University of Zurich[UZH]

Institute of Mathematics

Master's Thesis

Conditioning and Local Convergence of Galton Watson Trees

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematics.

Author:
Sina Maria Schlienger

Supervisors:
Dr. Alejandro Rosales Ortiz
Prof. Dr. Jean Bertoin

April 30, 2026

Abstract

After an introduction to branching processes, the notion of trees and random trees, Galton Watson trees are introduced. The main object of this thesis is to study how Galton Watson trees behave under different types of conditioning. Depending on the mean number of children denoted by m_π, Galton Watson trees can be classified into three groups: sub-critical ($m_\pi < 1$), critical ($m_\pi = 1$) or super-critical ($m_\pi > 1$). To warm-up, we focus on super-critical Galton Watson trees. This is the only regime where trees both survive and get extinct with a positive probability. Hence, conditioning super-critical Galton Watson trees on survival or extinction can be easily studied and yields to a two-type respectively sub-critical Galton Watson tree. Since for critical trees the probability of survival is zero, a new approach is needed. We will condition the trees on "being large", whether by setting the height, the cardinality or the maximal degree to be n, and then study the resulting limit as $n \to \infty$. To formalize this, we introduce local convergence of random trees. We will see that under suitable assumptions, critical Galton Watson trees conditioned on "being large" converge locally to Kesten's tree, which is a tree with an infinite spine and independent Galton Watson trees attached to it.

Acknowledgments

My sincere thanks to Dr. Alejandro Rosales Ortiz for his invaluable supervision and to Prof. Dr. Jean Bertoin for giving me the opportunity to complete my Master's thesis. I also want to thank my family and friends for supporting me throughout the whole journey.

Contents

1 Introduction **1**

2 Preliminaries: Branching processes **6**

3 The space of plane discrete trees **8**
3.1 Basic notions 8
3.2 Metric structure on $\mathbb{T}$ 12
3.3 Random trees 16

4 Galton Watson trees and Kesten's tree **17**
4.1 Galton Watson trees 17
4.2 Extinction and survival probabilities of Galton Watson trees 20
4.3 The branching property 26
4.4 Two-type Galton Watson trees 31
4.5 Kesten's tree 33

5 Conditioning of super-critical Galton Watson trees **36**
5.1 Conditioning on extinction 36
5.2 Conditioning on survival 38

6 Local limits of critical Galton Watson trees **42**
6.1 Preliminaries on weak convergence 42
6.2 Local convergence of random trees 45
6.3 Convergence towards Kesten's tree 52
6.4 Kesten's theorem and other applications 55
6.4.1 Kesten's theorem (height) 55
6.4.2 Cardinality 57
6.4.3 Maximal degree 60

7 Conclusion **62**

Chapter 1

Introduction

Did you ever wonder how long your surname will survive? F. Galton, a British scientist, asked himself what the probability for a noble surname of the UK to eventually die out would be. Together with the mathematician H. W. Watson, he published the article "On the probability of extinction of families", where the Galton Watson branching process was introduced and with it also Galton Watson trees. In this thesis, we will focus on Galton Watson trees and analyze how they behave under conditioning of different types.

In Chapter 2, we recall what a branching process is and connect it to Galton Watson trees. To then precisely define a Galton Watson tree, we first start in Chapter 3 with the notion of a tree. A tree is of the following structure: It has starts with a node called root, which then has children attached to it. Those children form a new generation and have children themselves, and so on. On the set of trees $\mathbb{T}$, we can define the truncation function, abbreviated by tr_n. It informally cuts a tree at height n. We introduce a metric structure on $\mathbb{T}$. If two trees are denoted $\mathtt{t}$ and $\mathtt{t}'$, the metric is given by $d_{\mathbb{T}}(\mathtt{t},\mathtt{t}') = 2^{-m(\mathtt{t},\mathtt{t}')}$, where $m(\mathtt{t},\mathtt{t}') = \sup\left\{n \geq 0 : tr_n(\mathtt{t}) = tr_n(\mathtt{t}')\right\}$. Then, $(\mathbb{T}, d_{\mathbb{T}})$ is a Polish metric space, what will be important later. We also endow the space with its Borel sigma-algenra. This is enough to define random trees, that are measurable functions defined on a probability space with values in the set of trees. A Galton Watson tree is a random tree, where each individual reproduces independently by a given offspring distribution π and it satisfies the branching property, for an example see Figure 1.1. Informally, this means that at each node, we again see something of the form of a Galton Watson tree.

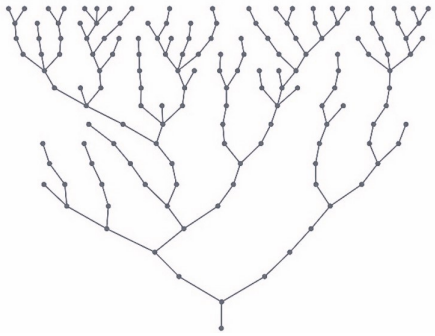

Figure 1.1: An example of a Galton Watson tree [Bar17].

A Galton Watson tree can survive, in other words grow until infinity, or can get extinct, which means that its height is finite. Using the probability generating function of π, denoted by φ_π, we derive in Chapter 4 the probabilities of survival and extinction of a Galton Watson tree. In this context, we also find an important trichotomy. Given that m_π is the expected number of children, a Galton Watson tree and its offspring distribution π are called sub-critical if $m_\pi < 1$, critical if $m_\pi = 1$ and super-critical if $m_\pi > 1$.

We will focus on super-critical and critical Galton Watson trees. The sub-critical case is similar to the critical case and results can be found in [AD20]. We are interested in how Galton Watson trees behave under conditioning on survival or extinction. To warm-up, we consider super-critical Galton Watson trees in Chapter 5. Out of the three cases, this is the only where survival and extinction both happen with positive probability. Contrarily, in the critical and sub-critical case, the Galton Watson trees die with probability one. Back to the super-critical case, we can condition super-critical Galton Watson on both, extinction and survival, without too much technicalities as all of the probabilities are positive. We will find that conditioning a super-critical Galton Watson tree with offspring distribution π on extinction gives a sub-critical Galton Watson tree with offspring distribution $\tilde{\pi}(k) = q_\pi^{k-1}\pi(k)$, where q_π is the smallest solution for $\varphi_\pi(r) = r$ and φ_π being again the probability generating function of π. On the other hand, conditioning on survival yields to a two-type Galton Watson tree. It is called two-type, because the nodes can be divided in two different types e and s and they reproduce differently. Nodes of kind e reproduce according to the offspring distribution $\tilde{\pi}$ and only generate children of the same type e. The other one, type s, has a different possibly unrelated offspring distribution and can have children of both e and s, but at least one child must be of type s. For a picture, look at Figure 1.2.

Analyzing how critical Galton Watson trees behave under conditioning on survival needs a bit more work. Since the probability of survival is zero, so we cannot just condition naively on survival. We proceed as follows: We condition the tree on "being

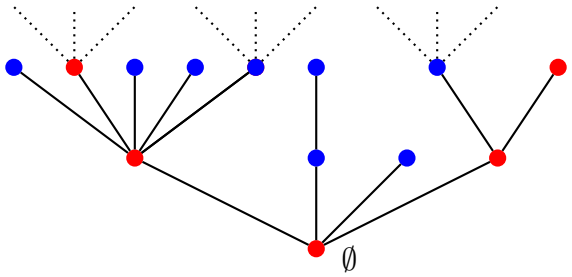

Figure 1.2: Two-type Galton Watson tree with two types of nodes (blue e and red s).

large" for some fixed n, for example by setting the height to be n or the cardinality to be n. After conditioning, we let n go to infinity and consider the limits in distribution of the conditioned trees. To make sense of this, we need to introduce local convergence of trees. Chapter 6 starts with a recall of preliminaries on weak convergence: A sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ on the a metric space S is said to converge weakly to μ if for all bounded and continuous functions

$$\int_S f d\mu_n \xrightarrow{n\to\infty} \int_S f d\mu.$$

If $(X_n)_{n\in\mathbb{N}}$ is a sequence of S-valued random variables with laws $(\mu_n)_{n\in\mathbb{N}}$, we say that it converges in distribution to a S-valued random variable X with law μ if and only if $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ. Thanks to Portmanteau's theorem, this is equivalent to

$$\mu_n(A) \xrightarrow{n\to\infty} \mu(A)$$

for all μ-continuous sets A, that is a set with a boundary of mass zero. As the space of trees is a metric space, we can apply all those results to sequences of trees. We will see that a sequence of random trees $(\mathbf{t}_n)_{n\in\mathbb{N}}$ with law $(\mu_n)_{n\in\mathbb{N}}$ converges in distribution respectively converges locally to another random tree $\mathbf{t}$ with law μ if

$$\mu_n(tr_k(\mathbf{t}) = tr_k(\mathtt{t})) \xrightarrow{n\to\infty} \mu(tr_k(\mathbf{t}) = tr_k(\mathtt{t})),$$

where tr_k denotes again the truncation function.

We will see that critical Galton Watson trees will converge to Kesten's tree $\mathcal{K}$. Informally, it is a random tree with an infinite spine. Onto this spine many independent Galton Watson trees are attached. See Figure 1.3 for an illustration.

To formalize "conditioning on being large", we introduce a functional A that takes values from the space of finite trees to positive natural numbers. Depending on the theorem and context, this functional fulfills either the Monotonicity property, the Additivity property or the Identity property, for more details see Chapter 6. Examples for

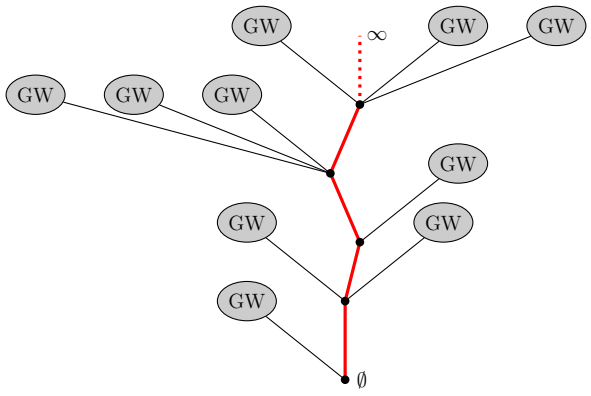

Figure 1.3: Example of a Kesten's tree with the infinite spine.

such A are the height of the tree, the cardinality or the maximal degree, which is the maximal number of children a node can have. We define the set

$$\mathbb{A}_n = \{A(\mathbf{T}) \in [n, n + n_1)\},$$

with $n_1 \in \mathbb{N}^*$. Using the notion of local convergence of trees and the definition of the functional we find the following important theorem:

Theorem. *Let* $\mathbf{T}$ *be a critical Galton Watson tree with offspring distribution* π *such that* $\pi(1) < 1$. *Let* $\mathbf{T}_n$ *be a random tree distributed according to* $\mathbf{T}$ *conditionally on* $\mathbb{A}_n$ *for a fixed* n_1 *and* $\mathbb{P}(\mathbb{A}_n) > 0$ *for* n *large enough. Denote by* $\mathcal{K}$ *Kesten's tree associated to* π. *Suppose one of the following conditions is fulfilled*

1. $n_1 = \infty$ *and* A *satisfies the Monotonicity property.*
2. $n_1 \in \mathbb{N}^* \cup \{\infty\}$ *and* A *satisfies the Identity property.*
3. $n_1 \in \mathbb{N}^* \cup \{\infty\}$, A *satisfies the Additivity property and*

$$\limsup_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} \leq 1.$$

Then $\mathbf{T}_n \xrightarrow[n\to\infty]{(d)} \mathcal{K}$, *so* $\mathbf{T}_n$ *converges locally to* $\mathcal{K}$.

With this theorem, we can see how critical Galton Watson trees behave under different types of conditioning: the height, the cardinality and the maximal degree. In each case, with suited assumptions, we find that critical Galton Watson trees conditioned on "being large" converge locally to Kesten's tree.

