\gamma\mu(X) = \gamma < \infty$, this construction can only construct PPP's of finite intensity measure. However, in practice this often is insufficient. For example a PPP of intensity 1 on $\mathbb{R}$ already cannot be created using this construction, as $\lambda(\mathbb{R}) = \infty$. Next, we introduce the superposition principle for PPP's and show how this mitigates the above limitation.

Lemma 2.2.11. Let $\eta_1, \eta_2, \ldots$ be a sequence of independent Poisson point processes with intensity measures $\lambda_1, \lambda_2, \ldots$ taking values in $(X, \mathcal{F})$. Then,

$$\eta(A) = \sum_{i=1}^{\infty} \eta_i(A) \qquad A \in \mathcal{F},$$

is a PPP with intensity measure $\lambda = \sum_{i=1}^{\infty} \lambda_i$.

Proof. Define $\nu_n(A) = \sum_{i=1}^{n} \mu_i(A)$. We have,

$$\mathbb{P}(\nu_n(A) \le k) = \mathbb{P}\left(\sum_{i=1}^{n} \eta_i(A) \le k\right) = \mathbb{P}\left(\sum_{i=1}^{n} \mathrm{Poi}(\lambda_i(A)) \le k\right) = \mathbb{P}\left(\mathrm{Poi}\left(\sum_{i=1}^{n} \lambda_i\right) \le k\right).$$

Then, by continuity of probability of increasing sets and continuity of the Poisson distribution, we obtain,

$$\begin{aligned}\mathbb{P}(\eta(A) \le k) = \lim_{n\to\infty} \mathbb{P}(\nu_n(A) \le k) &= \lim_{n\to\infty} \mathbb{P}\left(\mathrm{Poi}\left(\sum_{i=1}^{n} \lambda_i\right) \le k\right) \\ &= \mathbb{P}\left(\mathrm{Poi}\left(\sum_{i=1}^{\infty} \lambda_i\right) \le k\right) = \mathbb{P}\left(\mathrm{Poi}\left(\lambda\right) \le k\right),\end{aligned}$$

and thus $\eta(A) \sim \mathrm{Poi}(\lambda(A))$ distributed. Next, let $B_1, \ldots, B_m \in \mathcal{F}$ be mutually disjoint. Then $\eta_i(B_j)$ are independent for all $1 \le j \le m$ and $i \in \mathbb{N}$ and thus by the grouping property of independence, we obtain the independence of,

$$\sum_{i=1}^{\infty} \eta_i(B_1), \ldots, \sum_{i=1}^{\infty} \eta(B_m).$$

This finishes the proof. □

Now we are in a position to prove the general existence of the Poisson point process.

Lemma 2.2.12. Let μ be an s-finite measure on measurable space $(X, \mathcal{F})$. Then there exists PPP on X with intensity measure μ

Proof. Suppose $\mu(X) < \infty$. Let $\gamma = \mu(X)$ and $\lambda(\cdot) = \frac{\mu(\cdot)}{\gamma}$ so that λ is a probability measure. By Lemma 2.2.10, we obtain the existence of a PPP with measure $\gamma \cdot \lambda = \mu$ as desired.

In the case $\mu(X) = \infty$, we use s-finiteness of μ to write $\mu = \sum_{i=1}^{\infty} \mu_i$ with $\mu_i(X) < \infty$ for all i. Then let η_i be a PPP on X with intensity μ_i. By Lemma 2.2.11, we obtain that $\sum_{i=1}^{\infty} \eta_i$ is a PPP of intensity $\sum_{i=1}^{\infty} \mu_i = \mu$ as desired. □

This rigorously shows the existence of the Poisson point process on measurable space $(X, \mathcal{F})$ with s-finite intensity measure. Recall Example 2.2.2 in which we showed that not every nonnegative integer valued measure can be realised as Dirac measure. We show this is not a limitation for PPP's. That is, a PPP of any intensity measure can be obtained as a random sum of dirac measures located at random variables. We formalize this below.

Theorem 2.2.13. Let $(X, \mathcal{F})$ be a measurable space, and μ be an s-finite measure on X. There exists a probability space $(\Omega, A, \mathbb{P})$ supporting random variables $X_1, X_2, \dots$ taking values in X and random variable K taking values in $\overline{\mathbb{N}}_0$ such that,

$$\eta = \sum_{i=1}^{K} \delta_{X_i}$$

is a PPP of intensity μ.

Proof. If $\mu(X) < \infty$, then Lemma 2.2.10 gives us the desired result with $N \sim \text{Poi}(\mu(X))$ and $X_1, \dots, X_N \sim \frac{\mu}{\mu(X)}$, which, using Theorem 2.1.15, can be defined on the countably infinite product space,

$$\left(\mathbb{N}_0, 2^{\mathbb{N}_0}, \mathbb{P}_N\right) \times \prod_{i=1}^{\infty} \left(X, \mathcal{F}, \frac{\mu}{\mu(X)}\right),$$

where $\mathbb{P}_N$ is the Poisson probability measure. I.e. $\mathbb{P}_N(\{n\}) = e^{-\mu(X)} \frac{(\mu(X))^n}{n!}$.

In case $\mu(X)$ is infinite, we use that μ is s-finite and decompose $\mu = \sum_{i=1}^{\infty} \mu_i$ with $\mu_i(X) < \infty$. Define $\lambda_i = \frac{\mu_i}{\mu_i(X)}$. For each i, we construct probability space,

$$(\Omega_i, \mathcal{A}_i, \mathbb{P}_i) = \left(\mathbb{N}_0, 2^{\mathbb{N}_0}, \mathbb{P}_{K_i}\right) \times \prod_{i=1}^{\infty} (X, \mathcal{F}, \lambda_i),$$

on which $K_i \sim \text{Poi}(\mu_i(X))$ and $X_{i1}, X_{i2}, \dots \sim \lambda_i$ are all independently defined. Let,

$$(\Omega, \mathcal{A}, \mathbb{P}) = \prod_{i=1}^{\infty} (\Omega_i, \mathcal{A}_i, \mathbb{P}_i).$$

On this product space, define the random variable $K = \sum_{i=1}^{\infty} K_i$ and $n(i,j) = \sum_{l=1}^{i-1} K_i + j$, which both are measurable (as countable sums of random variables are measurable). Lastly, we relabel the random variables $Y_{n(i,j)} = X_{ij}$. We see,

$$\sum_{i=1}^{K} Y_i = \sum_{i=1}^{\infty} \sum_{j=1}^{K_i} \delta_{X_{ij}} = \sum_{i=1}^{\infty} \eta_i = \eta,$$

where η_i is a PPP of intensity μ_i and η is a PPP of intensity $\sum_{i=1}^{\infty} \mu_i = \mu$ as desired. □

Remark 2.2.14. Although formally the Poisson point process is a measure-valued random variable, the above theorem also justifies the common view that the PPP is a collection of random points. Indeed, for a PPP of any intensity, we can find a countable collection of random points such that the sum of Dirac measures at those points share the same intensity measure and thus by Lemma 2.2.8 are equal in distribution. This is not to say that any PPP is actually a sum of Dirac measures, but if we talk on the level of distributions, there is no difference.

2.2.3 The PPP as Renewal Process

Theorem 2.2.13 tells us that a PPP of arbitrary intensity can be realized as sum of Dirac measures. In this section, we explore some implications of this, in particular when we work with PPP's on $(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}))$. We first introduce some natural definitions.

Definition 2.2.15. Let $(X, \mathcal{F})$ be a measurable space and let suppose there exist X-valued random variables $X_1, X_2, \dots$ and $\overline{\mathbb{N}}_0$ valued random variable K such that,

$$\eta = \sum_{i=1}^{K} \delta_{X_i}$$

is a PPP of intensity measure μ on $(X, \mathcal{F})$. We call $X_1, X_2, \dots$ the *points* of η. Note that the points are random variables. We also write $x \in \eta$ for the event that $\eta(\{x\}) > 0$.

If $(X, \mathcal{F}) = (\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}))$ and there are random variables $0 < C_1 < C_2 < \cdots \in \mathbb{R}_{\geq 0}$ so that,

$$\eta = \sum_{i=1}^{\infty} \delta_{C_i}$$

is a PPP of intensity μ on $(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}))$, then $0 < C_1 < \cdots < C_k$ are the *first k points* of η.

In this section, we will find random variables $0 < C_1 < C_2 < \dots$ such that $\eta = \sum_{i=1}^{\infty} \delta_{C_i}$ is a PPP of intensity 1 on $(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}))$. We then generalize this construction to PPP's of more general intensity measure on $(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}))$ and use this to find the distribution of the first k ordered points for PPP's of intensity $f(t)dt$.

Lemma 2.2.16. Let $T_i \sim \text{Exp}(1)$ be i.i.d. and let $C_i = T_1 + \cdots + T_i$ for $i \in \mathbb{N}$. We have,

$$\eta = \sum_{i=1}^{\infty} \delta_{C_i},$$

is a PPP on $\big(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0})\big)$ of intensity 1.

Proof. We first define $N_a = \#\{C_i \text{ s.t. } C_i < a\}$ to count the number of points C_i in the interval $[0, a]$. For $b > a$, we have $\mathbb{P}(N_a = m \text{ and } N_b = m + k) = \frac{e^{-a} a^m}{m!} \cdot \frac{e^{-(b-a)}(b-a)^k}{k!}$. Indeed,

$$N_a = m \text{ and } N_b = m + k \iff \begin{cases} C_m \leq a, \\ a - C_m \leq T_{m+1} \leq b - C_m, \\ T_{m+2} + \cdots + T_{m+k} \leq b - (C_m + T_{m+1}), \\ b - (C_m + T_{m+1} + \cdots + T_{m+k}) \leq T_{m+k+1}. \end{cases}$$

Recall that summing n i.i.d. Exp(1) random variables yields a Gamma$(n,1)$ random variable. Thus C_m follows a Gamma$(m,1)$ distribution, and $T_{m+2}+\cdots+T_{m+k}$ follows a Gamma$(k-1,1)$ distribution. Also note that $C_m, T_{m+1}, T_{m+2}+\cdots+T_{m+k}$ and T_{m+k+1} are independent. Thus,

$$\mathbb{P}(N_b = m+k, N_a = m) = \int_0^a \int_{a-t_1}^{b-t_1} \int_0^{b-t_1-t_2} \int_{b-t_1-t_2-t_3}^{\infty} f(t_1,t_2,t_3,t_4)dt_4dt_3dt_2dt_1,$$

where $f(t_1,t_2,t_3,t_4) = \frac{t_1^{m-1}e^{-t_1}}{(m-1)!}e^{-t_2}\frac{t_3^{k-2}e^{-t_3}}{(k-2)!}e^{-t_4}$. Some computations show,

$$\begin{aligned}
&\int_0^a \int_{a-t_1}^{b-t_1} \int_0^{b-t_1-t_2} \int_{b-t_1-t_2-t_3}^{\infty} \frac{t_1^{m-1}e^{-t_1}}{(m-1)!}e^{-t_2}\frac{t_3^{k-2}e^{-t_3}}{(k-2)!}e^{-t_4}dt_4dt_3dt_2dt_1 \\
&= \int_0^a \int_{a-t_1}^{b-t_1} \int_0^{b-t_1-t_2} \frac{t_1^{m-1}}{(m-1)!}\frac{t_3^{k-2}}{(k-2)!}e^{-b}dt_3dt_2dt_1 \\
&= \int_0^a \int_{a-t_1}^{b-t_1} \frac{t_1^{m-1}}{(m-1)!}\frac{(b-t_1-t_2)^{k-1}}{(k-1)!}e^{-b}dt_2dt_1 \\
&= \int_0^a \int_0^{b-a} \frac{t_1^{m-1}}{(m-1)!}\frac{u^{k-1}}{(k-1)!}e^{-b}dudt_1 \\
&= \frac{a^m}{m!}\frac{(b-a)^k}{k!}e^{-b} \\
&= \frac{a^m e^{-a}}{m!} \cdot \frac{(b-a)^k e^{-(b-a)}}{k!}.
\end{aligned}$$

This implies that,

$$\mathbb{P}(N_b = m+k \mid N_a = m) = \frac{(b-a)^k e^{-(b-a)}}{k!}.$$

Hence, $\eta([a,b]) \sim \text{Poi}(b-a)$, this being independent of the number of points prior to the interval $[a,b]$. Basic inductive reasoning implies that for independent intervals $[a_1,b_1], \ldots [a_n,b_n]$ we have that $\eta([a_1,b_1]), \ldots, \eta([a_n,b_n])$ are independent.

This shows that *i)* and *ii)* in Definition 2.2.7 hold for all closed intervals in $\mathbb{R}_{\geq 0}$. For η to be a PPP, this needs to hold for all $A \in \mathcal{B}(\mathbb{R})$.

To show this, one typically shows that given $\eta([0,T]) = n$, the points $C_1, \ldots, C_k$ are the order statistics of uniformly distributed in $[0,T]$. Once this is established, a proof identical to that of Lemma 2.2.10 can be used to obtain that $\eta(A) \sim \text{Poi}(\lambda(A))$ and disjoint sets giving independent counts. We leave the details to [13, Section 5.3.5]. □

The above construction coincides with a PPP on $\mathbb{R}_{\geq 0}$ with Lebesgue intensity measure. Next, we show how we can obtain PPP on the positive real line with different intensity measures by applying transformations to the above point process.

Lemma 2.2.17. Let $(X,\mathcal{F}_X), (Y,\mathcal{F}_Y)$ be measurable spaces and η be a PPP of intensity measure μ on $(X,\mathcal{F}_X)$. Let $f : X \to Y$ be measurable. Then $f(\eta) = \eta \circ f^{-1}$ is a PPP of intensity measure $\mu \circ f^{-1}$ on $(Y,\mathcal{F}_Y)$.

Proof. Let $(X, \mathcal{F}_X), (Y, \mathcal{F}_Y), \eta$ and f be as above. Fix $A \in \mathcal{F}_Y$. Then,

$$f(\eta)(A) = \eta(f^{-1}(A)) \sim \text{Poi}(\lambda(f^{-1}(A))) = \text{Poi}((\lambda \circ f^{-1})(A)).$$

Lastly, $A_1, \dots, A_n$ mutually disjoint implies $f^{-1}(A_1), \dots, f^{-1}(A_n)$ mutually disjoint, hence,

$$f(\eta))(A_1), \dots, f(\eta)(A_n) = \eta(f^{-1}(A_1)), \dots, \eta(f^{-1}(A_n)),$$

are independent. This shows that $f(\eta)$ is a PPP of intensity measure $\lambda \circ f^{-1}$ on $(Y, \mathcal{F}_Y)$ $\square$

Example 2.2.18. Let η be a PPP of intensity 1 on $\mathbb{R}_{\geq 0}$ and let $f(x) = \sqrt{2x}$. Then $f(\eta)$ is a PPP with intensity tdt. Indeed, note that $f^{-1}(x) = \frac{1}{2}x^2$. Hence,

$$(\lambda \circ f^{-1})(A) = \int_{f^{-1}(A)} dt = \int_A tdt.$$

Thus the intensity measure of $f(\eta)$ is indeed $(\lambda \circ f^{-1})(t) = tdt$.

Remark 2.2.19. Let $T_i \sim \text{Exp}(1)$ be i.i.d. and $C_i = T_1 + \dots + T_i$. Then, $\eta = \sum_{i=1}^{\infty} \delta_{C_i}$ is a PPP of intensity 1 (Lemma 2.2.16). We have that $f(\eta) \overset{(d)}{=} \sum_{i=1}^{\infty} \delta_{f(C_i)}$ and thus $\sum_{i=1}^{\infty} \delta_{f(C_i)}$ is a PPP of intensity measure $\lambda \circ f^{-1}$ for λ the Lebesgue measure.

Example 2.2.20. Suppose $f(t) : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is measurable and set,

$$T = \{(x, y) \in \mathbb{R}^2 : x \geq 0, 0 \leq y \leq f(x)\}.$$

Let η be a PPP of intensity 1 on $(T, \mathcal{B}(T))$ and let $g : T \to \mathbb{R}_{\geq 0}$ be given by $g(x, y) = x$. By Lemma 2.2.17, $g(\eta)(A)$ is a PPP of intensity $\lambda \circ g^{-1}$ on $(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}))$ for λ the Lebesgue measure. In particular $g(\eta)$ is a PPP of intensity $f(t)dt$ since,

$$(\lambda \circ g^{-1})(A) = \int \int_{g^{-1}(A)} dxdy = \int_A f(t)dt.$$

In particular, if we order the points of η as $\{(C_1, B_1), (C_2, B_2), \dots\}$ with $0 < C_1 < C_2, \dots$ then $(C_1, \dots, C_k)$ follow the same law as the first k ordered points of a PPP of intensity $f(t)dt$ and $B_i \sim \text{Unif}([0, f(C_i)])$. The last claim is intuitive but can be formalized using Marked point processes. We link [12, Section 7.2].

We now proceed by giving the distribution of the first k points of a PPP on $(\mathbb{R}_{\geq 0}, \mathcal{B}(R_{\geq 0}))$. First for PPP's of intensity 1 and later for arbitrary intensity measures.

Lemma 2.2.21. Let η be a PPP of intensity 1 on $\mathbb{R}_{\geq 0}$ and let $0 < C_1 < C_2 < \dots < C_k$ be the first k points of η. Then $(C_1, \dots, C_k)$ has pdf $f_{C_1,\dots,C_k}(s_1, \dots, s_k) = e^{-s_k}$ for $s_1 < \dots < s_k$.

Proof. Recall from the construction of the homogeneous PPP on $\mathbb{R}_{\geq 0}$ that if $C_1 < \dots < C_k$ denote the first k ordered points of a PPP, then we have $\vec{C} = A\vec{T}$ where,

$$\vec{C} = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{pmatrix}, \qquad A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \qquad \vec{T} = \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_k \end{pmatrix}, \text{ with } T_i \sim \text{Exp}(1).$$

This linear transformation is invertible with inverse,

$$A^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

Observe that $f_{\vec{T}} = e^{-(t_1+\cdots+t_k)}$, thus we obtain,

$$f_{\vec{C}}(s_1, \ldots, s_k) = f_{\vec{T}}(A^{-1}(s_1, \ldots, s_k))\det(A^{-1}) = e^{-s_k} \cdot 1,$$

as desired. $\square$

We now upgrade this proof to the probability density function of the first k ordered points of PPP's of intensity $f(t)dt$.

Lemma 2.2.22. Let η be a PPP of intensity $f(t)dt$ on $(\mathbb{R}_{\geq 0}, \mathcal{B}(\mathbb{R}_{\geq 0}))$ for some increasing f and let $0 < C_1 < \cdots < C_k$ be the first k ordered points of η. Then,

$$f_{C_1,\ldots,C_k}(s_1, \ldots, s_k) = \prod_{i=1}^{k} f(s_i) \exp\left(-\int_0^{s_k} f(t)dt\right), \qquad \text{where } s_1 < \cdots < s_k.$$

Proof. Using 2.2.17, we aim to find g such that $(\lambda \circ g^{-1})(A) = \int_A f dt$ as then $g(\eta)$ is a homogeneous PPP. We define $F(x) = \int_0^x f(t)dt$, which is invertible since f is increasing. Then,

$$(\lambda \circ F)(A) = \int_{F(A)} dt = \int_A F'(t)dt = \int_A f(t)dt.$$

Thus we take $g = F^{-1}$. If $0 < X_1 < \cdots < X_k$ are the first k points of a PPP of intensity 1, then $C_1 = F^{-1}(X_1), \ldots, C_k = F^{-1}(X_k)$ are the first k points of η with intensity $f(t)dt$. We define $g(X_1, \ldots, X_k) = (F^{-1}(X_1), \ldots F^{-1}(X_k))$ so $g^{-1}(C_1, \ldots, C_k) = (F(C_1), \ldots, F(C_k))$. We get,

$$\det\left(\frac{\partial g^{-1}(x_1, \ldots, x_k)}{\partial(x_1, \ldots, x_k)}\Big|_{(x_1,\ldots,x_k)=(s_1,\ldots,s_k)}\right) = \prod_{i=1}^{k} f(s_i).$$

Thus,

$$\begin{aligned} f_{C_1,\ldots,C_k}(s_1, \ldots, s_k) &= f_{X_1,\ldots,X_k}(g^{-1}(s_1, \ldots, s_k))\det\left(\frac{\partial g^{-1}(x_1, \ldots, x_k)}{\partial(x_1, \ldots, x_k)}\Big|_{(x_1,\ldots,x_k)=(s_1,\ldots,s_k)}\right), \\ &= \prod_{i=1}^{k} f(s_i) \exp\left(-\int_0^{s_k} f(t)dt\right), \end{aligned}$$

as desired. $\square$

This concludes the section on Poisson point processes. We continue with a description of the Gromov–Hausdorff–Prokhorov topology.

2.3 The Gromov–Hausdorff–Prokhorov Topology

In essence, the Gromov–Hausdorff–Prokhorov topology is a topology on the space of compact measure metric spaces. More precisely, the Hausdorff distance compares distances between two subsets in a common metric space. The Gromov–Hausdorff metric assign distances between metric spaces by isometrically embedding them in an optimal common space and then assigning distances using the Hausdorff metric. Lastly, the Prokhorov distance is a metric on measures defined on a common metric space. Below, we formalize these notations.

2.3.1 The Gromov–Hausdorff Distance

We start with introducing the Gromov–Hausdorff distance, a metric that assigns distances between compact metric spaces.

Definition 2.3.1. Let (X, d) be a metric space, and let $A, B \subseteq X$ be two non-empty subsets. The *Hausdorff distance* $d_H(A, B)$ between A and B is defined as:

$$d_H(A, B) = \max\left\{\sup_{a\in A} d(a, B),\ \sup_{b\in B} d(A, b)\right\}.$$

Where the distance between a point a and set B is given by $d(a, B) = \inf_{b\in B} d(a, b)$.

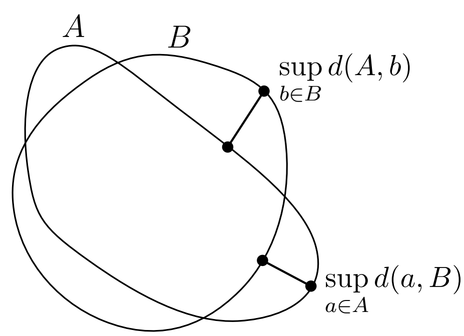


Figure 12: Hausdorff distance

This notion is useful for comparing subsets inside a common metric space, but cannot be used to compute distance between two metric spaces. To extend this definition to compute distances between two metric spaces, we first embed them in an optimal common space, and then compare them using the Hausdorff distance defined above.

Definition 2.3.2. Let (X, d_X) and (Y, d_Y) be two compact metric spaces. The *Gromov–Hausdorff distance* $d_{GH}(X, Y)$ between X and Y is defined as:

$$d_{GH}(X, Y) = \inf_{Z,\phi,\psi} d_H^Z(\phi(X), \psi(Y)),$$

where the infimum is taken over all metric spaces Z and all isometric embeddings $\phi : X \to Z$ and $\psi : Y \to Z$. Here, d_H^Z denotes the Hausdorff distance in the space Z. If Z is clear from context, we simply write d_H instead of d_H^Z.

Remark 2.3.3. We claim the Gromov–Hausdorff distance is a metric on the space of isometry classes of compact metric spaces. However, first note that we define the Gromov–Hausdorff distance using *compact* metric spaces. It is intuitive that we want totally bounded metric spaces as unbounded metric spaces lead to infinite Gromov–Hausdorff distance. However, completeness of the metric space is also important as is illustrated with the following example. If we compare the metric spaces $(0,1)$ and $[0,1]$ (say, both with Euclidean distance) then clearly $d_{GH}((0,1),[0,1]) = 0$. However, $(0,1)$ and $[0,1]$ are not isometric. Restricting to complete metric spaces removes this problem.

Theorem 2.3.4. The Gromov–Hausdorff distance is a pseudo-metric on the space of compact metric spaces and is a metric on the space of isometry classes of compact metric spaces.

Proof. We refer to [5, Theorem 7.3.30] for a proof. □

An equivalent definition of the Gromov–Hausdorff distance can be stated in terms of correspondences and distortions.

Definition 2.3.5. Let (X, d_X) and (Y, d_Y) be two compact metric spaces. We say $R \subseteq X \times Y$ is a *correspondence* if for all $x \in X$ there exists at least one $y \in Y$ with $(x, y) \in R$, and vice versa.

The *distortion* of a correspondence R is defined as,

$$\mathrm{dis}(R) = \sup_{(x,y),(x',y') \in R} |d_X(x, x') - d_Y(y, y')| \, ,$$

where the supremum is taken over all pairs of elements in the distortion.

Theorem 2.3.6. We have,

$$d_{GH}(X, Y) = \frac{1}{2} \inf_R \mathrm{dis}(R).$$

where the infimum is taken over all correspondences R between X and Y.

Proof. This is a standard result and can be found in [5, Theorem 7.3.25]. □

2.3.2 The Prokhorov Distance

Before continuing with the Gromov–Hausdorff–Prokhorov distance, we define the Prokhorov distance between two measures defined on a common metric space.

Definition 2.3.7. Let (M, d) be a metric space and $A \subset M$. Then,

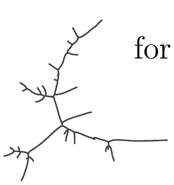
$$A^\epsilon = \{x \in M \text{ s.t. } d(x, A) < \epsilon\},$$

is called the *ϵ-thickening* of A.

Definition 2.3.8. Let (M, d) be a metric space with associated Borel sigma algebra $\mathcal{B}(M)$. Let $\mathcal{P}(M)$ denote the set of all probability measures on $(M, \mathcal{B}(M))$. Then,

$$d_P(\mu, \nu) = \inf\{\epsilon > 0 : \mu(A) \leq \nu(A^\epsilon) + \epsilon \text{ for all } A \in \mathcal{B}(M)\},$$

for $\mu, \nu \in \mathcal{P}(M)$ is called the *Prokhorov distance* between μ and ν.



Remark 2.3.9. To motivate this definition, suppose μ is a metric in some measure metric space. There are two ways we can perturb μ: We can slightly change the sets to which μ gives measure, and slightly change the mass size μ gives to each set. The ϵ enlarging of A allows for the former, while adding ϵ accounts for the latter.

This is different from total variation distance, which accounts for perturbations in measures given to sets, but does not differentiate between distances of sets in the underlying metric space. This makes total variation distance flexible, as it does not have to be defined on metric spaces (after all, it does not need a notion of two sets being close). However, total variation distance is not the right metric to go with Gromov–Hausdorff distance, precisely since it cannot differentiate between sets that are close in terms of Hausdorff distance.

Theorem 2.3.10. The function d_P is a metric on $\mathcal{P}(M)$.

Proof. Below, we check symmetry, positivity and the triangle inequality,

i) For showing symmetry, suppose that $\mu(A) \leq \nu(A^\epsilon) + \epsilon$ for some $\epsilon > 0$ and all $A \in \mathcal{B}(M)$. Then $\nu(A) = 1 - \nu\left((\overline{A^\epsilon})^\epsilon\right) \leq 1 - \mu(\overline{A^\epsilon}) + \epsilon = \mu(A^\epsilon) + \epsilon$, as $\overline{A} = \left(\overline{A^\epsilon}\right)^\epsilon$ and $\overline{A^\epsilon} \in \mathcal{B}(M)$.

ii) It is clear that $d_P(\mu, \nu) \geq 0$ and $d_P(\mu, \mu) = 0$ for all $\mu, \nu \in \mathcal{P}(M)$. Suppose $d_P(\mu, \nu) = 0$. Then, for all $A \in \mathcal{B}(M)$, $\mu(A) \leq \nu(A^\epsilon) + \epsilon$. Taking A closed, and letting $\epsilon \to 0$, we see that $\mu(A) \leq \nu(A)$ for all closed $A \in \mathcal{B}(M)$. By symmetry, $\mu(A) \geq \nu(A)$. Since $\mathcal{B}(M)$ is generated by the open (and hence closed) sets, we may conclude $\mu = \nu$.

iii) For the triangle inequality, suppose $d_P(\mu, \nu) \leq \epsilon_1$ and $d_P(\nu, \pi) \leq \epsilon_2$. For all $A \in \mathcal{B}(M)$,

$$\mu(A) \leq \nu(A^{\epsilon_1}) + \epsilon_1 \leq \pi\left((A^{\epsilon_1})^{\epsilon_2}\right) + \epsilon_1 + \epsilon_2 = \pi(A^{\epsilon_1+\epsilon_2}) + \epsilon_1 + \epsilon_2.$$

We conclude, $d_P(\mu, \pi) \leq d_P(\mu, \nu) + d_P(\nu, \pi)$.

This shows that d_P is a metric on $\mathcal{P}(M)$. □

If the metric space (M, d) is Polish (separable and complete), then $(\mathcal{P}(M), d_P)$ inherits the same property.

Theorem 2.3.11. Let (M, d) be a Polish metric space. Then $(\mathcal{P}(M), d_P)$ is also Polish.

Proof. We reference [7, Appendix 2.5.III]. □

We introduce two useful lemmas that bound the Prokhorov distance between two measures.

Lemma 2.3.12. Let (M, d) be a metric space with $\mu, \nu \in \mathcal{P}(M)$. If $X \sim \mu, Y \sim \nu$ are two random variables defined on a common space such that $\mathbb{P}(d(X, Y) > \epsilon) < \epsilon$, then $d_P(\mu, \nu) < \epsilon$.

Proof. Let $A \in \mathcal{B}(M)$. Then,

$$\begin{aligned} \mu(A) = \mathbb{P}(X \in A) &= \mathbb{P}(X \in A, d(X, Y) < \epsilon) + \mathbb{P}(X \in A, d(X, Y) \geq \epsilon), \\ &\leq \mathbb{P}(Y \in A^\epsilon) + \epsilon, \\ &= \mu(A^\epsilon) + \epsilon \end{aligned}$$

Since this holds for all $A \in \mathcal{B}(M)$, we conclude $d_P(\mu, \nu) < \epsilon$. □

Lemma 2.3.13. Let μ, ν be two measures, and suppose $K_1, \dots, K_N$ is a partition of the support of μ such that $\operatorname{diam}(K_i) < \epsilon$ for all $i = 1, \dots, N$. We have,

$$d_P(\mu, \nu) \le \max\left\{\epsilon, \sum_{i=1}^{N} |\mu(K_i) - \nu(K_i)|\right\} \le \epsilon + \sum_{i=1}^{N} |\mu(K_i) - \nu(K_i)|.$$

Proof. Let A be measurable, and define $I = \{i \in \{1, \dots, N\} \text{ s.t. } K_i \cap A \neq \emptyset\}$. We have,

$$\begin{aligned}
\mu(A) &\le \sum_{i \in I} \mu(K_i) \\
&\le \sum_{i \in I} (\mu(K_i) - \nu(K_i)) + \sum_{i \in I} \nu(K_i) \\
&\le \sum_{i \in I} |\mu(K_i) - \nu(K_i)| + \nu\left(\bigcup_{i \in I} K_i\right) \\
&\le \sum_{i=1}^{N} |\mu(K_i) - \nu(K_i)| + \nu(A^\epsilon).
\end{aligned}$$

Where $\sum_{i \in I} \nu(K_i) = \nu(\bigcup_{i \in I} K_i)$ by disjointness of the K_i's and $\nu(\bigcup_{i \in I} K_i) \le \nu(A^\epsilon)$ follows since $\bigcup_{i \in I} K_i \subset A^\epsilon$. This shows $d_P(\mu, \nu) \le \max\left\{\epsilon, \sum_{i=1}^{N} |\mu(K_i), \nu(K_i)|\right\}$ as desired. □

2.3.3 The Gromov–Hausdorff–Prokhorov Distance

It remains to combine the Gromov–Hausdorff distance and the Prokhorov distance to obtain a metric on the space of compact measure metric spaces.

Definition 2.3.14. Let (X, d_X, μ_X) and (Y, d_Y, μ_Y) be two compact measure metric spaces. The *Gromov–Hausdorff–Prokhorov distance*, or GHP-distance for short, is defined as,

$$d_{GHP}(X, Y) = \inf_{\varphi, \psi, Z} \left\{\max\left(d_H^Z(\varphi(X), \psi(Y)), d_P^Z(\mu_X \circ \phi^{-1}, \mu_Y \circ \psi^{-1})\right)\right\}.$$

The infimum is over all metric spaces Z and isometric embeddings $\varphi : X \to Z$, $\psi : Y \to Z$ and d_H^Z, d_P^Z are the Hausdorff and Prokhorov distance in Z respectively.

Remark 2.3.15. Let (Y, d, μ_Y), $(X, d|_X, \mu_X)$ be two measure metric spaces for $X \subset Y$ and probability measures μ_X, μ_Y. Then $d_{GHP}(X, Y) \le \max(d_Y(X, Y), d_P(\mu_X, \mu_Y))$. This follows immediately by taking $Z = Y$ and $\phi : X \to Y, \psi : Y \to Y$ to be the identity embeddings.

In the case $X \subset Y$, one might intuitively expect the identity embeddings to always be optimal, which would turn the above inequality into an equality. This however is not true. For example, take $X = [0, 1]$ and $Y = [0, 5]$, both with normalized Lebesgue measure. Then $d_Y(X, Y) = 4$. To bound $d_{GH}(X, Y)$, take $Z = [0, 5], \phi(x) = x + 2$ and $\psi(x) = x$. This shows, $d_{GH}(X, Y) \le 2$. Also, $d_P^Z(\mu_X \circ \phi^{-1}, \mu_Y \circ \psi^{-1}) \le 1$ for any choice Z, ϕ, ψ. Hence we may conclude, $d_{GHP}(X, Y) = \max(d_Y(X, Y), d_P(\mu_X, \mu_Y))$ cannot hold for all $X \subset Y$.

Similar to the Gromov–Hausdorff distance, we can restate the Gromov–Hausdorff–Prokhorov distance in terms of correspondences.

Definition 2.3.16. Let $(X, \mathcal{F}_X, \mu_X), (Y, \mathcal{F}_Y, \mu_Y)$ be two measure spaces, let π be a measure on the product space $(X \times Y)$ and let p_i be the projection to i'th coordinate. We define,

$$D(\pi; \mu_X, \mu_Y) = ||\pi \circ p_1^{-1} - \mu_X||_{TV} + ||\pi \circ p_2^{-1} - \mu_Y||_{TV},$$

where $||\mu - \nu||_{TV} = \sup\{|\mu(A) - \nu(A)| : A \subset X \text{ measurable}\}$ for μ, ν measures on measurable space $(X, \mathcal{F}_X)$.

Theorem 2.3.17. Let $(X, d_X, \mu_X), (Y, d_Y, \mu_Y)$ be compact measure metric spaces. Then,

$$d_{GHP}(X, Y) = \inf_{R,\pi} \left\{ \max \left(\frac{1}{2}\text{dis}(R), D(\pi; \mu_X, \mu_Y) + \pi(R^c) \right) \right\},$$

where the infimum is taken over all correspondences R between X and Y and all measures on $X \times Y$. Lastly, R^c denotes the complement of R in $X \times Y$.

Proof. A proof of the result can be found in [10, Theorem 3.6]. □

Remark 2.3.18. Note that $D(\pi; \mu_X, \mu_Y) = 0$ when $\pi(A, Y) = \mu_X(A)$ and $\pi(X, B) = \mu_Y(B)$, for all $A \in \mathcal{F}_X, B \in \mathcal{F}_Y$. Such a measure π on the product space $X \times Y$ is called a *coupling*.

Example 2.3.19. Let $L_n = (V, E)$ with $V = [n]$ and $E = \{\{i, i+1\} : i \in [n-1]\}$ be the line graph on n vertices and Let d_n be the graph distance. For $A \subset V(L_n)$, define $\mu_n(A) = \frac{\#A}{n}$ to be the uniform probability measure on the vertices of L_n. Define metric space $\big([0,1], d\big)$ with d Euclidean distance and let λ be the Lebesgue measure on $[0, 1]$. Then,

$$(L_n, n^{-1}d_n, \mu_n) \xrightarrow[n\to\infty]{} ([0, 1], d, \lambda),$$

in the GHP topology. In Figure 13, we visualize how $n^{-1}L_n$ approximates $[0, 1]$ by n equally spaced points.

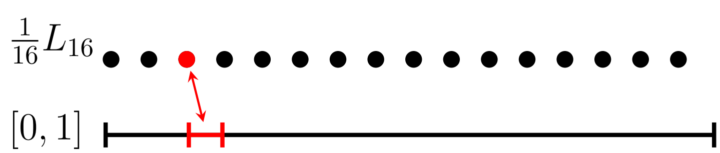


Figure 13: In red, we visualize the idea behind the correspondence R_n: we pair to each vertex in L_n the corresponding segment in R_n.

We make the correspondence described in Figure 13 concrete as subset of $L_n \times [0, 1]$,

$$R_n = \left\{ (x, y) \in L_1 \times [0, 1] : x \in [n] \text{ and } \frac{x-1}{n} \le y \le \frac{x}{n} \right\}.$$

Then for arbitrary $(x, y), (x', y') \in R_n$, note that $n^{-1}d(x, x') - d(y, y') \le 2n^{-1}$ and thus we found a correspondence with $\text{dis}(R_n) \xrightarrow[n\to\infty]{} 0$. Hence $n^{-1}L_n \xrightarrow[n\to\infty]{} [0, 1]$ in the GH-topology.

We first give some intuition into the convergence of μ_n to λ. Let $A \subset [0, 1]$ be an interval and let $B_n = \{x \in L_n : (x, y) \in R_n \text{ for } y \in A\}$ be the set approximating A in L_n. It is intuitive that $\mu_n(B_n) \to \lambda(A)$ as the left side computes the proportion of B_n in L_n and the right side the proportion of A in $[0,1]$. This is the intuition behind GHP-convergence. The correspondence used for GH-convergence gives us a way to approximate sets in $[0,1]$ by sets in $n^{-1}L_n$. The measures μ_n converge to λ since μ_n converges on approximations of sets.

Below we make this reasoning formal by finding a measure π_n on $L_n \times [0,1]$ such that we have, $D(\pi_n; \mu_X, \mu_Y) \to 0$ and $\pi_n(R_n^c) \to 0$. For this, define,

$$\pi_n(A, B) = \lambda\left(\{y \in [0,1] : y \in B, (x,y) \in R_n \text{ for some } x \in A\}\right),$$

where this is well defined since R^n is measurable in the product space. Note that $\pi_n(R_n^c) = 0$ for all n since $\pi_n(R_n^c) = \lambda(\emptyset)$ by construction. Then,

$$\pi_n(L_n, B) = \lambda(\{y \in B : (x,y) \in R_n \text{ for some } x \in L_n\}) = \lambda(B).$$

Then lastly,

$$\begin{aligned}\pi_n(A, [0,1]) &= \lambda(\{y \in [0,1] : (x,y) \in R_n \text{ for some } x \in A\}), \\ &= \sum_{i \in A} \lambda\left(\left[\frac{i-1}{n}, \frac{i}{n}\right]\right) = \frac{\#A}{n} = \mu_n(A)\end{aligned}$$

Thus, we have shown that $D(\pi; \mu_n, \lambda) = 0$ as π is a coupling of μ_n and λ. This concludes the example and shows $(L_n, n^{-1}d_n, \mu_n) \xrightarrow[n\to\infty]{} ([0,1], d, \lambda)$ in the GHP-topology.

2.4 Formalization of Proof Strategy

In this section, we justify the proof outline as given in Section 1.5. We formalize how the convergence of finite dimensional distributions (Theorem 1.5.1), together with tightness arguments (Corollary 1.5.4) are sufficient to deduce our main result, Theorem 1.5.2.

Lemma 2.4.1. Let $X_1, X_2, \ldots$ and X be random variables taking values on metric space M. The following are equivalent,

i) X_n converges in distribution to X,

ii) $\limsup_{n\to\infty} \mathbb{P}(X_n \in A) \leq \mathbb{P}(X \in A)$ for all closed sets $A \subset M$.

Proof. This is Portmanteua's Lemma, we reference [8, Theorem 3.2.11]. □

We use Portmanteau's lemma to prove the following,

Lemma 2.4.2. Suppose that $X_{n,k}, X_n, X_k$ and X are random variables living in the same metric space and for all $\epsilon > 0$,

$$X_{n,k} \xrightarrow[n\to\infty]{d} X_k, \quad \lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(d(X_{n,k}, X_n) > \epsilon) = 0 \text{ and } \lim_{k\to\infty} \mathbb{P}(d(X_k, X) > \epsilon) = 0,$$

then we have $X_n \xrightarrow[n\to\infty]{d} X$.

