UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Pós-graduação em Matemática

Mariana Pereira Lopes

An introduction to convergence of random trees.

Belo Horizonte
2025

Mariana Pereira Lopes

An introduction to convergence of random trees.

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Matemática.

Orientador: Renato Soares dos Santos

Belo Horizonte
2025

Lopes, Mariana Pereira.

L864i An introduction to convergence of random trees [recurso eletrônico] / Mariana Pereira Lopes. Belo Horizonte — 2025.
1 recurso online (62 f. il.): pdf.

Orientador: Renato Soares dos Santos.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Matemática.
Referências: f. 53-54

1. Matemática – Teses. 2. Árvores (Teoria dos grafos) – Teses. 3. Hausdorff, Medidas de – Teses. 4. Processo de Galton-Watson – Teses. 5. Convergência de Gromov-Hausdorff – Teses. I. Santos, Renato Soares dos. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Matemática. III. Título.

CDU 51(043)

Ficha catalográfica elaborada pela bibliotecária Irénquer Vismeg Lucas Cruz - CRB 6ª Região - nº 819.

Universidade Federal de Minas Gerais
Departamento de Matemática
Programa de Pós-Graduação em Matemática

FOLHA DE APROVAÇÃO

An introduction to convergence of random trees

MARIANA PEREIRA LOPES

Dissertação defendida e aprovada pela banca examinadora constituída por:

Prof. Renato Soares dos Santos
Orientador - UFMG

Prof. Bernardo Nunes Borges de Lima
UFMG

Profa. Eleanor Archer
Universidade de Paris-Dauphine

Belo Horizonte, 11 de agosto de 2025.

Av. Antônio Carlos, 6627 – Campus Pampulha - Caixa Postal: 702
CEP-31270-901 - Belo Horizonte – Minas Gerais - Fone (31) 3409-5963
e-mail: pgmat@mat.ufmg.br - home page: http://www.mat.ufmg.br/pgmat

Ao caminho que percorri e todos
que participaram dessa jornada.

Agradecimentos

À minha família, por sempre terem sido os primeiros a me apoiarem e acreditarem em mim, em especial à minha mãe, Ivone, que sempre deu seu máximo para tornar tudo possível, à meus irmãos Marina e Anderson, por sempre estarem lá quando precisei de apoio.

Ao meu orientador, Renato, pela paciência e por sempre estar disposto a me ajudar em minhas (várias) indecisões acadêmicas e me orientar durante praticamente toda minha trajetória na UFMG.

Aos professores e professoras do departamento de matemática, que fizeram todo meu aprendizado possível me apresentando às belezas da matemática. Em especial aos professores da área de probabilidade, que me inspiraram e me guiaram no processo de conhecer e escolher seguir esta área.

Aos meus amigos e colegas do departamento, que tornaram tudo mais leve e divertido em todos os momentos da minha trajetória acadêmica. Em especial a meu namorado Carlos, por me incentivar e me ajudar a seguir em frente nos momentos necessários.

Às secretárias do departamento de matemática, Andréa e Kelli, que me ajudaram a completar cada etapa do mestrado de maneira apropriada.

E à CAPES pelo suporte financeiro durante o mestrado.

Resumo

Neste trabalho visamos introduzir algumas noções de convergência de árvores aleatórias por meio de teoria e exemplos em três perspectivas. A primeira é a convergência de funções contorno e altura de árvores relacionadas a árvores de Galton-Watson críticas não degeneradas com variância finita. A segunda é a introdução ao conjunto dos espaços métricos conhecidos como árvores reais, à codificação destes elementos por excursões contínuas e à distância de Gromov-Hausdorff entre espaços métricos. Por fim, quando munimos as árvores reais de uma medida Boreliana, nós introduzimos a convergência Gromov-Hausdorff vaga de espaços métricos Heine- Borel.

Palavras-chave: árvores reais; árvores de Galton-Watson; convergência de árvores aleatórias; convergência Gromov-Hausdorff vaga; função altura; função contorno; métrica Gromov-Hausdorff.

Abstract

In this work we aim to introduce some notions of convergence of random trees by introducing theory and giving examples in three different perspectives. The first one is the weak convergence in path space of contour and height functions of trees related to a critical non-degenerate Galton-Watson trees with finite variance. The second is by introducing the set of metric spaces known as real trees, the coding of real trees by continuous excursions and the Gromov-Hausdorff distance between metric spaces. Finally, by equipping the real trees with a Borelian measure, we introduce the Gromov-Hausdorff vague convergence of Heine-Borel metric spaces.

Keywords: contour function; convergence of random trees; Galton-Watson trees; Gromov-Hausdorff metric; Gromov-Hausdorff vague convergence; height function; real trees.

Contents

Introduction

In this work, we aim to introduce a notion of convergence of random trees. In order to do so, we proceed by presenting central ideas that permeate the theory behind classical examples in this field.

In our first chapter, following the ideas of [19], we focus on rooted ordered discrete trees and some methods of coding them as simpler objects: graphs, càdlàg and continuous functions, or sequences of numbers in $\mathbb{Z}$. After that we recall the definition and useful properties of Galton-Watson trees. Finally we begin to discuss convergence of contour functions of Galton-Watson trees in specific examples with $\mathrm{Geom}(\frac{1}{2})$ offspring distribution, presenting a detailed proof based on [19], which follows itself ideas presented in [3]. Then following [19] and [20], the generalized result regarding the rescaled convergence of the height function of any critical non-degenerate Galton-Watson trees with finite variance is stated and proved.

In the second chapter, still following [19], we present the definition of the metric spaces known as real trees, a generalization of the notion of a discrete tree. By coding real trees as continuous excursions and equipping the space of compact real trees with the Gromov-Hausdorff metric, we show the relation between convergence of continuous functions and convergence of compact metric spaces. As an illustration of this theory, we introduce the Continuum Random Tree (CRT) as the weak rescaled limit of a sequence of discrete trees uniformly chosen in the set of trees with n vertices. This real tree, first introduced in [1] as the scaling limit of a uniformly chosen labelled tree with n vertices embedded in l^1, is the most classical example of real tree, as it is the scaling limit of several sequences $(t_n)_{n\in\mathbb{N}}$ of trees such that t_n is randomly chosen in the set of trees with n vertices.

Then in section 2.3, now based on the work [6], we change our focus by equipping rooted real trees with a Borelian measures, what we call a measure metric tree. In the set of Heine-Borel metric measured spaces, we introduce the notion of Gromov-Hausdorff vague convergence, a combination between the already presented Gromov-Hausdorff metric, and vague convergence of measures. As before, we present a relation between the Gromov-Hausdorff vague convergence of measured metric trees and the convergence in path space of the continuous excursions that codes them.

Finally, as an application of this theory, we show the Gromov-Hausdorff vague convergence of the rescaled Kallenberg-Kesten tree with offspring distribution $Geom(\frac{1}{2})$ (i.e., the corresponding Galton-Watson tree conditioned on survival). The limit in this example is equal to the self-similar continuum random tree introduced by Aldous in [1] as the limit of a sequence of uniformly chosen labelled trees with n vertices and distance rescaled by $n^{-\alpha}$, for $\alpha \in (0, \frac{1}{2})$, when embedded in l^1. The importance of this example is related to the work of Kesten in [16], where it is stated that the rescaled height function of the nearest-neighbor random walk on the Kallenberg-Kesten tree converges weakly in path space (under the annealed law) to a non zero function.

Further works related to Kesten's theorem are mentioned in subsection 2.4. There we discuss applications of the theory presented in the past chapters regarding convergence of random walks on random trees and, scaling limits of other types of random trees.

Chapter 1

Discrete trees

1.1 Discrete trees

Consider the following set

$$\mathcal{U} = \cup_{n=0}^{\infty} \mathbb{N}^n$$

with the conventions that $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{N}^0 = \{\emptyset\}$. If $u \in \mathcal{U}$ is such that $u = (u_1, u_2, \ldots, u_n)$, for $u_i \in \mathbb{N}$, $1 \leq i \leq n$, we set $|u| = n$ the *generation* of u.

We will put a kinship relation between elements of $\mathcal{U}$ using the function $\pi : \mathcal{U} \setminus \{\emptyset\} \to \mathcal{U}$ defined by $\pi(u_1, u_2, \ldots, u_n) = (u_1, u_2, \ldots, u_{n-1})$, and saying that $\pi(u)$ is the *father* of u. We denote by π^k the k th iteration of π.

Definition 1.1. A *finite rooted ordered tree* t is a finite subset of $\mathcal{U}$ such that:

(i) $\emptyset \in t$.

(ii) $u \in t \setminus \{\emptyset\} \Rightarrow \pi(u) \in t$.

(iii) $\forall u \in t \; \exists \; k_u(t) \geq 0$ such that $\forall j \in \mathbb{N}, (u, j) \in t \Leftrightarrow 1 \leq j \leq k_u(t)$.

Let us denote by $\mathbb{A}$ the set of all finite rooted ordered trees.

Similarly we can define a infinite rooted ordered tree as a infinite countable subset of $\mathcal{U}$ that satisfies the same properties. In this chapter we will consider only the finite case.

Remark 1.2. In general, we interpret each vertex of the tree t as a member of a family whose family tree is given by t. In this situation we can define $\prec$ a genealogical order on the tree, so, given $u, v \in t$, $u \prec v$ if v is a descendant of u.

We also interpret the number $k_u(t)$ as the number of children of u in t.

From now on we will denote by $\#t$ the total number of vertices in a tree and by $u_0 = \emptyset, u_1, \ldots, u_{\#t-1}$ the elements of t listed in lexicographical order.

We also will use the notation $u \leq v$ for the lexicographical order in $\mathcal{U}$. Note that this defines a total order on the tree, while the genealogical order defines only a partial order, but, for $u, v \in t$, $u \prec v$ implies that $u \leq v$. For example in the tree $t = \{\emptyset, 1, (1,1), 2\}$ we have that $\emptyset \leq 1 \leq (1,1) \leq 2$ in the lexicographical order but in the genealogical order we can only state the relations $\emptyset \prec 1 \prec (1,1)$ and $\emptyset \prec 2$.

Discrete trees can be represented graphically in several different ways. Here we will highlight some of them.

The first one is the usual representation of a tree as a graph with the vertices being the elements of t, and edges connecting each vertex $u \in t$ with vertex of the form (u, j) where $j \in \mathbb{N}$.

Now we will code a discrete tree by functions on its elements.

Definition 1.3. For t a discrete tree and $0 \leq n \leq \#t - 1$ define the *height function* of t at n as $h_t(n) = |u_n|$. That means that the height function is the sequence of the generations of the individuals $(u_n)_{n=0}^{\#t-1}$ listed in lexicographical order.

Definition 1.4. Let us construct the *contour function* C_s of a tree t. First, we embed the graph representation of the tree in the plane in such a way that all the edges have length one, so we imagine a particle that begins in the root at time $s = 0$ and then explores the tree walking continuously on the edges with unit speed in such a way that it visits new vertices according to the lexicographical order, following the clockwise direction.

In this exploration we cross each edge exactly two times, then the *total time of exploration* is $\zeta(t) = 2(\#t - 1)$.

The *contour function* C_s, for $s \in [0, \zeta(t)]$, is given by the graph distance on the tree between the position of the particle at time s and the root.

In figure 1.1 we illustrate the constructions given above.

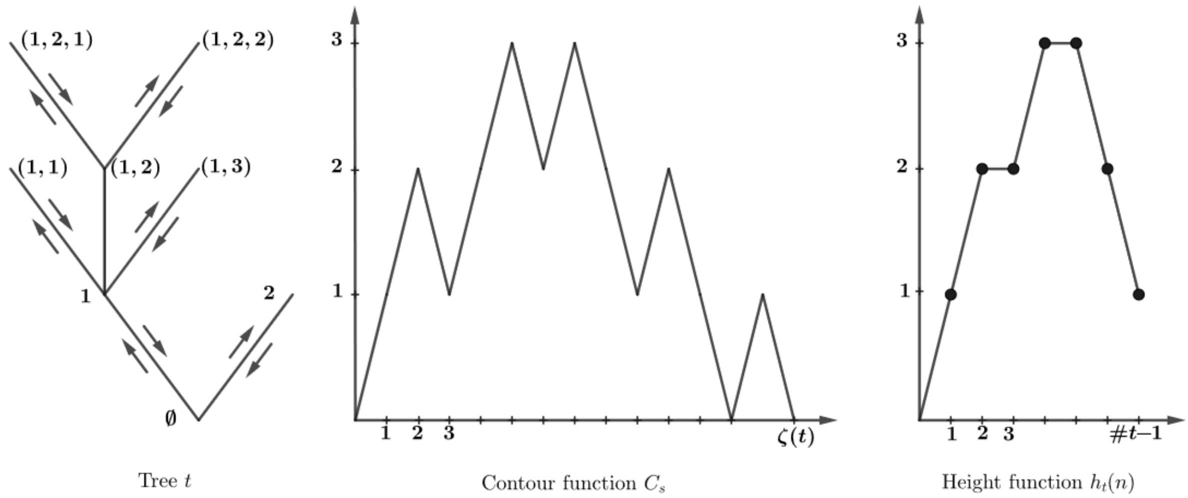

Figure 1.1: Graphical representations of a rooted ordered tree.

Note that in the height function we make the same path as the one of the contour function but we ignore the vertices that had already been explored before in the particle path, and the time spend by walking between them.

As said in remark 1.2 we can see trees as family trees. One way to formalize that is given below.

Denote by $\mathcal{S}$ the set of all finite sequences of nonnegative integers $m_1, \ldots, m_p$, with $p \geq 0$, such that

- $m_1 + m_2 + \ldots + m_i \geq i$, $\forall i \in \{1, \ldots, p-1\}$
- $m_1 + m_2 + \ldots + m_p = p - 1$.

We now prove that there is indeed an identification between $\mathcal{S}$ and $\mathbb{A}$

Proposition 1.5. The mapping

$$\Phi : t \to (k_{u_0}(t), k_{u_1}(t), \ldots, k_{u_{\#t-1}}(t))$$

defines a bijection from $\mathbb{A}$ onto $\mathcal{S}$.

Proof: Indeed, if $p = \#t$, it follows that

- $k_{u_0}(t) + k_{u_1}(t) + \ldots + k_{u_i}(t) \geq i + 1$, $\forall i \in \{0, \ldots, p-1\}$, since this sum counts all the children of $u_0, \ldots, u_i$, which contains at least the elements $u_1, \ldots, u_i, u_{i+1}$.

- $k_{u_0}(t) + k_{u_1}(t) + \ldots + k_{u_{\#t-1}}(t) = \#t - 1 = p - 1$, because it counts all elements of the tree except by the root.

so Φ is a well-defined function. It is injective by item (iii) on the definition of a finite rooted ordered tree using the lexicographical order.

To prove surjectivity, consider $(m_1, \ldots, m_p) \in \mathcal{S}$. We construct $t \in \mathbb{A}$ such that $\Phi(t) = (m_1, \ldots, m_p)$ following the following algorithm: consider $(t^{(n)})_{n=0}^{p}$ a sequence of trees indexed by time $n \in \mathbb{N}_0$ that will be modified by adding new elements at each time. Begin with $t^{(0)} = \{\emptyset\}$, we define $t^{(1)}$ by adding $1, 2, \ldots, m_1$ to t^0 (the children of the root, u_0, in lexicographical order). Next we define $t^{(2)}$ as $t^{(1)}$ plus the elements $(1,1), (1,2), \ldots, (1, m_2)$ as sons of 1 (u_1), with the convention that if $m_2 = 0$, then $t^{(2)} = t^{(1)}$. We keep defining $t^{(n)}$ inductively by adding the elements $(u_{n-1}, 1), \ldots, (u_{n-1}, m_n)$ to the tree $t^{(n-1)}$ until we reach the time $n = p$, in which $\#t^{(p)} = 1 + m_1 + \ldots + m_p = p$.

By construction it easily follows that $t^{(p)}$ is a finite rooted ordered tree and $\Phi(t^{(p)}) = (m_1, \ldots, m_p)$. Therefore the proposition is proved. ■

For our goals it will be useful to consider another way to code trees by finite sequences.

Definition 1.6. Let t be a rooted ordered tree. A *Lukasiewicz path* is a sequence of integers with indices $0 \leq n \leq \#t$ which satisfies the properties

- $x_n = \sum_{i=1}^{n} (k_{u_{i-1}}(t) - 1)$.

- $x_0 = 0$ and $x_{\#t} = -1$.

- $x_n \geq 0$ for every $0 \leq n \leq \#t - 1$.

- $x_i - x_{i-1} \geq -1$ for every $1 \leq i \leq \#t$.

The mapping Φ of the previous proposition induces a bijection between $\mathbb{A}$ and Lukasiewicz paths.

Now we introduce a relation between the Lukasiewicz path and the height function of a tree.

Proposition 1.7. The height function h_t and the Lukasiewicz path of a tree t are related by the formula

$$h_t(n) = \#\{j \in \{0, 1, \ldots, n-1\} : x_j = \inf_{j \leq k \leq n} x_k\}$$

for every $n \in \{0, \ldots, \#t - 1\}$.

Proof: By definition

$$h_t(n) = \#\{j \in \{0, 1, \ldots, n-1\} : u_j \prec u_n\}.$$

We will prove that $u_j \prec u_n$ iff $x_j = \inf_{j \leq k \leq n} x_k$.

We verify that

$$\inf\{k \geq j : x_k < x_j\} = \begin{cases} \#t & \text{, if for all } k > j, u_j \prec u_k \\ \inf\{k > j : u_j \nprec u_k\} & \text{, otherwise} \end{cases}$$

In fact, writing

$$S_k^j = x_k - x_j = \sum_{i=j}^{k-1} (k_{u_i}(t) - 1)$$

setting $S_j^j = 0$ by convention, $S_k^j = S_{k-1}^j + (k_{u_{k-1}}(t) - 1)$, that is, for each k, S_k^j counts the "profit" of descendants left by the vertices $u_j, u_{j+1}, \ldots, u_{k-1}$, as illustrated by the example below: so we have $S_k^j \geq 0$ for every $k > j$ such that $u_j \prec u_k$, and $S_m^j = -1$ if m is the first $k \geq j$ such that u_k is not a descendant of j. ■

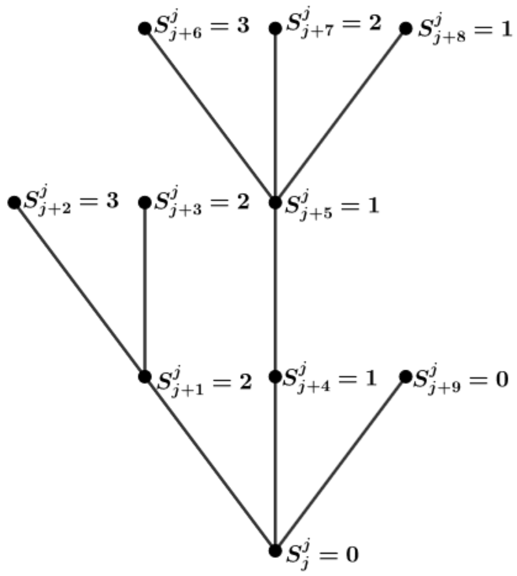

Figure 1.2: Example of the calculation of S_k^j.

1.2 Galton-Watson trees.

Let $(X_{n.j})_{n,j \in \mathbb{N}}$ be a set of independent and identically distributed random variables with distribution μ that take values on $\mathbb{N}_0$.

A *Galton-Watson process* is a stochastic process $(Z_n)_{n \in \mathbb{N}}$ defined by

$$Z_0 = 1 \quad \text{and} \quad Z_n = \sum_{j=1}^{Z_{n-1}} X_{n,j}, \text{ for all } n \in \mathbb{N}.$$

A classical interpretation of this model is as a population growth: each individual gives birth to a random i.i.d. number of children of the next generation and dies. We then see Z_n as the total number of elements in the n-th generation.

Now we present a classical result regarding this model.

Theorem 1.8. Let $\lambda = \sum_{k\in\mathbb{N}_0} k\mu(k) = \mathbb{E}_\mu(X_{1,1})$ and let us consider $\mathbb{P}_\mu(\lim_{n\to\infty} Z_n = 0)$ the probability of extinction of the process. If μ is non-degenerate we have:

- If $\lambda \leq 1$, then $\mathbb{P}_\mu(\lim_{n\to\infty} Z_n = 0) = 1$.
- If $\lambda > 1$, then $\mathbb{P}_\mu(\lim_{n\to\infty} Z_n = 0) < 1$.

That means that the population only has a positive probability of survival if $\lambda > 1$ or $\mu(k) = 1$, for some $k \geq 1$.

This theorem can be proved by arguments related to generating functions, see [14, Theorem 5.1], or with martingales theory, see [7, Example 12.29].

From now on we will refer to the distribution μ as

- Subcritical, if $\sum_{k\in\mathbb{N}_0} k\mu(k) < 1$.
- Critical, if $\sum_{k\in\mathbb{N}_0} k\mu(k) = 1$.
- Supercritical, if $\sum_{k\in\mathbb{N}_0} k\mu(k) > 1$.

Another classical result related to Galton-Watson trees is given below.

Theorem 1.9. Consider μ a critical offspring distribution such that $0 < \sigma^2 < \infty$, where $\sigma^2 = \sum_{k\in\mathbb{N}_0} k^2\mu(k) - (\sum_{k\in\mathbb{N}_0} k\mu(k))^2$. We introduce the notations

$$f(s) = \sum_{k\geq 0} s^k\mu(k), \text{ for } s \in \mathbb{C}, |s| \leq 1$$

and f_n, the n-th iteration of f.

Then $1 - f_n(0) = P(Z_n > 0)$, that is, the probability of the μ-Galton-Watson tree survive for at least n generations. Additionally, as $n \to \infty$

$$1 - f_n(0) - \frac{2}{n\sigma^2} = o\left(1/n\right).$$

The proof of this theorem can be found in [14, Theorem 9.1].

Henceforth we consider only subcritical or critical non-degenerate offspring distributions.

Proposition 1.10. Let $(K_u)_{u\in\mathcal{U}}$ be a collection of independent and identically distributed random variables with distribution μ. Denote by θ the random subset of $\mathcal{U}$ given by

$$\theta = \{u = (u_1, \ldots, u_n) \in \mathcal{U} : u_j \leq K_{\pi^{n-j+1}(u)} \text{ for every } 1 \leq j \leq n\}.$$

Then θ is almost surely a tree. Furthermore $(Z_n)_{n\in\mathbb{N}_0}$ where

$$Z_n = \#\{u \in \theta : |u| = n\}, \quad n \in \mathbb{N}_0,$$

is a Galton-Watson process with offspring distribution μ and initial value $Z_0 = 1$. In this context, we will call θ a μ-Galton-Watson tree.