The main part of this thesis is based on the sources [Duq24] and [AD20], where the first source is more used for the theory about metric of trees and the construction of a Galton Watson tree. The rest strongly depends on [AD20]. For Section 4.1, [Bar17] is also considered and information including Kesten's tree is found in [C24]. The theory of weak convergence can be found in [Bil99] or [Gal22]. Section 6.4.2 follows [AD13]. All figures without a reference are made using tikz.

Chapter 2

Preliminaries: Branching processes

Branching processes are closely linked to Galton Watson trees, which are the central objects in this thesis. A Galton Watson tree $\mathbf{T}$ starts with a single node called root. The root then gives birth to a random number of children according to a fixed probability distribution π known as offspring distribution. These children form the next generation and each child then produces independently a random number of descendants following the same offspring distribution π. Repeating this process results in an entire random tree, for an example see Figure 2.1.

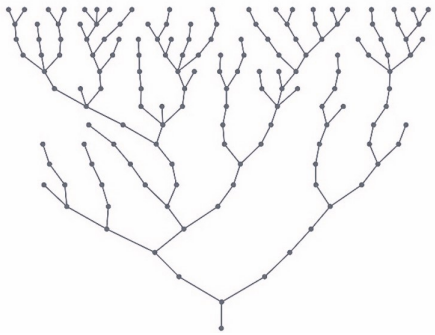

Figure 2.1: An example of the first generations of a Galton Watson tree [Bar17].

The Galton Watson branching process $(Z_n)_{n\geq 0}$ encodes exactly the population size of a Galton Watson tree. In particular, Z_n tracks the number of individuals at generation n of the tree and thus $(Z_n)_{n\geq 0}$ describes the population evolution of the Galton Watson tree. So both the process and the tree encode the population size at each generation, but the Galton Watson tree additionally includes information about the genealogical structure. This connection is the reason why we first recall the definition of a branching process with reproduction law π and remind that it is a Markov chain.

Definition 2.1. Let $\pi = (\pi(k))_{k\,\geq 0}$ be a probability law on $\mathbb{N}$, which is seen as the reproduction law. So for all $k \geq 0$ we have $\pi(k) \geq 0$ and $\sum_{k\geq 0} \pi(k) = 1$. Let $(X_{n,j})_{n\geq 0,\, j\geq 1}$ be a family of independent random variables with values in $\mathbb{N}$ and law π. Let $Z_0 \colon \Omega \to \mathbb{N}$ be a random variable with law ν independent of $(X_{n,j})_{n\geq 0,\, j\geq 1}$. We recursively define

$$Z_{n+1} = \begin{cases} \sum_{j=1}^{Z_n} X_{n,j} & \text{if } Z_n \geq 1 \\ 0 & \text{if } Z_n = 0, \end{cases}$$

where Z_n denotes the number of individuals of generation n. The process $(Z_n)_{n\geq 0}$ is called the *branching process* with reproduction law π and initial distribution ν.

Proposition 2.2. *A branching process with reproduction law π and initial distribution ν is a Markov chain with transition matrix*

$$Q(i,j) = \pi^{*i}(j)$$

for all $i, j \in \mathbb{Z}^+$. The convolution π^{*i} is defined by $\pi^{*1} = \pi$ and $\pi^{*(i+1)} = \pi^{*i} * \pi$ with the convention that $\pi^{*0} = \delta_0$, the Dirac measure.

Proof. To prove this statement we need to show that for any $z_0, \ldots, z_n$ such that $\mathbb{P}(Z_0 = z_0, \ldots, Z_n = z_n) > 0$, we have

$$\mathbb{P}\big(Z_{n+1} = y \mid Z_0 = z_0, \ldots, Z_n = z_n\big) = Q(z_n, y).$$

Consider the case where $z_n = 0$. Then, under $\mathbb{P}(\cdot \mid Z_n = z_n)$, we have $Z_{n+1} = 0$ a.s. by definition. Thus,

$$\mathbb{P}\big(Z_{n+1} = y \mid Z_0 = z_0, \ldots, Z_n = 0\big) = \delta_0(y) = \pi^{*0}(y) = Q(0, y).$$

Now suppose $z_n \geq 1$. Remark that for n fixed, the family $(X_{n,j})_{j\geq 1}$ is independent of $Z_0, \ldots, Z_n$. Then

$$\begin{aligned} \mathbb{P}\left(Z_{n+1} = y \mid Z_0 = z_0, \ldots, Z_n = z_n\right) &= \mathbb{P}\left(\sum_{j=1}^{Z_n} X_{n,j} = y \;\middle|\; Z_0 = z_0, \ldots, Z_n = z_n\right) \\ &= \mathbb{P}\left(\sum_{j=1}^{Z_n} X_{n,j} = y \;\middle|\; Z_n = z_n\right) \\ &= \mathbb{P}\left(\sum_{j=1}^{z_n} X_{n,j} = y\right) \\ &= \pi^{*\,z_n}(y) = Q(z_n, y). \end{aligned}$$

Thus, a branching process is a Markov chain with transition matrix $Q(i,j) = \pi^{*i}(j)$. □

Chapter 3

The space of plane discrete trees

3.1 Basic notions

In order to study Galton Watson trees we introduce a mathematical framework well ruled to define them. In this direction, we will rely on the notion of a plane tree and its properties. The set of finite sequences of positive integers denoted by

$$\mathbb{U} = \bigcup_{n\geq 0} (\mathbb{N}^*)^n,$$

with the convention that $(\mathbb{N}^*)^0 = \{\emptyset\}$, is called *Ulam's tree* and a sequence of $\mathbb{U}$ is denoted as $u = (u_1, \ldots, u_n) \in \mathbb{U}$. The *height* or also called generation of u is

$$H(u) = n$$

with the convention that $H(\emptyset) = 0$. For two sequences $u \in \mathbb{U}$ and $v \in \mathbb{U}$, the *concatenation* of them is denoted by uv with $uv = vu = v$ if $u = \emptyset$. The *parent* of $u = (u_1, \ldots, u_n)$ is $(u_1, \ldots, u_{n-1})$ and set the parent of $\emptyset$ to be again $\emptyset$. We can now define a plane tree, an example is drawn in Figure 3.1.

Definition 3.1. A *plane tree* $\mathtt{t}$ is a subset of $\mathbb{U}$ such that

1. $\emptyset \in \mathtt{t}$.
2. If $v \in \mathtt{t}$, $v \neq \emptyset$, then the parent of v also belongs to $\mathtt{t}$.
3. For all $u \in \mathtt{t}$, there exists $k_u(\mathtt{t}) \in \mathbb{N} \cup \{\infty\}$ such that for all $i \in \mathbb{N}^*$ we have $ui \in \mathtt{t}$ if and only if $1 \leq i \leq k_u(\mathtt{t})$.

The *set of trees* is denoted by $\mathbb{T}$.

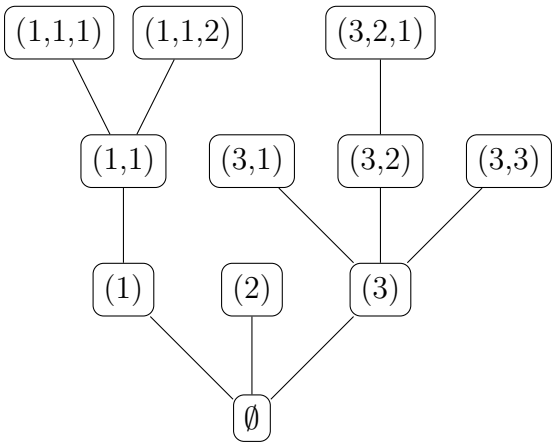

Figure 3.1: Example of a plane tree $\mathtt{t}$.

We say that $\emptyset$ is the *root* of the tree. The number $k_u(\mathtt{t})$ is the *number of offspring* of u. An offspring is often also called child. If a node $u \in \mathtt{t}$ has no children, so $k_u(\mathtt{t}) = 0$, it is called *leaf*. The set of leaves is denoted as $\mathscr{L}_0(\mathtt{t}) = \big\{u \in \mathtt{t} : k_u(\mathtt{t}) = 0\big\}$. In Figure 3.1, for instance, we have $\mathscr{L}_0(\mathtt{t}) = \big\{(1,1,1),(1,1,2),(3,1),(3,2,1),(3,3)\big\}$. If $k_u(\mathtt{t}) \geq 2$, then u is a *branching point*. Again looking at Figure 3.1, the nodes $(1,1)$ and (3) are such points. To denote the number of nodes of a trees $\mathtt{t}$, we write $\#\mathtt{t}$. More generally, $\#$ is used for the cardinality of a set. We can observe that

$$\sum_{u \in \mathtt{t}} k_u(\mathtt{t}) = \#\mathtt{t} - 1. \tag{3.1}$$

Indeed, the sum counts all children in the tree. The root itself is not a child and every other node is a child of exactly one parent. Therefore, the sum gives the number of all nodes of the tree minus the root, that is $\#\mathtt{t} - 1$. The *height of a tree* is

$$\mathcal{H}(\mathtt{t}) = \sup\big\{H(u) : u \in \mathtt{t}\big\} \tag{3.2}$$

and the *width* at height n is given by

$$Z_n(\mathtt{t}) = \#\big\{u \in \mathtt{t} : H(u) = n\big\}.$$

Notice that the width directly describes the number of individuals at the height n of the tree. For example in Figure 3.1, the height of node $(3,1)$ is 2 and the width at this height is 4, since there are four elements with height 2. The height of the example-tree is 3. Further, $\mathbb{T}^f$ is the set of finite trees, namely

$$\mathbb{T}^f = \{\mathtt{t} \in \mathbb{T} : \#\mathtt{t} < \infty\}.$$

And denote the set of finite trees with height less than n as

$$\mathbb{T}_n = \left\{ \mathtt{t} \in \mathbb{T} : \mathcal{H}(\mathtt{t}) \leq n \right\}.$$

Remark that the set of finite trees $\mathbb{T}^f$ is countable as each tree can be represented by a finite sequence of positive integers and there are countably many such sequences. In contrast, $\mathbb{T}$ is uncountable. An *infinite spine* is a subset of $\mathbb{U}$ of the form $\bar{v} = \{\bar{v}_n : n \in \mathbb{N}\}$, where $\bar{v}_n = (u_1, \ldots, u_n)$ with $\bar{v}_0 = \emptyset$. The set of all trees with exactly one such infinite spine is

$$\mathbb{T}^1 = \{\mathtt{t} \in \mathbb{T} : \mathtt{t} \text{ has exactly one infinite spine}\}$$

We introduce a partial order on $\mathbb{U}$ called *genealogical order*: We write $v \preceq u$ if there exists $w \in \mathbb{U}$ such that $u = vw$. If $v \preceq u$ and $v \neq u$, then v is an *ancestor* of u, denoted by $v \prec u$. The set of all ancestors of $u \in \mathbb{U}$ is

$$A_u = \{v \in \mathbb{U} : v \prec u\}.$$

Also denote $\bar{A}_u = A_u \cup \{u\}$. Notice that $\emptyset \notin A_u$ unless $u = \emptyset$. The *most recent common ancestor* of a subset U of $\mathbb{U}$, denoted $MRCA(U)$, is the unique element v of $\bigcap_{u \in U} \bar{A}_u$ with maximal height. If U is the set of just two nodes u and v, $MRCA(U)$ is symbolized as $u \wedge v$. In Figure 3.2, the ancestors of $(3, 2, 1)$ are marked gray and the most recent common ancestor $(1, 1, 1) \wedge (1, 2)$ is (1). Another order on the set $\mathbb{U}$ is the *lexographic order*. We write $v < u$ if either $v \prec u$ or $v = wjv'$ and $u = wiu'$ with $u = v \wedge v$. Alternatively, denote $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_m)$. Then $v < u$ if there exist $j \in \mathbb{N}$ such that $v_j = u_j$ but $v_{j+1} < u_{j+1}$.

