Random Trees

Convergence of Discrete Random Trees to the Continuum Random Tree

Pawel Morzywolek

Bachelor Thesis, Department of Mathematics, ETH Zürich

June 27, 2013

Supervised by Prof. Dr. Pierre Nolin

Contents

0 Introduction **3**

1 Discrete Trees **4**

1.1 First definitions 4
1.2 Coding of discrete trees 5
1.3 Galton-Watson trees 8
1.4 Convergence to Brownian Motion 13
1.5 Galton-Watson trees with a fixed progeny 22
1.6 Convergence of contour functions 24

2 Real Trees and their Coding by Brownian Excursions **28**

2.1 Real trees 28
2.2 Coding of real trees 32
2.3 The continuum random trees 41

0 Introduction

In this thesis we will discuss some recent results from the theory of random trees. This is a relatively young research field with many applications for example in Computer Science or in the theory of partial differential equations.
We will begin by considering discrete random trees, which may be viewed as a model of population in which each individual has a random number of children. Further we will learn about methods of coding those trees, which will allow us to obtain some convergence results.
The next step will be to define real random trees and to find the way of coding them with suitable functions. As a special case we will get the continuum random tree (CRT) - a real random tree coded by the normalized Brownian excursion.
Finally, we will show the convergence of (rescaled) discrete random trees to the CRT.

This work is fully based on the paper "Random Trees and Applications" by Jean-François Le Galle with few references to other sources.

I would like to thank Prof. Dr. Pierre Nolin for his supervision and Pierre-François Rodriguez for his valuable comments and remarks.

1 Discrete Trees

1.1 First definitions

Definition 1.

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

is a set of labels, where $\mathbb{N} = \{1, 2, \dots\}, \mathbb{N}^0 = \{\emptyset\}$.

Notice:

- An element of $\mathcal{U}$, $u = (u^1, \dots, u^n)$ is a sequence of elements of $\mathbb{N}$.
- For $u = (u^1, \dots, u^n) \in \mathcal{U}$ we set $|u| \stackrel{\text{def}}{=} n$, the *generation* of u.
- Let $u = (u^1, \dots, u^n)$, $v = (v^1, \dots, v^m)$ be two elements of $\mathcal{U}$, then $uv = (u^1, \dots, u^n, v^1, \dots, v^m)$ is called the *concatenation* of u and v (in particular $\emptyset u = u\emptyset = u$).
- Let $u = (u^1, \dots, u^n) \in \mathcal{U} \setminus \{\emptyset\}$. Then we define a map $\Pi\colon \mathcal{U} \setminus \{\emptyset\} \to \mathcal{U}$ (father of u), as $\Pi((u^1, \dots, u^n)) = (u^1, \dots, u^{n-1})$.

Definition 2. *A (finite) rooted ordered tree t is a finite subset $\mathcal{U}$, such that*

1. $\emptyset \in t$.
2. $u \in t \setminus \{\emptyset\} \Rightarrow \Pi(u) \in t$.
3. *for all $u \in t$ there exists an integer $k_u(t) \geq 0$, such that for all $j \in \mathbb{N}$*

   $$uj \in t \text{ iff } 1 \leq j \leq k_u(t)$$

   (we interpret $k_u(t)$ as a number of children of u).

Remark 1.1.

- *We denote by $\mathbb{A}$ the set of all rooted ordered trees.*
- *We consider t as the family tree, where each vertex represents an individual.*
- *The cardinality $\#(t)$ is the total progeny.*

1.2 Coding of discrete trees

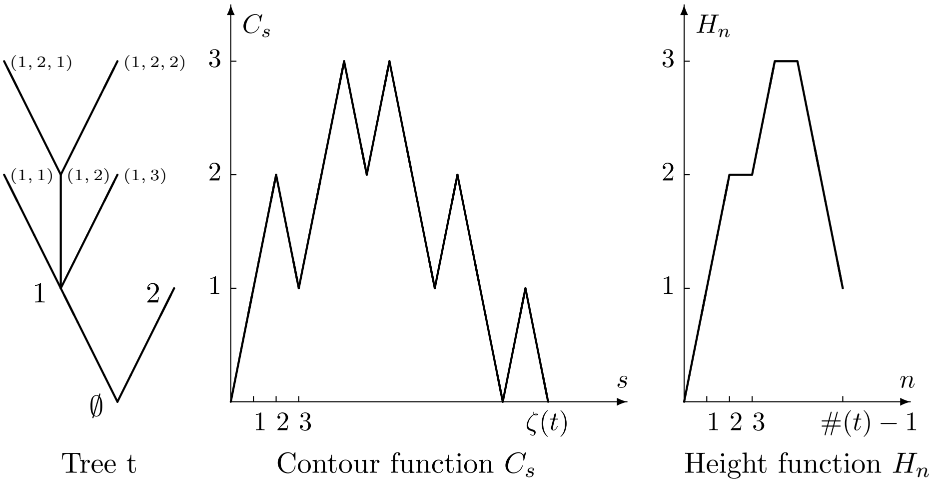


Figure 1

Remark 1.2. *We will denote by $u_0 = \emptyset, u_1, \ldots, u_{\#(t)-1}$ elements of t in lexicographical order.*

Definition 3. *The height function $h_t(n)$ of a tree t is a sequence of generations of elements in t listed in lexicographical order ($h_t(n)\colon 0 \le t < \#(t)$), such that $h_t(n) = |u_n|$ for $0 \le n < \#(t)$.*

Definition 4. *Suppose the tree is embedded in the half plane, such that edges have length one. Imagine a particle moving at unit speed on the tree in a continuous way, starting at the root and visiting all vertices in lexicographical order. The value of the contour function C_s at time $s \in [0, \zeta(t)]$ (where $\zeta(t) = 2(\#(t)-1)$) is the distance (on the tree) between the position of a particle at time s and the root of a tree.*

Let $\mathbb{S}$ denote the set of all finite sequences of non-negative integers $m_1, ..., m_p$ for $p \ge 1$, such that

- $m_1 + \ldots + m_i \ge i$ for all $i \in \{1, \ldots, p-1\}$;
- $m_1 + \ldots + m_p = p - 1$.

Proposition 1.3. *The mapping*

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

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

Proof. • Φ is well-defined

We need to check that for a given $t \in \mathbb{A}$ we have $(k_{u_0}(t), \ldots, k_{u_{\#(t)-1}}) \in \mathbb{S}$. The sequence is clearly finite (since t is a finite tree) and all values are non-negative. Moreover

$$k_{u_0}(t) + \ldots + k_{u_{\#(t)-1}} = \#(t) - 1,$$

because $k_{u_0} + \ldots + k_{u_{\#(t)-1}}$ is the number of children in t, hence it counts all individuals except for the root. Furthermore, for all $i \in \{0, \ldots \#(t) - 2\}$

$$k_{u_0} + \cdots + k_{u_i} > i.$$

This is because $u_1, \ldots, u_i$ are counted among children and the father of u_{i+1} is among $u_0, \ldots, u_i$ as well, hence

$$k_{u_0}(t) + \ldots + k_{u_i}(t) \geq i + 1.$$

- Φ is injective
Let $t \neq t^{'}$, then without loss of generality there exists the first element u in lexicographical order of t, such that $u \in t \setminus t^{'}$. But then $k_{\Pi(u)}(t) > k_{\Pi(u)}(t^{'}) \Rightarrow (k_{u_0}(t), \ldots, k_{u_{\#(t)-1}}(t)) \neq (k_{u_0}(t^{'}), \ldots, k_{u_{\#(t^{'})-1}}(t^{'})) \Rightarrow \Phi(t) \neq \Phi(t^{'})$.

- Φ is surjective
Let $(m_1, \ldots, m_p) \in \mathbb{S}$ for $p \geq 1$. We want to show that there exists $t \in \mathbb{A}$, such that $\Phi(t) = (m_1, \ldots, m_p)$. Note that while constructing a tree we have to make sure that two conditions are fulfilled, namely
  - in each moment, except for the end, (# of children) $\geq$ (# of fathers);
  - at the end (# of children) = (# of fathers) - 1.

  But those are exactly the conditions, which $(m_1, \ldots, m_p)$ satisfies. Hence, setting $k_{u_0} := m_1, \ldots, k_{u_{p-1}} := m_p$ we get a tree $t \in \mathbb{A}$.

$\square$

Definition 5. *Let $(m_1, \ldots, m_p) \in \mathbb{S}$. Then we define a finite sequence of integers*

$$x_n \overset{\text{def}}{=} \sum_{i=1}^{n} (m_i - 1) \quad \textit{for} \quad 0 \leq n \leq p$$

called Lukasiewicz path.

Remark 1.4. *The Lukasiewicz path has the following properties:*

1. $x_0 = 0,\ x_p = -1.$
2. $x_n \geq 0$ *for all* $0 \leq n \leq p - 1.$
3. $x_i - x_{i-1} \geq -1$ *for all* $1 \leq i \leq p.$

Proof. 1. Follows, since $m_1 + \ldots + m_p = p - 1$.

2. Follows, since $m_1 + \cdots + m_i \geq i$ for all $i \in \{1, \ldots, p-1\}$ and $x_0 = 0$.

3. Follows, since $x_i - x_{i-1} = \sum_{j=1}^{i}(m_j - 1) - \sum_{k=1}^{i-1}(m_k - 1) = m_i - 1 \geq -1$ (because m_i are non-negative).

$\square$

Claim 1.5. *There exists a bijection between trees (in $\mathbb{A}$) and the Lukasiewicz paths.*

Proof. It is easy to see that there is a bijection between the set of Lukasiewicz paths $\mathbb{L}$ and elements $\mathbb{S}$. By construction of the Lukasiewicz path we have $\#(\mathbb{S}) \leq \#(\mathbb{L})$, since for a given element in $\mathbb{S}$ we construct the Lukasiewicz path. Furthermore, given $(x_0, \ldots, x_p) \in \mathbb{L}$ we can uniquely reconstruct $(m_1, \ldots, m_p)$ knowing that $x_i - x_{i-1} = m_i - 1$, hence $\#(\mathbb{L}) \leq \#(\mathbb{S})$, so there is a bijection between both sets. Hence $\mathbb{L}$ is bijective to $\mathbb{A}$ by Proposition 1.3. $\square$

Note that we can interpret the value x_m as the number of elements "put at the stack", while travelling along the tree.

Proposition 1.6. *The height function h_t is related to the Lukasiewicz path of t by the formula*

$$h_t(n) = \#\left\{ j \in \{0, 1, \ldots, n-1\} : x_j = \inf_{j \leq l \leq n} x_l \right\}$$

for all $n \in \{0, 1, \ldots, \#(t) - 1\}$.

Proof. We can see that $h_t(n) = \#\left\{ j \in \{0, 1, \ldots, n-1\} : u_j \prec u_n \right\}$. Hence it is enough to show that

$$u_j \prec u_n \Leftrightarrow x_j = \inf_{j \leq l \leq n} x_l.$$

We will need the following definition and lemma:

Definition 6. *If t is a tree and $v \in t$, then we denote by $T_v t$ the tree shifted at v*

$$T_v t = \{u \in \mathcal{U} : vu \in t\}.$$

Lemma 1.7. *Let T_{u_j} be a shifted tree, then*

$$x_i^{T_{u_j}} = x_{j+i}^t - x_j^t$$

for all $i \in \left\{0, ..., \#(T_{u_j}) - 1\right\}$, where $x_i^{T_{u_j}}$ and x_i^t are the Lukasiewicz paths of T_{u_j} and t, respectively.

Proof. We observe the following:

$$k_{u_i}(T_{u_j} t) = k_{u_{j+i}}(t)$$

(note that T_{u_j} inherits the lexicographical order from t). Hence we have

$$x_i^{T_{u_j}} = \sum_{l=0}^{i-1}(k_{u_i}(T_{u_j}) - 1) = \sum_{l=0}^{i-1}(k_{u_{j+i}}(t) - 1) = x_{j+i}^t - x_j^t.$$

□

Back to the proof of Proposition 1.6.
"⇒" Let $u_j \prec u_n$. Consider T_{u_j}. Since T_{u_j} is a tree we know that

$$x_0^{T_{u_j}t} \leq x_i^{T_{u_j}t} \text{ for all } i \in \left\{0, \ldots, \#(T_{u_j}t) - 1\right\}.$$

Hence

$$x_0^{T_{u_j}t} = \inf_{i \in \left\{0,\ldots,\#(T_{u_j}t)-1\right\}} x_i^{T_{u_j}t}$$

and by Lemma 1.7 we obtain

$$x_j^t = \inf_{i \in \left\{0,\ldots,\#(T_{u_j}t)-1\right\}} x_{j+i}^t = \inf_{i \in \left\{j,\ldots,\#(T_{u_j}t)-1\right\}} x_i^t.$$

In particular

$$x_j^t = \inf_{j \leq i \leq n} x_i^t$$

(because $n \leq j + \#(T_{u_j}t) - 1$, since $u_j \prec u_n$).
"⇐" Let $x_j = \inf_{j \leq l \leq n} x_l$ and assume that $u_j \not\prec u_n$. Consider again the tree $T_{u_j}t$. Since $u_j \not\prec u_n$ we have $n > j + \#(T_{u_j}t) - 1$ (by a property of the lexicographical order). Moreover, we know that for $u_{j+\#(T_{u_j}t)}$ (the first element in lexicographical order of t not being descendant of u_j) holds

$$x_j^t - x_{j+\#(T_{u_j}t)}^t = x_0^{T_{u_j}t} - x_{\#(T_{u_j}t)}^{T_{u_j}t} = -1 \Rightarrow x_j^t > x_{j+\#(T_{u_j}t)}^t$$

$\Rightarrow x_j \neq \inf_{j \leq l \leq n} x_l$. Contradiction. □

1.3 Galton-Watson trees

Let μ be a critical or subcritical offspring distribution, i.e. a probability measure on $\mathbb{Z}_+$, such that $\sum_{k=0}^{\infty} k\mu(k) \leq 1$ (we exclude $\mu(1) = 1$ to avoid a trivial case).
We will make the following construction of a μ-Galton-Watson tree. Consider $(K_u, u \in \mathcal{U})$ a collection of independent random variables with distribution μ. Furthermore, let θ be a random subset of $\mathcal{U}$, defined as

$$\theta = \left\{u = \left(u^1, \ldots, u^n\right) \in \mathcal{U} : u^j \leq K_{(u^1,\ldots,u^{j-1})} \; \forall 1 \leq j \leq n\right\}$$

(Note that we can view θ as a random variable taking values in the space of discrete random trees $(\mathbb{A}, \mathcal{P}(\mathbb{A}))$.)