To prove this statement, it is sufficient to take $k_u(\theta) = K_u$ for every $u \in \theta$.

Remember that by proposition 1.8 we have that θ is almost surely a finite tree.

Definition 1.11. Also, for t a tree and $1 \leq j \leq k_\emptyset(t)$, we introduce the notation

$$T_j t = \{u \in \mathcal{U} : (j, u) \in t\}$$

that is, $T_j t$ is the tree t shifted at j.

Let us denote by Π_μ as the distribution of a μ-Galton-Watson tree on the space $\mathbb{A}$. Π_μ is characterized by

- $\Pi_\mu(k_\emptyset = j) = \mu(j)$, $j \in \mathbb{N}_0$.
- **(Branching property)** For every $j \geq 1$with $\mu(j) > 0$, the shifted trees $T_1 t, \ldots, T_j t$ are independent under the conditional probability $\Pi_\mu(\cdot | k_\emptyset = j)$, and their conditional distribution is Π_μ.

We now give a explicit formula for the distribution of a μ-Galton-Watson tree.

Proposition 1.12. Let us consider Π_μ as the distribution of θ, a μ-Galton-Watson tree on the space $\mathbb{A}$. Then, for every $t \in \mathbb{A}$,

$$\Pi_\mu(t) = \prod_{u \in t} \mu(k_u(t)).$$

Proof: Observe that

$$\{\theta = t\} = \bigcap_{u \in t} (K_u = k_u(t))$$

so we have

$$\Pi_\mu(t) = P(\theta = t) = \prod_{u \in t} P(K_u = k_u(t)) = \prod_{u \in t} \mu(k_u(t)).$$

■

Proposition 1.13. Let θ be a μ-Galton-Watson tree and consider Φ as defined in proposition 1.5. Then

$$\Phi(\theta) \overset{(d)}{=} (M_1, M_2, \ldots, M_T)$$

where $M_1, M_2, \ldots$ are independent random variables with distribution μ and

$$T = \inf\{n \geq 1 : M_1 + \ldots + M_n < n\}.$$

Proof: Assume that θ is given by the construction of the previous proposition.

Let $(U_n)_{0 \leq n \leq \#\theta - 1}$ be the elements of θ listed in lexicographical order, with the convention that $U_0 = \emptyset$.

By the construction of θ,

$$\Phi(\theta) = (K_{U_0}, \ldots, K_{U_{\#\theta-1}})$$

as $K_{U_0} + \ldots + K_{U_n} \geq n+1$ for all $n \in \{0, 1, \ldots, \#\theta - 2\}$ and $K_{U_0} + \ldots + K_{U_{\#\theta-1}} = \#\theta - 1$.

Define U_p for $p \geq \#\theta$ by adding $p - \#\theta + 1$ labels 1 to the right of $U_{\#\theta-1}$, as exemplified below

$$U_p = (U_{\#\theta-1}, 1, 1, \ldots, 1).$$

To prove the proposition we will check by induction on p that $K_{U_0}, \ldots, K_{U_p}$ are i.i.d. with distribution μ for every $p \geq 0$ (observe that $T \overset{(d)}{=} \#\theta - 1$ is already clear by the definition of Φ and the elements of $\mathcal{S}$). Observe that despite the variables $(K_u)_{u \in \mathcal{U}}$ being independent with distribution μ, the labels $(U_n)_{0\leq n\leq p}$ are random variables that depend on $(K_u)_{u\in\mathcal{U}}$.

For $p=0$ and $p=1$, the result follows since $U_0=\emptyset$ and $U_1=1$ deterministically.

Now take $p\geq 2$ and assume that the hypothesis holds for $k\leq p-1$. For every fixed $u\in\mathcal{U}$ consider $\mathcal{C}_u=\theta\cap\{v\in\mathcal{U}: v\leq u\}$, then $\#\mathcal{C}_u$ is measurable with respect to $\sigma(K_w: w<u)$. Indeed, taking $v=(v_1,\dots,v_n)\leq u$

$$\{v\in\theta\}=\cap_{m=0}^{n-1}\{K_{(v_1,\dots,v_m)}\geq v_{m+1}\}\in\sigma(K_w\colon\ w<v)\subset\sigma(K_w\colon\ w<u),$$

so we have

$$\{\#C_u\geq\ell\}=\bigcup_{\substack{v^1,\dots,v^\ell\leq u\\ v^i\neq v^j,\forall i\neq j}}\bigcap_{j=1}^{\ell}\{v^j\in\theta\}\quad\in\sigma(K_w\colon w<u).$$

We also note that $\{U_{\#\theta-1}=v\}$ is measurable with respect to $\sigma(K_v,v<u)$ when $v<u$:

$$\{U_{\#\theta-1}=(v_1,\dots,v_n)=v\}=\{K_v=0\}\cap\bigcap_{m=0}^{n-1}\{K_{(v_1,\dots,v_m)}=v_{m+1}\}\in\sigma(K_w\colon\ w<u).$$

As consequence, by our definition of the sequence $(U_n)_{n\in\mathbb{N}_0}$

$$\{U_p=u\}\cap\{\#\theta>p\}=\{u\in\theta\}\cap\{\#\mathcal{C}_u=p+1\}$$

so the left-hand side is measurable with respect to $\sigma(K_v,v<u)$.

Additionally the same holds for

$$\{U_p=u\}\cap\{\#\theta\leq p\}=\{\#C_u\leq p\}\cap\bigcup_{\mathcal{U}\ni v<u}\{U_{\#\theta-1}=v\}\cap\{u=(v,1,\dots,1)\}$$

Hence $\{U_p=u\}$ is $\sigma(K_v,v<u)$ measurable.

Now, in order to prove independence between K_{U_p} and $\sigma(K_{U_0},\dots,K_{U_{p-1}})$, take $g_0,g_1\dots,g_p$ nonnegative functions on $\mathbb{N}_0$ and note that

$$\begin{aligned}E[g_0(K_{U_0})g_1(K_{U_1})\cdots g_p(K_{U_p})]&=\\ &=\sum_{u_0<u_1<\cdots<u_p}E\left[\mathbb{1}_{\{U_0=u_0,\dots,U_p=u_p\}}g_0(K_{u_0})\cdots g_p(K_{u_p})\right]\\ &=\sum_{u_0<u_1<\cdots<u_p}E\left[\mathbb{1}_{\{U_0=u_0,\dots,U_p=u_p\}}g_0(K_{u_0})\cdots g_{p-1}(K_{u_{p-1}})\right]E[g_p(K_{u_p})]\end{aligned}$$

because, by the Galton-Watson tree construction, K_{u_p} is independent of $\sigma(K_v,v\leq u)\ni\{U_p=u\}$. Thus, as $E[g_p(K_{u_p})]=\int g_p\mathrm{d}\mu=\mu(g_p)$ does not depend on u_p,

$$E[g_0(K_{U_0})g_1(K_{U_1})\cdots g_p(K_{U_p})]=E[g_0(K_{U_0})g_1(K_{U_1})\cdots g_p(K_{U_p})]\mu(g_p),$$

and by applying the induction hypothesis we complete the proof. ∎

As immediate consequence of this proposition we obtain

Corollary 1.14. Let $(S_n)_{n\in\mathbb{N}_0}$ be a random walk on $\mathbb{Z}$ with initial value $S_0=0$ and jump distribution $\nu(k)=\mu(k+1)$ for every $k\geq -1$. Set

$$T=\inf\{n\geq 1: S_n=-1\}.$$

Then the Lukasiewicz path of a μ-Galton-Watson tree θ has the same distribution as $(S_0,S_1,\dots,S_T)$. In particular, $\#\theta$ and T have the same distribution.

1.3 Time-scale convergences

We have seen that a tree can be coded by a continuous function, the contour function. As the weak scale limit of continuous functions is well studied (see appendix A), we could first try to understand the scaling limit of the contour function. A simpler case is given below.

Example 1.15. Let $\mu \sim \text{Geom}(\frac{1}{2})$.

Note that, for $n \in \mathbb{N}$ and C_n the contour function at n of θ, the Galton-Watson tree with distribution μ, we have $C_{n+1} = C_n \pm 1$. We claim that, for $\mu \sim \text{Geom}(\frac{1}{2})$, the contour process $(C_n)_{n\in\mathbb{N}}$ behaves like a simple random walk.

To prove this, we begin by introducing some notations:

- i_n, the element $u \in \theta$ visited at time n by the particle that defines the contour function.
- $\tau_u^{(1)} = \inf_{k\geq 0}\{i_k = u\}$, the first visit time to $u \in \mathcal{U}$ and, for $\ell \geq 1$
$$\tau_u^{(\ell+1)} = \inf_{k>\tau_u^{(\ell)}}\{i_k = u\}$$
is the $\ell + 1$-th visit time to $u \in \mathcal{U}$. By convention, define $\inf\{\emptyset\} = \infty$.
- $A_u = \cup_{m\in\mathbb{N}}\pi^m(u)$, are the ancestors of $u \in \mathcal{U}$.
- $D_u^m(\theta) = D_u^m = \cup_{v\in\mathcal{U}}\{(u, m, v)\} \cap \theta$, the descendants of the m-th child of $u \in \mathcal{U}$ on the realization θ of the tree.
- $v_t^u = \sum_{m=0}^t \mathbb{1}_{\{i_m=u\}}$, the number of visits made to $u \in \mathcal{U}$ up to time t.
- $k_{i_n}(\theta) = k_{i_n}$, the total number of children of the vertex i_n in the realization θ of the Galton-Watson tree.

Suppose that $u = (u_1, \ldots, u_k) \in \theta$. Then

$$\tau_u^{(1)} = \sum_{\substack{v<u \\ v\notin A_u}} 2K_v(\theta) + \sum_{i=1}^{k}(2u_i - 1) \tag{1}$$

and, on the event $\{k_u \geq \ell - 1\}$ for $\ell > 1$

$$\tau_u^{(\ell)} = \tau_u^{(\ell-1)} + \sum_{v\in D_u^{\ell-1}} (2K_v(\theta)) + 2 \tag{2}$$

and observe that $\tau_u^{(\ell)} = \infty$ on $\{k_u < \ell - 1\} \cup \{u \notin \theta\}$.

Note that the sums

$$\sum_{\substack{v<u \\ v\notin A_u}} 2K_v(\theta) + \sum_{i=1}^{k}(2u_i - 1), \qquad \sum_{v\in D_u^{\ell-1}} (2K_v(\theta)) + 2, \quad \forall \ell \geq 2$$

are measurable with respect to $\sigma(k_v : v < u)$, and $\sigma(k_v : v > u)$, respectively, so the event $\{\tau_u^{(\ell)} = n\}$ can be decomposed as

$$\{\tau_u^{(\ell)} = n\} = \{k_u \geq \ell - 1\} \cap \{u \in \theta\} \cap A_n^\ell \tag{3}$$

where

$$A_n^\ell = \Big\{ \sum_{\substack{v<u \\ v\notin A_u}} 2K_v(\theta) + \sum_{i=1}^{k}(2u_i - 1) + \sum_{m=1}^{\ell-1}\sum_{v\in D_u^m}(2K_v(\theta)) + 2 = n \Big\}$$

is measurable with respect to $\sigma(k_v : v \neq u)$, therefore, it is independent of k_u.

We also have, for $u \in \mathcal{U}$

$$\{i_n = u\} = \bigcup_{m\in\mathbb{N}} \{\tau_u^{(m)} = n\}.$$

Note that $C_{n+1} = C_n + 1$ iff $k_{i_n} \geq v_n^n$, because as the vertex i_n is visited exactly $k_{i_n} + 1$ times by the explorer particle, if $k_{i_n} \geq v_n^n$ then i_n has at least one descendant that has not yet been visited by the particle, and, after this exploration, the particle will visit i_n at least once more. So we can write

$$\begin{aligned}
\mathbb{P}_\mu(C_{n+1} = C_n + 1) &= \mathbb{P}_\mu(k_{i_n} \geq v_n^{i_n}) \\
&= \sum_{k\geq 0} \mathbb{P}_\mu(k_{i_n} \geq k, v_n^n = k) \\
&= \sum_{k\geq 0}\sum_{u\in\mathcal{U}} \mathbb{P}_\mu(k_{i_n} \geq k, v_n^{i_n} = k, i_n = u) \\
&= \sum_{k\geq 0}\sum_{u\in\mathcal{U}} \mathbb{P}_\mu(k_u \geq k, \tau_u^{(k)} = n) \\
&= \sum_{k\geq 0}\sum_{u\in\mathcal{U}} \mathbb{P}_\mu(k_u \geq k, k_u \geq k-1, u \in \theta, A_n^k)
\end{aligned}$$

by the decomposition on (3). So, by independence of these events and the memoryless property of the geometric distribution

$$\begin{aligned}
\mathbb{P}_\mu(C_{n+1} = C_n + 1) &= \sum_{k\geq 0}\sum_{u\in\mathcal{U}} \mathbb{P}_\mu(k_u \geq k, k_u \geq k-1)\mathbb{P}_\mu(u \in \theta, A_n^k) \\
&= \sum_{k\geq 0}\sum_{u\in\mathcal{U}} \mathbb{P}_\mu(k_u \geq 1)\mathbb{P}_\mu(k_u \geq k-1)\mathbb{P}_\mu(u \in \theta, A_n^k) \\
&= \frac{1}{2}\sum_{k\geq 0}\sum_{u\in\mathcal{U}} \mathbb{P}_\mu(i_n = u, v_n^u = k) = \frac{1}{2}
\end{aligned}$$

as we claimed.

Let $(\theta_n)_{n\in\mathbb{N}}$ be a sequence of independent μ-Galton-Watson trees, and for each tree let us consider its contour function by $(C_{\theta_j}(s), 0 \leq s \leq \zeta(\theta_j))$. Then $\mathcal{C} = (\mathcal{C}_s)_{sz\in\mathbb{R}_+}$, where

$$\mathcal{C}_s = C_{\theta_j}(s - (\zeta(\theta_1) + \ldots + \zeta(\theta_{j-1})))$$

if $\zeta(\theta_1) + \ldots + \zeta(\theta_{j-1}) \leq s \leq \zeta(\theta_1) + \ldots + \zeta(\theta_j)$, is the concatenation of all contour functions.

As consequence of our discussion, $\mathcal{C}$ is exactly the linear interpolation of the graph of the reflected simple and symmetric random walk on $\mathbb{Z}$. Then, by Donsker's invariance principle and the continuous mapping theorem, as the reflection on $x = 0$ is a continuous map

$$\left(\frac{1}{\sqrt{n}}\mathcal{C}_{ns}\right)_{s\in\mathbb{R}^+} \xrightarrow[n\to\infty]{\text{(d)}} (\gamma_t)_{t\in\mathbb{R}^+} \tag{4}$$

where γ is a reflected Brownian motion.

In the next example, we consider a conditioned random tree also related to a Galton-Watson tree with geometric offspring distribution. In order to do so we introduce the following definition:

Definition 1.16. The *normalized Brownian excursion* $(e_t)_{t\in[0,1]}$ is the Brownian excursion conditioned to have length 1.

For a more precise definition of this process see [15, Section 2] or [8, Section 8.4].

Example 1.17. Let us prove time scale convergence for a tree uniformly chosen on the set of trees with a fixed number of vertices.

Denote the set of discrete rooted ordered trees with n vertices by $\mathbb{A}_n$, and consider $\mu \sim \text{Geom}(\frac{1}{2})$. We claim that the uniform distribution on $\mathbb{A}_n$ coincides with the distribution of a μ-Galton-Watson tree conditioned to have n vertices. Indeed, by proposition 1.12:

$$\Pi_\mu(t \cap \{\#t = n\}) = \prod_{u\in t} \mu(k_u(t)) = \frac{1}{2^{k_{u_0}+\ldots+k_{u_{\#t-1}}}} = \frac{1}{2^n}$$

and

$$\Pi_\mu(\#t = n) = \sum_{t\in\mathbb{A}_n} \Pi_\mu(t) = \sum_{t\in\mathbb{A}_n} \frac{1}{2^n} = \frac{\#\mathbb{A}_n}{2^n}$$

so we conclude our statement:

$$\Pi_\mu(t|\#t = n) = \frac{1}{\#\mathbb{A}_n}.$$

Let $\theta^{(n)}$ be a μ-Galton-Watson tree conditioned to have n vertices and $(C_t^{(n)})_{t\geq 0}$ the contour function associated to $\theta^{(n)}$ as in definition 1.4.

Consider $\mathcal{C}_0[a,b]$ the set of continuous functions with support on $[a,b]$ and let us define a sequence of transformations $K^n : \mathcal{C}_0[0,2n-2] \to \mathcal{C}_0[0,1]$ as

$$K^n(f)(s) = \begin{cases} 2ns - 1 & \text{, for } s \in [0, \frac{1}{2n}] \\ f(2ns-1) & \text{, for } s \in [\frac{1}{2n}, 1 - \frac{1}{2n}] \\ 2n - 1 - 2ns & \text{, for } s \in [1 - \frac{1}{2n}, 1]. \end{cases}$$

Note that $C^{(n)} \in \mathcal{C}_0[0,2n-2]$ so $K^n(C^{(n)})$ is a shifted rescaled version of the contour function.

Finally, consider $T_\ell = \inf\{k > 0 : S_k = \ell\}$ the first return time of $(S_k)_{k\in\mathbb{N}_0}$ to $\ell \in \mathbb{Z}$, where $(S_k)_{k\in\mathbb{N}_0}$ denotes the simple and symmetric random walk on $\mathbb{Z}$ with $S_0 = 0$. Therefore define $(S_t^{(n)})_{t\geq 0}$ as the linear interpolation of $(S_k)_{k\in\mathbb{N}_0}$ conditioned on $\{T_0 = n\}$.

By our previous example, the contour function of a μ-Galton-Watson forest has the same law as $(|S_t|)_{t\geq 0}$. It follows that $K^n(C^{(n)})$ has the same law as $K^n(|S^{(2n-2)}|)$, which coincides with the distribution of $(|S_{2nt}^{(2n)}| - 1)_{t\in[0,1]}$.

Lemma 1.18. The distribution of $(\frac{1}{\sigma\sqrt{n}} S_{[nt]})_{t\in[0,1]}$ under the conditional probability $P(\cdot|T_0 = n)$ converges as $n \to \infty$ to the law of the normalized Brownian excursion.

The proof of this lemma will be omitted and can be found in [15, Theorem 2.6].

By the continuous mapping theorem, $\left(\frac{|S^{(2n)}_{2nt}|-1}{\sqrt{2n}}\right)_{t\in[0,1]} \xrightarrow[n\to\infty]{(d)} (|e_t|)_{t\in[0,1]}$, therefore

$$\left(\frac{1}{\sqrt{2n}}K^n(C^{(n)})_t\right)_{t\in[0,1]} \xrightarrow[n\to\infty]{(d)} (|e_t|)_{t\in[0,1]}. \tag{5}$$

Noting that

$$\lim_{n\to\infty}\sup_{t\in[0,1]}\left|\frac{K^n(C^{(n)})_t}{\sqrt{2n}} - \frac{C^{(n)}_{2nt}}{\sqrt{2n}}\right| = 0$$

we conclude that

$$\left(\frac{1}{\sqrt{2n}}C^{(n)}_{2nt}\right)_{t\in[0,1]} \xrightarrow[n\to\infty]{(d)} (e_t)_{t\in[0,1]}$$

that is, the contour function of the tree uniformly chosen random tree in $\mathbb{A}_n$ converges weakly to the normalized Brownian excursion.

Usually, dealing with the transition probability of the contour function is not as simple as in these examples, in which the memoryless property played an essential role in relating the contour process to the behavior of a simple random walk.

Because of that, from now on we investigate the time scale convergence of trees by their coding by height functions, similarly to what was made in the previous examples. The reason for that is because proposition 1.7 and corollary 1.14 provide a relation between the height function, the Lukasiewicz path and a simple random walk.

1.3.1 Convergence of height functions

In the following sections, we consider μ a critical probability measure on $\mathbb{N}_0$ such that $0 < \sigma^2 < \infty$, where

$$\sigma^2 = \sum_{k=0}^{\infty} k^2\mu(k) - \left(\sum_{k\in\mathbb{N}_0} k\mu(k)\right)^2.$$

Definition 1.19. Consider $(\theta_n)_{n\in\mathbb{N}}$ a sequence of independent μ-Galton-Watson trees. As in the example 1.15, for each tree we consider its height function denoted by $(h_{\theta_i}(n), 0 \leq n \leq \#\theta_i - 1)$.

We define the *height process* $(H_n)_{n\in\mathbb{N}_0}$ of the trees by concatenating their height functions, i.e.,

$$H_n = h_{\theta_i}(n - (\#\theta_1 + \ldots + \#\theta_{i-1}))$$

if $\#\theta_1 + \ldots + \#\theta_{i-1}) \leq n \leq \#\theta_1 + \ldots + \#\theta_i$.

The height process determines the sequence of trees. In fact, the values of H between its k-th and $(k+1)$-th zero determines the height function of the k-th tree of the sequence.

As consequence of proposition 1.7 and corollary 1.14, we obtain the following description of the height process.

Proposition 1.20. For every $n \in \mathbb{N}_0$,

$$H_n \stackrel{d}{=} \#\{k \in \{0,1,\ldots,n-1\} : S_k = \inf_{k\leq j\leq n} S_j\}$$

where $(S_n)_{n\in\mathbb{N}_0}$ is a random walk with initial value $S_0 = 0$ and jump distribution $\nu(k) = \mu(k+1)$ for every $k \geq -1$.

Theorem 1.21. Let $(\theta_n)_{n\in\mathbb{N}}$ be a sequence of μ-Galton-Watson trees, and $(H_n)_{n\in\mathbb{N}_0}$ be the associated height process. Then:

$$\left(\frac{1}{\sqrt{n}} H_{[nt]}\right)_{t\in\mathbb{R}^+} \xrightarrow[n\to\infty]{\text{(d)}} \left(\frac{2}{\sigma}\gamma_t\right)_{t\in\mathbb{R}^+}$$

where γ is a reflected Brownian motion. The convergence holds in the sense of weak convergence on the Skorohod space $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+)$.