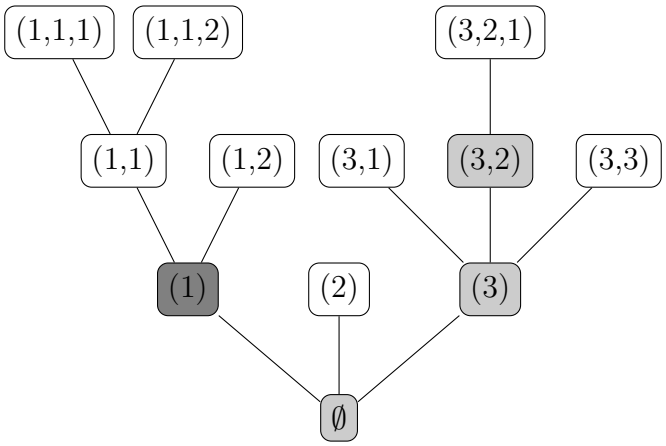

Figure 3.2: Ancestors of $(3, 2, 1)$ are marked gray and the most recent common ancestor of two nodes $(1, 1, 1) \wedge (1, 2)$ is dark gray.

The *subtree* of $\mathtt{t}$ above the node u is defined as

$$\mathcal{S}_u(\mathtt{t}) = \{v \in \mathbb{U} : uv \in \mathtt{t}\} = \left\{v \in \mathbb{U} : u \in \bar{A}_{uv}\right\}.$$

All the elements of the subtree $\mathcal{S}_u(\mathtt{t})$ are the *descendants* of u in $\mathtt{t}$, so those are all the nodes $v \in \mathbb{U}$ such that $uv \in \mathtt{t}$. They are marked gray in the example in Figure 3.3. Remark that the entries of the ancestors have to be shifted to form an actual subtree, see Figure 3.4. Also observe that $\mathcal{S}_\emptyset(\mathtt{t}) = \mathtt{t}$.

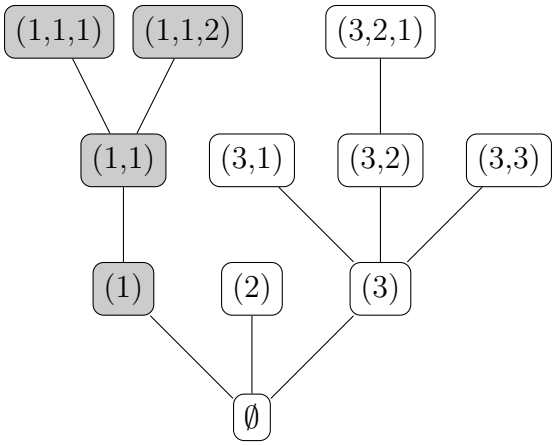

Figure 3.3: The descendants of node (1) are marked gray. Looking at them as a tree, it describes the subtree $\mathcal{S}_{(1)}(\mathtt{t})$.

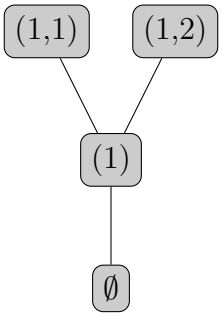

Figure 3.4: The actual subtree $\mathcal{S}_{(1)}(\mathtt{t})$ of Figure 3.3, where the entries got shifted to match the definition of a tree (see Definition 3.1).

Now we will introduce a first function on the set of trees, the so-called *restriction* or *truncation function* $tr_n(\mathtt{t})\colon \mathbb{T} \to \mathbb{T}_n$ given by

$$tr_n(\mathtt{t}) = \left\{u \in \mathtt{t} : H(u) \leq n\right\}.$$

It gives the subtree of $\mathtt{t}$ obtained by cutting at height n and only consider everything below. For a visualization, look at Figure 3.5.

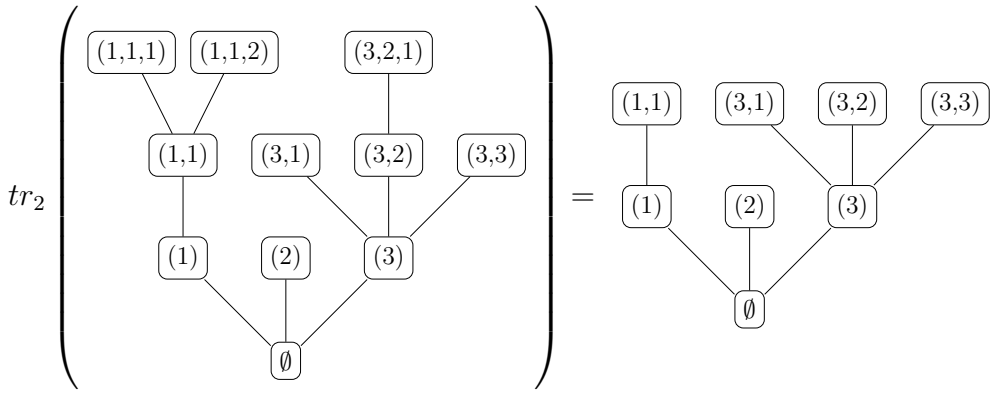

Figure 3.5: A visualization of the truncation function at height $n = 2$.

3.2 Metric structure on $\mathbb{T}$

In order to talk about convergence and limits of sequences of trees in Chapter 6, we need to endow the space $\mathbb{T}$ with a metric. Define a metric on the set $\mathbb{T}$ as

$$d_{\mathbb{T}}(\mathtt{t}, \mathtt{t}') = 2^{-m(\mathtt{t},\mathtt{t}')},$$

where $m(\mathtt{t}, \mathtt{t}') = \sup\left\{n \geq 0 : tr_n(\mathtt{t}) = tr_n(\mathtt{t}')\right\}$. So $m(\mathtt{t}, \mathtt{t}')$ is the last height where the two trees $\mathtt{t}$ and $\mathtt{t}'$ coincide. In Figure 3.6, we have $d_\pi(\mathtt{t}, \mathtt{t}') = 2^{-1} = \frac{1}{2}$, since the trees differ at height 2, so the last height where they were identical is 1. The function $d_{\mathbb{T}}$ is not only a metric but a so-called ultrametric, that is, it satisfies a stronger condition than the normal triangle inequality, namely $d_{\mathbb{T}}(\mathtt{t}, \mathtt{t}') \leq \max\left\{d_{\mathbb{T}}(\mathtt{t}, \mathtt{t}''), d_{\mathbb{T}}(\mathtt{t}'', \mathtt{t}')\right\}$.

Define open balls in the set of trees $\mathbb{T}$ as

$$B(\mathtt{t}, r) = \left\{\mathtt{t}' : d_{\mathbb{T}}(\mathtt{t}, \mathtt{t}') < r\right\}.$$

It suffices to consider r of the form $r = 2^{-n}$ for some $n \in \mathbb{N}$, as the distance between two trees can only be of the form 2^{-n}.

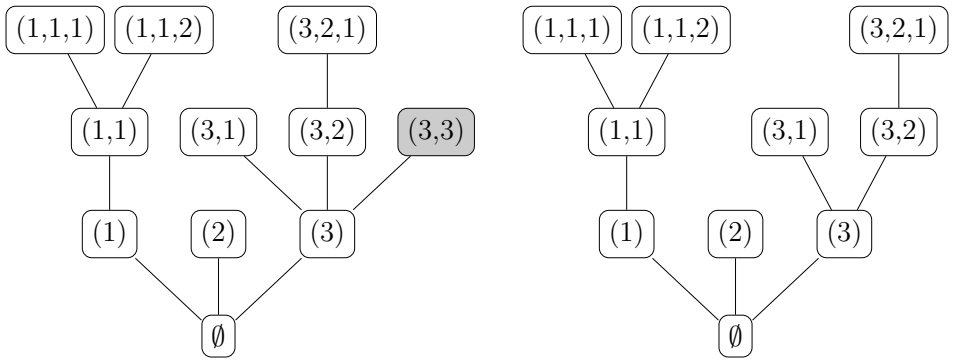

Figure 3.6: The distance of the two trees is $d_\pi(\mathtt{t}, \mathtt{t}') = 2^{-1} = \frac{1}{2}$, since they differ at height 2 and therefore the last height where they are identical is 1.

Lemma 3.2. *Every singleton* $\{\mathtt{t}\}$*, for* $\mathtt{t} \in \mathbb{T}^f$*, is both open and closed.*

Proof. Fix an arbitrary $\mathtt{t} \in \mathbb{T}^f$. The fact that $\{\mathtt{t}\}$ is open will immediately follow from verifying that we can write it as an open ball

$$\{\mathtt{t}\} = B(\mathtt{t}, 2^{-\mathcal{H}(\mathtt{t})}).$$

It is clear that $\{\mathtt{t}\} \subseteq B(\mathtt{t}, 2^{-\mathcal{H}(\mathtt{t})})$. We now show that $\{\mathtt{t}\}$ is the only element in the open ball. It suffices to check that $m(\mathtt{t}, \mathtt{t}') \leq \mathcal{H}(\mathtt{t})$ for two trees $\mathtt{t} \neq \mathtt{t}'$, because this then implies that $d(\mathtt{t}, \mathtt{t}') \geq \mathcal{H}(\mathtt{t})$ and therefore $\{\mathtt{t}\}$ is the only element in $B(\mathtt{t}, 2^{-\mathcal{H}(\mathtt{t})})$. Let us prove it by cases. If $\mathcal{H}(\mathtt{t}) > \mathcal{H}(\mathtt{t}')$, then $\mathtt{t}'$ has no vertices at height $\mathcal{H}(\mathtt{t}')+1$ while $\mathtt{t}$ does. Thus they differ at that height or even before, so $m(\mathtt{t}, \mathtt{t}') \leq \mathcal{H}(\mathtt{t}') + 1 \leq \mathcal{H}(\mathtt{t})$. For $\mathcal{H}(\mathtt{t}) < \mathcal{H}(\mathtt{t}')$ we use the same argument, just switching the labels. Consider the last case $\mathcal{H}(\mathtt{t}) = \mathcal{H}(\mathtt{t}')$. Since the trees are different by assumption, they must differ at a height smaller than $\mathcal{H}(\mathtt{t})$. So $m(\mathtt{t}, \mathtt{t}') \leq \mathcal{H}(\mathtt{t})$. Hence, singletons are open. To prove closeness, consider any $B(\mathtt{t}, 2^{-n})$ and pick $\mathtt{t}'$ in its complement, so $\mathtt{t}' \in \mathbb{T} \setminus B(\mathtt{t}, 2^{-n})$. Since singletons are open, there exists an open neighborhood around $\mathtt{t}'$, which lies entirely in the complement. This proves that $\mathbb{T} \setminus B(\mathtt{t}, 2^{-n})$ is open and by definition $B(\mathtt{t}, 2^{-n})$ is closed. □

Since we have defined a metric and the notion of open sets, we can now show that the truncation function is continuous by showing that the preimage of every open set is open.

Lemma 3.3. *The truncation function* $tr_n(\mathtt{t})\colon \mathbb{T} \to \mathbb{T}_n$ *is continuous.*

Proof. Let $U \subseteq \mathbb{T}_n$ be an open set. We can write U as the union of its singletons. As each element in $\mathbb{T}_n$ is of the form $tr_n(\mathtt{t})$ for some $\mathtt{t} \in \mathbb{T}$, we can write

$U = \bigcup_{tr_n(\mathtt{t})\in U} \{tr_n(\mathtt{t})\}$, where the union is taken over all $\mathtt{t} \in \mathbb{T}$ such that $tr_n(\mathtt{t}) \in U$. Remark that

$$tr_n^{-1}\left(\{tr_n(\mathtt{t})\}\right) = \{\mathtt{t}' \in \mathbb{T} : d_{\mathbb{T}}(\mathtt{t}, \mathtt{t}') \leq 2^{-n}\}. \tag{3.3}$$

This is precisely the open ball around $\mathtt{t}$ with radius 2^{-n}. Since the preimage commutes with arbitrary unions, we get

$$tr_n^{-1}(U) = \bigcup_{tr_n(\mathtt{t})\in U} tr_n^{-1}\left(\{tr_n(\mathtt{t})\}\right) = \bigcup_{tr_n(\mathtt{t})\in U} \{\mathtt{t}' \in \mathbb{T} : d_{\mathbb{T}}(\mathtt{t}, \mathtt{t}') \leq 2^{-n}\}$$

by making use of (3.3) in the last equality. Each set in the last union is an open ball in $\mathbb{T}$ and a union of open sets is again open, hence $tr_n^{-1}(U)$ is open. Since the preimage of each open set of $\mathbb{T}_n$ is open in $\mathbb{T}$, the truncation function is continuous. □

The fact that the space $(\mathbb{T}, d_{\mathbb{T}})$ is Polish will be crucial to make use of general theory on weak convergences of random variables with values in Polish spaces (see Chapter 6).