Proposition 1.8. *θ is almost surely a tree. Moreover, if $Z_n \stackrel{\text{def}}{=} \#\{u \in \theta : |u| = n\}$, then $(Z_n, n \geq 0)$ is a Galton-Watson process with offspring distribution μ and initial value $Z_0 = 1$.*

Proof.

- θ is almost surely a tree
  By construction of θ we see that $\theta(\omega)$ satisfies the conditions in the definition of a tree almost surely. Furthermore, the finiteness of $\theta(\omega)$ follows from our approach, since we consider θ as a μ-Galton-Watson tree with critical and subcritical offspring distribution. (see [2] for a general discussion).

- $(Z_n, n \geq 0)$ is a Galton-Watson process
  This fact follows from the construction above. We have $Z_0 = 1$ (because $\emptyset$ is the only element in the 0^{th} generation). Furthermore, $Z_n = \sum_{u \in \theta, |u|=n-1} K_u$ (K_u represents the number of children of an individual u and are μ-distributed). Hence, we get a Galton-Watson process with offspring distribution μ.

□

Definition 7. *If t is a tree and $1 \leq j \leq k_{\emptyset}(t)$, then we denote by $T_j t$ the tree shifted at j*

$$T_j t = \{u \in \mathcal{U} : ju \in t\}.$$

Remark 1.9.

- *$T_j t$ is a tree.*
- *We will consider the probability distribution Π_μ of θ on $\mathbb{A}$, which has the following properties*
  1. *$\Pi_\mu(k_\emptyset = j) = \mu(j)$, $\forall j \in \mathbb{Z}_+$*
  2. *for every $j \geq 1$ with $\mu(j) \geq 0$, the shifted trees $T_1 t, \ldots, T_j t$ are independent under the conditional probability $\Pi_\mu(dt | k_\emptyset = j)$ and their conditional distribution is Π_μ (this is called branching property of the Galton-Watson tree).*

Proposition 1.10. *For every $t \in \mathbb{A}$, we have*

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

Proof. First we can notice that

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

(since by Proposition 1.3 $(k_{u_0}(t), \ldots, k_{u_{\#(t)-1}}(t))$ uniquely determines the tree t).
Hence

$$\Pi_\mu(t) = \mathbb{P}(\theta = t) = \mathbb{P}(\bigcap_{u \in t} \{K_u = k_u(t))\}$$
$$\stackrel{*}{=} \prod_{u \in t} \mathbb{P}(K_u = k_u(t)) \stackrel{**}{=} \prod_{u \in t} \mu(k_u(t))$$

(where * and ** holds, since K_u are independent and μ-ditributed). □

Proposition 1.11. *Let θ be a μ-Galton-Watson tree. Then:*

$$\Phi(\theta) \stackrel{\text{def}}{=} (M_1, M_2, \ldots, M_T)$$

where the random variables $M_1, M_2, \ldots$ are independent and μ-distributed and $T \stackrel{\text{def}}{=} \inf\{n \geq 1 : M_1 + \ldots + M_n < n\}$.

Proof. Let $U_0, U_1, \ldots, U_{\#(\theta)-1}$ be elements of θ in lexicographical order. We can write

$$\Phi(\theta) = (K_{U_0}, K_{U_1}, \ldots, K_{U_{\#(\theta)-1}}).$$

Note that $K_{U_0} + \ldots + K_{U_n} \geq n+1$ for all $n \in \{0, 1, \ldots, \#(\theta) - 2\}$ and $K_{U_0} + \ldots + K_{U_{\#(\theta)-1}} = \#(\theta) - 1$. Furthermore, we define U_p for $p \geq \#(\theta)$ by

$$U_p \stackrel{\text{def}}{=} U_{\#(\theta)-1} \quad \underbrace{1 \ldots 1}_{p-\#(\theta)+1 \text{ ones}} \quad .$$

Now it suffices to prove that $K_{U_0}, \ldots, K_{U_p}$ are independent and μ-distributed for all $p \geq 0$. Observe that we cannot just use the fact that $(K_u, u \in \mathcal{U})$ are independent and μ-distributed, since in our case the indices are also random. We will proceed by induction.
For $p = 0$ and $p = 1$ the result is clear, since $U_0 = \emptyset$ and $U_1 = 1$ are deterministic.
Take $p \geq 2$ fixed and assume that we have already proven the case $p-1$. We can observe that for all fixed $u \in \mathcal{U}$ the random set

$$\theta \cap \{v \in \mathcal{U} : v \leq u\}$$

is measurable with respect to σ-algebra $\sigma(K_v, v < u)$.
Note that $\theta \cap \{v \in \mathcal{U} : v \leq u\}$ is a random variable taking values in $\mathbb{A}_u$, where
$\mathbb{A}_u \stackrel{\text{def}}{=} \mathbb{A} \cap \{v \in \mathcal{U} : v \leq u\}$. The measurability follows, since for all $t \in \mathbb{A}_u$ we have

$$(\theta \cap \{v \in \mathcal{U} : v \leq u\})^{-1}(t) = \bigcap_{v<u} \{K_v = a_v\},$$

where $a_v = \begin{cases} u^j, & \text{if } v = (u^1, \ldots, u^{j-1}) \\ k_v(t), & \text{otherwise} \end{cases}$. So, clearly $(\theta \cap \{v \in \mathcal{U} : v \leq u\})$ is $\sigma(K_v, v < u)$-measurable. From that we obtain that the event

$$\{U_p = u\} \cap \{\#(\theta) > p\}$$

is measurable with respect to $\sigma(K_v, v < u)$. To show this note first that

$$\{U_p = u\} \cap \{\#(\theta) > p\}$$
$$= \bigcup_{u_0 < u_1 < \ldots < u_{p-1} < u} \{U_0 = u_0, \ldots, U_{p-1} = u_{p-1}, U_p = u, \#(\theta) > p\}.$$

Furthermore, we see that for $t = (u_0, u_1, \ldots, u_{p-1}, u)$ we have

$$
\begin{aligned}
&(\theta \cap \{v \in \mathcal{U} : v \leq u\})^{-1}(t) \\
&= \{U_0 = u_0, \ldots, U_{p-1} = u_{p-1}, U_p = u, \#(\theta) > p\}
\end{aligned}
$$

and this set is in $\sigma(K_v, v < u)$ by the consideration above. Hence, $\{U_p = u\} \cap \{\#(\theta) > p\}$ is in $\sigma(K_v, v < u)$ as a countable union of sets above.
Similarly, we see that

$$\{U_p = u\} \cap \{\#(\theta) \leq p\}$$

is also measurable with respect to $\sigma(K_v, v < u)$.
Note the following equality:

$$\{U_p = u\} \cap \{\#(\theta) \leq p\} = \bigcup_{k=1}^{p} (\{U_p = u\} \cap \{\#(\theta) = k)\}\,.$$

Furthermore

$$
\begin{aligned}
&\{U_p = u\} \cap \{\#(\theta) = k\} \\
&= \bigcap_{u_0 < \ldots < u_{k-1} < u} \left\{ U_0 = u_0, \ldots, U_{k-1} = u_{k-1}, u = u_{k-1} \underbrace{1 \ldots 1}_{p-k+1} \right\}
\end{aligned}
$$

is $\sigma(K_v, v < u)$-measurable. Hence

$$\{U_p = u\} \cap \{\#(\theta) \leq p\}$$

is $\sigma(K_v, v < u)$-measurable. From that it follows that $\{U_p = u\}$ is measurable with respect to $\sigma(K_v, v < u)$.
Now we will need the following lemma:

Lemma 1.12. *The random variables $Z_1, \ldots, Z_k$ are independent iff $\mathbb{E}[\prod_{i=1}^k f_i(Z_i)] = \prod_{i=1}^k \mathbb{E}[f_i(Z_i)]$ for all non-negative functions $f_i \colon \mathbb{R} \to \mathbb{R}$.*

Proof. "$\Rightarrow$" This result is well-known in Probability Theory (confirm with [8], Corollary (4.7)) and we will not prove it, since only the other implication will be of interest for us.
"$\Leftarrow$" We can set $f_i := \mathbb{1}_{B_i}$ with $B_i \in \mathcal{B}(\mathbb{R})$, $i = 1, 2, \ldots, k$, hence we obtain independence of $Z_1, \ldots, Z_k$. ☐

Let $g_0, g_1, \ldots, g_p$ be non-negative functions on $\{0, 1, \ldots\}$ then

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

The last equality holds, since K_{u_p} is independent of $(K_v, v < u_p)$ together with the fact that $\{U_p = u_p\}$ is $\sigma(K_v, v < u)$ measurable. Furthermore, noticing that $\mathbb{E}[g_p(K_{u_p})] = \mu(g_p)$ does not depend on u_p we can write

$$(*) = \mu(g_p) \sum_{u_0 < \ldots < u_p} \mathbb{E}[\mathbb{1}_{\{U_0 = u_0, \ldots, U_p = u_p\}} g_0(K_{u_0}) \ldots g_{p-1}(K_{u_{p-1}})].$$

In particular, setting $g_p \equiv 1$ we obtain

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

Finally, we get

$$\begin{aligned}
&\mathbb{E}[g_0(K_{U_0})g_1(K_{U_1}) \ldots g_p(K_{U_p})] \\
&= \mathbb{E}[g_0(K_{U_0})g_1(K_{U_1}) \ldots g_{p-1}(K_{U_{p-1}})]\mathbb{E}[g_p(K_{U_p})].
\end{aligned}$$

Now the statement follows by the induction assumption. □

As a direct Corollary we get

Corollary 1.13. *Let $(S_n, n \geq 0)$ be a random walk on $\mathbb{Z}$ with initial value S_0 and jump distribution $\nu(k) = \mu(k+1)$ for all $k \geq -1$. Set $T = \inf\{n \geq 1 : S_n = -1\}$. Then the Lukasiewicz path of a μ-Galton-Watson process has the same distribution as $(S_0, S_1, \ldots, S_T)$. In particular $\#(\theta)$ and T has the same distribution.*

Proof. By Proposition 1.11 we have

$$\Phi(\theta) = (K_{U_0}, \ldots, K_{U_T})$$

where $T = \inf\{n \geq 0 : K_{U_0} + \ldots + K_{U_n} = n\}$. Moreover, set

$$\begin{aligned}
S_0 &= 0, \\
S_k &= \sum_{i=0}^{k-1} (K_{U_i} - 1).
\end{aligned}$$

Hence clearly jumps $(K_{U_i} - 1)$ of S_k are ν-distributed and

$$\begin{aligned}
T &= \inf\{n \geq 0 : K_{U_0} + \ldots + K_{U_n} = n\} \\
&= \inf\{n \geq 0 : (K_{U_0} - 1) + \ldots + (K_{U_n} - 1) = -1\} \\
&= \inf\{n \geq 1 : S_n = -1\}.
\end{aligned}$$

□

1.4 Convergence to Brownian Motion

We will show that suitably rescaled height functions (respectively contour functions) converge in distribution to the normalized Brownian excursion.
From now on, we will consider the critical offspring distribution μ (i.e. $\sum_{k=0}^{\infty} k\mu(k) = 1$) with finite variance $\sigma^2 > 0$. Moreover, we assume that μ has finite exponential moments, i.e. exists $\lambda > 0$, such that $\sum_{k=0}^{\infty} e^{\lambda k}\mu(k) < \infty$.

Definition 8. *Let $\theta_1, \theta_2, \dots$ be a sequence of μ-Galton-Watson trees. With each tree we associate a height function $(h_{\theta_i}(n), 0 \leq n \leq \#(\theta_i) - 1)$. We define the height process $(H_n, n \geq 0)$ as follows*

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

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

Note:

- $(H_n, n \geq 0)$ determines the sequence of trees.
- k^{th} excursion of H_n from 0 is the height function of the k^{th} tree.

Proposition 1.14. *For every $n \geq 0$ it holds:*

$$H_n = \#\left\{k \in \{0, 1, \dots, n-1\} : S_k = \inf_{k \leq j \leq n} S_j\right\}$$

where $(S_n, n \geq 0)$ is a random walk defined in Corollary 1.13.

Proof. The proposition follows directly from the fact that the random walk $(S_n, n \geq 0)$ has the same distribution as the Lukasiewicz path of sequence of random trees $\theta_1, \theta_2, \dots$ (by Corollary 1.13) together with Proposition 1.6. $\square$

Now we come to our main convergence result.

Theorem 1.15. *Let $\theta_1, \theta_2, \dots$ be a sequence of independent μ-Galton-Watson trees and $(H_n, n \geq 0)$ the associated height process. Then*

$$\left(\frac{1}{\sqrt{p}} H_{[pt]}, t \geq 0\right) \xrightarrow[p \to \infty]{(d)} \left(\frac{2}{\sigma}\gamma_t, t \geq 0\right)$$

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

Proof. Let $S = (S_n, n \geq 0)$ be as in Corollary 1.13. Note that S is recurrent. This holds because its jumps have mean zero, since

$$\sum_{k=0}^{\infty} k\mu(k) = \sum_{k=0}^{\infty} k\nu(k-1) = \sum_{k=-1}^{\infty} (k+1)\nu(k) = 1$$

$$\Rightarrow \sum_{k=-1}^{\infty} k\nu(k) = 0,$$

therefore S also has mean zero. Hence either it takes only values 0 from some point on or it takes infinitely many positive and negative values. But, since the only negative jump of S is -1 we have that S must cross 0 each time. Furthermore, we introduce

$$M_n = \sup_{0\le k\le n} S_k, \quad I_n = \inf_{0\le k\le n} S_k.$$

By Donsker's invariance theorem we get

$$(\frac{1}{\sqrt{p}} S_{[pt]},\, t \ge 0) \xrightarrow[p\to\infty]{(d)} (\sigma B_t,\, t \ge 0)$$

where B is a Brownian Motion started at the origin.
Furthermore, introduce for all $n \ge 0$ time-reversed random walk $\hat{S^n}$

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

Note that $(\hat{S^n},\ 0 \le k \le n)$ has the same distribution as $(S_k,\ 0 \le k \le n)$. Hence from Proposition 1.14 we have

$$H_n = \# \left\{ k \in \{0, 1, \dots, n-1\} : S_k = \inf_{k\le j\le n} S_j \right\} = \Phi_n(\hat{S^n})$$

where Φ is defined as $\Phi_n(\omega) = \# \left\{ k \in \{1, \dots, n\} : \omega(k) = \sup_{0\le j\le k} \omega(j) \right\}$ for any discrete trajectory $\omega = (\omega(0), \omega(1), \dots)$.
Furthermore, set

$$K_n = \Phi_n(S) = \# \{ k \in \{1, \dots, n\} : S_k = M_k \}.$$

Lemma 1.16. *Define a sequence of stopping times T_j, $j = 0, 1, \dots$ by setting T_0 and $T_j = \inf \{ n > T_{j-1} : S_n = M_n \}$. Then the random variables $S_{T_j} - S_{T_{j-1}}$, $j = 1, 2, \dots$ are iid with distribution*

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

Proof.