Proof: We begin by establishing weak convergence for the finite-dimensional marginals.

Let $S = (S_n)_{n\in\mathbb{N}_0}$ be the random walk described in proposition 1.20. Note that the jump distribution ν has mean 0 and finite variance, hence S is recurrent (by theorem 1.3.1 in [25]).

Donsker's invariance principle gives

$$\left(\frac{1}{\sqrt{n}} S_{[nt]}\right)_{t\in\mathbb{R}_+} \xrightarrow[n\to\infty]{\text{(d)}} (\sigma B_t)_{t\in\mathbb{R}^+} \tag{6}$$

where $S_{[nt]}$ is a time-scaled version of the linear interpolation of the random walk S and B is the standard linear Brownian motion started at the origin.

For every $n \geq 0$, define the time-reversed random walk $\hat{S}^n$ by

$$\hat{S}^n_k = S_n - S_{(n-k)^+}.$$

Note that $(\hat{S}^n_k)^n_{k=0}$ has the same distribution as $(S_k)^n_{k=0}$ as both walks begin and end in the same positions and perform the same steps, but in reversed order.

Additionally for any discrete trajectory $X = (X_n)_{n\in\mathbb{N}_0}$ set

$$\Phi_n(X) = \#\Big\{k \in \{1, 2, \ldots, n\} : X_k = \sup_{0\leq j\leq k} X_j\Big\}$$

and

$$K_n = \Phi_n(S) = \#\Big\{k \in \{1, 2, \ldots, n\} : S_k = M_k\Big\}$$

where $M_n = \sup_{0\leq k\leq n} S_k$.

From the formula on the proposition 1.20 we can see that

$$H_n = \#\Big\{k \in \{0, 1, \ldots, n-1\} : S_k = \inf_{k\leq j\leq n} S_j\Big\} = \Phi_n(\hat{S}^n).$$

Lemma 1.22. Define $(T_j)_{j\in\mathbb{N}_0}$ a sequence of stopping times by setting $T_0 = 0$ and inductively

$$T_j = \inf\{n > T_{j-1} : S_n = M_n\}, \quad j \geq 1.$$

Then, for $j \geq 1$, the random variables $S_{T_j} - S_{T_{j-1}}$ are independent and identically distributed, with distribution

$$P(S_{T_1} = k) = \nu([k, \infty]), \quad k \geq 0.$$

Proof:

It follows from the strong Markov property (see appendix C.8) and properties of the random walk that the random variables $S_{T_j} - S_{T_{j-1}}$, $j \in \mathbb{N}$, are independent and identically distributed.

We know that the invariant measure of the recurrent random walk S is the counting measure on $\mathbb{Z}$. Indeed, for a fixed $j \in \mathbb{Z}$ and any $k \in \mathbb{N}_0$,

$$\sum_{i\in\mathbb{Z}} \#(i) P(S_k = i | S_{k+1} = j) = \sum_{i\in\mathbb{Z}} \nu(i-j) = \sum_{l=-1}^{\infty} \nu(l) = 1 = \#(j).$$

Since ν has mean 0 and variance $0 < \sigma^2 < \infty$, it follows that $\nu(-1) > 0$ and there exists some $k \in \mathbb{N}$ such that $\nu(k) \geq 0$. So for every $i, j \in \mathbb{Z}$ with $j \geq i$,

$$P(S_{n+j-i} = i | S_n = j) \geq (\nu(-1))^{j-1} > 0$$

for any $n \in \mathbb{N}_0$. On the other hand taking $j - i = qk + r$, for some $q \in \mathbb{N}$ and $0 \leq r \leq k-1$

$$P(S_{n+q+k+1-r} = j | S_n = i) \geq (\nu(k))^{q+1} (\nu(-1))^{k-r} > 0$$

for any $n \in \mathbb{N}_0$. So $(S_n)_{n\in\mathbb{N}_0}$ is irreducible.

Then, by theorem C.12, for $R_0 = \inf\{n \geq 1 : S_n = 0\}$ the measure

$$\gamma(i) = E\left[\sum_{n=0}^{R_0-1} \mathbb{1}_{\{S_n=i\}}\right] = E\left[\sum_{n\geq 0} \mathbb{1}_{\{S_n=i, n\leq R_0-1\}}\right]$$

is also a invariant measure for S, so it is equal the counting measure up to multiplication by a positive constant. Since $\gamma(0) = \#\{0\} = 1$, for every $j \in \mathbb{Z}$

$$E\left[\sum_{n=0}^{R_0-1} \mathbb{1}_{S_n=j}\right] = \#(j) = 1.$$

Notice that, almost surely, $T_1 \leq R_0$ and $S_n > 0$ for $T_1 \leq n < R_0$. So, for every $i \leq 0$

$$E\left[\sum_{n=0}^{T_1-1} \mathbb{1}_{S_n=i}\right] = 1.$$

Therefore, for any function $g : \mathbb{Z} \to \mathbb{Z}_+$

$$E\left[\sum_{n=0}^{T_1-1} g(S_n)\right] = E\left[\sum_{i\leq 0}\sum_{n\geq 0} g(i) \mathbb{1}_{S_n=i, n\leq T_1-1}\right] = \sum_{i\leq 0} g(i). \tag{7}$$

Then, for any function $f : \mathbb{Z} \to \mathbb{Z}_+$,

$$\begin{aligned} E[f(S_{T_1})] &= E\left[\sum_{k\geq 0} f(S_{k+1}) \mathbb{1}_{T_1=k+1}\right] \\ &= E\left[\sum_{k\geq 0} f(S_{k+1}) \mathbb{1}_{T_1>k, S_{k+1}\geq 0}\right] \\ &= \sum_{k\geq 0} E\left[f(S_{k+1}) \mathbb{1}_{T_1>k, S_{k+1}\geq 0}\right] \end{aligned}$$

by monotone convergence and linearity. Using the Markov property at time k and (7) respectively

$$\begin{aligned} E[f(S_{T_1})] &= \sum_{k\geq 0} E\left[\mathbb{1}_{T_1>k} \sum_{j\geq 0} \nu(j) f(S_k + j) \mathbb{1}_{S_k+j\geq 0}\right] \\ &= \sum_{i\leq 0}\sum_{j\geq 0} \nu(j) f(i+j) \mathbb{1}_{i+j\geq 0} = \sum_{m\geq 0} f(m) \sum_{j\geq m} \nu(j) \end{aligned}$$

taking $f = \mathbb{1}_k$ the result follows. ■

Note that S_{T_1} have finite first moment:

$$E[S_{T_1}] = \sum_{k\geq 0} k\nu([k,\infty)) = \sum_{k\geq 0}\sum_{j\geq k} k\nu(j) = \sum_{j\geq 0} \frac{j(j+1)}{2}\nu(j) = \frac{\sigma^2}{2}$$

because each $\nu(j)$ is summed $\sum_{k=0}^{j} k$ times.

Let us introduce the notation

$$I_n = \inf_{0\leq k\leq n} S_k.$$

To establish the convergence of the finite-dimensional marginals we will need the following lemma.

Lemma 1.23. We have

$$\frac{H_n}{S_n - I_n} \xrightarrow[n\to\infty]{(\text{ P})} \frac{2}{\sigma^2}$$

where $\xrightarrow{(\text{ P})}$ means convergence in probability.

Proof: From our definitions, since $T_{K_n} \leq n$ almost surely, by definition of K_n and T_j

$$M_n = \sum_{n\geq T_k} (S_{T_k} - S_{T_{k-1}}) = \sum_{k=1}^{K_n} (S_{T_k} - S_{T_{k-1}}).$$

As $K_n \xrightarrow[n\to\infty]{} \infty$ almost surely, using the previous lemma and the law of large numbers we get

$$\frac{M_n}{K_n} \xrightarrow[n\to\infty]{} E[S_{T_1}] = \frac{\sigma^2}{2} \quad \text{a.s.}$$

Since $(\hat{S}_k^n)_{k=0}^n \overset{(d)}{=} (S_k)_{k=0}^n$, replacing S with $\hat{S}^n$ we have that

$$(M_n, K_n) \overset{(d)}{=} \left(\sup_{0\leq k\leq n} \hat{S}_k^n, \Phi_n(\hat{S}^n)\right) = (S_n - I_n, H_n).$$

Hence it holds

$$\frac{S_n - I_n}{H_n} \xrightarrow[n\to\infty]{(P)} E[S_{T_1}] = \frac{\sigma^2}{2}.$$

■

From (6), we have that for every choice of $0 \leq t_1 \leq t_2 \leq \ldots \leq t_m$

$$\frac{1}{\sqrt{n}}\left(S_{[nt_1]} - I_{[nt_1]}, \ldots, S_{[nt_m]} - I_{[nt_m]}\right) \xrightarrow[n\to\infty]{(\text{d})} \sigma\left(B_{t_1} - \inf_{0\leq s\leq t_1} B_s, \ldots, B_{t_m} - \inf_{0\leq s\leq t_m}\right)$$

Thus it follows from lemma 1.23 and Slutsky's theorem that

$$\frac{1}{\sqrt{n}}\left(H_{[nt_1]}, \ldots, H_{[nt_m]}\right) \xrightarrow[n\to\infty]{(\text{d})} \frac{2}{\sigma}\left(B_{t_1} - \inf_{0\leq s\leq t_1} B_s, \ldots, B_{t_m} - \inf_{0\leq s\leq t_m}\right).$$

By using the reflection principle and the fact that a time reversal standard Brownian motion still is a Brownian motion we can state that the process

$$\beta_t = B_t - \inf_{0\leq s\leq t} B_s$$

is a reflected Brownian motion. This complete the proof of convergence of finite-dimensional marginals in theorem 1.21.

Now we will proof tightness property on the law of the processes

$$H_t^{(n)} = \frac{H_{[nt]}}{\sqrt{n}}$$

in the set of all probability measures on the Skorohod space $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$. By theorem A.15, it is enough to verify these two properties:

(i) For every $t \geq 0$ and $\eta > 0$ there exists a constant $K \geq 0$ such that

$$\liminf_{n\to\infty} P(H_t^{(n)} \leq K) \geq 1 - \eta.$$

(ii) For every fixed $T > 0$ and $\delta > 0$, define $I_i^{(k)} = \left[\frac{(i-1)}{2^k}T, \frac{i}{2^k}T\right]$, $1 \leq i \leq 2^n$. Then

$$\lim_{k\to\infty} \limsup_{n\to\infty} P\left(\max_{1\leq i\leq 2^k} \sup_{t\in I_i^{(k)}} |H_t^{(n)} - H_{i2^{-k}T}^{(n)}| > \delta\right) = 0.$$

Property (i) follows from convergence of finite-dimensional marginals. So let us prove property (ii). Fix $T > 0$ and $\delta > 0$ and observe that

$$P\left(\max_{1\leq i\leq 2^n} \sup_{t\in I_i^{(n)}} |H_t^{(n)} - H_{i2^{-k}T}^{(n)}| > \delta\right) \leq A_1(n,k) + A_2(n,k) + A_3(n,k) \tag{8}$$

where

$$\begin{aligned}
A_1(n,k) &= P\left(\max_{1\leq i\leq 2^k} |H_{i2^{-k}T}^{(n)} - H_{(i-1)2^{-k}T}^{(n)}| > \frac{\delta}{5}\right)\\
A_2(n,k) &= P\left(\sup_{t\in I_i^{(k)}} H_t^{(n)} > H_{(i-1)2^{-k}T}^{(n)} + \frac{4\delta}{5} \text{ for some } 1 \leq i \leq 2^k\right)\\
A_3(n,k) &= P\left(\inf_{t\in I_i^{(k)}} H_t^{(n)} < H_{(i-1)2^{-k}T}^{(n)} - \frac{4\delta}{5} \text{ for some } 1 \leq i \leq 2^k\right).
\end{aligned}$$

Indeed, adding $\pm H_{(i-1)2^{-k}T}$, using the triangular inequality and that $P(X + Y > \delta) \leq P(X > \delta_1) + P(Y > \delta_2)$, where $\delta_1 + \delta_2 = \delta$, it follows that

$$\begin{aligned}
P\left(\max_{1\leq i\leq 2^n} \sup_{t\in I_i^{(n)}} |H_t^{(n)} - H_{i2^{-k}T}^{(n)}| > \delta\right) \leq & P\left(\max_{1\leq i\leq 2^k} |H_{i2^{-k}T}^{(n)} - H_{(i-1)2^{-k}T}^{(n)}| > \frac{\delta}{5}\right)\\
& + P\left(\max_{1\leq i\leq 2^n} \sup_{t\in I_i^{(n)}} |H_t^{(n)} - H_{(i-1)2^{-k}T}^{(n)}| > \frac{4\delta}{5}\right)
\end{aligned}$$

moreover the last term of the sum above is bounded by $A_2(n,k) + A_3(n,k)$.

Let us bound A_1. As consequence of the convergence of finite-dimensional marginals:

$$\limsup_{n\to\infty} A_1(n,k) = P\left(\max_{1\leq i\leq 2^k} \frac{2}{\sigma}|\beta_{i2^{-k}T} - \beta_{\frac{(i-1)}{2^k}T}| > \frac{\delta}{5}\right)$$

by continuity of the process β_t, the right-hand size tends to 0 when $k \to \infty$.

In order to bound A_2, consider $(\tau_k^{(n)})_{k\in\mathbb{N}_0}$ a sequence of stopping times defined by $\tau_0^{(n)} = 0$ and

$$\tau_{l+1}^{(n)} = \inf\Big\{t \geq \tau_l^{(n)} : H_t^{(n)} \geq \inf_{\tau_l^{(n)} \leq r \leq t} H_r^{(n)} + \frac{\delta}{5}\Big\}.$$

Let $1 \leq i \leq 2^n$ be such that

$$\sup_{t\in I_i^{(k)}} H_t^{(n)} > H_{(i-1)2^{-k}T}^{(n)} + \frac{4\delta}{5}$$

then the interval $I_i^{(k)} = \left[\frac{(i-1)}{2^k}T, \frac{i}{2^k}T\right]$ contains at least one of the random times $\tau_l^{(n)}, l \geq 0$. Let $\tau_j^{(n)}$ be the first of such times in this interval. By minimality of $\tau_j^{(n)}$ we have

$$\sup_{t\in\left[\frac{(i-1)}{2^k}T, \tau_j^{(n)}\right)} H_t^{(n)} \leq H_{(i-1)2^{-k}T}^{(n)} + \frac{\delta}{5}.$$

Since the upward jumps of $H^{(n)}$ are at most of size $\frac{1}{\sqrt{n}}$, we also got, taking $n > (\frac{5}{\delta})^2$

$$H_{\tau_j^{(n)}}^{(n)} \leq H_{(i-1)2^{-k}T}^{(n)} + \frac{\delta}{5} + \frac{1}{\sqrt{n}} < H_{(i-1)2^{-k}T}^{(n)} + \frac{2\delta}{5}.$$

From our choice of i we have that

$$\sup_{t\in\left[\tau_j^{(n)}, \frac{i}{2^k}T\right]} H_t^{(n)} > H_{\tau_j^{(n)}}^{(n)} + \frac{\delta}{5}$$

which means that $\tau_{j+1}^{(n)} \leq i2^{-k}T$. From all this we conclude that for $n > (\frac{5}{\delta})^2$

$$A_2(n,k) \leq P\left(\tau_l^{(n)} < T \text{ and } \tau_{l+1}^{(n)} - \tau_l^{(n)} \leq 2^{-k}T \text{ for some } l \geq 0\right) \tag{9}$$

An analogous argument gives a similar bound for $A_3(n,k)$. Indeed, consider $(\hat{\tau}_k^{(n)})_{k\in\mathbb{N}_0}$ a sequence of stopping times defined by $\hat{\tau}_0^{(n)} = 0$ and

$$\hat{\tau}_{l+1}^{(n)} = \inf\Big\{t \geq \hat{\tau}_l^{(n)} : H_t^{(n)} \leq \sup_{\hat{\tau}_l^{(n)} \leq r \leq t} H_r^{(n)} - \frac{\delta}{5}\Big\}.$$

Now choose $1 \leq i \leq 2^n$ such that

$$\inf_{t\in I_i^{(k)}} H_t^{(n)} < H_{(i-1)2^{-k}T}^{(n)} - \frac{4\delta}{5}$$

this implies that the interval $I_i^{(k)}$ contains at least one $\hat{\tau}_l^{(n)}$, for $l \geq 0$. Let us denote by $\hat{\tau}_j^{(n)}$ the last of such times on this interval. We want to show that $\hat{\tau}_{j-1}^{(n)} \in I_i^{(k)}$.

If $\inf_{r\in[(i-1)2^{-k}T, \hat{\tau}_j^{(n)})} H_r^{(n)} = \inf_{t\in I_i^{(n)}} H_t^{(n)}$, then, by our choice of i

$$\inf_{r\in[(i-1)2^{-k}T, \hat{\tau}_j^{(n)})} H_r^{(n)} < H_{(i-1)2^{-k}}^{(n)} - \frac{4\delta}{5}$$

since we are taking the linear interpolation of $H^{(n)}$, the result follows.

On the other hand, maximality of $\hat{\tau}_j^{(n)}$ implies that

$$\inf_{r\in[\hat{\tau}_j^{(n)},i2^{-k}T]} H_r^{(n)} \geq \sup_{r\in[\hat{\tau}_j^{(n)},i2^{-k}T]} H_r^{(n)} - \frac{\delta}{5} \geq H_{\hat{\tau}_j^{(n)}}^{(n)} - \frac{\delta}{5}$$

by our choice of i, if $\inf_{t\in I_i^{(n)}} H_t^{(n)} = \inf_{r\in[\hat{\tau}_j^{(n)},i2^{-k}T]} H_r^{(n)}$, it follows that

$$H_{(i-1)2^{-k}} \geq H_{\hat{\tau}_j^{(n)}}^{(n)} + \frac{3\delta}{5}$$

that is, $\hat{\tau}_{j-1}^{(n)} \in I_i^{(k)}$. Then

$$A_3(n,k) \leq P\left(\hat{\tau}_l^{(n)} < T \text{ and } \hat{\tau}_{l+1}^{(n)} - \hat{\tau}_l^{(n)} \leq 2^{-k}T \text{ for some } l \geq 0\right). \tag{10}$$

Now we will state and prove the next lemmas in terms of the variables $(\tau_k^{(n)})_{k\in\mathbb{N}_0}$, but note that the same results are easily obtained in a similar manner when we consider $(\hat{\tau}_k^{(n)})_{k\in\mathbb{N}_0}$ instead.

Lemma 1.24. For every $x > 0$ and $n \geq 1$, set

$$G_n(x) = P\left(\tau_l^{(n)} < T \text{ and } \tau_{l+1}^{(n)} - \tau_l^{(n)} < x \text{ for some } l \geq 0\right)$$

and

$$F_n(x) = \sup_{l\geq 0} P\left(\tau_l^{(n)} < T \text{ and } \tau_{l+1}^{(n)} - \tau_l^{(n)} < x\right).$$

Then, for every integer $K \geq 1$

$$G_n(x) \leq KF_n(x) + Ke^T \int_0^\infty e^{-Ky} F_n(y)\,\mathrm{d}y.$$

Proof: For every $K \in \mathbb{N}$ we have

$$\begin{aligned} G_n(x) &\leq \sum_{k=0}^{K-1} P\left(\tau_l^{(n)} < T \text{ and } \tau_{l+1}^{(n)} - \tau_l^{(n)} < x\right) + P\left(\tau_K^{(n)} < T\right) \\ &\leq KF_n(x) + e^T E\left[\mathbb{1}_{\tau_K^{(n)}<T} \exp\left(-\sum_{l=0}^{K-1}(\tau_{l+1}^{(n)} - \tau_l^{(n)})\right)\right] \\ &\leq KF_n(x) + e^T \prod_{l=0}^{K-1} E\left[\mathbb{1}_{\tau_K^{(n)}<T} \exp\left(-K(\tau_{l+1}^{(n)} - \tau_l^{(n)})\right)\right]^{\frac{1}{K}} \end{aligned}$$

using Holder inequality for the last step. Finally, for every $l \in \{0, 1, \ldots, K-1\}$

$$\begin{aligned} E\left[\mathbb{1}_{\tau_K^{(n)}<T} \exp\left(-K(\tau_{l+1}^{(n)} - \tau_l^{(n)})\right)\right] &\leq E\left[\mathbb{1}_{\tau_K^{(n)}<T} \int_{\tau_{l+1}^{(n)}-\tau_l^{(n)}}^\infty Ke^{-Ky}\,\mathrm{d}y\right] \\ &\leq \int_0^\infty Ke^{-Ky} F_n(y)\,\mathrm{d}y \end{aligned}$$

by Fubini's theorem. ■

As a consequence of this lemma, in order to bound (9) and (10) we only need to bound the value of $F_n(x)$.

Lemma 1.25. The random variables $\tau_{l+1}^{(n)} - \tau_l^{(n)}$ are independent and identically distributed. Furthermore

$$\lim_{x\downarrow 0}\left(\limsup_{n\to\infty} P(\tau_1^{(n)} \le x)\right) = 0.$$

Proof: We will prove that

$$\left(H^{(n)}_{\tau_l^{(n)}+p} - \inf_{\tau_l^{(n)}\le t\le \tau_l^{(n)}+p} H_t^{(n)}\right)_{p\ge 0}$$

is independent of $\mathcal{F}_{\tau_l^{(n)}}$ and has the same distribution of $(H_p^{(n)})_{p\ge 0}$.

To simplify the notation let us denote $\tau_l^{(n)} = \tau$ and consider $n = 1$. By proposition 1.20

$$\inf_{\tau\le k\le \tau+n} H_k = \#\Big\{k \in \{0, 1, \dots, \tau - 1\} : S_k = \inf_{k\le j\le \tau+n} S_j\Big\}$$

then we get

$$\begin{aligned} H_{\tau+n} - \inf_{\tau\le k\le \tau+n} H_k &= \#\Big\{k \in \{\tau, \dots, \tau + n - 1\} : S_k = \inf_{k\le j\le \tau+n} S_j\Big\} \\ &= \#\Big\{k \in \{0, 1, \dots, n - 1\} : S_k^\tau = \inf_{k\le j\le n} S_j^\tau\Big\} \end{aligned}$$

where $S_j^\tau = S_{\tau+j} - S_\tau$. As consequence of the strong Markov property, S^τ is independent of $\mathcal{F}_\tau$ and has the same distribution as S, concluding the proof of the affirmation.