Proof. First observe that using Lemma 2.4.1, it suffices to show,

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in A) \leq \mathbb{P}(X \in A) \text{ for all } A \text{ closed.}$$

Now observe that for all $\epsilon > 0$, and A closed, we have,

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}(X_n \in A) &\leq \limsup_{n\to\infty} \mathbb{P}(X_{n,k} \in A^\epsilon) + \limsup_{n\to\infty} \mathbb{P}(d(X_n, X_{n,k}) > \epsilon),\\ &\leq \lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(X_{n,k} \in A^\epsilon),\\ &\leq \lim_{k\to\infty} \mathbb{P}(X_k \in A^\epsilon),\end{aligned}$$

where the last step follows from Lemma 2.4.1 together with the assumption $X_{n,k} \xrightarrow{d} X_k$ and the observation that A^ϵ is closed. Similarly, for all B and all $\delta > 0$ we have,

$$\lim_{k\to\infty} \mathbb{P}(X_k \in B) \leq \mathbb{P}(X \in B^\delta) + \lim_{k\to\infty} \mathbb{P}(d(X_k, X) > \delta) = \mathbb{P}(X \in B^\delta).$$

By using $B = A^\epsilon$, we combine both bounds into,

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in A) \leq \mathbb{P}(X \in A^{\epsilon+\delta}), \text{ for all } A \text{ closed.}$$

Let $r = \epsilon + \delta$. Since the result holds for all $\epsilon, \delta > 0$, we let $r \to 0$ and thus $A^r \downarrow A$ (Recall that A is closed, thus for all $\left(r_i\right)_{i\in\mathbb{N}}$ with $r_i \to 0$, we have $\bigcap_{i=1}^\infty A^{(r_i)} = A$). By continuity of $\mathbb{P}$ from above, we find,

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in A) \leq \mathbb{P}(X \in A), \text{ for all } A \text{ closed,}$$

as desired. □

Proposition 2.4.3. Suppose we have convergence $\left(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}} d_n, \nu_n^{(k)}\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}^{(k)}, d, \mu^{(k)}\right)$ in GHP-topology, and for all $\epsilon > 0$, we have,

$$i) \lim_{t\to\infty} \mathbb{P}\left(d_H\left(\mathcal{T}(t), \mathcal{T}\right) > \epsilon\right) = 0, \qquad ii) \lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(d_H\left(\mathcal{T}_n(tn^{\frac{1}{2}}), \mathcal{T}_n\right) > \epsilon\right) = 0,$$

$$iii) \lim_{k\to\infty} \mathbb{P}\left(d_P\left(\mu^{(k)}, \mu\right) > \epsilon\right)\right) = 0, \qquad iv) \lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(d_P\left(\nu_n^{(k)}, \nu_n\right) > \epsilon\right) = 0.$$

Then, $\left(\mathcal{T}_n, n^{-\frac{1}{2}} d_n, \nu_n\right) \xrightarrow[n\to\infty]{} (\mathcal{T}, d, \mu)$ in the GHP-topology.

Proof. We aim to use Lemma 2.4.2. To this end, we define,

$$X = (\mathcal{T}, d, \mu), \ X_n = \left(\mathcal{T}_n, n^{-\frac{1}{2}} d_n, \nu_n\right), \ X_k = \left(\mathcal{T}^{(k)}, d, \mu^{(k)}\right) \text{ and } X_{k,n} = \left(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}} d_n, \nu_n^{(k)}\right).$$

Thus we need to show that for all $\epsilon > 0$,

$$\lim_{k\to\infty} \mathbb{P}(d_{GHP}(X_k, X) > \epsilon) = 0 \text{ and } \lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(d_{GHP}(X_{k,n}, X_n) > \epsilon) = 0.$$

Observe $\mathcal{T}^{(k)} \subset \mathcal{T}$. Hence $d_{GHP}(X_k, X) \leq \max\left(d_H(\mathcal{T}^{(k)}, \mathcal{T}), d_P(\mu^{(k)}, \mu)\right)$, (Remark 2.3.15). By an application of the union bound, we obtain,

$$\begin{aligned}\mathbb{P}(d_{GHP}(X_k, X) > \epsilon) &\leq \mathbb{P}(\max(d_H(\mathcal{T}^{(k)}, \mathcal{T}), d_P(\mu^{(k)}, \mu)) > \epsilon), \\ &\leq \mathbb{P}(d_H(\mathcal{T}^{(k)}, \mathcal{T}) > \epsilon) + \mathbb{P}(d_P(\mu^{(k)}, \mu) > \epsilon).\end{aligned}$$

By taking limits and substituting $iii)$ we see,

$$\lim_{k\to\infty} \mathbb{P}(d_{GHP}(X_k, X) > \epsilon) \leq \lim_{k\to\infty} \mathbb{P}(d_H(\mathcal{T}^{(k)}, \mathcal{T}) > \epsilon).$$

Thus it suffices to show $\mathbb{P}(d_H(\mathcal{T}^{(k)}, \mathcal{T}) > \epsilon) \to 0$ as $k \to \infty$. This statement follows from $i)$. Indeed, assuming $i)$ we have that for any $\delta > 0$, we can pick $T > 0$ such that,

$$\mathbb{P}(d_H(\mathcal{T}(T), \mathcal{T}) > \epsilon) < \frac{\delta}{2}.$$

Let K be large enough so that $\mathbb{P}(\eta([0,T]) > K) < \frac{\delta}{2}$ for η a PPP of intensity tdt. On the event $\{\eta([0,T]) \leq K\}$, we have $\mathcal{T}(T) \subset \mathcal{T}^{(K)}$ as $C_K > T$. Thus by another application of the union bound,

$$\mathbb{P}(d_H(\mathcal{T}^{(K)}, \mathcal{T}) > \epsilon) \leq \mathbb{P}(\eta([0,T]) > K) + \mathbb{P}(d_H(\mathcal{T}(T), \mathcal{T}) > \epsilon) < \delta.$$

This shows that $i)$ implies $\lim_{k\to\infty} \mathbb{P}(d_H(\mathcal{T}^{(k)}, \mathcal{T}) > \epsilon) = 0$ for all $\epsilon > 0$. We conclude that for all $\epsilon > 0$,

$$\lim_{k\to\infty} \mathbb{P}(d_{GHP}(X_k, X) > \epsilon) = 0.$$

An analogous argument shows that for all $\epsilon > 0$,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(d_{GHP}(X_{k,n}, X_n) > \epsilon) = 0.$$

Thus we may invoke Lemma 2.4.2 to get,

$$(\mathcal{T}_n, n^{-\frac{1}{2}} d_n, \nu_n) \xrightarrow[n\to\infty]{d} (\mathcal{T}, d, \mu),$$

in the GHP-topology. This finishes the proof. □

Remark 2.4.4. For now we consider $\mathcal{T}$ to be the tree constructed via stick-breaking where the length of the sticks are determined by a PPP of intensity tdt. During this thesis, we also consider various other intensities of PPP's and their corresponding trees. The above proof is sufficient in the more general cases as long as $\lim_{k\to\infty} \mathbb{P}(\eta([0,T] > k)) = 0$. All PPP's considered in this thesis satisfy this criteria.

Proposition 2.4.3 gives us the precise layout of the proof that the CRT is the scaling limit of uniform labeled rooted trees, and justifies the proof layout given in Section 1.5.

3 The Finite Dimensional Distribution

This section is aimed at proving Theorem 1.5.1.

3.1 Convergence of the Repeat and Attachment Points

Recall that for $S_n \in_u [n]^{n-1}$, we write,

$$C_i^n \text{ is the index of the } i\text{'th repeat} \quad \text{and} \quad B_i^n = \min\{l \in [n] \text{ s.t. } S_n(l) = S_n(C_k^n)\}.$$

In Section 1.3, we saw a heuristic for the fact that the scaled repeat points $n^{-\frac{1}{2}}\left(C_1^n, \dots, C_k^n\right)$ jointly converge to the k first ordered points of a Poisson point process of intensity tdt. In this section, we make this convergence formal. Furthermore, we saw a heuristic that B_k^n is almost uniform on $\{1, \dots, C_k^n - 1\}$. Thes we may expect $n^{-\frac{1}{2}} \xrightarrow[n\to\infty]{d} \text{Unif}([0, C_k])$.

This section is dedicated to proving,

Theorem 3.1.1. For any $k \in \mathbb{N}$, we have,

$$n^{-\frac{1}{2}}(C_1^n, \dots, C_k^n, B_1^n, \dots, B_k^n) \xrightarrow[n\to\infty]{d} (C_1, \dots, C_k, B_1, \dots, B_k),$$

where $0 < C_1 < \dots < C_k$ are the first k points of a PPP of intensity tdt and $B_i \sim \text{unif}([0, C_i])$.

Remark 3.1.2. It suffices to show that for $0 < s_1 < \dots < s_k$ and $t_1, \dots, t_k \in [0, 1]$, we have,

$$\begin{aligned}
&i)\ \mathbb{P}\Big(C_1^n \le s_1 n^{1/2}, \dots, C_k^n \le s_k n^{1/2}\Big) \xrightarrow[n\to\infty]{} \mathbb{P}\big(C_1 \le s_1, \dots, C_k \le s_k\big),\\
&ii)\ \mathbb{P}\Big(B_1^n \le t_1 C_1^n, \dots, B_k \le t_k C_k^n \ \Big|\ C_1^n \le s_1 n^{1/2}, \dots, C_k^n \le s_{k-1} n^{1/2}\Big),\\
&\xrightarrow[n\to\infty]{} \mathbb{P}\Big(B_1 \le t_1 C_1, \dots, B_k \le t_k C_k \ \Big|\ C_1 \le s_1, \dots, C_k^n \le s_{k-1}\Big).
\end{aligned}$$

We start with $i)$, for which we introduce the following result.

Lemma 3.1.3. For all $k \in \mathbb{N}$ and $0 < s_1 < \dots < s_k$, we have,

$$n^{\frac{k}{2}} \mathbb{P}\Big(C_1^n = \lfloor s_1 n^{\frac{1}{2}} \rfloor, \dots, C_k^n = \lfloor s_k n^{\frac{1}{2}} \rfloor\Big) \xrightarrow[n\to\infty]{\text{u.c.}} f_{(C_1,\dots,C_k)}(s_1, \dots, s_k),$$

where $f_{C_1,\dots,C_k}(s_1, \dots, s_k)$ is the joint pdf of the first k points in a PPP of intensity tdt. Here u.c. denotes that the convergence is uniform over compact sets.

Proof. Recall from Lemma 2.2.22 that $f_{C_1,\dots C_k}(s_1, \dots, s_k) = s_1 \dots s_k e^{-\frac{s_k^2}{2}}$. We apply induction on k. For $k = 1$, observe $C_1^n = \lfloor s_1 n^{\frac{1}{2}} \rfloor$ precisely when the first $\lfloor s_1 n^{\frac{1}{2}} \rfloor - 1$ entries of S_n are unique, and entry $\lfloor s_1 n^{\frac{1}{2}} \rfloor$ is a repeat. Note that,

$$\mathbb{P}\Big(S_n(i) \text{ is a repeat} \ \Big|\ S_n(1), \dots, S_n(i-1) \text{ are not repeats}\Big) = \frac{i-1}{n},$$

since there are $i - 1$ distinct entries in $S_n(1), \dots, S_n(i-1)$.

Thus,

$$n^{\frac{1}{2}}\mathbb{P}\big(C_1^n = \lfloor s_1 n^{\frac{1}{2}}\rfloor\big) = n^{\frac{1}{2}}\frac{\lfloor s_1 n^{\frac{1}{2}}\rfloor - 1}{n}\prod_{i=1}^{\lfloor s_1 n^{\frac{1}{2}}\rfloor - 2}\Big(1-\frac{i}{n}\Big),$$

$$= \frac{\lfloor s_1 n^{\frac{1}{2}}\rfloor - 1}{n^{\frac{1}{2}}}\exp\left(\sum_{i=1}^{\lfloor s_1 n^{\frac{1}{2}}\rfloor - 2}\log\Big(1-\frac{i}{n}\Big)\right).$$

Since $\frac{\lfloor s_1 n^{\frac{1}{2}} - 1\rfloor}{n^{\frac{1}{2}}} \xrightarrow[n\to\infty]{\text{u.c.}} s_1$, it only remains to show that,

$$\sum_{i=1}^{\lfloor s_1 n^{\frac{1}{2}}\rfloor - 2}\log\Big(1-\frac{i}{n}\Big) \xrightarrow[n\to\infty]{\text{u.c.}} -\frac{s_1^2}{2}.$$

For this, we use the Taylor expansion $\log(1-x) = -\sum_{j=1}^{\infty}\frac{x^j}{j}$ for $|x|<1$ to obtain,

$$\sum_{i=1}^{\lfloor s_1 n^{\frac{1}{2}}\rfloor - 2}\log\Big(1-\frac{i}{n}\Big) = -\frac{1}{n}\sum_{i=1}^{\lfloor s_1 n^{\frac{1}{2}}\rfloor - 1} i - \sum_{i=1}^{\lfloor s_1 n^{\frac{1}{2}}\rfloor - 1}\sum_{j=2}^{\infty}\Big(\frac{i}{n}\Big)^j\frac{1}{j}.$$

Using that $\sum_{i=1}^{k} i = \frac{k(k+1)}{2}$, we obtain that the first sum converges uniformly over compacts to $-\frac{s_1^2}{2}$. For the second sum, we have,

$$\sum_{i=1}^{\lfloor s_1 n^{\frac{1}{2}}\rfloor - 1}\sum_{j=2}^{\infty}\Big(\frac{i}{n}\Big)^j\frac{1}{j} \le s_1 n^{\frac{1}{2}}\sum_{j=2}^{\infty}\Big(\frac{s_1}{n^{\frac{1}{2}}}\Big)^j = s_1^3 n^{-\frac{1}{2}}\sum_{j=0}^{\infty}\Big(\frac{s_1}{n^{\frac{1}{2}}}\Big)^j = s_1^3 n^{-\frac{1}{2}}\frac{1}{1-\frac{s_1}{n^{\frac{1}{2}}}} \xrightarrow[n\to\infty]{\text{u.c.}} 0,$$

which concludes the base case of the induction proof. Before continuing we ease notation by writing,

$$C_k^n(s) = \big\{C_1^n = \lfloor s_1 n^{\frac{1}{2}}\rfloor, \dots, C_k^n = \lfloor s_k n^{\frac{1}{2}}\rfloor\big\}, \qquad \text{where } s = (s_1,\dots,s_l) \text{ with } l > k,$$

so that the induction hypothesis reads, $n^{\frac{k}{2}}\mathbb{P}(C_k^n(s)) \xrightarrow[n\to\infty]{\text{u.c.}} s_1\dots s_k e^{-\frac{s_k^2}{2}}$. Then, for the induction step, we see,

$$n^{\frac{k+1}{2}}\mathbb{P}\big(C_{k+1}^n(s)\big) = n^{\frac{1}{2}}\mathbb{P}\big(C_{k+1}^n = \lfloor s_{k+1}n^{\frac{1}{2}}\rfloor \;\big|\; C_k^n(s)\big) n^{\frac{k}{2}}\mathbb{P}\big(C_k^n(s)\big),$$

By the induction hypothesis, $n^{\frac{k}{2}}\mathbb{P}\big(C_k^n(s)\big) \xrightarrow[n\to\infty]{\text{u.c.}} s_1\dots s_k e^{-\frac{s_k^2}{2}}$, so that it suffices to show,

$$n^{\frac{1}{2}}\mathbb{P}\big(C_{k+1}^n = \lfloor s_{k+1}n^{\frac{1}{2}}\rfloor \;\big|\; C_k^n(s)\big) \xrightarrow[n\to\infty]{\text{u.c.}} s_{k+1}e^{-\frac{s_{k+1}^2 - s_k^2}{2}}.$$

For this, we follow identical steps to the base case. Observe $\big\{C_{k+1}^n = \lfloor s_{k+1}n^{\frac{1}{2}}\rfloor \;\big|\; C_k^n(s)\big\}$ happens precisely when $S_n(i)$ is not a repeat for $i = \lfloor s_k n^{\frac{1}{2}}\rfloor + 1$ to $\lfloor s_{k+1}n^{\frac{1}{2}}\rfloor - 1$ and $S_n(\lfloor s_{k+1}n^{\frac{1}{2}}\rfloor)$

is a repeat. There are precisely $\lfloor s_k n^{\frac{1}{2}} \rfloor - 1 - k$ unique entries in $S_n(1), \dots, S_n(\lfloor s_k n^{\frac{1}{2}} \rfloor)$ and thus $\mathbb{P}(S_n(i) \text{ is a repeat}) = \frac{\lfloor s_k n^{\frac{1}{2}} - 1 - k + i \rfloor}{n}$ for $i = C_k^n + 1, \dots, C_{k+1}^n - 1$. For notational clarity, write $I_n = \{i \in \mathbb{N} : \lfloor s_k n^{\frac{1}{2}} \rfloor + 1 \leq i \leq \lfloor s_{k+1} n^{\frac{1}{2}} \rfloor - 1\}$. We compute,

$$
\begin{aligned}
n^{\frac{1}{2}} \mathbb{P}\big(C_{k+1}^n = \lfloor s_{k+1} n^{\frac{1}{2}} \rfloor \bigm| C_k^n(s)\big) &= \frac{\lfloor s_{k+1} n^{\frac{1}{2}} \rfloor - 1 - k}{n^{\frac{1}{2}}} \prod_{i \in I_n} \left(1 - \frac{i - k - 1}{n}\right), \\
&= s_{k+1} \exp\left(O(n^{-1/2}) - \frac{1}{n} \sum_{i \in I_n} \big(i - k - 1\big)\right) + o(1),
\end{aligned}
$$

where we again used the Taylor expansion $\log(1-x)$. Also note that $\frac{\lfloor s_{k+1} n^{\frac{1}{2}} \rfloor - k - 1}{n^{\frac{1}{2}}} \xrightarrow[n \to \infty]{\text{u.c.}} s_{k+1}$ so that $o(1)$ is uniformly small on compact sets,

$$
\begin{aligned}
\frac{1}{n} \sum_{i \in I_n} \big(i - k - 1\big) &= \frac{1}{n} \left(-|I_n|(1+k) + \sum_{i \in I_n} i\right), \\
&= \frac{1}{n} \left(-O\big(n^{\frac{1}{2}}\big) + \sum_{i=1}^{\lfloor s_{k+1} n^{\frac{1}{2}} \rfloor - 1} i - \sum_{i=1}^{\lfloor s_k n^{\frac{1}{2}} \rfloor} i\right), \\
&= o(1) + \frac{\big(\lfloor s_{k+1} n^{\frac{1}{2}} \rfloor - 1\big) \lfloor s_{k+1} n^{\frac{1}{2}} \rfloor}{2n} - \frac{\lfloor s_{k+1} n^{\frac{1}{2}} \rfloor \big(\lfloor s_{k+1} n^{\frac{1}{2}} \rfloor + 1\big)}{2n}, \\
&= \frac{s_{k+1}^2 - s_k^2}{2} + o(1),
\end{aligned}
$$

with all convergence being uniform over compact sets. This shows that,

$$
n^{\frac{1}{2}} \mathbb{P}\big(C_{k+1}^n = \lfloor s_{k+1} n^{\frac{1}{2}} \rfloor \bigm| C_k^n(s)\big) \xrightarrow[n \to \infty]{\text{u.c.}} s_{k+1} e^{-\frac{s_{k+1}^2 - s_k^2}{2}},
$$

and concludes the induction proof. □

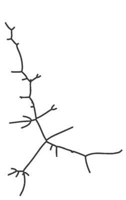
In order to use this result to prove *i*) in Remark 3.1.2, we introduce the following lemma.

Lemma 3.1.4. Let $K \subset \mathbb{R}^k$ be compact and $A, A_n \subset K$ be Borel measurable sets. Furthermore, suppose $\lambda(A_n \Delta A) \to 0$ as $n \to \infty$ and assume $g_n : K \to \mathbb{R}$ is measurable and converges uniformly over compact sets to measurable and integrable $g : K \to \mathbb{R}$. Then,

$$
\int_{A_n} g_n(x) dx \xrightarrow[n \to \infty]{} \int_A g(x) dx
$$

Proof. Let A^* be the closure of $A \cup \bigcup_{i=1}^{\infty} A_n$ so that A^* is compact (here we use that A_n, A are bounded by a compact set K). We have,

$$
\int_{A_n} g_n(x) dx = \int_{A_n} \big(g_n(x) - g(x)\big) dx + \int_{A_n} g(x) dx.
$$



Since $\lambda(A_n \Delta A) \to 0$, we have $\int_{A_n} g(x)dx \to \int_A g(x)dx$. Similarly,

$$\int_{A_n} |g_n(x) - g(x)|dx \leq \int_{A^*} |g_n(x) - g(x)|dx \xrightarrow[n\to\infty]{} 0,$$

since $g_n \to g$ uniformly on A^*. Thus $\int_{A_n} \big(g_n(x) - g(x)\big)dx \to 0$, concluding the proof. $\square$

Using the lemma above, we show,

Lemma 3.1.5. For $0 < s_1 < \cdots < s_k$, we have,

$$\mathbb{P}\Big(C_1^n \leq s_1 n^{1/2}, \dots, C_k^n \leq s_{k-1} n^{1/2}\Big) \xrightarrow[n\to\infty]{} \mathbb{P}\big(C_1 \leq s_1, \dots, C_k \leq s_k\big).$$

Proof. Using Lemma 3.1.4 together with Lemma 3.1.3, we get,

$$\begin{aligned}
\mathbb{P}\big(C_1^n \leq s_1 n^{\frac{1}{2}}, \dots, C_k^n \leq s_k n^{\frac{1}{2}}\big) &= \sum_{x_1=1}^{\lfloor s_1 n^{\frac{1}{2}}\rfloor} \sum_{x_2=x_1+1}^{\lfloor s_2 n^{\frac{1}{2}}\rfloor} \cdots \sum_{x_k=x_{k-1}+1}^{\lfloor s_k n^{\frac{1}{2}}\rfloor} \mathbb{P}\big(C_1^n = x_1, \dots C_k^n = x_k\big), \\
&= \int_1^{\lfloor s_1 n^{\frac{1}{2}}\rfloor} \int_{x_1+1}^{\lfloor s_2 n^{\frac{1}{2}}\rfloor} \cdots \int_{x_{k-1}+1}^{\lfloor s_k n^{\frac{1}{2}}\rfloor} \mathbb{P}\big(C_1^n = \lfloor x_1 \rfloor, \dots C_k^n = \lfloor x_k \rfloor\big) dx_k \dots dx_2 dx_1, \\
&= \int_{n^{-\frac{1}{2}}}^{\lfloor s_1 n^{\frac{1}{2}}\rfloor n^{-\frac{1}{2}}} \int_{y_1+n^{-\frac{1}{2}}}^{\lfloor s_2 n^{\frac{1}{2}}\rfloor n^{-\frac{1}{2}}} \cdots \int_{y_{k-1}+n^{-\frac{1}{2}}}^{\lfloor s_k n^{\frac{1}{2}}\rfloor n^{-\frac{1}{2}}} n^{\frac{k}{2}} \mathbb{P}\big(C_1^n = \lfloor y_1 n^{\frac{1}{2}} \rfloor, \dots C_k^n = \lfloor y_k n^{\frac{1}{2}} \rfloor\big) dy_k \dots dy_2 dy_1, \\
&\xrightarrow[n\to\infty]{} \int_0^{s_1} \int_{y_1}^{s_2} \cdots \int_{y_{k-1}}^{s_k} y_1 \dots y_k e^{-\frac{y_k^2}{2}} dy_k \dots dy_2 dy_1, \\
&= \mathbb{P}(C_1 \leq s_1, \dots, C_k \leq s_k),
\end{aligned}$$

where we used the substitution $x_i = n^{\frac{1}{2}} y_i$ and recognize $y_1 \dots y_k e^{-\frac{y_k^2}{2}}$ as the pdf of the first k ordered points of a PPP of intensity tdt. This finishes the proof of $i)$ in Remark 3.1.2. $\square$

We continue with a proof of $ii)$.

Lemma 3.1.6. For $t_1, \dots, t_k \in [0, 1]$ and $0 < s_1 \cdots < s_k$, we have,

$$\mathbb{P}\big(B_1^n \leq t_1 C_1^n, \dots, B_k^n \leq t_k C_k^n \;\big|\; C_1^n = \lfloor x_1 n^{\frac{1}{2}} \rfloor, \dots, C_k^n = \lfloor x_k n^{\frac{1}{2}} \rfloor\big) \xrightarrow[n\to\infty]{\text{u.c.}} t_1 \dots t_k$$

Proof. We introduce new shorthand notation,

$$B_k^n(t) = \big\{B_1^n \leq t_1 C_1^n, \dots, B_k^n \leq t_k C_k^n\big\}, \qquad \text{where } t = (t_1, \dots t_l) \text{ with } l \geq k.$$

To show $\mathbb{P}\big(B_k^n(t) \mid C_k^n(x)\big) \xrightarrow[n\to\infty]{\text{u.c.}} t_1 \dots t_k$, we condition on the event that the first k repeats are unique. That is, we condition on $P_k^n = \big\{S_n(C_i^n) \neq S_n(C_j^n) : \text{ for } i, j \in \{1, \dots, k\}, i \neq j\big\}$. Conditional on P_k^n, we have,

$$S_n(C_i^n) \in_u S = \big\{S_n(1), \dots, S_n(C_i^n - 1)\big\} \setminus \big\{S_n(C_1^n), \dots, S_n(C_{i-1}^n)\big\},$$

for all $i \in \{1,2,\dots,k\}$. Since S (as multiset) does not contain repeats under P_k^n, we conclude that B_i^n is uniformly distributed on $\{1,2,\dots,C_i^n-1\} \setminus \{C_1^n,\dots,C_{i-1}^n\}$ and thus,

$$\mathbb{P}\big(B_i^n \le t_i C_i^n \bigm| C_k^n(x), B_{i-1}^n(t), P_k^n\big) = \frac{\lfloor t_i C_i^n \rfloor - \#\{j : C_j^n \le t_i C_i^n\}}{C_i - i}.$$

Given that $0 \le \#\{j : C_j^n \le t_i C_i^n\} \le i-1$ and $C_i^n = \lfloor x_i n^{\frac{1}{2}} \rfloor$ we obtain,

$$\frac{\lfloor t_i \lfloor x_i n^{\frac{1}{2}} \rfloor \rfloor - (i-1)}{\lfloor x_i n^{\frac{1}{2}} \rfloor - i} \le \mathbb{P}\big(B_i^n \le t_i C_i^n \bigm| C_k^n(x), B_{i-1}^n(t), P_k^n\big) \le \frac{\lfloor t_i \lfloor x_i n^{\frac{1}{2}} \rfloor \rfloor}{\lfloor x_i n^{\frac{1}{2}} \rfloor - i}.$$

Thus we find, $\mathbb{P}\big(B_i^n \le t_i C_i^n \bigm| C_k^n(x), B_{i-1}^n(t), P_k^n\big) \xrightarrow[n\to\infty]{\text{u.c.}} t_i$ as both bounds converge to t_i uniformly over compacts. We conclude,

$$\mathbb{P}\big(B_k(t) \bigm| C_k^n(x), P_k^n\big) = \prod_{i=1}^{k} \mathbb{P}\big(B_i^n \le t_i C_i^n \bigm| C_k^n(x), B_{i-1}^n(t), P_k^n\big) \xrightarrow[n\to\infty]{\text{u.c.}} t_1 \dots t_k.$$

Thus it suffices to show $\mathbb{P}\big(P_k^n\big) \to 1$ as $n \to \infty$. For this let $\epsilon > 0$ and apply a union bound,

$$\begin{aligned}\mathbb{P}\big((P_k^n)^c\big) &\le \mathbb{P}\left(\left\{C_1^n < \epsilon n^{\frac{1}{2}}\right\} \cup \bigcup_{i=1}^{k} \left\{\exists j \in [k]\backslash\{i\} : S_n(C_j^n) = S_n(C_i^n), C_1^n \ge \epsilon n^{\frac{1}{2}}\right\}\right), \\ &\le \mathbb{P}\big(C_1^n < \epsilon n^{\frac{1}{2}}\big) + \sum_{i=1}^{k} \mathbb{P}\big(\exists j \in [k]\backslash\{i\} : S_n(C_j^n) = S_n(C_i^n) \bigm| C_1^n > \epsilon n^{\frac{1}{2}}\big).\end{aligned}$$

From Lemma 3.1.5, we have $\mathbb{P}(C_1^n < \epsilon n^{\frac{1}{2}}) \xrightarrow[n\to\infty]{} \mathbb{P}(C_1 < \epsilon)$. Furthermore, for any fixed i, j we get, $\mathbb{P}\left(S_n(C_i^n) = S_n(C_j^n), i \ne j\right) \le \frac{1}{\epsilon n^{\frac{1}{2}} - 1}$ as there are at least $C_j^n - j - 1 \ge C_1^n - 1 \ge \epsilon n^{\frac{1}{2}} - 1$ distinct entries in $S_n(1), \dots, S_n(C_j^n - 1)$ and we need $S_n(C_j^n)$ to be the unique value equaling $S_n(C_i^n)$. By another union bound we get,

$$\mathbb{P}\left(\exists j \in [k]\backslash\{i\} : S_n(C_j^n) = S_n(C_i^n) \bigm| C_1^n > \epsilon n^{\frac{1}{2}}\right) \le (k-1)\frac{1}{\epsilon n^{\frac{1}{2}} - 1}$$

Hence for all $\epsilon > 0$, we have,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\big((P_k^n)^c\big) &\le \lim_{n\to\infty} \mathbb{P}\big(C_1^n < \epsilon n^{\frac{1}{2}}\big) + \sum_{i=1}^{k} \mathbb{P}\left(\left\{S_n(C_j^n) \ne S_n(C_i^n), j \in [k]\backslash\{i\} \bigm| C_1^n > \epsilon n^{\frac{1}{2}}\right\}\right), \\ &\le \mathbb{P}\big(C_1 < \epsilon\big) + k \lim_{n\to\infty} (k-1)\frac{1}{\epsilon n^{\frac{1}{2}}} = \mathbb{P}(C_1 < \epsilon).\end{aligned}$$

By letting $\epsilon \to 0$, we see $\mathbb{P}\big((P_k^n)^c\big) \xrightarrow[n\to\infty]{} 0$ and thus $\mathbb{P}(P_k^n) \xrightarrow[n\to\infty]{} 1$, concluding the proof. $\square$

We are now in a position to prove *ii)* in Remark 3.1.2.

Lemma 3.1.7. For $t_1, \dots, t_k \in [0,1]$ and $0 < s_1 < \cdots < s_k$, we have,

$$\mathbb{P}\big(B_1^n \le t_1 C_1^n, \dots, B_k \le t_k C_k^n \;\big|\; C_1^n \le s_1 n^{1/2}, \dots, C_k^n \le s_k n^{\frac{1}{2}}\big) \xrightarrow[n\to\infty]{} \mathbb{P}\big(B_1 \le t_1 C_1, \dots, B_k \le t_k C_k \;\big|\; C_1 \le s_1, \dots, C_k \le s_k\big).$$

Proof. By using the total law for conditional probabilities followed by the same approach as the proof of Lemma 3.1.5, we see,

$$\begin{aligned}
&\mathbb{P}\big(B_k^n(t) \;\big|\; C_1^n \le s_1 n^{\frac{1}{2}}, \dots, C_k^n \le s_k n^{\frac{1}{2}}\big),\\
&= \sum_{x_1=1}^{\lfloor s_1 n^{\frac{1}{2}} \rfloor} \cdots \sum_{x_k=x_{k-1}+1}^{\lfloor s_k n^{\frac{1}{2}} \rfloor} \frac{\mathbb{P}\big(B_k^n(t) \;\big|\; C_1^n = x_1, \dots, C_k^n = x_k\big)\mathbb{P}\big(C_1^n = x_1, \dots, C_k^n = x_k\big)}{\mathbb{P}\big(C_1^n \le s_1 n^{\frac{1}{2}}, \dots, C_k^n \le s_k n^{\frac{1}{2}}\big)},\\
&= \int_{a_1}^{b_1} \cdots \int_{a_k}^{b_k} \frac{\mathbb{P}\big(B_k^n(t) \;\big|\; C_k^n(x)\big) n^{\frac{k}{2}} \mathbb{P}\big(C_k^n(x)\big)}{\mathbb{P}\big(C_1^n \le s_1 n^{\frac{1}{2}}, \dots, C_k^n \le s_k n^{\frac{1}{2}}\big)} dx_k \dots dx_1,
\end{aligned}$$

where $a_1 = n^{-\frac{1}{2}}, a_i = x_i + n^{-\frac{1}{2}}$ for $i \in \{2, \dots, k\}$ and $b_i = n^{-\frac{1}{2}}\big(\lfloor s_i n^{\frac{1}{2}} \rfloor + 1\big)$ for $i \in [k]$. Furthermore, by Lemma 3.1.3, 3.1.6 and 3.1.5, we have,

$$\frac{\mathbb{P}\big(B_k^n(t) \;\big|\; C_k^n(x)\big) n^{\frac{k}{2}} \mathbb{P}\big(C_k^n(x)\big)}{\mathbb{P}\big(C_1^n \le s_1 n^{\frac{1}{2}}, \dots, C_k^n \le s_k n^{\frac{1}{2}}\big)} \xrightarrow[n\to\infty]{\text{u.c.}} \frac{t_1 \dots t_k f_{C_1,\dots,C_k}(x_1, \dots, x_k)}{\mathbb{P}\big(C_1 \le s_1, \dots, C_k \le s_k\big)}.$$

By Lemma 3.1.4 we conclude,

$$\begin{aligned}
&\mathbb{P}\big(B_k^n(t) \;\big|\; C_1^n \le s_1 n^{\frac{1}{2}}, \dots, C_k^n \le s_k n^{\frac{1}{2}}\big),\\
&\qquad \xrightarrow[n\to\infty]{} \int_0^{s_1} \cdots \int_{x_{k-1}}^{s_k} \frac{t_1 \dots t_k f_{C_1,\dots,C_k}(x_1, \dots, x_k)}{\mathbb{P}\big(C_1 \le s_1, \dots, C_k \le s_k\big)} dx_k \dots dx_1,\\
&\qquad = \mathbb{P}\big(B_1 \le t_1 C_1, \dots, B_k \le t_k C_k \;\big|\; C_1 \le s_1, \dots, C_k \le s_k\big),
\end{aligned}$$

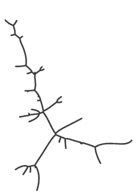
as desired. □

This shows that for all $0 < s_1 < \cdots < s_k$ and $t_1, \dots, t_k \in [0,1]$, we have

$$\mathbb{P}\big(C_1^n \le s_1 n^{\frac{1}{2}}, \dots C_k^n \le s_k n^{\frac{1}{2}}, B_1^n \le t_1 C_1^n, \dots, B_k^n \le t_k C_k^n\big) \xrightarrow[n\to\infty]{} \mathbb{P}\big(C_1 \le s_1, \dots, C_k \le s_k, B_1 \le t_1 C_1, \dots, B_k \le t_k C_k\big),$$

and hence finishes the proof of Theorem 3.1.1. □

Remark 3.1.8. We can visualize the points $\big(C_i, B_i\big)_{i\in\mathbb{N}}$ as a two dimensional point process. Indeed, let $\eta = \sum_{i=1}^{\infty} \delta_{(C_i,B_i)}$. By Example 2.2.20, we know that η is a PPP of intensity 1 on the region $T = \{(x,y) \in \mathbb{R}^2 : 0 \le x, 0 \le y \le x\}$. In particular, this means that for large n, the points $n^{-\frac{1}{2}}\big(C_1^n, B_1^n\big)$ are roughly distributed as a homogeneous PPP of intensity 1 on T. See Figure 14.



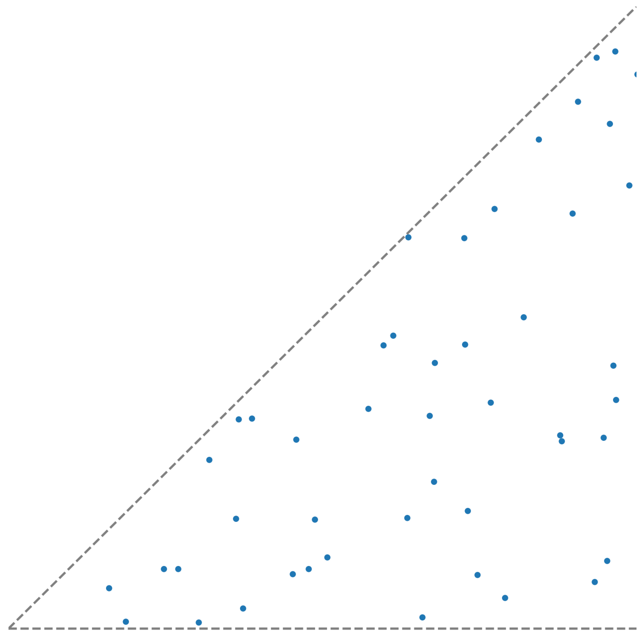


Figure 14: The points $n^{-\frac{1}{2}}\big(C_i^n, B_i^n\big)$ for a sequence S_n with $n = 5000$.

3.2 Convergence of the Partial Trees $\mathcal{T}_n^{(k)} \to \mathcal{T}^{(k)}$

In the prior section, we saw,

$$n^{-\frac{1}{2}}\big(C_1^n, \ldots, C_k^n, B_1^n, \ldots B_k^n\big) \xrightarrow[n\to\infty]{d} \big(C_1, \ldots, C_k, B_1, \ldots, B_k\big).$$

Since C_i^n denotes the endpoint of the i'th stick in $\mathcal{T}_n$, and C_i denotes the endpoint of the i'th stick in $\mathcal{T}$, this means that the length of the first k sticks in $(\mathcal{T}_n, n^{-\frac{1}{2}}d_n)$ converge in distribution to the length of the first k sticks in $(\mathcal{T}, d)$. Furthermore, in both cases the attachment points are roughly uniform over the already constructed trees. Thus we should expect that for any k, the trees $(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}}d_n) \xrightarrow[n\to\infty]{d} (\mathcal{T}^{(k)}, d)$ converge in distribution in the Gromov–Hausdorff topology. In this section, we show this is indeed the case.

From Section 2.3.1, we know that it is enough to find a relation R_n between $\mathcal{T}_n^{(k)}$ and $\mathcal{T}^{(k)}$ such that $\operatorname{dis}(R_n) \to 0$ as $n \to \infty$. However, we can only define relations on metric spaces, not on probability distributions. Thus we need to work with coupled realized 'values' (i.e. metric spaces) of the random variables $\mathcal{T}_n^{(k)}$ and $\mathcal{T}^{(k)}$ and show that on the coupling, the Gromov–Hausdorff distance goes to 0 in probability or almost surely.

An easy coupling is obtained via Skorohod's representation theorem. In short, this theorem allows us to go from convergence in distribution of,

$$\big(C_1^n, \ldots, C_k^n, B_1^n, \ldots, B_k^n\big) \xrightarrow[n\to\infty]{d} \big(C_1, \ldots, C_k, B_1, \ldots, B_k\big),$$

to almost sure convergence by defining the random variables on a common probability space,

$$\mathbb{P}\Big(\lim_{n\to\infty} n^{-\frac{1}{2}}\big(C_1^n, \ldots, C_k^n, B_1^n, \ldots, B_k^n\big) = \big(C_1, \ldots, C_k, B_1, \ldots, B_k\big)\Big) = 1.$$

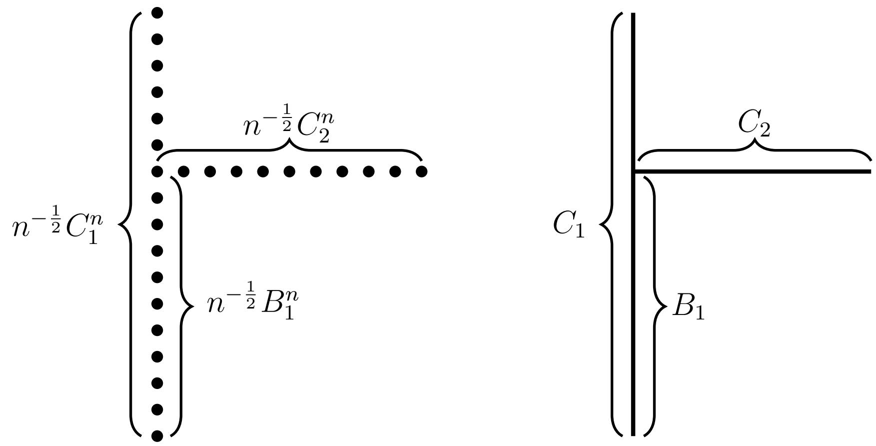


Figure 15: Trees $\mathcal{T}_n^{(2)}$ (left) and $\mathcal{T}^{(2)}$ (right).

On this new probability space, we can then show $\mathbb{P}\big(\lim_{n\to\infty} d_{GH}\big(n^{-\frac{1}{2}}\mathcal{T}_n^{(k)}, \mathcal{T}^{(k)}\big) = 0\big) = 1$, i.e. $n^{-\frac{1}{2}}\mathcal{T}_n^{(k)}$ converges to $\mathcal{T}^{(k)}$ almost surely in Gromov–Hausdorff topology. This immediately implies the desired convergence in distribution.

We first state Skorohod's representation theorem and then show, by constructing an explicit correspondence, that the above almost sure convergence of repeat and attachment points also leads to almost sure convergence of $\mathcal{T}_n^{(k)}$ to $\mathcal{T}^{(k)}$ in the GH-topology.