- $(S_{T_j} - S_{T_{j-1}})$ are iid by the Strong Markov Property.
- We need to compute the distribution of S_{T_1}.
  For $R_0 = \inf \{ n \ge 1 : S_n = 0 \}$ we have the following claim:

  Claim 1.17.

  $$\mathbb{E}[\sum_{n=0}^{R_0-1} \mathbb{1}_{\{S_n=i\}}] = 1 \quad \textit{for all } i \in \mathbb{Z}.$$

  Proof. (A reference for this proof is [3], in particular Proposition 31.)
  We will deduce the result from more general considerations for Markov processes.

Let $(X_n)_{n\geq 0}$ be the irreducible Markov chain on the countable state space E, with transition probabilities $P = (p_{xy})_{x,y\in E}$. Let τ_x^+ be defined as

$$\tau_x^+ \overset{\text{def}}{=} \inf\{n \geq 1 : X_n = x\} \quad \text{(return time to x).}$$

Furthermore, we can introduce for $x \in E$

$$\mu_x(y) \overset{\text{def}}{=} \mathbb{E}_x[\sum_{n=1}^{\tau_x^+} \mathbb{1}_{\{X_n=y\}}] \quad \text{(number of visits in y).}$$

Note that we obtain setting from the claim if we set

- Markov process X_n to be random walk S_n;
- state space E to be $\mathbb{Z}$;
- starting point $x = 0$, visited point $y = i$;
- stopping time $\tau_x^+ = R_0$
  (note that: $\mathbb{E}_x[\sum_{n=1}^{\tau_x^+} \mathbb{1}_{\{X_n=y\}}] = \mathbb{E}_x[\sum_{n=0}^{\tau_x^+-1} \mathbb{1}_{\{X_n=y\}}]$).

Lemma 1.18. *Let $x \in E$ be a recurent state of $P = (p_{xy})_{x,y\in E}$, then $\mu_x(\cdot)$ is an invariant measure of P.*

Proof. Recall:
A probability measure Π on E, such that $\Pi(x) \geq 0$ for all $x \in E$ is called invariant measure of the Markov chain with transition probabilities $P = (p_{xy})_{x,y\in E}$, if

$$\forall y \in E \;\; \Pi(y) = \sum_{x\in E} \Pi(x) p_{xy}.$$

We have to check the condition above

$$\begin{aligned}
\mu_x(y) &= \mathbb{E}_x[\sum_{n=1}^{\tau_x^+} \mathbb{1}_{\{X_n=y\}}] = \mathbb{E}_x[\sum_{n=1}^{\infty} \mathbb{1}_{\left\{X_n=y, n\leq\tau_x^+\right\}}] \\
&\overset{\text{Fubini}}{=} \sum_{n=1}^{\infty} \mathbb{P}_x[X_n = y, n \leq \tau_x^+] = \sum_{n=1}^{\infty}\sum_{z\in E} \mathbb{P}_x[X_{n-1} = z, X_n = y, n \leq \tau_x^+] \\
&\overset{\text{Markov}}{=} \sum_{n=1}^{\infty}\sum_{z\in E} p_{zy}\mathbb{P}_x[X_{n-1} = z, n \leq \tau_x^+] \overset{\text{Fubini}}{=} \sum_{z\in E} p_{zy} \sum_{n=1}^{\infty} \mathbb{P}_x[X_{n-1} = z, n \leq \tau_x^+] \\
&= \sum_{z\in E} p_{zy} \sum_{n=1}^{\infty} \mathbb{E}_x[\mathbb{1}_{\left\{X_{n-1}=z, n\leq\tau_x^+\right\}}] \overset{\text{Fubini}}{=} \sum_{z\in E} p_{zy}\mathbb{E}_x[\sum_{n=1}^{\tau_x^+} \mathbb{1}_{\{X_{n-1}=z\}}] \\
&= \sum_{z\in E} p_{zy}\mathbb{E}_x[\sum_{n=0}^{\tau_x^+-1} \mathbb{1}_{\{X_n=z\}} = \sum_{z\in E} p_{zy}\mu_x(z) \quad \forall y \in E.
\end{aligned}$$

Hence μ_x is invariant. ∎

Furthermore, we can consider an invariant measure on E defined as

$$\nu(x) = 1, \quad \forall x \in E.$$

By uniqueness (up to multiplication with a constant) of an invariant measure for recurrent, irreducible Markov chain we have

$$\mu_x = c \cdot \mu \quad (\forall y \in E \text{ being recurrent state}).$$

Coming back to our original setting we can find a constant value c

$$\mu_0(i) = \mathbb{E}[\sum_{n=0}^{R_0-1} \mathbb{1}_{\{S_n=i\}}] = c\nu(i), \quad \forall i \in \mathbb{Z} \quad \text{(because S is recurrent)}.$$

In particular for $i = 0$

$$\mu_0(0) = 1 = c\nu(0) = c \Rightarrow c = 1.$$

Hence

$$\mathbb{E}[\sum_{n=0}^{R_0-1} \mathbb{1}_{\{S_n=i\}}] = 1.$$

□

Now we come back to the Proof of Lemma 1.16.
Note, that $T_1 \leq R_0$ and that S takes only positive values on (T_1, R_0). For $i \leq 0$ it follows

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

This is because

$$1 = \mathbb{E}[\sum_{n=0}^{R_0-1} \mathbb{1}_{\{S_n=i\}}] = \mathbb{E}[\sum_{n=0}^{T_1-1} \mathbb{1}_{\{S_n=i\}}] + \underbrace{\mathbb{E}[\sum_{n=T_1}^{R_0-1} \mathbb{1}_{\{S_n=i\}}]}_{=0}).$$

Therefore for all $g \colon \mathbb{Z} \to \mathbb{Z}_+$:

$$\mathbb{E}[\sum_{n=0}^{T_1-1} g(S_n)] = \sum_{i=-\infty}^{0} g(i).$$

This is because

$$\mathbb{E}[\sum_{n=0}^{T_1-1} g(S_n)] = \mathbb{E}[\sum_{n=0}^{T_1-1} \sum_{i=-\infty}^{\infty} g(i)\mathbb{1}_{\{S_n=i\}}] = \mathbb{E}[\sum_{n=0}^{T_1-1} \sum_{i=-\infty}^{0} g(i)\mathbb{1}_{\{S_n=i\}}]$$
$$\stackrel{\text{Fubini}}{=} \sum_{i=-\infty}^{0} g(i)\mathbb{E}[\sum_{n=0}^{T_1-1} \mathbb{1}_{\{S_n=i\}}] = \sum_{i=-\infty}^{0} g(i).$$

Then for all $f : \mathbb{Z} \to \mathbb{Z}_+$:

$$\begin{aligned}
\mathbb{E}[f(S_{T_1})] &\overset{*}{=} \mathbb{E}[\sum_{k=0}^{\infty} \mathbb{1}_{\{k<T_1\}} f(S_{k+1}) \mathbb{1}_{\{S_{k+1}\geq 0\}}] \\
&\overset{\text{Fubini}}{=} \sum_{k=0}^{\infty} \mathbb{E}[\mathbb{1}_{\{k<T_1\}} f(S_{k+1}) \mathbb{1}_{\{S_{k+1}\geq 0\}}] \\
&= \sum_{k=0}^{\infty} \mathbb{E}[\mathbb{1}_{\{k<T_1\}} \sum_{j=0}^{\infty} \nu(j) f(S_k + j) \mathbb{1}_{\{S_k+j\}}] \\
&\overset{\text{Fubini}}{=} \mathbb{E}[\sum_{k=0}^{\infty} \mathbb{1}_{\{k<T_1\}} \sum_{j=0}^{\infty} \nu(j) f(S_k + j) \mathbb{1}_{\{S_k+j\}}] \\
&= \mathbb{E}[\sum_{k=0}^{T_1-1} (\sum_{j=0}^{\infty} \nu(j) f(S_k + j) \mathbb{1}_{\{S_k+j\}})] \\
&= \sum_{i=-\infty}^{0} \sum_{j=0}^{\infty} \nu(j) f(i+j) \mathbb{1}_{\{i+j\geq\}} \\
&\overset{**}{=} \sum_{m=0}^{\infty} f(m) \sum_{j=m}^{\infty} \nu(j)
\end{aligned}$$

(where (*) holds, since we want to pick first k, such that k+1 is nonnegative and take $f(S_{k+1})$;
(**) holds because we have

$$\begin{aligned}
&\sum_{i=-\infty}^{0} \sum_{j=0}^{\infty} \nu(j) f(i+j) \mathbb{1}_{\{i+j\}} \overset{\text{Fubini}}{=} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \nu(j) f(j-i) \mathbb{1}_{\{j-i\geq 0\}} \\
&= \sum_{j=0}^{\infty} \sum_{i=0}^{j} \nu(j) f(j-i) = \sum_{j=0}^{\infty} \sum_{m=0}^{j} \nu(j) f(m) = \sum_{m=0}^{\infty} f(m) \sum_{j=m}^{\infty} \nu(j)).
\end{aligned}$$

Now take $f(x) = \mathbb{1}_{\{x=k\}}$ in order to get the statement of the lemma.

□

Furthermore, we can note that S_{T_1} has finite first moment

$$\begin{aligned}
&\mathbb{E}[S_{T_1}] = \sum_{k=0}^{\infty} k\nu([k,\infty)) = \sum_{k=0}^{\infty} k \sum_{j=k}^{\infty} \nu(j) = \sum_{j=0}^{\infty} \sum_{k=0}^{j} k\nu(j) \\
&= \sum_{j=0}^{\infty} \frac{j(j+1)}{2} \nu(j) \overset{*}{=} \frac{\sigma^2}{2}.
\end{aligned}$$

In order to see that (*) holds, consider μ a ditribution with variance $\sigma^2 > 0$ and mean 1. Let X be a μ-ditributed random variable. Then we have

$$\sigma^2 + 1 = \mathbb{E}[X^2] = \sum_{k=0}^{\infty} k^2 \mu(k) = \sum_{k=0}^{\infty} k^2 \nu(k-1) = \sum_{k=-1}^{\infty} (k+1)^2 \nu(k)$$
$$= \sum_{k=0}^{\infty} (k+1)^2 \nu(k) = \sum_{k=0}^{\infty} k(k+1)\nu(k) + \underbrace{\sum_{k=0}^{\infty} (k+1)\nu(k)}_{=\sum_{k=0}^{\infty}(k+1)\mu(k+1))=1}$$
$$\Rightarrow \sum_{k=0}^{\infty} k(k+1)\nu(k) = \sigma^2.$$

We will now need the following lemma:

Lemma 1.19. *Let $\varepsilon \in (0, \frac{1}{4})$. We can find $\varepsilon' > 0$ and an integer $N \geq 1$ such that, for every $n \geq N$ and $l \in \{0, 1, \ldots, n\}$*

$$\mathbb{P}[|M_l - \frac{\sigma^2}{2} K_l| > n^{\frac{1}{4}+\varepsilon}] < \exp(-n^{\varepsilon'}).$$

We continue with the proof of Theorem 1.15. We will prove the lemma afterwards. Take $\varepsilon = \frac{1}{8}$. (M_n, K_n) has the same distribution as $(S_n - I_n, H_n)$. To see that it is enough to replace S with $\hat{S^n}$ and note that $M_n - S_n = -\hat{I^n}$, $S_n = \hat{S^n_n}$, since S and $\hat{S^n}$ have the same distribution. Hence, we can apply Lemma 1.19 in order to get (for n sufficiently large and $l \in \{1, \ldots, n\}$):

$$\mathbb{P}[|S_l - I_l - \frac{\sigma^2}{2} H_l| > n^{\frac{3}{8}}] < \exp(-n^{\varepsilon'})$$
$$\Rightarrow \mathbb{P}[\sup_{0 \leq l \leq n} |S_l - I_l - \frac{\sigma^2}{2} H_l| > n^{\frac{3}{8}}] < n \exp(-n^{\varepsilon'}).$$

Take an integer $A \geq 1$:

$$\mathbb{P}[\sup_{0 \leq t \leq A} |S_{[pt]} - I_{[pt]} - \frac{\sigma^2}{2} H_{[pt]}| > (Ap)^{\frac{3}{8}}] < (Ap) \exp(-(Ap)^{\varepsilon'}) \tag{1}$$
$$\Rightarrow \mathbb{P}[\underbrace{\sup_{0 \leq t \leq A} |\frac{S_{[pt]} - I_{[pt]}}{\sqrt{p}} - \frac{\sigma^2}{2} \frac{H_{[pt]}}{\sqrt{p}}| > \frac{A^{\frac{3}{8}}}{p^{\frac{1}{8}}}}_{A_p :=}] < (Ap) \exp(-(Ap)^{\varepsilon'}). \tag{2}$$