Lemma 3.4. *$(\mathbb{T}, d_{\mathbb{T}})$ is a non compact Polish metric space.*

Proof. In a metric space compactness implies that every sequence has a convergent subsequence. So to prove that $(\mathbb{T}, d_{\mathbb{T}})$ is non compact, it suffice to show that there exists a sequence that has no convergent subsequence. Consider the sequence $(\mathtt{t}_k)_{k\geq 0}$, which describes trees with exactly k leaves connected directly to the root and nothing else. Then for $k \neq l$

$$m(\mathtt{t}_k, \mathtt{t}_l) = \sup\{n \geq 0 : tr_n(\mathtt{t}_k) = tr_n(\mathtt{t}_l)\} = 0,$$

since they differ already at the first generation. Thus $d_{\mathbb{T}}(\mathtt{t}_k, \mathtt{t}_l) = 2^{-0} = 1$. So $(\mathtt{t}_k)_{k\geq 0}$ has no convergent subsequence and hence, $(\mathbb{T}, d_{\mathbb{T}})$ is non compact.

A metric space is Polish if it is separable and complete. We want to prove this two properties, starting with separable. We need to show that $(\mathbb{T}, d_{\mathbb{T}})$ contains a countable and dense subset. Recall that $\mathbb{T}^f$ is countable and notice that it is also dense in $\mathbb{T}$ since every $\mathtt{t} \in \mathbb{T}$ either belongs to $\mathbb{T}^f$ or is a limit point of $\mathbb{T}^f$, as the sequence $(tr_n(\mathtt{t}))_{n\in\mathbb{N}}$ converges to $\mathtt{t}$. To show completeness, we need to prove that every Cauchy sequence of elements in $\mathbb{T}$ has a limit that lies again in $\mathbb{T}$. Let $(\mathtt{t}_j)_{j\in\mathbb{N}}$ be a Cauchy sequence in $\mathbb{T}$, so for any $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that for all $j, k > N$ we have $d_{\mathbb{T}}(\mathtt{t}_j, \mathtt{t}_k) < \varepsilon$. Then also $(tr_n(\mathtt{t}_j))_{j\in\mathbb{N}}$ is Cauchy in $\mathbb{T}_n$ as $d_{\mathbb{T}}(tr_n(\mathtt{t}_j), tr_m(\mathtt{t}_k)) \leq d_{\mathbb{T}}(\mathtt{t}_j, \mathtt{t}_k) < \varepsilon$. If $\mathtt{t}, \mathtt{t}' \in \mathbb{T}_j$ and $d_{\mathbb{T}}(\mathtt{t}, \mathtt{t}') \leq 2^{-j}$, then $\mathtt{t} = \mathtt{t}'$, since two distinct trees would have a distance larger than 2^{-j}. Hence, the sequence $(tr_n(\mathtt{t}_j))_{j\in\mathbb{N}}$ is eventually constant for j large enough. Define this constant as $\mathtt{t}^n$. Remark that by continuity of the truncation function, for all $k > n$

$$tr_n(\mathtt{t}^k) = \lim_{j\to\infty} tr_n(tr_k(\mathtt{t}_j)) = \lim_{j\to\infty} tr_n(\mathtt{t}_j) = \mathtt{t}^n.$$

Define the tree $\mathtt{t} = \bigcup_{n\in N} \mathtt{t}^n$ and by construction $tr_n(\mathtt{t}) = \mathtt{t}^n$. Since $(\mathtt{t}_j)_{j\in\mathbb{N}}$ is Cauchy, there exists $N \in \mathbb{N}$ such that for all $n \geq N$ $tr_n(\mathtt{t}_j) = \mathtt{t}^n$. As the two trees $\mathtt{t}_j$ and $\mathtt{t}$ match up to height n for $n \geq N$, we have $d_{\mathbb{T}}(\mathtt{t}_j, \mathtt{t}) \leq 2^{-n}$. Since this holds for n arbitrary large, we deduce that $d_{\mathbb{T}}(\mathtt{t}_j, \mathtt{t}) \to 0$. This proves that the space is complete. □

To be able to introduce random trees in the next section, we endow the space $\mathbb{T}$ with its Borel sigma-algebra $\mathscr{B}(\mathbb{T})$, which is the aigma-algebra generated by all open sets. Since the open balls $B(\mathtt{t}, 2^{-n})$ form a basis, we have

$$\mathscr{B}(\mathbb{T}) = \sigma(\mathscr{C}) \text{ with } \mathscr{C} = \left\{\emptyset\right\} \cup \left\{B(\mathtt{t}, 2^{-n}), \mathtt{t} \in \mathbb{T}, n \in \mathbb{N}\right\}.$$

We find different generators of the Borel sigma-algebra.

Lemma 3.5. *The following holds:*

1. *$\mathscr{C}$ is a pi-system generating $\mathscr{B}(\mathbb{T})$.*
2. *The collection of sets*
$$C_u = \{\mathtt{t} \in \mathbb{T} : u \in \mathtt{t}\}$$
for $u \in \mathbb{U}$ generates $\mathscr{B}(\mathbb{T})$, so $\mathscr{B}(\mathbb{T}) = \sigma(C_u : u \in \mathbb{U})$.

Proof. Let us prove the two claims.

1. $\mathscr{C}$ is not empty as $\mathtt{t} = B(\mathtt{t}, 0) \in \mathscr{C}$. Further, it is closed under intersection. Indeed, let $\mathtt{t}_1, \mathtt{t}_2 \in \mathbb{T}$ and $B(\mathtt{t}_1, 2^{-n_1}), B(\mathtt{t}_2, 2^{-n_2}) \in \mathscr{C}$. Assume $n_1 \leq n_2$, otherwise just switch the numeration. Then
$$B(\mathtt{t}_1, 2^{-n_1}) \cap B(\mathtt{t}_2, 2^{-n_2}) = \begin{cases} B(\mathtt{t}_2, 2^{-n_2}) & \text{if } tr_{n_1}(t_1) = tr_{n_1}(t_2) \\ \emptyset & \text{else} \end{cases}$$
Since both are in $\mathscr{C}$, it is stable under intersections. Open and closed under intersections implies that it is a pi-system. It generates $\mathscr{B}(\mathbb{T})$ by definition.

2. To prove that the sets C_u generate the sigma algebra, we show the two inclusions.
   - $\sigma(C_u : u \in \mathbb{U}) \subseteq \mathscr{B}(\mathbb{T})$: Remark that $C_u = \bigcup_{\mathtt{t}\in\mathbb{T}^f\cap C_u} B(\mathtt{t}, H(u))$. Since $\mathbb{T}^f$ is countable, $C_u \in \sigma(\mathscr{C}) = \mathscr{B}(\mathbb{T})$. Since $\sigma(C_u : u \in \mathbb{U})$ is by definition the smallest sigma algebra containing C_u, we get this inclusion.
   - $\mathscr{B}(\mathbb{T}) \subseteq \sigma(C_u : u \in \mathbb{U})$: It suffices to check that $\mathscr{C} \subseteq \sigma(C_u : u \in \mathbb{U})$. Let $A \in \mathscr{C}$. Then either $A = \emptyset$ and clearly in $\sigma(C_u : u \in \mathbb{U})$, since it is a sigma algebra, or $A = B(\mathtt{t}, 2^{-n})$ for some $\mathtt{t} \in \mathbb{T}, n \in \mathbb{N}$. In the second scenario we can write $B(\mathtt{t}, 2^{-n}) = \bigcap_{u\in tr_n(\mathtt{t})} C_u$. Since $tr_n(\mathtt{t})$ is finite, we get $A = B(\mathtt{t}, 2^{-n}) \in \sigma(C_u : u \in \mathbb{U})$.

□

3.3 Random trees

Until now we only considered deterministic trees. Now we want to include some randomness. Having the metric and measurable framework, we are now able introduce random trees. A Galton Watson tree is a typical model of such trees. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

Definition 3.6. A measurable function

$$\mathbf{t} : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{T}, \mathscr{B}(\mathbb{T}))$$

is called a *random tree*.

In other words, a random tree is a $\mathbb{T}$-valued random variable. The law of the random tree $\mathbf{t}$, denoted as $\mathbb{P}_{\mathbf{t}}$, is defined as the push-forward of $\mathbb{P}$ by $\mathbf{t}$. We will show that the law is determined by its truncations. In other words, it is enough to know the probabilities $\mathbb{P}(tr_n(\mathbf{t}) = tr_n(\mathtt{t}))$ to characterize it.

Lemma 3.7. *The law of a random tree* $\mathbf{t}$ *is characterized by* $\mathbb{P}(tr_n(\mathbf{t}) = tr_n(\mathtt{t}))$.

Proof. Let μ be a probability measure on $(\mathbb{T}, \mathscr{B}(\mathbb{T}))$. It is fully determined by its value on $\mu(B(\mathtt{t}, 2^{-n})), \mathtt{t} \in \mathbb{T}, n \geq 0$, since $\mathscr{C}$ is a pi-system by Lemma 3.5. Indeed, assume μ and μ' are two different measures with $\mu(B(\mathtt{t}, 2^{-n})) = \mu'(B(\mathtt{t}, 2^{-n}))$ for all $\mathtt{t} \in \mathbb{T}, n \geq 0$. Since $\mathscr{C}$ generates $\mathscr{B}(\mathbb{T})$, by Carathéodory's extension theorem $\mu = \mu'$. Returning now to random trees, we denote the law of $\mathbf{t}$ under $\mathbb{P}$ as μ. Notice that for all $\mathtt{t} \in \mathbb{T}, n \in \mathbb{N}$

$$\left\{tr_n(\mathbf{t}) = tr_n(\mathtt{t})\right\} = \left\{\mathbf{t} \in B(\mathtt{t}, 2^{-n})\right\}.$$

Let $\mathbf{t}' : (\Omega', \mathscr{F}', \mathbb{P}') \to (\mathbb{T}, \mathscr{B}(\mathbb{T}))$ be another random tree and μ' the law of $\mathbf{t}'$ under $\mathbb{P}'$. If now $\mathbb{P}(\mathbf{t} \in B(\mathtt{t}, 2^{-n})) = \mathbb{P}'(\mathbf{t}' \in B(\mathtt{t}, 2^{-n}))$, then

$$\mu(B(\mathtt{t}, 2^{-n})) = \mathbb{P}(\mathbf{t} \in B(\mathtt{t}, 2^{-n})) = \mathbb{P}'(\mathbf{t}' \in B(\mathtt{t}', 2^{-n})) = \mu'(B(\mathtt{t}', 2^{-n})).$$

Since μ is characterized by $B(\mathtt{t}, 2^{-n})$, we get $\mu = \mu'$. Thus, the law of $\mathbf{t}$ is characterized by $\mathbb{P}(tr_n(\mathbf{t}) = tr_n(\mathtt{t}))$ for $\mathtt{t} \in \mathbb{T}$. □

From now on we will often work in the canonical space of trees $(\mathbb{T}, \mathscr{B}(\mathbb{T}), \mu)$, where μ is a probability measure on $\mathbb{T}$. We define $\mathbf{t} : \mathbb{T} \to \mathbb{T}$ to be the identity map. In this case, the law of $\mathbf{t}$ under μ is exactly μ.