Theorem 3.2.1. Let (X_n) be a sequence of random vectors such that $X_n \xrightarrow[n\to\infty]{d} X$. Then there exist a probability space with random vectors Y_n, Y, such that,

i) The distribution of Y_n, Y is the same as the distribution of X_n, X respectively.

ii) We have convergence, $Y_n \xrightarrow{a.s.} Y$ as $n \to \infty$.

I.e. whenever we have convergence in distribution, we may change probability spaces to get almost sure convergence.

Proof. We reference to [4][Theorem 25.6.] □

By Skorohod's representation theorem, we may work in a probability space where,

$$\mathbb{P}\Big(\lim_{n\to\infty} \big(n^{-\frac{1}{2}}C_1^n, \dots, n^{-\frac{1}{2}}C_k^n, B_1^n, \dots, B_k^n\big) = \big(C_1, \dots, C_k, B_1, \dots, B_k\big)\Big) = 1.$$

We will show that whenever $n^{-\frac{1}{2}}\big(C_1^n, \dots, C_k^n, B_1^n, \dots, B_k^n\big) \xrightarrow[n\to\infty]{} \big(C_1, \dots, C_k, B_1, \dots, B_k\big)$ and $B_i \neq C_i$ for all $i, j \in [k]$, which is a probability 1 event, then the corresponding trees $\mathcal{T}_n^{(k)}$ (unique up to relabeling) converge to $\mathcal{T}^{(k)}$ (deterministic) in GH-distance. I.e. we show the following,

Theorem 3.2.2. Suppose $n^{-\frac{1}{2}}\big(C_i^n, B_i^n\big) \xrightarrow[n\to\infty]{} \big(C_i, B_i\big)$ with $B_i \neq C_j$ for all $i, j \in [k]$. Then,

$$d_{GH}\big(\mathcal{T}_n^{(k)}, \mathcal{T}^{(k)}\big) \xrightarrow[n\to\infty]{} 0.$$

To show convergence in the Gromov–Hausdorff topology, we first define a correspondence R_n between the metric $\big(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}}d_n\big)$ and $\big(\mathcal{T}^{(k)}, d\big)$.

Remark 3.2.3. We could define the correspondence as follows: For $s \in [0, C_k]$, we pair $\rho(s)$ with vertex $v_s \in \mathcal{T}_n^{(k)}$ where v_s has label $S_n\big(\lfloor sn^{\frac{1}{2}} \rfloor\big)$, I.e. $R = \big\{(\rho(s), v_s) \text{ where } s \in [0, C_k]\big\}$. However, let $s = C_1$, so $\rho(s)$ is on the second stick of $\mathcal{T}$. Then, $\lfloor sn^{\frac{1}{2}} \rfloor = C_1^n + o(n^{\frac{1}{2}})$ and thus we cannot prevent say $\lfloor sn^{\frac{1}{2}} \rfloor \leq C_1^n - 1$ for all n, in which case v_s is on the first stick.

The problem is that this relation links vertices in $\mathcal{T}_n$ up to scale $\sqrt{n}$: we have no control over vertices within a distance $o(\sqrt{n})$ of a repeat index and these vertices might be linked to the wrong stick in $\mathcal{T}$. Since the distortion of a correspondence is based on the supremum over all pairs in the correspondence, we need to make sure that the correspondence links all vertices between C_{k-1}^n and C_k^n to the vertices between C_{k-1} and C_k. Below, we introduce a working correspondence, based on an interpolation between the discrete and continuous sticks.

Definition 3.2.4. Let S_n have at least k repeats and define $\phi_n : [0, C_k) \to \mathcal{T}_n^{(k)}$ as,

$$\phi_n(x) = \Big\{ F\left(\left\lfloor C_{i-1}^n + \tfrac{x - C_{i-1}}{C_i - C_{i-1}}\big(C_i^n - C_{i-1}^n\big)\right\rfloor\right) \qquad \text{for } x \in [C_{i-1}, C_i) \quad ,$$

to be the projection of the first k sticks of $\mathcal{T}$ (as subset of $\mathbb{R}$) onto the first k sticks of $\mathcal{T}_n^{(k)}$. Define correspondence $R_n \subset \big(\mathcal{T}^{(k)} \times \mathcal{T}_n^{(k)}\big)$ as,

$$R_n = \big\{\big(\rho(x), \phi_n(x)\big) \text{ for } x \in [0, C_k)\big\}.$$

For all $n \geq 1$, R_n is a correspondence. Indeed, for all $x \in [0, C_k)$, we have $(\rho(x), \phi_n(x)) \in R$. For the other direction, let $v \in \mathcal{T}_n^{(k)}$ be on stick i. Note $f(x) = C_{i-1}^n + \frac{x - C_{i-1}}{C_i - C_{i-1}}(C_i^n - C_{i-1}^n)$ is continuous with $f(C_{i-1}) = C_{i-1}^n$ and $f(C_i) = C_i^n$. Hence, $\lfloor f(x) \rfloor$ with $x \in [C_{i-1}, C_i)$ takes every integer between C_{i-1}^n and C_i^n. Unlike the naive correspondence from Remark 3.2.3, correspondence R_n by definition only contains pairs of vertices on the same stick.

Lemma 3.2.5. Suppose $\lim_{n\to\infty}(C_i^n, B_i^n) = (C_i, B_i)$, with $C_i \neq B_j$ for $i, j \in [k]$. We have,

$$\lim_{n\to\infty} \mathrm{Dis}(R_n) = \lim_{n\to\infty} \sup_{x,y \in [0, C_k)} |d(\rho(x), \rho(y)) - n^{-\frac{1}{2}} d_n(\phi(x), \phi(y))| = 0.$$

Proof. Let $0 \leq x \leq y < C_k$, so that $\rho(x), \rho(y) \in \mathcal{T}^{(k)}$ and suppose $\rho(x)$ and $\rho(y)$ are on stick i, j respectively. Then for all $n \geq 1$, $\phi_n(x)$ and $\phi_n(y)$ are also on stick i, j in $\mathcal{T}_n^{(k)}$. To ease notation, let $\mathrm{Dis}^n(x, y) = |d(\rho(x), \rho(y)) - n^{-\frac{1}{2}} d_n(\phi_n(x), \phi_n(y))|$. A simple computation shows,

$$\begin{aligned} d(\rho(x), \rho(y)) &= d(\rho(C_i), \rho(C_j)) - d(\rho(x), \rho(C_i)) - d(\rho(y), \rho(C_j)), \\ d_n(\phi_n(x), \phi_n(y)) &= d_n(\phi_n(C_i), \phi_n(C_j)) - d_n(\phi_n(x), \phi_n(X_i)) - d_n(\phi_n(C_j), \phi_n(y)), \end{aligned}$$

and thus we obtain the bound,

$$\mathrm{Dis}^n(x, y) \leq \mathrm{Dis}^n(x, C_i) + \mathrm{Dis}^n(C_i, C_j) + \mathrm{Dis}^n(C_j, y).$$

Since x and C_i are on the same stick, we have $d(\rho(x), \rho(C_i)) = C_i - x$. Similarly, $\phi_n(x)$ and $\phi_n(C_i)$ are on the same stick, therefore,

$$\begin{aligned} d_n(\phi_n(x), \phi_n(C_i)) &= \left\lfloor C_{i-1}^n + \frac{C_i - C_{i-1}}{C_i - C_{i-1}}(C_{i-1}^n - C_i^n) \right\rfloor - \left\lfloor C_{i-1}^n + \frac{x - C_{i-1}}{C_i - C_{i-1}}(C_{i-1}^n - C_i^n) \right\rfloor, \\ &= \delta + (C_i - x)\frac{C_{i-1}^n - C_i^n}{C_i - C_{i-1}}, \end{aligned}$$

with $\delta \in (-1, 1)$. We obtain,

$$\begin{aligned} \mathrm{Dis}^n(x, C_i) &= |d(\rho(x), \rho(C_i)) - n^{-\frac{1}{2}} d_n(\phi_n(x), \phi_n(y))|, \\ &= \left| C_i - x - n^{-\frac{1}{2}} \left(\delta + (C_i - x)\frac{C_{i-1}^n - C_i^n}{C_i - C_{i-1}} \right) \right|, \\ &\leq n^{-\frac{1}{2}} + \left| (C_i - x) \left(1 - n^{-\frac{1}{2}} \frac{C_i^n - C_{i-1}^n}{C_i - C_{i-1}} \right) \right|, \\ &\leq n^{-\frac{1}{2}} + \max_{1 \leq i \leq k} \left(|C_i - C_{i-1}| \left(1 - n^{-\frac{1}{2}} \frac{C_i^n - C_{i-1}^n}{C_i - C_{i-1}} \right) \right). \end{aligned}$$

Observe that this bound is independent of x. Thus we have,

$$\begin{aligned} \lim_{n\to\infty} \sup_{x,y\in[0,C_k)} \mathrm{Dis}^n(x, y) \leq \lim_{n\to\infty} 2n^{-\frac{1}{2}} + \lim_{n\to\infty} 2 \max_{1\leq i\leq k} \left(|C_i - C_{i-1}| \left(1 - n^{-\frac{1}{2}} \frac{C_i^n - C_{i-1}^n}{C_i - C_{i-1}} \right) \right) \\ + \lim_{n\to\infty} \max_{1\leq i,j\leq k} \mathrm{Dis}^n(C_i, C_j) = \lim_{n\to\infty} \max_{1\leq i,j\leq k} \mathrm{Dis}^n(C_i, C_j). \end{aligned}$$

Hence, the lemma is proven upon showing $\lim_{n\to\infty} \max_{1\leq i,j\leq k} \mathrm{Dis}^n(C_i, C_j) = 0$. Note that,

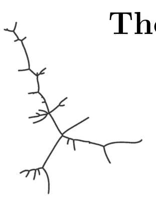
$$\mathrm{Dis}^n(C_i, C_j) = \left| d(\rho(C_i), \rho(C_j)) - n^{-\frac{1}{2}} d_n(F(C_i^n), F(C_j^n)) \right|.$$

We turn the assumptions $\lim_{n\to\infty}(n^{-\frac{1}{2}} C_i^n, n^{-\frac{1}{2}} B_i^n) = (C_i, B_i)$ and $B_i \neq C_j$ for $1 \leq i, j \leq k$ into: $\forall \epsilon \in \left(0, \frac{|C_i - B_i|}{4}\right)$, there exists n large enough such that,

$$|n^{-\frac{1}{2}} C_i^n - C_i| < \epsilon, \qquad |n^{-\frac{1}{2}} B_i^n - B_i| < \epsilon, \qquad |n^{-\frac{1}{2}} B_i^n, C_i| > 2\epsilon.$$

Thus, for large enough n, we have $C_j^n \leq B_i^n \leq C_{j+1}^n$ if and only if $C_j \leq B_i \leq C_{j+1}$. Hence, any path from C_i^n to C_j^n in $\mathcal{T}_n$ must follow the same segments as the path from C_i to C_j in $\mathcal{T}$. Given that $i, j \leq k$, we have that $d_n(C_i^n, C_j^n)$ equals the length of at most k branch segments, which all take the form $|C_i^n - C_{i-1}^n|, |C_i^n - B_j^n|$ or $|B_i^n - B_j^n|$. Similarly $d(C_i, C_j)$ consists of the corresponding branches with lengths $|C_i - C_{i-1}|, |C_i - B_j|$ or $|B_i - B_j|$. Since $C_i^n \to C_i$ and $B_i^n \to B_i$ as $n \to \infty$, we apply the triangle inequality to the at most k branches to find that for n large enough we have $|n^{-\frac{1}{2}} d_n(C_i^n, C_j^n) - d(C_i, C_j)| < 2\epsilon k$ for all $1 \leq i, j \leq k$. We obtain $\lim_{n\to\infty} \max_{1\leq i,j\leq k} \mathrm{Dis}^n(C_i, C_j) = 0$. This concludes the proof. □

Theorem 3.2.6. We have $\left(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}} d_n\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}^{(k)}, d\right)$ in the Gromov–Hausdorff topology.



Proof. From Theorem 3.1.1, we know that $(C_i^n, B_i^n) \xrightarrow[n\to\infty]{d} (C_i, B_i)$ for $1 \le i \le k$. Then using Theorem 3.2.1, we may work in a probability space where this convergence is almost sure, that is:

$$\mathbb{P}\left(\lim_{n\to\infty}(C_i^n, B_i^n) = (C_i, B_i) \text{ for } 1 \le i \le k\right) = 1.$$

Recall that each $B_i \sim \text{Unif}([0, C_i])$, hence $\mathbb{P}(B_i \neq C_j \text{ for all } 1 \le i, j \le k) = 1$. Combining both events yields,

$$\mathbb{P}\left(\lim_{n\to\infty}(C_i^n, B_i^n) = (C_i, B_i) \text{ and } C_i \neq B_j \text{ for all } 1 \le i, j \le k\right) = 1.$$

By Lemma 3.2.5, we obtain that almost surely,

$$\lim_{n\to\infty} \sup_{s,u\in[0,t]} |n^{-\frac{1}{2}} d_n(v_s, v_u) - d(\rho(s), \rho(u))| = 0.$$

Thus, we have found a relation R_n between $(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}} d_n)$ and $(\mathcal{T}^{(k)}, d)$ for which almost surely $\text{Dis}(R_n) \xrightarrow[n\to\infty]{} 0$. Hence we obtain $(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}} d_n) \xrightarrow[n\to\infty]{\text{a.s.}} (\mathcal{T}^{(k)}, d)$ in the GH-topology, which implies the desired convergence in distribution. □

3.3 Finite Dimensional Convergence of the Measures

In the prior section, we saw how we can construct a correspondence that shows

$$\left(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}} d_n\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}^{(k)}, d\right),$$

in the GH-topology. In the current section, we extend this to the GHP-topology by including the measures $\nu_n^{(k)}$ and $\mu^{(k)}$. Using Skorohods representation Theorem 3.2.1, we again work in a probability space on which $\left(C_i^n, B_i^n\right)$ converges to $\left(C_i, B_i\right)$ almost surely and extend this space to also include measures $\nu_n^{(k)}$ and $\mu^{(k)}$. From Definition 2.3.17, it suffices to find a measure $\pi_n : \mathcal{T}_n^{(k)} \times \mathcal{T}^{(k)} \to [0, 1]$ so that $D\left(\pi; \nu_n(k), \mu^{(k)}\right) \to 0$ and $\pi_n\left(R_n^c\right) \to 0$ as $n \to \infty$ where R_n is the correspondence from Definition 3.2.4.

Definition 3.3.1. For $A \subset \mathcal{T}_n^{(k)}, B \subset \mathcal{T}^{(k)}$, and ϕ_n as in Definition 3.2.4, we define,

$$\pi_n(A, B) = \mu^{(k)}\left(\left\{x \in B : \phi_n(\rho^{-1}(x)) \in A\right\}\right).$$

Lemma 3.3.2. With R_n as in Definition 3.2.4, we claim,

$$i)\ \pi_n\left(R_n^c\right) = 0 \quad \text{and} \quad ii)\ D\left(\pi; \nu_n^{(k)}, \mu^{(k)}\right) \xrightarrow[n\to\infty]{} 0.$$

Proof. We start with $i)$. Since $\{x : \phi_n(\rho^{-1}(x)) \in R_n^c)\} = \emptyset$, we find,

$$\pi_n(R_n^c) = \mu^{(k)}(\emptyset) = 0, \text{ for all } n.$$

We continue with $ii)$. For this, let $B \subset \mathcal{T}^{(k)}$. Then,

$$\pi_n(\mathcal{T}_n^{(k)}, B) = \mu^{(k)}\left(\left\{x \in B : \phi_n(\rho^{-1}(x)) \in \mathcal{T}_n^{(k)}\right\}\right) = \mu^{(k)}(B).$$

Thus, $||\pi_n \circ p_1^{-1} - \mu^{(k)}||_{TV} = 0$ for all n. It remains to show that $||\pi_n \circ p_2^{-1} - \nu_n^{(k)}||_{TV} \to 0$ as $n \to \infty$. For this, let $A \subset \mathcal{T}_n^{(k)}$. We compute,

$$\pi_n\big(A, \mathcal{T}^{(k)}\big) = \mu^{(k)}\big(\{x \in \mathcal{T}^{(k)} : \phi_n(\rho^{-1}(x)) \in A\}\big) = \sum_{a \in A} \mu^{(k)}\big(\{x \in \mathcal{T}^{(k)} : \phi_n(\rho^{-1}(x)) = a\}\big).$$

Let $a \in \mathcal{T}_n^{(k)}$ be a vertex on the i'th branch and let $s \in \{C_{i-1}^n, \ldots, C_i^n - 1\}$ be the index corresponding to a. Recall that $\mu^{(k)}$ is the pushforward by ρ of the normalized Lebesgue measure on $[0, C_k]$, denoted λ. Hence,

$$\begin{aligned}
\mu^{(k)}\big(\{x \in \mathcal{T}^{(k)} : \phi_n(\rho^{-1}(x)) = a\}\big) &= \lambda\big(\{x \in [0, C_k] : \phi_n(x) = a\}\big), \\
&= \lambda\Big(\Big\{x : \Big\lfloor C_{i-1}^n + \frac{x - C_{i-1}}{C_i - C_{i-1}}(C_i^n - C_{i-1}^n)\Big\rfloor = s\Big\}\Big), \\
&= \lambda\Big(\Big\{x : C_{i-1}^n + \frac{x - C_{i-1}}{C_i - C_{i-1}}(C_i^n - C_{i-1}^n) \in [s, s+1)\Big\}\Big), \\
&= \lambda\Big(\Big[0, \frac{C_i - C_{i-1}}{C_i^n - C_{i-1}^n}\Big)\Big), \\
&= \frac{1}{C_k}\frac{C_i - C_{i-1}}{C_i^n - C_{i-1}^n}.
\end{aligned}$$

Observe that,

$$\frac{1}{C_k}\frac{C_i - C_{i-1}}{C_i^n - C_{i-1}^n} = \frac{1}{C_k}\frac{C_i - C_{i-1}}{n^{\frac{1}{2}}(C_i - C_{i-1}) + o(n^{\frac{1}{2}})} = \frac{1}{C_k n^{\frac{1}{2}}} + o\big(n^{-\frac{1}{2}}\big) = \frac{1}{C_k^n} + o\big(n^{-\frac{1}{2}}\big).$$

We conclude,

$$\pi_n(A, \mathcal{T}^{(k)}) = \sum_{a \in A} \mu^{(k)}\big(\{x \in \mathcal{T}^{(k)} : \phi_n(x) = a\}\big) = \frac{\#A}{C_k^n} + \#A o\big(n^{-\frac{1}{2}}\big) = \nu_n^{(k)}(A) + o(1),$$

Since $\frac{\#A}{C_k^n} = \nu_n^{(k)}(A)$ and $\#A \le C_k^n = O(n^{\frac{1}{2}})$. We conclude that $D\big(\pi_n; \nu_n^{(k)}, \mu^{(k)}\big) \xrightarrow[n\to\infty]{} 0$. This verifies $ii)$ and concludes the proof. $\square$

This allows us to prove $d_{GHP}\big((\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}}, \nu_n^{(k)}), (\mathcal{T}^{(k)}, d, \mu^{(k)}\big) \to 0$ as $n \to \infty$.

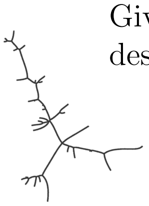
Lemma 3.3.3. For all $k \in \mathbb{N}$, we have,

$$\mathbb{P}\Big(d_{GHP}\big((\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}}, \nu_n^{(k)}), (\mathcal{T}^{(k)}, d, \mu^{(k)}\big) \xrightarrow[n\to\infty]{} \Big) = 1$$

Proof. Suppose $\lim_{n\to\infty}(C_i^n, B_i^n) = (C_i, B_i)$, with $C_i \neq B_j$ for $i, j \in [k]$. Then we have,

$$\begin{aligned}
d_{GHP}\big((\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}}, \nu_n^{(k)}), (\mathcal{T}^{(k)}, d, \mu^{(k)})\big) &= \inf_{R,\pi}\left\{\max\left(\frac{1}{2}\mathrm{dis}(R), D(\pi; \nu_n^{(k)}, \mu^{(k)}) + \pi(R^c)\right)\right\}, \\
&\le \max\left(\frac{1}{2}\mathrm{dis}(R_n), D(\pi_n; \nu_n^{(k)}, \mu^{(k)}) + \pi_n(R_n^c)\right), \\
&\xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

Given that $\lim_{n\to\infty}(C_i^n, B_i^n) = (C_i, B_i)$, with $C_i \neq B_j$ for $i, j \in [k]$ holds almost surely, the desired result follows. $\square$



Since almost sure convergence implies convergence in distribution, we find that,

$$\left(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}}, \nu_n^{(k)}\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}^{(k)}, d, \mu^{(k)}\right),$$

which finishes the proof of Theorem 1.5.1.

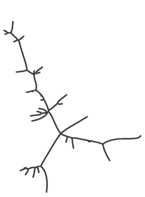
Remark 3.3.4. The results in Section 3.2 and Section 3.3 relied only on the convergence of,

$$n^{-\frac{1}{2}}\left(C_1^n, \ldots, C_k^n, B_1^n, \ldots, B_k^n\right) \to \left(C_1, \ldots, C_k, B_1, \ldots, B_k\right),$$

and the fact that $C_i \neq B_j$ for all $i, j \in [k]$ is true with probability 1. In Section 6 and Section 7 we consider finite dimensional convergence $g(n)T_n^{(k)} \to \mathcal{T}^{(k)}$ for more general trees $\mathcal{T}_n$ and $\mathcal{T}$ and scaling functions $g : \mathbb{N} \to [0, \infty)$. In each of these cases, we will show,

$$g(n)\left(C_1^n, \ldots, C_k^n, B_1^n, \ldots, B_k^n\right) \xrightarrow[n\to\infty]{d} \left(C_1, \ldots, C_k, B_1, \ldots, B_k\right),$$

for their respective repeat and attachment indices C_i^n, B_i^n, C_i, B_i and $C_i \neq B_j$ for all $i, j \in [k]$ is true with probability 1. This allows us to immediately apply the result in this section to obtain convergence in the GHP-topology of the trees and uniform measures obtained from the first k branches.



4 Tightness of $\left(\mathcal{T}_n, n^{-\frac{1}{2}}d_n\right)_{n\in\mathbb{N}}$

This section is dedicated towards showing *i*) and *ii*) from Proposition 2.4.3. That is, we aim to show that for all $\epsilon > 0$,

$$i)\lim_{t\to\infty}\mathbb{P}\left(d_H\left(\mathcal{T}(t),\mathcal{T}\right) > \epsilon\right) = 0 \quad \text{and} \quad ii)\lim_{t\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(d_H\left(\mathcal{T}_n\big(tn^{\frac{1}{2}}\big),\mathcal{T}_n\right) > \epsilon\right) = 0.$$

Remark 4.0.1. Observe that *i*) implies that $\mathcal{T}$ is compact almost surely. This follows from a diagonalization argument. To see this, *i*) implies that we can find a sequence t_i such that,

$$\sum_{i=1}^{\infty}\mathbb{P}\big(d_H\big(\mathcal{T}(t_i),\mathcal{T}\big) > \epsilon\big) < \sum_{i=1}^{\infty}2^{-i} < \infty.$$

By the first Borel–Cantelli lemma, this yields,

$$\mathbb{P}\left(\limsup_{i\to\infty} d_H\big(\mathcal{T}(t_i),\mathcal{T}\big) > \epsilon\right) = 0.$$

Since the metric space $\big(\mathcal{T}(t), d\big)$ is compact for every t, we have $\mathcal{T}$ is compact almost surely.

4.1 Compactness of the Continuous Tree

Theorem 4.1.1. For all $\epsilon > 0$, we have,

$$\lim_{t\to\infty}\mathbb{P}\big(d_H\big(\mathcal{T}(t),\mathcal{T}\big) > \epsilon\big) = 0.$$

The idea behind the proof is to bound the growth $d_H(\mathcal{T}(t),\mathcal{T})$ by controlling the growth of infinitely many intermediate sections of finite length. That is, we bound the distances,

$$d_H(\mathcal{T}(t),\mathcal{T}(2t)), d_H(\mathcal{T}(2t),\mathcal{T}(4t)), d_H(\mathcal{T}(4t),\mathcal{T}(8t)), \ldots.$$

Theorem 4.1.1 is proven upon showing the combined growth on these infinitely many sections is smaller than ϵ with probability 1 as $t\to\infty$. We illustrate this in Figure 16.

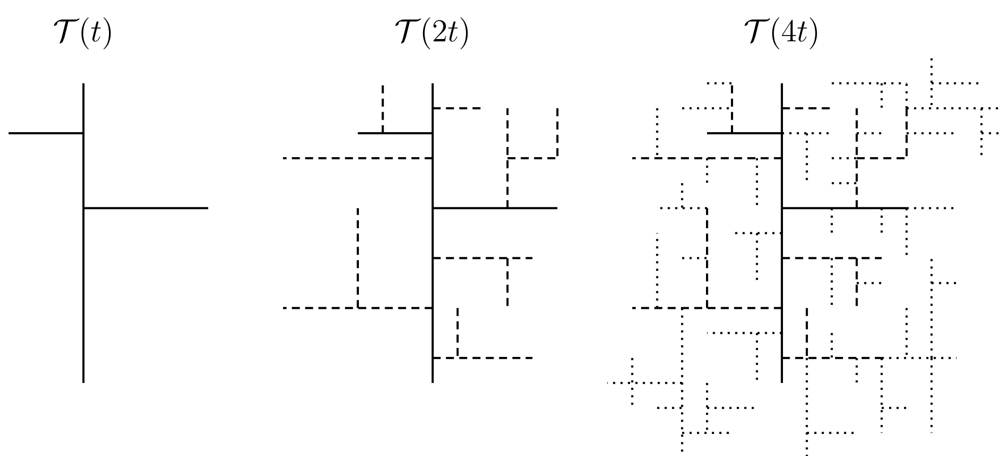


Figure 16: Illustration of $\mathcal{T}(t), \mathcal{T}(2t)$ and $\mathcal{T}(4t)$.

Before proving Theorem 4.1, we first introduce various intermediate results.

Lemma 4.1.2. Fix $\epsilon > 0$ and let $\epsilon_i \equiv \epsilon_i(T) > 0$, for some $T > 0$ such that $\sum_{i=0}^{\infty} \epsilon_i < \epsilon$. Then for any $t > 0$,

$$\mathbb{P}\big(d_H\big(\mathcal{T}(t), \mathcal{T}\big) > \epsilon\big) \leq \sum_{i=0}^{\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1} t)\big) > \epsilon_i\big).$$

Proof. By repeatedly applying the triangle inequality, we obtain,

$$d_H\big(\mathcal{T}(t), \mathcal{T}\big) = \lim_{k\to\infty} d_H\big(\mathcal{T}(t), \mathcal{T}(2^k t)\big) \leq \lim_{k\to\infty} \sum_{i=0}^{k} d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1} t)\big).$$

Observe that,

$$d_H\big(\mathcal{T}(t), \mathcal{T}\big) > \epsilon \implies d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1} t)\big) > \epsilon_i \text{ for some } i \in \mathbb{N}_0.$$

Thus we apply a union bound to obtain,

$$\begin{aligned}\mathbb{P}\big(d_H\big(\mathcal{T}(t), \mathcal{T}\big) > \epsilon\big) &\leq \mathbb{P}\left(\bigcup_{i=0}^{\infty} \left\{d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1} t)\big) > \epsilon_i\right\}\right), \\ &\leq \sum_{i=0}^{\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1} t)\big) > \epsilon_i\big),\end{aligned}$$

as desired. □

Corollary 4.1.3. Theorem 4.1.1 is proven upon finding $\epsilon_i(t) : [0, \infty) \to [0, \infty)$ for which,

$$i)\ \lim_{t\to\infty} \sum_{i=0}^{\infty} \epsilon_i(t) = 0 \quad \text{and} \quad ii)\ \lim_{t\to\infty} \sum_{i=0}^{\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1} t)\big) > \epsilon_i\big) = 0.$$

Proof. Condition $i)$ ensures that for all $\epsilon > 0$, there exists $T > 0$ such that $\sum_{i=0}^{\infty} \epsilon_i(T) < \epsilon$ and condition $ii)$ ensures that $\lim_{t\to\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(t), \mathcal{T}\big) > \epsilon\big) = 0$. □

We aim to bound $\mathbb{P}\big(d_H\big(\mathcal{T}(a), \mathcal{T}(2a)\big) > c\big)$ and substitute $a = 2^i t$ and $c = \epsilon_i$ for a suitable ϵ_i. To bound $d_H(\mathcal{T}(a), \mathcal{T}(2a))$, we first bound $d_H(\mathcal{T}(a), \rho(s))$ for some $s \in (a, 2a]$.

Remark 4.1.4. We give intuition behind bounding $\mathbb{P}\big(d_H\big(\mathcal{T}(a), \rho(s)\big) > c\big)$. As explained in Figure 17, we may write $d_H\big(\mathcal{T}(a), \rho(s)\big) = \sum_{i=1}^{N} d_i$, where N is the random number of sticks traversed on the path from $\rho(s)$ to $\mathcal{T}(a)$ and d_i denotes the length of the path traversed on the i'th stick. Note that $d_1 > c$ is only possible if $s - c > a$. Hence,

$$\mathbb{P}(d_1 > c) = \mathbb{1}_{\{s-c>a\}} \mathbb{P}\big(\eta([s-c, s]) = 0\big) = \mathbb{1}_{\{s-c>a\}} \exp\left(-\int_{s-c}^{s} t dt\right) \leq \exp\left(-ca\right),$$

where η is a PPP of intensity tdt. Thus d_1 is stochastically dominated by an Exp(a) random variable. Similar reasoning shows that d_i is also stochastically dominated by an Exp(a)

random variable for $i \in \{1, \ldots, N\}$. To dominate N, any stick $\rho((C_j, C_{j+1}]) \subset \mathcal{T}(2a)$ is attached at $\rho(B_j)$ with $B_j \sim \text{Unif}([0, C_j])$. Since $C_j \leq 2a$, we have $\rho((C_j, C_{j+1}])$ is attached to $\mathcal{T}(a)$ with probability at least $\frac{1}{2}$. This gives a heuristic why N is stochastically dominated by a $\text{Geom}\left(\frac{1}{2}\right)$ random variable. If we assume the stochastic domination of d_i and N can be done by independent random variables, we see,

$$\mathbb{P}(d_H(\mathcal{T}(a), \rho(s)) > c) = \mathbb{P}\left(\sum_{i=1}^{N} d_i > c\right) \leq \mathbb{P}\left(\sum_{i=1}^{M} X_i > c\right) \leq \exp\left(-\frac{ca}{2}\right),$$

where $M \sim \text{Geom}\left(\frac{1}{2}\right)$ and $X_i \sim \text{Exp}(a)$ and the last inequality uses that $\sum_{i=1}^{M} X_i \sim \text{Exp}\left(\frac{a}{2}\right)$. Unfortunately, we do not have independence between N and d_i. However, negative correlations (large number of sticks N implies d_i smaller and vise versa) could allow us to formalize this argument using couplings. This approach is taken in Section 7.4.

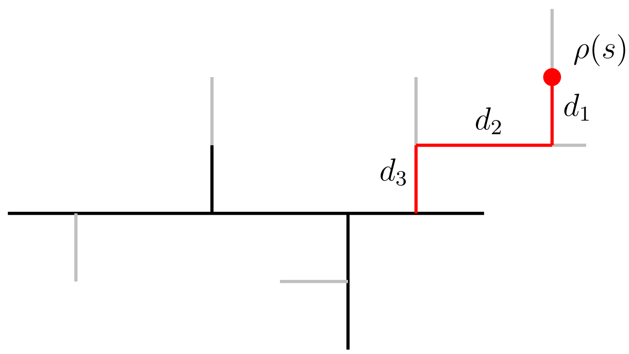


Figure 17: Illustration of $d_H\big(\mathcal{T}(a), \rho(s)\big)$. Here $\rho([0, a])$ is drawn in black, while $\rho((a, 2a])$ is gray. Red denotes the path from $\rho(s)$ to $\mathcal{T}(a)$. This path can be divided into partial paths along each branch in $\mathcal{T}(2a) \setminus \mathcal{T}(a)$ as indicated. We have $d_H\big(\mathcal{T}(a), \rho(s)\big) = d_1 + d_2 + d_3$.

In this section, we use an approach relying on sampling the repeat points $0 < C_1 < C_2 < \ldots$ and attachment points $B_i \sim \text{Unif}([0, C_i])$ using one PPP. Recall Example 2.2.20, where for,

$$T = \{(x, y) \in \mathbb{R}^2 : x \geq 0, y \leq x\},$$

and η a PPP of intensity 1 on T, we may write $\eta = \sum_{i=1}^{\infty} \delta_{(C_i, B_i)}$, where $0 < C_1 < C_2 < \ldots$ are the ordered points of a PPP of intensity tdt on $\mathbb{R}_{\geq 0}$ and $B_i \sim \text{unif}([0, C_i])$. Thus, to construct $\mathcal{T}$, we may take the ordered x-coordinates of η as endpoints of the sticks, and the corresponding y-values as attachment point of the sticks. We illustrate this in a small example in Figure 18, where for readability, we scale up the constructed tree by a factor 2.

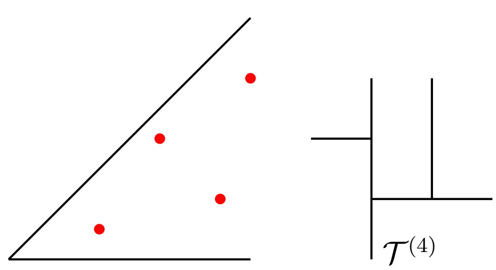


Figure 18: Tree from PPP on wedge T

By using this coupled sampling approach of $0 < C_1 < C_2 < \dots$ and $B_1, B_2, \dots$, we show the following.

Lemma 4.1.5. Fix $a, c > 0$ and let $s \in [a, 2a]$. Then,

$$\mathbb{P}\big(d_H\big(\mathcal{T}(a), \rho(s)\big) > c\big) \leq e^{-ac}.$$

Remark 4.1.6. Before proving this, we explain how to recover the path from $\rho(s)$ to $\mathcal{T}(a)$ from the PPP η on T. For this, let $\eta' = \eta + \delta_{(a,a)}$. The tree constructed from η and η' is the same, only difference being that the stick added to exceed $\mathcal{T}(a)$ is broken in two. This means that we always start the construction $\mathcal{T}(2a)\backslash\mathcal{T}(a)$ with a new stick, ensuring the definitions below are consistent.

The distance from $\rho(s)$ to the start of the stick on which $\rho(s)$ lies equals the distance between the line $x = s$ and the closest point in η to the left $x = s$. If the y-value of this point is smaller than a, we attach this stick to $\mathcal{T}(a)$ and found $d_H(\mathcal{T}(a), \rho(s))$. If the corresponding y-value is larger than a, we attach to a stick in $\mathcal{T}(2a)\backslash\mathcal{T}(a)$ and need to look further down the tree for the full path from $\rho(s)$ to $\mathcal{T}(a)$. This description explains the definition below.

Definition 4.1.7. Let $\eta = \sum_{i=1}^{\infty} \delta_{(C_i, B_i)}$ be a PPP of intensity 1 on T and set $\eta' = \eta + \delta_{(a,a)}$. We define,

$$p_1 \equiv \big(p_x^1(s), p_y^1(s)\big) = \underset{(x,y)\in\eta' :\, x \leq s}{\arg\max}\ x, \quad d_1 \equiv d_1(s) = s - p_x^1(s) \ \text{ and } \ s_2 = p_y^1(s).$$

Note that p_1 exists as η has no accumulation points almost surely. Also $p_1 \geq a$ as we added the points (a, a) to the point process. Recursively, whenever $s_{i-1} > a$, we set,

$$p_i \equiv p_i(s) = p(p_{i-1}(s)), \qquad d_i \equiv d_i(s) = s_i - p_x^i(s) \quad \text{and} \quad s_{i+1} = p_y^i(s_i).$$

Set $N = \min\{i \in \mathbb{N} : s_{i+1} \leq a\}$ for the number of sticks on the path from $\rho(s)$ to $\mathcal{T}(a)$ and let $S = \bigcup_{i=1}^{N} [p_x^i(s), s_i]$ to be the subset of $[a, 2a]$ corresponding to the parts of the sticks on the path from $\rho(s)$ to $\mathcal{T}(a)$ as stated in Lemma 4.1.8.

Lemma 4.1.8. Let η be a PPP of intensity 1 in T. The path from $\rho(s)$ to $\mathcal{T}(a)$ given η equals $\rho(S)$ and hence $d_H\big(\mathcal{T}(\rho(s)), \mathcal{T}(2a)\big) = \lambda(S)$ with λ the Lebesgue measure on $\mathbb{R}$.

Proof. The result is immediate as Definition 4.1.7 is a mathematical formalization from the explanation in Remark 4.1.6. We also give a visual explanation in Figure 19. □

Remark 4.1.9. We make two observations,

i) $d_H\big(\mathcal{T}(a), \rho(s)\big) > c$ given η happens exactly when $\lambda(S) = d_1 + \dots + d_N > c$ where λ is the Lebesgue measure on $\mathbb{R}$.

ii) By removing points from η in $(a, 2a] \times [0, a)$, we cannot decrease $d_H(\mathcal{T}(a), \rho(s))$.

We combine these observations to show $d_H(\mathcal{T}(a), \rho(s)) > c$ is possible only if $\eta(R) = 0$ for $R \subset T$ with $\lambda(R) = ac$ for λ the Lebesgue measure on $\mathbb{R}^2$.

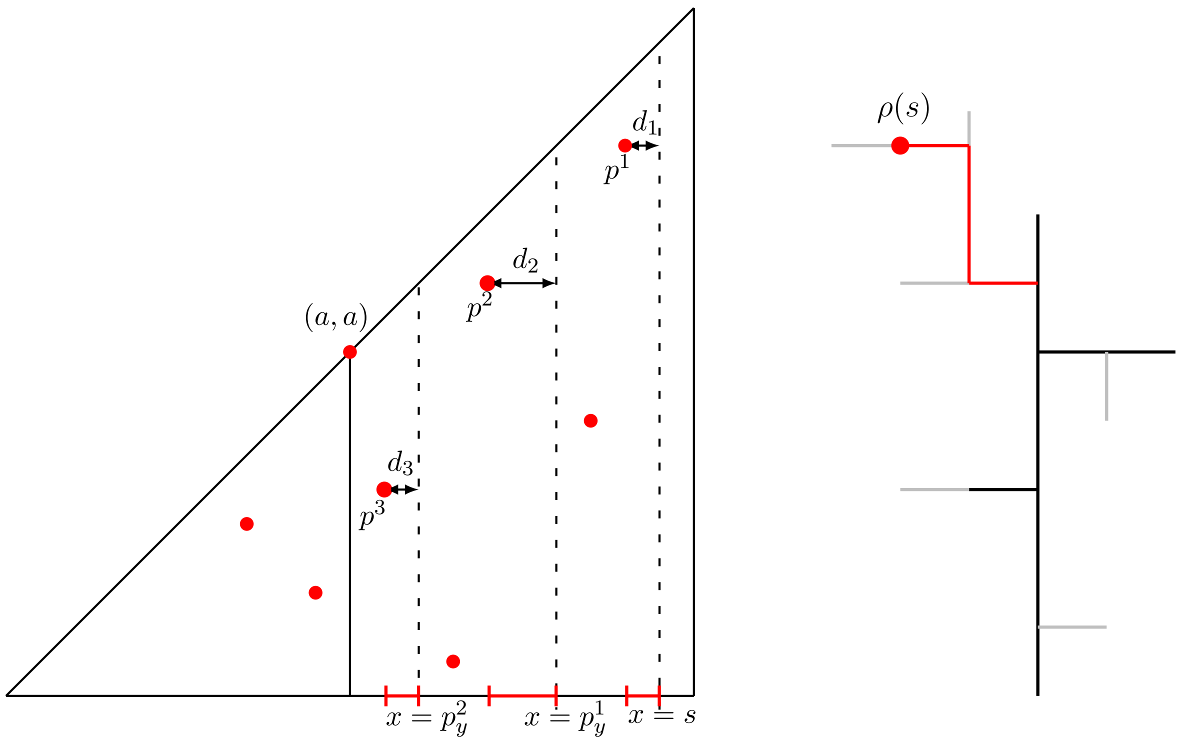


Figure 19: Explanation Definition 4.1.7. The red intervals on the x-axis is the set $S \subset [a, 2a]$. On the right, we see the tree constructed from this sample of η.

proof of Lemma 4.1.5. Define the point process $\eta^* = \{(x, y) \in \eta' : x \leq a \text{ or } y \geq a\}$ for the set of points in η in $T \setminus \big((a, 2a] \times [0, a)\big)$. Let S^*, N^* and p_i^*, d_i^* for $i \in [N^*]$ be as defined in Definition 4.1.7, but using η^* instead of η'. Since η^* is obtained from η' by removing points only in $(a, 2a] \times [0, a)$, we have, $d_H(\mathcal{T}(a), \rho(s)) \leq \lambda(S^*)$ given η. We define,

$$S_r^* = \{x \in S^* : x \geq r\} \text{ and } r^* = \underset{a \leq r \leq 2a}{\operatorname{argmax}} \{\eta'(S_r^* \times [0, a]) \geq 1\}$$

so that r^* is the rightmost point in η inside $S^* \times [0, a]$. Observe $S = S_{r^*}^* = \{x \in S^* : x \geq r^*\}$. Indeed, suppose $r^* \in [(p_x^k)^*, s_k^*] \subset S^*$ for some $k \in [N^*]$. Then, $p_i = p_i^*$ for $i < k$ and $p_k^x = r^*$ with $s_k < a$. This shows $S = \{x \in S^* : x \geq r^*\}$. See Figure 20 for an explanation with $k = 3$.