Furthermore, from

$$\sum_{p \geq 1} \mathbb{P}[A_p] \leq \sum_{p \geq 1} (Ap) \exp(-(Ap)^{\varepsilon'}) < \infty$$

(last inequality holds, since $n\exp(-n^{\varepsilon'}) \leq \frac{1}{n^2}$ for all $\varepsilon' > 0$ and n sufficiently large). Hence by the Borel-Cantelli Lemma we have

$$\mathbb{P}[\limsup_{p\to\infty} A_p] = 0$$

(i.e. the probability that infinitely many A_p's occur is 0) and together with the fact that $A^{\frac{3}{8}}p^{-\frac{1}{8}} \xrightarrow[p\to\infty]{} 0$ we obtain

$$\sup_{0\leq t\leq A} |\frac{S_{[pt]} - I_{[pt]}}{\sqrt{p}} - \frac{\sigma^2}{2}\frac{H_{[pt]}}{\sqrt{p}}| \xrightarrow[p\to\infty]{} 0 \quad \text{almost surely.}$$

Hence by Donsker's invariance theorem we obtain

$$(\frac{1}{\sqrt{p}}(S_{[pt]} - I_{[pt]}),\ t \geq 0) \xrightarrow[p\to\infty]{(d)} (\sigma(B_t - \inf_{0\leq s\leq t} B_s),\ t \geq 0)$$

and by a theorem due to Paul Lévy we know that $(B_t - \inf_{0\leq s\leq t} B_s,\ t \geq 0)$ is a reflected Brownian Motion. (Confirm with Thm. 3.6.17. in [4].) □

Let us now prove Lemma 1.19. We will need one more lemma:

Lemma 1.20. *Let $Y_1, Y_2, \ldots$ be a sequence of iid real random variables. We assume that there exists a number $\lambda > 0$, such that $\mathbb{E}[\exp(\lambda|Y_1|)] < \infty$ and that $\mathbb{E}[Y_1] = 0$. Then for all $\alpha > 0$, we can choose N sufficiently large, such that for all $n \geq N$ and $l \in \{1, 2, \ldots, n\}$*

$$\mathbb{P}[|Y_1 + \ldots + Y_l| > n^{\frac{1}{2}+\alpha}] \leq \exp(-n^{\frac{\alpha}{2}}).$$

Proof. First observe the following fact:

$$\mathbb{P}[|Y_1 + \ldots + Y_l| > n^{\frac{1}{2}+\alpha}] = \underbrace{\mathbb{P}[Y_1 + \ldots + Y_l > n^{\frac{1}{2}+\alpha}]}_{1} + \underbrace{\mathbb{P}[-(Y_1 + \ldots + Y_l) > n^{\frac{1}{2}+\alpha}]}_{2}$$

1. Since $\mathbb{E}[\exp(\lambda|Y_1|)] < \infty$ (in particular all moments of Y_1 are finite) we can Taylor-expand the function $\mathbb{E}[\exp(\lambda Y_1)]$:

   $$\mathbb{E}[\exp(\lambda Y_1)] = \mathbb{E}[1 + \lambda Y_1 + \frac{1}{2}\lambda^2 Y_1^2 + o(\lambda^2)] = 1 + c\lambda^2 + o(\lambda^2) \ \ (\lambda \to 0)$$

   (where $c := \frac{1}{2}var(Y_1)$). Hence there exists a constact C, such that for sufficiently small $\lambda > 0$:

   $$\mathbb{E}[\exp(\lambda Y_1)] \leq e^{C\lambda^2}.$$

   Furthermore, we have

   $$\mathbb{P}[Y_1 + \ldots + Y_l > n^{\frac{1}{2}+\alpha}] \overset{*}{\leq} e^{-n^{\frac{1}{2}+\alpha}\lambda}\mathbb{E}[e^{\lambda(Y_1+\ldots+Y_l)}] \leq e^{-n^{\frac{1}{2}+\alpha}\lambda}e^{C\lambda^2 n}$$

((*) holds because $\mathbb{1}_{\left\{Y_1+\ldots+Y_l-n^{\frac{1}{2}+\alpha}\geq 0\right\}} \leq e^{\lambda(Y_1+\ldots+Y_l-n^{\frac{1}{2}+\alpha})}$ almost surely).

Taking $\lambda := n^{-\frac{1}{2}}$ (for n large enough) we have

$$\mathbb{P}[Y_1+\cdots+Y_l > n^{\frac{1}{2}+\alpha}] \leq e^C e^{-n^\alpha}.$$

2. We proceed analogously as in 1. in order to obtain a similar bound

$$\mathbb{E}[\exp(-\lambda Y_1)] \leq e^{C\lambda^2}.$$

And further

$$\mathbb{P}[-(Y_1+\ldots+Y_l) > n^{\frac{1}{2}+\alpha}] \leq e^{-n^{\frac{1}{2}+\alpha}\lambda}\mathbb{E}[e^{-\lambda(Y_1+\ldots+Y_l)}] \leq e^{-n^{\frac{1}{2}+\alpha}\lambda}e^{C\lambda^2 n}$$
$$\xRightarrow{\lambda:=n^{-\frac{1}{2}}} \mathbb{P}[-(Y_1+\ldots+Y_l) > n^{\frac{1}{2}+\alpha}] \leq e^C e^{-n^\alpha}.$$

Hence, from 1. and 2. we finally obtain

$$\mathbb{P}[|Y_1+\cdots+Y_l| > n^{\frac{1}{2}+\alpha}] \leq 2e^C e^{-n^\alpha} \leq e^{-n^{\frac{\alpha}{2}}}$$

where the last inequality holds for n sufficiently large. □

We can finally prove Lemma 1.19.

Proof. Take $\alpha \in (0, \frac{\varepsilon}{2})$ and introduce $m_n \overset{\text{def}}{=} [n^{\frac{1}{2}+\alpha}]$. Then for all $l \in \{0, 1, \ldots, n\}$

$$\mathbb{P}[|M_l - \frac{\sigma^2}{2}K_l| > n^{\frac{1}{4}+\varepsilon}] \leq \underbrace{\mathbb{P}[K_l > m_n]}_{1} + \underbrace{\mathbb{P}[|M_l - \frac{\sigma^2}{2}K_l| > n^{\frac{1}{4}+\varepsilon},\ K_l \leq m_n]}_{2}.$$

It is enough to find bounds for 1. and 2.:

$$\begin{aligned}
2:\quad & \mathbb{P}[|M_l - \frac{\sigma^2}{2}K_l| > n^{\frac{1}{4}+\varepsilon},\ K_l \leq m_n] \\
& \leq \mathbb{P}[\sup_{0\leq k\leq m_n} |\sum_{j=1}^{k}((S_{T_j} - S_{T_{j-1}}) - \frac{\sigma^2}{2})| > n^{\frac{1}{4}+\varepsilon}] \\
& \overset{*}{\leq} \mathbb{P}[\sup 0 \leq k \leq m_n| \sum_{j=1}^{k}((S_{T_j} - S_{T_{j-1}}) - \frac{\sigma^2}{2})| > m_n^{\frac{1}{2}+\varepsilon}] \\
& \overset{**}{\leq} m_n \exp(-m_n^{\frac{\varepsilon}{2}})
\end{aligned}$$

(*) holds, because

$$m_n^{\frac{1}{2}+\varepsilon} = [n^{\frac{1}{2}+\varepsilon}]^{\frac{1}{2}+\varepsilon} \leq [n^{\frac{1}{2}+\frac{\varepsilon}{2}}]^{\frac{1}{2}+\varepsilon} \leq n^{\frac{1}{4}+\frac{3}{4}\varepsilon+\frac{\varepsilon^2}{2}} \leq n^{\frac{1}{4}+\varepsilon}$$

for ε small enough;
(**) holds by Lemma 1.20. Note, that we can use the lemma, since

$$\mathbb{E}[S_{T_1} - \frac{\sigma^2}{2}] = 0$$

and further

$$\mathbb{E}[\exp(\lambda|S_{T_1} - \frac{\sigma^2}{2}|)] \leq \max(\mathbb{E}[\exp(\lambda S_{T_1})], \underbrace{\exp(\lambda\frac{\sigma^2}{2})}_{<\infty})$$

(because S_{T_1} is non-negative). Hence, we get

$$\mathbb{E}[\exp(\lambda S_{T_1})] \overset{\text{L. 1.16}}{=} \sum_{k=0}^{\infty} e^{\lambda k}\nu([k,\infty))$$
$$= \sum_{j=0}^{\infty} e^{\lambda k} \sum_{k=0}^{j} e^{\lambda k}\nu(j) = \sum_{j=0}^{\infty} \nu(j) \frac{e^{\lambda(j+1)}-1}{e^{\lambda}-1}$$
$$= \frac{e^{\lambda}}{e^{\lambda}-1} \underbrace{\sum_{j=0}^{\infty} \nu(j)e^{\lambda j}}_{<\infty \text{ (finite exp. moments)}} - \frac{1}{e^{\lambda}-1} \underbrace{\sum_{j=0}^{\infty} \nu(j)}_{<1} < \infty.$$

Thus, the assumptions of Lemma 1.20 are satisfied.

$$1: \ \ \mathbb{P}[K_l > m_n] \leq \mathbb{P}[K_n > m_n] \leq \mathbb{P}[S_{T_{m_n}} \leq M_n].$$

Hence

$$\mathbb{P}[K_l > m_n] \leq \underbrace{\mathbb{P}[S_{T_{m_n}} \leq n^{\frac{1}{2}+\frac{\alpha}{2}}]}_{I} + \underbrace{\mathbb{P}[M_n > n^{\frac{1}{2}+\frac{\alpha}{2}}]}_{II}.$$

For n large we can apply Lemma 1.20 (in order to get bound for II):

$$\mathbb{P}[M_n > n^{\frac{1}{2}+\frac{\alpha}{2}}] = \mathbb{P}[\sup_{0\leq k\leq n} S_k > n^{\frac{1}{2}+\frac{\alpha}{2}}] \leq n\exp(-n^{\frac{\alpha}{4}})$$

(the last equality holds by Lemma 1.20, since $S_k = Y_1 + \ldots + Y_k$ has mean 0 and finite exponential moments).
So now we only need to find a bound for I.

$$\mathbb{P}[S_{T_{m_n}} \leq n^{\frac{1}{2}+\frac{\alpha}{2}}] = \mathbb{P}[S_{T_{m_n}} - \frac{\sigma^2}{2}m_n \leq n^{\frac{1}{2}+\frac{\alpha}{2}} - \frac{\sigma^2}{2}m_n]$$
$$\leq \mathbb{P}[S_{T_{m_n}} - \frac{\sigma^2}{2}m_n \leq n^{\frac{1}{2}+\frac{\alpha}{2}}] \leq \exp(n^{-\frac{\alpha}{4}})$$

(the last inequality holds by Lemma 1.20, since $(S_{T_{m_n}} - \frac{\sigma^2}{2}m_n)$ is a sum of centered iid random variables, hence

$$\mathbb{E}[S_{T_{m_n}} - \frac{\sigma^2}{2}m_n] = 0$$

and

$$\mathbb{E}[\exp(\lambda|S_{T_{m_n}} - \frac{\sigma^2}{2}m_n|)] \leq \mathbb{E}[\exp(\lambda S_{T_{m_n}})] = \mathbb{E}[\exp(\lambda S_{T_1})]^{m_n} < \infty).$$

Hence, we get

$$\begin{aligned}\mathbb{P}[|M_l - \frac{\sigma^2}{2}K_l| > n^{\frac{1}{4}+\varepsilon}] &\leq m_n \exp(-m_n^{\frac{\varepsilon}{2}}) + n\exp(-n^{\frac{\alpha}{4}}) + \exp(-n^{\alpha 4}) \\ &\overset{*}{\leq} (m_n + n + 1)\exp(-n^{\frac{\alpha}{4}}) \leq \exp(-n^{\varepsilon'}) \quad \text{for some } \varepsilon' > 0\end{aligned}$$

((*) holds, because $\exp(-m_n^{\frac{\varepsilon}{2}}) \leq \exp(-n^{\frac{\alpha}{4}})$, since $[n^{\alpha+\frac{1}{2}}]^{\frac{\varepsilon}{2}} > n^{\frac{\alpha}{4}}$). □

1.5 Galton-Watson trees with a fixed progeny

In this section we will prove a similar result as in Theorem 1.15, but for the tree having fixed number of individuals.
Furthermore, we assume that μ has finite exponential moments.
For every $p \geq 1$ we denote by $\theta^{(p)}$ a μ-Galton-Watson tree conditioned to have $\#(\theta) = p$.
For this to hold we need an assumption that $\mathbb{P}[\#(\theta) = p] > 0$ for all $p \geq 1$ (which holds, if $\mu(1) \geq 0$).
Furthermore, we denote by $(H_k^{(p)})_{0 \leq k \leq p}$ the height process of $\theta^{(p)}$, with the convention $H_p^{(p)} = 0$.

Theorem 1.21. *We have*

$$(\frac{1}{\sqrt{p}}H_{[pt]}^{(p)}, 0 \leq t \leq 1) \xrightarrow[p\to\infty]{(d)} (\frac{2}{\sigma}e_t,\ 0 \leq t \leq 1).$$

In order to prove this theorem we will need a lemma that can viewed as a conditional Donsker's invariance theorem (hence, we will not prove it).

Lemma 1.22. *The distribution of the process $(\frac{1}{\sqrt{p}}S_{[pt]}, 0 \leq t \leq 1)$ under the conditional probability $\mathbb{P}(\cdot|T_1 = p)$ converges as p tends to ∞ to the law of $(\sigma e_t, 0 \leq t \leq 1)$.*

Let's prove now Theorem 1.21.