Then, writing

$$\begin{aligned} \tau_{l+1}^{(n)} - \tau_l^{(n)} &= \inf\Big\{t \ge \tau_l^{(n)} : H_t^{(n)} - \inf_{\tau_l^{(n)}\le r\le t} H_r^{(n)} \ge \frac{\delta}{5}\Big\} - \tau_l^{(n)} \\ &= \inf\Big\{t \ge 0 : H_t^{(n)} - \inf_{\tau_l^{(n)}\le r\le t} H_r^{(n)} \ge \frac{\delta}{5}\Big\} \end{aligned}$$

the first statement follows from our affirmation.

Now we prove the second assertion. For every $\eta > 0$ set

$$T_\eta^{(p)} = \inf\Big\{t \ge 0 : \frac{S_{[nt]}}{\sqrt{n}} < -\eta\Big\}$$

Then

$$P(\tau_1^{(n)} \le x) = P\left(\sup_{s\le x} H_s^{(n)} > \frac{\delta}{5}\right) \le P\left(\sup_{s\le T_\eta^{(p)}} H_s^{(n)} > \frac{\delta}{5}\right) + P\left(T_\eta^{(p)} < x\right) \tag{11}$$

equation (6) implies that

$$\limsup_{n\to\infty} P(T_\eta^{(n)} < x) \le \limsup_{n\to\infty} P\left(\inf_{t\le x} \frac{S_{[nt]}}{\sqrt{n}} < -\eta\right) \le P\left(\inf_{t\le x} \beta_t \le -\eta\right)$$

and $\lim_{x\downarrow 0} P\left(\inf_{t\le x} \beta_t \le -\eta\right) = 0$, because of the continuity of the Brownian motion.

By construction of the height process and the random walk S on proposition 1.20 and corollary 1.14

$$\sup_{s\le T_\eta^{(n)}} H_s^{(n)} \overset{(d)}{=} \frac{\mathcal{E}_n - 1}{\sqrt{n}}$$

where $\mathcal{E}_n$ is the extinction time of a μ-Galton-Watson tree started at $Z_0 = [\eta\sqrt{n}]+1$ (when beginning at 0 the process ends when the rescaled random walk achieves the height $-\frac{1}{\sqrt{n}}$, so starting from $[\eta\sqrt{n}]+1$ the process ends when the rescaled random walk achieves $-\eta$).

As consequence of theorem 1.9

$$P\left(\sup_{s\leq T_\eta^{(n)}} H_s^{(n)} > \frac{\delta}{5}\right) = 1 - [f_{[(\delta\sqrt{n})/5]+1}(0)]^{[\eta\sqrt{n}]+1}$$

as each one of the $[\eta\sqrt{n}]+1$ individuals of the first generation will built its family independently. Using theorem 1.9 again

$$\lim_{\eta\to 0}\left(\liminf_{n\to\infty}[f_{[(\delta\sqrt{n})/5]+1}(0)]^{[\eta\sqrt{n}]+1}\right) = 1$$

thus $\lim_{\eta\to 0}\left(\liminf_{n\to\infty} P\left(\sup_{s\leq T_\eta^{(n)}} H_s^{(n)} > \delta/5\right)\right) = 0$ and the result follows. ■

Remark 1.26. Note that

$$P(\hat{\tau}_1^{(n)} \leq x) \leq P\left(\sup_{s\leq x} H_s^{(n)} > \frac{\delta}{5}\right)$$

implying that inequality (11) also holds for $\hat{\tau}_1^{(n)}$.

Finally, in order to complete the proof of the theorem, set

$$F(x) = \lim_{n\to\infty} F_n(x), \quad G(x) = \limsup_{n\to\infty} G_n(x)$$

By lemma 1.25, $F(x) \downarrow 0$ as $x \downarrow 0$. On the other hand, lemma 1.24 states that for every $K \in \mathbb{N}$

$$G(x) \leq KF(x) + Ke^T \int_0^\infty e^{-Ky} F(y)\, \mathrm{d}y$$

then, by the bounded convergence theorem, $G(x) \downarrow 0$ when $x \downarrow 0$. By the bound in (9) we achieved

$$\lim_{k\to\infty}\left(\limsup_{n\to\infty} A_2(n,k)\right) = 0$$

and the same property holds for $A_3(n,k)$.

This proved tightness on the law of $H^{(n)}$. By Prohorov's theorem and the finite-dimensional marginals convergence we finished the proof of the theorem. ■

Note that in the previous proof as we used recurrence multiple times, that is, we used that the expectation of the random walk is 0. This property of the random walk depends only on the average of μ, so it is important to take a critical distribution.

1.3.2 Convergence of contour functions

We can use theorem 1.21 to show scale convergence of contour functions in a more general setting than the one of the example 1.15.

Let μ be a critical probability distribution on $\mathbb{N}_0$ with finite variance, and $(\theta_n)_{n\in\mathbb{N}}$ a sequence of independent μ-Galton-Watson trees.

As done for the height function and for our first example, we now want to introduce the contour process of the forest $(\theta_n)_{n\in\mathbb{N}}$. However the contour function of a tree θ is supported on $[0,\zeta(\theta)]$, where $\zeta(\theta) = 2(\#\theta - 1)$, so the contour function of the trees $\theta = \{\emptyset\}$ is trivial. Therefore, instead of just concatenating the contour functions as made before we will make the convention of defining the contour function $C_t(\theta)$ on the interval $[0,\zeta(\theta)+1]$ by taking $C_t(\theta) = 0$ if $\zeta(\theta) \leq t \leq \zeta(\theta)+1$.

Now we obtain the *contour process* $(C_t)_{t\in\mathbb{R}_+}$ of $(\theta_n)_{n\in\mathbb{N}}$ by concatenating the functions $(C_t(\theta_n), 0 \leq t \leq \zeta(\theta_n)+1)$.

Finally we state the convergence theorem for the rescaled contour process.

Theorem 1.27. Let μ be a critical distribution for the Galton-Watson tree, $(\theta_n)_{n\in\mathbb{N}}$ be a sequence of independent μ-Galton-Watson trees, and $(C_t)_{t\in\mathbb{R}_+}$ its associated contour process. Then

$$\left(\frac{1}{\sqrt{n}}C_{2nt}\right)_{t\in\mathbb{R}_+} \xrightarrow[n\to\infty]{(\mathrm{d})} \left(\frac{2}{\sigma}\gamma_t\right)_{t\in\mathbb{R}^+} \tag{12}$$

Proof: Our goal is to write the contour process in terms of the height process. For this purpose remember that H_n corresponds to the generation of the n-th individual on a list organized by the lexicographical order, one tree after another. Additionaly remember that $I_n = \inf_{0\leq k\leq n} S_k$, where $(S_n)_{n\in\mathbb{N}_0}$ is the random walk defined in 1.14, decreases if and only if $n = \sum_{\ell=1}^{k} \#\theta_\ell$ for some $k \in \mathbb{N}$ or when the individual U_n is a trivial tree.

Then consider

$$J_n = 2n - H_n + I_n, \text{ for } n \in \mathbb{N}_0.$$

By our observations on the definitions of $(H_n)_{n\in\mathbb{N}_0}$ and $(I_n)_{n\in\mathbb{N}_0}$, $(J_n)_{n\in\mathbb{N}_0}$ is a strictly increasing sequence such that $J_n \geq n$.

Now let us check that $[J_n, J_{n+1}]$ is the time interval when the contour process goes from the individual n to the individual $n+1$. First we will give a heuristic proof of this claim.

Note that the estimated time for the contour process to reach the $n-$th vertex if it is in θ_1 is $2n - H_n$. Indeed the particle must cross at most n edges twice in order to reach U_n, with the exception of the H_n ancestors of U_n, that must be visited again when the returns to the root. On the other hand, if $U_n \notin \theta_1$, then we must add the I_n horizontal lines placed by convention at the end of the contour function of each tree crossed up to time n. Now we will give an alternative proof by induction.

As $[J_0, J_1] = [0,1]$ the result is true for $n = 0$.

Assume that the hypothesis is true for $n \leq k-1$, $k \in \mathbb{N}$. Then J_k is the first time of arrival of the contour process on the k th individual of the forest.

The possible cases are the following:

- $J_{k+1} - J_k = 1$ iff $U_k \prec U_{k+1}$ ($\Rightarrow H_{k+1} = H_k + 1$ and $I_{k+1} = I_k$), or if $k+1 = \#\theta_1$ or $k+1 < \#\theta_1$ and U_{k+1} is a trivial tree ($\Rightarrow H_{k+1} = H_k$ and $I_{k+1} = I_k - 1$).

- $J_{k+1} - J_k = 2$ iff U_{k+1} and U_k are siblings.

- $J_{k+1} - J_k = 2 + j$ iff $\pi^j(U_k)$ and U_{k+1} are siblings or if the last individual of a tree is on the j th generation (or on the $j+1$ th, if $\#\theta_1 = k+1$)

In each case, $[J_k, J_{k+1}]$ is the time needed for the contour process to go to from the individual k to the individual $k+1$, completing the proof of the claim.

Heuristically, we can think that

From this observation, we get

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

Define a random function $\varphi : \mathbb{R}_+ \to \mathbb{N}_0$ by $\varphi(t) = n$ if and only if $t \in [J_n, J_{n+1})$. From the previous bound, for all $m \in \mathbb{N}$

$$\sup_{t\in[0,m]} |C_t - H_{\varphi(t)}| \leq \sup_{t\in[0,J_m]} |C_t - H_{\varphi(t)}| \leq 1 + \sup_{n\leq m} |H_{n+1} - H_n| \tag{13}$$

Similarly it follows from the definition of J_n that

$$\sup_{t\in[0,m]} \left|\varphi(t) - \frac{t}{2}\right| \leq \sup_{t\in[0,J_m]} \left|\varphi(t) - \frac{t}{2}\right| \leq \frac{1}{2} \sup_{n\leq m} H_n + \frac{1}{2}|I_m| + 1. \tag{14}$$

Gathering all this information we can now associate the behavior of the contour function to the height function.

For every $n \in \mathbb{N}$, denote $\varphi_n(t) = n^{-1}\varphi(nt)$. By equation (13) we have that for every $m \geq 1$

$$\sup_{t\in[0,m]} \left|\frac{1}{\sqrt{n}} C_{2nt} - \frac{1}{\sqrt{n}} H_{n\varphi_n(2t)}\right| \leq \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} \sup_{t\leq 2m} |H_{[nt]+1} - H_{[nt]}| \xrightarrow[n\to\infty]{\text{(d)}} 0 \tag{15}$$

by theorem 1.21.

On the other hand, convergence (6) given by Donsker's theorem implies that for every $m \geq 1$

$$\frac{1}{\sqrt{n}} I_{mn} \xrightarrow[n\to\infty]{\text{(d)}} \sigma \inf_{t\leq m} B_t$$

then, putting the previous equation together with (14) and theorem 1.21

$$\sup_{t\in[0,m]} |\varphi_n(2t) - t| \leq \frac{1}{n} \sup_{k\leq 2mn} H_k + \frac{1}{n}|I_{2mn}| + \frac{2}{n} \xrightarrow[n\to\infty]{\text{(d)}} 0$$

combining this convergence with theorem 1.21 give us

$$\left(\frac{1}{\sqrt{n}} H_{[n\varphi_n(2t)]}\right)_{t\in\mathbb{R}^+} \xrightarrow[n\to\infty]{\text{(d)}} \left(\frac{2}{\sigma}\gamma_t\right)_{t\in\mathbb{R}^+}$$

and equation (15) implies the desired result. ■

Chapter 2

Real trees

2.1 Real trees

In this section we aim to build metric spaces that generalize the notion of a discrete tree.

Definition 2.1. A metric space $(\mathcal{T}, d)$ is a *real tree* if it holds for all a, b in $\mathcal{T}$

(i) there exists a unique isometric map $f_{a,b} : [0, d(a, b)] \to \mathcal{T}$ such that $f_{a,b}(0) = a$ and $f_{a,b}(d(a, b)) = b$,

(ii) If $q : [0, 1] \to \mathcal{T}$ is a continuous injective map such that $q(0) = a$ and $q(1) = b$ we have $q([0, 1]) = f_{a,b}([0, d(a, b)])$.

When a real tree is equipped with a distinguished vertex $\rho = \rho(\mathcal{T})$ (the root of $\mathcal{T}$) we call $(\mathcal{T}, d, \rho)$ a rooted real tree.

Remark 2.2. Condition (i) implies that between every pair of points $a, b \in \mathcal{T}$ there exists a unique path with length $d(a, b)$. Condition (ii) implies that this is the unique simple path between a and b in $\mathcal{T}$.

Consequently $\mathcal{T}$ is a connected set with only one simple path between every pair of points modulo reparametrizations.

Now we set some notations that will be useful later.

Definition 2.3. Consider $(\mathcal{T}, d, \rho)$ a rooted real tree and $a, b \in \mathcal{T}$. We define

(a) The *path between* a *and* b given by $f_{a,b}([0, d(a, b)])$ will be denoted by $[\![a, b]\!]$.

(b) We say that a is an *ancestor* of b ($a \preceq b$) if $a \in [\![\rho, b]\!]$.

(c) $a \wedge b$, the *most recent common ancestor of* a *and* b, is the unique $c \in \mathcal{T}$ such that $[\![\rho, c]\!] = [\![\rho, a]\!] \cap [\![\rho, b]\!]$.

(d) The *multiplicity of a vertex* $a \in \mathcal{T}$ is the number of connected components of $\mathcal{T} \setminus \{a\}$. Vertices with multiplicity one are called leaves (we denote the set of leaves of T by $\mathrm{lf}(T)$).

We are interested in building real trees from continuous functions, as done for discrete trees via their contour functions. Therefore we consider a continuous excursion $g : \mathbb{R}_+ \to$

$\mathbb{R}_+$, that is, a continuous function not identically 0 such that $g(0) = 0$. Let us introduce the notation

$$\mathscr{e} = \{g : \mathbb{R}_+ \to \mathbb{R}_+ : g \text{ is a continuous excursion}\}$$

and define for each $g \in \mathscr{e}$ its *excursion length*

$$\eta(g) = \sup\{s > 0; g(s) > 0\}.$$

For $s, t \geq 0$ we define the pseudometric

$$d'_g(s,t) = g(s) + g(t) - 2m_g(s,t) \tag{16}$$

where $m_g(s,t) = \inf\{g(r); r \in [s \wedge t, s \vee t]\}$.

Introducing the equivalence relation $s \sim_g t$ iff $d'_g(s,t) = 0$, we then define the quotient space

$$\mathcal{T}_g = \mathbb{R}_+ / \sim_g . \tag{17}$$

Note that d'_g now induces a metric in $\mathcal{T}_g$, that we will denote by d_g.

Finally denote by $p_g : \mathbb{R}_+ \to \mathcal{T}_g$ the canonical projection. Intuitively, $\mathcal{T}_g$ is constructed by gluing the parts of the graph of g that can be joined by a line segment parallel to the x axis that does not pass through the graph of g. This idea is illustrated by the figure 2.1.

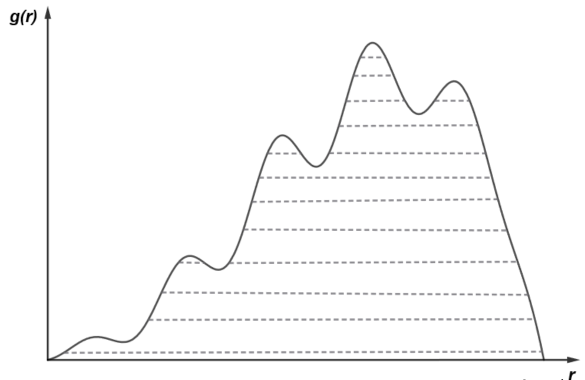

Figure 2.1: Intuitive construction of $\mathcal{T}_g$.

Note that when $\mathbb{R}_+$ is equipped with the Euclidean metric and $\mathcal{T}_g$ with the metric d_g, the map p_g is continuous.

Henceforth consider that g is always a compactly supported excursion, that is $\eta(g) < \infty$.

Theorem 2.4. The metric space $(\mathcal{T}_g, d_g, \rho_g)$ is a compact rooted real tree with $\rho_g = p_g(0)$.

In fact, any rooted compact real tree can be represented in the form $\mathcal{T}_g$.

Remark 2.5. To prove this theorem, it is useful to first understand intuitively the construction of a subtree of $\mathcal{T}_g$ given the graph of the excursion g. In order to do this, let us take a look at Figure 2.2, that shows the construction of a subtree of $\mathcal{T}_g$ generated by the points $p_g(s)$, $p_g(t)$, $p_g(u)$ and $p_g(v)$, that is, the union of the ancestral lines of these points in $\mathcal{T}_g$. Note that this subtree corresponds to the union of the bold segments that are constructed from the graph of g as showed on the figure. We can observe that the ancestral line of $p_g(x)$ is a line segment of length $g(x)$ for every $x \in \{s,t,u,v\}$. The intersection between the ancestors lines of $p_g(s)$ and $p_g(t)$ has length $m_g(s,t)$, and a similar property holds for any other pair of values, so the distance $d_g(p_g(s), p_g(t)) = d'_g(s,t)$ is the length of the path between $p_g(s)$ and $p_g(t)$.

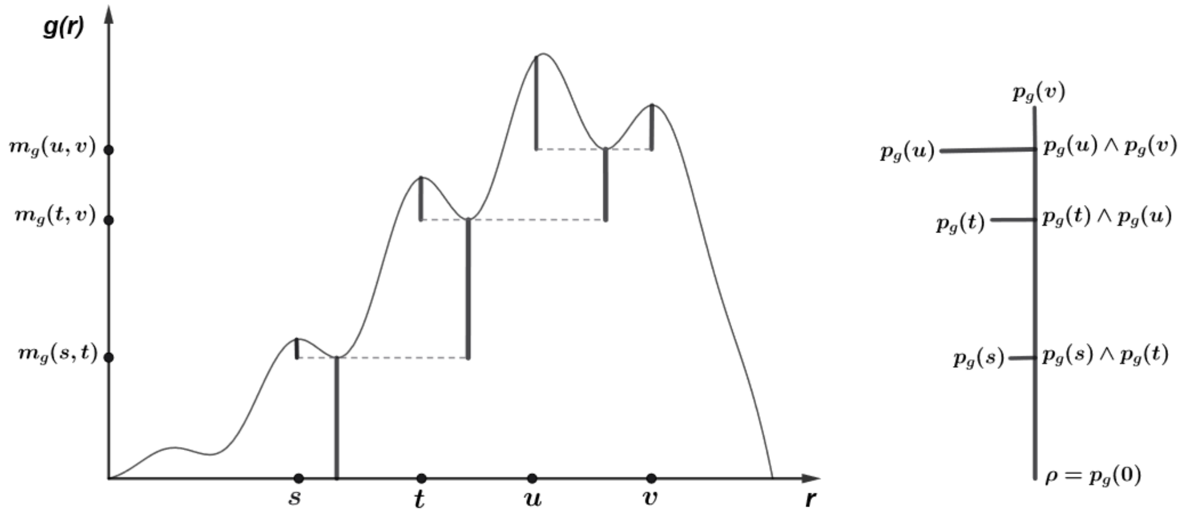

Figure 2.2: Construction of a subtree of $\mathcal{T}_g$.

Given that intuition let us proceed to the proof of the theorem.
Proof: We begin by stating and proving a root change lemma.

Lemma 2.6. Let $0 \leq s_0 < \eta(g)$. For any $r \in \mathbb{R}_+$, denote by $\bar{r}$ the unique element of $[0, \eta(g))$ such that $r - \bar{r} = k \cdot \eta(g)$ for some $k \in \mathbb{N}_)$. Set

$$g'(s) = \begin{cases} g(s_0) + g(\overline{s_0 + s}) - 2m_g(s_0, \overline{s_0 + s}) & \text{, if } s \in [0, \eta(g)] \\ 0 & \text{, if } s > \eta(g). \end{cases}$$

Then, the function g' is a compactly supported excursion, so we can define $\mathcal{T}_{g'}$. Furthermore, for every $s, t \in [0, \eta(g)]$, we have

$$d'_{g'}(s,t) = d'_g(\overline{s_0 + s}, \overline{s_0 + t}) \tag{18}$$

and there exists a isometry R from $\mathcal{T}_{g'}$ onto $\mathcal{T}_g$ such that, for every $s \in [0, \eta(g)]$

$$R(p_{g'}(s)) = p_g(\overline{s_0 + s}). \tag{19}$$

Assuming that the theorem 2.4 is proved, we see that $\mathcal{T}_{g'}$ coincides with the real tree $\mathcal{T}_g$ re-rooted at $p_g(s_0)$. Thus the lemma tells us that the function g' codes $\mathcal{T}_g$ re-rooted at an arbitrary vertex.

Proof:
We begin by proving relation (18). If $s < t$ are elements of $[0, \eta(g) - s_0]$, then $s_0 + s \in [0, \eta(g))$, so $\overline{s_0 + s} = s_0 + s$, and the same holds for $s_0 + t$. Moreover we have two possibilities for the value of $m_g(s_0 + s, s_0 + t)$:

1. $m_g(s_0 + s, s_0 + t) \geq m_g(s_0, s_0 + s)$.

2. $m_g(s_0 + s, s_0 + t) < m_g(s_0, s_0 + s)$.