Chapter 4

Galton Watson trees and Kesten's tree

4.1 Galton Watson trees

We are prepared to introduce formally Galton Watson trees and prove their existence. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A $\mathbb{T}$-valued random variable $\mathbf{t}$ with law π is said to satisfy the branching property if under $\mathbb{P}(\cdot \mid k_\emptyset(\mathbf{t}) = k)$ the subtrees $\mathcal{S}_1(\mathbf{t}), \dots, \mathcal{S}_k(\mathbf{t})$ are independent and distributed as the original tree for all $k \in \mathbb{N}^*$. In other words, conditionally on $k_\emptyset(\mathbf{t})$, $\mathcal{S}_1(\mathbf{t}), \dots, \mathcal{S}_{k_\emptyset}(\mathbf{t})$ are independent with law π.

Definition 4.1. Let $\pi = (\pi(k))_{k\in\mathbb{N}}$ be a reproduction law and $\mathbf{T} : \Omega \to \mathbb{T}$ an $\mathscr{F}$-measurable random tree. $\mathbf{T}$ is a *Galton Watson tree* with offspring distribution π if

1. For all $k \in \mathbb{N}$, $\mathbb{P}(k_\emptyset(\mathbf{T}) = k) = \pi(k)$ and
2. $\mathbf{T}$ satisfies the branching property.

We start by constructing Galton Watson trees by hand to show its existence.

Existence: We work in an appropriate auxiliary probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Denote $u = (u_1, \dots, u_n) \in \mathbb{U}$ an element of Ulam's tree and $\pi = (\pi(k))_{k\in\mathbb{N}}$ the reproduction law. Let $(X(u))_{u\in\mathbb{U}}$ be a collection of independent identically distributed random variables with law π and set

$$\mathbf{T} \coloneqq \{\emptyset\} \cup \left\{ u \in \mathbb{U} \setminus \{\emptyset\} : \forall k \in \{1, \dots H(u)\}, u_k \leq X(u_{k-1}) \right\}.$$

The goal is to show that $\mathbf{T}$ really defines a Galton Watson tree. Starting with the fact that $\mathbf{T} \in \mathbb{T}$, which means that it is ordered and rooted. By definition, $\emptyset \in \mathbf{T}$ and for all

$u \in \mathbf{T}$, its parent is also in $\mathbf{T}$. Further, if we set $k_u(\mathbf{T}) = X(u)$, then for all $u \in \mathbf{T}$ and all $i \in \mathbb{N}^*$, $ui \in \mathbf{T}$ if and only if $1 \leq i \leq k_u(\mathbf{T})$. Thus $\mathbf{T} \in \mathbb{T}$. To prove measurability, it is enough to show that $\mathbf{T}^{-1}(C_u) \in \mathscr{F}$, since those generate $\mathscr{B}(\mathbb{T})$ by Lemma 3.5. Remark that

$$\mathbf{T}^{-1}(C_u) = \{\omega \in \Omega : \mathbf{T}(\omega) \in C_u\} = \{u \in \mathbf{T}\} = \bigcap_{1 \leq k \leq H(u)} \left\{u_k \leq X(u_{k-1})\right\}.$$

Since each set of the intersection is $\mathscr{F}$-measurable, the whole in intersection is. Thus $\mathbf{T}$ is measurable, so $\mathbf{T}$ is a $\mathbb{T}$-valued random variable.

We now show that $\mathbf{T}$ satisfies the two properties characterizing Galton Watson trees. First, for $k \in N$, by definition $\mathbb{P}(k_\emptyset(\mathbf{T}) = k) = \mathbb{P}(X(\emptyset) = k) = \pi(k)$. To prove the second condition, we denote $X_j(u) = X((j)u)$ for all $j \in \mathbb{N}^*, u \in \mathbb{U}$. Remark that the families $(X_j(u))_{u \in \mathbb{U}}, j \in \mathbb{N}^*$ are independent of each other and of $X(\emptyset)$. Further, for fixed j, $(X_j(u))_{u \in \mathbb{U}}$ has the same law as $(X(u))_{u \in \mathbb{U}}$, since all $X(u)$ are independent and identically distributed with law π and the $X_j(u)$ are just a relabeling. We can now define

$$\mathbf{T}_j = \left\{\emptyset\right\} \cup \left\{u \in \mathbb{U} \setminus \left\{\emptyset\right\} : \forall k \in \left\{1, \dots H(u)\right\}, u_k \leq X_j(u_{k-1})\right\}.$$

Since $(X(u))_{u \in \mathbb{U}} \overset{(d)}{=} (X_j(u))_{u \in \mathbb{U}}$ we also get that $\mathbf{T} \overset{(d)}{=} \mathbf{T}_j$, conditionally on $k_\emptyset(\mathbf{T}) = k$. We now want to show that $\mathbf{T}_j$ are actual subtrees of $\mathbf{T}$. By definition a subtree is $\mathcal{S}_{(j)}(\mathbf{T}) = \left\{u \in \mathbb{U} : (j)u \in \mathbf{T}\right\}$. Remark that $\mathcal{S}_{(j)}(\mathbf{T}) = \mathbb{1}_{j \leq X(\emptyset)} \mathbf{T}_j$. Thus $\mathbf{T}_j$ describe indeed the subtrees and since they are independent and distributed as the original tree, $\mathbf{T}$ defines a Galton Watson tree with reproduction law π.

So Galton Watson trees as defined in Definition 4.1 exist. Next, we focus on characterizing the law of $\mathbf{T}$.

Lemma 4.2. *Let $\mathbf{T} : \Omega \to \mathbb{T}$ be a Galton Watson tree with offspring distribution π. Suppose that $\pi(0) + \pi(1) < 1$. For all $\mathtt{t} \in \mathbb{T}^f$,*

$$\mathbb{P}(\mathbf{T} = \mathtt{t}) = \prod_{u \in \mathtt{t}} \pi(k_u(\mathtt{t})) \tag{4.1}$$

and the law of $\mathbf{T}$ is characterized by π.

Remark that the prerequisite $\pi(0) + \pi(1) < 1$ is also called the non-triviality condition of π. If it would not hold, so it would be equal to one, then each node would have at most one child and there would be no branching. This is not really interesting to study.

Proof. First, we show by induction that

$$\mathbb{P}\left(tr_n(\mathbf{T}) = tr_n(\mathtt{t})\right) = \prod_{\substack{u\in\mathtt{t}\\ H(u)<n}} \pi(k_u(\mathtt{t})) \tag{4.2}$$

for $\mathtt{t} \in \mathbb{T}$ and for all $n \in \mathbb{N}^*$. Let us start with the case $n = 1$. Remark that $\left\{tr_1(\mathbf{T}) = tr_1(\mathtt{t})\right\} = \left\{k_\emptyset(\mathbf{T}) = k_\emptyset(\mathtt{t})\right\}$ and thus by the first point in the definition of a Galton Watson tree (see Definition 4.1)

$$\mathbb{P}\left(tr_1(\mathbf{T}) = tr_1(\mathtt{t})\right) = \mathbb{P}\left(k_\emptyset(\mathbf{T}) = k_\emptyset(\mathtt{t})\right) = \pi(k_\emptyset(\mathtt{t})) = \prod_{\substack{u\in\mathtt{t}\\ H(u)<1}} \pi(k_u(\mathtt{t})).$$

Suppose now that the identity holds for $n \geq 1$. Observe that

$$\left\{tr_{n+1}(\mathbf{T}) = tr_{n+1}(\mathtt{t})\right\} = \left\{k_\emptyset(\mathbf{T}) = k_\emptyset(\mathtt{t})\right\} \bigcap_{1\leq j\leq k_\emptyset(\mathtt{t})} \left\{tr_n(\mathcal{S}_{(j)}(\mathbf{T})) = tr_n(\mathcal{S}_{(j)}(\mathtt{t}))\right\}$$

By the second property of the definition of Galton Watson trees, the subtrees are independent conditionally on $k_\emptyset(\mathbf{T})$. So by also considering probabilities we get

$$\mathbb{P}\left(tr_{n+1}(\mathbf{T}) = tr_{n+1}(\mathtt{t})\right) = \mathbb{P}\left(k_\emptyset(\mathbf{T}) = k_\emptyset(\mathtt{t})\right) \prod_{1\leq j\leq k_\emptyset(\mathtt{t})} \mathbb{P}\left(tr_n(\mathbf{T}) = tr_n(\mathcal{S}_{(j)}(\mathtt{t}))\right).$$

Define the sets $U_j = \left\{(j)u \in \mathbb{U} : H(u) < n \text{ and } (j)u \in \mathtt{t}\right\}$. By induction hypothesis, $\mathbb{P}\left(tr_n(\mathbf{T}) = tr_n(\mathcal{S}_{(j)}(\mathtt{t}))\right) = \prod_{u\in U_j} \pi(k_u(\mathtt{t}))$. Since $U_1, \ldots, U_{k_\emptyset(\mathtt{t})}$ form a partition of the set $\left\{u \in \mathtt{t} : 1 \leq H(u) < n+1\right\}$, we get

$$\mathbb{P}\left(tr_{n+1}(\mathbf{T}) = tr_{n+1}(\mathtt{t})\right) = \pi(k_\emptyset(\mathtt{t})) \prod_{1\leq j\leq k_\emptyset(\mathtt{t})} \prod_{u\in U_j} \pi(k_u(\mathtt{t})) = \prod_{\substack{u\in\mathtt{t}\\ H(u)<n+1}} \pi(k_u(\mathtt{t}))$$

By induction, (4.2) holds. Notice that for any $\mathtt{t} \in \mathbb{T}^f$, there exists $n \in \mathbb{N}^*$ such that $H(u) < n$ for all $u \in \mathtt{t}$. For this n, $tr_n(\mathtt{t}) = \mathtt{t}$ and so $\left\{tr_n(\mathbf{T}) = tr_n(\mathtt{t})\right\} = \{\mathbf{T} = \mathtt{t}\}$. Thus, in this setting (4.2) is exactly (4.1). Further, by Lemma 3.7, the law of a random tree is determined by its truncations. Since $\mathbf{T}$ is a random tree by definition, its law is characterized by $\mathbb{P}(tr_n(\mathbf{T}) = tr_n(\mathtt{t}))$ and thanks to the proven Equation (4.1), it is therefore characterized by π. □

4.2 Extinction and survival probabilities of Galton Watson trees

Galton Watson trees can extinct or survive and we are interested how they behave under conditioning on these two events. For this, we need to know the probability of survival and extinction. We start by introducing some notions.

We write the *mean offspring*, which describes the expected number of children, as

$$m_\pi = \mathbb{E}[X] = \sum_{k\geq 0} k\pi(k),$$

where X is the random variable with probabilities given by the offspring distribution π. The *probability generating function* of the offspring function π, denoted by

$$\varphi_\pi(r) = \mathbf{E}[r^X] = \sum_{k\geq 0} \pi(k) r^k$$

with $r \in [0,1]$, is a very important and useful tool for the study of Galton Watson trees. If the offspring distribution is clear, we sometimes write φ for φ_π. We can also define successive iterations of the probability generating function and denote them as

$$\varphi_n(r) = \mathbb{E}[r^{Z_n}].$$

Lemma 4.3. *For all $n \in \mathbb{N}$ and $r \in [0,1]$ we have*

$$\mathbb{E}[r^{Z_{n+1}} \mid Z_n] = \varphi_\pi(r)^{Z_n} \tag{4.3}$$

and in particular $\mathbb{E}[r^{Z_n} \mid Z_0] = \varphi_n(r)^{Z_0}$.