To conclude, suppose $\lambda(S^*) > c$ and take $j \in [a, 2a]$ such that $\lambda(S_j^*) = c$. Then,

$$\mathbb{P}\big(d_H\big(\mathcal{T}(a), \rho(s)\big) > c\big) \leq \mathbb{P}\big(\lambda(S_{r^*}^*) > c\big) \leq \mathbb{1}_{\{\lambda(S^*) > c\}} \mathbb{P}\big(\eta\big(S_j^* \times [0, a]\big) = 0\big) \leq e^{-ac},$$

as desired. □

We continue by upgrading $\mathbb{P}\big(d_H\big(\mathcal{T}(a), \rho(s)\big) > c\big) \leq e^{-ac}$ to a bound $\mathbb{P}(d_H(\mathcal{T}(a), \mathcal{T}(2a)) > c)$. We will give two approaches. First a more standard reasoning, and then a trick-based proof.

Remark 4.1.10. For the standard reasoning, we observe,

$$\mathbb{P}\big(d_H\big(\mathcal{T}(a), \mathcal{T}(2a)\big) > c\big) = \mathbb{P}\left(\sup_{s \in [a, 2a]} d_H\big(\mathcal{T}(a), \rho(s)\big) > c\right) \leq \mathbb{P}\left(\bigcup_{s \in [a, 2a]} d_H\big(\mathcal{T}(a), \rho(s)\big) > c\right).$$

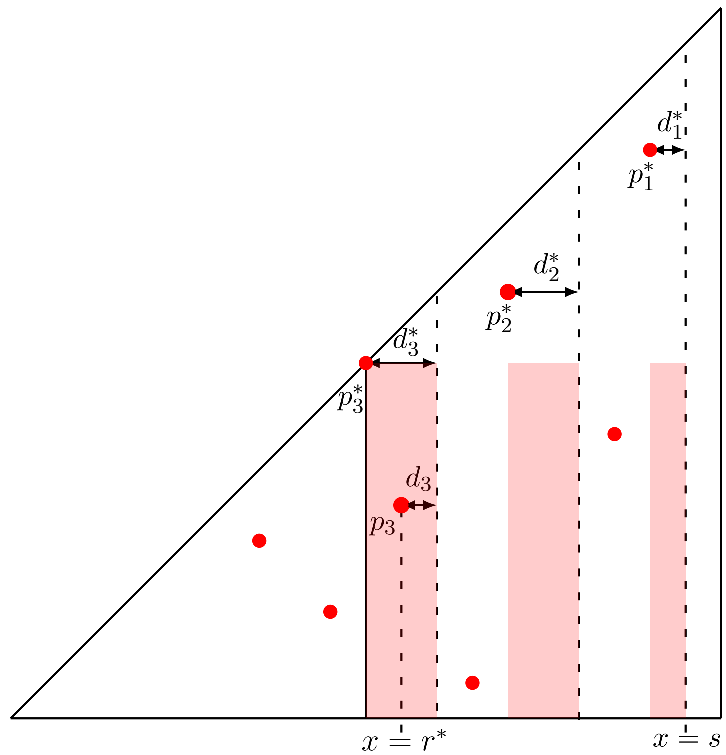


Figure 20: Explanation relation η^* and η. Red denotes the region $S^* \times [0, a]$.

We would like to use a union bound, but cannot apply this technique as we are taking the union over an uncountable set. However, $d_H\big(\mathcal{T}(a), \rho(s)\big)$ must take its maximum value at the endpoint of a stick, or at $s = 2a$. The number of sticks on the interval $[a, 2a]$ follows a $\text{Poi}\left(\int_a^{2a} t dt\right) = \text{Poi}\left(\frac{3}{2}a^2\right)$ distribution, so the number of $s \in [a, 2a]$ where $d_H(\mathcal{T}(a), \rho(s))$ can attain its maximum value is nicely behaved. This can be turned into a bound, as seen below.

Lemma 4.1.11. Let $a, c > 0$. We have,

$$\mathbb{P}\big(d_H\big(\mathcal{T}(a), \mathcal{T}(2a)\big) > c\big) \leq \left(\frac{3}{2}a^2 + 1\right) e^{-ac}.$$

Proof. Let $K = \{C_i : C_i \in (a, 2a]\} \cup \{2a\}$ and note $|K| \sim \text{Poi}\left(\int_a^{2a} t dt\right) + 1 = \text{Poi}\left(\frac{3}{2}a^2\right) + 1$. By the discussion in Remark 4.1.10, and the law of total probability, we obtain,

$$\begin{aligned} \mathbb{P}(d_H(\mathcal{T}(a), \mathcal{T}(2a)) > c) &= \sum_{k=1}^{\infty} \mathbb{P}\left(\max_{x \in K}\{d_H(\mathcal{T}(a), \rho(x))\} > c \,\middle|\, |K| = k\right) \mathbb{P}(|K| = k), \\ &\leq \sum_{k=1}^{\infty} \mathbb{P}\left(\bigcup_{x \in K} \{d_H(\mathcal{T}(a), \rho(x)) > c\} \,\middle|\, |K| = k\right) \mathbb{P}(|K| = k), \\ &\leq e^{-ac} \sum_{k=1}^{\infty} k \mathbb{P}(|K| = k). \end{aligned}$$

This completes the proof as $\sum_{k=1}^{\infty} k\mathbb{P}(|K| = k) = \mathbb{E}[|K|] = \left(\frac{3}{2}a^2 + 1\right)$. $\square$

Remark 4.1.12. The second proof relies on the fact that if there exists $s \in [a, 2a]$ so that $d_H(\mathcal{T}(a), \mathcal{T}(2a)) > c$, then every point in a segment of length at least $\frac{c}{2}$ in $\mathcal{T}(2a) \backslash \mathcal{T}(a)$ is at distance at least $\frac{c}{2}$ from $\mathcal{T}(a)$. This is illustrated in Figure 21, where in black we draw $\mathcal{T}(a)$ and $\mathcal{T}(2a) \setminus \mathcal{T}(a)$ is drawn in gray. red denotes all points that are at least half as far from $\mathcal{T}(a)$ as the (furthest) point $\rho(x)$. Using this, we obtain the following bound.

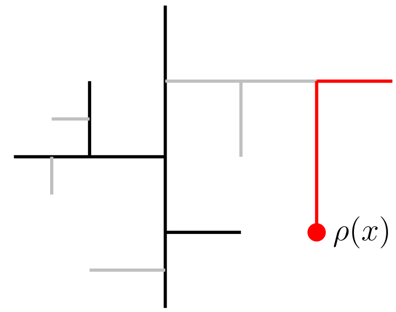


Figure 21: Distances in $\mathcal{T}(2a)$.

Lemma 4.1.13. Let $a, c > 0$. Then we have,

$$\mathbb{P}(d_H(\mathcal{T}(a), \mathcal{T}(2a)) > c) \leq \frac{2a}{c} e^{-\frac{ac}{2}}.$$

Proof. By Lemma 4.1.5, we get,

$$\mathbb{E}\left[\lambda\left(\left\{x \in [a, 2a] : d_H\big(\mathcal{T}(a), \rho(x)\big) > \frac{c}{2}\right\}\right)\right] \leq a e^{-\frac{ac}{2}}.$$

Then if $d_H(\mathcal{T}(a), \mathcal{T}(2a)) > c$, we have $\lambda\left(\left\{x \in [a, 2a] : d_H\big(\mathcal{T}(a), \rho(x)\big) > \frac{c}{2}\right\}\right) > \frac{c}{2}$ as seen in Remark 4.1.12. Using this, we obtain the bound,

$$\mathbb{E}\left[\lambda\left(\left\{x \in [a, 2a] : d_H\big(\mathcal{T}(a), \rho(x)\big) > \frac{c}{2}\right\}\right)\right] \geq \frac{c}{2}\mathbb{P}(d_H(\mathcal{T}(a), \mathcal{T}(2a)) > c).$$

Combining both inequalities yields $\mathbb{P}\big(d_H(\mathcal{T}(a), \mathcal{T}(2a)) > c\big) \leq \frac{2a}{c} e^{-\frac{ac}{2}}$, the desired result. □

Proof of Theorem 4.1. From Corollary 4.1.3, it is enough to find $\epsilon_i(t)$ such that,

$$i) \lim_{t\to\infty} \sum_{i=0}^{\infty} \epsilon_i(t) = 0 \quad \text{and} \quad ii) \lim_{t\to\infty} \sum_{i=0}^{\infty} \mathbb{P}\Big(d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1} t)\big) > \epsilon_i\Big) = 0.$$

Substitute $a = 2^i t$ and $c = \epsilon_i$ in Lemma 4.1.13. This yields,

$$\mathbb{P}\Big(d_H\big(\mathcal{T}(t), \mathcal{T}\big) > \epsilon\Big) \leq \sum_{i=0}^{\infty} \mathbb{P}\Big(d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1} t)\big) > \epsilon_i\Big) \leq \frac{2^{i+1} t}{\epsilon_i} \exp\left(-2^{i-1} t \epsilon_i\right).$$

Set $\epsilon_i = (i+1)\frac{2^{1-i}}{\sqrt{t}}$. Then $ii)$ is satisfied. Indeed, assume t is large enough so $\ln(4) - \sqrt{t} \leq -1$. Then,

$$\begin{aligned}
\sum_{i=0}^{\infty} \mathbb{P}\Big(d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1} t)\big) > \epsilon_i\Big) &= \sum_{i=0}^{\infty} \frac{2^{i+1} 2^{i-1}}{i+1} t^{\frac{3}{2}} \exp\Big(-\sqrt{t}(i+1)\Big), \\
&\leq t^{\frac{3}{2}} \sum_{i=0}^{\infty} \exp\Big(\ln(4) - \sqrt{t}\Big)^{i+1}, \\
&\leq t^{\frac{3}{2}} \frac{4\exp(-\sqrt{t})}{1 - 4\exp(-\sqrt{t})}, \\
&\xrightarrow[t\to\infty]{} 0.
\end{aligned}$$

Clearly, $i)$ is satisfied too as,

$$\lim_{t\to\infty}\sum_{i=0}^{\infty}\epsilon_i(t)=\lim_{t\to\infty}\frac{1}{\sqrt{t}}\sum_{i=0}^{\infty}\frac{i+1}{2^{i-1}}=\lim_{t\to\infty}\frac{8}{\sqrt{t}}=0.$$

This concludes the proof. □

Remark 4.1.14. In the above reasoning, we used the bound from Lemma 4.1.13. If instead we used the bound from Lemma 4.1.11, a similar approach works with the same choice of ϵ_i. We leave filling in the details to the interested reader.

4.2 Compactness of the Discrete Tree

In this section, we prove $ii)$ of Proposition 2.4.3. That is, we aim to show,

Theorem 4.2.1. For all $\epsilon > 0$, we have,

$$\lim_{t\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(d_H\left(\mathcal{T}_n(tn^{\frac{1}{2}}),\mathcal{T}_n\right)>\epsilon n^{\frac{1}{2}}\right)=0.$$

Due to the presence of both n and t, this result might seem more difficult to obtain. However, we will see that we can follow similar reasoning to Section 4.1, with minor modifications to various proofs to adapt them to the discrete setting. One such modification is immediately made. We work with the event,

$$A_n=\left\{C_j^n>\alpha j \text{ for all } j\in\{1,\dots,N\}\right\},$$

where $\alpha > 1$ is a constant to be determined and N denotes the random number of repeats in S_n. Instead, we try to show,

Lemma 4.2.2. For all $\epsilon > 0$, we have,

$$\lim_{t\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(d_H\left(\mathcal{T}_n(tn^{\frac{1}{2}}),\mathcal{T}_n\right)>\epsilon n^{\frac{1}{2}},A_n\right)=0.$$

Remark 4.2.3. To show that Lemma 4.2.2 implies Theorem 4.2.1, it is enough to show $\mathbb{P}(A_n^c)\to 0$ as $n\to\infty$, as,

$$\mathbb{P}\big(d_H\big(\mathcal{T}_n(tn^{\frac{1}{2}}),\mathcal{T}_n\big)>\epsilon\big)\le\mathbb{P}\big(d_H\big(\mathcal{T}_n(tn^{\frac{1}{2}}),\mathcal{T}_n\big)>\epsilon,A_n\big)+\mathbb{P}\big(A_n^c\big).$$

We show this claim in Lemma 4.2.14 at the end of this section. The reason for working on A_n will become clear later. We start with the discrete counterpart of Lemma 4.1.2.

Lemma 4.2.4. Let $\epsilon > 0$ and let $\epsilon_i \equiv \epsilon_i(T) > 0$, for some $T > 0$ such that $\sum_{i=0}^{\infty}\epsilon_i < \epsilon$. Then, for any $t > 0$,

$$\mathbb{P}\big(d_H\big(\mathcal{T}_n\big(tn^{\frac{1}{2}}\big),\mathcal{T}_n\big)>\epsilon n^{\frac{1}{2}},A_n\big)\le\sum_{i=0}^{\infty}\mathbb{P}\big(d_H\big(\mathcal{T}_n\big(2^itn^{\frac{1}{2}}\big),\mathcal{T}_n\big(2^{i+1}tn^{\frac{1}{2}}\big)\big)>\epsilon_i n^{\frac{1}{2}},A_n\big).$$

Proof. The proof is analogous to that of 4.1.2. □

Corollary 4.2.5. Lemma 4.2.2 is shown upon finding $\epsilon_i(t) : [0, \infty) \to [0, \infty)$ such that,

$$i) \lim_{t\to\infty} \sum_{i=0}^{\infty} \epsilon_i(t) = 0 \text{ and } ii) \lim_{t\to\infty} \limsup_{n\to\infty} \sum_{i=0}^{\infty} \mathbb{P}\Big(d_H\big(\mathcal{T}(2^i t n^{\frac{1}{2}}), \mathcal{T}(2^{i+1} t n^{\frac{1}{2}})\big) > \epsilon_i n^{\frac{1}{2}}, A_n\Big) = 0.$$

Fix $a, c \in \mathbb{N}$ and $s \in \{a+1, \dots, 2a\}$. Let v_s denote the vertex in $\mathcal{T}_n$ corresponding to $S_n(s)$. We aim to bound $\mathbb{P}\big(d_H\big(\mathcal{T}_n(a), \mathcal{T}_n(v_s)\big) > c\big)$. For this, we introduce notation for the ancestral line of a vertex in $\mathcal{T}_n$.

Definition 4.2.6. For $s \in [n-1]$, let v_s be the label of the vertex corresponding to $S_n(s)$. Let $p(v_s) \in [n]$ be the label of the parent of v_s where the parent of the root is considered to be the root itself. Inductively, set $p^k(v_s) = p(p^{k-1}(v_s))$.

Example 4.2.7. In Figure 22, we draw the tree corresponding to $S_8 = \{1, 8, 4, 4, 8, 1, 7\}$. We have $v_3 = 4$ and $p(4) = 8$. Similarly, $v_4 = 3$ and $p^3(3) = 1$.

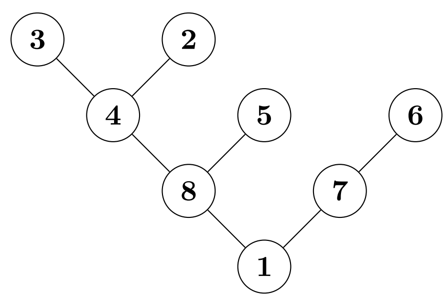


Figure 22: Tree T_8 corresponding to $S_n = \{1, 8, 4, 4, 8, 1, 7\}$.

Lemma 4.2.8. Let $s \in [n]$ and v_s be the vertex corresponding to s. We have,

$$\mathbb{P}\Big(p^i(v_s) = k \;\Big|\; v_s, \dots, p^{i-1}(v_s) \text{ and } p^{i-1}(v_s) \neq S_n(1)\Big) = \begin{cases} \frac{1}{n-i} & \text{if } k \notin \{v_s, \dots, p^{i-1}(v_s)\}, \\ 0 & \text{if } k \in \{v_s, \dots, p^{i-1}(v_s)\}. \end{cases}$$

I.e. given that $p^{i-1}(v_s)$ is not the root, $p^i(v_s)$ is uniform over the vertices not yet seen on the ancestral line of v_s.

Proof. This follows from vertex exchangeability. Indeed, for $a_0, \dots, a_{i-1}, x \in [n]$, define,

$$T_x = \{T_n : v_s = a_0, p(v_s) = a_1, \dots, p_{i-1}(v_s) = a_{i-1}, p_i(v_s) = x\},$$

for the set of trees with ancestral line $v_s = a_0, \dots, p^{i-1}(v_s) = a_{i-1}$ for which $p^i(v_s) = x$. Note,

$$\mathbb{P}\Big(p^i(v_s) = x \;\Big|\; v_s = a_0, \dots, p^{i-1}(v_s) = a_{i-1} \text{ and } a_{i-1} \neq S_n(1)\Big) = \frac{|T_x|}{\sum_{y\in I} |T_y|},$$

where $I = [n] \setminus \{a_0, \dots, a_{i-1}\}$. Next, let $y \in I$ and define $\sigma : [n] \to [n]$ to be the permutation swapping x and y. Permutation σ is its own inverse, and hence a bijection on $[n]$ which leaves $a_0, \dots, a_{i-1}$ untouched. Observe $\sigma(T_x) = T_y$, hence, $|T_x| = |T_y|$. In particular,

$$\mathbb{P}\Big(p^i(v_s) = x \mid v_s = a_0, \dots, p^{i-1}(v_s) = a_{i-1} \text{ and } a_{i-1} \neq S_n(1)\Big),$$

is identical for all $x \in [n] \setminus \{a_0, \dots, a_{i-1}\}$. This proves the lemma as for any realization of $v_s, p(v_s), \dots, p^{i-1}(v_s)$ with $p^{i-1}(v_s) \neq S_n(1)$, we see $p^i(v_s) \sim \mathrm{Unif}([n] \setminus \{v_s, \dots, p^{i-1}(v_s)\})$. □

Lemma 4.2.9. We have $d_H\big(\mathcal{T}_n(a), v_s\big) \leq m = \min_{k\in\mathbb{N}} \big\{p^k(v_s) \in \big\{S_n(1), \dots, S_n(a)\big\}\big\}$.

Proof. Observe that $\{v_s, p(v_s), \dots, p^m(v_s)\}$ is a path of length m from v_s to $\mathcal{T}_n(a)$. Hence we obtain $d_H(\mathcal{T}_n(a), v_s) \leq \min_{k\in\mathbb{N}} \big\{p^k(v_s) \in \{S_n(1), \dots, S_n(a)\}\big\}$. □

Remark 4.2.10. In Lemma 4.2.9, we cannot in general have an equality. Indeed, if $p^k(v)$ is a leaf of $\mathcal{T}_n(a)$, then $p^k(v) \notin \{S_n(1), \dots, S_n(a)\}$ and thus we need to look a step further before finding an element in $\{S_n(1), \dots, S_n(a)\}$. The Lemma explains why we work on A_n: Event A_n ensures $\{S_n(1), \dots, S_n(a)\}$ can have at most $\frac{a}{\alpha}$ repeats and thus $\#\{S_n(1), \dots, S_n(a)\} \geq a\frac{\alpha-1}{\alpha}$, which lower bounds $\mathbb{P}\big(p(v) \in \{S_n(1), \dots, S_n(a)\}\big)$.

Lemma 4.2.11. Let $a, c \in \mathbb{N}$ and fix $s \in \{a+1, \dots, 2a\}$. Let v_s be the vertex corresponding to $S_n(s)$. Then,

$$\mathbb{P}\big(d_H\big(\mathcal{T}_n(a), \mathcal{T}_n(v_s)\big) > c, A_n\big) \leq \exp\left(-\frac{ac}{n}\frac{\alpha-1}{\alpha}\right).$$

Proof. From Lemma 4.2.9, we get,

$$\begin{aligned}
\mathbb{P}\big(d_H\big(\mathcal{T}_n(a), v_s\big) > c, A_n\big) &\leq \mathbb{P}\Big(\min_{k\in\mathbb{N}} \big\{p^k(v_s) \in \big\{S_n(1), \dots, S_n(a)\big\}\big\} > c, A_n\Big), \\
&\leq \mathbb{P}\left(\min_{k\in\mathbb{N}} \left\{p^k(v_s) \in \left\{1, \dots, \left\lceil a\frac{\alpha-1}{\alpha}\right\rceil\right\}\right\} > c\right),
\end{aligned}$$

where working on A_n guarantees $\#\{S_n(1), \dots, S_n(a)\} > a\frac{\alpha-1}{\alpha}$ and where we used vertex exchangeability. Recall from Lemma 4.2.8 that $p^k(v_s) \in_u [n] \setminus \{v_s, \dots, p^{k-1}(v_s)\}$ for $p^{k-1}(v_s)$ not the root (which is guaranteed since $p^{k-1}(v_s) = S_n(1) \implies p^{k-1}(v_s) \in \mathcal{T}(a)$). Thus,

$$\begin{aligned}
\mathbb{P}\big(d_H\big(\mathcal{T}_n(a), v_s\big) > c, A_n\big) &\leq \mathbb{P}\left(\min_{k\in\mathbb{N}} \left\{p^k(v_s) \in \left\{1, \dots, \left\lceil a\frac{\alpha}{\alpha-1}\right\rceil\right\}\right\} > c\right), \\
&= \prod_{i=0}^{c} \left(1 - \frac{a}{(n-i)}\frac{\alpha-1}{\alpha}\right), \\
&\leq \left(1 - \frac{a}{n}\frac{\alpha-1}{\alpha}\right)^{c+1}, \\
&\leq \exp\left(-\frac{ac}{n}\frac{\alpha-1}{\alpha}\right),
\end{aligned}$$

where we used $(1-x) \leq e^{-x}$ in the last step. This concludes the proof. □

We upgrade this to a bound for $\mathbb{P}\big(d_H\big(\mathcal{T}_n(a), \mathcal{T}_n(2a)\big) > c\big)$ using the same trick used in the proof of Lemma 4.1.13.

Lemma 4.2.12. Fix $a, c \in \mathbb{N}$. We have,

$$\mathbb{P}\big(d_H\big(\mathcal{T}_n(a), \mathcal{T}_n(2a)\big) > c, A_n\big) \leq \frac{2a}{c} \exp\left(-\frac{ac}{2n}\frac{\alpha-1}{\alpha}\right)$$

Proof. From Lemma 4.2.11, we immediately get,

$$\mathbb{E}\left[\#\left\{s \in \{a+1, \ldots, 2a\} : d_H\big(\mathcal{T}_n(a), v_s\big)\right\} > \frac{c}{2}, A_n\right] \leq a \exp\left(-\frac{ac}{2n}\frac{\alpha-1}{\alpha}\right).$$

If vertex $v \in \mathcal{T}_n(2a) \setminus \mathcal{T}_n(a)$ satisfies $d_H\big(\mathcal{T}_n(a), v\big) > c$, then all $x \in \big\{v, p(v), \ldots, p^{\lfloor c/2 \rfloor + 1}(v)\big\}$ satisfy $d_H\big(\mathcal{T}_n(a), x\big) > \frac{c}{2}$. Hence,

$$\mathbb{E}\left[\#\left\{s \in \{a+1, \ldots, 2a\} : d_H\big(\mathcal{T}_n(a), v_s\big)\right\} > \frac{c}{2}, A_n\right] \geq \frac{c}{2}\mathbb{P}\big(d_H\big(\mathcal{T}_n(a), \mathcal{T}_n(2a)\big) > c, A_n\big).$$

Combining both bounds gives the desired result. □

We are now in a position to prove Lemma 4.2.2.

Proof of Lemma 4.2.2. By substituting $a = 2^i t n^{\frac{1}{2}}$ and $c = \epsilon_i n^{\frac{1}{2}}$ into the above bound, we get

$$\begin{aligned}\sum_{i=0}^{\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(2^i t n^{\frac{1}{2}}), \mathcal{T}(2^{i+1} t n^{\frac{1}{2}})\big) > \epsilon_i n^{\frac{1}{2}}, A_n\big) &\leq \sum_{i=0}^{\infty} \frac{2^{i+1} t n^{\frac{1}{2}}}{\epsilon_i n^{\frac{1}{2}}} \exp\left(-\frac{2^i t n^{\frac{1}{2}} \epsilon_i n^{\frac{1}{2}}}{4n}\right), \\ &= \sum_{i=0}^{\infty} \frac{2^{i+1} t}{\epsilon_i} \exp\left(-2^{i-2} t \epsilon_i\right)\end{aligned}$$

Thus, Corollary 4.2.5 is transformed into finding $\epsilon_i(t)$ such that,

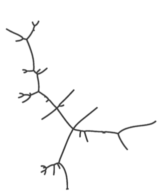
$$i) \lim_{t\to\infty} \sum_{i=0}^{\infty} \epsilon_i(t) = 0, \qquad \text{and} \qquad ii) \lim_{t\to\infty} \sum_{i=0}^{\infty} \frac{2^{i+1} t}{\epsilon_i} \exp\left(-2^{i-2} t \epsilon_i\right).$$

We choose $\epsilon_i(t) = (i+1)\frac{2^{2-i}}{\sqrt{t}}$. Analogous computations to that of Proof 4.1 show this choice of ϵ_i satisfies both $i)$ and $ii)$, finishing the proof. □

It remains to show Lemma 4.2.2 implies Theorem 4.2.1. It is enough to show $\mathbb{P}(A_n^c) \xrightarrow[n\to\infty]{} 0$. For this, we bound the lower tail of C_j^n.

Lemma 4.2.13. For $x < n$, we have,

$$\mathbb{P}(C_j^n < x) \leq e\left(\frac{x^2 e}{2nj}\right)^j \exp\left(-\frac{x^2}{2n}\right) \leq e\left(\frac{x^2 e}{2nj}\right)^j.$$



Proof. Observe that $C_j^n < x$ happens when $S_n(1), \dots, S_n(x-1)$ contains at least j repeats. Also observe that,

$$\mathbb{P}\big(S_n(i) \text{ is a repeat } \big| \, S_n(1), \dots, S_n(i-1)\big) \leq \frac{i-1}{n}.$$

Hence, $\mathbb{P}(C_j^n < x) \leq \mathbb{P}(X \geq j)$ where $X \sim \mathrm{Ber}\big(\frac{1}{n}\big) + \dots + \mathrm{Ber}\big(\frac{x-1}{n}\big)$ with independent Bernoulli random variables. Recall the Chernoff bound for the sum of independent Bernoulli random variables with mean μ and where $\delta > -1$,

$$\mathbb{P}(X > (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} = \frac{e^{(\delta+1)\mu}}{(1+\delta)^{(1+\delta)\mu}} e^{-\mu}.$$

In our case, $\mu = \sum_{i=1}^{x-1} \frac{i}{n} = \frac{(x-1)x}{2n}$ and $(1+\delta)\mu = j$. Filling this in yields the bound,

$$\begin{aligned}
\mathbb{P}(C_j^n < x) \leq \mathbb{P}(X > j) \leq \left(\frac{\mu e}{j}\right)^j e^{-\mu} &\leq \left(\frac{x^2 e}{2nj}\right)^j \exp\left(-\frac{(x-1)^2}{2n}\right), \\
&\leq e\left(\frac{x^2 e}{2nj}\right)^j \exp\left(-\frac{x^2}{2n}\right),
\end{aligned}$$

where we used $\exp\big(-\frac{(x-1)^2}{2n}\big) = \exp\big(-\frac{x^2}{2n} + \frac{x}{n} - \frac{1}{2n}\big) \leq \exp\big(-\frac{x^2}{2n} + 1\big)$ since $x < n$. This gives the desired bound. □

With this tool, we show the following lemma, from which it follows that $\mathbb{P}(A_n^c) \to 0$.

Lemma 4.2.14. Let N be denote the number of repeats in S_n. Then, there exists $\beta > \frac{1}{2}$ and $\alpha > 1$ such that for n large enough and $k \in \mathbb{N}$,

$$i)\ \mathbb{P}(N > \beta n) \leq \exp(-c_1 n) \quad \text{and} \quad ii)\ \sum_{j=k}^{\lfloor \beta n \rfloor} \mathbb{P}(C_j^n < \alpha j) \leq n^{-\frac{k}{3}}.$$

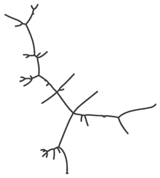
for c_1 a positive constant.

Proof. For i), Observe $N > \beta n \implies C_{\lfloor \beta n \rfloor}^n < n$, as the $\lfloor \beta n \rfloor$'th repeat must happen at or before index $n-1$ in S_n. By Lemma 4.2.13, we obtain,

$$\mathbb{P}\Big(N > \beta n\Big) \leq \mathbb{P}(C_{\beta n}^n < n) \leq e\left(\frac{n^2 e}{2n\beta n}\right)^{\beta n} \exp\left(-\frac{n^2}{2n}\right) \leq e \exp\Big(n\Big(\beta\big(1 - \ln(2\beta)\big) - \frac{1}{2}\Big)\Big),$$

Since $\beta\big(1 - \ln(2\beta)\big) - \frac{1}{2} < 0$ for $\beta > \frac{1}{2}$, we have shown i). For ii) we use Lemma 4.2.13 to get,

$$\begin{aligned}
\sum_{j=k}^{\lfloor \beta n \rfloor} \mathbb{P}(C_j^n > \alpha j) &\leq \sum_{j=k}^{\lfloor \beta n \rfloor} \left(\frac{j\alpha^2 e}{2n}\right)^j \exp\left(\frac{-j^2\alpha^2 e}{2n}\right), \\
&\leq \sum_{j=k}^{\lfloor n^{\frac{2}{3}} \rfloor} \left(\frac{j\alpha^2 e}{2n}\right)^j + \sum_{j=\lceil n^{\frac{2}{3}} \rceil}^{\beta n} \left(\frac{\beta\alpha^2 e}{2}\right)^j \exp\left(-\frac{j^2\alpha^2 e}{2n}\right), \\
&\leq C_1 n^{-\frac{k}{3}} + \beta n \exp\left(-C_2 n^{\frac{1}{3}}\right),
\end{aligned}$$



where $C_1, C_2 > 0$ are positive constants. Here we used $\left(\frac{\beta\alpha^2 e}{2}\right) < 1$ in the last step. This is possible for $\beta > \frac{1}{2}$ and $\alpha > 1$ as for example the choice $\beta = 0.6$ and $\alpha = 1.1$ works. □

Corollary 4.2.15. We can find $\alpha > 1$ such that $\mathbb{P}(A_n^c) \to 0$ as $n \to \infty$.

Proof. By a union bound, we have,

$$\mathbb{P}(A_n^c) \leq \mathbb{P}\big(\exists j \in \{1, \dots, \lfloor \beta n \rfloor\} : C_j^n < \alpha j \text{ or } N > \lfloor \beta n \rfloor\big) \leq \mathbb{P}(N > \beta n) + \sum_{j=1}^{\lfloor \beta n \rfloor} \mathbb{P}(C_j^n < \alpha j).$$

By Lemma 4.2.14, we can find β and $\alpha > 1$ such that both terms go to zero. □

Thus we have shown criteria *i)* and *ii)* of Proposition 2.4.3 are satisfied. Together with convergence of finite dimensional distributions, Theorem 1.5.1, this shows that,

$$\big(\mathcal{T}_n, n^{-\frac{1}{2}} d_n\big) \xrightarrow[n\to\infty]{d} \big(\mathcal{T}, d\big),$$

in the Gromov–Hausdorff topology. In the next section, we will add the measures ν_n and μ to this convergence, and upgrade the result to Gromov–Hausdorff–Prokhorov convergence.

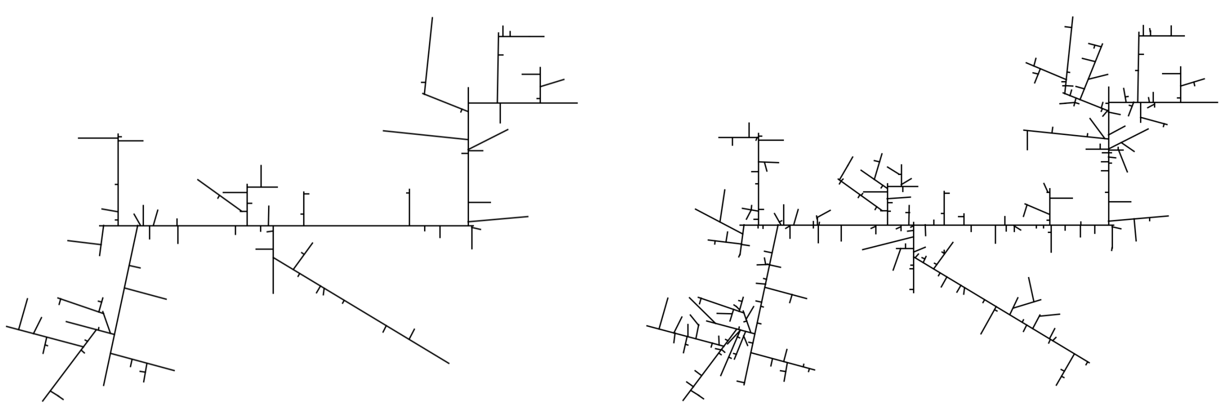
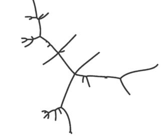


Figure 23: Difference between $\mathcal{T}^{(100)}$ (left) and $\mathcal{T}^{(300)}$ (right).



5 Tightness of the Measures

In this section, we show *iii*) and *iv*) of Proposition 2.4.3. That is, there exists a probability measure μ on $\mathcal{T}$ such that for all $\epsilon > 0$,

$$iii) \lim_{k\to\infty} \mathbb{P}\left(d_P\left(\mu^{(k)}, \mu\right) > \epsilon)\right) = 0 \quad \text{and} \quad iv) \lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(d_P\left(\nu_n^{(k)}, \nu_n\right) > \epsilon\right) = 0.$$

Recall that $\mu^{(k)}$ is the pushforward of the normalized Lebesgue measure on $[0, C_k]$ by ρ, and that $\nu_n(A) = \frac{|A|}{n}$ and $\nu_n^{(k)}(A) = \frac{|A|}{|\mathcal{T}_n^{(k)}|}$ are the uniform probability measure on the vertices of $\mathcal{T}_n$ and $\mathcal{T}_n^{(k)}$ respectively. We start with proving *iii*).

5.1 Convergence of $\mu^{(k)}$ to μ

Recall Theorem 2.3.11. Since ℓ^1 is Polish, we have that $(\mathcal{P}(\ell^1), d_P)$ is Polish. As $\mu^{(k)} \in \mathcal{P}(\ell^1)$, it is enough to show $\left(\mu^{(k)}\right)_{k\in\mathbb{N}}$ is Cauchy as then $\mu^{(k)}$ has a unique limit which we can take as μ. Thus our aim is to show that for all $\epsilon > 0$, there exists $N > 0$ such that $k, m > N$ implies $\mathbb{P}(d_P\left(\mu^{(k)}, \mu^{(m)}\right) < \epsilon) = 1$.

By definition of the Prokhorov distance, it suffices to show that for large enough k, m and for all measurable $A \subset \mathcal{T}$,

$$\mu^{(k)}(A) \leq \mu^{(m)}(A^\epsilon) + \epsilon.$$

Here, we take $A \subset \mathcal{T}$ instead of $A \subset \ell^1$ since $\mu^{(k)}(B) = \mu^{(m)}(B) = 0$ whenever $B \cap \mathcal{T} = \emptyset$. We give some intuition why $\mu^{(k)}(A) \leq \mu^{(m)}(A^\epsilon) + \epsilon$ holds true. For this, the measures $\mu^{(k)}$ and $\mu^{(m)}$ may differ by ϵ in two different ways: they should assign assign roughly the same measure (i.e. allowed to differ by ϵ) to roughly the same sets (can enlarge set by ϵ).

Note that $\mu^{(k)}(A)$ is the proportion of A in $\mathcal{T}^{(k)}$. We will show that the expected proportion of tree $\mathcal{T} \setminus \mathcal{T}^{(k)}$ attached to $A \subset \mathcal{T}^{(k)}$ is the same as the proportion of A in $\mathcal{T}^{(k)}$. We illustrate this in Figure 24, where $A^\uparrow$ denotes A together with the branches attached to A.

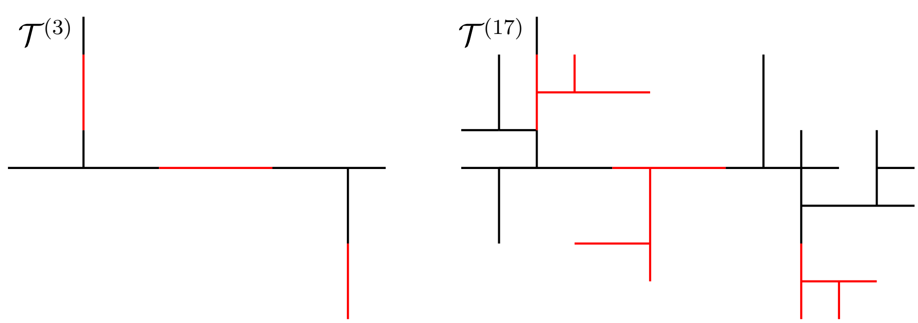


Figure 24: Left: $A \subset \mathcal{T}^{(3)}$ drawn in red. Right: $A^\uparrow$ drawn in red.

This property will allow us to conclude that $\mu^k(A^\uparrow)$ converges as $k \to \infty$. Thus for large enough k, m we have $|\mu^{(k)}(A^\uparrow) - \mu^{(m)}(A^\uparrow)| \leq \epsilon$. This shows that $\mu^{(k)}$ and $\mu^{(m)}$ are roughly the same on sets of the form $A^\uparrow$. Lastly, we use that $\mathbb{P}(d_H(\mathcal{T}^{(k)}, \mathcal{T}) < \epsilon) \to 1$ as $k \to \infty$ to conclude that for k large enough, $A \subset \mathcal{T}^{(k)}$ is such that $A^\uparrow \subset A^\epsilon$. This will allow us to conclude $d_P(\mu^{(k)}, \mu^{(m)}) \leq \epsilon$. Below we make this reasoning formal. We start with some definitions and a result on martingales.

Definition 5.1.1. For $A \subset \mathcal{T}^{(k)}$, let $A^{\uparrow} = \pi_k^{-1}(A) \cap \mathcal{T}$, with $\pi_k : \ell^1 \to \mathbb{R}^k$ the projection map. Observe that $A^{\uparrow}$ consists of A together with points in $\mathcal{T} \setminus \mathcal{T}^{(k)}$ whose path to $\mathcal{T}^{(k)}$ ends in A.

To shorten notation, we write $c_i = C_i - C_{i-1}$.

Remark 5.1.2. Note that $A^{\uparrow}$ for $A \subset \mathcal{T}^{(k)}$ depends on k as seen in Figure 25. In both the top and bottom figure, we have $A = \mathcal{T}^{(1)}$. However, in the top case, we view $A \subset \mathcal{T}^{(1)}$ and in the bottom case, we view $A \subset \mathcal{T}^{(2)}$. We always write 'let $A \subset \mathcal{T}^{(k)}$' to indicate that $A^{\uparrow}$ is taken with respect to k.

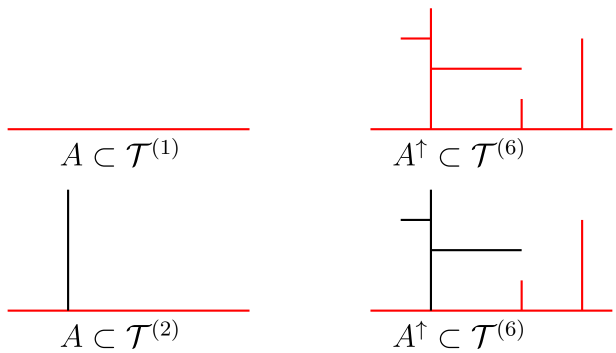


Figure 25: $A^{\uparrow}$ depends on k

Theorem 5.1.3. Suppose $(X_1, X_2, \dots)$ is a martingale such that $\sup_n |X_n| < \infty$, then there is a random variable X such that $\lim_{n\to\infty} X_n = X$ with probability 1.