In the first case, for every $r \in [s, t]$, $m_g(s_0, s_0 + r) = m_g(s_0, s_0 + s)$ as the intervals in which the minimum is taken are increasing and we suppose that the minimum of g on $[s_0, s_0 + t]$ is achieved in $[s_0, s_0 + s]$. So the minimum of g' on $[s, t]$ is

$$m_{g'}(s,t) = g(s_0) + m_g(s + s_0, s_0 + t) - 2m_g(s_0, s_0 + s)$$

it follows that

$$\begin{aligned}
d'_{g'}(s,t) &= g'(s) + g'(t) - 2m_{g'}(s,t) \\
&= 2g(s_0) + g(s_0+s) - 2m_g(s_0, s_0+s) + g(s_0+t) - 2m_g(s_0, s_0+t) - 2m_{g'}(s,t) \\
&= g(s_0+s) - 2m_g(s_0, s_0+s) + g(s_0+t) - 2m_g(s_0, s_0+t) \\
&\qquad - 2(m_g(s_0+s, s_0+t) - 2m_g(s_0, s_0+s)) \\
&= g(s_0+s) + g(s_0+t) - 2m_g(s_0+s, s_o+t) \\
&= d'_g(s_0+s, s_0+t) = d'_g(\overline{s_0+s}, \overline{s_0+t}).
\end{aligned}$$

In the second case, the minimum in the definition of $m_{g'}(s,t)$ is attained at

$$r_1 = \inf\{r \in [s,t] : g(s_0+r) \leq m_g(s_0, s_0+s)\}.$$

Indeed, for $r \in [s, r_1]$,

$$m_g(s_0, s_0+s) \geq m_g(s_0, s_0+r) \geq m_g(s_0, s_0+r_1),$$

so $m_g(s_0, s_0+r) = m_g(s_0, s_0+s)$ and $g(s_0+r) \geq g(s_0+r_1)$. Moreover for $r \in [r_1, t]$,

$$g(s_0+r) - 2m_g(s_0, s_0+r) \geq -m_g(s_0, s_0+r) \geq -m_g(s_0, s_0+r_1)$$

and $m_g(s_0, s_0+r_1) = g(s_0+r_1)$. Then $g'(r) \geq g'(r_1)$ for every $r \in [s,t]$.

Therefore

$$m_{g'}(s,t) = g(s_0) + g(s_0+r_1) - 2m_g(s_0, s_0+r_1) = g(s_0) - m_g(s_0, s_0+s)$$

and, using that

$$m_g(s_0+s, s_0+t) = m_g(s_0, s_0+t),$$

$$\begin{aligned}
d'_{g'}(s,t) &= g(s_0+s) - 2m_g(s_0, s_0+s) + g(s_0+t) - 2m_g(s_0, s_0+t) + 2m_g(s_0, s_0+s) \\
&= g(s_0+s) + g(s_0+t) - 2m_g(s_0+s, s_0+t) \\
&= d'_g(s_0+s, s_0+t) = d'_g(\overline{s_0+s}, \overline{s_0+t}).
\end{aligned}$$

The other cases are treated in a analogous way, just replacing s_0+s and s_0+t by the values $\overline{s_0+s}$ or $\overline{s_0+t}$.

By the equation (18), if $s,t \in [0, \eta(g)]$ are such that $d'_{g'}(s,t) = 0$, then $d'_g(\overline{s_0+s}, \overline{s_0+t}) = 0$, implying that $p_g(\overline{s_0+s}) = p_g(\overline{s_0+t})$.

As $\eta(g') \leq \eta(g)$ it follows that $\mathcal{T}_{g'} = p_{g'}([0, \eta(g)])$. So we can define R by (19). From this definition, it is immediate that R takes elements of $\mathcal{T}_{g'}$ onto $\mathcal{T}_g$, and, by (18), R is an isometry. ■

We begin the proof of the theorem with some preliminaries definitions. Note that assuming the result of this theorem, by Remark 2.5, each of these notations makes sense with our previous definitions for vertices of a real tree. Therefore we will maintain the same notation although they are initially defined in different context.

For $\sigma, \sigma' \in \mathcal{T}_g$ we define

- $\sigma \preceq \sigma'$ iff $d_g(\rho, \sigma') = d_g(\rho, \sigma) + d_g(\sigma, \sigma')$. This defines a partial order on $\mathcal{T}_g$.

  If $\sigma = p_g(s)$ and $\sigma' = p_g(t)$ then $\sigma \preceq \sigma'$ iff $m_g(s,t) = g(s)$.

- $[\![\sigma', \sigma]\!] = \{\sigma'' \in \mathcal{T}_g : d_g(\sigma', \sigma) = d_g(\sigma', \sigma'') + d_g(\sigma'', \sigma)\}$.
  It follows that if $\sigma = p_g(s)$ and $\sigma' = p_g(t)$
  $$[\![\rho, \sigma]\!] \cap [\![\rho, \sigma']\!] = [\![\rho, \gamma]\!]$$
  where $\gamma = p_g(r)$, for $r \in \{u \in [s,t] : g(u) = m_g(t,s)\}$. We then put $\gamma = \sigma \wedge \sigma'$.
- $\mathcal{T}_g[\sigma] = \{\sigma' \in \mathcal{T}_g : \sigma \preceq \sigma'\}$, the "offspring" of σ.
  If $\mathcal{T}_g[\sigma] \neq \{\sigma\}$ and $\sigma \neq \rho$, then $\mathcal{T}_g \setminus \mathcal{T}_g[\sigma]$ and $\mathcal{T}_g[\sigma] \setminus \{\sigma\}$ are two nonempty disjoint open sets.

The statements in the first two definitions follow from the construction of $\mathcal{T}_g$ and from the ideas given in remark 2.5.

To see that $\mathcal{T}_g \setminus \mathcal{T}_g[\sigma]$ is open, let s be such that $p_g(s) = \sigma$ and note that $\mathcal{T}_g[\sigma]$ is the image under p_g (a continuous function) of the compact set
$$\{u \in [0, \eta(g)] : m_g(s,u) = g(s)\}.$$
The set $\mathcal{T}_g[\sigma] \setminus \{\sigma\}$ is open because if $\sigma' \in \mathcal{T}_g[\sigma]$ and $\sigma' \neq \sigma$, it follows that the open ball centered at σ' with radius $d_g(\sigma, \sigma')$ is contained in $\mathcal{T}_g[\sigma] \setminus \{\sigma\}$.

We now prove property (i) of the definition of a real tree. So fixed σ_1 and σ_2 in $\mathcal{T}_g$ let us prove the existence and uniqueness of the mapping f_{σ_1,σ_2}. Using Lemma 2.6 with s_0 such that $p_g(s_0) = \sigma_1$, we may assume without loss of generality that $\sigma_1 = \rho$.

If $\sigma \in \mathcal{T}_g$ is fixed, we will prove that there exists a unique isometric mapping $f = f_{\rho,\sigma}$: $[0, d_g(\rho, \sigma)] \to \mathcal{T}_g$ such that $f(0) = \rho$ and $f(d_g(\rho, \sigma)) = \sigma$.

Let $s \in p_g^{-1}(\{\sigma\})$, so that $g(s) = d_g(\rho, \sigma)$. Then, for every $a \in [0, d_g(\rho, \sigma)]$ define
$$v(a) = \inf\{r \in [0,s] : m_g(r,s) = a\}.$$
Note that $g(v(a)) = a$ because as $0 = g(0) \leq a \leq g(s)$, by continuity of g, this infimum is achieved in some element of $\{t \in [0,s] : g(t) = a\}$. We define
$$f(a) = p_g(v(a)). \tag{20}$$

Let us verify the properties of f. Observe that $v(0) = 0$ so that $f(0) = \rho$, and, as $d_g(\rho, \sigma) = g(s)$ and $d'_g\big(v(g(s)), s\big) = 0$ it follows that $f(d_g(\rho, \sigma)) = p_g\big(v(g(s))\big) = p_g(s) = \sigma$. Now we verify that f is an isometry: if $a, b \in [0, d_g(\rho, \sigma)]$ with $a \leq b$, by definition $m_g(v(a), v(b)) = a$, and
$$d_g(f(a), f(b)) = g(v(b)) + g(v(a)) - 2a = b - a.$$
Note that the property (i) of the definition of a real tree implies that $[\![\rho, \sigma]\!] \supset f([0, d_g(\rho, \sigma)])$. Indeed, for $a \in [0, d_g(\rho, \sigma)]$
$$d_g(f(a), \sigma) = d_g(f(a), f(d_g(\rho, \sigma))) = d_g(\rho, \sigma) - a = d_g(\rho, \sigma) - d_g(\rho, f(a)).$$
then $f(a) \preceq \sigma$. On the other hand, if $\eta \preceq \sigma$ we saw in the construction of f by the equation (20) that $\eta = f(d_g(\rho, \eta))$. It follows that $f([0, d_g(\rho, \sigma)]) = [\![\rho, \sigma]\!]$.

To get uniqueness, suppose $\tilde{f}$ is another such isometry. For $a \in [0, d_g(\rho, \sigma)]$, let t be such that $p_g(t) = \tilde{f}(a)$. The properties of $\tilde{f}$ imply that $\tilde{f}(a) \preceq \sigma$, so, as $\sigma = p_g(s)$ and $\tilde{f}(s) = p_g(t)$, $g(t) = m_g(t,s)$. Moreover
$$g(t) = d_g(\rho, p_g(t)) = d_g(\tilde{f}(0), \tilde{f}(a)) = a.$$

Then $t \geq v(a) = v(g(t))$. On the other hand we have that

$$a = g(v(a)) = m_g(v(a), s)$$

then, as $[v(a), s] \subset [t, s]$,

$$a = g(t) = m_g(t, s) \geq m_g(v(a), s) = g(v(a)) = a$$

so $d'_g(t, v(a)) = 0$, and $\tilde{f}(a) = p_g(t) = p_g(v(a)) = f(a)$ proving uniqueness of f.

Now we turn to the property (ii) of the definition of a real tree. Let $q : [0, 1] \to \mathcal{T}_g$ be a continuous injective mapping, and we want to prove that:

$$q([0, 1]) = f_{q(0),q(1)}([0, d_g(q(0), q(1))]).$$

From Lemma 2.6 again, assume without loss of generality $q(0) = \rho$ and set $\sigma = q(1)$. We already have noticed that $f_{\rho,\sigma}([0, d_g(\rho, \sigma)]) = [\![\rho, \sigma]\!]$.

Assume for contradiction that $[\![\rho, \sigma]\!] \not\subset q([0, 1])$. Then let $\eta \in [\![\rho, \sigma]\!] \setminus q([0, 1])$ be such that $\eta \neq \rho, \sigma$. Then $q([0, 1])$ is contained in the union of two disjoint open sets $\mathcal{T}_g \setminus \mathcal{T}_g[\eta]$ and $\mathcal{T}_g[\eta] \setminus \{\eta\}$, with $q(0) = \rho \in \mathcal{T}_g \setminus \mathcal{T}_g[\eta]$ and $q(1) = \sigma \in \mathcal{T}_g[\eta] \setminus \{\eta\}$. This contradicts the fact that $q([0, 1])$ is connected.

Conversely, for $a \in (0, 1)$ set $\eta = q(a)$ and let $\gamma = \sigma \wedge \eta$. Note that

$$d_g(\eta, \sigma) = d_g(\eta, \gamma) + d_g(\gamma, \sigma)$$

then $\gamma \in [\![\rho, \eta]\!] \cap [\![\rho, \sigma]\!]$. So from the first part of the proof of property (ii), $\gamma \in q([0, a])$ and, by changing the root to η, $\gamma \in q([a, 1])$. Since q is injective, this implies that $\gamma = \eta = q(a)$, and $\eta \in [\![\rho, \sigma]\!]$. Hence, $q([0, 1]) = [\![\rho, \sigma]\!]$. ■

2.2 Gromov-Hausdorff distance.

We want to study the convergence of compact real trees, so it is interesting to define a notion of distance between two compact metric spaces. Fist we introduce a concept of equivalence between rooted metric spaces.

Definition 2.7. We say that two rooted compact metric spaces $(\mathcal{T}_1, d_1, \rho_1)$ and $(\mathcal{T}_2, d_2, \rho_2)$ are *equivalent* if there is a root-preserving isometry that maps $\mathcal{T}_1$ into $\mathcal{T}_2$.

Let us denote by $\mathbb{T}$ the set of all equivalence classes of compact rooted real trees.

In this work we will use the Gromov-Hausdorff distance between (equivalent classes of) compact metric spaces, that was first introduced by Gromov (see for example [12]) in view of geometric applications. We begin by introducing a distance between closed subsets of a bounded metric space.

Definition 2.8. For a bounded metric space (M, d), the *Hausdorff metric* between closed subsets of M is given by

$$d_H(A, B) = \inf\{\varepsilon > 0; A \subset U_\varepsilon(B) \text{ and } B \subset U_\varepsilon(A)\}$$

where $U_\varepsilon(A) := \{x \in M; d(x, A) < \varepsilon\}$.

Now we introduce a notion of distance between two different compact metric spaces.

Definition 2.9. For $\mathcal{T}$ and $\mathcal{T}'$ two rooted compact metric spaces, with respective roots ρ and ρ', we define the pointed *Gromov-Hausdorff distance* by

$$d_{GH}(\mathcal{T},\mathcal{T}') = \inf\{d_H(\varphi(\mathcal{T}),\varphi(\mathcal{T}')) \vee d(\varphi(\rho),\varphi(\rho'))\} \tag{21}$$

where the infimum is over all choices of a metric space (M, d) and all isometric embeddings $\varphi : \mathcal{T} \to M$ and $\varphi' : \mathcal{T}' \to M$ of $\mathcal{T}$ and $\mathcal{T}'$ into (M, d).

This distance depends only on the equivalent classes of $\mathcal{T}$ and $\mathcal{T}'$, so it defines a metric on $\mathbb{T}$, the set of all equivalent classes of rooted compact metric spaces (see [10, Theorem 7.3.30]). To illustrate this definition we give an example on the calculation of the Gromov-Hausdorff distance between two spaces.

Example 2.10. Let P be a metric space consisting of one point and $\mathcal{X} = (X, d, \rho)$ be any compact pointed metric space. Then we have that $d_{G,H}(P,\mathcal{X}) = \frac{\text{diam}(X)}{2}$.

Indeed, if we denote by $p \in X$ the point such that $d(p, \partial X) = \frac{\text{diam}(X)}{2}$ and we define $\varphi : P \to X$ by $\varphi(P) = p$, then $d(\rho, p) \le \frac{\text{diam}(X)}{2}$ and $d_H(X, p) \le \frac{\text{diam}(X)}{2}$, implying that $d_{G,H}(P,\mathcal{X}) \le \frac{\text{diam}(X)}{2}$.

We can also note that if φ' is any other embedding of P into X, then $d_H(X, \varphi'(P)) > \frac{\text{diam}(X)}{2}$, so the infimum on the definition 2.9 is achieved exactly at $\frac{\text{diam}(X)}{2}$.

An alternative definition for the Gromov-Hausdorff distance is given below.

Theorem 2.11. Let $(\mathcal{T}, d, \rho)$ and $(\mathcal{T}', d', \rho')$ be two rooted compact metric spaces.

We call $\mathcal{R} \subseteq \mathcal{T} \times \mathcal{T}'$ a *correspondence* between $\mathcal{T}$ and $\mathcal{T}'$ if, for all $x \in \mathcal{T}$, there exists at least one $x' \in \mathcal{T}'$ such that $(x, x') \in \mathcal{R}$ and, for every $y' \in \mathcal{T}'$, there exists at least one $y \in \mathcal{T}$ such that $(y, y') \in \mathcal{R}$. The *distortion of the correspondence* $\mathcal{R}$ is defined by:

$$\text{dis}(\mathcal{R}) = \sup\{|d(x, y) - d'(x', y')| : (x, x'), (y, y') \in \mathcal{R}\}$$

Let $\mathcal{C}(\mathcal{T},\mathcal{T}')$ denote the set of all correspondences between $\mathcal{T}$ and $\mathcal{T}'$. Then we have:

$$d_{GH}(\mathcal{T},\mathcal{T}') = \frac{1}{2}\inf\{\text{dis}(\mathcal{R}); \mathcal{R} \in \mathcal{C}(\mathcal{T},\mathcal{T}'), (\rho, \rho') \in \mathcal{R}\}. \tag{22}$$

The proof of this theorem can be found in [10, Theorem 7.3.25].

An important result is that $\mathbb{T}$ equipped with the Gromov-Hausdorff distance is a Polish space.

Theorem 2.12. The metric space $(\mathbb{T}, d_{GH})$ is complete and separable.

Despite the probabilistic importance of such topological property we will not use this theorem. So we give a reference for its proof in [23, Theorem 1]. Now that we have introduced a distance between the (equivalent classes of) rooted compact real trees as a subset of $(\mathbb{T}, d_{GH})$, we want to understand how this notion relates to the coding of these trees by continuous excursions. Fortunately, by the following lemma, there is a strong relationship between these concepts, which allows us to translate the results of chapter 1 to the context of Gromov-Hausdorff convergence of compact metric spaces.

Lemma 2.13. If g and g' are two continuous excursions with compact support then

$$d_{GH}(\mathcal{T}_g, \mathcal{T}_{g'}) \le 2\|g - g'\|,$$

where $\|g - g'\|$ is the uniform norm of $g - g'$.

Proof: Let us construct a correspondence between $\mathcal{T}_g$ and $T_{g'}$ by setting

$$\mathcal{R} = \{(\sigma, \sigma'); \sigma = p_g(t) \text{ and } \sigma' = p_{g'}(t) \text{ for some } t \geq 0\}.$$

Now we will bound the distortion of $\mathcal{R}$. Take (σ, σ') and (η, η') in $\mathcal{R}$. By the definition of $\mathcal{R}$ there exists $s, t \geq 0$ such that $p_g(s) = \sigma, p_{g'}(s) = \sigma'$ and $p_g(t) = \eta, p_{g'}(t) = \eta'$. So, by equation (16):

$$|d_g(\sigma, \eta) - d_g(\sigma', \eta')| \leq |g(s) - g'(s)| + |g(t) - g'(t)| + 2|m_g(s,t) - m_{g'}(s,t)| \leq 4\|g - g'\|.$$

Therefore $\text{dis}(\mathcal{R}) \leq 4\|g - g'\|$, and the result follows from the definition of Gromov-Hausdorff distance by correspondences in equation (22). ■

We could have proved theorem 2.4 using the previous proposition and theorem 2.12, by using that each compactly supported continuous excursion can be approximated by a rescaled contour of a discrete tree, but the chosen proof gives us a better understanding of the construction of $\mathcal{T}_g$.

Now we introduce the most classical example of a compact real tree. Recall the definition of Brownian excursion on example 1.17.

Definition 2.14. The *continuum random tree* (CRT) is the random tree $\mathcal{T}_e$ coded by the normalized Brownian excursion $e = (e_t)_{t \in [0,1]}$.

By lemma 2.13 $g \mapsto \mathcal{T}_g$ is a continuous application, therefore it is measurable. Then the CRT is a random variable taking values on $\mathbb{T}$.

Now we want to relate example 1.17 to a weak convergence result on $\mathbb{T}$. In order to do this, we must see each element of $t \in \mathbb{A}$, the set of all finite rooted ordered trees, as a rooted real tree. One way to do this is by thinking of t as the union of line segments of length 1, as represented by the left part of figure 1.1, equipped with the distance of the shortest path in the tree. Alternatively, if $C = (C_t)_{t \in \mathbb{R}_+}$ is the contour function of t, we can identify t as $\mathcal{T}_C$.

Theorem 2.15. For every $n \in \mathbb{N}$, let $\mathcal{T}_{(n)}$ be a random tree distributed uniformly over $\mathbb{A}_n$. Then $\frac{1}{\sqrt{2n}}\mathcal{T}_{(n)}$ converges in distribution to the CRT $\mathcal{T}_e$ in the space $\mathbb{T}$.

Proof: As proved in example 1.17, the contour functions $C_t^{(n)}$ of $\mathcal{T}_{(n)}$ converge weakly, after appropriate rescaling, to the normalized Brownian excursion $(e_t)_{t \in [0,1]}$. Thus, denoting

$$\tilde{C}_t^{(n)} = \frac{1}{\sqrt{2n}} C_{2nt}^{(n)}, \quad t \geq 0,$$

it follows that $\frac{1}{\sqrt{2n}}\mathcal{T}_{(n)}$ (that is, the tree $\mathcal{T}_{(n)}$ with all distances multiplied by $\frac{1}{\sqrt{2n}}$) has the same distribution as $\mathcal{T}_{\tilde{C}^{(n)}}$. So the result follows from lemma 2.13. ■

In the next section, we introduce another type of metric between metric spaces, but now they carry additional information in the form of a measure.

2.3 Gromov-Hausdorff vague topology

Henceforth we begin to work in rooted metric spaces with a measure. Unless said otherwise we will use [6] as reference.

Definition 2.16. A *rooted, complete, separable metric measure space* (X, d, ρ, μ) is a complete separable metric space (X, d) equipped with a distinguished point $\rho \in X$ (the root), and a Borel measure μ on X.

From now on we call it simply a metric measure space.

Similarly to definition 2.7 regarding rooted compact metric spaces, we have the following notion of equivalence of metric measure spaces. However, unlike that case, the equivalences now will depend only on the roots and on the support of the measures.

Definition 2.17. We say that two metric measure spaces $(X_1, d_1, \rho_1, \mu_1)$ and $(X_2, d_2, \rho_2, \mu_2)$ are *equivalent* ($(X_1, d_1, \rho_1, \mu_1) \simeq (X_2, d_2, \rho_2, \mu_2)$) if there is an isometry $\phi : \text{supp}(\mu_1) \cup \{\rho_1\} \to \text{supp}(\mu_2) \cup \{\rho_2\}$ such that

- $\phi(\rho_1) = \rho_2$
- $\phi_*\mu_1 = \mu_2$, where $\phi_*\mu = \mu \circ \phi^{-1}$ is the pushforward of μ under ϕ.

In general we will not distinguish between a metric measure space and its equivalence class.

Now we highlight a subclass of equivalent classes of metric measure spaces.

Definition 2.18. Let $\chi = (X, d, \rho, \mu)$ be a metric measure space. We say that χ is *boundedly finite* if μ is finite on all bounded subsets of X and $\rho \in \text{supp}(\mu)$. Noting that this property is invariant under equivalence of metric measure spaces, we can extend this definition to equivalence classes of metric spaces. Let us denote by $\mathbb{X}$ the set of boundedly finite equivalence classes of metric measure space.

Remark 2.19. In [6] the definition of $\mathbb{X}$ does not assume that the root is contained in the support of the measure. Since our examples always satisfy that assumption we will suppose that in order to simplify some notations.

Definition 2.20. An equivalence class of boundedly finite metric measure spaces is called a *Heine-Borel locally finite measure space* if it contains a representative $\chi' = (X', d', \rho', \mu')$ such that (X', d') is a Heine-Borel space, that is, a metric space in which every bounded closed set is compact. Let $\mathbb{X}_{\text{HB}}$ be the subspace of Heine-Borel spaces in $\mathbb{X}$.

As in the previous section, we want to introduce a useful notion of convergence in these classes of metric measure spaces. In order to do this, we will use again the notion of Gromov-Hausdorff metric together with a metric between the measures. The latter is given below.