Proof. Define $\mathscr{F}_0 = \sigma(Z_0)$ and for all $n \geq 1$, let $\mathscr{F}_n = \sigma(Z_0, X_{p,j} : j \in \mathbb{N}^*, 0 \leq p \leq n-1)$. We immediately see that $(\mathscr{F}_n)_{n\geq 0}$ is a filtration: $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$ as we just get more elements $X_{p,j}$ inside the sigma-algebra. We then can write

$$\begin{aligned}
\mathbb{E}\left[r^{Z_{n+1}} \;\middle|\; \mathscr{F}_n\right] &= \mathbb{E}\left[r^{\sum_{1\leq j\leq Z_n} X_{n,j}} \;\middle|\; \mathscr{F}_n\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\{Z_n=0\}} + \sum_{k\geq 1} \mathbb{1}_{\{Z_n=k\}} \prod_{1\leq j\leq k} r^{X_{n,j}} \;\middle|\; \mathscr{F}_n\right],
\end{aligned}$$

where we first use the definition of Z_n and in the second equation we applied exponent laws and the fact that we can partition the event according to the value of Z_n. Since those random variables are nonnegative, we can switch the series and the expectation.

Moreover, the indicator functions are $\mathscr{F}_n$-measurable, so they can be taken ouside of the expectation. Therefore,

$$\begin{aligned}\mathbb{E}\left[r^{Z_{n+1}} \;\middle|\; \mathscr{F}_n\right] &= \mathbb{1}_{\{Z_n=0\}} + \sum_{k\geq 1} \mathbb{1}_{\{Z_n=k\}} \mathbb{E}\left[\prod_{1\leq j\leq k} r^{X_{n,j}} \;\middle|\; \mathscr{F}_n\right] \\ &= \mathbb{1}_{\{Z_n=0\}} + \sum_{k\geq 1} \mathbb{1}_{\{Z_n=k\}} \prod_{1\leq j\leq k} \mathbb{E}\left[r^{X_{n,j}}\right] \qquad (4.4)\end{aligned}$$

and in the last equation we use the fact that for all $k \geq 1$, the random variables $r^{X_{n,k}}$ are independent of $\mathscr{F}_n$, since by definition $\mathscr{F}_n$ only contains information until height $n-1$ and $\{X_{n,j} : j \geq 1\}$ are of generation n. Observe further that $\varphi(r) = \mathbb{E}[r^{X_{n,j}}]$ for all j and n, since $X_{n,j}$ are identically distributed as X. Substituting this into Equation (4.4), we get

$$\mathbb{E}\left[r^{Z_{n+1}} \;\middle|\; \mathscr{F}_n\right] = \mathbb{1}_{\{Z_n=0\}} + \sum_{k\geq 1} \mathbb{1}_{\{Z_n=k\}} \varphi(r)^k = \varphi(r)^{Z_n}.$$

This proves the first part. We prove the follow up by induction. For $n = 1$, we get the result by looking at the definition of $\varphi_n(r)$ and using the fact that Z_0 is independent. Let the identity be true for n. Then

$$\begin{aligned}\mathbb{E}[r^{Z_{n+1}} \mid Z_0] &= \mathbb{E}\left[\mathbb{E}[r^{Z_{n+1}} \mid Z_n] \mid Z_0\right] \\ &= \mathbb{E}\left[\varphi(r)^{Z_n} \mid Z_0\right] \\ &= \left(\varphi_n(\varphi(r))\right)^{Z_0}\end{aligned}$$

where we first used the tower property, then Equation (4.3) and in the last step the induction hypothesis. To conclude, observe that

$$\varphi_n(\varphi(r)) = \mathbb{E}[\varphi(r)^{Z_n}] = \mathbb{E}\left[\mathbb{E}[r^{Z_{n+1}} \mid Z_n]\right] = \mathbb{E}[r^{Z_{n+1}}] = \varphi_{n+1}(r). \qquad (4.5)$$

□

Looking at Equation (4.5) in the proof of Lemma 4.3, we see that $\varphi_n(r)$ can also be defined recursively with

$$\begin{aligned}\varphi_0(r) &= r \\ \varphi_{n+1}(r) &= \varphi(\varphi_n(r)) = \varphi_n(\varphi(r)).\end{aligned}$$

Later, we prove that the extinction probability is a fixed point of the probability generating function φ_π. Therefore, we first investigate the solutions of the fixed-point equation $\varphi_\pi(r) = r$ in the interval $[0, 1]$.

Lemma 4.4. *Suppose non-triviality of π, so $\pi(0) + \pi(1) < 1$. Then the function φ_π is strictly convex and this implies that φ'_π is strictly increasing. Consider the equation $\varphi_\pi(r) = r$ for $r \in [0,1]$ and denote q_π as its smallest solution.*

1. *If $m_\pi \leq 1$, then there exists exactly one solution for $r \in [0,1]$, namely $r = 1$, so $q_\pi = 1$.*
2. *If $m_\pi > 1$, then there are two distinct solutions for $r \in [0,1]$. One is $r = 1$ and the other one is strictly smaller than 1, so $q_\pi < 1$. In particular, if $\pi(0) = 0$, then $q_\pi = 0$ and if $\pi(0) > 0$, then $q_\pi \in (0,1)$.*

Moreover, $(\varphi_n(0))_{n\geq 0}$ is an increasing sequence which converges to q_π.

Proof. In this proof, we will write $\varphi = \varphi_\pi$. Note that

$$\varphi'(r) = \frac{d}{dr}\left(\sum_{k\geq 0} \pi(k) r^k\right) = \sum_{k\geq 1} k\pi(k) r^{k-1}$$

where we can change the differential and the integral since the series converges. The second derivative is

$$\varphi''(r) = \frac{d}{dr}\left(\sum_{k\geq 1} k\pi(k) r^{k-1}\right) = \sum_{k\geq 2} k(k-1)\pi(k) r^{k-2}. \tag{4.6}$$

First, let's prove that φ is strictly convex under the non-triviality assumption. Since $\pi(0) + \pi(1) < 1$, there exists $k \geq 2$ such that $\pi(k) > 0$. So $\varphi''(r) > 0$ for all $r \in (0,1]$ using (4.6). By a result from analysis, a function is strictly convex if its second derivative is positive. Thus, the function φ is strictly convex on $(0,1]$ and by continuity on the entire interval $[0,1]$. Strict convexity and differentiability of a function implies its derivative is strictly increasing is a known fact from analysis.

Remark that $\varphi(r) = r$ has always a solution, namely $r = 1$, since

$$\varphi(1) = \sum_{k\geq 0} \pi(k) 1^k = \sum_{k\geq 0} \pi(k) = 1.$$

Further, by looking at the first derivative of φ, we observe that

$$m_\pi = \sum_{k\geq 0} k\pi(k) = \varphi'(1). \tag{4.7}$$

Define $f(r) = \varphi(r) - r$. f is still strictly convex, since the second derivative is still positive as r just vanishes, and f' is strictly increasing. Let's look at the two different cases.

- Suppose $m_\pi \leq 1$. Then for all $r < 1$, we have

$$f'(r) < f'(1) = m_\pi - 1 \leq 0$$

where the first inequality holds since f' is strictly increasing. Hence, f is strictly decreasing on $[0, 1)$ as the derivative is negative. That is for all $r \in [0, 1)$,

$$f(r) > f(1) = \varphi(1) - 1 = 0.$$

So $\varphi(r) > r$ for all $r \in [0, 1)$. Thus there exits no other solution than 1, which is therefore also the smallest solution, so $q_\pi = 1$.

- Suppose now that $m_\pi > 1$. We want to find a second solution and also show that there are no more. Assume first that $\pi(0) = 0$. Then, $\varphi(0) = \pi(0) = 0$, so $r = 0$ is a solution and directly the smallest in the interval $[0, 1]$, hence $q_\pi = 0$. Now we assume that $\pi(0) > 0$. As $m_\pi > 1$, we have

$$f'(1) = \varphi'(1) - 1 = m_\pi - 1 > 0.$$

Consider now a point near 1, so $r \in (1 - \varepsilon, 1)$ for a small ε. By the mean value theorem, we have for all $\xi \in (r, 1)$

$$f(1) - f(r) = (1 - r)f'(\xi).$$

Since also $\xi \in (1 - \varepsilon, 1)$, we know that $f'(\xi) > 0$. Thus,

$$0 - f(r) = (1 - r)f'(\xi) > 0,$$

which implies $f(r) < 0$ for $r \in (1-\varepsilon, 1)$. Notice that $f(0) = \varphi(0) - 0 = \pi(0) > 0$. By the intermediate value theorem, there exists $\hat{r} \in (0, 1)$ such that $f(\hat{r}) = 0$. This is our second solution to $\varphi(r) = r$, which is between 0 and 1, so $q_\pi \in (0, 1)$. To show that there exist only two solutions and not more, remark that convex functions intersect a horizontal at most twice.

To prove that $(\varphi_n(0))_{n\geq 0}$ converges to q_π, we start by noticing that $\varphi_n(0)$ is bounded since $\varphi([0, 1]) \subseteq [0, 1]$. Further, it is monotone and non decreasing by induction. So by the monotone sequence theorem, the sequence has a limit, denote it by q. It is indeed a solution of $\varphi(r) = r$, since by continuity

$$q = \lim_{n\to\infty} \varphi_n(0) = \varphi\left(\lim_{n\to\infty} \varphi_{n-1}(0)\right) = \varphi(q).$$

We want to show that $q = q_\pi$, so that the limit q is the smallest solution. Suppose p is another solution of $\varphi(r) = r$. Since $0 \leq p$ and φ is non decreasing, we get

$\varphi(0) \leq \varphi(p) = p$ and by induction then $\varphi_n(0) \leq p$ for all $n \geq 0$. Taking the limit leads to

$$q = \lim_{n\to} \varphi_n(0) \leq p.$$

Thus, q is the smallest solution, so $q_\pi = q = \lim_{n\to\infty} \varphi_n(0)$. □

In the previous Lemma 4.4, we distinguished several possibilities for solutions of $\varphi_\pi(r) = r$ depending on the value of the mean offspring: $m_\pi > 1$, $m_\pi < 1$ and $m_\pi = 1$. Those three cases naturally lead to the following important trichotomy:

Definition 4.5. A Galton Watson tree and its offspring distribution π are called

- *sub-critical* if $m_\pi < 1$,
- *critical* if $m_\pi = 1$,
- *super-critical* if $m_\pi > 1$.

Introducing now the *time of extinction*

$$T = \inf \{n \in \mathbb{N} : Z_n = 0\}$$

with the convention that $\inf(\emptyset) = \infty$. The event of extinction of the population is then $\{T < \infty\}$ and the event that the population will survive is given by $\{T = \infty\}$. Let us now connect these events with the smallest solution of the fixed-point equation $\varphi_\pi(r) = r$.

Lemma 4.6. *If T denotes the time of extinction and q_π is the smallest solution of $\varphi(r) = r$, then*

- $\mathbb{P}(T < \infty) = q_\pi$
- $\mathbb{P}(T = \infty) = 1 - q_\pi$.

Proof. Remark that $\{Z_n = 0\} \subset \{Z_{n+1} = 0\}$, as $Z_n = 0$ implies that $Z_m = 0$ for all $m \geq n$, and that $\{T < \infty\} = \bigcup_{n\in\mathbb{N}} \{Z_n = 0\}$. So since for $n \to \infty$ the events $\{Z_n = 0\}$ are increasing, $\mathbb{P}(T < \infty) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0)$. We also know

$$\varphi_n(r) = \mathbb{E}[r^{Z_n}] = \mathbb{E}\left[r^{Z_n} \mathbb{1}_{\{Z_n \geq 1\}}\right] + \mathbb{P}(Z_n = 0),$$

which directly implies $\varphi_n(0) = \mathbb{P}(Z_n = 0)$. Thus, we have

$$\mathbb{P}(T < \infty) = \lim_{n\to\infty} \varphi_n(0) = q_\pi,$$

where for the last equality Lemma 4.4 was used. It follows immediately that

$$\mathbb{P}(T = \infty) = 1 - \mathbb{P}(T < \infty) = 1 - q_\pi.$$

□

As the process $(Z_n)_{n\geq 0}$ is directly connected with a Galton Watson tree, the population survives if and only if the corresponding tree survives. We say that a Galton Watson tree survives if the tree is infinite or in other words if the height is infinite. Thus, the probability of survival for a Galton Watson tree is exactly the probability that the extinction time satisfies $T = \infty$, so $1 - q_\pi$. In contrast, extinction occurs when the Galton Watson tree is finite or equivalently has a finite height. So the event of extinction has the same probability as the event $\{T < \infty\}$, which is given by q_π.