Lemma 5.1.4. Let $A \subset \mathcal{T}^{(k)}$. Then $\mu^{(j)}(A^{\uparrow})$ is a martingale for $j \geq k$ in filtration $\sigma(\mathcal{T}^{(j)})$.

Proof. It is clear that $\mu^{(j)}$ is $\sigma(\mathcal{T}^{(j)})$ measurable and $\mathbb{E}[|\mu^{(j)}(A^{\uparrow})|] < \infty$ as $\mu^{(j)}(A^{\uparrow}) \in [0, 1]$. It remains to show $\mathbb{E}[\mu^{(j+1)}(A^{\uparrow}) \mid \mathcal{F}_j] = \mu^{(j)}(A^{\uparrow})$. For this, define $G_j = \sigma(F_j, c_{j+1}, C_{j+1})$. Observe that conditional on $\mathcal{T}^{(j)}$ the $j + 1$'th branch is part of $A^{\uparrow}$ with probability $\mu^{(j)}(A^{\uparrow})$. Depending on weather branch $j + 1$ is added to $A^{\uparrow}$ or not, we see,

$$\mu^{(j+1)}(A^{\uparrow}) = \begin{cases} (C_j\mu^{(j)} + c_{j+1})/C_{j+1} & \text{with probability } \mu^{(j)}(A^{\uparrow}), \\ C_j\mu^{(j)}/C_{j+1} & \text{with probability } 1 - \mu^{(j)}(A^{\uparrow}). \end{cases}$$

By putting the two cases together, we obtain,

$$\mathbb{E}\left[\mu^{(j+1)}(A^{\uparrow}) \mid G_j\right] = \mu^{(j)}(A^{\uparrow})\frac{C_j\mu^{(j)}(A^{\uparrow}) + c_{j+1}}{C_{j+1}} + (1 - \mu^{(j)}(A^{\uparrow}))\frac{C_j\mu^{(j)}(A^{\uparrow})}{C_{j+1}} = \mu^{(j)}(A^{\uparrow}).$$

By the tower property, we obtain,

$$\mathbb{E}\left[\mu^{(j+1)}(A^{\uparrow}) \mid \sigma(\mathcal{T}^{(j)})\right] = \mathbb{E}\left[\mathbb{E}\left[\mu^{(j+1)}(A^{\uparrow}) \mid G_j\right] \mid \sigma(\mathcal{T}^{(j)})\right] = \mu^{(j)}(A^{\uparrow}),$$

where we used that $\mu^{(j)}(A^{\uparrow})$ is $\sigma(\mathcal{T}^{(j)})$ measurable. This concludes the proof. ☐

Remark 5.1.5. Since $\mu^{(j)}(A^{\uparrow}) \in [0,1]$, we apply Theorem 5.1.3 to find that there must exists some random variable $\mu(A^{\uparrow})$ such that $\lim_{j\to\infty} \mu^{(j)}(A^{\uparrow}) = \mu(A^{\uparrow})$ almost surely. In particular, this means that for sufficiently large k, m we have $|\mu^{(j)}(A^{\uparrow}) - \mu^{(m)}(A^{\uparrow})| < \epsilon$ almost surely.

Theorem 5.1.6. The sequence $\left(\mu^{(j)}\right)_{j\in\mathbb{N}}$ is almost surely Cauchy in $(\mathcal{P}(\ell^1), d_P)$.

Proof. Fix $\epsilon > 0$ and let K be large enough so that $\mathbb{P}(d_H(\mathcal{T}^{(K)}, \mathcal{T}) > \epsilon) < \epsilon$, which is possible by Theorem 4.1.1. Condition on $\mathcal{T}^{(K)}$ and let $J_1, \dots, J_{N_\epsilon}$ be a measurable partition of $\mathcal{T}^{(K)}$ for which $\mathrm{diam}(J_i) < \epsilon$ for all i. Conditional on the event $E = \left\{d_H\left(\mathcal{T}^{(K)}, \mathcal{T}\right) < \epsilon\right\}$, we have $\mathrm{diam}(J_i^{\uparrow}) \leq 3\epsilon$ since for any $x, y \in J_i^{\uparrow}$, we can bound,

$$\begin{aligned} d(x,y) &\leq d(x, \pi_K(x)) + d(\pi_K(x), \pi_K(y)) + d(\pi_K(y), y), \\ &\leq d_H(\mathcal{T}^{(K)}, \mathcal{T}) + \mathrm{diam}(J_i) + d_H(\mathcal{T}^{(K)}, \mathcal{T}) = 3\epsilon, \end{aligned}$$

where d denotes the distance in $\mathcal{T}$ and π_K is the projection map from $\ell^1 \to \mathbb{R}^k$. The proof is illustrated in Figure 26. Be aware this is a two-dimensional representation of $\mathcal{T} \subset \ell^1$, thus $d(x,y)$ should not be interpreted as Euclidean distance but instead be seen as distance traversed over the drawn branches.

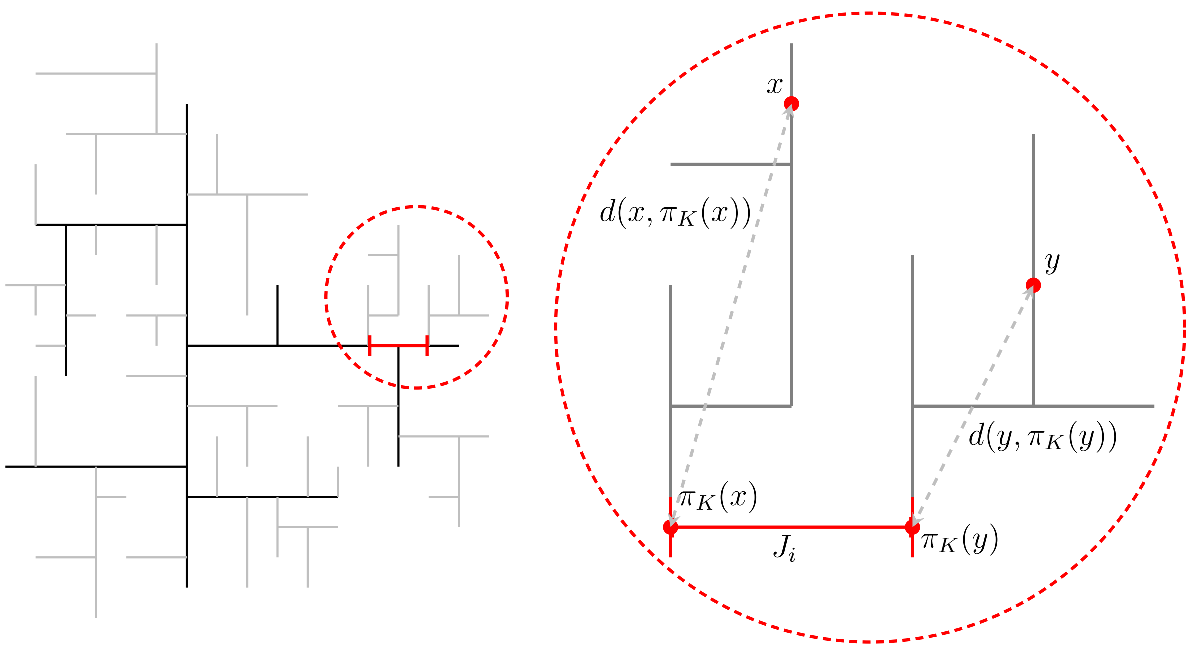


Figure 26: Illustration $\mathrm{diam}(J_i^{\uparrow}) \leq 3\epsilon$

Conditional on E, the family $\left(J_i^{\uparrow}\right)_{i\in[N_\epsilon]}$ is a partition of $\mathcal{T}$ of sets of diameter at most 3ϵ. From Lemma 2.3.13, we get,

$$d_P\left(\mu^{(j)}, \mu^{(m)}\right) \leq \sum_{i=1}^{N_\epsilon} |\mu^{(j)}(J_i^{\uparrow}) - \mu^{(m)}(J_i^{\uparrow})| + 3\epsilon + \mathbb{1}_{\{E^c\}}.$$

By applying union bounds, we get for all $\delta > 0$ and $b > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(d_P(\mu^{(j)}, \mu^{(m)}) \geq \delta N_\epsilon + 3\epsilon \;\middle|\; \mathcal{T}^{(K)}\right), \\
\leq\, &\mathbb{P}\left(\sum_{i=1}^{N_\epsilon} |\mu^{(j)}(J_i^{\uparrow}) - \mu^{(m)}(J_i^{\uparrow})| \geq \delta N_\epsilon \;\middle|\; \mathcal{T}^{(K)}\right) \mathbb{1}_{\{N_\epsilon \leq b\}} + \mathbb{1}_{\{N_\epsilon > b\}} + \mathbb{1}_{\{E^c\}} \\
\leq\, &\sum_{i=1}^{N_\epsilon} \mathbb{P}\left(|\mu^{(j)}(J_i^{\uparrow}) - \mu^{(m)}(J_i^{\uparrow})| > \delta \;\middle|\; \mathcal{T}^{(K)}\right) \mathbb{1}_{\{N_\epsilon \leq b\}} + \mathbb{1}_{\{N_\epsilon > b\}} + \mathbb{1}_{\{E^c\}}.
\end{aligned}
$$

By setting $\delta = \frac{\epsilon}{N_\epsilon}$ and taking expectations, we obtain for all $b > 0$,

$$
\begin{aligned}
&\mathbb{P}(d_P(\mu^{(j)}, \mu^{(m)}) \geq 4\epsilon), \\
\leq\, &\mathbb{E}\left[\sum_{i=1}^{N_\epsilon} \mathbb{P}\left(|\mu^{(j)}(J_i^{\uparrow}) - \mu^{(m)}(J_i^{\uparrow})| > \frac{\epsilon}{N_\epsilon} \;\middle|\; \mathcal{T}^{(K)}\right) \mathbb{1}_{\{N_\epsilon \leq b\}} + \mathbb{1}_{\{N_\epsilon > b\}} + \mathbb{1}_{\{E^c\}}\right], \\
\leq\, &\sum_{i=1}^{b} \mathbb{P}\left(|\mu^{(j)}(J_i^{\uparrow}) - \mu^{(m)}(J_i^{\uparrow})| > \frac{\epsilon}{b}\right) + \mathbb{P}(N_\epsilon > b) + \mathbb{P}\big(d_H\big(\mathcal{T}^{(K)}, \mathcal{T}\big) > \epsilon\big).
\end{aligned}
$$

By choice of K, we have $\mathbb{P}\big(d_H\big(\mathcal{T}^{(K)}, \mathcal{T}\big) > \epsilon\big) < \epsilon$. Recall that N_ϵ is the minimal number of sets J_i of diameter ϵ needed to partition $\mathcal{T}^{(K)}$. Since $\mathcal{T}^{(K)} \subset \mathcal{T}$ is compact almost surely, we can make $\mathbb{P}(N_\epsilon > b)$ arbitrarily small with finite b. By Remark 5.1.5, $\mathbb{P}\left(|\mu^{(j)}(J_i^{\uparrow}) - \mu^{(m)}(J_i^{\uparrow})| > \frac{\epsilon}{b}\right)$ can be made arbitrarily small by taking n, m large enough. We conclude, for n, m large enough $\mathbb{P}(d_P(\mu^{(j)}, \mu^{(m)}) > 4\epsilon) < \epsilon$, which shoes $\left(\mu^{(j)}\right)_{j \in \mathbb{N}}$ is Cauchy in probability. □

Since $\mu^{(j)}$ is Cauchy in probability in complete space $(\mathcal{P}(\ell^1), d_P)$, we conclude $\mu^{(k)} \xrightarrow[k\to\infty]{} \mu$ in the Prokhorov metric for some measure μ in probability. Hence, there exists a measure μ such that

$$iii) \lim_{k\to\infty} \mathbb{P}\left(d_P\left(\mu^{(k)}, \mu\right) > \epsilon\right)\big) = 0,$$

which shows $iii)$ of Proposition 2.4.3 and finishes this subsection.

5.2 Convergence of $\nu_n^{(k)}$ to ν_n

This section is dedicated towards showing $iv)$ of Proposition 2.4.3. I.e. we show, for all $\epsilon > 0$, we have ,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(d_P(\nu_n^{(k)}, \nu_n) > \epsilon) = 0.$$

For this, we follow similar reasoning to [2], pages 18-20. Due to our different definition of $\mathcal{T}_n$, we use a different derivation of equation 40 and do not use urn models to describe the problem. Also the proof of Theorem 6.3.4 has been streamlined. After introducing some notation, we start with the discrete analogue of Lemma 5.1.4. The paper [3] was of important help in writing this section.

Definition 5.2.1. Color $A \subset \mathcal{T}_n^{(k)}$ red. Inductively, color branch $j > k$ red if it is attached to a red vertex. Let $A^\uparrow \subset \mathcal{T}_n$ be all vertices colored red after adding all branches. Alternatively, for $A \subset \mathcal{T}_n^{(k)}$, we see $A^\uparrow$ consists of A together with all vertices $v \in \mathcal{T}_n \setminus \mathcal{T}_n^{(k)}$ such that the path from v to $\mathcal{T}_n^{(k)}$ ends in v. Observe that $A^\uparrow$ does not have to equal the set of descendants of A, as highlighted in Figure. 27.

We write $c_j^n = C_j^n - C_{j-1}^n$ for the number of vertices on branch j.

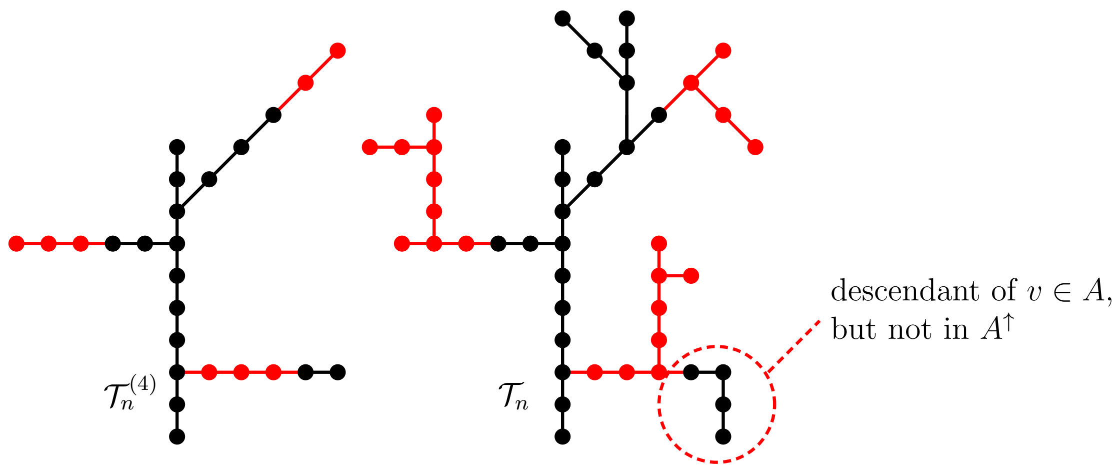


Figure 27: On the left, $A \subset \mathcal{T}_n^{(4)}$ is denoted in red. On the right, we constructed $A^\uparrow \subset \mathcal{T}_n$.

Lemma 5.2.2. Let $A \subset \mathcal{T}_n^{(k)}$. For $j \geq k$, $\nu_n^{(j)}(A^\uparrow)$ is a martingale in filtration $F_j = \sigma(\mathcal{T}_n^{(j)})$.

Proof. Condition on $G_j = \sigma(\mathcal{T}_n^{(j)}, c_{j+1}^n, C_{j+1}^n)$. Branch $j+1$ is attached to a vertex that is already colored red with probability $\nu^{(j)}(A^\uparrow)$. Thus we obtain,

$$\nu_n^{(j+1)}(A^\uparrow) = \begin{cases} (C_j^n \nu_n^{(j)}(A^\uparrow) + c_{j+1}^n)/C_{j+1}^n, & \text{with probability } \nu_n^{(j)}(A^\uparrow), \\ C_j^n \nu_n^{(j)}(A^\uparrow)/C_{j+1}^n, & \text{with probability } 1 - \nu_n^{(j)}(A^\uparrow). \end{cases}$$

Using this, we get,

$$\mathbb{E}\left[\nu_n^{(j+1)}(A^\uparrow) \mid G_j\right] = \nu_n^{(j)}(A^\uparrow) \frac{C_j^n \nu_n^{(j)}(A^\uparrow) + c_{j+1}^n}{C_{j+1}^n} + (1 - \nu_n^{(j)}(A^\uparrow)) \frac{\nu_n^{(j)}(A^\uparrow) C_j^n}{C_{j+1}^n} = \nu_n^{(j)}(A^\uparrow).$$

By the tower property of expectation, we get,

$$\mathbb{E}\left[\nu_n^{(j+1)}(A^\uparrow) \mid F_j\right] = \mathbb{E}\left[\mathbb{E}\left[\nu_n^{(j+1)}(A^\uparrow) \mid G_j\right] \mid F_j\right] = \mathbb{E}\left[\nu_n^{(j)}(A^\uparrow) \mid F_j\right] = \nu_n^{(j)}(A^\uparrow),$$

where the last step follows as $\nu_n^{(j)}$ is F_j-measurable. We conclude $\nu_n^j(A^\uparrow)$ is a martingale. $\square$

Remark 5.2.3. Given that $\nu_n^{(j)}(A^\uparrow)$ is a bounded martingale, we expect $\nu_n^{(j)}(A^\uparrow)$ to converge. This convergence is trivial for any finite n as $\nu_n^{(j)}(A^\uparrow) = \nu_n(A^\uparrow)$ for j greater than the number of sticks in $\mathcal{T}_n$ (which is bounded by n). To show $\limsup_{n\to\infty} \nu_n^{(j)}(A^\uparrow)$, also converges almost surely as $j \to \infty$, we use a more quantitative tail bound given below.

Lemma 5.2.4. Let $A \subset \mathcal{T}_n^{(k)}$. We have,

$$\mathbb{P}\left((\nu_n^{(k)}(A^{\uparrow}) - \nu_n(A^{\uparrow}))^2 \geq C^2 \mid \mathcal{T}_n^{(k)}\right) \leq \frac{X_{k,n}}{C^2},$$

where $X_{k,n}$ is a random variable measurable with respect to $\mathcal{T}_n^{(k)}$ such that,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{E}[X_{k,n}] = 0.$$

By assuming the above lemma holds, we prove,

Theorem 5.2.5. For all $\epsilon > 0$, we have,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(d_p(\nu_n^{(k)}, \nu_n) \geq \epsilon\right) = 0.$$

Proof. Fix $\epsilon > 0$ and take $K > 0$ large enough so that $\mathbb{P}\big(d_H\big(n^{-\frac{1}{2}}\mathcal{T}_n^{(K)}, n^{-\frac{1}{2}}\mathcal{T}_n\big) > \epsilon\big) < \epsilon$ uniformly in n, which is possible by Theorem 4.2.1. Condition on $\mathcal{T}^{(K)}$ and let N_ϵ be the size of the smallest partition $C_1, \ldots, C_{N_\epsilon}$ of $n^{-\frac{1}{2}}\mathcal{T}_n^{(K)}$ where $\operatorname{diam}(C_i) < \epsilon$ for all i. On the event $E = \big\{d_H\big(\mathcal{T}_n^{(K)}, \mathcal{T}_n\big) < \epsilon\big\}$, we have $\operatorname{diam}(C_i^{\uparrow}) < \epsilon + 2d_H(n^{-\frac{1}{2}}\mathcal{T}_n^{(K)}, n^{-\frac{1}{2}}\mathcal{T}_n) \leq 3\epsilon$.

By Lemma 2.3.13 and union bounds, we obtain that for all $\delta > 0$ and $b > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(d_p(\nu_n, \nu_n^{(K)}) \geq \delta N_\epsilon + 3\epsilon \mid \mathcal{T}_n^{(K)}\Big) \\
&\leq \mathbb{P}\left(\sum_{j=1}^{N(\epsilon)} |\nu_n^{(K)}(C_j^{\uparrow}) - \nu_n C_j^{\uparrow}| \geq \delta N_\epsilon \;\middle|\; \mathcal{T}_n^{(K)}\right) + \mathbb{1}_{\{E^c\}} \\
&\leq \sum_{j=1}^{N_\epsilon} \mathbb{P}\left(|\nu_n^{(K)}(C_j^{\uparrow}) - \nu_n(C_j^{\uparrow})| \geq \delta \;\middle|\; \mathcal{T}_n^{(K)}\right) \mathbb{1}_{\{N_\epsilon \leq b\}} + \mathbb{1}_{\{N_\epsilon > b\}} + \mathbb{1}_{\{E^c\}},
\end{aligned}$$

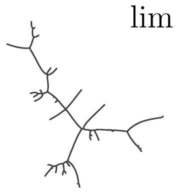
By setting $\delta = \frac{\epsilon}{N(\epsilon)}$, we obtain,

$$\begin{aligned}
&\mathbb{P}\big(d_p(\nu_n, \nu_n^{(K)}) \geq 4\epsilon \mid \mathcal{T}_n^{(K)}\big), \\
&\leq \sum_{j=1}^{N(\epsilon)} \mathbb{P}\left(|\nu_n^{(K)}(C_j^{\uparrow}) - \nu_n(C_j^{\uparrow})| \geq \frac{\epsilon}{N(\epsilon)} \;\middle|\; \mathcal{T}_n^{(K)}\right) \mathbb{1}_{\{N_\epsilon \leq b\}} + \mathbb{1}_{\{N_\epsilon > b\}} + \mathbb{1}_{\{E^c\}}. \\
&\leq \sum_{j=1}^{b} \mathbb{P}\left(|\nu_n^{(K)}(C_j^{\uparrow}) - \nu_n(C_j^{\uparrow})| \geq \frac{\epsilon}{b} \;\middle|\; \mathcal{T}_n^{(K)}\right) + \mathbb{1}_{\{N_\epsilon > b\}} + \mathbb{1}_{\{E^c\}}.
\end{aligned}$$

By taking expectations and using Lemma 5.2.4, we find,

$$\limsup_{n\to\infty} \mathbb{P}\left(d_p(\nu_n, \nu_n^{(K)}) \geq 4\epsilon\right) \leq \frac{b^3}{\epsilon^2}\frac{C}{K-1} + \mathbb{P}(N(\epsilon) > b) + \limsup_{n\to\infty} \mathbb{P}\big(d_H\big(\mathcal{T}_n, \mathcal{T}_n^{(K)}\big) > \epsilon\big).$$

By choice of K, we have $\limsup_{n\to\infty} \mathbb{P}\big(d_H\big(\mathcal{T}_n, \mathcal{T}_n^{(K)}\big) > \epsilon\big) < \epsilon$. Recall that N_ϵ is the number of sets of diameter at most ϵ needed to cover $n^{-\frac{1}{2}}\mathcal{T}_n^{(K)}$. Since $\mathcal{T}_n^{(K)} \subset \mathcal{T}_n$ and $n^{-\frac{1}{2}}\mathcal{T}_n$ is compact almost surely, we can choose b to make $\mathbb{P}(N_\epsilon > b)$ arbitrarily small. We conclude $\limsup_{n\to\infty} \mathbb{P}\left(d_p(\nu_n, \nu_n^{(K)}) \geq 4\epsilon\right) \xrightarrow[K\to\infty]{} 0$ for all $\epsilon > 0$. □



To finish this section, it only remains to verify Lemma 5.2.4. This is done in two steps,

Lemma 5.2.6. For $A \subset \mathcal{T}_n^{(k)}$, we have,

$$\mathbb{P}\left((\nu_n^{(k)}(A^\uparrow) - \nu_n(A^\uparrow))^2 \geq C^2 \mid \mathcal{T}_n^{(k)}\right) \leq \frac{X_{n,k}}{C^2} = \frac{2n^2}{C^2} \sum_{j=k}^{N} \mathbb{E}\left[\frac{1}{(C_j^n - j)^4} \,\middle|\, \mathcal{T}_n^{(k)}\right],$$

where N denotes the random number of branches of $\mathcal{T}_n$. Note that $X_{n,k}$ is a random variable measurable with respect to $\mathcal{T}_n^{(k)}$.

Proof. We apply Markov's inequality to obtain,

$$\mathbb{P}\left((\nu_n^{(k)}(A^\uparrow) - \nu_n(A^\uparrow))^2 \geq C^2 \mid \mathcal{T}_n^{(k)}\right) \leq \frac{1}{C^2}\mathbb{E}\left[(\nu_n^{(k)}(A^\uparrow) - \nu_n(A^\uparrow))^2 \mid \mathcal{T}_n^{(k)}\right].$$

Let N be the number of repeats in S_n so $\nu_n(A^\uparrow) = \nu_n^{(N)}(A^\uparrow)$. Since $(\nu_n^{(k)}(A^\uparrow) - \nu_n(A^\uparrow))^2 \leq 2$ is bounded for all $k \in \mathbb{N}$, we apply the optional stopping theorem to obtain,

$$\begin{aligned}
\mathbb{E}\left[(\nu_n^{(k)}(A^\uparrow) - \nu_n(A^\uparrow))^2 \mid \mathcal{T}_n^{(k)}\right] &= \mathbb{E}\left[\nu_n^{(k)}(A^\uparrow)^2 + \nu_n^{(N)}(A^\uparrow)^2 - 2\nu_n^{(k)}(A^\uparrow)\nu_n^{(N)}(A^\uparrow) \mid \mathcal{T}_n^{(k)}\right], \\
&= \mathbb{E}\left[\nu_n^{(N)}(A^\uparrow)^2 - \nu_n^{(k)}(A^\uparrow)^2 \mid \mathcal{T}_n^{(k)}\right], \\
&= \sum_{j=k}^{N-1} \mathbb{E}\left[\nu_n^{(j+1)}(A^\uparrow)^2 - \nu_n^{(j)}(A^\uparrow)^2 \mid \mathcal{T}_n^{(k)}\right].
\end{aligned}$$

Where the second line follows from,

$$\mathbb{E}\left[2\nu_n^{(k)}(A^\uparrow)\nu_n^{(N)}(A^\uparrow) \mid \mathcal{T}_n^{(k)}\right] = 2\nu_n^{(k)}(A^\uparrow)^2 = \mathbb{E}\left[\nu_n^{(k)}(A^\uparrow)^2 \mid \mathcal{T}_n^{(k)}\right],$$

which again uses the optional stopping theorem. Next, we bound,

$$\begin{aligned}
&\mathbb{E}\left[\nu_n^{(j+1)}(A^\uparrow)^2 - \nu_n^{(j)}(A^\uparrow)^2 \mid \mathcal{T}_n^{(k)}\right] \\
&= \mathbb{E}\left[\nu_n^{(j)}(A^\uparrow)\left(\frac{C_j^n \nu_n^{(j)}(A^\uparrow) + c_{j+1}^n}{C_{j+1}^n}\right)^2 + (1 - \nu_n^{(j)}(A^\uparrow)\left(\frac{C_j^n \nu_n^{(j)}(A^\uparrow)}{C_{j+1}^n}\right)^2 - \nu_n^{(j)}(A^\uparrow)^2 \,\middle|\, \mathcal{T}_n^{(k)}\right] \\
&= \mathbb{E}\left[\frac{(c_{j+1}^n)^2 \nu_n^{(j)}(A^\uparrow)(1 - \nu_n^{(j)}(A^\uparrow))}{(C_{j+1}^n)^2} \,\middle|\, \mathcal{T}_n^{(k)}\right] \\
&\leq \mathbb{E}\left[\frac{(c_{j+1}^n)^2}{(C_j^n - j)^2} \,\middle|\, \mathcal{T}_n^{(k)}\right],
\end{aligned}$$

For the last expectation, condition on $\mathcal{T}_n^{(j)}$. Then $c_{j+1}^n > x$ when $S_n(C_j^n + 1), \dots, S_n(C_j^n + x)$ are unique. There are $C_j^n - j$ unique entries in $S_n(1), \dots, S_n(C_j^n)$. Thus,

$$\mathbb{P}(c_{j+1}^n > x \mid \mathcal{T}_n^{(j)}) = \prod_{i=1}^{x}\left(1 - \frac{C_j^n - j + i}{n}\right) \leq \left(1 - \frac{C_j^n - j}{n}\right)^x.$$

Hence c_{j+1}^n is dominated by a $Y \sim \text{Geom}\left(\frac{C_j^n - j}{n}\right)$ random variable. In particular,

$$\mathbb{E}\left[\left(c_{j+1}^n\right)^2 \mid \mathcal{T}_n^{(j)}\right] \leq \mathbb{E}[Y^2] \leq \frac{2n^2}{(C_j^n - j)^2}$$

By using the tower property of expectation, we compute,

$$\mathbb{E}\left[\frac{\left(c_{j+1}^n\right)^2}{(C_j^n - j)^2} \,\middle|\, \mathcal{T}_n^{(k)}\right] \leq \mathbb{E}\left[\frac{1}{(C_j^n - j)^2}\mathbb{E}\left[\left(c_{j+1}^n\right)^2 \,\middle|\, \mathcal{T}_n^{(j)}\right] \,\middle|\, \mathcal{T}_n^{(k)}\right] \leq 2n^2 \mathbb{E}\left[\frac{1}{(C_j^n - j)^2} \,\middle|\, \mathcal{T}_n^{(k)}\right].$$

Putting everything together, we obtain the desired statement. □

Lemma 5.2.7. We have,

$$\limsup_{n\to\infty} \mathbb{E}[X_{n,k}] = \limsup_{n\to\infty} \mathbb{E}\left[\sum_{j=k}^{N} \mathbb{E}\left[\frac{2n^2}{(C_j^n - j)^4} \,\middle|\, \mathcal{T}_n^{(k)}\right]\right] \xrightarrow[k\to\infty]{} 0$$

Remark 5.2.8. This is heuristically clear. Indeed $C_j^n \approx n^{\frac{1}{2}} j^{\frac{1}{2}}$, since $n^{-\frac{1}{2}} C_j^n \to C_j$, the j'th point of a PPP of intensity tdt so that $C_j \approx j^{\frac{1}{2}}$. Thus we expect $\frac{1}{(C_j^n - j)^4} \approx \frac{1}{(C_j^n)^4} \approx \frac{1}{n^2 j^2}$. In particular, $\limsup_{n\to\infty} 2n^2 \mathbb{E}[X_{n,k}] \approx \sum_{i=k}^{n/e} \frac{2}{j^2} \approx \frac{2}{k} \to 0$ as $k \to \infty$.

Proof of Lemma 5.2.7. Recall $C_j^n - j$ is the number of unique elements of $S_n(1), \ldots, S_n(C_j^n)$. Thus we have, $C_j^n - j \geq 1$ and thus $\frac{2n^2}{(C_j^n - j)^4} \leq 2n^2$. In particular for all $1 \leq m \leq n$ we have,

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=k}^{N} \mathbb{E}\left[\frac{2n^2}{(C_j^n - j)^4} \,\middle|\, \mathcal{T}_n^{(k)}\right]\right] &\leq \mathbb{E}\left[\sum_{j=k}^{m} \mathbb{E}\left[\frac{2n^2}{(C_j^n - j)^4} \,\middle|\, \mathcal{T}_n^{(k)}\right]\right] + \mathbb{E}\left[(N - m)2n^2 \mathbb{1}_{\{N>m\}}\right], \\
&\leq \sum_{j=k}^{m} \mathbb{E}\left[\frac{2n^2}{(C_j^n - j)^4}\right] + 2n^3 \mathbb{P}(N > m).
\end{aligned}$$

Note that N is typically of order $\frac{n}{e}$ (a well known fact coming from the birthday paradox type problems). Hence, we set $m = \lfloor \beta n \rfloor$ so that m is of the order n. Fix $\alpha > 1$. Then,

$$\begin{aligned}
\sum_{j=k}^{\beta n} \mathbb{E}\left[\frac{2n^2}{(C_j^n - j)^4}\right] &\leq \sum_{j=k}^{\beta n} \left(\mathbb{E}\left[\frac{2n^2}{(C_j^n - j)^4} \,\middle|\, C_j^n > \alpha j\right] + 2n^2 \mathbb{P}(C_j^n > \alpha j)\right), \\
&\leq \frac{2\alpha^4}{(\alpha - 1)^4} \sum_{j=k}^{\beta n} \left(\mathbb{E}\left[\frac{n^2}{(C_j^n)^4} \,\middle|\, C_j^n > \alpha j\right]\right) + 2n^2 \sum_{j=k}^{\beta n} \mathbb{P}(C_j^n > \alpha j), \\
&\leq C \sum_{j=k}^{\beta n} \mathbb{E}\left[\frac{n^2}{(C_j^n)^4}\right] + 2n^2 \sum_{j=k}^{\beta n} \mathbb{P}(C_j^n > \alpha j)
\end{aligned}$$

Hence we have shown $\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{E}[X_{k,n}] = 0$ if we find $\alpha > 1$ and $\beta > \frac{1}{e}$ so that,

$$i) \lim_{n\to\infty} 2n^3 \mathbb{P}(N > \beta n) = 0,\ ii) \lim_{n\to\infty} 2n^2 \sum_{j=k}^{\beta n} \mathbb{P}(C_j^n > \alpha j) = 0,\ iii) \lim_{n\to\infty} \sum_{j=k}^{\beta n} \mathbb{E}\left[\frac{n^2}{(C_j^n)^4}\right] \xrightarrow[k\to\infty]{} 0.$$

Statement $i)$ and $ii)$ follow directly from Lemma 4.2.14. (Statement $i)$ since we have exponential decay of the tail $\mathbb{P}(N > \beta)$ for $\beta > \frac{1}{2}$ and $ii)$ by choosing $k \geq 7$). To show $iii)$, recall if X is a positive discrete random variable taking values $0 < x_1 < x_2 < \dots$ then,

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} (x_k - x_{k-1}) \mathbb{P}(X > x_{k-1}).$$

Using this in our setting, together with Lemma 4.2.13, we obtain,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{(C_j^n)^4}\right] &= \sum_{k=1}^{\infty} \left(\frac{1}{k^4} - \frac{1}{(k+1)^4}\right) \mathbb{P}\left(\frac{1}{(C_j^n)^4} > \frac{1}{k^4}\right) \\
&\leq \sum_{k=1}^{\infty} \left(\frac{1}{k^4} - \frac{1}{(k+1)^4}\right) \mathbb{P}\left(C_j^n < k\right) \\
&\leq 4e \sum_{k=1}^{M} k^{-5} \left(\frac{k^2 e}{2nj}\right)^j + \sum_{k=M+1}^{\infty} \left(\frac{1}{k^4} - \frac{1}{(k+1)^4}\right), \\
&\leq 4e \left(\frac{e}{2nj}\right)^j \sum_{k=1}^{M} k^{-5+2j} + \frac{1}{M^4}, \\
&\leq 4e \left(\frac{e}{2nj}\right)^j M^{2j-4} + \frac{1}{M^4}
\end{aligned}$$

Setting $M = \left(\frac{2nj}{e}\right)^{\frac{1}{2}}$ shows that $\mathbb{E}\left[\frac{1}{(C_j^n)^4}\right] \leq C(nj)^{-2}$ for some constant C. In particular,

$$\lim_{n\to\infty} \sum_{j=k}^{\beta n} \mathbb{E}\left[\frac{n^2}{(C_j^n)^4}\right] \leq \lim_{n\to\infty} C \sum_{j=k}^{\infty} \frac{1}{j^2} \leq \frac{1}{k-1} \xrightarrow[k\to\infty]{} 0.$$

This shows $iii)$ and concludes the proof of Lemma 5.2.7. □

In particular, this shows both $iii)$ and $iv)$ of Proposition 2.4.3 and finishes the proof of Theorem 1.5.2.

6 Generalization of $\mathcal{T}_n$

We saw for $S_n \in_u [n]^{n-1}$, the corresponding random tree constructed via the Foata–Fuchs bijection $\mathcal{T}_n$, converges to the continuum random tree $\mathcal{T}$ constructed via stick-breaking. The tree $\mathcal{T}$ has two sources of randomness: the length of the sticks, determined by the PPP of intensity tdt and the attachment points of the sticks, which was uniform over the already existing tree. In this section, we construct a tree $\mathcal{T}_\beta \subset \ell^1$ with a different distribution for the length of the sticks parametrized by β, and we show this object is also a scaling limit of discrete random trees constructed by the Foata–Fuchs bijection.

The convergence of $n^{-\frac{1}{2}}\mathcal{T}_n$ to $\mathcal{T}$ was a consequence of the scaled repeat points $n^{-\frac{1}{2}}C_1^n$ in S_n converging to a PPP of intensity tdt, and B_i^n being roughly uniform over $\{1, \dots, C_i^n - 1\}$, together with tightness arguments. For i on the scale $n^{\frac{1}{2}}$, the probability that $S_n(i)$ is a repeat is roughly $\frac{i}{n}$. This is a good starting point for a new model: what happens to trees constructed from random sequences $S_n^f \in [n]^{n-1}$ where $\mathbb{P}(S_n^f(i) \text{ is a repeat}) = f\big(\frac{i}{n}\big)$, for some function $f : [0,1] \to [0,1]$? In this section, we compute the finite dimensional distributions for trees constructed from S_n^f when $f(x) = x^\beta + o(x^\beta)$ as $x \to 0$ for parameter $\beta > 0$ and show tightness for the specific choice $f(x) = x^\beta$. We formalize the definition of S_n^f first.

Definition 6.0.1. Let $f : [0,1] \to [0,1]$ and define the random variable $S_n^f \in [n]^{n-1}$ as: $S_n(1) \in_u [n]$ and for $i \in \{2, \dots, n-1\}$,

$$S_n^f(i) \sim \begin{cases} S_n(j) \text{ with } j \in_u \{1, \dots, i-1\}, & \text{with probability } f\left(\frac{i}{n}\right), \\ \text{Unif}\big([n] \setminus \{S_n(1), \dots, S_n(i-1)\}\big), & \text{with probability } 1 - f\left(\frac{i}{n}\right). \end{cases}$$

Let $\mathcal{T}_{n,f}$ denote the random tree obtained by applying the Foata–Fuchs bijection to S_n^f.

Remark 6.0.2. Conditional on $S_n^f(i)$ being a repeat, we let $S_n^f(i) \in_u \{\{S_n(1), \dots, S_n(i-1)\}\}$ (as multiset). This is different from the uniform case, where conditional on $S_n(i)$ being a repeat, $S_n \in_u \{S_n(1), \dots, S_n(i-1)\}$ (as set, i.e. we do not take multiplicity into account). This change is made to simplify the proof corresponding to Lemma 4.2.11. In this proof, we condition on $\{S_n(1), \dots, S_n(a)\}$ having few repeats for all $a \in [n-1]$, else $\mathbb{P}\big(p(v) \in \mathcal{T}_n(a)\big)$ could become too small. Proving $\mathbb{P}(A_n) \to 0$ as $n \to \infty$ was done in Lemma 4.2.14, however this proof cannot immediately be generalized to the current case, it fails for $\beta \leq \frac{e}{2} - 1$.

Instead of adapting the proof, we simply change the sampling. This does not change the finite dimensional distributions: as we will see, the first k repeats happen at a scale $n^{\frac{\beta}{\beta+1}}$. Thus, for any $a > 0$, the proportion of repeats in $S_n(1), \dots, S_n\big(an^{\frac{\beta}{\beta+1}}\big)$ goes to zero and sampling while taking multiplicities into account tends towards sampling uniformly on the distinct values. This gives intuition why both choices of sampling repeats yield the same finite dimensional distributions. Only when we prove compactness, and need properties of $S_n(i)$ with i on a scale larger than $n^{\frac{\beta}{\beta+1}}$, does the choice of model impact the proofs. Sampling repeats uniformly from the multiset $\{\{S_n(1), \dots, S_n(i-1)\}\}$ makes these proofs easier.

We continue with more definitions. We define,

$$C_i^n = \text{index of } i\text{'th repeat in } S_n^f \quad \text{and} \quad B_i^n = \min\{l \in [n] \text{ s.t. } S_n(l) = S_n(C_i^n)\},$$

to be the repeat and attachment indices. For notational clarity, we do not denote C_i^n and B_i^n to depend on choice of f as throughout this section, it is clear that C_i^n, B_i^n refer to S_n^f. We define $\mathcal{T}_{n,f}^{(k)}$ to be the tree obtained from $\{S_n^f(1), \dots, S_n^f(C_k^n - 1)\}$. For the measures, let ν_n be the uniform probability measure on $\mathcal{T}_{n,f}$ so that $\nu_n(A) = \frac{|A|}{n}$ for all $A \subset \mathcal{T}_{n,f}$. Similarly, $\nu_n^{(k)}$ denotes the uniform probability measure on $\mathcal{T}_{n,f}^{(k)}$.

In the continuous setting, let $0 < C_1 < \dots$ be the ordered points of a PPP of intensity $t^\beta dt$ and $B_i \sim \text{Unif}([0, C_i])$. Let $\mathcal{T}_\beta$ denote the tree obtained by applying the stick-breaking construction to the points $(C_i, B_i)_{i\in\mathbb{N}}$ and set $\mathcal{T}_\beta^{(k)} = \rho([0, C_k])$ and $\mu^{(k)} = \rho^*\overline{\lambda}$ where $\overline{\lambda}$ is the normalized Lebesgue measure on $[0, C_k]$. We will show $\mu^{(k)}$ converges to some probability measure μ on $\mathcal{T}_\beta$.