Proof. Let $(H_n, n \geq 0)$ be the height process associated with the sequence of independent μ-Galton-Watson trees. We will assume that H is given in terms of the random walk as in Proposition 1.14. By T_1 we will denote the number of vertices of a first tree

$$T_1 = \inf\{n \geq 1 : \ H_n = 0\} = \inf\{n \geq 0 : \ S_n = -1\}.$$

By a combinatorial argument we obtain

$$\mathbb{P}[T_1 = p] = \frac{1}{p}\mathbb{P}[S_p = -1].$$

In order to get this formula consider a random walk on $[0, p]$ with ν-distributed jumps, such that $S_0 = 0$ and $S_p = -1$. It is easy to see that there is only one cyclic permutation of increments of S, such that p is the first time, when S reaches -1. It is namely the one that shifts the first time, when S reaches its minimum to p. The set of similar combinatorial results for random walks is known in the literature as the "Ballot Theorem".
On the other hand from the local Central Limit Theorem we have

$$\lim_{p\to\infty} \sqrt{p}\,\mathbb{P}[S_p = -1] = \frac{1}{\sigma\sqrt{2\pi}}.$$

As a reference to this result we use [5]. Recall:

Definition 9. *(D1, Ch.1, p.42,[5]) A random walk with transition function $P(x, y)$ on R is called strongly aperiodic if it has the property that for each x in R, the smallest subgroup of R which contains the set*

$$x + \Sigma = [y|y = x + z, \textit{ where } P(0, z) > 0]$$

is R itself.

Note, that the condition $\mu(1) > 0$ that we assumed at the beginning of this subsection guarantees us the strong aperiodicity.
Furthermore, we have the following result:

Proposition 1.23. *(P9, Ch.2, p.75,[5]) For strongly aperiodic random walk of dimension $d \geq 1$ with mean $\mu = 0$ and finite second moments,*

$$\lim_{n\to\infty} (2\pi n)^{\frac{d}{2}} P_n(0, x) = |Q|^{-\frac{1}{2}},\ x \in R,$$

where $|Q|$ is the determinant of the quadratic form

$$Q(\theta) = \sum x \in R(x \cdot \theta)^2 P(0, x).$$

It is easy to see that in the one-dimensional case we have $|Q| = \sigma^2$ and hence, we directly get the formula used above.

From that we obtain

$$\mathbb{P}[T_1 = p] \sim \frac{1}{\sigma\sqrt{2\pi p^3}} \text{ as } p \to \infty. \tag{3}$$

From the proof of Theorem 1.15 ((1) with $A = 1$) we recall that we can find $\varepsilon > 0$, such that for p large enough

$$\mathbb{P}[\sup_{0\leq t\leq 1} |\frac{\sigma^2}{2}\frac{H_{[pt]}}{\sqrt{p}} - \frac{S_{[pt]} - I_{[pt]}}{\sqrt{p}}| > p^{-\frac{1}{8}}] < \exp(-p^{\varepsilon}).$$

Comparing with (3) we get

$$\mathbb{P}[\sup_{0\le t\le 1} |\frac{\sigma^2}{2}\frac{H_{[pt]}}{\sqrt{p}} - \frac{S_{[pt]} - I_{[pt]}}{\sqrt{p}}| > p^{-\frac{1}{8}} \mid T_1 = p] < \exp(-p^{\varepsilon'}) \ \ (\text{for } \varepsilon' < \varepsilon).$$

Furthermore, since $I_n = 0$ for $0 \le n < T_1$, hence, for p large enough

$$\mathbb{P}[\sup_{0\le t\le 1} |\frac{\sigma^2}{2}\frac{H_{[pt]}}{\sqrt{p}} - \frac{S_{[pt]}}{\sqrt{p}}| > p^{-\frac{1}{8}} \mid T_1 = p] < \exp(-p^{\varepsilon'}).$$

Note, that $(H_p^{(k)},\ 0 \le k \le p)$ has the same distribution as $(H_k,\ 0 \le k \le p)$ under $\mathbb{P}[\cdot|T_1 = p]$. Hence, Theorem 1.21 follows from the last bound together with Lemma 1.22 (by the same argument as in the proof of Theorem 1.15). □

1.6 Convergence of contour functions

We will see that the convergence result for height functions from the previous section can also be shown for contour functions.
Let us consider again a sequence $\theta_1, \theta_2, \dots$ of independent μ-Galton-Watson trees. Let $(C_t,\ t \ge 0)$ denote the contour process obtained by concatenating the contour functions of $\theta_1, \theta_2, \dots$. We will use the following convention:
$C_t(\theta)$ is a contour function defined for $0 \le t \le \xi(\theta) := 2\#(\theta) - 1$, by taking $C_t = 0$, if $\zeta(\theta) \le t \le \xi(\theta)$ (where $\zeta(\theta) = 2(\#(\theta) - 1)$). We obtain $(C_t,\ t \ge 0)$ by concatenating the functions $(C_t(\theta_1),\ 0 \le t \le \xi(\theta_1)),\ (C_t(\theta_2),\ 0 \le t \le \xi(\theta_2)), \dots$.
For $n \ge 0$ define a sequence:

$$J_n = 2n - H_n + I_n.$$

Note, that J_n is strictly increasing and that $J_n \ge n$ for all n. This is because $J_0 \ge 0$ and $J_{n+1} - J_n = 2 - \underbrace{(H_{n+1} - H_n)}_{\le 1} + \underbrace{(I_{n+1} - I_n)}_{\in\{-1,0\}} \ge 1$ (because cases 1 and -1 cannot occur at the same time).

Claim 1.24. $[J_n, J_{n+1}]$ *is the time interval during which the contour process goes from the individual n to the individual n+1.*

Proof. We will proceed by induction on n. Let $T_{n,n+1}$ be time needed to move from the individual n to n+1. We need to show $T_{n,n+1} = J_{n+1} - J_n$.

- For $[J_0, J_1]$ we have $J_1 - J_0 = 1$ and clearly $T_{0,1} = 1$.
- Assume that the statement holds for $[J_{n-1}, J_n]$. We have to show it for $[J_n, J_{n+1}]$ (by the induction assumption it is enough to show $T_{n,n+1} = J_{n+1} - J_n$). There are two cases, which we need to consider
  1. n and n+1 are in the same tree (so $I_n = I_{n+1}$). Then
     - $H_{n+1} > H_n$ (so $H_{n+1} - H_n = 1$), hence, $J_{n+1} - J_n = 1$ and $T_{n,n+1} = 1$

– $H_{n+1} \leq H_n$, so $J_{n+1} - J_n = 2 - (H_{n+1} - H_n)$ and $T_{n,n+1} = H_n - (H_{n+1} - 1) + 1 = 2 - (H_{n+1} - H_n)$, hence, $T_{n,n+1} = J_{n+1} - J_n$.

2. n and n+1 are in different trees (so $I_{n+1} - I_n = -1$, $H_{n+1} = 0$). From that we have $J_{n+1} - J_n = 2 + H_n + (-1) = 1 + H_n$. On the other hand $T_{n,n+1} = H_n + 1$ (by the construction of the contour process).

From the computation above together with the induction assumption we get the statement.

□

Furthermore, we get the following formulas for C_t:

Claim 1.25. *For* $t \in [J_n, J_{n+1}]$ *we have*

$$C_t = H_n - (t - J_n), \text{ if } t \in [J_n, J_{n+1} - 1]$$
$$C_t = (H_{n+1} - (J_{n+1} - t))^+, \text{ if } t \in [J_{n+1} - 1, J_{n+1}].$$

Proof. We will use induction to prove this claim.

- For $t \in [J_0, J_1]$ we have $C_t = \begin{cases} t, & \text{if } H_1 = 0 \\ 0, & \text{if } H_1 = 0 \end{cases} = (H_1 - (J_1 - t))^+$ (because $J_1 = 1$).
- Assume the statement holds for $t \in [J_{n-1}, J_n]$. We want to show it for $t \in [J_n, J_{n+1}]$.
  By the induction hypothesis we know that at the time J_n we are at u_n. Then by the construction of the contour process we know that during the time $[J_n, J_{n+1} - 1]$ the contour process goes back to the last common ancestor of u_n and u_{n+1}, hence, $C_t = H_n - (t - J_n)$ (since we move with unit speed).
  Furthermore, on the interval $[J_{n+1} - 1, J_{n+1}]$ we move from the last common ancestor of u_n and u_{n+1} to u_{n+1}, hence, $C_t = H_{n+1} - (J_{n+1} - t)^+$.
  The statement follows by the induction assumption.

□

From the formulas above we get the following bound:

Claim 1.26.

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

Proof. There two cases that we need to consider:

1. Let $t \in [J_n, J_{n+1} - 1]$, then

$$\begin{aligned} |C_t - H_n| &= |H_n - (t - J_n) - H_n| = |t - J_n| \leq |(J_{n+1} - 1) - J_n| \\ &= |1 - (H_{n+1} - H_n) + \underbrace{(I_{n+1} - I_n)}_{\in\{-1,0\}}| \leq |H_{n+1} - H_n| + 1. \end{aligned}$$

2. Let $t \in [J_{n+1}-1, J_{n+1}]$, then

$$
\begin{aligned}
|C_t - H_n| &= |(H_{n+1} - (J_{n+1} - t))^+ - H_n| \\
&= \begin{cases} H_n, & \text{if } (H_{n+1} - (J_{n+1} - t))^+ = 0 \\ 3n+1, & \text{if } (H_{n+1} - (J_{n+1} - t))^+ > 0 \end{cases} \\
&\le \begin{cases} |H_n - H_{n+1}| + 1, & \text{if } (H_{n+1} - (J_{n+1} - t))^+ = 0 \\ |H_{n+1} - H_n| + |J_{n+1} - t|, & \text{if } (H_{n+1} - (J_{n+1} - t))^+ > 0 \end{cases} \\
&\le |H_{n+1} - H_n| + 1.
\end{aligned}
$$

Hence, the claim follows. □

Furthermore, we define a random function $\varphi\colon \mathbb{R}_+ \to \{0, 1, \ldots\}$ by

$$\varphi(t) = n \ \text{ iff } \ t \in [J_n, J_{n+1}).$$

From Claim 1.26 we obtain

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

Moreover, we get the following bound

Claim 1.27.

$$\sup_{t\in[0,m]} |\varphi(t) - \frac{t}{2}| \le \sup_{t\in[0,J_m]} |\varphi(t) - \frac{t}{2}| \le \frac{1}{2} \sup_{n\le m} H_n + \frac{1}{2}|I_m| + 1. \tag{4}$$

Proof. The first inequality is clear, since $J_n \ge n$ for all n.
To show the second inequality consider $t \in [J_n, J_{n+1})$ $(n+1 \le m)$, then

$$|\varphi(t) - \frac{t}{2}| = |n - \frac{t}{2}| = \frac{1}{2} \underbrace{|J_n + H_n - I_n - t|}_{=:f(t)}.$$

Note that f is linear in t, hence, $\sup_{t\in[J_n,J_{n+1})} f(t)$ is obtained in one of the endpoints of the interval, hence

$$|J_n + H_n - I_n - t| \le \max\{|J_n + H_n - I_n - J_n|, |J_n + H_n - I_n - J_{n+1}|\}.$$

Furthermore, it holds

$$
\begin{aligned}
&|J_n + H_n - I_n - J_{n+1}| = |J_{n+1} - J_n - H_n + I_n| \\
&= |2 - (H_{n+1} - H_n) + (I_{n+1} - I_n) - H_n + I_n| \\
&= |2 - H_{n+1} + I_{n+1}| \le 2 + |H_{n+1}| + |I_{n+1}|.
\end{aligned}
$$

Hence, we get a bound

$$|J_n + H_n - I_n - t| \le 2 + \max\{H_n, H_{n+1}\} + |I_n|.$$

Furthermore, we obtain

$$|\varphi(t) - \frac{t}{2}| \leq \frac{1}{2}\max\{H_n, H_{n+1}\} + \frac{1}{2}|I_n| + 1.$$

Finally putting all results together we have

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

$\square$

Now we come to the main result of this section (it is an analogon of Theorem 1.21 for contour functions).

Theorem 1.28.

$$(\frac{1}{\sqrt{p}}C^{(p)}_{2pt},\ t \geq 0) \xrightarrow[p\to\infty]{(d)} (\frac{2}{\sigma}e_t,\ t \geq 0)$$

where e_t is a normalized Brownian excursion.

Proof. For $p \geq 1$, set $\varphi_p(t) = p^{-1}\varphi(pt)$. Using bound (4) obtained above we have for all $m \geq 1$

$$\sup_{t\leq m} |\frac{1}{\sqrt{p}}C^{(p)}_{2pt} - \frac{1}{\sqrt{p}}H^{(p)}_{p\varphi_p(2t)}| \leq \frac{1}{\sqrt{p}} + \frac{1}{\sqrt{p}}\sup_{t\leq 2m} |H^{(p)}_{[pt]+1} - H^{(p)}_{[pt]}| \xrightarrow[p\to\infty]{(P)} 0 \tag{5}$$

by Theorem 1.21.
Furthermore, by a conditional Donsker's invariance theorem we get

$$\frac{1}{\sqrt{p}}I^{(p)}_{mp} \xrightarrow[p\to\infty]{(d)} \sigma \inf_{t\leq m} e_t.$$

Then by Claim 1.27 we obtain

$$\sup_{t\leq m} |\varphi(2t) - t| \leq \frac{1}{p}\sup_{k\leq 2mp} H^{(p)}_k + \frac{1}{p}|I^{(p)}_{2mp}| + \frac{2}{p} \xrightarrow[p\to\infty]{(P)} 0. \tag{6}$$

Hence, the statement follows by (5), (6) and Theorem 1.21. $\square$

We have shown that the rescaled height processes (and contour processes) of the large Galton-Watson trees converge in distribution towards the normalized Brownian excursion.
In the next section we will try to reinterpret those results in order to get the convergence of trees, but for that we will need to define first how the limiting trees should look like.

2 Real Trees and their Coding by Brownian Excursions

2.1 Real trees

Definition 10. *A compact metric space $(\mathcal{T}, d)$ is a real tree, if the following two properties hold for every $a, b \in \mathcal{T}$*

1. *there is a unique isometric map $f_{a,b}$ from $[0, d(a,b)]$ into $\mathcal{T}$, such that $f_{a,b}(0) = a$ and $f_{a,b}(d(a,b)) = b$;*
2. *if q is continuous, injective map from $[0,1]$ into $\mathcal{T}$, such that $q(0) = a$ and $q(1) = b$, then we have $q([0,1]) = ([0, d(a,b)])$.*

A rooted tree is a real tree $(\mathcal{T}, d)$ with a distinguished vertex $\rho = \rho(\mathcal{T})$ called the root.