Definition 2.21. Fix a metric space (X, d). We equip the space of all finite measures on $(X, \mathcal{B}(X))$, with the *Prohorov metric*, which is given by

$$d_{Pr}^{(X,d)}(\mu, \mu') = \inf\{\varepsilon > 0 : \mu(A) \leq \mu'(A^\varepsilon) + \varepsilon, \mu'(A) \leq \mu(A^\varepsilon) + \varepsilon \;\; \forall A \text{ closed}\}$$

where $A^\varepsilon = \{x : d(x, A) \leq \varepsilon\}$ is the closed ε-neighborhood of A.

This metric induces weak convergence, which we will denote by $\underset{n\to\infty}{\Longrightarrow}$.

Given a rooted metric measure space (X, d, ρ, μ), denote by $B_d(x, R)$ and $\bar{B}_d(x, R)$ respectively the open and closed balls induced by the distance d centered in $x \in X$ with radius $R \geq 0$. We will denote by $\mu\big|_R(\cdot) = \mu(\cdot \cap \bar{B}_d(\rho, R))$, that is, the restriction of μ to $\bar{B}_d(\rho, R)$.

Definition 2.22. For each $n \in \mathbb{N} \cup \{\infty\}$, let $\chi_n = (X_n, d_n, \rho_n, \mu_n) \in \mathbb{X}_{\text{HB}}$. We say that $(\chi_n)_{n\in\mathbb{N}}$ converges to χ_∞ in *Gromov-Hausdorff vague topology* if and only if there exists a rooted Heine-Borel space (E, d_E, ρ_E) and, for each $n \in \mathbb{N} \cup \{\infty\}$, isometries φ_n : $\text{supp}(\mu_n) \to E$ with $\varphi_n(\rho_n) = \rho_E$ such that, for all but countably $R \geq 0$,

$$((\varphi_n)_*\mu_n)\Big|_{\bar{B}_{d_E}(\rho_E,R)} \xRightarrow[n\to\infty]{} ((\varphi_\infty)_*\mu_\infty)\Big|_{\bar{B}_{d_E}(\rho_E,R)} \tag{23}$$

and additionally

$$d_H\Big(\varphi_n(\text{supp}(\mu_n)) \cap \bar{B}_{d_E}(\rho_E, R), \varphi_\infty(\text{supp}(\mu_\infty)) \cap \bar{B}_{d_E}(\rho_E, R)\Big) \xrightarrow{n\to\infty} 0. \tag{24}$$

Remark 2.23. In [6] a more general approach is considered for Gromov-Hausdorff vague convergence: it is defined for sequences in $\mathbb{X}$ via localization of another topology, the so-called Gromov-Hausdorff weak topology. The definition given here is equivalent to this one when restricted to $\mathbb{X}_{\text{HB}}$ (see [6, Definition 5.8 and Proposition 5.9]).

Remark 2.24. The space $\mathbb{X}_{\text{HB}}$ equipped with the topology generated by the Gromov-Hausdorff vague convergence is complete and separable (see [6, Proposition 5.12]).

In order to prove the next lemma we will introduce some theory about a topology in $\mathbb{X}$ that, unlike the Gromov-Hausdorff vague topology, ignores some information about the convergence of the support of the measures. Here we choose to define it via isometric embeddings.

Definition 2.25. For each $n \in \mathbb{N} \cup \{\infty\}$, let $\chi_n = (X_n, d_n, \rho_n, \mu_n) \in \mathbb{X}$.

We say that $(\chi_n)_{n\in\mathbb{N}}$ converges to χ_∞ in *Gromov vague topology* if and only if there exists a rooted, complete separable metric space (E, d_E, ρ_E) and, for each $n \in \mathbb{N} \cup \{\infty\}$, isometries $\varphi_n : \text{supp}(\mu_n) \to E$ with $\varphi_n(\rho_n) = \rho_E$ such that, for all but countably $R \geq 0$,

$$((\varphi_n)_*\mu_n)\Big|_{\bar{B}_{d_E}(\rho_E,R)} \xRightarrow[n\to\infty]{} ((\varphi_\infty)_*\mu_\infty)\Big|_{\bar{B}_{d_E}(\rho_E,R)}. \tag{25}$$

We highlight two properties of this topology that will be used in this work. Since there is a lot of theory behind these properties that will not be discussed here, we will just refer to a source for their proof.

Lemma 2.26. Consider $\chi = (X, d, \rho, \mu)$, $\chi_n = (X_n, d_n, \rho_n, \mu_n)$ elements of $\mathbb{X}$, and another boundedly finite measure μ'_n in X_n for each $n \in \mathbb{N}$. Assume that the following conditions are satisfied:

1. $\chi_n \xrightarrow[n\to\infty]{} \chi$ Gromov vaguely;
2. There exists a sequence $(R_k)_{k\in\mathbb{N}}$ such that $R_k \xrightarrow[k\to\infty]{} \infty$ and, for all $k \in \mathbb{N}$,

$$d_{Pr}^{(X_n,d_n)}\Big(\mu_n|_{R_k}, \mu'_n|_{R_k}\Big) \xrightarrow{n\to\infty} 0.$$

Then $(X_n, d_n, \rho_n, \mu'_n)$ converges Gromov vaguely to χ.

The proof of this lemma can be found in [6, lemma 2.9].

Definition 2.27. For $\delta, R > 0$ let us define $m_\delta^R : \mathbb{X} \to \mathbb{R}_+ \cup \{\infty\}$ as

$$m_\delta^R(\chi_n) = \inf\{\mu_n(\bar{B}_{d_n}\Big(x,\delta)\Big) : x \in \text{supp}(\mu_n) \cap \bar{B}_{d_n}(\rho_n, R)\}.$$

We say that $(\chi_n)_{n\in\mathbb{N}}$ a sequence on $\mathbb{X}$ satisfies the *local lower mass bound property* if for every $\delta > 0$ and $R > 0$

$$\liminf_{n\to\infty} m_\delta^R(\chi_n) > 0. \tag{26}$$

Finally, we state a relation between the Gromov-Hausdorff vague and the Gromov vague convergences.

Theorem 2.28. Consider $\chi = (X, d, \rho, \mu)$, $\chi_n = (X_n, d_n, \rho_n, \mu_n)$ elements of $\mathbb{X}_{\text{HB}}$. Then the following are equivalent

1. $\chi_n \xrightarrow[n\to\infty]{} \chi$ Gromov vaguely and it satisfies the *local lower mass bound property.*

2. $\chi_n \xrightarrow[n\to\infty]{} \chi$ Gromov-Hausdorff vaguely.

The proof of this proposition can be found in [6, corollary 6.2].

Using this relation we prove the perturbation of measure property for the Gromov-Hausdorff vague topology.

Lemma 2.29. Consider $\chi = (X, d, \rho, \mu)$, $\chi_n = (X_n, d_n, \rho_n, \mu_n)$ elements of $\mathbb{X}_{\text{HB}}$, and another boundedly finite measure μ'_n in X_n for each $n \in \mathbb{N}$. Assume that the following conditions are satisfied:

1. $\chi_n \xrightarrow[n\to\infty]{} \chi$ Gromov-Hausdorff-vaguely;

2. There exists a sequence $(R_k)_{k\in\mathbb{N}}$ such that $R_k \xrightarrow[k\to\infty]{} \infty$ and, for all $k \in \mathbb{N}$,

$$d_H^{(X_n,d_n)}\Big(\text{supp}(\mu_n) \cap \bar{B}_{d_n}(\rho_n, R_k), \text{supp}(\mu'_n) \cap \bar{B}_{d_n}(\rho_n, R_k)\Big) \xrightarrow{n\to\infty} 0$$

and

$$d_{Pr}^{(X_n,d_n)}\Big(\mu_n|_{R_k}, \mu'_n|_{R_k}\Big) \xrightarrow{n\to\infty} 0.$$

Then $\chi'_n = (X_n, d_n, \rho_n, \mu'_n)$ converges Gromov-Hausdorff vaguely to χ.

Proof: Lemma 2.26 implies that χ'_n converges Gromov vaguely to χ. Therefore, by theorem 2.28 we only need to check that, for every $\delta > 0$ and $R > 0$, the sequence $(\chi'_n)_{n\in\mathbb{N}}$ satisfies the lower local mass bound property.

Note that it is sufficient to prove that $(\chi'_n)_{n\in\mathbb{N}}$ satisfies equation (26) for every $\delta > 0$ and for a sequence $R_k \xrightarrow[k\to\infty]{} \infty$. Indeed, if for every $R > 0$ we can take a $R_k > R$, then $m_\delta^R(\chi'_n) \geq m_\delta^{R_k}(\chi'_n)$ because

$$\{x : x \in \text{supp}(\mu'_n) \cap \bar{B}_{d_n}(\rho_n, R_k)\} \supset \{x : x \in \text{supp}(\mu'_n) \cap \bar{B}_{d_n}(\rho_n, R)\}.$$

Now fix $\delta > 0$, $k \in \mathbb{N}$ and $0 < \eta < (\delta \wedge \varepsilon_n/2$, where $\varepsilon_n = m_{\delta/2}^{R_k}(\chi_n) > 0$. By our assumptions and the definition of Hausdorff distance, there exists $N_1 \in \mathbb{N}$ such that for every $n > N_1$ it follows that

$$\text{supp}(\mu'_n) \cap \bar{B}_{d_n}(\rho_n, R_k) \subset \Big(\text{supp}(\mu_n) \cap \bar{B}_{d_n}(\rho_n, R_k)\Big)^{\frac{\delta}{2}-\eta}.$$

Then for every $x \in \text{supp}(\mu'_n) \cap \bar{B}_{d_n}(\rho_n, R_k)$, there exist some $y \in \text{supp}(\mu_n) \cap \bar{B}_{d_n}(\rho_n, R_k)$ such that $x \in \bar{B}_{d_n}(y, \frac{\delta}{2} - \eta)$. Thus

$$\mu'_n \left(\bar{B}_{d_n}(x, \delta)\right) \geq \mu'_n \left(\bar{B}_{d_n}(y, \delta/2 + \eta)\right),$$

by our assumption on the convergence of the Prohorov distance there exists $N_2 \in \mathbb{N}$ such that for every $n > N_2$ and $x \in \text{supp}(\mu'_n) \cap \bar{B}_{d_n}(\rho_n, R_k)$

$$\mu'_n \left(\bar{B}_{d_n}(y, \delta/2 + \eta)\right) \geq \mu_n(\bar{B}_{d_n}(y, \delta/2)) - \eta > \varepsilon_n/2$$

where the last inequality follows from our choice of ε. Hence, for $n > N_1 \vee N_2$

$$\mu'_n \left(\bar{B}_{d_n}(x, \delta)\right) \geq \varepsilon_n/2$$

then

$$m_\delta^{R_k}(\chi'_n) \geq \frac{1}{2} m_{\delta/2}^{R_k}(\chi_n).$$

Since $\chi_n \xrightarrow[n\to\infty]{} \chi$ Gromov-Hausdorff vaguely, theorem 2.28 implies that $(\chi_n)_{n\in\mathbb{N}}$ satisfy the local lower mass bound property, so, by taking $\limsup_{n\to\infty}$ in both sides of the last inequality we complete the proof.

■

Henceforth we return to work specifically on the space of real rooted trees. Consider t' a discrete tree with root ρ', and equip t' with the graph distance d', that is, the length of the shortest path. Then, as made in the comment before 2.15 (t', d', ρ') can be embedded isometrically into a complete, locally compact real tree (T, d, ρ) in a unique way, where T is obtained by "filling the edges" of t'. We denote the image of t' by $\text{nod}(T)$, so $\text{nod}(T)$ is a discrete subset of T, and we can identify t' with $\text{nod}(T_n)$. A natural measure on t' in the context of random walks is given below.

Definition 2.30. Consider a discrete tree (t', d', ρ') and the correspondent embedding into a complete, locally compact real tree (T, d, ρ).

The *degree measure* $\mu_{t'}^{\text{deg}}$ on $\text{nod}(T)$ is given by

$$\mu_T^{\text{deg}} = \frac{1}{2} \sum_{x \in \text{nod}(T)} \deg(x) \delta_x \tag{27}$$

where $\deg(x)$ is the multiplicity of the vertex x.

We could have defined $\mu_{t'}^{\text{deg}}$ on t' and considered μ_T^{deg} as the pushforward of the degree measure on t'. Note that in this case $(t', d', \rho', \mu_{t'}^{\text{deg}}) \simeq (T, d, \rho, \mu_T^{\text{deg}})$ (remember that, by definition 2.17, the equivalence only consider the support of the measures).

Another relevant measure on a real tree is defined below.

Definition 2.31. Consider a metric space (T, d) and define, for $S \subseteq T$ and $\delta > 0$,

$$H_\delta^1(S) = \inf\Big\{\sum_{i\in\mathbb{N}} \text{diam}(U_i) : \bigcup_{i\in\mathbb{N}} U_i \supseteq S,\ \text{diam}(U_i) < \delta\Big\},$$

where the infimum is over all countable covers of S by sets $U_i \subset T$ satisfying $\text{diam}(U_i) < \delta$.

As $H_\delta^1(S)$ is nonincreasing in δ, we consider

$$H^1(S) = \lim_{\delta \downarrow 0} H_\delta^1(S),$$

which is an outer measure on T. By Carathéodory's extension theorem, its restriction to the σ-algebra of Carathéodory-measurable sets is a measure. We call this measure the 1-*dimensional Hausdorff measure* on T.

The 1-Hausdorff measure on (T, d) coincides with our concept of length of the space, therefore it defines a natural measure on T. We will call it the *length measure* and denote it by $\lambda = \lambda_{(T,d)}$.

The Gromov-Hausdorff vague convergence give us a relation between these two measures.

Proposition 2.32. Consider $(t_n)_{n\in\mathbb{N}}$ a sequence of discrete rooted trees, and the corresponding sequence of rooted real trees (T_n, d_n, ρ_n) constructed as above. Assume that there exists two sequences of positive numbers $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ that converge to 0, such that

$$(T_n, \alpha_n d_n, \rho_n, \beta_n \lambda_{(T_n,d_n)}) \xrightarrow{n\to\infty} \chi$$

Gromov-Hausdorff vaguely for some real tree $\chi = (T, d, \rho, \mu) \in \mathbb{X}_{\mathrm{HB}}$. Then

$$(T_n, \alpha_n d_n, \rho_n, \beta_n \mu^{\deg}_{T_n}) \xrightarrow{n\to\infty} \chi$$

Gromov-Hausdorff vaguely.

Let us give an idea of the proof of the previous proposition. We have that $\mathrm{supp}(\mu^{\deg}_{T_n}) = \mathrm{nod}(T_n)$ and $\mathrm{supp}(\lambda_{(T_n,d_n)}) = T_n$. Since, by the embedding of a discrete tree into a real tree, the 1−neighborhood of the set $\mathrm{nod}(T_n)$ is T_n, we have that, for every $R > 0$,

$$d_H\Big(\mathrm{nod}(T_n) \cap \bar{B}_{\alpha_n d_n}(\rho_n, R), \bar{B}_{\alpha_n d_n}(\rho_n, R)\Big) \le \alpha_n \xrightarrow{n\to\infty} 0.$$

Furthermore, assuming that the diameter of T_n is smaller than R, it follows that

$$d_{Pr}^{(T_n,d_n)}(\beta_n \mu^{\deg}_{T_n}, \beta_n \lambda_{(T_n,d_n)}) \le \frac{1}{2}\alpha_n.$$

Indeed if we take any closed subset A of T_n in the definition of the Prohorov distance for each x in $\mathrm{nod}(T_n) \cap A$ the set $A^{\frac{1}{2}\alpha_n} = \{x \in T_n : \alpha_n d_n(x, A) \le \frac{\alpha_n}{2}\}$ will contain $\bar{B}_{\alpha_n d_n}(x, \alpha_n/2)$. Then, as $\lambda_{(T_n,d_n)}(\bar{B}_{\alpha_n d_n}(x, \alpha_n/2)) = \frac{\deg(x)}{2}$ and these balls are disjoint

$$\lambda_{(T_n,d_n)}(A^{\frac{1}{2}\alpha_n}) \ge \mu^{\deg}_{T_n}(A).$$

Moreover for each $x \in A \setminus \mathrm{nod}(T_n)$ there exists $u, v \in \mathrm{nod}(T_n)$ such that $x \in]\!]u, v[\![$, where $]\!]u, v[\![= [\![u, v]\!] \setminus \{u, v\}$. Then u or v must be contained in $A^{\frac{1}{2}\alpha_n}$ implying that

$$\mu^{\deg}_{T_n}(A^{\frac{1}{2}\alpha_n} \cap [\![u, v]\!]) = \mathbb{1}_{u\in A^{\alpha_n/2}} \frac{\deg(u)}{2} + \mathbb{1}_{v\in A^{\alpha_n/2}} \frac{\deg(v)}{2}$$

while

$$\lambda_{(T_n,d_n)}(A \cap [\![u, v]\!]) \le \mathbb{1}_{u\in A^{\alpha_n/2}} \frac{1}{2} + \mathbb{1}_{v\in A^{\alpha_n/2}} \frac{1}{2}.$$

It follows that

$$\mu^{\deg}_{T_n}(A^{\frac{1}{2}\alpha_n}) \ge \lambda_{(T_n,d_n)}(A).$$

In the general case, we have to take the boundary effects into account. So consider the annulus $S^\varepsilon(\rho_n, R) = \bar{B}_{\alpha_n d_n}(\rho_n, R + \frac{1}{2}\varepsilon) \setminus B_{\alpha_n d_n}(\rho_n, R - \frac{1}{2}\varepsilon)$. We have that

$$d_{Pr}^{(T_n,\alpha_n d_n)}\Big(\beta_n \mu^{\deg}_{T_n}\Big|_R, \beta_n \lambda_{(T_n,d_n)}\Big|_R\Big) \le \Big(\frac{1}{2}\alpha_n\Big) \vee \Big(\beta_n \lambda_{(T_n,d_n)}\Big(S^{\alpha_n}(\rho_n, R)\Big)\Big) \tag{28}$$

In order to give an idea of the proof of this inequality let us denote $B = \bar{B}_{d_n}(\rho_n, R)$ and, for A a closed subset of B consider the cases

1. $A \subset B_{\alpha_n d_n}(\rho_n, R - \frac{1}{2}\alpha_n)$,
2. $A \cap \mathrm{nod}(T_n) \setminus B_{\alpha_n d_n}(\rho_n, R - \frac{1}{2}\alpha_n) = \emptyset$,
3. $A \cap \mathrm{nod}(T_n) \setminus B_{\alpha_n d_n}(\rho_n, R - \frac{1}{2}\alpha_n) \neq \emptyset$.

In the first case, since $A^{\frac{1}{2}\alpha_n} \cap B = A^{\frac{1}{2}\alpha_n}$ we can apply the same ideas as before and show that $\mu^{\deg}_{T_n}(A^{\frac{1}{2}\alpha_n}) \geq \lambda_{(T_n,d_n)}(A)$ and $\lambda_{(T_n,d_n)}(A^{\frac{1}{2}\alpha_n}) \geq \mu^{\deg}_{T_n}(A)$.

In the second case every $x \in A \setminus B_{\alpha_n d_n}(\rho_n, R - \frac{1}{2}\alpha_n)$ is outside $\mathrm{nod}(T)$, so it may happen that $\beta_n \lambda_{(T_n,d_n)}(A \setminus B_{\alpha_n d_n}(\rho_n, R - \frac{1}{2}\alpha_n)) > \alpha_n/2$ while $\beta_n \mu^{\deg}_{T_n}(A^{\frac{1}{2}\alpha_n} \cap B) = 0$. But it is true that

$$\beta_n \mu^{\deg}_{T_n}\Big(A^{\beta_n \lambda_{(T_n,d_n)}\left(S^{\alpha_n}(\rho_n,R)\right)} \cap \Big(B_{\alpha_n d_n}(\rho_n, R - \alpha_n/2)\Big)^c \cap B\Big) + \beta_n \lambda_{(T_n,d_n)}\Big(S^{\alpha_n}(\rho_n, R)\Big) \geq \beta_n \lambda_{(T_n,d_n)}\Big(A \cap \Big(B_{\alpha_n d_n}(\rho_n, R - \alpha_n/2)\Big)^c \cap B\Big).$$

Similarly, in the third case it may happen that $\frac{\deg(x)}{2} > \lambda_{(T_n,d_n)}(\bar{B}_{d_n}(x, \alpha_n/2) \cap B)$ for some $x \in A \cap \mathrm{nod}(T_n) \setminus B_{\alpha_n d_n}(\rho_n, R - \frac{1}{2}\alpha_n)$, but we have that

$$\beta_n \lambda_{(T_n,d_n)}\Big(A^{\beta_n \lambda_{(T_n,d_n)}\left(S^{\alpha_n}(\rho_n,R)\right)} \cap \Big(B_{\alpha_n d_n}(\rho_n, R - \alpha_n/2)\Big)^c \cap B\Big) + \beta_n \lambda_{(T_n,d_n)}\Big(S^{\alpha_n}(\rho_n, R)\Big) \geq \beta_n \mu^{\deg}_{T_n}\Big(A \cap \Big(B_{\alpha_n d_n}(\rho_n, R - \alpha_n/2)\Big)^c \cap B\Big)$$

what give us equation (28).

By assumption $\beta_n \lambda_{(T_n,d_n)}\Big(S^{\alpha_n}(\rho_n, R)\Big) \xrightarrow{n\to\infty} 0$ in equation (28) for every $R > 0$ with $\mu(S(\rho, R)) = 0$. Therefore, by lemma 2.29 the result follows.

We again use the notations $\mathcal{T}_g, d'_g, d_g$ and p_g introduced in section 2.1 in order to build a real tree from a continuous excursion $g \in \mathscr{e}$. In the context of metric measure real trees we can make a similar construction and define the following:

Definition 2.33. The glue map G

$$G(g) = (\mathcal{T}_g, d_g, \rho_g, \mu_g) \tag{29}$$

sends an excursion to a complete, separable, rooted real tree where d_g, ρ_g, μ_g are the pushforwards of d'_g, the Lebesgue measure on $[0, \eta(g)]$ $(\lambda_{([0,\eta(g)])})$, and 0, respectively, under the canonical projection $p_g : \mathbb{R}_+ \to \mathcal{T}_g$.