We want to stick to the notation $\{\mathcal{H}(\mathtt{t}) < \infty\}$ respectively $\{\mathcal{H}(\mathtt{t}) = \infty\}$ when we talk about extinction and survival, but will sometimes also use others. This connection between the genealogical trees and their populations, together with the two previous lemmas, leads to the following corollary.

Corollary 4.7. *Let π be an offspring distribution of the Galton Watson tree $\mathbf{T}$. If π is sub-critical or if π is critical with $\pi(1) < 1$, then $\mathbf{T}$ gets extincted almost surely, that is $\mathbb{P}(\mathcal{H}(\mathbf{T}) < \infty) = 1$.*

Proof. Let π be a sub-critical offspring distribution or critical one with $\pi(1) < 1$. In the critical case, this extra condition excludes the case where every individual has exactly one offspring and therefore extinction cannot happen. Then, by Lemma 4.6, we have $\mathbb{P}(\mathcal{H}(\mathbf{T}) < \infty) = \mathbb{P}(T < \infty) = q_\pi$. Further, Lemma 4.4 states that for $m_\pi \leq 1$, which is true for the sub-critical and critical case, we have $q_\pi = 1$. Putting it together gives the result. □

With this corollary, observe that super-critical Galton Watson trees are the only ones that survive with a positive probability smaller than one. This implies that for such trees, we get both extinction and survival with a positive probability. Thus we can condition on both and analyze the resulting law of the tree, which will be done in Chapter 5. Contrarily, in the sub-critical and critical scenario, the tree gets extincted almost surely. So condition on extinction will be trivial, as it happens with probability one, and gives no new information. On the other hand, conditioning on survival is not well-defined as the probability is zero. We will have a closer look at the critical case in Chapter 6 using a different approach.

4.3 The branching property

We have identified the probabilities of extinction and survival for Galton Watson trees. To understand the structure of Galton Watson trees more deeply, especially later when we look at conditioned trees, we need to show a fundamental property of Galton Watson trees, which is called the *branching property*. Roughly, it says that if you look at any node in a Galton Watson tree, the subtree growing from this node is again a Galton Watson tree with the same offspring distribution. Further, the subtrees rooted at all the different children belonging to the same generation are independent of each other. This branching property also directly implies, for $\mathbf{T}$ a Galton Watson tree, that the associated process $(Z_n(\mathbf{T}))_{n\in\mathbb{N}}$ is a branching process with the same reproduction law as $\mathbf{T}$.

To do all of this formally, define first the sigma-algebras

$$\mathscr{F}_n = \sigma(tr_n(\mathtt{t})).$$

Since for $n < m$ we can write $tr_n(tr_m(\mathtt{t})) = tr_n(\mathtt{t})$, as truncating at a higher height followed by a lower is exactly the same as just cutting at the lower height, $tr_n(\mathtt{t})$ is a measurable function of $tr_m(\mathtt{t})$. Then, $\sigma(tr_n(\mathtt{t})) \subseteq \sigma(tr_m(\mathtt{t}))$ and by definition $\mathscr{F}_n \subseteq \mathscr{F}_m$. Hence, $(\mathscr{F}_n)_{n\geq 0}$ is a filtration.

Recall the definition of the number of individuals at height n,

$$Z_n(\mathtt{t}) = \#\{u \in \mathtt{t} : H(u) = n\}. \tag{4.8}$$

We can also rewrite this expression as

$$Z_n(\mathtt{t}) = \sum_{\substack{u\in\mathbb{U}\\ H(u)=n}} \mathbb{1}_{C_u}$$

where $C_u = \{\mathtt{t} \in \mathbb{T} : u \in \mathtt{t}\}$. To see this, observe that $\mathtt{t} \in C_u$ if and only if $u \in \mathtt{t}$. So we add 1 to $Z_n(\mathtt{t})$ if and only if $u \in \mathtt{t}$ and $H(u) = n$, hence the sum counts exactly all nodes at height n. If we set $H(u) = n$, then the set C_u is $\mathscr{F}_n$-measurable since we can write

$$C_u = \{\mathtt{t} \in \mathbb{T} : u \in tr_n(\mathtt{t})\} = tr_n^{-1}\left(\{\mathtt{t} \in \mathbb{T} : u \in \mathtt{t}\}\right) = tr_n^{-1}(C_u).$$

Thus, also $\mathbb{1}_{C_u}$ is measurable and since the sum is countable, each Z_n is $\mathscr{F}_n$-measurable. In other words, $(Z_n)_{n\geq 0}$ is adapted to $(\mathscr{F}_n)_{n\geq 0}$.

For the next lemma we need some technicalities and definitions. To be precise, we need to add a cemetery point † to $\mathbb{U}$. So from now on, we have

$$\mathbb{U} = \bigcup_{n\geq 0} (\mathbb{N}^*)^n \cup \{\dagger\}.$$

All the previous definitions and results are still valid. Fix $n \in \mathbb{N}$ and a tree $\mathtt{t}$. We now enumerate the vertices at generation n as follows:

$$\tau_1^n(t) \coloneqq \inf\{u \in \mathtt{t} : H(u) = n\}$$
$$\tau_k^n(t) \coloneqq \inf\{u \in \mathtt{t}, \tau_{k-1}(\mathtt{t}) < u : H(u) = n\} \text{ for } k \geq 2,$$

where the infimum is with respect to the lexical order on $\mathbb{U}$ and with the convention that $\inf\{\emptyset\} = \dagger$. See Figure 4.1 for an example of the numeration.

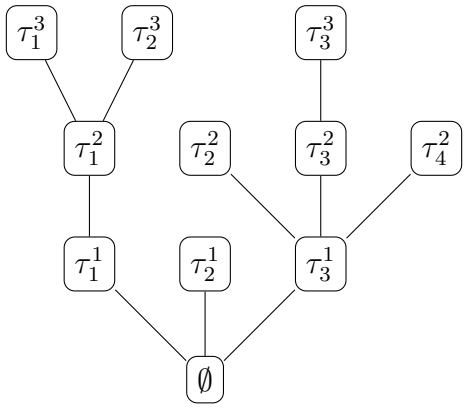

Figure 4.1: Numeration of the nodes.

We identify any point u with its spine. So each node $u \in \mathbb{U}$ is now its spine tree spine$(u) \in \mathbb{T}$, which is the tree consisting of u and all its ancestors. Under this identification, nodes can be seen as trees and so $\tau_j^n : \mathbb{T} \to \mathbb{T}$ is a $\mathscr{B}(\mathbb{T})$-measurable function. Further, it is $\mathscr{F}_n$-measurable, since $\tau_j^n(\mathtt{t}) = \tau_j^n(tr_n(\mathtt{t}))$. We want to show that $\mathcal{S}_u(\mathbf{T})$ is a random variable, such that all the following definitions and equations make sense. To do this, extend the definition of $\mathcal{S}_u(\mathtt{t})$ to

$$\mathcal{S}_u(\mathtt{t}) = \begin{cases} \mathcal{S}_u(\mathtt{t}) & u \in \textbf{Spine} \text{ and } u \in \mathtt{t} \\ \dagger & \text{else.} \end{cases}$$

We can look at $(\mathtt{t}, u) \to \mathcal{S}_u(\mathtt{t})$ as a function from $\mathbb{T} \times \mathbb{T} \to \mathbb{T}$ instead of $\mathbb{T} \times \mathbb{U} \to \mathbb{T}$, since we can identify each $u \in \mathbb{U}$ with its spine. It allows us to show that the function is measurable, as we have defined a sigma-algebra on $\mathbb{T}$. Recall that the Borel sigma-algebra is generated by the sets $C_v = \{\mathtt{t} \in \mathbb{T} : v \in \mathtt{t}\}$, so it suffices to show that $\mathcal{S}_u^{-1}(C_v)$ is measurable. Consider the case where $u \in \textbf{Spine}$ and $u \in \mathtt{t}$. The preimage

is given by

$$\begin{aligned}\mathcal{S}_u^{-1}(C_v) &= \{(\mathtt{t}, u) \in \mathbb{T} \times \mathbb{T} : \mathcal{S}_u(\mathtt{t}) \in C_v\} \\ &= \{(\mathtt{t}, u) \in \mathbb{T} \times \mathbb{T} : v \in \mathcal{S}_u(\mathtt{t})\} \\ &= \{(\mathtt{t}, u) \in \mathbb{T} \times \mathbb{T} : \exists w \in \mathbb{U} : u = \text{spine}(w) \text{ and } wv \in \mathtt{t}\} \\ &= \bigcup_{w \in \mathbb{U}} (C_{wv} \times \{\text{spine}(w)\})\end{aligned}$$

The sets C_{wv} are measurable by definition. Further, $\{\text{spine}(w)\}$ describes a finite tree and by Lemma 3.2, it is open and therefore also measurable. As $\mathbb{U}$ is countable and a countable union of measurable sets is again measurable, $\mathcal{S}_u^{-1}(C_v)$ is measurable.

Let us now investigate the branching property. Recall that, informally, we want to show that at each node of the tree, we see again the same construct as at the root: a Galton Watson tree with offspring distribution π.

Lemma 4.8. *The branching property tells us that conditional on $\mathscr{F}_n$, the subtrees $\mathcal{S}_{\tau_1}(\mathbf{T}), \ldots, \mathcal{S}_{\tau_{Z_n}}(\mathbf{T})$ are independent and with the same law π as the Galton Watson tree $\mathbf{T}$ under $\mathbb{P}$. In other words, for any non-negative $\mathscr{F}_n$-measurable function $\Phi : \Omega \to \mathbb{R}$ and measurable non-negative functions $F_j : \mathbb{T} \to \mathbb{R}_+$, the following equality is fulfilled*

$$\mathbb{E}\left[\Phi \cdot \prod_{j=1}^{Z_n} F_j(\mathcal{S}_{\tau_j}(\mathbf{T}))\right] = \mathbb{E}\left[\Phi \cdot \prod_{j=1}^{Z_n} \mathbb{E}[F_j(\mathbf{T})]\right] \tag{4.9}$$

Proof. We show it by induction. It suffices to show it for $\Phi = \mathbb{1}_{\{tr_n(\mathbf{T})=\mathtt{t}\}}$, because they generate the sigma-algebra and with usual arguments, the property extents to general Φ. Beginning with $n = 1$, remark that this is exactly the second point in Definition 4.1. So it is fulfilled. Suppose now the identity holds for n. Observe that

$$\mathbb{1}_{\{tr_{n+1}(\mathbf{T})=\mathtt{t}\}} = \mathbb{1}_{\{k_\emptyset(\mathbf{T})=k_\emptyset(\mathtt{t})\}} \prod_{l=1}^{Z_1} \mathbb{1}_{\left\{tr_n\left(\mathcal{S}_{\tau_l^1}(\mathbf{T})\right)=\mathcal{S}_{\tau_l^1}(\mathtt{t})\right\}}, \tag{4.10}$$

since checking if the trees are the same up to height $n+1$ is equivalent to verifying that their roots have the same amount of children and that the subtrees rooted at each of those children of coincide up to height n. We can also rewrite

$$\prod_{j=1}^{Z_{n+1}} F_{\tau_j^{n+1}}(\mathcal{S}_{\tau_j}^{n+1}(\mathbf{T})) = \prod_{l=1}^{Z_1} \left(\prod_{j=1}^{Z_n(\mathcal{S}_{\tau_l^1}(\mathtt{t}))} F_{\tau_l^1 \tau_j^n}(\mathcal{S}_{\tau_j^n}(\mathcal{S}_{\tau_l^1}(\mathbf{T}))) \right), \tag{4.11}$$

where for easier notation we renamed $F_j = F_{\tau_j^n}$ in the statement. Both sides of the equation go over all subtrees $\mathcal{S}_{\tau_j^{n+1}}$. Combining the two Equations (4.10) and (4.11), then taking the expectation yields to

$$\mathbb{E}\left[\mathbb{1}_{\{tr_{n+1}(\mathbf{T})=\mathtt{t}\}}\prod_{j=1}^{Z_{n+1}} F_{\tau_j^{n+1}}(\mathcal{S}_{\tau_j}^{n+1}(\mathbf{T}))\right]$$
$$= \mathbb{E}\left[\mathbb{1}_{\{k_\emptyset(\mathbf{T})=k_\emptyset(\mathtt{t})\}}\prod_{l=1}^{Z_1}\left(\mathbb{1}_{\{tr_n(\mathcal{S}_{\tau_l^1}(\mathbf{T}))=\mathcal{S}_{\tau_l^1}(\mathtt{t})\}}\prod_{j=1}^{Z_n(\mathcal{S}_{\tau_l^1}(\mathtt{t}))} F_{\tau_l^1\tau_j^n}\big(\mathcal{S}_{\tau_j^n}(\mathcal{S}_{\tau_l^1}(\mathbf{T}))\big)\right)\right]. \quad (4.12)$$