Further in the text we will only consider rooted trees.

Remark 2.1. *Consider a rooted real tree $(\mathcal{T},d)$:*

- *the range of the mapping $f_{a,b}$ in 1. is $[[a,b]]$ (the line segment between a and b in the tree);*
- *$[[\rho, a]]$ (the line segment between the root and a in the tree) is the ancestral line of vertex a;*
- *we define a partial order on the tree by $a \prec b$ (a is an ancestor of b) iff $a \in [[\rho, b]]$;*
- *if $a, b \in \mathcal{T}$, there is a unique $c \in \mathcal{T}$, such that $[[\rho, a]] \cap [[\rho, b]] = [[\rho, c]]$. c is called the most recent ancestor to a and b;*
- *we define the multiplicity of a vertex $a \in \mathcal{T}$ as the number of connected components of $\mathcal{T} \setminus \{a\}$;*
- *vertices of $\mathcal{T} \setminus \{\rho\}$ with multiplicity one are called leaves.*

We want to study the convergence of real trees, for this reason we need a notion of distance between them. We will introduce the Gromov-Hausdorff distance between compact metric spaces.
Let (E, δ) be a metric space. Denote as $\delta_{Haus}(K, K')$ the usual Hausdorff metric between compact subsets of E

$$\delta_{Haus}(K, K') = \inf\left\{\varepsilon > 0 : \ K \subset U_\varepsilon(K') \text{ and } K' \subset U_\varepsilon(K)\right\}$$

where $U_\varepsilon(K) \overset{\text{def}}{=} \{x \in E : \ \delta(x, K) \leq \varepsilon\}$.
Then we define the Gromov-Hausdorff distance as following

Definition 11. *Let $\mathcal{T}$,$\mathcal{T}'$ be two rooted compact metric spaces, with respective roots ρ and ρ'. We define a distance between $\mathcal{T}$,$\mathcal{T}'$ (Gromov-Hausdorff distance) by*

$$d_{GH}(\mathcal{T}, \mathcal{T}') = \inf\left\{\delta_{Haus}(\phi(\mathcal{T}, \mathcal{T}')) \vee \delta(\phi(\rho), \phi'(\rho'))\right\}$$

where the infimum is over all choices of a metric space (E, δ) and all isometric embeddings $\phi\colon \mathcal{T} \to E$ and $\phi'\colon \mathcal{T}' \to E$ of $\mathcal{T}$ and $\mathcal{T}'$ into (E, δ).

Remark 2.2. • *Two rooted compact metric spaces $\mathcal{T}_1$ and $\mathcal{T}_2$ are called equivalent, if there is a root preserving isometry that maps $\mathcal{T}_1$ onto $\mathcal{T}_2$.*

- *We denote $\mathbb{T}$ the set of all (equivalence classes of) rooted real trees.*
- *d_{GH} defines a metric on $\mathbb{T}$*
  - *$d_{GH}(\mathcal{T},\mathcal{T}') = 0$ iff $\mathcal{T} = \mathcal{T}'$ (by definition of equivalence classes);*
  - *$d_{GH}(\mathcal{T},\mathcal{T}') = d_{GH}(\mathcal{T}',\mathcal{T})$ (because δ_{Haus} and δ are metrics);*
  - *$d_{GH}(\mathcal{T},\mathcal{T}') + d_{GH}(\mathcal{T}',\mathcal{T}'') \leq d_{GH}(\mathcal{T},\mathcal{T}'')$ (because δ_{Haus} and δ are metrics).*

We will also use the alternative definition of d_{GH}.
For two given compact metric spaces $(\mathcal{T}_1, d_1)$ and $(\mathcal{T}_2, d_2)$ we define a correspondence between $\mathcal{T}_1$ and $\mathcal{T}_2$ as a subset $\mathcal{R}$ of $\mathcal{T}_1 \times \mathcal{T}_2$, such that for all $x_1 \in \mathcal{T}_1$ there exists $x_2 \in \mathcal{T}_2$, such that $(x_1, x_2) \in \mathcal{R}$ and conversely for all $y \in \mathcal{T}_2$ there is $y_1 \in \mathcal{T}_1$, such that $(y_1, y_2) \in \mathcal{R}$.
The distorsion of the correspondence $\mathcal{R}$ is defined as

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

Then we have following claim:

Claim 2.3. *Let $\mathcal{T}$, $\mathcal{T}'$ be two rooted compact metric spaces with respective roots ρ and ρ'. Then*

$$d_{GH}(\mathcal{T},\mathcal{T}') = \frac{1}{2} \inf_{\mathcal{R}\in\mathcal{C}(\mathcal{T},\mathcal{T}'),\,(\rho,\rho')\in\mathcal{R}} dis(\mathcal{R})$$

(where $\mathcal{C}(\mathcal{T},\mathcal{T}')$ denotes all correspondences between $\mathcal{T}$ and $\mathcal{T}'$).

Proof. (see [4], Lemma 2.3.)
We show first $\frac{1}{2}\inf \mathrm{dis}(\mathcal{R}) \leq d_{GH}(\mathcal{T},\mathcal{T}')$.
For any root-invariant isometric copies $\phi(\mathcal{T})$,$\phi'(\mathcal{T}')$ embedded in a metric space (E,δ) and such that $r > d_{GH}(\mathcal{T},\mathcal{T}')$. We can define

$$\mathcal{R} = \left\{(x,x') \in \mathcal{T}\times\mathcal{T}' \colon \delta(\phi(x),\phi'(x')) < r\right\}.$$

Note, that $\mathcal{R}$ is a correspondence between $\mathcal{T}$ and $\mathcal{T}'$, such that $(\rho,\rho') \in \mathcal{R}$.
Furthermore, we have

$$dis(\mathcal{R}) < 2r.$$

This is because for $x_1, y_1 \in \mathcal{T}$, $x_2, y_2 \in \mathcal{T}'$ by the triangle inequality we have

$$\begin{aligned}&|d_{\mathcal{T}}(x_1,y_1) - d_{\mathcal{T}'}(x_2,y_2)| = |\delta(\phi(x_1),\phi(y_1)) - \delta(\phi'(x_2),\phi'(y_2))|\\ &\leq \delta(\phi(x_1),\phi(y_1)) + \delta(\phi'(x_2),\phi'(y_2)) < 2r.\end{aligned}$$

Since we can choose r arbitrarily small, such that $r > d_{GH}(\mathcal{T}, \mathcal{T}')$, hence, we obtain

$$\frac{1}{2} \inf dis(\mathcal{R}) \leq d_{GH}(\mathcal{T}, \mathcal{T}').$$

On the other hand we have to show

$$d_{GH}(\mathcal{T}, \mathcal{T}') \leq \frac{1}{2} \inf dis(\mathcal{R}).$$

Given a correspondence $\mathcal{R}$ containing (ρ, ρ') we can define a distance d on $\mathcal{T} \bigsqcup \mathcal{T}'$ (a disjoint union of $\mathcal{T}$ and $\mathcal{T}'$) as

- $d|_{\mathcal{T}\times\mathcal{T}} = d_{\mathcal{T}}$;
- $d|_{\mathcal{T}'\times\mathcal{T}'} = d_{\mathcal{T}'}$;
- for $x \in \mathcal{T}$, $x' \in \mathcal{T}'$

$$d(x, x') = \inf \left\{ d_{\mathcal{T}}(x, y) + d_{\mathcal{T}'}(x', y') + \frac{1}{2} dis(\mathcal{R})\colon (y, y') \in \mathcal{R} \right\}.$$

Note, that

- if $(x, x') \in \mathcal{R}$, then $d(x, x') = \frac{1}{2} dis(\mathcal{R})$
- d is a metric
  - $d(x, y) = 0$ iff $x = y$;
  - d is symmetric (it follows since $d_{\mathcal{T}}$ and $d_{\mathcal{T}'}$ are metrics);
  - d satisfies the triangle inequality
    In order to see this we need to check two cases (all other follow by symmetry).
    1. Let $x, y \in \mathcal{T}$ and $x' \in \mathcal{T}'$. We want to show

$$d(x, x') + d(x', y) \geq d(x, y).$$

$$\begin{aligned} P &= d_{\mathcal{T}}(x, y); \\ L &= \inf \left\{ d_{\mathcal{T}}(x, z) + d_{\mathcal{T}'}(x', z')\colon (z, z') \in \mathcal{R} \right\} + \frac{1}{2} dis(\mathcal{R}) \\ &\quad + \inf \left\{ d_{\mathcal{T}}(y, \tilde{z}) + d_{\mathcal{T}'}(x', \tilde{z'})\colon (\tilde{z}, \tilde{z'}) \in \mathcal{R} \right\} + \frac{1}{2} dis(\mathcal{R}). \end{aligned}$$

Hence, we need to show (for any $z, \tilde{z} \in \mathcal{T}$; $z', \tilde{z'} \in \mathcal{T}'$, such that $(z, z'), (\tilde{z}, \tilde{z'}) \in \mathcal{R}$)

$$d_{\mathcal{T}}(x, z) + d_{\mathcal{T}'}(x', z') + d_{\mathcal{T}}(y, \tilde{z}) + d_{\mathcal{T}'}(x', \tilde{z'}) + dis(\mathcal{R}) \geq d_{\mathcal{T}}(x, y).$$

Note further

$$\begin{aligned}
&d_{\mathcal{T}'}(x', z') + d_{\mathcal{T}'}(x', \tilde{z'}) + dis(\mathcal{R}) \geq d_{\mathcal{T}'}(z', \tilde{z'}) + dis(\mathcal{R}) \\
&\geq d_{\mathcal{T}'}(z', \tilde{z'}) + |d_{\mathcal{T}}(z, \tilde{z}) - d_{\mathcal{T}'}(z', \tilde{z'})| \geq d_{\mathcal{T}}(z, \tilde{z}).
\end{aligned}$$

Hence, finally we obtain

$$\begin{aligned}
&d_{\mathcal{T}}(x, z) + d_{\mathcal{T}'}(x', z') + d_{\mathcal{T}}(y, \tilde{z}) + d_{\mathcal{T}'}(x', \tilde{z'}) + dis(\mathcal{R}) \\
&\geq d_{\mathcal{T}}(x, z) + d_{\mathcal{T}}(z, \tilde{z}) + d_{\mathcal{T}}(y, \tilde{z}) \geq d_{\mathcal{T}}(x, y).
\end{aligned}$$

2. Let $x, y \in \mathcal{T}$, $x' \in \mathcal{T}'$. We want to show

$$d(x, y) + d(y, x') \geq d(x, x').$$

$$\begin{aligned}
&L = d_{\mathcal{T}}(x, y) + \inf\left\{d_{\mathcal{T}}(y, z) + d_{\mathcal{T}'}(x', z') \colon (z, z') \in \mathcal{R}\right\} + \frac{1}{2}dis(\mathcal{R}); \\
&P = \inf\left\{d_{\mathcal{T}}(x, \tilde{z}) + d_{\mathcal{T}'}(x', \tilde{z'}) \colon (\tilde{z}, \tilde{z'}) \in \mathcal{R}\right\} + \frac{1}{2}dis(\mathcal{R}).
\end{aligned}$$

Let $z, \tilde{z} \in \mathcal{T}$, $z', \tilde{z'} \in \mathcal{T}'$, such that

$$\begin{aligned}
&d(x', y) = d_{\mathcal{T}}(y, z) + d_{\mathcal{T}'}(x', z') + \frac{1}{2}dis(\mathcal{R}); \\
&d(x, x') = d_{\mathcal{T}}(x, \tilde{z}) + d_{\mathcal{T}'}(x', \tilde{z'}) + \frac{1}{2}dis(\mathcal{R}).
\end{aligned}$$

We need to show that

$$d_{\mathcal{T}}(x, y) + d_{\mathcal{T}}(y, z) + d_{\mathcal{T}'}(x', z') \geq d_{\mathcal{T}}(x, \tilde{z}) + d_{\mathcal{T}'}(x', \tilde{z'}).$$

Note that we know

$$\begin{aligned}
&d_{\mathcal{T}}(y, z) + d_{\mathcal{T}'}(x', z') \leq d_{\mathcal{T}}(y, \tilde{z}) + d_{\mathcal{T}'}(x', \tilde{z'}); \\
&d_{\mathcal{T}}(x, \tilde{z}) + d_{\mathcal{T}'}(x', \tilde{z'}) \leq d_{\mathcal{T}}(x, z) + d_{\mathcal{T}'}(x', z').
\end{aligned}$$

Hence, it is enough to show

$$d_{\mathcal{T}}(x, y) + d_{\mathcal{T}}(y, z) + d_{\mathcal{T}'}(x', z') \geq d_{\mathcal{T}}(x, z) + d_{\mathcal{T}'}(x', z').$$

But this is the triangle inequality for $d_{\mathcal{T}}$.

Hence, it follows that d satisfies the triangle inequality.

So we have shown that d is a metric.

Furthermore, computing the Hausdorff distance with the metric d we get

$$d_{Haus}(\mathcal{T}, \mathcal{T}') \leq \frac{1}{2} dis(\mathcal{R})$$

(since by the definition of a correspondence $\forall x \in \mathcal{T}\ \exists x' \in \mathcal{T}'$, such that $(x, x') \in \mathcal{R}$ and conversely $\forall y' \in \mathcal{T}'\ \exists y \in \mathcal{T}$, such that $(y, y') \in \mathcal{R}$). Moreover

$$d(\rho, \rho') = \frac{1}{2} dis(\mathcal{R}),$$

because $(\rho, \rho') \in \mathcal{R}$. Hence, we have

$$d_{Haus}(\mathcal{T}, \mathcal{T}') \vee d(\rho, \rho') \leq \frac{1}{2} dis(\mathcal{R}).$$

So we obtain

$$d_{GH}(\mathcal{T}, \mathcal{T}') \leq \frac{1}{2} \inf dis(\mathcal{R}),$$

which finally proves our statement. $\square$

2.2 Coding of real trees

We will describe a method of coding real random trees well-suited for proving further convergence results.
Consider a (deterministic) continuous function $g\colon [0, \infty) \to [0, \infty)$ with compact support, such that $g(0) = 0$ (we exclude the case $g \equiv 0$).
For every $s, t \geq 0$ set

$$m_g(s, t) = \inf_{r \in [s \wedge t, s \vee t]} g(r)$$

and

$$d_g(s, t) = g(s) + g(t) - 2m_g(s, t).$$

Furthermore, introduce an equivalence relation

$$s \sim t \text{ iff } d_g(s, t) = 0.$$

Hence, we can define a quotient space $\mathcal{T}_g$:

$$\mathcal{T}_g \stackrel{\text{def}}{=} [0, \infty)/\sim.$$

Furthermore, we introduce a canonical projection $p_g\colon [0, \infty) \to \mathcal{T}_g$.