Let us distinguish the elements $g \in \mathscr{e}$ by its excursion lengths. If $\eta(g) < \infty$ we say that g is *compactly supported*. If $\eta(g) = \infty$ and $\lim_{x\to\infty} g(x) = \infty$ we say that g is *transient*. In this case we define for $R > 0$ the last visit to height R by

$$\xi_g(R) = \sup\{t \geq 0 : g(t) \leq R\} < \infty. \tag{30}$$

Then the glue map acts on $\mathscr{e}$ as follows.

Lemma 2.34. Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous excursion.

1. If g is compactly supported, then $G(g)$ is a rooted compact finite measure real tree, in particular $G(g) \in \mathbb{X}_{\mathrm{HB}}$.

2. If g is transient, then $G(g) \in \mathbb{X}_{\mathrm{HB}}$.
3. If g is neither compactly supported nor transient, then $G(g) \notin \mathbb{X}_{\mathrm{HB}}$.

Proof:

1. Follows from the proof of theorem 2.4.
2. Assume that g is transient.

   Then for all $R > 0$, $\xi_g(R) < \infty$, and, by continuity of g, $A_R = \{s \in [0, \infty) : g(s) \leq R\}$ is a closed subset of $[0, \xi_g(R)]$, hence it is compact. As previously stated the continuity of g implies continuity of p_g. Therefore, by construction of $\mathcal{T}_g$, $\bar{B}_{d_g}(\rho_g, R) = p_g(A_R)$ is a compact subset of $\mathcal{T}_g$.

   Consequently, $\mu_g\Big(\bar{B}_{d_g}(\rho_g, R)\Big) \leq \lambda([0, \xi_g(R)]) = \xi_g(R) < \infty$, as μ_g is the pushforward of the Lebesgue measure on $\mathbb{R}_+$. Since any bounded closed subset of $(\mathcal{T}_g, d_g)$ is a closed subset of a closed ball $\bar{B}_{d_g}(\rho_g, R)$ for some $R > 0$, it is compact as well. Thus $G(g) \in \mathbb{X}_{\mathrm{HB}}$.

3. Assume that g is such that $\eta(g) = \infty$ but $a = \liminf_{t\to\infty} g(t) < \infty$, and define $b = \limsup_{t\to\infty} g(t)$.

   If $b > a$, $(\mathcal{T}_g, d_g)$ is not Heine-Borel (and therefore not locally compact). Indeed, there is an $\varepsilon > 0$ with $a + 3\varepsilon < b$, and an increasing sequence $(t_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_+$ with $g(t_n) \in [a + 2\varepsilon, a + 3\varepsilon]$ and $\inf_{u\in[t_n,t_{n+1}]} g(u) \leq a + \varepsilon$ for all $n \in \mathbb{N}$.

   Then the sequence $(x_n)_{n\in\mathbb{N}}$ defined by $x_n = p_g(t_n)$ defines a sequence of points in $\bar{B}_{d_g}(\rho_g, a + 3\varepsilon)$ with $d_g(x_i, x_j) = d'_g(g(t_i), g(t_j)) > \varepsilon$ for all $i \neq j$.

   Moreover if $b = a$,

   $$\mu_g\Big(\bar{B}_{d_g}(\rho_g, b + 1)\Big) = \lambda\Big(\{s \in \mathbb{R}_+ : g(s) \leq b + 1\}\Big) = \infty, \tag{31}$$

   which means that μ_g is not boundedly finite. In both cases $G(g) \notin \mathbb{X}_{\mathrm{HB}}$. ■

2.3.1 A transient excursion example.

Now we will focus on present a example related to a non-compact real tree in $\mathbb{X}_{\mathrm{HB}}$, therefore we will use a transient continuous excursion.

A interesting fact about the trees coded by these functions follows.

Proposition 2.35. Let g be a transient excursion. Then $G(g) \in \mathbb{X}_{\mathrm{HB}}$ is a real tree with exactly one end at infinity, i.e., there is a unique bijective isometry $\varphi : [0, \infty) \to \mathcal{T}_g$ with $\varphi(0) = \rho_g$.

Proof: We will show that $\varphi_g = p_g \circ \xi_g$, where ξ_g is defined as in equation (30). To prove that φ_g is a isometry we use again the observations in remark 2.5. So taking $r \leq t$ in $\mathbb{R}$

$$d_g(\varphi_g(r), \varphi_g(t)) = g(\xi_g(r)) + g(\xi_g(t)) - 2m_g(\xi_g(r), \xi_g(t)) = r + t - 2r = t - r$$

because $\xi_g(x) = x$ for every $x \in \mathbb{R}_+$, which also implies that $\varphi_g(0) = \rho_g$.

To prove uniqueness, assume that $\psi : [0, \infty) \to \mathcal{T}_g$ is another isometry that satisfies $\psi(0) = \rho_g$ and fix $R > 0$. Choose $t \in \mathbb{R}_+$ with $p_g(t) = \psi(R)$. As ψ is an isometry, we have that

$$d_g(p_g(t), \rho_g) = d_g(\psi(R), \psi(0)) = R$$

so $g(t) = R$, and then $t \leq \xi_g(R)$.

Now take $S > \sup_{u \in [0, \xi_g(R)]}\{g(u)\}$ and $s \in p_g^{-1}\Big(\psi(S)\Big)$. Note that $s > \xi_g(R)$ and $g(s) = S$, so

$$S - R = d_g\Big(\psi(S), \psi(R)\Big) = S + R - 2 \inf_{u \in [t,s]} g(u),$$

hence $\inf_{u \in [t, \xi_g(R)]}\{g(u)\} \geq \inf_{u \in [t,s]}\{g(u)\} = R$. This implies that

$$d_g\Big(\psi(R), \varphi_g(R)\Big) = d_g\Big(p_g(t), p_g(\xi_g(R))\Big) = 2R - 2 \inf_{u \in [t, \xi_g(R)]}\{g(u)\} = 0.$$

■

Consider $\mathscr{e}_\infty \subset \mathscr{e}$ as the set of continuous transient excursions. Then a result analogous to lemma 2.13 follows.

Proposition 2.36. The glue map $G : \mathscr{e}_\infty \to \mathbb{X}_{\text{HB}}$ is continuous if $\mathscr{e}_\infty$ is equipped with the topology of uniform convergence on compact sets, and $\mathbb{X}_{\text{HB}}$ with the Gromov-Hausdorff vague topology.

The proof of this theorem will be omitted as it involves some topics that were not discussed here. We refer to [6, Proposition 7.5] for a proof.

The remainder of this work will focus on showing convergences related to a specific class of examples of random locally compact real trees in $\mathbb{X}_{\text{HB}}$.

Definition 2.37. The Kallenberg-Kesten tree t^∞ is the tree associated to a critical μ-Galton-Watson process $(Z_n)_{n \in \mathbb{N}}$ with variance $0 < \sigma^2 < \infty$ conditioned to survive.

A more precise definition based on [16] follows.

For a discrete tree t, write t_k for the set of vertices of t that have height equal to k. Then, for $k \in \mathbb{N}_0$, denote by $\mathbb{T}_k$ the set of all rooted ordered trees t of k generations, that is, such that $\#t_k \neq 0$ but $\#t_{k+1} = 0$. And denote by $\mathbb{T}_\infty$ the set of infinite trees, that is, t such that $\#t_k \neq 0$ for all $k \in \mathbb{N}$.

Moreover, consider the function $T_{[k]} : \mathbb{T}_\infty \cup \bigcup_{n \geq k} \mathbb{T}_n \to \mathbb{T}_k$ such that $T_{[k]}(t)$ consists of the k first generations of the tree t.

Finally, let μ be a critical nondegenerate offspring distribution with finite variance, and consider $(Z_n)_{n \in \mathbb{N}_0}$ a μ-Galton-Watson process with tree representation θ. It follows from the branching property (remember definition 1.11) and the estimate in theorem 1.9 that, for $t \in \mathbb{T}_k$,

$$\lim_{n \to \infty} P(T_{[k]}(\theta) = t | Z_n \neq 0) = \#t_k P(T_{[k]}(\theta) = t).$$

Then, by linearity of the limit, $\nu(T_{[k]}(\theta) = t) = \#t_k P(T_{[k]}(\theta) = t)$ defines a probability measure on $\mathbb{T}_k$. By Kolmogorov's extension theorem, since the product measures $\prod_{k \in J} T_{[k]*}\nu$ for $J \subset \mathbb{N}$ satisfy the consistency condition (it is sufficient to check on the generating sets of the product σ-algebra), there exists a unique extension of $T_{[k]*}\nu$ to a probability measure on $\mathbb{T}_\infty$ the set of rooted ordered infinite discrete trees. Let us denote this extension by ν.

Then, ν is the distribution of the μ-Galton-Watson tree conditioned on no-extinction.

Remark 2.38. An alternative construction of t^∞, based on [16, Lemma 2.2] and [13, Section 5] is made as follows. Consider that we can classify the vertices of a tree into two types: the normal nodes and the special nodes, with the root being special. A normal node has offspring distribution μ, and all its children are normal, while the special node has offspring distribution

$$\hat{\mu}(k) = k\mu(k) \quad , \text{ for } k \in \mathbb{N}_0,$$

the *size-biased distribution of* μ. Moreover, we choose uniformly at random one of the children of the special node to be special, while the others are normal.

Since all the special nodes have exactly one special child, there exist, almost surely, a unique infinite path from the root, formed by the special vertices. Let us call this path the *spine* of the tree.

Note that the contour function of t^∞ is a transient excursion, so the last assertion agrees with proposition 2.35.

Now we intend to show Gromov-Hausdorff vague convergence of the rescaled Kallenberg-Kesten tree.

In order to do so, let us consider a Kallenberg-Kesten tree (t^∞, d', ρ') as defined in 2.37 where μ will be considered a $Geom(\frac{1}{2})$ distribution and d is the graph distance on t^∞. Henceforth we deal with the embedding of (t^∞, d', ρ') into a complete, locally compact rooted real tree $(\mathcal{T}^\infty, d, \rho)$. Then we state the following:

Theorem 2.39. Consider $(\mathcal{T}^\infty, d, \rho)$ a embedded μ-Kallenberg-Kesten tree with $\mu \sim Geom(\frac{1}{2})$ and $\tilde{B}_t = B_t - 2\inf_{s\in[0,t]}\{B_s\}$, where B is the standard Brownian motion on $\mathbb{R}$. Then

$$(\mathcal{T}^\infty, n^{-1}d, \rho, n^{-2}\lambda_{(\mathcal{T}^\infty,d)}) \xrightarrow[n\to\infty]{(d)} G(\tilde{B}),$$

$$(\mathcal{T}^\infty, n^{-1}d, \rho, n^{-2}\mu^{deg}_{\mathcal{T}^\infty}) \xrightarrow[n\to\infty]{(d)} G(\tilde{B})$$

Gromov-Hausdorff vaguely.

Proof: Let us consider a Kallenberg-Kesten tree (t^∞, d', ρ') as defined in 2.37 where μ will be considered a $Geom(\frac{1}{2})$ distribution and d is the graph distance on t^∞. Henceforth we deal with the embedding of (t^∞, d', ρ') into a complete, locally compact rooted real tree $(\mathcal{T}^\infty, d, \rho)$. Equip this space with the lengh measure $\lambda_{(\mathcal{T}^\infty,d)}$.

Based the construction of the Kallenberg-Kesten tree in remark 2.38 and on the behavior of the contour function of geometric critical Galton-Watson trees in example 1.15, it follows that $(\mathcal{T}^\infty, d, \rho, \lambda_{(\mathcal{T}^\infty,d)})$ has the same law as $G(\tilde{S})$, where, for $t \geq 0$, $\tilde{S}_t = S_t - 2\inf_{s\in[0,t]}\{S_s\}$, a process that is obtained from the path of S, the linear interpolation of a simple random walk on $\mathbb{Z}$, by reflecting this path at each time point in the level of its previous minimum. The relation between these two process is more clear if we note that the branches attached to the normal children of a special vertex behaves like a μ-Galton-Watson tree. Then, $\inf_{k\leq n} S_k$ coincides with the $\mathcal{E}_n - 1$ where $\mathcal{E}_n$ is the number of elements of the spine already visited by the particle that defines the contour process up to time n.

By Donsker's invariance theorem and the continuous mapping theorem,

$$(\tilde{S}^{(n)}_t)_{t\geq 0} = (n^{-1}\tilde{S}_{n^2 t})_{t\geq 0} \xrightarrow[n\to\infty]{(d)} (\tilde{B}_t)_{t\geq 0}$$

where, for $t \geq 0$, $\tilde{B}_t = B_t - 2\inf_{s\in[0,t]}\{B_s\}$, and B is the standard Brownian motion on $\mathbb{R}$. By [22, Theorem 1.3], $\tilde{B}$ is a three-dimensional Bessel process, defined as the process that coincides in law with the radial part of a three-dimensional Brownian motion.

From now on we refer to $G(\tilde{B})$ as the continuum Kallenberg-Kesten tree, χ^∞.

Because of the rescaling, the elements $G(\tilde{S}^{(n)}) = (\mathcal{T}_{\tilde{S}^{(n)}}, d_{\tilde{S}^{(n)}}, \rho_{\tilde{S}^{(n)}}, \lambda_{\tilde{S}^{(n)}})$ in the definition 2.33 can be written in terms of $d_{\tilde{S}}$, $\rho_{\tilde{S}}$ and $\lambda_{(\mathcal{T}_{\tilde{S}}, d_{\tilde{S}})}$ as $(\mathcal{T}_{\tilde{S}}, n^{-1}d_{\tilde{S}}, , \rho_{\tilde{S}}, n^{-2}\lambda_{(\mathcal{T}_{\tilde{S}}, d_{\tilde{S}})})$.

Hence by the previous observations and the continuity of the glue map

$$(\mathcal{T}^\infty, n^{-1}d, \rho, n^{-2}\lambda_{(T^\infty,d)}) \sim G(\tilde{S}^{(n)}) \xrightarrow[n\to\infty]{(d)} G(\tilde{B}) = \chi^\infty$$

Gromov-Hausdorff vaguely.

By proposition 2.32, this also implies

$$(\mathcal{T}^\infty, n^{-1}d, \rho, n^{-2}\mu^{deg}_{\mathcal{T}^\infty}) \xrightarrow[n\to\infty]{(d)} \chi^\infty$$

Gromov-Hausdorff vaguely.

That is, we conclude that if we take a discrete geometrical Kallenberg-Kesten tree with edge length rescaled to become n^{-1} and equip it with the measure that assigns mass $\frac{1}{2}n^{-2}\deg(x)$ to each vertex x, this rescaled discrete measure tree $(t^\infty, n^{-1}d, n^{-2}\mu^{\deg}_{t^\infty})$ converges weakly with respect to the Gromov-Hausdorff-vague topology to the continuum Kallenberg-Kesten tree. ■

2.4 Further related works.

The Gromov-Hausdorff vague convergence can be used to develop a great variety of works on real trees by treating them as the limit of a sequence of embedded discrete trees. Here we describe some of them. We begin by mentioning some works related to the convergence of random walks on random trees.

Definition 2.40. Let t a discrete tree in which all vertices have finite degree. We denote by $(X^T_n)_{n\in\mathbb{N}_0}$ the *neareast-neighbor random walk on* t, that is, the Markov chain on t with transition probabilities $p(x,y) = \begin{cases} \frac{1}{\deg(x)} & \text{if } x \text{ is connected to } y \text{ in } t, \\ 0 & \text{otherwise.} \end{cases}$

In our previous example, let us consider $(X_k)_{k\in\mathbb{N}_0}$ the random walk on $(\mathcal{T}^\infty, d, \rho, \mu^{\deg}_T)$ under its annealed distribution. That is, first we choose T, a realization of the graph $\mathcal{T}^\infty$, according to ν (defined on the comments after definition 2.37), and then X^n starts at $X_0 = \rho$ and moves on T as $(X^T_n)_{n\in\mathbb{N}_0}$.

In [16, Theorem 1.16], Kesten stated that if $(Y_t)_{t\in\mathbb{R}_+}$ is the linear interpolation of the height function of the random walk on the Kallenberg-Kesten tree, then $n^{-\frac{1}{3}}Y_{nt}$ converges weakly in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}_+)$ (under the annealed law) to a non-zero process.

Later, in [2, section 5.1], now focused on $(\mathcal{T}_e, d_e, \mu_e, \rho_e)$, the Continuum Random Tree (remember 2.14) equipped with a probability measure μ_e and a distance d_e, Aldous conjectured the existence of a strong Markov process[1] $X = (X_t)_{t\geq 0}$ on $\mathcal{T}_e$ with continuous paths such that μ_e is its reversible equilibrium and it satisfies

(i) for each path $[\![a,b]\!] \subset \mathcal{T}_e$ and each $x \in [\![a,b]\!]$,

$$P_x(\tau_a < \tau_b) = \frac{d_e(x,b)}{d_e(a,b)},$$

where, for $x \in \mathcal{T}_e$, $\tau_x = \inf\{t \geq 0 : X_t = x\}$.

[1] We refer to [18] for a precise definition of this process.

(ii) for all $f : \mathcal{T}_e \to \mathbb{R}_+$ bounded measurable function and x, y on $\mathcal{T}_e$,

$$\mathbb{E}_x\Big[\int_0^{\tau_z} f(X_t)\,\mathrm{d}t\Big] = 2\int_{\mathcal{T}_K} \mu_e(\mathrm{d}y)\,\Big(f(y)\cdot d_e\big(z, c(x,z,y)\big)\Big) \tag{32}$$

where for $x, y, z \in \mathcal{T}_K$ $c(x,z,y)$ denotes the unique point in $\mathcal{T}_K$ that satisfies

$$[\![x, c(x,z,y)]\!] = [\![x,z]\!] \cap [\![x,y]\!].$$

He called it the Brownian motion on the CRT. In a further work [17], Krebs showed a construction of this process based on Dirichlet forms.

Later, in [11], Croydon considered a triple (t_n, μ_n, X^n), where, for each $n \in \mathbb{N}$, t_n is a discrete ordered tree with n vertices, μ_n is the uniform measure on the vertices of t_n and X_n is the simple random walk on t_n started from the root $\rho \in t_n$. Denoting by $(\tilde{t}_n, \tilde{\mu}_n, \tilde{X}^n)$ its random embedding on l^1, the Banach spaces of sequences $(x_n)_{n\in\mathbb{N}_0}$ such that $x_n \in \mathbb{R}$ for each $n \in \mathbb{N}$ and $\sum_{n\in\mathbb{N}} |x_i| < \infty$, and by $\mathbb{P}_n$ the law of that triple on $\mathcal{K}(l^1) \times \mathcal{M}_1(l^1) \times C([0,1], l^1)$, where $\mathcal{K}(l^1)$ is the space of compact sets of l^1, he concludes that, if the sequence of contour functions $n^{-1/2}C(t_n)$ converges weakly to the reflected Brownian excursion on $C([0,1], \mathbb{R}_+)$, then $\Big(n^{-1/2}\tilde{t}_n, \tilde{\mu}_n(\sqrt{n}\cdot), (n^{-1/2}X^n_{tn^{3/2}})_{t\in[0,1]}\Big)$ converges in distribution to $(\tilde{\mathcal{T}_\rceil}, \tilde{\mu_e}, \tilde{X})$, where this triple is a random embedding of the CRT, the length measure and Brownian Motion X into l^1.

Finally, in [5, theorem 1], a very general result is proved about scaling limits of random walks on Heine-Borel tree-like spaces that converge Gromov-Hausdorff vaguely. Here we state a simplified version of this theorem.

Definition 2.41. Consider $\chi = (\mathcal{T}, d, \rho, \mu)$ a discrete ordered rooted tree embedded as a real measure tree.

The *continuous-time nearest-neighbor random walk* on χ $(X_n^{\mathcal{T}})_{n\in\mathbb{N}_0}$ is the Markov chain with initial state $X_0 = \rho$ that jumps from $x \in \mathrm{nod}(\mathcal{T})$ to $y \in \mathrm{nod}(\mathcal{T})$ if $d(x,y) = 1$ at rate $r(x,y) = \frac{1}{2\cdot\mu(\{y\})\cdot d(x,y)}$.

Theorem 2.42. Let, for each $n \in \mathbb{N}$, $\chi^n = (\mathcal{T}_n, d_n, \rho_n, \mu_n)$ be a discrete tree embedded into a real tree, and $\chi = (\mathcal{T}, d, \rho, \mu)$ be a Heine-Borel real tree.

If χ^n converges to χ Gromov-Hausdorff vaguely, then there exists a metric space (E, d_E) and isometric embeddings $\phi_n : \mathcal{T}_n \to E$, $\phi : \mathcal{T} \to E$ such that $(\phi \circ X^n)_{n\in\mathbb{N}}$ converges weakly in path space to $\phi \circ X$, where X the unique strong Markov process on $\mathcal{T}$ that, when restricted to compact subtrees, satisfies (32).

In this work we focused on scaling limits of trees related to a critical Galton-Watson process, but the topological results introduced in this chapter can be applied to a wider variety of random trees. Here we cite a result related to spanning trees in high-dimensional graphs.

A *spanning tree* of a connected finite graph $G = (E, V)$ is a subgraph $t^G = (E', V)$, where $E' \subset E$ connects every vertex in V and does not contain any cycle. The *uniform spanning tree* of G is a random graph uniformly chosen from the set of spanning trees of G. In [4], a scaling convergence theorem is obtained regarding uniform spanning trees of a sequence of graphs that satisfy certain conditions (see [4, Assumption 1.4]). Examples of graphs that satisfy their conditions are the d-dimensional torus $\mathbb{Z}_n^d$ with $d > 4$ and the hypercube $\{0,1\}^n$. The statement of [4, Theorem 1.8] is as follows.

Let G_n be a sequence of graphs that satisfy [4, Assumption 1.4], and let $\mathcal{T}_n$ be a uniform spanning tree of G_n. Equip $\mathcal{T}_n$ with d_n, the graph distance, with μ_n, the uniform measure on the vertices of $\mathcal{T}_n$, and with O_n, an arbitrary vertex of G_n. Then there exists a sequence β_n satisfying $0 < \inf_{n\in\mathbb{N}} \beta_n \leq \sup_{n\in\mathbb{N}} \beta_n < \infty$ such that

$$\left(\mathcal{T}_n, \frac{d_n}{\beta(d)n^{d/2}}, \mu_n, O_n\right) \xrightarrow[n\to\infty]{(d)} (\mathcal{T}_e, , d_e, \mu_e, O)$$

where $\mathcal{T}_e$ is the CRT equipped with μ_e, and root O. This convergence is a convergence in distribution on the space of (equivalent classes of) compact metric spaces equipped with the so-called Gromov-Hausdorff-Prohorov metric, which coincides with the Gromov-Hausdorff vague convergence as the measures μ_n and μ_e have full support (see [6, Remark 5.2]).