The first subsection is dedicated towards showing,

Theorem 6.0.3. For $\beta > 0$, let $f : [0, 1] \to [0, 1]$ be such that $f(x) = x^\beta + o(x^\beta)$ as $x \to 0$. For any $k \in \mathbb{N}$, we have convergence of finite dimensional distributions in the GHP-topology,

$$\left(\mathcal{T}_{n,f}^{(k)}, n^{-\frac{\beta}{\beta+1}} d_n, \nu_n^{(k)}\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}_\beta^{(k)}, d, \mu^{(k)}\right),$$

where d_n is the graph distance on $\mathcal{T}_{n,f}^{(k)}$ and d is the ℓ^1 metric.

Remark 6.0.4. The only condition specified on f is $f(x) = x^\beta + o(x^\beta)$ as $x \to 0$. It is not surprising this suffices to determine the finite dimensional distributions. Indeed, as will be shown, repeats C_i^n happen at the scale $n^{\frac{\beta}{\beta+1}}$. As $n^{\frac{\beta}{\beta+1}}/n \to 0$ as $n \to \infty$, we see that the repeats $C_1^n < \cdots < C_k^n$, and thus the trees $\mathcal{T}_{n,f}^{(k)}$ are determined by the behavior of f around 0.

However, it cannot be the case that convergence of the full tree,

$$\left(\mathcal{T}_{n,f}, n^{-\frac{\beta}{\beta+1}} d_n, \nu_n\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}_\beta, d, \mu\right),$$

holds for any f with $f(x) = x^\beta + o(x^\beta)$. As an example, take $f(x) = x\mathbb{1}_{x\le\frac{1}{2}}$. Almost surely, the set $\left\{S_n^f(\frac{n}{2}), \dots, S_n^f(n-1)\right\}$ does not contain any repeats. Assuming Theorem 6.0.3, we must scale $\mathcal{T}_{n,f}$ by a factor $n^{-\frac{1}{2}}$ which means that the $\frac{n}{2}$ vertices $\left\{S_n^f(\frac{n}{2}), \dots, S_n^f(n-1)\right\}$ are all on one stick approaching infinite length and $n^{-\frac{1}{2}}\mathcal{T}_{n,f}$ cannot converge to an almost sure compact metric space.

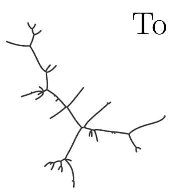
In subsection 6.1, we prove Theorem 6.0.3. Then in subsection 6.2 and 6.3, we show that for the choice $f = \left(\frac{i}{n}\right)^\beta$, we have,

$$\left(\mathcal{T}_{n,f}, n^{-\frac{\beta}{\beta+1}} d_n, \nu_n\right) \xrightarrow[n\to\infty]{d} \left(\mathcal{T}_\beta, d, \mu\right),$$

In the GHP-topology.

6.1 The Finite Dimensional Distribution

In this section, we prove Theorem 6.0.3. We largely follow the same reasoning as Section 3. To that end, we aim to show,



Theorem 6.1.1. With notation as in Definition 6.0.1, we have for all $k \in \mathbb{N}$,

$$n^{-\frac{\beta}{\beta+1}}\left(C_1^n, \ldots, C_k^n, B_1^n, \ldots, B_k^n\right) \xrightarrow[n\to\infty]{d} \left(C_1, \ldots, C_k, B_1, \ldots, B_k\right),$$

where $C_1, \ldots, C_k$ are the first k points in a PPP of intensity $t^\beta dt$ and $B_i \sim \mathrm{Unif}([0, C_i])$.

To show this, we first prove the following lemma,

Lemma 6.1.2. With notation as above, we have for all $0 < s_1 < \cdots < s_k$,

$$n^{\frac{k\beta}{\beta+1}}\mathbb{P}\left(C_1^n = \lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor, \ldots, C_k^n = \lfloor s_k n^{\frac{\beta}{\beta+1}} \rfloor\right) \xrightarrow[n\to\infty]{\text{u.c.}} f_{C_1,\ldots,C_k}(s_1, \ldots, s_k),$$

where the convergence is uniform over compact sets and $f_{C_1,\ldots,C_k}(s_1, \ldots, s_k)$ is the pdf of the first k ordered points of a PPP of intensity $t^\beta dt$.

Proof. Recall from Lemma 2.2.22 that $f_{C_1,\ldots,C_k}(s_1, \ldots, s_k) = s_1^\beta \ldots s_k^\beta e^{-\frac{s_k}{\beta+1}}$. We reason by induction on k.

For the base case, observe $C_1^n = \lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor$ exactly when index 1 to $\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor - 1$ of S_n are not repeats, while index $\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor$ is a repeat. Since $S_n^f(i)$ is a repeat with probability $f\left(\frac{i}{n}\right)$, we obtain,

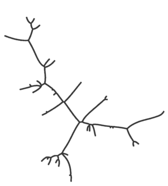
$$n^{\frac{\beta}{\beta+1}}\mathbb{P}\left(C_1^n = \lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor\right) = n^{\frac{\beta}{\beta+1}} f\left(\frac{\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor}{n}\right) \prod_{i=2}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor} \left(1 - f\left(\frac{i}{n}\right)\right).$$

Recall f satisfies $f(x) = x^\beta + o(x^\beta)$ as $x \to 0$. Since $\frac{\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor}{n} \to 0$, we obtain,

$$\begin{aligned} n^{\frac{\beta}{\beta+1}} f\left(\frac{\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor}{n}\right) &= n^{\frac{\beta}{\beta+1}} \left(\frac{\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor}{n}\right)^\beta + o\left(n^{\frac{\beta}{\beta+1}}\left(\frac{\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor}{n}\right)^\beta\right), \\ &= s_1^\beta + o(1) + o\left(n^{-\frac{\beta}{\beta+1}} \lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor\right), \\ &= s_1^\beta + o(1), \end{aligned}$$

where the convergence is uniform for s_1 in compact sets since $\frac{\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor}{n} \to 0$ is uniform over compact sets. We continue with the product term. From the Taylor expansion of $\ln(1-x)$, we see $\ln(1-x) = x + O(x^2)$ for x around 0. We substitute $x = f\left(\frac{i}{n}\right)$, to get,

$$\begin{aligned} \prod_{i=2}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor} \left(1 - f\left(\frac{i}{n}\right)\right) &= \exp\left[\sum_{i=2}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor} \log\left(1 - f\left(\frac{i}{n}\right)\right)\right], \\ &= \exp\left[- \sum_{i=2}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor} f\left(\frac{i}{n}\right) - \sum_{i=1}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}} \rfloor} O\left(f^2\left(\frac{i}{n}\right)\right)\right] \end{aligned}$$



We focus on the two sums above separately. For the first sum,

$$\begin{aligned}
\sum_{i=2}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor} f\left(\frac{i}{n}\right) &= -f\left(\frac{1}{n}\right) + \sum_{i=1}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor} \left[\left(\frac{i}{n}\right)^{\beta} + o\left(\left(\frac{i}{n}\right)^{\beta}\right)\right], \\
&= o(1) + n^{-\beta} \sum_{i=1}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor} i^{\beta} + \sum_{i=1}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor} o\left(\left(\frac{i}{n}\right)^{\beta}\right), \\
&= o(1) + n^{-\beta}\left(\frac{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor^{\beta+1}}{\beta+1} + O\left(\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor^{\beta}\right)\right) + o\left(\sum_{i=1}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor} \left(\frac{i}{n}\right)^{\beta}\right), \\
&= o(1) + \frac{s_1}{\beta+1} + O\left(n^{-\frac{1}{\beta}}\right) + o\left(\frac{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor^{\beta+1}}{n^{\beta}}\right), \\
&= o(1) + \frac{s_1}{\beta+1}.
\end{aligned}$$

Where the above convergence holds uniformly for s_1 in compact set. For the second sum,

$$\begin{aligned}
\sum_{i=1}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor} O\left(f^2\left(\frac{i}{n}\right)\right) &= \sum_{i=1}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor} O\left(\left(\frac{i}{n}\right)^{2\beta}\right), \\
&= \lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor O\left(\left(\frac{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor}{n}\right)^{2\beta}\right), \\
&= O\left(n^{-\frac{\beta}{\beta+1}}\right), \\
&= o(1).
\end{aligned}$$

Again, convergence being uniform for s_1 in compact sets. By substituting all intermediate results, we obtain,

$$\begin{aligned}
n^{\frac{\beta}{\beta+1}}\mathbb{P}\left(C_1^n = \lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor\right) &= n^{\frac{\beta}{\beta+1}} f\left(\frac{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor}{n}\right) \prod_{i=2}^{\lfloor s_1 n^{\frac{\beta}{\beta+1}}\rfloor} \left(1 - f\left(\frac{i}{n}\right)\right), \\
&= \left(s_1^{\beta} + o(1)\right) \exp\left[-\frac{s_1}{\beta+1} + o(1)\right],
\end{aligned}$$

which shows the base case of the induction proof. We continue with the induction step. To ease notation, write,

$$C_k^n(x) = \left\{C_1^n = \lfloor x_1 n^{\frac{\beta}{\beta+1}}\rfloor, \ldots, C_k^n = \lfloor x_k n^{\frac{\beta}{\beta+1}}\rfloor\right\}, \qquad \text{where } x = (x_1, \ldots, x_l) \text{ for } l \geq k,$$

so the induction hypothesis reads $n^{\frac{k\beta}{\beta+1}}\mathbb{P}(C_k^n(s)) \xrightarrow[n\to\infty]{\text{u.c.}} s_1^{\beta}\ldots s_k^{\beta} e^{-\frac{s_k}{\beta+1}}$. Since,

$$n^{\frac{(k+1)\beta}{\beta+1}}\mathbb{P}\left(C_{k+1} = \lfloor s_{k+1} n^{\frac{\beta}{\beta+1}}\rfloor, C_k^n(s)\right) = n^{\frac{\beta}{\beta+1}}\mathbb{P}\left(C_{k+1}^n = \lfloor s_{k+1} n^{\frac{\beta}{\beta+1}}\rfloor \;\middle|\; C_k^n(s)\right) n^{\frac{k\beta}{\beta+1}}\mathbb{P}\left(C_k^n(s)\right),$$

it suffices to show,

$$n^{\frac{\beta}{\beta+1}}\mathbb{P}\left(C_{k+1}^n=\lfloor s_{k+1}n^{\frac{\beta}{\beta+1}}\rfloor \mid C_k^n(s)\right)\xrightarrow[n\to\infty]{\text{u.c.}} s_{k+1}^{\beta}\exp\left(-\frac{s_{k+1}^2-s_k^2}{\beta+1}\right).$$

The event $\left\{C_{k+1}^n=\lfloor s_{k+1}n^{\frac{\beta}{\beta+1}}\rfloor\right\}$ given $C_k^n(s)$ occurs precisely when S_n^f contains no repeats between indices $\lfloor s_k n^{\frac{\beta}{\beta+1}}\rfloor+1$ and $\lfloor s_{k+1}n^{\frac{\beta}{\beta+1}}\rfloor-1$, while index $\lfloor s_{k+1}n^{\frac{\beta}{\beta+1}}\rfloor$ is a repeat. We write $I_n=\{\lfloor s_k\rfloor n^{\frac{\beta}{\beta+1}}+1,\dots,\lfloor s_{k+1}n^{\frac{\beta}{\beta+1}}\rfloor-1\}$. Then,

$$\begin{aligned}
n^{\frac{\beta}{\beta+1}}\mathbb{P}\left(C_{k+1}^n=\lfloor s_{k+1}n^{\frac{\beta}{\beta+1}}\rfloor \mid C_k^n(s)\right)&=n^{\frac{\beta}{\beta+1}}f\left(\frac{\lfloor s_{k+1}n^{\frac{\beta}{\beta+1}}\rfloor}{n}\right)\prod_{i=I_n}\left(1-f\left(\frac{i}{n}\right)\right),\\
&=(s_{k+1}+o(1))\exp\left[-\sum_{i\in I_n}f\left(\frac{i}{n}\right)+O\left(f^2\left(\frac{i}{n}\right)\right)\right],\\
&=(s_{k+1}+o(1))\exp\left[-\frac{s_{k+1}^{\beta+1}-s_k^{\beta+1}}{\beta+1}+o(1)\right],
\end{aligned}$$

where we omit computations as they follow the same reasoning as the base case. This concludes the proof. □

We continue with the proof of Theorem 6.1.1.

Proof of Theorem 6.1.1. It suffices to show,

$$\begin{aligned}
&i)\ \mathbb{P}\big(C_1^n\le s_1n^{\frac{\beta}{\beta+1}},\dots,C_k^n\le s_kn^{\frac{\beta}{\beta+1}}\big)\xrightarrow[n\to\infty]{}\mathbb{P}\big(C_1\le s_1,\dots,C_k\le s_k\big),\\
&ii)\ \mathbb{P}(B_1^n\le t_1C_1^n,\dots,B_k^n\le t_kC_k^n\mid C_1^n\le s_1n^{\frac12},\dots,C_k^n\le s_kn^{\frac12}),\\
&\xrightarrow[n\to\infty]{}\mathbb{P}\big(B_1\le t_1C_1,\dots,B_k\le t_kC_k\mid C_1\le s_1,\dots,C_k\le s_k\big)
\end{aligned}$$

Statement $i)$ follows from identical reasoning as the proof of Lemma 3.1.5. Statement $ii)$ is shown by copying the proof of Lemma 3.1.7 and 3.1.5 but replacing $n^{-\frac12}$ with $n^{-\frac{\beta}{\beta+1}}$. □

Hence we see that the rescaled first k repeat and attachment points C_i^n and B_i^n converge to C_i and B_i where $0<C_1<\cdots<C_k$ are the first k ordered points of a PPP with intensity $t^\beta dt$. The result of Theorem 6.0.3 is immediate.

Proof of Theorem 6.0.3. The result immediately follows from the convergence,

$$n^{-\frac{\beta}{\beta+1}}\big(C_1^n,\dots,C_k^n,B_1^n,\dots,B_k^n\big)\xrightarrow[n\to\infty]{d}\big(C_1,\dots,C_k,B_1,\dots,B_k\big),$$

together with the work in Section 3.2 and Section 3.3 as seen in Remark 3.3.4. □

Thus we have shown that for any $f:[0,1]\to[0,1]$ with $f(x)=x^\beta+o(x^\beta)$ as $x\to 0$, we have convergence of finite dimensional distributions,

$$\left(\mathcal{T}_{n,f}^{(k)},n^{-\frac{\beta}{\beta+1}}d_n,\nu_n^{(k)}\right)\xrightarrow[n\to\infty]{d}\big(\mathcal{T}_\beta^{(k)},d,\mu^{(k)}\big),$$

in the GHP-topology. As seen in Remark 6.0.4, setting $f(x) = x^\beta + o(x^\beta)$ as $x \to 0$ is not sufficient for tightness of $\left(n^{-\frac{\beta}{\beta+1}}\mathcal{T}_{n,f}\right)_{n\in\mathbb{N}}$. We will show that for the choice $f(x) = x^\beta$, we do have tightness of $\left(n^{-\frac{\beta}{\beta+1}}\mathcal{T}_{n,f}\right)_{n\in\mathbb{N}}$. In particular,

Theorem 6.1.3. Let $f(x) = x^\beta$. Then,

$$\left(\mathcal{T}_{n,f}, n^{-\frac{\beta}{\beta+1}}d_n, \nu_n\right) \xrightarrow[n\to\infty]{d} (\mathcal{T}_\beta, d, \mu),$$

convergence being in the GHP-topology.

Remark 6.1.4. Throughout the following two subsections, we write $\mathcal{T}_n \equiv \mathcal{T}_{n,f}$ for $f(x) = x^\beta$ and $\mathcal{T} \equiv \mathcal{T}_\beta$. By analogous reasoning to the proof of Proposition 2.4.3, we have shown Theorem 6.1.3 upon proving,

$$i) \lim_{t\to\infty} \mathbb{P}\left(d_H\left(\mathcal{T}(t), \mathcal{T}\right) > \epsilon\right) = 0, \qquad ii) \lim_{t\to\infty}\limsup_{n\to\infty} \mathbb{P}\left(d_H\left(\mathcal{T}_n(tn^{\frac{\beta}{\beta+1}}), \mathcal{T}_n\right) > \epsilon n^{\frac{\beta}{\beta+1}}\right) = 0,$$
$$iii) \lim_{k\to\infty} \mathbb{P}\left(d_P\left(\mu^{(k)}, \mu\right) > \epsilon\right)\right) = 0, \qquad iv) \lim_{k\to\infty}\limsup_{n\to\infty} \mathbb{P}\left(d_P\left(\nu_n^{(k)}, \nu_n\right) > \epsilon\right) = 0$$

We show $i)$ and $ii)$ in the next subsection and $iii)$ and $iv)$ in Subsection 6.3.

6.2 Tightness of $\left(\mathcal{T}_{n,f}, n^{-\frac{1}{2}}d_n\right)_{n\in\mathbb{N}}$

6.2.1 compactness of $\mathcal{T}_\beta$

In this section, we show $i)$, that is,

Theorem 6.2.1. We have,

$$\lim_{t\to\infty} \mathbb{P}\left(d_H\left(\mathcal{T}(t), \mathcal{T}\right) > \epsilon\right) = 0.$$

The proof will be mostly identical to Section 4.1. We focus on highlighting the differences rather than proving this result from scratch. By following identical reasoning to that of the proof of Lemma 4.1.2 and Corollary 4.1.3, we immediately obtain,

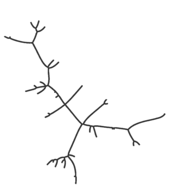
Lemma 6.2.2. Theorem 6.2.1 is proven upon finding $\epsilon_i(t) : [0,\infty) \to [0,\infty)$ for which,

$$i) \lim_{t\to\infty}\sum_{i=0}^{\infty} \epsilon_i(t) = 0 \quad \text{and} \quad ii) \lim_{t\to\infty}\sum_{i=0}^{\infty} \mathbb{P}\Big(d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1}t)\big) > \epsilon_i\Big) = 0.$$

We aim to bound $\mathbb{P}\Big(d_H\big(\mathcal{T}(a), \mathcal{T}(2a)\big) > c\Big)$ for which we first bound $\mathbb{P}\Big(d_H\big(\mathcal{T}(a), \rho(s)\big) > c\Big)$, for some $s \in [a, 2a]$.

Lemma 6.2.3. Fix $a, c > 0$ and $s \in [a, 2a]$. Then,

$$\mathbb{P}(d_H(\mathcal{T}(a), \mathcal{T}(2a)) > c) \le \exp\left(-\frac{ca^\beta}{2}\right).$$



Proof. Similar to the proof of Lemma 4.1.5, we sample the random variables (C_i, B_i) jointly from one PPP on a subset of $\mathbb{R}^2$. In the original Lemma, we took a PPP of intensity 1 on the wedge $\{(x, y) \in \mathbb{R}^2 : 0 \leq x, 0 \leq y \leq x\}$ as the point process can be written as $\sum_{i=1}^{\infty} \delta_{(C_i,B_i)}$ with $0 < C_1 < C_2 < \dots$ ordered points of a PPP of intensity tdt and $B_i \sim \text{Unif}([0, C_i])$.

In this proof, we set η to be a homogeneous PPP of intensity 1 in,

$$T_\beta = \{(x, y) \in \mathbb{R}^2 : 0 \leq x, 0 \leq y \leq x^\beta\}.$$

Write $\eta = \sum_{i=1}^{\infty} \delta_{(C_i,B_i')}$ with $0 < C_1 < \dots$ It follows from Example 2.2.20 that $0 < C_1 < \dots$ are the ordered points of a PPP of intensity $t^\beta dt$. However, note that $B_i' \sim \text{Unif}\big([0, (C_i)^\beta]\big)$ has the wrong distribution to represent the attachment points. Thus we set $B_i = (C_i)^{1-\beta} B_i'$ so that $B_i \sim \text{Unif}([0, C_i])$.

Observe that branch i is attached to $\mathcal{T}(2a)$ at point $\rho(B_i) = \rho\left(B_i'(C_i)^{1-\beta}\right)$. In Figure 28, we highlight in red the region where points of η in that region imply the corresponding stick is attached to $\mathcal{T}(a)$. Note that the upper boundary of this shape is concave and decreasing for $\beta < 1$ and concave increasing for $\beta > 1$.

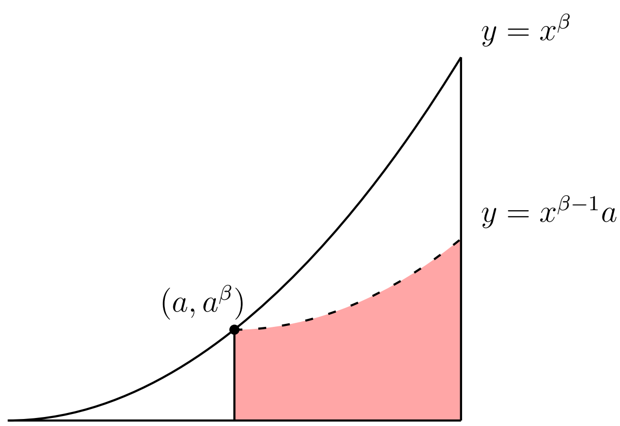


Figure 28: Attachment to $\mathcal{T}(a)$ for $\beta > 1$.

As in Remark 4.1.6, we may find the subset $S \subset [a, 2a]$ such that $\rho(S)$ corresponds to the path from $\rho(s)$ to $\mathcal{T}(a)$. We define,

Definition 6.2.4. Let η be a PPP of intensity 1 in T_β, set $\eta' = \eta + \delta_{(a,a^\beta)}$. For $s > a$, define,

$$p_1 \equiv \left(p_x^1(s), p_y^1(s)\right) = \underset{(x,y)\in\eta' :\, x \leq s}{\arg\max}\ x, \quad d_1 \equiv d_1(s) = s - p_x^1(s) \ \text{ and } \ s_2 \equiv s_2(s) = \frac{p_y^1(s)}{\left(p_x^1(s)\right)^{\beta-1}}.$$

where p_1 is well defined as η contains no accumulation points almost surely and $(a, a^\beta) \in \eta'$. Whenever $s_{i-1} > a$, we set,

$$p_i \equiv p_i(s) = p(p_{i-1}(s)), \quad d_i(s) \equiv d_i = s_i - p_x^i(s) \ \text{ and } \ s_{i+1} = \frac{p_y^i(s)}{(p_x^i(s))^{\beta-1}}.$$

Set $N = \min\{i \in \mathbb{N} : s_{i+1} \leq a\}$ for the number of sticks on the path from $\rho(s)$ to $\mathcal{T}(a)$ and define $S = \bigcup_{i=1}^{N} [p_x^i(s), s_i]$. As in Lemma 4.1.8, we also have that the path from $\rho(s)$ to $\mathcal{T}(a)$ is given by $\rho(S)$. We adjust Remark 4.1.9 to the new setting.

i) $d_H(\mathcal{T}(a), \rho(s)) > c$ given η happens exactly when $\lambda(S) = d_1 + \cdots + d_N > c$ where λ is the Lebesgue measure on $\mathbb{R}$.

ii) By removing points from η in $\{(x, y) : a < x \leq 2a, 0 \leq y \leq ca^{\beta-1}\}$, we cannot decrease $d_H(\mathcal{T}(a), \rho(s))$.

Set $\eta^* = \{(x, y) \in \eta : x \leq a \text{ or } y > x^{\beta-1}a\}$ and let S^* be defined using Definition 6.2.4 but with η^* instead of η and set $S_j^* = \{x \in S : x \geq j\}$. Similar to Proof 4.1, we see that $d_H\big(\mathcal{T}(a), \rho(s)\big) > c$ implies that $\lambda(S^*) > c$. Furthermore, if η' contains any point in the region $S_j^* \times [0, \min_{x\in[a,2a]} x^{\beta-1}a]$, then $d_H\big(\mathcal{T}(a), \rho(s)\big) < c$. Note that $\min_{x\in[a,2a]} x^{\beta-1}a \leq \frac{1}{2}a^\beta$, and thus we find

$$\mathbb{P}\big(d_H\big(\mathcal{T}(a), \rho(s)\big) > c\big) \leq \mathbb{1}_{\{\lambda(S^*)>j\}}\mathbb{P}\big(\eta\big(S_j^* \times [0, \min_{x\in[a,2a]} x^{\beta-1}a]\big) = 0\big) \leq e^{-\frac{c\alpha^\beta}{2}},$$

where we may condition on $\lambda(S^*) > c$ since otherwise $d_H(\mathcal{T}(a), \rho(s)) < c$. This concludes the desired result. We give an illustration in Figure 29 where $\beta < 1$ and $\lambda(S^*) > c$. □

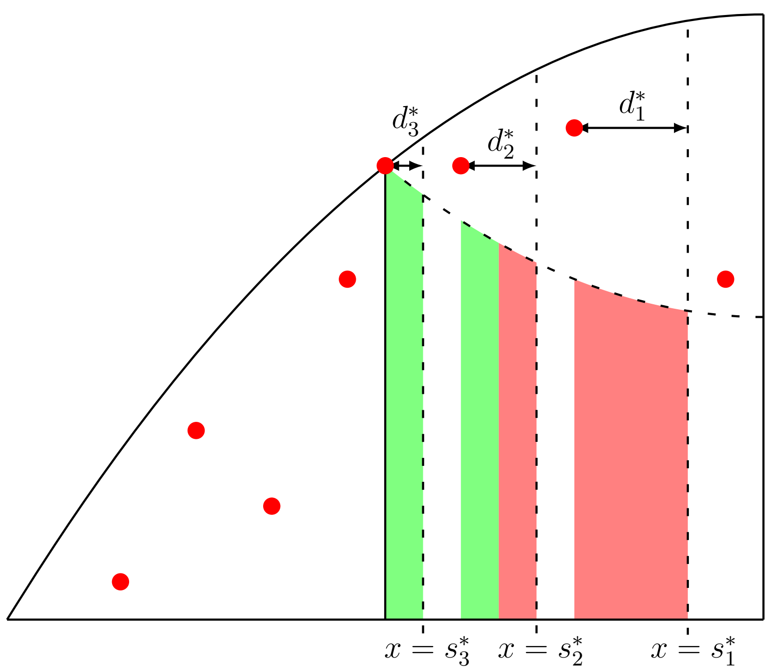


Figure 29: Visualization of area where η should be void of points. In red, we denote the area $\{(x, y) : x \in S_j^* \text{ and } 0 \leq y \leq x^{\beta-1}a\}$ and $\{(x, y) : x \in S^* \setminus S_j^* \text{ and } 0 \leq y \leq x^{\beta-1}a\}$ is denoted in green. In particular, if η were to contain any point in the red area, the distance $d_H\big(\mathcal{T}(a), \rho(s)\big)$ would be less than c. The height of this area is at least $\min_{x\in[a,2a]} x^{\beta-1}a \leq \frac{a^\beta}{2}$ and the with of the sum of red rectangles is $\lambda(S_j^*) = c$.

We upgrade the bound on $\mathbb{P}\big(d_H\big(\mathcal{T}(a), \rho(s)\big)\big) > c$ to a bound on $\mathbb{P}\big(d_H\big(\mathcal{T}(a), \mathcal{T}(2a)\big) > c\big)$ using the same trick as in Lemma 4.1.13.

Lemma 6.2.5. Let $a, c > 0$. We have,

$$\mathbb{P}\big(d_H\big(\mathcal{T}(a), \mathcal{T}(2a)\big) > c\big) \leq \frac{2a}{c} \exp\left(-\frac{ca^\beta}{4}\right)$$

Proof. Analogous to that of 4.1.13 □

Lastly, we show that $i)$ and $ii)$ in Lemma 6.2.2 hold.

Proof of Theorem 6.2.1. By substituting $a = 2^i t$ and $c = \epsilon_i$ in Lemma 6.2.5, we obtain,

$$\sum_{i=0}^{\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1} t)\big) > \epsilon_i\big) \leq \sum_{i=0}^{\infty} \frac{2^{i+1} t}{\epsilon_i} \exp\left(-\epsilon_i 2^{i\beta-2} t^\beta\right).$$

We set $\epsilon_i = (i+1+i\beta)2^{2-i\beta} t^{-\frac{\beta}{2}}$ and simplify under the assumption that $t^{\frac{\beta}{2}} > \ln(2)$.

$$\begin{aligned}
\sum_{i=0}^{\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1} t)\big) > \epsilon_i\big) &\leq \sum_{i=0}^{\infty} \frac{2^{(i+1+i\beta)+i\beta} t^{1+\frac{\beta}{2}}}{(i+1+i\beta)} \exp\left(-(i+1+i\beta) t^{\beta-\frac{\beta}{2}}\right), \\
&\leq t^{1+\frac{\beta}{2}} \sum_{i=0}^{\infty} \exp\left((i+1+i\beta)\Big(\ln(2) - t^{\frac{\beta}{2}}\Big)\right), \\
&\leq t^{1+\frac{\beta}{2}} \sum_{i=0}^{\infty} \exp\left(\ln(2) - t^{\frac{\beta}{2}}\right)^{i+1}, \\
&\leq t^{1+\frac{\beta}{2}} \frac{2\exp(-t^{\frac{\beta}{2}})}{1 - 2\exp(-t^{\frac{\beta}{2}})} \xrightarrow[t\to\infty]{} 0.
\end{aligned}$$

This verifies $ii)$ of Lemma 6.2.2. To verify $i)$, observe,

$$\sum_{i=0}^{\infty} \epsilon_i = t^{-\frac{\beta}{2}} \sum_{i=0}^{\infty} (i+1+i\beta)2^{2-i\beta} \leq t^{-\frac{\beta}{2}} \frac{4(1+\beta)}{(1-2^{-\beta})^2} \xrightarrow[t\to\infty]{} 0.$$

This verifies condition $i)$ and $ii)$ and in turn proves Theorem 6.2.1 finishing this section. □

6.2.2 Compactness of $n^{-\frac{1}{2}}\mathcal{T}_{n,f}$

In this section, we aim to show $ii)$ of Remark 6.1.4. That is,

Theorem 6.2.6. for all $\epsilon > 0$ we have,

$$ii) \lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(d_H\left(\mathcal{T}_n(tn^{\frac{1}{2}}), \mathcal{T}_n\right) > \epsilon n^{\frac{\beta}{\beta+1}}\right) = 0.$$

The proof of Theorem 6.2.6 closely mimics the reasoning in Subsection 4.2. However, the proof of Lemma 4.2.11 heavily relied on the uniform sampling of sequence S_n. As S_n^f is not uniform in $[n]^{n-1}$, we have to use a new approach for proving the analogous lemma in the current setting.

To ease notation, we write $S_n \equiv S_n^f$ throughout this section.

Lemma 6.2.7. Theorem 6.2.6 is shown upon finding $\epsilon_i(t) : [0,\infty) \to [0,\infty)$ such that,

$$i)\ \lim_{t\to\infty} \sum_{i=0}^{\infty} \epsilon_i(t) = 0 \text{ and } ii)\ \lim_{t\to\infty} \limsup_{n\to\infty} \sum_{i=0}^{\infty} \mathbb{P}\big(d_H\big(\mathcal{T}_n(2^i t n^{\frac{\beta}{\beta+1}}), \mathcal{T}_n(2^{i+1} t n^{\frac{\beta}{\beta+1}})\big) > \epsilon_i n^{\frac{\beta}{\beta+1}}\big) = 0.$$

Proof. Analogous to the reasoning in Lemma 4.2.4 and Corollary 4.2.5. □

Next, we bound the distance from a vertex $v_s \in \mathcal{T}_n(2a)\backslash\mathcal{T}_n(a)$ to tree $\mathcal{T}_n(a)$.

Lemma 6.2.8. Fix $a \in \mathbb{N}$ and $s \in \{a+1, \dots, 2a\}$. Let $v_s \in \mathcal{T}_n(2a)$ be the vertex corresponding to $S_n(s)$. Then,

$$\mathbb{P}\big(d_H\big(\mathcal{T}_n(a), v_s\big) > c\big) \le \exp\left(-\frac{ca^\beta}{2n^\beta}\right).$$

Proof. Recall that $p^k(v)$ denotes the label of the k'th parent of vertex $v \in \mathcal{T}_n$ and,

$$d_H\big(\mathcal{T}_n(a), v_s\big) \le \min_{k\in\mathbb{N}} \big\{p^k(v_s) \in \{S_n(1), \dots, S_n(a)\}\big\}.$$

Instead of sampling their vertices $p^k(v_s) \in \mathcal{T}(2a)$, we iteratively sample their indices in S_n. We define,

$$Q(s) = \begin{cases} s-1 & \text{if } S_n(s) \text{ is not a repeat,} \\ \min_{k\in\mathbb{N}}\{S_n(k) = S_n(J)\} \text{ with } J \in_u \{1, \dots, s-1\} & \text{if } S_n(s) \text{ is a repeat,} \end{cases}$$

and inductively set $Q^k(s) = Q(Q^{k-1}(s))$. Note that $p^k(v_s) \stackrel{d}{=} S_n(Q^k(s))$, since $Q^k(s)$ is the stochastic process of tracking the index of $p^k(v_s)$. In particular, we have,

$$d_H\big(\mathcal{T}(a), v_s\big) \le \min_{k\in\mathbb{N}} \big\{Q^k(s) \le a\big\}.$$

Fix $k \in \mathbb{N}$, we aim to lower bound $\mathbb{P}\big(Q^k(s) \le a \mid s, Q(s), \dots, Q^{k-1}(s)\big)$. Assume $Q^{k-1}(s) > a$ as otherwise $Q^k(s) \le a$ is deterministically true. We make two observation about the distribution of $Q^k(s)$ conditional on if $S_n(Q^{k-1}(s))$ is a repeat or not.

i) If $S_n(Q^{k-1}(s))$ is not a repeat, then $Q^k(s) = Q^{k-1}(s) - 1 > a$ (unless $Q^{k-1}(s) = a+1$).

ii) If $S_n(Q^{k-1}(s))$ is a repeat, then deterministically $Q^k(s) \le J$ where $J \in_u \{1, \dots, Q^{k-1}(s)\}$.

Both the probability of $S_n(Q^{k-1}(s))$ being a repeat conditional on $Q^{k-1}(s) > a$ and the distribution of J (which bounds $q^k(s)$ from below) only depend on. Specifically,

$$i)\ \mathbb{P}\big(S_n(Q^{k-1}(s)) \text{ is a repeat} \mid s, Q(s), \dots, Q^{k-1}(s)\big) = \left(\frac{Q^{k-1}(s)}{n}\right)^\beta \ge \left(\frac{a}{n}\right)^\beta,$$

$$ii)\ \mathbb{P}\big(Q^k(s) \le a \mid s, Q(s), \dots, Q^{k-1}(s) \text{ and } S_n(Q^{k-1}(s)) \text{ is a repeat}\big) \ge \frac{a}{Q^{k-1}(s) - 1} \ge \frac{1}{2},$$

where the last inequality follows since $a < Q^{k-1}(S) \leq 2a$. We combine both bounds to obtain,

$$\mathbb{P}\big(Q^k(s) \leq a \bigm| s, Q(s), \dots, Q^{k-1}(s)\big) \geq \frac{a^\beta}{2n^\beta}.$$

Putting all results together, we find,

$$\begin{aligned}\mathbb{P}\big(d_H\big(\mathcal{T}_n(a), v_s\big) > c\big) &\leq \mathbb{P}\Big(\min_{k\in\mathbb{Z}}\{Q^k(s) \leq a\} > c\Big),\\ &\leq \left(1 - \left(\frac{a^\beta}{2n^\beta}\right)\right)^c,\\ &\leq \exp\left(-\frac{ca^\beta}{2n^\beta}\right),\end{aligned}$$

which is the desired result. □

We upgrade this result to a bound on $\mathbb{P}\big(d_H\big(\mathcal{T}_n(a), \mathcal{T}_n(2a)\big) > c\big)$ using the exact same proof as Lemma 4.2.12.

Lemma 6.2.9. Let $a, c \in \mathbb{N}$. Then we have,

$$\mathbb{P}\big(d_H\big(\mathcal{T}_n(a), \mathcal{T}_n(2a)\big) > c\big) \leq \frac{2a}{c}\exp\left(-\frac{ca^\beta}{4n^\beta}\right).$$

Now we are in a position to prove Theorem 6.2.6.

Proof of Theorem 6.2.6. We substitute $a = 2^i t n^{\frac{\beta}{\beta+1}}$ and $c = \epsilon_i n^{\frac{\beta}{\beta+1}}$ into the result of Lemma 6.2.5 and simplify to obtain,

$$\sum_{i=0}^{\infty} \mathbb{P}\big(d_H\big(\mathcal{T}_n(2^i t n^{\frac{\beta}{\beta+1}}), \mathcal{T}_n(2^{i+1} t n^{\frac{\beta}{\beta+1}})\big) > \epsilon_i n^{\frac{\beta}{\beta+1}}\big) \leq \sum_{i=0}^{\infty} \frac{2^{i+1}t}{\epsilon_i}\exp\left(-\epsilon_i 2^{i\beta-2} t^\beta\right).$$

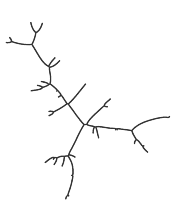
Thus the criteria in Lemma 6.2.7 are translated into finding $\epsilon_i(t) : [0,\infty) \to [0,\infty)$ for which,

$$\lim_{t\to\infty}\sum_{i=0}^{\infty}\epsilon_i = 0 \quad \text{and} \quad \lim_{t\to\infty}\sum_{i=0}^{\infty}\frac{2^{i+1}t}{\epsilon_i}\exp\left(-\epsilon_i 2^{i\beta-2}t^\beta\right) = 0.$$

These exactly coincide with the criteria in Lemma 6.2.2, which were shown to hold in Proof 6.2.1 by choosing $\epsilon_i = (i + 1 + i\beta)2^{2-i\beta}t^{-\frac{\beta}{2}}$. □

6.3 Tightness of the Measures

In this section, we show *iii)* and *iv)* of Remark 6.1.4. We start with *iii)* which follows identical reasoning to Section 5.1.



6.3.1 Convergence Measures on the Continuous Tree

In this section, we aim to show,

Theorem 6.3.1. For all $\epsilon > 0$, we have,

$$\lim_{k\to\infty} \mathbb{P}\left(d_P\left(\mu^{(k)}, \mu\right) > \epsilon\right)\right) = 0.$$

Recall that $(\mathcal{P}(\ell^1), d_P)$ is Polish. Hence it is enough to show that $\mu^{(k)}$ is Cauchy in probability. Recall the definition of $A^{\uparrow}$.

Definition 6.3.2. For $A \subset \mathcal{T}^{(k)}$, let $A^{\uparrow} = \pi_k^{-1}(A) \cap \mathcal{T}$, with $\pi_k : \ell^1 \to \mathbb{R}^k$ the projection map. Observe that $A^{\uparrow}$ consists of those points whose path to $\mathcal{T}^{(k)}$ ends in A.

Lemma 5.1.4 carries directly over to our current more general setting. (Intuitively, the proof only uses that branch j attaches to subset $A \subset \mathcal{T}^{(j)}$ with probability $\mu^{(j)}(A)$. This of course is not changed by changing the distribution of the stick lengths).

Lemma 6.3.3. Let $A \subset \mathcal{T}^{(k)}$. Then $\mu^{(j)}(A^{\uparrow})$ is a martingale for $j \geq k$ in filtration $\sigma(\mathcal{T}^{(j)})$.