Claim 2.4. *$(\mathcal{T}_g, d_g)$ is a metric space.*

Proof. We need to check that d_g is a metric. Let $\sigma = p_g(s),\ \sigma' = p_g(t),\ \sigma'' = p_g(u) \in \mathcal{T}_g$:

- $d_g(\sigma, \sigma') = 0$ iff $\sigma = \sigma'$ (by definition of the equivalence relation);
- $d_g(\sigma, \sigma') = d_g(\sigma', \sigma)$ (since $m_g(s,t) = m_g(t,s)$);
- $d_g(\sigma, \sigma') \leq d_g(\sigma, \sigma'') + d_g(\sigma'', \sigma')$.
  This is because

  $$d_g(\sigma, \sigma'') + d_g(\sigma'', \sigma') = g(s) + g(u) - 2m_g(s,u) + g(u) + g(t) - 2m_g(u,t)$$

  and

  $$d_g(\sigma, \sigma') = g(s) + g(t) - 2m_g(s,t),$$

  hence, we need to show

  $$g(u) - m_g(s,u) + m_g(s,t) - m_g(u,t) \geq 0.$$

  We can see that $m_g(s,t) = m_g(u,t)$ or $m_g(s,t) = m_g(s,u)$ (w.l.o.g. $m_g(s,t) = m_g(u,t)$). Then we need $g(u) - m_g(s,u) \geq 0$, but it is clear by the definition of m_g.

□

Claim 2.5. *p_g is continuous (when $[0,\infty)$ is equipped with the Euclidean metric and $\mathcal{T}_g$ with d_g).*

Proof. Note that

$$\begin{aligned} d_g(p_g(s), p_g(t)) &= g(s) + g(t) - 2m_g(s,t) = d(g(s), m_g(s,t)) + d(g(t), m_g(s,t)) \\ &\leq d(g(s), g(t)) \text{ (where d is a Euclidean metric).} \end{aligned}$$

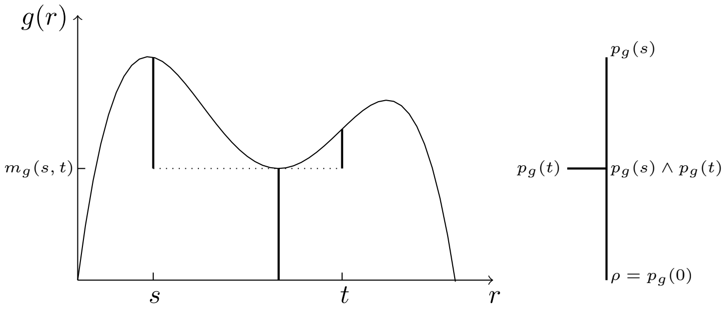
Hence, the claim follows by the continuity of g. □

Remark 2.6. *We set $\rho = p_g(0)$. If $\zeta > 0$ is the supremum of the support of g, then we have $p_g(t) = \rho$ for every $t \geq \zeta$. In particular $\mathcal{T}_g$ is compact, since $\mathcal{T}_g = p_g([0,\zeta])$.*

Now we will prove a theorem which gives a reason why we have introduced the quotient space $\mathcal{T}_g$.

Theorem 2.7. *The metric space $(\mathcal{T}_g, d_g)$ is a real tree. (We will view $(\mathcal{T}_g, d_g)$ as a rooted tree with a root $\rho = p_g(0)$.)*

To get a better understanding of the construction of $\mathcal{T}_g$ please have a look at the figure.



The figure shows the construction of a subtree of $\mathcal{T}_g$ consisting of the union of the ancestral lines of the vertices $p_g(s)$ and $p_g(t)$ corresponding to the times $s, t \in [0, \zeta]$.

Before proving Theorem 2.7 let us have a look on a lemma, which allows us to re-root considered real tree.

Lemma 2.8. *Let $s_0 \in [0, \zeta)$. For any real $r \geq 0$, denote by $\overline{r}$ the unique element of $[0, \zeta]$, such that $r - \overline{r}$ is an integer multiple of ζ. Set*

$$g'(s) = g(s_0) + g(\overline{s_0 + s}) - 2m_g(s_0, \overline{s_0 + s}) \text{ for all } s \in [0, \zeta]$$

and $g'(s) = 0$ for $s > \zeta$.
Then the function g' is continuous with compact support and satisfies $g'(0) = 0$, so that we can define $\mathcal{T}_{g'}$. Furthermore, for all $s, t \in [0, \zeta]$ we have

$$d_{g'}(s, t) = d_g(\overline{s_0 + s}, \overline{s_0 + t}) \tag{7}$$

and there exists a unique isometry R from $\mathcal{T}_{g'}$ onto $\mathcal{T}_g$, such that for all $s \in [0, \zeta]$

$$R(p_{g'}(s)) = p_g(\overline{s_0 + s}). \tag{8}$$

Remark 2.9. *Assuming Theorem 2.7 we see that $\mathcal{T}_{g'}$ coincides with the real tree $\mathcal{T}_g$ re-rooted at $p_g(s_0)$.*

Let us prove Lemma 2.8.

Proof. It is easy to see from the definition of g' that it is continuous, compactly supported and that $g' = 0$. Let us check now the relation (7). There are three cases that we need to consider.

1. First consider the case when $s, t \in [0, \zeta - s_0)$. The following two possibilities may occur
   - $m_g(s_0 + s, s_0 + t) \geq m_g(s_0, s_0 + s)$
     Note that then we have

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

for all $r \in [s,t]$. Hence

$$
\begin{aligned}
m_{g'}(s,t) &= \inf_{r\in[s\wedge t, s\vee t]} g'(r) \\
&= \inf_{r\in[s\wedge t, s\vee t]} (g(s_0) + g(s_0+r) - 2m_g(s_0, s_0+r)) \\
&= g(s_0) + m_g(s_0+s, s_0+t) - 2m_g(s_0, s_0+s).
\end{aligned}
$$

Furthermore, we can compute

$$
\begin{aligned}
& d_{g'}(s,t) = g'(s) + g'(t) - 2m_{g'}(s,t) \\
& = g(s_0) + g(s_0+s) - 2m_g(s_0, s_0+s) \\
& + g(s_0) + g(s_0+t) - 2m_g(s_0, s_0+t) \\
& - 2g(s_0) - 2m_g(s_0+s, s_0+t) + 4m_g(s_0, s_0+s) \\
& = g(s_0+s) + g(s_0+t) - 2m_g(s_0+s, s_0+t) \\
& = d_g(s_0+s, s_0+t).
\end{aligned}
$$

- $m_g(s_0+s, s_0+t) < m_g(s_0, s_0+s)$ Note first that

  $$m_{g'}(s,t) = g'(r_1)$$

  where r_1 is the first $r \in [s,t]$, such that $g(s_0+r) = m_g(s_0, s_0+s)$, because for $r \in [r_1, t]$ we have $g(s_0+r) - 2m_g(s_0, s_0+r) \geq -m_g(s_0, s_0+r) \geq -m_g(s_0, s_0+r_1)$. Therefore

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

  And further

  $$
  \begin{aligned}
  & d_{g'}(s,t) = g'(s) + g'(t) - 2m_{g'}(s,t) \\
  & = g(s_0) + g(s_0+s) - 2m_g(s_0, s_0+s) \\
  & + g(s_0) + g(s_0+t) - 2m_g(s_0, s_0+t) \\
  & - 2g(s_0) + 2m_g(s_0, s_0+s) \\
  & = g(s_0+s) + g(s_0+t) - 2m_g(s_0+s, s_0+t) \\
  & = d_g(s_0+s, s_0+t).
  \end{aligned}
  $$

2. The second case to consider is when $s,t \in [\zeta - s_0, \zeta)$. Note that the situation is symmetric to the first case (changing places of s and t). We have two further cases.
   - $m_g(\overline{s_0+s}, \overline{s_0+t}) \geq m_g(s_0, \overline{s_0+t})$
     We can note that

     $$m_g(s_0, \overline{s_0+r}) = m_g(s_0, \overline{s_0+t}) = m_g(\overline{s_0+s}, s_0)$$

     for all $r \in [s,t]$. Hence, we obtain

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

And it follows

$$d_{g'}(s,t) = d_g(\overline{s_0+s}, \overline{s_0+t})$$

by the same computation as above.

- $m_g(\overline{s_0+s}, \overline{s_0+t}) < m_g(\overline{s_0+t}, s_0)$
  We can see that

  $$m_{g'}(s,t) = g'(r_1)$$

  where r_1 is the last $r \in [s,t]$, such that $g(\overline{s_0+r}) = m_g(s_0, \overline{s_0+t})$, because for $r \in [s, r_1]$ we have $g(\overline{s_0+r}) - 2m_g(s_0, \overline{s_0+r}) \geq -m_g(s_0, \overline{s_0+r}) \geq -m_g(s_0, \overline{s_0+r_1})$. Therefore

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

  Hence, we obtain in a similar way as above

  $$d_{g'}(s,t) = d_g(\overline{s_0+s}, \overline{s_0+t}).$$

  (because $m_g(\overline{s_0+s}, \overline{s_0+t}) = m_g(s_0, \overline{s_0+s})$).

3. Let $s \in [0, \zeta - s_0)$ and $t \in [\zeta - s_0, \zeta)$. There are two cases that we need to take into account.

- $m_g(\overline{s_0+t}, s_0) \leq m_g(s_0, s_0+s)$
  Then

  $$m_{g'}(s,t) = g'(r_1)$$

  where r_1 is the first $r \in [s,t]$, such that $g(\overline{s_0+r}) = m_g(s_0, s_0+s)$, since for $r \in [r_1, t]$ we have $g(\overline{s_0+r}) - 2m_g(s_0, \overline{s_0+r}) \geq -m_g(s_0, \overline{s_0+r}) \geq -m_g(s_0, \overline{s_0+r})$. Hence

  $$\begin{aligned} m_{g'}(s,t) &= \inf_{r \in [s \wedge t, s \vee t]} g'(r) \\ &= \inf_{r \in [s \wedge t, s \vee t} (g(s_0) + g(\overline{s_0+r}) - 2m_g(s_0, \overline{s_0+r})) \\ &= g(s_0) - m_g(s_0, s_0+s). \end{aligned}$$

  And further we obtain

  $$\begin{aligned} d_{g'}(s,t) &= g'(s) + g'(t) - 2m_{g'}(s,t) \\ &= g(s_0) + g(s_0+s) - 2m_g(s_0, s_0+s) \\ &+ g(s_0) + g(\overline{s_0+t}) - 2m_g(s_0, \overline{s_0+t}) \\ &- 2g(s_0) + 2m_g(s_0, s_0+s) \\ &= g(s_0+s) + g(\overline{s_0+t}) - 2m_g(s_0+s, \overline{s_0+t}) \\ &= d_g(s_0+s, \overline{s_0+t}) \end{aligned}$$

  (because $m_g(s_0, \overline{s_0+t}) = m_g(s_0+s, \overline{s_0+t})$).

- $m_g(\overline{s_0+t}, s_0) > m_g(s_0, s_0+s)$
  This case is symmetrical to the one that we have just considered (switching s and t). We have

  $$m_{g'}(s,t) = g'(r_1)$$

  where r_1 is the last $r \in [s,t]$, such that $g(\overline{r+s_0}) = m_g(\overline{s_0+t}, s_0)$, since for $r \in [s, r_1]$ we have $g(\overline{s_0+r}) - 2m_g(s_0, \overline{s_0+r}) \geq -m_g(s_0, \overline{s_0+r}) \geq -m_g(s_0, \overline{s_0+r_1})$. Hence, we get

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

  And finally we obtain

  $$d_{g'}(s,t) = d_g(s_0+s, \overline{s_0+t})$$

  by the same computation as above, since $m_g(s_0, s_0+s) = m_g(s_0+s, \overline{s_0+t})$.

Hence, we have shown that (7) holds.
Now we can conclude that (8) is defined uniquely. Assume namely that $p_{g'}(s) = p_{g'}(t)$. We want to show that $R(p_{g'}(s)) = R(p_{g'}(t))$. But we have

$$\begin{aligned}
&p_{g'}(s) = p_{g'}(t) \Rightarrow d_{g'}(p_{g'}(s), p_{g'}(t)) = 0 \\
&\Rightarrow d_g(p_g(\overline{s_0+s}), p_g(\overline{s_0+t})) = 0 \\
&\Rightarrow p_g(\overline{s_0+s}) = p_g(\overline{s_0+t}) \\
&\Rightarrow R(p_{g'}(s)) = R(p_{g'}(t)).
\end{aligned}$$

Furthermore, R is an isometry, because

$$d_{g'}(p_{g'}(s), p_{g'}(t)) = d_g(p_g(\overline{s_0+s}), p_g(\overline{s_0+t})) = d_g(R(p_{g'}(s)), R(p_{g'}(t))).$$

The fact that R is surjective is also immediate, since let $\sigma = p_g(s) \in \mathcal{T}_g$, then $\sigma = R(p_{g'}(\overline{s-s_0}))$.

□

Now we will prove Theorem 2.7.