A corollary of this result related to our previous discussion is [4, theorem 1.11], which states the scaling limit of the simple random walk on $\mathcal{T}_n$ to the Brownian motion on the CRT.

Bibliography

[1] Aldous, David. “The continuum random tree I”. In: *The Annals of Probability* 19.1 (1991), pp. 1–28.

[2] Aldous, David. “The Continuum random tree II: an overview”. In: *Stochastic Analysis: Proceedings of the Durham Symposium on Stochastic Analysis, 1990*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1991, pp. 23–70.

[3] Aldous, David. “The continuum random tree III”. In: *Annals of Probability* 21 (1993), pp. 248–289.

[4] Archer, Eleanor, Nachmias, Asaf, and Shalev, Matan. “The GHP Scaling Limit of Uniform Spanning Trees in High Dimensions”. In: *Communications in Mathematical Physics* 405.3, 73 (2024).

[5] Athreya, Siva, Löhr, Wolfgang, and Winter, Anita. “Invariance principle for variable speed random walks on trees”. In: *The Annals of Probability* 45.2 (Mar. 2017).

[6] Athreya, Siva, Löhr, Wolfgang, and Winter, Anita. “The gap between Gromov-vague and Gromov–Hausdorff-vague topology”. In: *Stochastic Processes and their Applications* 126.9 (2016), pp. 2527–2553.

[7] Bernardo N. B. de Lima, Leonardo T. Rolla. “Probabilidade”. 2024. URL: `https://www.ime.usp.br/~leorolla/probabilidade/`.

[8] Bertoin, J. *Lévy Processes*. Cambridge University Press, 1998.

[9] Billingsley, Patrick. *Convergence of Probability Measures (Second Edition)*. Wiley Series in Probability and Statistics, 1999.

[10] Burago, Dmitri, Burago, Yuri, and Ivanov, Sergei. *A Course in Metric Geometry*. American Mathematical Society, 2001.

[11] Croydon, David. “Convergence of simple random walks on random discrete trees to Brownian motion on the continuum random tree”. In: *Ann. Inst. H. Poincaré Probab. Statist.* 44.2 (2008), pp. 987–1019.

[12] Gromov, M. *Metric Structures for Riemannian and Non-Riemannian Spaces*. Birkhäuster, 1999.

[13] Janson, Svante. “Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation”. In: *Probability Surveys* 9.none (2012), pp. 103–252.

[14] K.B. Athreya, P. E. Ney. *Branching Processes*. Springer, Berlin, 1972.

[15] Kaigh, W. D. “An Invariance Principle for Random Walk Conditioned by a Late Return to Zero”. In: *The Annals of Probability* 4.1 (1976), pp. 115–121.

[16] Kesten, Harry. "Subdiffusive behavior of random walk on a random cluster". In: *Annales de l'I.H.P. Probabilités et statistiques* 22.4 (1986), pp. 425–487.

[17] Krebs, W.B. "Brownian motion on the continuum tree". In: *Probability Theory and Related Fields 101* (1995), pp. 421–433.

[18] Le Gall, J. F. *Brownian Motion, Martingales, and Stochastic Calculus*. Springer Publishing Company, Incorporated, 2018.

[19] Le Gall, J.F. "Random trees and applications". In: *Probability Surveys* 2 (2005), pp. 245–311.

[20] Le Gall, J.F. *Random trees and spatial branching processes.* Maphisto Lecture Notes Series (University of Aarhus), 2000.

[21] Norris, J. R. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics, 1997.

[22] Pitman, James W. "One-Dimensional Brownian Motion and the Three-Dimensional Bessel Process". In: *Advances in Applied Probability* 7.3 (1975), pp. 511–526. (Visited on 07/05/2025).

[23] Pitman, Jim and Winter, Anita. "Rayleigh processes, real trees, and root growth with re-grafting". In: *Probability Theory and Related Fields* 134 (Mar. 2004), pp. 81–126.

[24] Shiryaev, A. N. *Probability (2nd ed.)* Berlin, Heidelberg: Springer-Verlag, 1996.

[25] Spitzer, F. *Principles of Random Walk*. Van Nostrand, 1963.

Appendix A

Convergence of probability measures

In this chapter we will give a brief introduction to some the theory of convergence of probability measures. The details and proofs can be found at [9]. Henceforth we always consider E a metric space and $\mathcal{B}(E)$ its Borel σ-algebra.

Definition A.1. Take $(P_n)_{n\in\mathbb{N}}$ and P probability measures on $(E, \mathcal{B}(E))$, where E is a metric space. We say that P_n converges weakly to P, denoted by $P_n \Rightarrow P$, if $\lim_{n\to\infty} \int_E f \, \mathrm{d}P_n = \int_E f \, \mathrm{d}P$ for any f bounded, continuous real function f on E.

Let $\mathcal{M}_1(E)$ be the set of all probability measures on E. We will denote weak convergence of a sequence $(P_n)_{n\in\mathbb{N}}$ to P on $\mathcal{M}_1(E)$ by $P_n \xRightarrow{n\to\infty} P$.

Definition A.2. Let E and E' be two metric spaces and $h : E \to E'$ a measurable function, and P a probability measure on E. We call h_*P *the pushforward of P by h* as the probability measure on E' defined as $h_*P_n(A) = P_n(h^{-1}(A))$, for every $A \in \mathcal{B}(E')$.

Theorem A.3. (Continuous mapping theorem) If $(P_n)_{n\in\mathbb{N}}$ is a weakly convergent sequence in E that converges to P and $h : E \to E'$ is a continuous function, then $h_*P_n \Rightarrow h_*P$.

Definition A.4. Let Π be a family of probability measures on $(E, \mathcal{B}(E))$. We call Π *relatively compact* if every sequence of elements of Π contains a weakly convergent subsequence.

Definition A.5. A family Π of probability measures on $(E, \mathcal{B}(E))$ is *tight* if for every $\eta > 0$ there exists a compact set $K \subset E$ such that

$$P(K) \geq 1 - \eta \quad \forall P \in \Pi.$$

Theorem A.6. (Prohorov theorem) If E is a polish space, that is, separable and complete, then Π is relatively compact if, and only if, it is tight.

Definition A.7. The *Prohorov distance* in $\in \mathcal{M}_1(E)$ is defined by

$$d_{Pr}^{(E,d)}(\mu, \mu') = \inf\{\varepsilon > 0 : \mu(A) \leq \mu'(A^\varepsilon) + \varepsilon, \mu'(A) \leq \mu(A^\varepsilon) + \varepsilon \;\; \forall A \text{ closed}\}$$

where $A^\varepsilon = \{x : d(x, A) \leq \varepsilon\}$ is the closed ε-neighborhood of A.

Theorem A.8. Let E be a polish space, $(P_n)_{n\in\mathbb{N}}$ and P be probability measures on E. Then $\lim_{n\to\infty} d_{Pr}^{(E,d))}(P_n, P) = 0$ iff $P_n \xRightarrow{n\to\infty} P$.

A.1 Weak convergence on D_∞

Consider the space $D([0,\infty)) = D_\infty$ of functions $f : \mathbb{R}_+ \to \mathbb{R}$ that are right-continuous with left limits (càdlàg). We want to define the metric that will be used in D_∞, in order to do so we introduce some notations.

For every $m \in \mathbb{N}$ define $g_m : \mathbb{R}_+ \to \mathbb{R}_+$ as

$$g_m(t) = \begin{cases} 1 & \text{, for } t \leq m-1, \\ m-t & \text{, for } m-1 \leq t \leq m, \\ 0 & \text{, for } t \geq m. \end{cases}$$

For $x \in D_\infty$, let x^m be the element of D_∞ defined by

$$x^m(t) = g_m(t)x(t), \quad t \geq 0.$$

Now we introduce the set

$$\Lambda_m = \{\lambda : [0,m] \to [0,m] : \lambda \text{ is a strictly increasing, continuous function}\}$$

and, for each $\lambda \in \Lambda_m$ consider

$$\|\lambda\| = \sup_{s \neq t} \left| \log\left(\frac{\lambda t - \lambda s}{t-s} \right) \right|.$$

Using these notations we define, for each $m \in \mathbb{N}$, d_m a metric on the subset of elements of D_∞ with support on $[0,m]$

$$d_m(x^m, y^m) = \inf_{\varepsilon > 0} \Big\{ \exists \lambda \in \Lambda_m : \|\lambda\| < \varepsilon \text{ and } \sup_{t \in [0,m]} |x^m(t) - y^m(\lambda(t))| \leq \varepsilon \Big\}.$$

Finally, we are able to define the Skorohod metric in D_∞.

Definition A.9. Take $x, y \in D_\infty$. We define the metric that induces the $J_1 Skorohod topology on D_\infty$ by

$$d_\infty(x,y) = \sum_{m=1}^{\infty} 2^{-m}(1 \wedge d_m(x^m, y^m)).$$

Proposition A.10. D_∞ equipped with the metric d_∞ is a polish space.

So we can apply the result of Prohorov's theorem in D_∞ equipped with the Skorohod topology.

Definition A.11. Given an metric space E, we say that $\mathcal{K} \subset \mathcal{B}(E)$ is a *separating class* if μ, ν probability measures on $(E, \mathcal{B}(E))$ satisfy $\mu(K) = \nu(K)$ for every $K \in \mathcal{K}$, then $\mu(A) = \nu(A)$ for every $A \in \mathcal{B}(E)$.

Definition A.12. Let P be a probability measure on $(D_\infty, \mathcal{B}(D_\infty))$. For $t_1, \ldots t_p \in \mathbb{R}_+$, denote by $\pi_{t_1,\ldots,t_p}$ the projection of D_∞ in $\mathbb{R}^p$ such that $\pi_{t_1,\ldots,t_p}(x) = (x(t_1), \ldots, x(t_p))$.

Then the elements of

$$\{(\pi_{t_1,\ldots t_p})_* P : (t_1, \ldots, t_p) \in \mathbb{R}_+^k, p \in \mathbb{N}\}$$

are called the *finite-dimensional distributions* of P.

Proposition A.13. The set $\{\pi_{t_1,\ldots t_p} : (t_1, \ldots, t_p) \in \mathbb{R}^k_+, p \in \mathbb{N}\}$ is a separating class of D_∞.

As a consequence of the previous proposition we have that

Proposition A.14. If $(P_n)_{n\in\mathbb{N}}$ is a relatively compact sequence of probability measures on D_∞ and the finite-dimensional distributions of P_n converges weakly to those of P, then P_n converges weakly to P.

Theorem A.15. A sequence $\{P_n \,; n \in \mathbb{N}\}$ of probability measures on D_∞ is tight if and only if the following conditions holds for all m:

(i) For each t in a dense subset of R_+,

$$\lim_{a\to\infty} \limsup_{n\to\infty} P_n \left(\{x \in D_\infty : |x(t)| \geq a\}\right) = 0.$$

(ii) For each $m \in \mathbb{N}$ and $\varepsilon > 0$,

$$\lim_{\delta\to\infty} \limsup_{n\to\infty} P_n \left(\{x \in D_\infty : w'_m(x,\delta) \geq \varepsilon\}\right) = 0$$

where

$$w'_m(x,\delta) = \inf_{\{t_i\}}\{\max_{1\leq i\leq v} w(x, [t_{i-1}, t_i))\}$$

in which

$$w(x, [t_{i-1}, t_i)) = \sup_{s,t\in[t_{i-1},t_i)} \{|x(s) - x(t)|\}$$

and the infimum is taken over all decompositions $0 = t_0 < t_1 < \ldots < t_v = m$ of $[0, m] \in \mathbb{R}_+$ such that $t_i - t_{i-1} > \delta$ for $1 \leq i < v$, $v \in \mathbb{N}$.

Appendix B

Brownian motion

In this chapter we will give a brief introduction to some of the results related to Brownian motion. The details and proofs can be found at [9] and [18].

Definition B.1. The (standard) Brownian motion on $\mathbb{R}$, $(B_t)_{t\in\mathbb{R}_+}$, is the random element that takes values on $\mathcal{C}(\mathbb{R}_+,\mathbb{R})$, the space of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}$ and satisfies:

- $B_0 = 0$ almost surely.
- For every $0 \le s < t$, the random variable $B_t - B_s$ is independent of $\sigma(B_r, r \le s)$ and distributed according to $\mathcal{N}(0, t-s)$.

Equip the space $\mathcal{C}_\infty = \mathcal{C}(\mathbb{R}_+,\mathbb{R})$ with the topology of uniform convergence on compact sets and the corresponding Borel $\sigma-$algebra C.

Definition B.2. The *Wiener measure* is a measure on $(\mathcal{C}_\infty, C)$ defined by

$$W(A) = P(B \in A)$$

for $A \in C$.

When A is a cylinder set, that is $A = \{w \in \mathcal{C}_\infty : w(t_0) \in A_0, \dots, w(t_k) \in A_k\}$ for some $k \in \mathbb{N}_0$, $0 = t_0 < t_1 < \ldots < t_k$ and $A_0, \dots, A_k \in \mathcal{B}(\mathbb{R})$, the Wiener measure can be written as

$$W(A) = \mathbb{1}_{A_0}(0) \int_{A_1\times\ldots\times A_k} \frac{dx_1 \ldots dx_k}{(2\pi)^{\frac{n}{2}}\sqrt{t_1(t_2-t_1)\ldots(t_n-t_{n-1})}} \exp\left(-\sum_{i=1}^{k}\frac{(x_i-x_{i-1})^2}{2(t_i-t_{i-1})}\right).$$

Theorem B.3. (Donsker's invariance theorem) Let $(\xi_i)_{i\in\mathbb{N}_0}$ be a sequence of independent and identically distributed random variables with mean 0 and variance $\sigma^2 \in (0,\infty)$. Consider $(S_n)_{n\in\mathbb{N}_0}$, where $S_n = \sum_{i=0}^{n}\xi_i$, a random walk on $\mathbb{Z}$. It follows that

$$\left(\frac{1}{\sigma\sqrt{n}}S_{[nt]}\right)_{t\in\mathbb{R}_+} \xrightarrow[n\to\infty]{(d)} (B_t)_{t\in\mathbb{R}_+}$$

where $[nt]$ is the integer part of nt, so $(S_{[nt]})_{t\in\mathbb{R}_+}$ is a process in D_∞.

There is also a version of Donsker's theorem for the random walk as a random element of $\mathcal{C}(\mathbb{R}_+,\mathbb{R})$ by taking $(S_{[nt]})_{t\in\mathbb{R}_+}$ as the linear interpolation of the rescaled random walk.

Remark B.4. Actually in [9] it is proved the weak convergence on the Skorohod space $D[0,1]$ of P^n, the distributions of $\left(\frac{1}{\sigma\sqrt{n}} S_{[nt]}\right)_{t\in[0,1]}$, to the Wiener measure on $\mathcal{C}[0,1]$. The theory behind this proof can be easily generalized to show convergence on $D[0,m]$ for any $m \in \mathbb{N}$, since, by [9, lemma 3, chapter 16], this is a sufficient and necessary condition for weak convergence on D_∞ the stated result follows.

Now we list some useful properties of the Brownian motion on $\mathbb{R}$.

Proposition B.5. Let B be a Brownian motion. Then

1. $-B$ is also a Brownian motion.
2. For every $\lambda > 0$, $(\frac{1}{\lambda} B_{\lambda^2 t})_{t\in\mathbb{R}_+}$ is also a Brownian motion.
3. For every $s \geq 0$ the process $(B_{s+t} - B_s)_{t\in\mathbb{R}_+}$ is a Brownian motion independent of $\sigma(B_r, r \leq s)$.

Theorem B.6. (Reflection principle) For every $t > 0$, set $S_t = \sup_{s\leq t} B_s$. Then, if $a \geq 0$ and $b \leq a$ we have

$$P(S_t \geq a, B_t \leq b) = P(B_t \geq 2a - b).$$

Moreover, S_t has the same distribution as $|B_t|$.

Appendix C

Markov Chains basics

In this chapter we will give a brief introduction to some Markov chain theory. The details and proofs can be found in [21]. Some of our formulation is taken from [24].

Consider I a countable set which we will call the *state space*. Each element of I is called a *state*. Also let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a stochastic process, that is, a sequence in which each X_n is a random variable with values in I. Consider $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ the canonical filtration of X, meaning that each $\mathcal{F}_n$ is the σ-algebra generated by $(X_0, X_1, \ldots, X_n)$.

In this context we define:

Definition C.1. We say that $X = (X_n)_{n\in\mathbb{N}_0}$ is a *discrete-time Markov chain* if X satisfies the **Markov property**, that is, for every $f : I \to \mathbb{R}$ bounded and measurable and $n \geq m \geq 0$

$$\mathbb{E}[f(X_n)|\mathcal{F}_m] = \mathbb{E}[f(X_n)|X_m] \qquad \mathbb{P} \text{ almost surely.}$$

Here X_m corresponds to the σ-algebra generated by X_m.

The Markov property can be reformulated in a number of ways. One equivalent ways is given below: If $(X_n)_{n\in\mathbb{N}_0}$ is a Markov chain with $X_0 = x$, then, conditioned on $X_m = i$, $(X_{n+m})_{n\in\mathbb{N}_0}$ is a Markov chain started at i and is independent of the random variables $X_0, \ldots, X_m$.

Another way of defining Markov chains is by taking $Q = (q_{i,j})_{i,j\in I}$ a matrix that satisfies $\sum_{j\in I} q_{i,j} = 1$ and λ a probability measure on I.

Then $X = (X_n)_{n\in\mathbb{N}_0}$ is a time homogeneous Markov chain with initial distribution λ and transition matrix Q if

(i) X_0 has distribution λ.

(ii) For $n \geq 0$ and $i_1, \ldots, i_{n+1} \in I$

$$\mathbb{P}(X_{n+1} = i_{n+1}|X_0 = i_1, \ldots, X_n = i_n) = q_{i_n, i_{n+1}}.$$

$p_{i,j}$ will be called transition probability from i to j.

Remark C.2. In the first chapter of this work we will deal only with time-homogeneous Markov chains, so we will refer to them simply as Markov chains.

One advantage of the definition by transition matrix is that the elements of $Q^n = (q_{i,j}^{(n)})_{i,j\in I}$ have the following property

$$q_{i,j}^{(n)} = \mathbb{P}(X_n = j|X_0 = i).$$

Definition C.3. Let $(Y_k)_{k\in\mathbb{N}_0}$ be a sequence of i.i.d. sequence of random variables with common distribution μ, a probability distribution on $\mathbb{Z}$.

A random walk $(S_n)_{n\in\mathbb{N}_0}$ on $\mathbb{Z}$ is the random process defined by:

$$S_n = \sum_{k=0}^{n} Y_k.$$

This random walk also can be defined as the Markov chain with initial distribution μ and transition matrix $Q = (q_{i,j})_{i,j\in I}$ where $q_{i,j} = \mu(j-i)$. Sometimes it is convenient to assume that $S_0 = 0$.

Proposition C.4. A Markov chain $(X_n)_{n\in\mathbb{N}_0}$ with initial distribution λ and transition matrix Q satisfies, for $N \geq 0$ and $i_1, \ldots, i_N \in I$

$$\mathbb{P}(X_0 = i_1, \ldots, X_N = i_N) = \lambda(i_0) q_{i_0,i_1} \ldots q_{i_{N-1},i_N}.$$

There exists an analogous version of the Markov property related to certain types of random times. Now we concentrate on describing this property.

Definition C.5. A random variable $T : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ is called a *stopping time* if $\{T = n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}_0$.

We define the σ-algebra $\mathcal{F}_\tau$ by

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau = n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}_0\}.$$

Example C.6. The *first passage time* to state i is defined by

$$T_i(\omega) = \inf\{n \geq 1 : X_n(\omega) = i\}$$

where $\inf \emptyset = \infty$ is a stopping time.

Theorem C.7. Let $(X_n)_{n\in\mathbb{N}_0}$ be an Markov chain, τ be a stopping time and θ the shift operator, that is $\theta(X) = (X_1, X_2, \ldots)$. Then, for every $F : I^{\mathbb{N}_0} \to \mathbb{R}$ a bounded measurable function, the following *strong Markov property* is valid:

$$\mathbb{E}_x[F(\theta^\tau(X))\mathbb{1}_{\tau<\infty} | \mathcal{F}_\tau] = \mathbb{E}_{X_\tau}[F(X)]\mathbb{1}_{\tau<\infty} \quad \mathbb{P}_x \text{ almost surely.}$$

Theorem C.8. Let $(X_n)_{n\in\mathbb{N}_0}$ be a Markov chain with with initial distribution λ and transition matrix Q, and let T be a stopping time. Then, conditioned on $T < \infty$ and on $X_T = i$, $(X_{T+n})_{n\in\mathbb{N}_0}$ is a Markov chain started at i and is independent of the random variables $X_0, \ldots, X_T$.

Definition C.9. A Markov chain with transition matrix $Q = (q_{i,j})_{i,j\in I}$ is called *irreducible* if, for every $i, j \in I$ there exists $m \in \mathbb{N}_0$ such that $q_{i,j}^{(m)} > 0$.

Definition C.10. A state $i \in I$ is recurrent if

$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 1.$$

Definition C.11. Let λ be a measure on I and $Q = (q_{i,j})_{i,j\in I}$ be a transition matrix of a Markov chain. We say that λ is *invariant* if

$$\lambda(j) = \sum_{i\in I} \lambda(i) q_{i,j} \quad \text{for all } j \in I.$$

Theorem C.12. Let X be a Markov chain with transition matrix Q and λ be an invariant measure for Q with $\lambda_k = 1$ for some $k \in I$. Then if X is irreducible and recurrent we conclude that

- The expected time spent in i between visits to k

$$\gamma_i^k = \mathbb{E}_k \left[\sum_{n=0}^{T_k - 1} \mathbb{1}_{\{X_n = i\}} \right]$$

  is a invariant measure for X.

- $0 < \gamma_i^k < \infty$ for all $i \in I$.
- $\lambda = \gamma^k$.

That means that a recurrent irreducible Markov chain has an invariant measure that is unique up to multiplication by a positive constant.