In particular, this ensures that for j, m large enough, $|\mu^{(j)}(A^{\uparrow}) - \mu^{(m)}(A^{\uparrow})| < \epsilon$ almost surely. We now proof Theorem 6.3.1

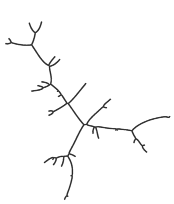
Proof of Theorem 6.3.1. Fix $\epsilon > 0$. From Theorem 6.2.1, we may take K large enough so that $\mathbb{P}\big(d_H\big(\mathcal{T}^{(K)}, \mathcal{T}\big) > \epsilon\big) < \epsilon$. Condition on $\mathcal{T}^{(K)}$ and Let $J_1, \ldots, J_{N_\epsilon}$ be the smallest partition of $\mathcal{T}^{(K)}$ of sets with diameter at most ϵ. On event $E = \{d_H\big(\mathcal{T}^{(K)}, \mathcal{T}\big) > \epsilon\}$, we have that $J_1^{\uparrow} \cap \mathcal{T}^{(j)}, \ldots, J_{N_\epsilon}^{\uparrow} \cap \mathcal{T}^{(j)}$ forms a partition of $\mathcal{T}^{(j)}$ of sets of diameter at most 3ϵ. Hence, via the exact same reasoning as the proof of Theorem 5.1.6, we get for all $\delta, b > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(d_P(\mu^{(j)}, \mu^{(m)}) \geq \delta N_\epsilon + 3\epsilon \;\Big|\; \mathcal{T}^{(K)}\Big), \\
&\qquad\qquad \leq \sum_{i=1}^{N_\epsilon} \mathbb{P}\left(|\mu^{(j)}(J_i^{\uparrow}) - \mu^{(m)}(J_i^{\uparrow})| > \delta \;\middle|\; \mathcal{T}^{(K)}\right) \mathbb{1}_{\{N_\epsilon \leq b\}} + \mathbb{1}_{\{N_\epsilon > b\}} + \mathbb{1}_{\{E^c\}}.
\end{aligned}$$

By taking expectations, and setting $\delta = \frac{\epsilon}{N_\epsilon}$, this turns into,

$$\mathbb{P}(d_P(\mu^{(j)}, \mu^{(m)}) \geq 4\epsilon) \leq \sum_{i=1}^{b} \mathbb{P}\left(|\mu^{(j)}(J_i^{\uparrow}) - \mu^{(m)}(J_i^{\uparrow})| > \frac{\epsilon}{b}\right) + \mathbb{P}(N_\epsilon > b) + \epsilon.$$

Given that N_ϵ is the least number of sets of diameter ϵ needed to partition $\mathcal{T}^{(K)}$, and $\mathcal{T}^{(K)} \subset \mathcal{T}$ is compact almost surely, we can make $\mathbb{P}(N_\epsilon > b)$ arbitrarily small with finite b. This in turn means that $\mathbb{P}\left(|\mu^{(j)}(J_i^{\uparrow}) - \mu^{(m)}(J_i^{\uparrow})| > \frac{\epsilon}{b}\right)$ can be made arbitrarily small since $\mu^{(j)}(J_i^{\uparrow})$ is a bounded martingale. It follows that $\mathbb{P}(d_P(\mu^{(j)}, \mu^{(m)}) \geq 4\epsilon)$ is arbitrarily small for n, m large enough. Hence $\big(\mu^{(j)}\big)$ is Cauchy in probability, which shows Theorem 6.3.1. ☐



6.3.2 Convergence Measure on the Discrete Tree

In this section, we show *iv)* of Remark 6.1.4. That is,

Theorem 6.3.4. For all $\epsilon > 0$, we have,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(d_P\big(\nu_n^{(k)}, \nu_n\big) > \epsilon\big) = 0.$$

We follow largely the reasoning in Section 5.2. However, the reasoning in this section turns out to be easier as the probability of $S_n^f(i)$ being a repeat is given by $f\left(\frac{i}{n}\right)$ which is independent of the number of repeats that occoured in $S_n^f(1), \dots, S_n^f(i-1)$. This was different in Section 5.2, where uniform sampling of sequence S_n meant that $\mathbb{P}\big(S_n(i) \text{ is a repeat }\big) = \frac{|\{S_n(1),\dots,S_n(i-1)|}{n}$ which depends on the number of repeats in $S_n(1), \dots, S_n(i-1)$.

Recall that for $A \subset \mathcal{T}_n^{(k)}$, we define $A^\uparrow \subset \mathcal{T}_n$ to be A together with all vertices $v \in \mathcal{T}_n \setminus \mathcal{T}_n^{(k)}$ for which the path from v to $\mathcal{T}^{(k)}$ ends in A. We have,

Lemma 6.3.5. Let $A \subset \mathcal{T}_n^{(k)}$. For $j \geq k$, $\nu_n^{(j)}(A^\uparrow)$ is a martingale in filtration $F_j = \sigma(\mathcal{T}_n^{(j)})$.

Proof. This follows from the same reasoning used to prove Lemma 5.2.2. □

Lemma 6.3.6. For $A \subset \mathcal{T}_n^{(k)}$, we have,

$$\mathbb{P}\left((\nu_n^{(k)}(A^\uparrow) - \nu_n(A^\uparrow))^2 \geq C^2 \mid \mathcal{T}_n^{(k)}\right) \leq \frac{X_{n,k}}{C^2} = \frac{1}{C^2}\sum_{j=k}^{N} \mathbb{E}\left[\frac{2n^{2\beta}}{(C_j^n)^{2\beta+2}} \,\middle|\, \mathcal{T}_n^{(k)}\right],$$

where N denotes the random number of branches of $\mathcal{T}_n$. Note that $X_{n,k}$ is a random variable.

Proof. This proof is largely the same as that of Lemma 5.2.6. We focus on the differences. Via identical reasoning, we obtain,

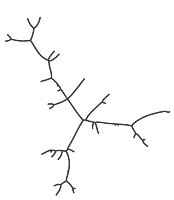
$$\mathbb{P}\left((\nu_n^{(k)}(A^\uparrow) - \nu_n(A^\uparrow))^2 \geq C^2 \mid \mathcal{T}_n^{(k)}\right) \leq \frac{1}{C^2}\sum_{j=k}^{N} \mathbb{E}\left[\frac{(c_{j+1}^n)^2}{(C_j^n)^2} \,\middle|\, \mathcal{T}_n^{(k)}\right],$$

with N being the random number of sticks in $\mathcal{T}_n$ and $c_j^n = C_j^n - C_{j-1}^n$. Observe that $c_{j+1}^n > x$ given $\mathcal{T}_n^{(j)}$ happens precisely when $S_n^f(C_j^n + 1), \dots, S_n^f(C_j^n + x)$ are all not repeats. This happens with probability,

$$\mathbb{P}\big(c_{j+1}^n > x \mid \mathcal{T}_n^{(j)}\big) = \prod_{i=1}^{x}\left(1 - \left(\frac{C_j^n + x}{n}\right)^\beta\right) \leq \left(1 - \left(\frac{C_j^n}{n}\right)^\beta\right)^x,$$

and thus $c_{j+1}^n \mid \mathcal{T}_n^{(j)}$ is dominated by a $Y \sim \text{Geom}\left(\left(\frac{C_j^n}{n}\right)^\beta\right)$ random variable. In particular, this means,

$$\mathbb{E}\left[\left(c_{j+1}^n\right)^2 \mid \mathcal{T}_n^{(j)}\right] \leq \mathbb{E}[Y^2] \leq \frac{2n^{2\beta}}{(C_j^n)^{2\beta}}.$$



By using the tower property of expectation, we obtain,

$$\begin{aligned}\mathbb{P}\left((\nu_n^{(k)}(A^{\uparrow}) - \nu_n(A^{\uparrow}))^2 \geq C^2 \mid \mathcal{T}_n^{(k)}\right) &\leq \frac{1}{C^2}\sum_{j=k}^{N}\mathbb{E}\left[\frac{(c_{j+1}^n)^2}{(C_j^n)^2} \,\middle|\, \mathcal{T}_n^{(k)}\right], \\ &\leq \frac{1}{C^2}\sum_{j=k}^{N}\mathbb{E}\left[\frac{2n^{2\beta}}{(C_j^n)^{2\beta+2}} \,\middle|\, \mathcal{T}_n^{(k)}\right],\end{aligned}$$

as desired. □

We continue with showing $\limsup_{n\to\infty}\mathbb{E}\big[X_{n,k}\big] \to 0$ as $k \to \infty$.

Lemma 6.3.7. We have,

$$\limsup_{n\to\infty}\mathbb{E}\left[\sum_{j=k}^{N}\mathbb{E}\left[\frac{2n^{2\beta}}{(C_j^n)^{2\beta+2}} \,\middle|\, \mathcal{T}_n^{(k)}\right]\right] \xrightarrow[k\to\infty]{} 0.$$

Proof. We have $N \leq n$ as a tree with n vertices can have at most n branches. Thus,

$$\sum_{j=k}^{N}\mathbb{E}\left[\frac{2n^{2\beta}}{(C_j^n)^{2\beta+2}} \,\middle|\, \mathcal{T}_n^{(k)}\right] \leq \sum_{j=k}^{n}\mathbb{E}\left[\frac{2n^{2\beta}}{(C_j^n)^{2\beta+2}} \,\middle|\, \mathcal{T}_n^{(k)}\right],$$

where we set $C_j^n = n$ for $j > N$. Next, observe that,

$$\mathbb{P}(C_j^n < x) \leq \mathbb{P}(X \geq j) \text{ where } X \sim \text{Binom}\left(x, \left(\frac{x}{n}\right)^{\beta}\right).$$

Indeed, $C_j^n < x$ can only happen if $S_n^f(1), \ldots, S_n^f(x)$ contains at least j repeats. The bound follows since $\mathbb{P}(S_n^f(i) \text{ is a repeat}) \leq \left(\frac{x}{n}\right)^{\beta}$. By using a Chernoff bound for the binomial distribution with $\mu = \frac{x^{\beta+1}}{n^{\beta}}$ and $(1+\delta)\mu = j$, (valid for $\delta > -1$), we obtain,

$$\mathbb{P}\big(C_j^n \leq x\big) \leq \mathbb{P}(X \geq j) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \leq \left(\frac{e\mu}{j}\right)^{j} = \left(\frac{ex^{\beta+1}}{jn^{\beta}}\right)^{j}.$$

Using this, we compute,

$$\begin{aligned}\mathbb{E}\left[\frac{1}{(C_j^n)^{2\beta+2}}\right] &\leq \sum_{k=1}^{\infty}\left(\frac{1}{k^{2\beta+2}} - \frac{1}{(k+1)^{2\beta+2}}\right)\mathbb{P}(C_j^n < k), \\ &\leq C\sum_{k=1}^{M}k^{-2\beta-3}\left(\frac{ek^{\beta+1}}{jn^{\beta}}\right)^{j} + \sum_{k=M+1}^{\infty}\left(\frac{1}{k^{2\beta+2}} - \frac{1}{(k+1)^{2\beta+2}}\right), \\ &\leq C\left(\frac{e}{jn^{\beta}}\right)^{j}\sum_{k=1}^{M}k^{j(\beta+1)-2\beta-3} + M^{-2\beta-2}, \\ &\leq C\left(\frac{e}{jn^{\beta}}\right)^{j}\left(M^{\beta+1}\right)^{(j-2)} + \left(M^{(\beta+1)}\right)^{-2}.\end{aligned}$$

By choosing $M^{\beta+1} = n^{\beta} j e^{-1}$, we obtain,

$$\mathbb{E}\left[\frac{1}{(C_j^n)^{2\beta+2}}\right] \leq C j^{-2} n^{-2\beta},$$

for some constant C. By putting everything together, we obtain,

$$\limsup_{n\to\infty} \mathbb{E}\left[\sum_{j=k}^{N} \mathbb{E}\left[\frac{2n^{2\beta}}{(C_j^n)^{2\beta+2}} \;\middle|\; \mathcal{T}_n^{(k)}\right]\right] \leq \limsup_{n\to\infty} \sum_{j=k}^{n} \mathbb{E}\left[\frac{2n^{2\beta}}{(C_j^n)^{2\beta+2}}\right] \leq C \sum_{j=k}^{\infty} \frac{1}{j^2}.$$

This conclude the proof as the last sum goes to zero as $k \to \infty$. $\square$

Theorem 6.3.4 can now be shown following identical steps as the proof of Theorem 5.2.5. $\square$

7 A Threshold for Compactness

In Section 6, we saw that sampling S_n non-uniformly by setting $\mathbb{P}(S_n(i)) = f\left(\frac{i}{n}\right)$ for some f allows for convergence to trees $\mathcal{T}$ where the stick lengths are determined by a PPP of intensity $t^\beta dt$. Importantly, the stick lengths obtained from these PPP's still decrease quick enough to be able to show tightness of $\left(\mathcal{T}_n, n^{-\frac{1}{2}}, \nu_n\right)_{n\in\mathbb{N}}$ in the GHP-topology (Section 6.2 and Section 6.3). In this section, we find where this tightness argument breaks down. We do this by considering functions of the form $f_n(i) : \{2, \dots, n\} \to [0,1]$. More specifically, we use

$$f(i,n) \equiv f_n^\gamma(i) = \ln^\gamma(in^{-\frac{1}{2}} + 1)n^{-\frac{1}{2}},$$

parametrized by $\gamma > 0$ and we let,

$$S_n \equiv S_n^f(i) \sim \begin{cases} S_n(j) \text{ where } j \in_u \{1, \dots, i-1\}, & \text{with probability } f(i,n), \\ \text{Unif}\big([n] \setminus \{S_n(1), \dots, S_n(i-1)\}\big), & \text{with probability } 1 - f(i,n). \end{cases}$$

Let $\mathcal{T}_n \equiv \mathcal{T}_{n,f}$ be the random tree constructed by applying the Foata–Fuchs bijection to S_n.

On the continuous side, we let η be a PPP on $\mathbb{R}_{\geq 0}$ of intensity $\ln^\gamma(t+1)dt$ and let $\mathcal{T} \equiv \mathcal{T}_\gamma$ be the continuum random tree constructed from applying the stick-breaking construction to sticks with the points in η as end points.

The definitions of $C_i^n, B_i^n, C_i, B_i, \mathcal{T}_n^{(k)}, \mathcal{T}^{(k)}, \nu_n^{(k)}, \mu^{(k)}, \nu_n$ and μ are taken as in Definition 6.0.1 but applied to the trees $\mathcal{T}_n$ and $\mathcal{T}$ in the current setting. The aim of this section is to show.

Theorem 7.0.1. For $\gamma > 1$, we have convergence in the GHP-topology,

$$\big(\mathcal{T}_n, n^{-\frac{1}{2}} d_n, \nu_n\big) \xrightarrow[n\to\infty]{d} \big(\mathcal{T}, d, \mu\big).$$

If $\gamma \leq 1$, then $\mathcal{T}$ is not compact almost surely.

Remark 7.0.2. It should be noted that $\mathcal{T}$ not being compact automatically implies that we cannot have convergence in the GHP-topology as the metric d_{GHP} is defined on the space of compact measure metric spaces (up to isometries).

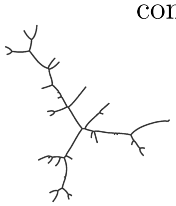
We start this section by showing convergence of finite dimensional distributions.

Theorem 7.0.3. For all $k \in \mathbb{N}$ and $\gamma > 0$, we have convergence in the GHP-topology,

$$\big(\mathcal{T}_n^{(k)}, n^{-\frac{1}{2}} d_n, \nu_n^{(k)}\big) \xrightarrow[n\to\infty]{d} \big(\mathcal{T}^{(k)}, d, \mu^{(k)}\big),$$

Remark 7.0.4. Observe that this result is independent of $\gamma > 1$. Intuitively, the convergence of $n^{-\frac{1}{2}}\mathcal{T}_n$ to $\mathcal{T}$ fails for $\gamma \leq 1$ because the sticks do not become short quick enough (in fact, we will see they do not become short at all). Thus, the problem lies in showing tightness, not the finite dimensional distributions.

Remark 7.0.5. By the same reasoning as Proposition 2.4.3, Theorem 7.0.1 is proven by combining Theorem 7.0.3 together with the statements below.



For all $\epsilon > 0$ and $\gamma > 1$, we have,

$$\text{i)} \lim_{t\to\infty} \mathbb{P}\Big(d_H(\mathcal{T}(t), \mathcal{T}) > \epsilon\Big), \qquad \text{ii)} \lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big(d_H(\mathcal{T}_n(tn^{1/2}), \mathcal{T}_n) > \epsilon\Big),$$

$$\text{iii)} \lim_{k\to\infty} \mathbb{P}\Big(d_P(\mu^{(k)}, \mu) > \epsilon\Big), \qquad \text{iv)} \lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big(d_P(\nu_n^{(k)}, \nu_n) > \epsilon\Big).$$

7.1 Finite Dimensional Distribution

In this Section, we prove 7.0.3. For this, we aim to prove the following result.

Theorem 7.1.1. Take C_i^n and B_i^n as above. For all $k \in \mathbb{N}$, we have,

$$n^{-\frac{1}{2}}\big(C_1^n, \dots, C_k^n, B_1^n, \dots, B_k^n\big) \xrightarrow[n\to\infty]{d} \big(C_1, \dots, C_k, B_1, \dots, B_k\big),$$

where $C_1 < \cdots < C_k$ are the first k points of a PPP with intensity $\ln^\gamma(t+1)dt$ on $\mathbb{R}_{\geq 0}$ and $B_i \sim \text{Unif}([0, C_i])$.

Before proving the above theorem, we introduce three lemmas. Throughout the following lemmas, we write for $t = (t_1, \dots, t_l)$ and $x = (x_1, \dots, x_l)$, with $l \geq k$,

$$i)\ B^k(t) \text{ for } B_1^n \leq t_1 C_1^n, \dots, B_k^n \leq t_k C_k^n \ \text{ and } \ ii)\ C^k(x) \text{ for } C_1^n = \lfloor x_1 n^{\frac{1}{2}} \rfloor, \dots, C_k^n = \lfloor x_k n^{\frac{1}{2}} \rfloor,$$

to lighten notation.

Lemma 7.1.2. For arbitrary $k \in \mathbb{Z}_{\geq 0}$, we have,

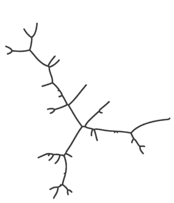
$$n^{\frac{k}{2}} \mathbb{P}\left(C_1^n = \lfloor x_1 n^{\frac{1}{2}} \rfloor, \dots, C_k^n = \lfloor x_k n^{\frac{1}{2}} \rfloor\right) \xrightarrow[n\to\infty]{\text{u.c.}} f_{C_1,\dots,C_k}(x_1, \dots, x_k),$$

where $f_{C_1,\dots,C_k}(x_1, \dots, x_k)$ is the pdf of the first k points of a PPP of intensity $\ln^\gamma(t+1)dt$.

Remark 7.1.3. In Section 3 and 6.1, we used Taylor expansions for approximations. Here, this also is the first step. Furthermore, $f(i, n)$ is chosen such that we can use Riemann sums to pass to the limit.

Proof. We know $f_{C_1,\dots,C_k}(x_1, \dots, x_k) = \prod_{i=1}^k \ln^\gamma(x_i + 1) \exp\left(-\int_0^{x_k} \ln^\gamma(t+1)dt\right)$, as seen in Lemma 2.2.22. We proceed by induction on k. For $k = 1$, we obtain,

$$\begin{aligned}
n^{\frac{1}{2}} \mathbb{P}(C_1^n = \lfloor x_1 n^{\frac{1}{2}} \rfloor) &= n^{\frac{1}{2}} \ln^\gamma\left(\lfloor x_1 n^{\frac{1}{2}} \rfloor n^{-\frac{1}{2}} + 1\right) n^{-\frac{1}{2}} \prod_{i=2}^{\lfloor x_1 n^{\frac{1}{2}} \rfloor - 1} \left(1 - \ln^\gamma(in^{-\frac{1}{2}} + 1)n^{-\frac{1}{2}}\right), \\
&= \ln^\gamma(x_1 + 1) \exp\left(\sum_{i=2}^{\lfloor x_1 n^{\frac{1}{2}} \rfloor - 1} \ln\left(1 - \ln^\gamma(in^{-\frac{1}{2}} + 1)n^{-\frac{1}{2}}\right)\right) + o(1), \\
&= \ln^\gamma(x_1 + 1) \exp\left(o(1) - \sum_{i=2}^{\lfloor x_1 n^{\frac{1}{2}} \rfloor - 1} \ln^\gamma\left(in^{-\frac{1}{2}} + 1\right)n^{-\frac{1}{2}}\right), \\
&= \ln^\gamma(x_1 + 1) \exp\left(o(1) - \int_0^{x_1} \ln^\gamma(t+1)dt\right) + o(1).
\end{aligned}$$



with $o(1)$ denotes being uniformly small on compact sets. We used a first order approximation of $\ln(1-x)$ to go from the second to third line and used that $\sum_{i=2}^{\lfloor x_1 n^{\frac{1}{2}} \rfloor -1} \ln^\gamma \left(in^{-\frac{1}{2}} + 1 \right) n^{-\frac{1}{2}}$ is a Riemann sum for $\int_0^{x_1} \ln^\gamma(t+1)dt$ with step size $n^{-\frac{1}{2}}$. The terms $i = 1$ and $i = \lfloor x_1 n^{\frac{1}{2}} \rfloor - 1$ are missing, but both terms are $o(1)$, and hence do not pose problems. Since $\ln^\gamma(t+1)$ is continuous for $t \geq 0$, the convergence to the integral is uniform on compact sets. This establishes the base case.

We continue with the induction step. Using the induction hypothesis, we obtain,

$$\begin{aligned} n^{\frac{k+1}{n}} \mathbb{P}\big(C^{k+1}(x)\big) &= n^{\frac{1}{2}} \mathbb{P}\big(C^n_{k+1} = \lfloor x_{k+1} n^{\frac{1}{2}} \rfloor \,\big|\, C^k(x)\big) \cdot n^{\frac{k}{2}} \mathbb{P}\big(C^k(x)\big), \\ &= n^{\frac{1}{2}} \mathbb{P}\big(C^n_{k+1} = \lfloor x_{k+1} n^{\frac{1}{2}} \rfloor \,\big|\, C^k(x)\big) \left(\prod_{i=1}^{k} \ln^\gamma(x_i+1) \exp\left(-\int_0^{x_k} \ln^\gamma(t+1)dt \right) + o(1) \right). \end{aligned}$$

with $o(1)$ denoting uniformly small for $(x_1, \ldots, x_k)$ in compact sets. Thus it suffices to show,

$$n^{\frac{1}{2}} \mathbb{P}\big(C^n_{k+1} = \lfloor x_{k+1} n^{\frac{1}{2}} \rfloor \,\big|\, C^k(x)\big) \xrightarrow[n\to\infty]{\text{u.c.}} \ln^\gamma\big(x_{k+1}+1\big) \exp\left(-\int_{x_k}^{x_{k+1}} \ln^\gamma(t+1)dt \right).$$

This is verified by following the same steps as in the base case,

$$\begin{aligned} & n^{\frac{1}{2}} \mathbb{P}\big(C^n_{k+1} = \lfloor x_{k+1} n^{\frac{1}{2}} \rfloor \,\big|\, C^k(x)\big), \\ &= n^{\frac{1}{2}} \ln\left(\lfloor x_{k+1} n^{\frac{1}{2}} \rfloor n^{-\frac{1}{2}} + 1 \right)^\gamma n^{-\frac{1}{2}} \prod_{i=\lfloor s_k n^{\frac{1}{2}} \rfloor + 1}^{\lfloor x_{k+1} n^{\frac{1}{2}} \rfloor - 1} \left(1 - \ln\left(in^{-\frac{1}{2}} + 1 \right)^\gamma n^{-\frac{1}{2}} \right), \\ &= \Big(\ln(x_{k+1}+1)^\gamma + o(1) \Big) \exp\left(o(1) - \sum_{i=\lfloor s_k n^{\frac{1}{2}} + 1 \rfloor}^{\lfloor x_{k+1} n^{\frac{1}{2}} \rfloor} \ln\left(in^{-\frac{1}{2}} + 1 \right)^\gamma n^{-\frac{1}{2}} \right), \\ &= \ln(x_{k+1}+1)^\gamma \exp\left(-\int_{x_k}^{x_{k+1}} \ln(t+1)^\gamma dt \right) + o(1), \end{aligned}$$

where convergence is uniformly small on compact sets. This concludes the proof. □

Lemma 7.1.4. We have uniform convergence over compact sets,

$$\mathbb{P}\Big(B^n_1 \leq t_1 C^n_1, \ldots, B^n_k \leq t_k C^n_k \,\Big|\, C^n_1 = \lfloor x_1 n^{\frac{1}{2}} \rfloor, \ldots, C^n_k = \lfloor x_k n^{\frac{1}{2}} \rfloor \Big) \xrightarrow[n\to\infty]{\text{u.c.}} \prod_{i=1}^{k} t_i.$$

Proof. This proof is identical to Lemma 3.1.6, with only a small modification for sampling repeat $S_n(i)$ uniformly in $\{\{S_n(1), \ldots, S_n(i-1)\}\}$ instead of in $\{S_n(1), \ldots, S_n(i-1)\}$. □

Proof of Theorem 7.1.1. It suffices to show,

$$i)\ \mathbb{P}\big(C^n_1 \leq s_1 n^{\frac{\gamma}{\gamma+1}}, \ldots, C^n_k \leq s_k n^{\frac{\gamma}{\gamma+1}}\big) \xrightarrow[n\to\infty]{} \mathbb{P}\big(C_1 \leq s_1, \ldots, C_k \leq s_k\big),$$

$$ii)\ \mathbb{P}(B^n_1 \leq t_1 C^n_1, \ldots, B^n_k \leq t_k C^n_k \mid C^n_1 \leq s_1 n^{\frac{1}{2}}, \ldots, C^n_k \leq s_k n^{\frac{1}{2}}),$$

$$\xrightarrow[n\to\infty]{} \mathbb{P}\big(B_1 \leq t_1 C_1, \ldots, B_k \leq t_k C_k \mid C_1 \leq s_1, \ldots, C_k \leq s_k\big)$$

Statement $i)$ follows from identical reasoning as the proof of Lemma 3.1.5. Statement $ii)$ follows from a similar proof to that of Lemma 3.1.7. □

As seen in Remark 3.3.4, having convergence of scaled repeat points,

$$n^{-\frac{1}{2}}\big(C_1^n, \dots, C_k^n, B_1^n, \dots, B_k^n\big) \xrightarrow[n\to\infty]{d} \big(C_1, \dots, C_k, B_1, \dots, B_k\big),$$

implies that the respective partial trees $\mathcal{T}_n^{(k)}$ and $\mathcal{T}^{(k)}$ and uniform measures $\nu_n^{(k)}$ and $\mu^{(k)}$ must also converge in distribution in the GHP-topology. Theorem 7.0.3 follows immediately. □

7.2 Where $\mathcal{T}_\gamma$ Fails to Be Compact

In this section, we show that $\gamma \le 1$ implies that $\mathcal{T}$ is not compact almost surely, by showing that $\mathcal{T}$ is constructed from an infinite number of sticks of length exceeding 1, or that $\mathcal{T}$ contains a stick of infinite length.

Let η be a PPP of intensity $\ln^\gamma(t+1)$ and for $n \in \mathbb{N}$. Define event $A_n = \big\{\eta([n-1, n)) = 0\big\}$. Note that,

$$\mathbb{P}(A_n) = \exp\left(-\int_{n-1}^{n} \ln(t+1)^\gamma dt\right) \ge \exp\left(-\ln(n+1)^\gamma\right).$$

Hence $\sum_{n=1}^\infty \mathbb{P}(A_n) \ge \sum_{n=2}^\infty \exp(-\ln(n)^\gamma)$ diverges whenever $\gamma \in (0, 1]$. Given that the events $(A_n)_{n\ge1}$ are independent, we may apply the Borel–Cantelli lemma, to obtain that infinitely many events A_n must occur with probability 1. We conclude there must be a stick of infinite length, or infinitely many sticks of length exceeding 1.

In the first case, $\mathcal{T}$ is clearly not compact. In the second case, let $I = \{i \in \mathbb{N} : C_i - C_{i-1} \ge 1\}$ be the set of indices of the sticks with length exceeding 1. It follows that $|I| = \infty$ almost surely. Define the sequence $\{\rho(C_i)\}_{i\in I} \subset \mathcal{T}$. For any $i < j \in \mathbb{N}$, we see $d_H\big(\rho(C_i), \rho(C_j)\big) \ge 1$ as the path from $\rho(C_j)$ to $\rho(C_i)$ must necessarily traverse all of $\rho([C_{j-1}, C_j])$. Hence $\{\rho(C_i)\}_{i\in I}$ cannot have a convergent subsequence and $\mathcal{T}$ is not compact.

In particular, recall that $\lim_{t\to\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(t), \mathcal{T}\big) > \epsilon\big) = 0$ implies $\mathcal{T}$ is compact almost surely. Hence, not all criteria in Remark 7.0.5 can be shown for $\gamma \le 1$. Furthermore, for $\gamma \le 1$, tree $\mathcal{T}$ cannot be seen as limiting random variable using the GHP-topology as this topology is defined on compact metric spaces.

The above argument fails for $\gamma > 1$ since $\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty$, and thus we may hope to show $\mathcal{T}$ is compact when $\gamma > 1$.

Remark 7.2.1. One might think that we can do better than having a threshold at $\gamma = 1$ for $\ln(t+1)^\gamma$. An obvious example to try would be $\delta \ln(t+1)dt$, we explore this below.

Let η be a PPP of intensity $\delta \ln(t+1)dt$ and again, let A_n be the event $\eta([n-1, n]) = 0$. Then,

$$\sum_{n=1}^\infty \mathbb{P}(A_n) = \sum_{n=1}^\infty \exp\left(-\delta \int_{n-1}^{n} \ln(t+1)dt\right) \ge \sum_{n=1}^\infty \exp\left(-\delta \ln(n+1)\right) = \sum_{n=1}^\infty \left(\frac{1}{n+1}\right)^\delta.$$

And thus, one could expect a threshold at $\delta = 1$. However, we can sharpen this reasoning by letting B_i^ϵ be the event $\eta([(i-1)\epsilon, i\epsilon) = 0)$. Following the computations above shows that,

$$\sum_{i=1}^{\infty} \mathbb{P}(B_i^\epsilon) = \sum_{i=1}^{\infty} \exp\left(-\delta \int_{(i-1)\epsilon}^{i} \epsilon \ln(t+1) dt\right) \geq \sum_{i=1}^{\infty} \exp(-\delta\epsilon \ln(i+1)) = \sum_{i=1}^{\infty} \left(\frac{1}{i+1}\right)^{\epsilon\delta}.$$

Thus for any $\delta > 0$, we may choose $\epsilon = \delta^{-1}$ and with probability 1, there will be either an infinitely long branch, or infinitely many sticks of length exceeding ϵ. For any δ the CRT constructed from a PPP of intensity $\delta \ln(t+1)$ cannot be compact.

7.3 Tightness of $\left(\mathcal{T}_{n,f}, n^{-\frac{1}{2}} d_n\right)_{n\in\mathbb{N}}$

In this section, we show $i)$ and $ii)$ of Remark 7.0.5. That is, we aim to show,

$$i) \lim_{t\to\infty} \mathbb{P}\left(d_H\left(\mathcal{T}(t), \mathcal{T}\right) > \epsilon\right) = 0 \quad \text{and} \quad ii) \lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(d_H\left(\mathcal{T}_n(tn^{\frac{1}{2}}), \mathcal{T}_n\right) > \epsilon\right) = 0.$$

We start with $i)$.

7.3.1 Compactness of $\mathcal{T}_\gamma$

Theorem 7.3.1. For all $\epsilon > 0$, we have,

$$\lim_{t\to\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(t), \mathcal{T}\big) > \epsilon\big) = 0.$$

Remark 7.3.2. We would like to follow the reasoning in Section 4.1 and Section 6.2.1. However, we run into a problem when trying to generalize Lemma 4.1.5 respectively Lemma 6.2.3. In this lemma, we aim to bound $\mathbb{P}\big(d_H\big(\mathcal{T}(a), \rho(s)\big) > c\big)$ for $a, c > 0$ and $s \in [a, 2a]$. To adapt the proof to the current setting, we would sample repeat/attachment points C_i, B_i via a homogeneous PPP of intensity 1 on the region $T = \{(x, y) \in \mathbb{R}^2 : x \geq 0, 0 \leq y \leq \ln^\gamma(x+1)\}$. Identical reasoning yields that $d_H\big(\mathcal{T}(a), \rho(s)\big) > c$ can only happen if $\eta(S) = 0$ where S is a region of measure $\lambda(S) = c\frac{\ln^\gamma(a+1)}{2}$, and thus $\mathbb{P}\big(d_H\big(\mathcal{T}(a), \rho(s)\big) > c\big) \leq \exp\left(-\frac{c\ln^\gamma(a+1)}{2}\right)$.

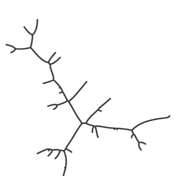
This can be extended to $\mathbb{P}\big(d_H\big(\mathcal{T}(a), \mathcal{T}(2a)\big) > c\big) \leq \frac{2a}{c} \exp\left(-\frac{c\ln^\gamma(a+1)}{4}\right)$ via the trick used in Lemma 4.2.12 and Lemma 6.2.9. Filling in $a = 2^i t$ and $c = \epsilon_i$ as before yields,

$$\mathbb{P}\big(d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1}t)\big) > \epsilon_i\big) \leq \exp\left(\ln\left(\frac{2^{i+1}t}{\epsilon_i}\right) - \frac{\epsilon_i \ln^\gamma(2^i t + 1)}{4}\right).$$

Since we require $\sum_{i=1}^{\infty} \epsilon_i(t) \to 0$ as $t \to \infty$, we need ϵ_i to be of the order $i^{-(1+\delta)}$ for $\delta > 0$. However, with such a choice of ϵ_i, we see $\ln\left(\frac{2^{i+1}t}{\epsilon_i}\right) - \frac{\epsilon_i \ln^\gamma(2^i t+1)}{4}$ stays positive and hence we have no hope of showing $\mathbb{P}\big(d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1}t)\big) > \epsilon_i\big)$ is summable.

We aim to solve this issue by looking at $d_H\big(\mathcal{T}(x_i t), \mathcal{T}(x_{i+1}t)\big)$ for some $x_i \to \infty$ as $i \to \infty$, instead of doublings $d_H\big(\mathcal{T}(2^i t), \mathcal{T}(2^{i+1}t)\big)$. By following identical computations, we obtain,

$$\mathbb{P}\big(d_H\big(\mathcal{T}(x_i t), \mathcal{T}(x_{i+1}t)\big) > \epsilon_i\big) \leq \exp\left(\ln\left(\frac{2x_{i+1}t}{\epsilon_i}\right) - \epsilon_i \frac{\ln^\gamma(x_i t + 1) x_i}{2x_{i+1}}\right).$$



We explore if this is possibly below. For this, consider t fixed and remove all constants, we see this quantity is summable in i only if $\epsilon_i \ln^{\gamma-1}(x_i) > \frac{x_{i+1}}{x_i} > 1$. Given that ϵ_i must be of the order $i^{-(1+\delta)}$, this requires $x_i = \exp(i^\alpha)$ with $\alpha(\gamma-1) = 1+\delta$, or a faster growing sequence x_i. However, with such a choice of x_i, the fraction $\frac{x_{i+1}}{x_i}$ grows much faster than $\epsilon_i \ln^{\gamma-1}(x_i)$ showing that no choice of sequence x_i salvages the reasoning from Section 4.1 and Section 6.2.1.

In the current section, we use a different technique for proving the analogous version of Lemma 4.1.5 and Lemma 6.2.3, yielding a tighter bound. The new bound allows for proving almost sure compactness of $\mathcal{T}$.

We start with a generalization of Lemma 4.1.2.

Lemma 7.3.3. Fix $\epsilon > 0$ and let $\big(x_i\big)_{i\in\mathbb{N}_0}$ be an increasing sequence such that $x_0 = 1$ and $x_i \to \infty$ as $i \to \infty$. Suppose $\epsilon_i \equiv \epsilon(T) > 0$ exist such that $\sum_{i=1}^{\infty} \epsilon_i < \epsilon$. Then for any $t > 0$,

$$\mathbb{P}\big(d_H\big(\mathcal{T}(t), \mathcal{T}\big) > \epsilon\big) \leq \sum_{i=0}^{\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(x_i t), \mathcal{T}(2x_{i+1}t)\big) > \epsilon_i\big).$$

Proof. Identical to the proof of Lemma 4.1.2 upon replacing 2^i with x_i. □

Corollary 7.3.4. Theorem 7.3.1 is shown upon finding $\epsilon_i(t) > 0$ such that,

$$i) \lim_{t\to\infty} \sum_{i=0}^{\infty} \epsilon_i(t) = 0 \quad \text{and} \quad ii) \lim_{t\to\infty} \sum_{i=0}^{\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(x_i t), \mathcal{T}(x_{i+1}t)\big) > \epsilon_i\big) = 0,$$

where x_i is some increasing, diverging sequence with $x_0 = 1$.

Throughout the remainder of this section, we fix $x_i = \exp\big(i^\alpha\big)$ for parameter $\alpha \equiv \alpha(\gamma) > 0$ to be determined. We start with proving a tighter bound for $\mathbb{P}\big(d_H\big(\mathcal{T}(a), \mathcal{T}(b)\big) > c\big)$ for some $1 \leq a < b$ and $c > 0$.

Lemma 7.3.5. Fix $1 \leq a < b$ and let $l \in [a, b]$, then for all $c > 0$ we have,

$$\mathbb{P}\left(d_H\left(\mathcal{T}(a), \rho(l)\right) > c\right) \leq 4\left(\frac{b}{a}\right)^2 \exp\left(-\frac{c\ln^\gamma(a+1)}{4}\right).$$

Proof. Let $U_1, U_2, \cdots \sim$ Unif$([0,1])$ i.i.d. random variables, and let η denote a PPP of intensity $\ln^\gamma(t+1)dt$. Set $\eta' = \eta + \delta_a$. Set $t_1 = l$ and define recursively:

$$p(t_i) = \max\{x \in \eta \text{ s.t. } x \leq t_i\}, \qquad d_i = t_i - p(t_i), \qquad t_{i+1} = \max\{a, U_i p(t_i)\},$$

so that $p(t_i)$ denotes the starting point of the stick on which t_i lies and $\rho(t_{i+1}) \in \mathcal{T}(2a)$ is the location this stick gets attached to. Let $N = \min_{i\in\mathbb{Z}}\{U_i p(t_i) \leq a\}$ denote the number of sticks on the path from $\rho(v)$ to $\mathcal{T}(a)$ and note that $d_{N+1}, d_{N+2}, \cdots = 0$.

By construction, $d_H\big(\mathcal{T}(a), \rho(l)\big) = \sum_{i=1}^{\infty} d_i = \sum_{i=1}^{N} d_i$. We aim to bound $\mathbb{P}\left(\sum_{i=1}^{j} d_i > c\right)$ and $\mathbb{P}(N > j)$ for some $j \in \mathbb{N}$ and then upgrade this to a bound on $\mathbb{P}(\sum_{i=1}^{N} d_i > c)$.

To bound $\mathbb{P}(N > j)$, recall $N = \min_{i \in \mathbb{Z}}\{U_i p(t_i) \le a\}$ and thus $t_i \le U_1 \dots U_{i-1} l \le U_1 \dots U_{i-1} b$ for all i. This implies $\mathbb{P}(N > j) \le \mathbb{P}(U_1 \dots U_j b > a)$. We obtain,

$$\mathbb{P}(N > j) \le \mathbb{P}\left(U_1 \dots U_j b > a\right) = \mathbb{P}\left(\sum_{i=1}^{j} -\ln(U_i) < \ln\left(\frac{b}{a}\right)\right) = \mathbb{P}(G_{j,1} < L),$$

where $G_{j,1} \sim \text{Gamma}(j, 1)$ and $L = \ln\left(\frac{b}{a}\right)$. Recall $\mathbb{E}\left[e^{-G_{j,1}s}\right] = \left(\frac{1}{1+s}\right)^{j-1}$ and thus for $s > 0$,

$$\mathbb{P}(N > j) \le \mathbb{P}\left(G_{j,1} < L\right) = \mathbb{P}\left(e^{-G_{j,1}s} > e^{-Ls}\right) \le \mathbb{E}\left[e^{-G_{j,1}s}\right] e^{Ls} = e^{Ls}\left(\frac{1}{1+s}\right)^{j}.$$

To bound $\mathbb{P}\left(\sum_{i=1}^{j} d_i > c\right)$, we first bound $\mathbb{P}\left(d_i > s_i \mid d_{i-1}, \dots, d_1\right)$. We have,

$$\begin{aligned}
\mathbb{P}(d_i > c \mid d_1, \dots, d_{i-1}) &= \mathbb{1}_{\{t_i - c \ge a\}} \mathbb{P}(\eta([t_i - c, t_i]) = 0) \\
&\le \mathbb{1}_{\{t_i - c \ge a\}} \exp\left(-\int_{t_i - c}^{t_i} \ln^{\gamma}(t+1) dt\right), \\
&\le \exp\left(-c \ln^{\gamma}(a+1)\right)
\end{aligned}$$

Define $S_j = \sum_{i=1}^{j} d_i$. Inductively, we obtain that S_j is stochastically dominated by $\sum_{i=1}^{j} X_i$ where $X_i \sim \text{Exp}(\lambda)$ i.i.d. with $\lambda = \ln^{\gamma}(a+1)$. Since e^{tx} is increasing in x for $t > 0$, we get,

$$\mathbb{E}\left[\exp\left(tS_j\right)\right] \le \mathbb{E}\left[\exp\left(t\sum_{i=1}^{j} X_i\right)\right],$$

Recall that $\mathbb{E}[e^{tX_i}] = \frac{\lambda}{\lambda - t}$, for $t \in (0, \lambda)$. In particular, by Markovs inequality, we get,

$$\mathbb{P}\left(\sum_{i=1}^{j} d_i > c\right) = \mathbb{P}\left(e^{S_j t} > e^{ct}\right) \le e^{-ct}\left(\frac{\lambda}{\lambda - t}\right)^{j}.$$

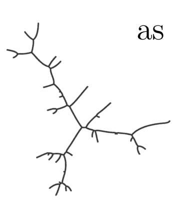
We apply a union bound to get stochastic domination for d.

$$\begin{aligned}
\mathbb{P}(d > c) = \mathbb{P}\left(\sum_{i=1}^{N} d_i > c\right) &= \mathbb{P}\left(\left(\sum_{i=1}^{N} d_i > c, N \le j\right) \bigcup \left(\sum_{i=1}^{N} d_i > c, N > j\right)\right), \\
&\le \mathbb{P}\left(\sum_{i=1}^{N} d_i > c, N \le j\right) + \mathbb{P}\left(\sum_{i=1}^{N} d_i > c, N > j\right) \le \mathbb{P}\left(\sum_{i=1}^{j} d_i > c\right) + \mathbb{P}\left(N > j\right), \\
&\le e^{-ct}\left(\frac{\lambda}{\lambda - t}\right)^{j} + e^{Ls}\left(\frac{1}{1+s}\right)^{j},
\end{aligned}$$

where the above inequality holds for all $s > 0, t \in (0, \lambda)$ and $j \in \mathbb{Z}_{\ge 0}$. By taking the values, $s = e - 1, t = \lambda\frac{e-1}{e}$ and $j = \lfloor c\lambda\frac{e-1}{2e}\rfloor$, we obtain,

$$\mathbb{P}(d > c) \le e^{-ct}\left(\frac{\lambda}{\lambda - t}\right)^{j} + e^{Ls}\left(\frac{1}{1+s}\right)^{j} \le \left(1 + e\left(\frac{b}{a}\right)^{e-1}\right) e^{-\lambda c\frac{e-1}{2e}} \le 4\left(\frac{b}{a}\right)^{2} e^{-\frac{\lambda c}{4}},$$

as desired. □



Lemma 7.3.6. For $1 < a < b$, we have, $\mathbb{P}\left(d_H(\mathcal{T}(a), \mathcal{T}(b)) > c\right) \leq 8\left(\frac{b}{a}\right)^2 \frac{b}{c} \exp\left(\frac{-c\ln^\gamma(a+1)}{8}\right)$.