Proof. Preliminaries:
For σ, $\sigma' \in \mathcal{T}_g$ we have

$$\sigma \prec \sigma' \text{ iff } d_g(\sigma, \sigma') = d_g(\rho, \sigma') - d_g(\rho, \sigma).$$

Moreover, if $\sigma = p_g(s)$, $\sigma' = p_g(t)$, then

$$\sigma \prec \sigma' \text{ iff } m_g(s,t) = g(s).$$

This follows directly from the condition above, since $d_g(\sigma,\sigma') = g(s) + g(t) - 2m_g(s,t)$ and $d_g(\rho,\sigma') - d_g(\rho,\sigma) = g(t) - g(s)$. Note that "$\prec$" defines a partial order on $\mathcal{T}_g$.
For any $\sigma_0, \sigma \in \mathcal{T}_g$ set

$$[[\sigma_0,\sigma]] = \left\{\sigma' \in \mathcal{T}_g : \; d_g(\sigma_0,\sigma) = d_g(\sigma_0,\sigma') + d_g(\sigma',\sigma)\right\}.$$

Furthermore, if $\sigma = p_g(s)$ and $\sigma' = p_g(t)$, then it is easy to verify that $[[\rho,\sigma]] \cap [[\rho,\sigma']] = [[\rho,\gamma]]$, where $\gamma = p_g(r)$, if r is the time when g achieves its minimum between s and t. To check this note first

$$\begin{aligned}[[\rho,\sigma]] \cap [[\rho,\sigma']] &= \left\{\tilde\sigma \in \mathcal{T}_g : \; \tilde\sigma \prec \sigma \text{ and } \tilde\sigma \prec \sigma'\right\}; \\ [[\rho,\gamma]] &= \{\tilde\sigma \in \mathcal{T}_g \; \tilde\sigma \prec \gamma\}.\end{aligned}$$

- $[[\rho,\gamma]] \subset [[\rho,\sigma]] \cap [[\rho,\sigma']]$
  because $\gamma = p_g(r)$, such that $g(r) = m_g(s,t)$, hence, $g(r) = m_g(s,r)$ and $g(r) = m_g(r,t)$ and from that we have $\gamma \prec \sigma$ and $\gamma \prec \sigma'$ and by transitivity $[[\rho,\gamma]] \subset [[\rho,\sigma]] \cap [[\rho,\sigma']]$.
- $[[\rho,\sigma]] \cap [[\rho,\sigma']] \subset [[\rho,\gamma]]$
  assume that there exists $\gamma' = p_g(r')$, such that $\gamma' \in ([[\rho,\sigma]] \cap [[\rho,\sigma']]) \setminus [[\rho,\gamma]]$. Then $\gamma' \prec \sigma$ and $\gamma' \prec \sigma'$. Hence, $g(r') = m_g(r',s)$ and $g(r') = m_g(s,t) \Rightarrow g(r') = m_g(s,t)$. Now, since $g(r) = m_g(s,t)$ we have $p_g(r) = p_g(r')$. Contradiction to $\gamma' \notin [[\rho,\gamma]]$.

Put $\gamma = \sigma \wedge \sigma'$.
Furthermore, set $\mathcal{T}_g[\sigma] \stackrel{\text{def}}{=} \{\sigma' \in \mathcal{T}_g : \; \sigma \prec \sigma'\}$. Note that, if $\mathcal{T}_g[\sigma] \neq \{\sigma\}$ and $\sigma \neq \rho$, then $\mathcal{T}_g[\sigma] \setminus \{\sigma\}$ and $\mathcal{T}_g \setminus \mathcal{T}_g[\sigma]$ are two non-empty, disjoint open sets.

- $\mathcal{T}_g \setminus \mathcal{T}_g[\sigma]$ is open, since $\mathcal{T}_g[\sigma] = p_g(\underbrace{\{u \in [0,\zeta] : \; m_g(s,u) = g(s)\}}_{\text{compact}})$. Hence, $\mathcal{T}_g[\sigma]$ is compact, so also closed.
- $\mathcal{T}_g[\sigma] \setminus \{\sigma\}$ is open. In order to see that consider $\sigma' \in \mathcal{T}_g[\sigma] \setminus \{\sigma\}$. Then $B_{d_g(\sigma,\sigma')}(\sigma') \subset \mathcal{T}_g[\sigma] \setminus \{\sigma\}$. Hence, $\mathcal{T}_g \setminus \{\sigma\}$ is open.

Let us check the property 1. from the definition of a real tree. Take $\sigma_1, \; \sigma_2 \in \mathcal{T}_g$. We need to show the existence and the uniqueness of f_{σ_1,σ_2}.

- existence
  Without loss of generality (by Lemma 2.8) set $\sigma_1 = \rho$ and further $\sigma = \sigma_2$. So we need to show that there exists an isometry $f_{\rho,\sigma} \colon [0, d_g(\rho,\sigma)] \to \mathcal{T}_g$, such that $f_{\rho,\sigma}(0) = \rho$ and $f_{\rho,\sigma}(d_g(\rho,\sigma)) = \sigma$.
  Set $s \in p_g^{-1}(\{\sigma\})$, such that $g(s) = d_g(\rho,\sigma)$. Then for all $a \in [0, d_g(\rho,\sigma)]$ we can set

  $$v(a) = \inf\{r \in [0,s] : \; m_g(r,s) = a\}.$$

Note that $g(v(a)) = a$. Setting $f(a) = p_g(v(a))$ we obtain $f(0) = \rho$ and $f(d_g(\rho, \sigma)) = \sigma$ (because $f(d_g(\rho, \sigma)) = p_g(v(d_g(\rho, \sigma))) = p_g(v(g(s))) = p_g(s) = \sigma$).
We still need to check that f is an isometry. Let $a,\ b \in [0, d_g(\rho)]$, such that $a \le b$. From $m_g(v(a), v(b)) = a$ (by the definition of v) we obtain

$$d_g(f(a), f(b)) = d_g(p_g(v(a)), p_g(v(b)))$$
$$= g(v(a)) + g(v(b)) - 2m_g(v(a), v(b)) = b - a,$$

so f is an isometry.

- uniqueness
Assume that $\tilde{f} \colon [0, d_g(\rho, \sigma)] \to \mathcal{T}_g$ is isometric, such that $\tilde{f}(0) = \rho$ and $\tilde{f}(d_g(\rho, \sigma)) = d_g(\rho, \sigma)$. Let $a \in [0, d_g(\rho, \sigma)]$, then

$$d_g(\sigma, \tilde{f}(a)) = d_g(\tilde{f}(d_g(\rho, \sigma)), \tilde{f}(a)) = d_g(\rho, \sigma) - a$$
$$= d_g(\rho, \sigma) - d_g(\rho, \tilde{f}(a)).$$

Hence, $\tilde{f}(a) \prec \sigma$.
Recall that $\sigma = p_g(s)$. Pick t, such that $p_g(t) = \tilde{f}(a)$. We can note that $g(t) = d_g(\rho, p_g(t)) = d_g(\tilde{f}(0), \tilde{f}(a)) = a$. Moreover, from $\tilde{f}(a) \prec \sigma$ we have $g(t) = m_g(s, t)$.
On the other hand

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

Hence

$$d_g(t, v(a)) = g(t) + g(v(a)) - 2m_g(v(a), t) = 0.$$

From that we finally get

$$\tilde{f}(a) = p_g(t) = p_g(v(a)) = f(a).$$

So the property 1. from the definition of a real tree is satisfied.
We still need to check the property 2. Let $q \colon [0, 1] \to \mathcal{T}_g$ be continuous and injective. We want to show that

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

By Lemma 2.8 we can again assume $q(0) = \rho$ and set $\sigma = q(1)$. So it is enough to show $q([0, 1]) = [[\rho, \sigma]]$ (since $f_{\rho,\sigma}([0, d_g(\rho, \sigma)]) = [[\rho, \sigma]]$).

- $[[\rho, \sigma]] \subset q([0, 1])$
Assume that $\gamma \in [[\rho, \sigma]] \backslash g([0, 1])$. It follows that $q([0, 1]) \subset (\mathcal{T}_g[\gamma] \backslash \{\gamma\}) \cup (\mathcal{T}_g \backslash \mathcal{T}_g[\gamma])$. Moreover, $\rho \in \mathcal{T}_g \backslash \mathcal{T}_g[\gamma]$ and $\sigma \in \mathcal{T}_g[\gamma] \backslash \{\gamma\}$ (because $\gamma \prec \sigma$). This is a contradiction to the fact that $q([0, 1])$ is connected.

- $[[\rho, \sigma]] \supset q([0,1])$

  Assume that there is $a \in (0,1)$, such that $q(a) \notin [[\rho, \sigma]]$. Set $\eta = q(a)$ and $\gamma = \sigma \wedge \eta$. Note that $\gamma \in [[\rho, \eta]] \cap [[\eta, \sigma]]$. Moreover, $d_g(\eta, \sigma) = d_g(\eta, \gamma) + d_g(\gamma, \sigma)$. By the similar argument as in the part 1. we have

  $$\gamma \in [[\rho, \eta]] \subset q([0, a])$$

  and analogously setting η to be a root, via root-change argument, we obtain

  $$\gamma \in [[\eta, \sigma]] \subset q([a, 1]).$$

  Since q is injective we finally get

  $$\gamma = q(a) = \eta.$$

  This is a contradiction to $\eta \notin [[\rho, \sigma]]$. Hence, it follows that the property 2. must necessarily be satisfied.

$\square$

Now we will show the result that will allow us to conclude the convergence of real trees from the convergences of coding functions.

Lemma 2.10. *Let g, $g' : [0, \infty) \to [0, \infty)$ be two continuous functions with compact support, such that $g(0) = 0 = g'(0)$. Then*

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

(where $\|\cdot\|$ is the uniform norm).

Proof. Construct a correspondence between $\mathcal{T}_g$ and $\mathcal{T}_{g'}$ by

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

Let $(\sigma, \sigma'),\ (\eta, \eta') \in \mathcal{R}$. By the definition of $\mathcal{R}$ there are $s, t \geq 0$, such that $p_g(s) = \sigma,\ p_g(t) = \eta,\ p_{g'}(s) = \sigma'$ and $p_{g'}(t) = \eta'$.
Furthermore, we have

$$\begin{aligned} d_g(\sigma, \eta) &= g(s) + g(t) - 2m_g(s,t); \\ d_{g'}(\sigma', \eta') &= g'(s) + g'(t) - 2m_{g'}(s,t). \end{aligned}$$

Hence

$$\begin{aligned} &|d_g(\sigma, \eta) - d_{g'}(\sigma', \eta')| \\ &= |g(s) + g(t) - 2m_g(s,t) - g'(s) - g'(t) + 2m_{g'}(s,t)| \leq 4\|g - g'\|. \end{aligned}$$

In particular it follows that

$$dis(\mathcal{R}) \leq 4\|g - g'\|.$$

Now by Claim 2.3 we obtain

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

$\square$

2.3 The continuum random trees

Let $e = (e_t,\ 0 \leq t \leq 1)$ be the normalized Brownian excursion.

Definition 12. *The continuum random tree (CRT) is the real random tree $\mathcal{T}_e$ coded by the normalized Brownian excursion.*

Remark 2.11. *The CRT $\mathcal{T}_e$ is a random variable taking values in the set $\mathbb{T}$ (by Lemma* 2.10 *CRT is continuous, thus measurable).*

Note that

- $\mathbb{A}$ is the set of all (finite) rooted, ordered trees. By $\mathbb{A}_n$ denote the subset of $\mathbb{A}$ consisting of trees with n vertices.
- We may view $t \in \mathbb{A}$ as a rooted real tree. If $(C_t,\ t \geq 0)$ is the contour function of the tree, then we identify $t = \mathcal{T}_C$ (note that the lexicographical structure of t disappears, when we consider it as a real tree).
- For $\lambda > 0$ and $\mathcal{T} \in \mathbb{T}$ the tree $\lambda\mathcal{T}$ is just "the same" tree rescaled by λ.

We come to our final result. We will show that the continuum random tree is the limit of rescaled discrete random trees.

Theorem 2.12. *For every $n \geq 1$, let $\mathcal{T}_{(n)}$ be a random tree distributed uniformly over $\mathbb{A}_n$. Then $\frac{1}{\sqrt{p}}\mathcal{T}_{(n)}$ converges in distribution to the CRT $\mathcal{T}_e$ in the space $\mathbb{T}$.*

Proof. We consider the case where θ is a Galton-Watson tree with the geometric offspring distribution $\mu(k) = 2^{-k-1}$ (Note that for $X \sim \mu$: $\mathbb{E}[X] = 1$, $var(X) = 2$). Moreover, for $n \geq 1$ let θ_n be distributed as θ, conditioned to have n vertices. Then θ_n has the same distribution as $\mathcal{T}_{(n)}$.
On the other hand, for $(C^n_t,\ t \geq 0)$ the contour function of θ_n define

$$\tilde{C}^n_t = \frac{1}{\sqrt{2n}} C^n_{2nt},\ t \geq 0.$$

By Theorem 1.28 $((\frac{1}{\sqrt{p}}C^{(p)}_{2pt},\ t \geq 0) \xrightarrow[p\to\infty]{(d)} (\frac{2}{\sigma}e_t,\ t \geq 0))$ we have

$$(\tilde{C}^n_t,\ t \geq 0) \xrightarrow[p\to\infty]{(d)} (e_t,\ t \geq 0).$$

(Note that $\sigma = \sqrt{2}$.) Since θ_n has the same distribution as $\mathcal{T}_{(n)}$, hence, the tree $\mathcal{T}_{\tilde{C}^n}$ coded by $\tilde{C}^n$ has the same distribution as $\frac{1}{\sqrt{2n}}\mathcal{T}_{(n)}$. Hence, the statement follows from Lemma 2.10. □

References

[1] J.F. Le Gall. Random Trees and Applications. (2005), 1-33.

[2] P. Nolin. Probability Theory. (2012)

[3] P. Nolin. Markov chains: mixing times and applications. (2012), 15-19.

[4] S.N. Evans, J. Pitman, A. Winter. Rayleigh processes, real trees, and root growth with re-grafting. (2004), 12.

[5] F. Spitzer. *Principles of Random Walk.* Springer, 1976.

[6] I. Karatzas, S.E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer, 1988.

[7] M. Drmota. *Random Trees. An interplay between Combinatorics and Probability.* Springer, 2009.

[8] R. Durrett. *Probability. Theory and Examples.* Wadsworth, 1991.

[9] K.B. Athreya, S.N. Lahiri. *Measure Theory and Probability Theory.* Springer, 2006.