Proof. Identical to the proof of Lemma 4.1.13. $\square$

We are now in a position to prove Theorem 7.3.1.

Proof of Theorem 7.3.1. We aim to show *i)* and *ii)* of Corollary 7.3.4. By making the substitutions $a = x_i t$, $b = x_{i+1}t$ and $c = \epsilon_i$ in Lemma 7.3.6 and simplifying, condition *ii)* is translated into,

$$\lim_{t\to\infty} \sum_{i=0}^{\infty} \frac{1}{\epsilon_i} \exp\left(3\ln(x_{i+1}) + \ln(t) - \frac{\epsilon_i \ln^\gamma(x_i t)}{8}\right) = 0.$$

We set $\epsilon_i = \frac{8}{\ln^\gamma(x_i t)}\left(3\ln(x_{i+1}) + 2\ln(t) + \ln(g(i))\right)$ for ansatz function $g(i)$ with $g(i) \geq 1$ for all $i \in \mathbb{N}_0$. Recall that $x_i = \exp\left(i^\alpha\right)$ so that we obtain,

$$\sum_{i=0}^{\infty} \frac{1}{\epsilon_i} \exp\left(3\ln(x_{i+1}) + \ln(t) - \frac{\epsilon_i \ln^\gamma(x_i t)}{8}\right) = \frac{1}{t}\sum_{i=0}^{\infty} \frac{1}{\epsilon_i g(i)},$$

so that it suffices to show $\frac{1}{t}\sum_{i=0}^{\infty} \frac{1}{\epsilon_i g(i)} \to 0$ as $t \to \infty$. For this, observe that $\epsilon_i \geq \frac{1}{\ln^\gamma(x_i t)}$ for $t \geq e$ and hence,

$$\begin{aligned}
\frac{1}{t}\sum_{i=0}^{\infty} \frac{1}{\epsilon_i g(i)} \leq \frac{1}{t}\sum_{i=0}^{\infty} \frac{\ln^\gamma(x_i t)}{g(i)} &= \frac{1}{t}\left(\sum_{\{i:i^\alpha \leq \ln(t)\}} \frac{(i^\alpha + \ln(t))^\gamma}{g(i)} + \sum_{\{i:i^\alpha > \ln(t)\}} \frac{(i^\alpha + \ln(t))^\gamma}{g(k)}\right), \\
&\leq \frac{1}{t}\left(2^\gamma \ln^\gamma(t)\ln^{\frac{1}{\alpha}}(t) + 2^\gamma \sum_{i^\alpha > \ln(t)} \frac{i^{\alpha\gamma}}{g(i)}\right) \xrightarrow[t\to\infty]{} 0,
\end{aligned}$$

where in the last step, we chose $g(i) = (i+1)^{2+\alpha\gamma}$, so that $\sum_{i^\alpha > \ln(t)} \frac{i^{\alpha\gamma}}{g(i)}$ is finite. In particular, this shows that,

$$\lim_{t\to\infty} \sum_{i=0}^{\infty} \mathbb{P}\big(d_H\big(\mathcal{T}(x_i t), \mathcal{T}(x_{i+1}t)\big) > \epsilon_i\big) = 0.$$

Lastly, we show $\sum_{i=0}^{\infty} \epsilon_i \to \infty$ as $t \to \infty$. For this, we set $I_1 = \{i \in \mathbb{N}_0 : (i+1)^\alpha \leq \ln(t)\}$ and $I_2 = \{i \in \mathbb{N}_0 : (i+1)^\alpha > \ln(t)\}$. We obtain,

$$\begin{aligned}
\sum_{i=0}^{\infty} \epsilon_i &= 8\sum_{i=1}^{\infty} \frac{3(i+1)^\alpha + 2\ln(t) + \ln\left((i+1)^{\alpha\gamma+2}\right)}{\left(i^\alpha + \ln(t)\right)^\gamma}, \\
&\leq 8\sum_{I_1} \frac{5\ln(t) + (\alpha\gamma+2)\ln(\ln(t))}{\ln^\gamma(t)} + 8\sum_{I_2} \frac{5(i+1)^\alpha + (\alpha\gamma+2)\ln(i+1)}{i^{\alpha\gamma}}, \\
&\leq C_1 \ln(t)^{1-\gamma+\frac{1}{\alpha}} + C_2 \sum_{I_2} i^{\alpha-\alpha\gamma},
\end{aligned}$$

where the final bound holds for t large enough and positive constants C_1 and C_2 possibly depending on α and γ. The final expression goes to zero when $1-\gamma+\frac{1}{\alpha}<0$ and $\alpha-\alpha\gamma<-1$. Both conditions are satisfied when $\alpha(\gamma-1)>1$. We pick $\alpha=\frac{2}{\gamma-1}$ which is positive precisely when $\gamma>1$. Also observe that the criteria $\alpha(\gamma-1)>1$ coincides with the observation made in Remark 7.3.2. □

This completes the proof of Theorem 7.3.1. We continue in the discrete setting.

7.3.2 Compactness of $n^{-\frac{1}{2}}\mathcal{T}_{n,f}$

Theorem 7.3.7. For all $\epsilon>0$, we have,

$$\lim_{t\to\infty}\limsup_{n\to\infty}\mathbb{P}\big(d_H\big(\mathcal{T}_n(t),\mathcal{T}_n\big)>\epsilon\big)=0.$$

As seen in Corollary 4.2.5 and Lemma 6.2.7, Theorem 7.3.7 is proven upon showing,

Lemma 7.3.8. Theorem 7.3.7 is proven upon finding $\epsilon_i(t):[0,\infty)\to[0,\infty)$ such that,

$$i)\lim_{t\to\infty}\sum_{i=0}^{\infty}\epsilon_i(t)=0\ \text{ and }\ ii)\lim_{t\to\infty}\limsup_{n\to\infty}\sum_{i=0}^{\infty}\mathbb{P}\big(d_H\big(\mathcal{T}(x_itn^{\frac{1}{2}}),\mathcal{T}(x_{i+1}tn^{\frac{1}{2}})\big)>\epsilon_in^{\frac{1}{2}}\big)=0.$$

We start with a tail bound on the distance of vertex $v\in\mathcal{T}_n(b)\backslash\mathcal{T}_n(a)$ to $\mathcal{T}_n(a)$, for some $b>a$.

Lemma 7.3.9. Let $b>a$ be positive integers. Let $l\in\{a,a+1,\dots,b\}$ and v_l the vertex corresponding to l in $\mathcal{T}_n(b)\setminus\mathcal{T}_n(a)$. Then,

$$\mathbb{P}\left(d_H\left(\mathcal{T}_n(a),v_l\right)>c\right)\le 2\left(\frac{b}{a}\right)^2\exp\left(-\frac{c\ln^{\gamma}(an^{-\frac{1}{2}})n^{-\frac{1}{2}}}{4}\right).$$

Proof. Let $U_1,U_2,\dots\sim\text{Unif}([0,1])$ i.i.d. and set $t_1=l$. Define recursively,

$$\begin{aligned} q(t_i)&=\max\{C_i^n,a:C_i^n\le t_i\}, &\qquad S_n(q(t_i))&=S_n(\lceil U_i(q(t_i)-1)\rceil),\\ d_i&=q(t_i)-t_i, & t_{i+1}&=\min\{k\in\mathbb{N}:S_n(k)=S_n(q(t_i))\}.\end{aligned}$$

Observe that $q(t_i)$ corresponds to the start of the stick on which t_i lies. Furthermore, notice that $\lceil U_i(q(t_i)-1)\rceil\in_u\{1,\dots,q(t_i)-1\}$ so that $S_n(q(t_i))$ has the correct law. Also, t_{i+1} is the index of the first appearance in S_n of the vertex to which this stick is attached to, so that the parent vertex of $S_n(q(t_i))$ is $S_n(t_{i+1})$.

Set $N=\min\{k:t_k\le a\}-1$ and observe $\sum_{i=1}^{N}d_i=d_H\big(\mathcal{T}_n(a),v_l\big)$, see Remark 7.3.10. We aim to stochastically dominate N. It deterministically holds that $t_{i+1}\le\lceil U_i(q(t_i)-1)\rceil$ with equality if $q(t_i)$ appears exactly once in $S_n(1),\dots,S_n(q(t_i))$ and strict inequality otherwise. Since $q(t_i)<t_i$ we see $t_{i+1}\le U_it_i$ is deterministically true. We get $t_{i+1}<U_i\dots U_1t_1$ is deterministically true by iteration. Hence,

$$\mathbb{P}(N>j)\le\mathbb{P}(t_{j+1}>a)\le\mathbb{P}(U_i\dots U_1b>a)=\mathbb{P}(G_{j,1}>L)\le\left(\frac{1}{1+s}\right)^je^{Ls},$$

where $L = \ln(b/a)$ and $G_{j,1} \sim \text{Gamma}(j, 1)$. The last inequality is obtained with a Chernoff bound and holds for $s > 0$. Next, for $a \leq t_i \leq b$ independent of $S_n(t_i+1), \ldots, S_n(b)$, we have,

$$\begin{aligned}\mathbb{P}\left(t_i - p(t_i) > y \mid d_1, \ldots, d_{i-1}\right) &= \mathbb{1}_{\{t_i - a > c\}} \prod_{i=1}^{y} \left(1 - \ln^{\gamma}\left(t_i n^{-\frac{1}{2}} + 1\right) n^{-\frac{1}{2}}\right)^{y}, \\ &\leq \mathbb{1}_{\{t_i - a > c\}} \exp\left(-y \ln^{\gamma}\left(t_i n^{-\frac{1}{2}} + 1\right) n^{-\frac{1}{2}}\right), \\ &\leq \exp\left(-y\lambda\right),\end{aligned}$$

where $\lambda = \ln^{\gamma}\big(a n^{\frac{1}{2}}\big) n^{-\frac{1}{2}}$. Hence $d_i \mid d_1, \ldots, d_{i-1}$ are stochastically dominated by i.i.d. $\text{Exp}(\lambda)$ random variables for. Thus, $\sum_{i=1}^{j} d_i$ is stochastically bounded by a $\text{Gamma}(j, \lambda)$ random variable. From a Chernoff bound, we obtain $\mathbb{P}\left(\sum_{i=1}^{j} d_i > c\right) \leq \left(\frac{\lambda}{\lambda - t}\right)^{j} e^{-\lambda c}$ for all $t \in (0, \lambda)$.

To bound $\mathbb{P}\left(\sum_{i=1}^{N} d_i > c\right)$, we apply a union bound,

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{N} d_i > c\right) &\leq \mathbb{P}\left(\sum_{i=1}^{j} d_i > c\right) + \mathbb{P}(N > j), \\ &\leq \left(\frac{1}{s+1}\right)^{j} e^{Ls} + \left(\frac{\lambda}{\lambda - t}\right)^{j} e^{-tc}, \\ &\leq 2\left(\frac{b}{a}\right)^{2} e^{-\frac{\lambda c}{4}}.\end{aligned}$$

where the last step follows from taking $s = e - 1, t = \lambda\frac{e-1}{e}$ and $j = c\lambda\frac{e-1}{2e}$ concluding the proof. □

Remark 7.3.10. The tree in Figure 30 corresponds to $S_n = \{8, 5, 23, 24, 22, 23, 20, 19, 18, 21, 14, 20, 22, 18, 18, 21, 16, 15, 30, 29, 25, 21, 22, 16, 16, 21, 19, 26, 30\}$. Take $a = 5, b = 30, l = 29$. Then $v_l = 27$ and $q(t_1) = 27$ as $S_n(27)$ is the repeat to the left of $S_n(29)$ and $d_1 = 29 - 27 = 2$. Also, $S_n(q(t_i)) = 19$ and 19 first appears in $S_n(8)$, so that $t_2 = 8$. Lastly, $q(t_2) = a = 5$ as there is no repeat in S_n with index smaller than 8. Thus $d_2 = 8 - 5 = 3$ and $N = 2$ as we reached a vertex in $\mathcal{T}_{30}(a)$. Also note $d_H\big(v_{29}, \mathcal{T}_{30}(a)\big) = 5 = d_1 + d_2 = \sum_{i=1}^{N} d_i$.

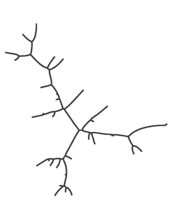
Lemma 7.3.11. For integers $1 \leq a < b$, we have,

$$\mathbb{P}\left(d\left(\mathcal{T}_n(a), \mathcal{T}_n(b)\right) > c\right) \leq 4\left(\frac{b}{a}\right)^{2} \frac{b}{c} \exp\left(-\frac{c \ln^{\gamma}\left(a n^{-\frac{1}{2}}\right) n^{-\frac{1}{2}}}{8}\right).$$

Proof. Identical to the proof of Lemma 4.2.12. □

Theorem 7.3.12. We have for all $\epsilon > 0$,

$$\lim_{t \to \infty} \limsup_{n \to \infty} \mathbb{P}(d(\mathcal{T}_n(t), \mathcal{T}_n) > \epsilon) = 0$$



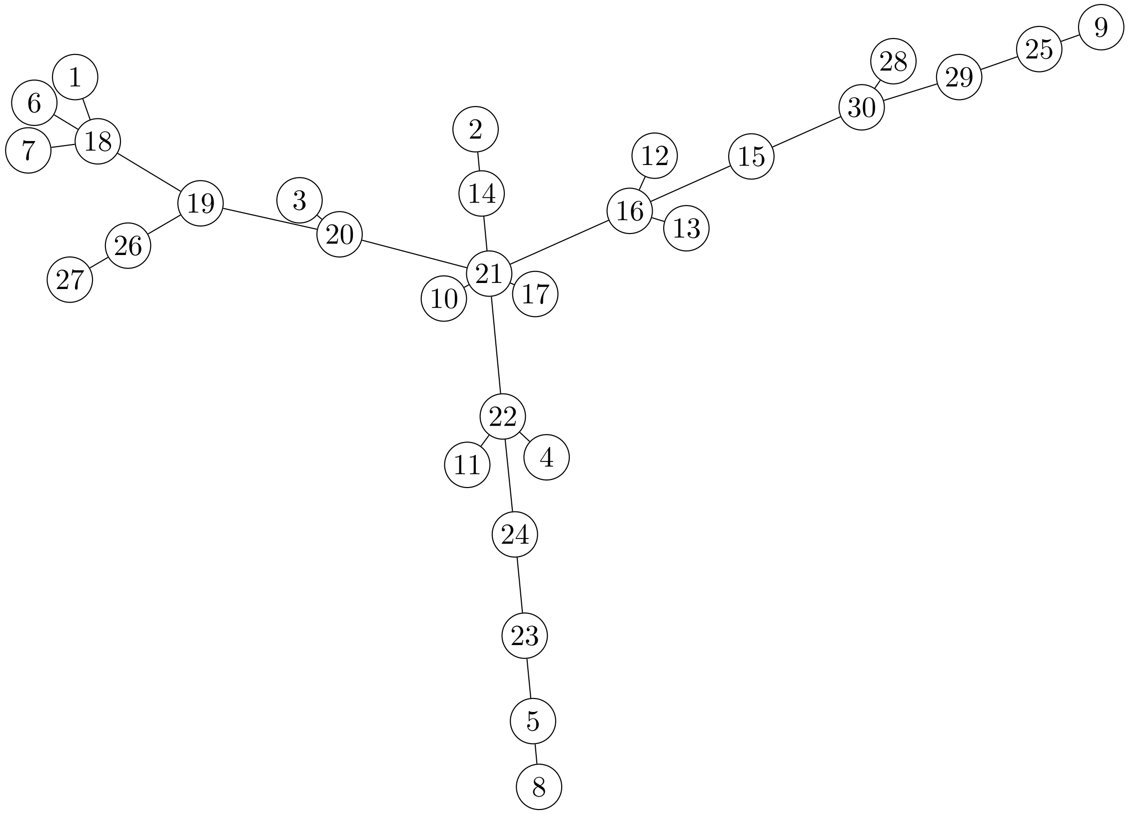


Figure 30: Tree T_{30}

Proof. By substituting $a = x_i t n^{\frac{1}{2}}, b = x_{i+1} t n^{\frac{1}{2}}$ and $c = \epsilon_i n^{\frac{1}{2}}$ into Lemma 7.3.11 and simplifying, we obtain that condition *ii*) of Lemma 7.3.8 is translated into,

$$\lim_{t\to\infty} \sum_{i=0}^{\infty} \frac{1}{\epsilon_i} \exp\left(3\ln(x_{i+1} + \ln(t) - \frac{\epsilon_i \ln^{\gamma}(x_i t)}{8}\right) = 0.$$

This condition also appeared in the proof of Theorem 7.3.1 and was shown to hold for the choice $\epsilon_i = \frac{8}{\ln^{\gamma}(x_i t)}\big(3\ln(x_{i+1}) + 2\ln(t) + (\alpha\gamma + 2)\ln(i+1)\big)$, concluding the proof. □

7.4 Tightness Measures

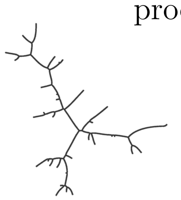
In this section, we prove *iii*) and *iv*) of Remark 7.0.5. We start with *iii*).

7.4.1 Convergence of Measure on the Continuous Tree

Theorem 7.4.1. For $\gamma > 1$ and $\epsilon > 0$, we have,

$$\lim_{k\to\infty} \mathbb{P}\left(d_P\left(\mu^{(k)}, \mu\right) > \epsilon\right)\right) = 0.$$

Proof. This follows from identical reasoning to Section 6.3 and Section 5.1. Indeed, these proofs use the following three properties of $\mathcal{T}$:



i) There exists a K such that with probability $1-\epsilon$, we have that $d_H\big(\mathcal{T}^{(K)}, \mathcal{T}\big) < \epsilon$.

ii) $\mathcal{T}$ is compact almost surely.

iii) Conditional on $\mathcal{T}^{(K)}$, for $A \subset \mathcal{T}^{(K)}$ we have $\mu^{(j)}(A^{\uparrow})$ is a martingale.

The first two statements are ensured by Theorem 7.3.1 and the second statement by the uniform attachment of the sticks. Thus, the exact same reasoning also goes through for the current tree $\mathcal{T}$ with $\gamma > 1$. □

7.4.2 Convergence of Measure on the Discrete Tree

In this section, we show,

Theorem 7.4.2. For $\gamma > 1$ and all $\epsilon > 0$, we have,

$$\lim_{k\to\infty} \mathbb{P}\big(d_P\big(\nu_n^{(k)}, \nu_n\big) > \epsilon\big) = 0.$$

We follow identical reasoning to Section 6.3.2 and to lesser extend Section 5.2. Recall the definition of $A^{\uparrow}$, see Definition 5.2.1. From identical reasoning to Lemma 5.2.2, it follows,

Lemma 7.4.3. Let $A \subset \mathcal{T}_n^{(k)}$. For $j \geq k$, $\nu_n^{(j)}(A^{\uparrow})$ is a martingale in filtration $F_j = \sigma(\mathcal{T}_n^{(j)})$.

Next, we translate 6.3.6 to the current setting.

Lemma 7.4.4. For $A \subset \mathcal{T}_n^{(k)}$, we have,

$$\mathbb{P}\left((\nu_n^{(k)}(A^{\uparrow}) - \nu_n(A^{\uparrow}))^2 \geq C^2 \mid \mathcal{T}_n^{(k)}\right) \leq \frac{X_{n,k}}{C^2} = \frac{1}{C^2}\sum_{j=k}^{N} \mathbb{E}\left[\frac{2n}{(C_j^n)^2 \ln^{2\gamma}\left(C_j^n n^{-\frac{1}{2}}+1\right)} \,\middle|\, \mathcal{T}_n^{(k)}\right],$$

where N denotes the random number of branches of $\mathcal{T}_n$. Note that $X_{n,k}$ is a random variable.

Proof. By following identical reasoning to Lemma 6.3.6, we obtain,

$$\mathbb{P}\left((\nu_n^{(k)}(A^{\uparrow}) - \nu_n(A^{\uparrow}))^2 \geq C^2 \mid \mathcal{T}_n^{(k)}\right) \leq \frac{1}{C^2}\sum_{j=k}^{N} \mathbb{E}\left[\frac{(c_{j+1}^n)^2}{(C_j^n)^2} \,\middle|\, \mathcal{T}_n^{(k)}\right],$$

with N being the random number of sticks in $\mathcal{T}_n$ and $c_j^n = C_j^n - C_{j-1}^n$. Observe that $c_{j+1}^n > x$ given $\mathcal{T}_n^{(j)}$ happens precisely when $S_n^f(C_j^n+1), \ldots, S_n^f(C_j^n+x)$ are all not repeats. This happens with probability,

$$\mathbb{P}\big(c_{j+1}^n > x \mid \mathcal{T}_n^{(j)}\big) = \prod_{i=1}^{x}\Big(1 - \ln^{\gamma}\big((C_j^n + x)n^{-\frac{1}{2}}+1\big)n^{-\frac{1}{2}}\Big) \leq \Big(1 - \ln^{\gamma}\big((C_j^n)n^{-\frac{1}{2}}+1\big)n^{-\frac{1}{2}}\Big)^{x}.$$

This gives stochastic domination of c_j^n by $Y \sim \text{Geom}\Big(\ln^{\gamma}\big((C_j^n)n^{-\frac{1}{2}}+1\big)n^{-\frac{1}{2}}\Big)$. In particular,

$$\mathbb{E}\left[\left(c_{j+1}^n\right)^2 \mid \mathcal{T}_n^{(j)}\right] \leq \mathbb{E}[Y^2] \leq \frac{2n^2}{\Big(\ln^{\gamma}\big((C_j^n)n^{-\frac{1}{2}}+1\big)n^{-\frac{1}{2}}\Big)^2} = \frac{2n}{\ln^{2\gamma}\big((C_j^n)n^{-\frac{1}{2}}+1\big)}.$$

The proof is finished by applying the tower property of expectations, identical to the proof of Lemma 6.3.6. □

We continue by showing,

Lemma 7.4.5. We have,

$$\limsup_{n\to\infty} \mathbb{E}\left[\sum_{j=k}^{N} \mathbb{E}\left[\frac{2n}{(C_j^n)^2 \ln^{2\gamma}\left(C_j^n n^{-\frac{1}{2}}+1\right)} \,\middle|\, \mathcal{T}_n^{(k)}\right]\right] \xrightarrow[k\to\infty]{} 0.$$

Proof. This proof follows similar reasoning to the proof of Lemma 6.3.7. but we use a different analysis of the final Chernoff bound. First, note that $N \le n$ as a tree with n vertices can have at most n branches. Thus,

$$\sum_{j=k}^{N} \mathbb{E}\left[\frac{2n}{(C_j^n)^2 \ln^{2\gamma}\left(C_j^n n^{-\frac{1}{2}}+1\right)} \,\middle|\, \mathcal{T}_n^{(k)}\right] \le \sum_{j=k}^{n} \mathbb{E}\left[\frac{2n}{(C_j^n)^2 \ln^{2\gamma}\left(C_j^n n^{-\frac{1}{2}}+1\right)} \,\middle|\, \mathcal{T}_n^{(k)}\right],$$

where we set $C_j^n = n$ for $j > N$. Next, observe that,

$$\mathbb{P}(C_j^n < x) \le \mathbb{P}(X \ge j), \text{ where } X \sim \mathrm{Binom}\big(x, \ln^{\gamma}\left(xn^{-\frac{1}{2}}+1\right)n^{-\frac{1}{2}}\big).$$

Indeed, $C_j^n < x$ can only happen if $S_n^f(1), \ldots, S_n^f(x)$ contains at least j repeats. The bound follows since $\mathbb{P}(S_n^f(i) \text{ is a repeat}) \le \ln^{\gamma}\left(xn^{-\frac{1}{2}}+1\right)n^{-\frac{1}{2}}$. By using a Chernoff bound and following the computations in the proof of Lemma 6.3.7, we obtain,

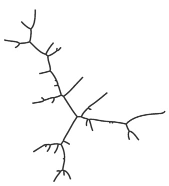
$$\mathbb{P}(C_j^n < x) \le \left(\frac{ex\ln^{\gamma}\left(xn^{-\frac{1}{2}}+1\right)}{jn^{\frac{1}{2}}}\right)^{j}.$$

Remark 7.4.6. It is heuristically clear that $\lim_{n\to\infty}\sum_{j=k}^{\infty}\mathbb{E}\left[\frac{n}{(C_j^n)^2\ln^{2\gamma}\left(C_j^n n^{-\frac{1}{2}}+1\right)}\right] \xrightarrow[k\to\infty]{} 0$. Indeed $\mathbb{E}[n^{-\frac{1}{2}}C_i^n] \xrightarrow[n\to\infty]{} \mathbb{E}[C_i] \approx \frac{i}{\ln(i)^{\gamma}}$. Thus, we may expect that

$$\mathbb{E}\left[\frac{n}{(C_j^n)^2 \ln^{2\gamma}\left(C_j^n n^{-\frac{1}{2}}+1\right)}\right] \approx \frac{\ln(j)^{2\gamma}}{j^2 \ln\left(\frac{j}{\ln(j)^{\gamma}}\right)^{2\gamma}} \approx \frac{1}{j^2}.$$

To show this formally, we define $g(x) = x^{-2}\ln^{-2\gamma}(x+1)$. We write the expectation in terms of the tail distribution,

$$\begin{aligned}
\mathbb{E}\left[\frac{n}{(C_j^n)^2 \ln^{2\gamma}\left(C_j^n n^{-\frac{1}{2}}+1\right)}\right] &= \mathbb{E}\left[g(n^{-\frac{1}{2}}C_j^n)\right],\\
&= \sum_{k=1}^{\infty}\Big(g\big((k+1)n^{-\frac{1}{2}}\big) - g\big(kn^{-\frac{1}{2}}\big)\Big)\mathbb{P}\big(g(n^{-1/2}C_j^n) > g(n^{-1/2}k)\big),\\
&\le \sum_{k\le h(j)n^{\frac{1}{2}}}\Big(g\big((k+1)n^{-\frac{1}{2}}\big) - g\big(kn^{-\frac{1}{2}}\big)\Big)\,\mathbb{P}(C_j^n < k) + \sum_{k>h(j)n^{\frac{1}{2}}} g\big((k+1)n^{-\frac{1}{2}}\big) - g\big(kn^{-\frac{1}{2}}\big),\\
&\le \left(\frac{e}{j}\right)^{j}\sum_{k\le h(j)n^{\frac{1}{2}}} -g'(x_k)n^{-\frac{1}{2}}\ln\left(kn^{-\frac{1}{2}}+1\right)^{j\gamma}(kn^{-\frac{1}{2}})^{j} + \sum_{k>h(j)n^{\frac{1}{2}}} g\big((k+1)n^{-\frac{1}{2}}\big) - g\big(kn^{-\frac{1}{2}}\big),
\end{aligned}$$



for some cutoff point $h(j)$ to be determined. Here $g'(x_k)$ comes from the mean value theorem and $x_k \in (kn^{-\frac{1}{2}}, (k+1)n^{-\frac{1}{2}})$. For ease of notation, define,

$$S_1 = \sum_{k \le h(j)n^{\frac{1}{2}}} g'(x_k)n^{-\frac{1}{2}} \ln\left(kn^{-\frac{1}{2}}+1\right)^{j\gamma}(kn^{-\frac{1}{2}})^j, \qquad S_2 = \sum_{k > h(j)n^{\frac{1}{2}}} g(kn^{-\frac{1}{2}}) - g((k+1)n^{-\frac{1}{2}}).$$

We work on S_1 first. Note $|kn^{-\frac{1}{2}} - x_k| \le n^{-\frac{1}{2}}$. Thus for all $k \in \{1, 2, \ldots, \lfloor h(j)n^{\frac{1}{2}} \rfloor\}$ we see,

$$|\ln(kn^{-\frac{1}{2}}+1)^{j\gamma}(kn^{-\frac{1}{2}})^j - \ln(x_k+1)^{j\gamma}x_k^j| \le n^{-\frac{1}{2}} \max_{x\in[0,h(j)]} \frac{d}{dx}\ln(x+1)^{j\gamma}x^j \le Cn^{-\frac{1}{2}},$$

for C some constant depending on j and γ but independent of n. Thus, we may write,

$$\sum_{k \le h(j)n^{\frac{1}{2}}} g'(x_k)n^{-\frac{1}{2}} \ln\left(kn^{-\frac{1}{2}}+1\right)^{j\gamma}(kn^{-\frac{1}{2}})^j = \sum_{k \le h(j)n^{\frac{1}{2}}} g'(x_k)n^{-\frac{1}{2}}\Big(\ln\left(x_k+1\right)^{j\gamma}(x_k)^j + E_k\Big),$$

where the error term E_k is $O(n^{-\frac{1}{2}})$. Note $g'(x)\ln(x+1)^{j\gamma}x^j$ is continuous on $[0,j]$ for $j \ge 3$ since $g'(x)$ has a pole of order $2\gamma+2$ at 0 which is removed by a zero of order $j + j\gamma$ from the term $x^j \ln(x+1)^{j\gamma}$. Thus we may recognize S_1 as a Riemann sum to obtain,

$$S_1 \le \sum_{k \le h(j)n^{\frac{1}{2}}} g'(x_k)n^{-\frac{1}{2}}\Big(\ln\left(x_k+1\right)^{j\gamma}(x_k)^j + E_k\Big) \xrightarrow[n\to\infty]{} \int_0^{h(j)} g'(x)\ln(x+1)^{j\gamma}x^j\,dx,$$
$$\le \ln(h(j)+1)^{j\gamma}h(j)^j g(h(j)).$$

Using this bound, we find,

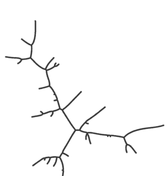
$$\limsup_{n\to\infty}\left(\frac{e}{j}\right)^j \sum_{k \le h(j)n^{\frac{1}{2}}} g'(x_k)n^{-\frac{1}{2}} \ln\left(kn^{-\frac{1}{2}}+1\right)^{j\gamma}(kn^{-\frac{1}{2}})^j \le \frac{e^2}{j^2}\left(\frac{e\ln(h(j)+1)^{\gamma}h(j)}{j}\right)^{j-2}.$$

Thus, $\limsup_{n\to\infty} S_1$ is $O\left(\frac{1}{j^2}\right)$ precisely when $e\ln(h(j)+1)^{\gamma}h(j) \le j$. We note $h(j) = \frac{j}{e\ln(j)^{\gamma}}$ works. With $h(j)$ defined, we bound S_2. By the telescoping nature of S_2, we obtain,

$$\sum_{k > h(j)n^{\frac{1}{2}}} g(kn^{-\frac{1}{2}}) - g((k+1)n^{-\frac{1}{2}}) \le g(\lceil h(j)n^{\frac{1}{2}}\rceil n^{-\frac{1}{2}}) - \lim_{k\to\infty} g(kn^{-\frac{1}{2}}) \xrightarrow[n\to\infty]{} g(h(j)),$$
$$= \frac{e^2\ln(j)^{2\gamma}}{j^2\ln\left(\frac{j}{e\ln(j)^{\gamma}}+1\right)^{2\gamma}}$$

By combining the bounds on S_1 and S_2, we obtain,

$$\limsup_{n\to\infty}\mathbb{E}\left[\frac{1}{(C_j^n n^{-\frac{1}{2}})^2\ln(C_j^n n^{-\frac{1}{2}}+1)^{2\gamma}}\right] \le \frac{e^2\ln(j)^{2\gamma}}{j^2\ln\left(\frac{j}{e\ln(j)^{\gamma}}+1\right)^{2\gamma}} + \frac{e^2}{j^2} = O\Big(\frac{1}{j^2}\Big).$$



Hence we have,

$$\limsup_{n\to\infty} \mathbb{E}\left[\sum_{j=k}^{N} \mathbb{E}\left[\frac{2n}{(C_j^n)^2 \ln^{2\gamma}\left(C_j^n n^{-\frac{1}{2}}+1\right)} \;\middle|\; \mathcal{T}_n^{(k)}\right]\right] \leq \limsup_{n\to\infty} \sum_{j=k}^{n} \mathbb{E}\left[\frac{2n}{(C_j^n)^2 \ln^{2\gamma}\left(C_j^n n^{-\frac{1}{2}}+1\right)}\right],$$
$$\leq C\sum_{j=k}^{\infty}\frac{1}{j^2} \xrightarrow[k\to\infty]{} 0,$$

as desired. □

We are now in position to prove Theorem 7.4.2.

Proof of Theorem 7.4.2. This follows from the exact same steps as the proof of Theorem 5.2.5 with Lemma 7.4.4 and Lemma 7.4.5 taking the position of Lemma 5.2.4. □

7.5 Comparison to Existing Literature

In [6], Nicolas Curien and Bénédicte Haas constructed random tree $\mathcal{T} \subset \ell^1$ by considering sticks of fixed length $a_1, a_2, \dots > 0$ and constructed $\mathcal{T}^{(k)}$ ($\mathcal{T}_n$ in their notation) inductively: set $\mathcal{T}^{(1)} = a_1$ and let $\mathcal{T}^{(k)} \subset \ell^1$ denote the random tree obtained from gluing stick a_k to a uniform point on $\mathcal{T}^{(k-1)}$. As with the stick-breaking construction, each branch is attached in the direction of a new basis vector of ℓ^1 and $\mathcal{T}$ denotes the closure of $\bigcup_{k=1}^{\infty} \mathcal{T}^{(k)}$. In other words, $\mathcal{T}$ is built using the stick-breaking construction, but the lengths of the sticks are deterministic. The only randomness comes from uniformly attaching new branches to the already existing tree.

In [6, Theorem 1], they showed,

Theorem 7.5.1. Suppose there exists $\alpha \in (0, 1]$ such that,

$$\alpha_i \leq i^{-\alpha+o(1)} \qquad \text{and} \qquad A_i = i^{1-\alpha+o(1)},$$

where $A_i = a_1 + \cdots + a_i$. Then $\mathcal{T}$ is almost surely compact.

Throughout this thesis, we work with random stick lengths and thus cannot apply this theorem directly. However, let η be a PPP on $\mathbb{R}_{\geq 0}$ of intensity $f(t)dt$ and let $N(t) = \eta([0, t])$ be the number of points in η with value less than t. Then, the strong law for PPP's states,

$$\frac{N(t)}{\Lambda(t)} \xrightarrow[t\to\infty]{} 1,$$

almost surely. Here $\Lambda(t) = \int_0^t f(s)ds$. In particular, this implies that for large i the i'th point of η happens at roughly $\Lambda^{-1}(i)$. That is, if $0 < C_1 < \dots$ are the ordered points of a PPP of intensity $f(t)dt$, we see $C_i = \Lambda^{-1}(i)(1 + o(1))$ almost surely. In our case, $f(x) = \ln^{\gamma}(x + 1)$ and one can check $\Lambda(t) = \Theta(t \ln^{\gamma}(t))$ so that $\Lambda^{-1}(i) = \Theta\big(\frac{i}{\ln^{\gamma}(i)}\big)$. In particular, if we want to write $C_i = i^{1-\alpha+o(1)}$ almost surely, we must choose $\alpha = 0$, and hence our tree $\mathcal{T}$ constructed from a PPP of intenstiy $\ln^{\gamma}(t + 1)dt$ is not covered by [6, Theorem 1].

In [6] they also considered $\mu^{(k)}$, the uniform measure on $\mathcal{T}$ (in the paper denoted μ_n). They showed that when $\sum_{i=1}^{\infty} \frac{a_i^2}{A_i} < \infty$, then μ_n converges to some measure μ on $\mathcal{T}$. In particular, a uniform point on $\mathcal{T}$ must hence be bounded. However, this is only a necessary condition for $\mathcal{T}$ to be bounded, not sufficient.

We already saw that $C_i = \Theta\big(\frac{i}{\ln^{\gamma}(i)}\big)$ almost surely. Similar reasoning shows that $a_i = \Theta\big(\frac{1}{\ln^{\gamma}(i)}\big)$ almost surely, and thus $\frac{a_i^2}{A_i} = \Theta\big(\frac{1}{i \ln^{\gamma}(i)}\big)$ almost surely. In particular, this gives that

$$\sum_{i=1}^{\infty} \frac{a_i^2}{A_i} < \infty \iff \gamma > 1.$$

This confirms that $\mathcal{T}$ constructed from a PPP of intensity $\ln^{\gamma}(t+1)dt$ cannot be compact for $\gamma \leq 1$. Section 7 partially answers the question at the bottom of [6, page 7], the maximal height of $\mathcal{T}$ constructed from $a_i \sim \ln^{-\gamma}(i)$ with $\gamma > 1$ stays stochastically bounded when the a_i's are sampled through a PPP of intensity $\ln^{\gamma}(t+1)dt$.

It remains to be checked if the computations in Section 7 can be generalized to show a variant of Theorem 7.5.1 that encompasses results when a_i are of the order $\frac{1}{\ln^{\gamma}(i)}$ for $\gamma > 1$. Another natural generalization includes answering if any tree constructed from a PPP with intensity $f(t)dt$ with $f = \omega(\ln(t+1))$ is compact. Furthermore, the Hausdorff dimensions and local limits of all trees covered in this thesis are yet to be considered.

8 Bibliography

[1] Louigi Addario-Berry, Arthur Blanc-Renaudie, Serte Donderwinkel, Mickaël Maazoun, and James B. Martin. The Foata–Fuchs proof of Cayley's formula, and its probabilistic uses. *Electronic Communications in Probability*, 28:1 – 13, 2023.

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

[3] Eleanor Archer and Matan Shalev. Random choice spanning trees. preprint arXiv:2402.05800, 2024.

[4] Patrick Billingsley. *Probability and measure*. Wiley, 3. edition, 1995.

[5] Dmitri Burago, Yurij D. Burago, and Sergei Ivanov. *A Course in Metric Geometry*. American Mathematical Society, 2001.

[6] Nicolas Curien and Bénédicte Haas. Random Trees Constructed by Aggregation. *Annales de l'Institute Fourier*, 2016.

[7] Daryl J. Daley and David Vere-Jones. *An introduction to the theory of point processes. Vol. I*. Springer-Verlag, 2. edition, 2003.

[8] Richard Durrett. *Probability: theory and examples*. Duxbury Press, 2. edition, 1996.

[9] Svante Janson. On the Gromov–Prohorov distance, 2020. arXiv:2005.13505.

[10] Ali Khezeli. Metrization of the Gromov–Hausdorff (-Prokhorov) topology for boundedly-compact metric spaces. *Stochastic Processes and their Applications*, 130(6):3842–3864, 2020.

[11] Takis Konstantopoulos. Review: An Introduction to Measure Theory. *The American Mathematical Monthly*, 120(8):762–768, 2013.

[12] Günter Last and Mathew Penrose. *Lectures on the Poisson Process*. Cambridge University Press, 2017.

[13] Sheldon M. Ross. *Introduction to Probability Models*. Academic Pres, 10. edition, 2009.

[14] Benedikt Stufler. Mass and radius of balls in Gromov-Hausdorff-Prokhorov convergent sequences, 2022. arXiv: 2201.12251.

[15] José Velhinho. Topics of Measure Theory on Infinite Dimensional Spaces. *Mathematics*, 5(3):44, 2017.