We can now apply the base case for the subtrees $\mathcal{S}_{\tau_l^1}$ and use independence. The right hand side of Equation (4.12) is then given by

$$\pi(k_\emptyset(\mathtt{t}))\prod_{l=1}^{Z_1}\mathbb{E}\left[\mathbb{1}_{\{tr_n(\mathbf{T})=\mathcal{S}_{\tau_l^1}(\mathtt{t})\}}\prod_{j=1}^{Z_n(\mathcal{S}_{\tau_l^1}(\mathtt{t}))} F_{\tau_l^1\tau_j^n}\big(\mathcal{S}_{\tau_j^n}(\mathbf{T})\big)\right].$$

By the induction hypothesis, we further can write this part as

$$\pi(k_\emptyset(\mathtt{t}))\prod_{l=1}^{Z_1}\mathbb{P}(tr_n(\mathbf{T})=\mathcal{S}_{\tau_l^1}(\mathtt{t}))\prod_{j=1}^{Z_n(\mathcal{S}_{\tau_l^1}(\mathtt{t}))}\mathbb{E}\Big[F_{\tau_l^1\tau_j^n}(\mathbf{T})\Big]. \quad (4.13)$$

Remark that $(\{\tau_l^1\tau_j^n : 1\le j\le Z_n(\mathcal{S}_{\tau_l^1})\})_{1\le l\le Z_1}$ describes exactly the same as the set $\{\tau_j^{n+1} : 1\le j\le Z_{n+1}(\mathtt{t})\}$ and $\mathbb{P}(tr_{n+1}(\mathbf{T})=\mathtt{t}) = \pi(k_\emptyset(\mathtt{t}))\prod_{l=1}^{Z_1}\mathbb{P}(tr_n(\mathbf{T})=\mathcal{S}_{\tau_l^1}(\mathtt{t}))$. Using those facts and inserting the right hand side (4.13) into Equation (4.12) we get

$$\mathbb{E}\left[\mathbb{1}_{\{tr_{n+1}(\mathbf{T})=\mathtt{t}\}}\prod_{j=1}^{Z_{n+1}(\mathtt{t})} F_{\tau_j^{n+1}}(\mathcal{S}_{\tau_j}^{n+1}(\mathbf{T}))\right] = \mathbb{P}(tr_{n+1}(\mathbf{T})=\mathtt{t})\prod_{j=1}^{Z_{n+1}(\mathtt{t})}\mathbb{E}\Big[F_{\tau_j^{n+1}}(\mathbf{T})\Big]$$
$$= \mathbb{E}\left[\mathbb{1}_{\{tr_{n+1}(\mathbf{T})=\mathtt{t}\}}\prod_{j=1}^{Z_{n+1}(\mathtt{t})}\mathbb{E}\left[F_{\tau_j^{n+1}}(\mathbf{T})\right]\right]$$

This is exactly the branching property for height $n+1$. Thus by induction, (4.9) is proven. □

Thanks to the branching property of Galton Watson trees, we can easily show that the sequence describing the number of children in each generation, denoted by $(Z_n(\mathbf{T}))_{n\ge0}$, is a branching process.

Lemma 4.9. *Let $\mathbf{T} : \Omega \to \mathbb{T}$ be a Galton Watson tree with reproduction law π. Then $(Z_n(\mathbf{T}))_{n\geq 0}$, with each $Z_n(\mathbf{T})$ defined as in Equation (4.8), is a branching process with offspring distribution π.*

Proof. Start by observing that

$$Z_{n+1}(\mathbf{T}) = \sum_{j=1}^{Z_n} k_{\tau_j}(\mathbf{T}) = \sum_{j=1}^{Z_n} k_\emptyset(\mathcal{S}_{\tau_j}(\mathbf{T})),$$

since we just shift everything. By Lemma 4.8, conditionally on $\mathscr{F}_n$, the subtrees $\mathcal{S}_{\tau_j}$ are independent and with the same law π as $\mathbf{T}$. Since $Z_{n+1}(\mathbf{T})$ is the sum of those random variables with law π, its law is $\pi^{*Z_n(\mathbf{T})}$. Finally, for all functions $f : \mathbb{N} \to [0, \infty)$, we have

$$\begin{aligned}\mathbb{E}[f(Z_{n+1}(\mathbf{T})) \mid \mathscr{F}_n] &= \sum_{k\in\mathbb{N}} f(k)\mathbb{P}(Z_{n+1}(\mathbf{T}) = k) \\ &= \sum_{k\in\mathbb{N}} f(k)\pi^{*Z_n(\mathbf{T})}(k) \\ &= (Qf)(Z_n(\mathbf{T})).\end{aligned}$$

So $(Z_n(\mathbf{T}))_{n\geq 0}$ is a Markov chain with transition matrix $Q(i, j) = \pi^{*i}(j)$. Therefore, by Proposition 2.2, $(Z_n(\mathbf{T}))_{n\geq 0}$ is a branching process with offspring distribution π. □

Using the branching property, one can give an alternative proof to show that the extinction probability is a solution to the fixed-point equation $\varphi_\pi(r) = r$. Suppose the probability of extinction is q_π. Let k be the number of children of the root. Remark that extinction occurs if all of the k subtrees rooted at the children become extinct. So

$$\begin{aligned}q_\pi &= \mathbb{P}(\mathcal{H}(\mathbf{T}) < \infty) \\ &= \sum_{k\in\mathbb{N}} \mathbb{P}(\mathcal{H}(\mathcal{S}_1(\mathbf{T}) < \infty), \ldots, \mathcal{H}(\mathcal{S}_k(\mathbf{T}) < \infty) \mid k_\emptyset(\mathbf{T}) = k)\pi(k) \\ &= \sum_{k\in\mathbb{N}} \mathbb{P}(\mathcal{H}(\mathbf{T}) < \infty)^k \pi(k) \\ &= \varphi_\pi(\mathbb{P}(\mathcal{H}(\mathbf{T}) < \infty)) = \varphi_\pi(q_\pi),\end{aligned}$$

where in the third equality the branching property was used. So the extinction probability really satisfies the fixed-point equation.

The branching property is not essential for this, as we also derived the probability using properties of the probability generating function. However, it becomes crucial later when we study conditioned Galton Watson trees.

4.4 Two-type Galton Watson trees

In all the previous chapters, we considered Galton Watson trees with single-type nodes, so all the nodes had the same probability distribution. We will now introduce a two-type Galton Watson tree, an example is drawn in Figure 4.2. These trees will appear in Chapter 5.2, when we condition super-critical Galton Watson trees on survival.

In our case, we consider the two types survivor and extinct individual. Let $u \in \mathtt{t}$ be a node of the tree. We say that u is a *survivor* if $\#\{\mathcal{S}_u(\mathtt{t})\} = \infty$ and call it *extinct* otherwise. Denote S as the set of all individuals which are survivors and E the set of all extinct individuals. Since $\mathcal{S}_\emptyset(\mathtt{t}) = \mathtt{t}$, the probability that the root $\emptyset$ is a survivor is the same as the probability of survival of the tree, which is given by $1 - q_\pi$ according to Lemma 4.6.

For the definition of a two-type Galton Watson tree, we first take a closer look at the generating function of (S, E) under the condition that the tree survives.

Lemma 4.10. *Let* $\mathbf{T}$ *be a super-critical Galton Watson tree with offspring distribution* π*. Recall* q_π *to be the the smallest solution for* $\varphi(r) = r$ *on* $[0, 1]$*. Then for* $r, l \in [0, 1]$*,*

$$\mathbb{E}[r^S l^E \mid \mathcal{H}(\mathbf{T}) = \infty] = \frac{\varphi((1 - q_\pi)r + q_\pi l) - \varphi(q_\pi l)}{1 - q_\pi}.$$

Proof. Remark that surviving is the same as saying that $S \geq 1$. By the definition of conditional expectation, we have

$$\mathbb{E}[r^S l^E \mid \mathcal{H}(\mathbf{T}) = \infty] = \frac{\mathbb{E}[r^S l^E \mathbb{1}_{\{S \geq 1\}}]}{\mathbb{P}(S \geq 1)} = \frac{1}{1 - q_\pi} \mathbb{E}[r^S l^E \mathbb{1}_{\{S \geq 1\}}].$$

By Lemma 4.4, $q_\pi < 1$ for a super-critical Galton Watson tree. This ensures that $1 - q_\pi > 0$ in order to not divide by 0. Further,

$$\begin{aligned}
\mathbb{E}[r^S l^E \mid \mathcal{H}(\mathbf{T}) = \infty] &= \frac{1}{1 - q_\pi} \sum_{n \in \mathbb{N}^*} \pi(n) \sum_{k=1}^{n} \binom{n}{k} (1 - q_\pi)^k q_\pi^{n-k} r^k l^{n-k} \\
&= \frac{1}{1 - q_\pi} \sum_{n \in \mathbb{N}^*} \pi(n) \left(((1 - q_\pi)r + q_\pi l)^n - (q_\pi l)^n \right) \\
&= \frac{\varphi((1 - q_\pi)r + q_\pi l) - \varphi(q_\pi l)}{1 - q_\pi}.
\end{aligned}$$

We get the first equation from the fact that $\#S$ and $\#E = n - \#S$ are binomially distributed and conditioning on survival eliminates the case $S = 0$, that is $k = 0$. □

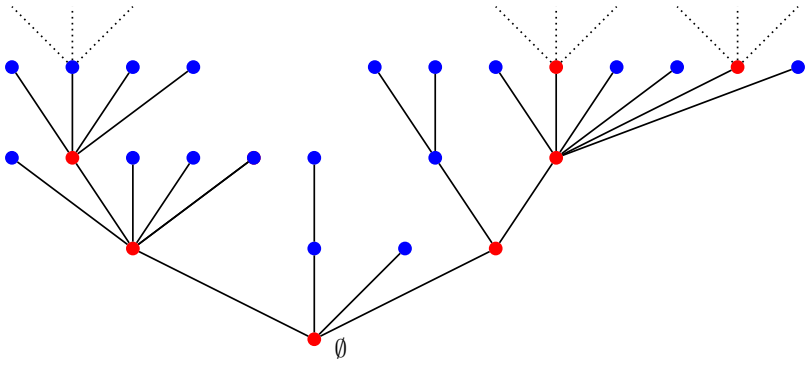

Figure 4.2: Two-type Galton Watson tree with type s nodes (red) and type e nodes (blue).

Definition 4.11. $\tilde{\mathbf{T}}^s$ is a *two-type Galton Watson tree* if its structure is as follows: There are two types of nodes: s and e.

- A node of type e only produces individual of type e with offspring distribution
$$\tilde{\pi}(k) = q_\pi^{n-1}\pi(k). \tag{4.14}$$
- An individual of type s produces a random number $S \geq 1$ of individuals of type s and E individuals of type e with generating function
$$\mathbb{E}[r^S l^E] = \frac{\varphi((1-q_\pi)r + q_\pi l) - \varphi(q_\pi l)}{1-q_\pi}. \tag{4.15}$$
Further, the probability for a node u of type s to have n children and where the children of type s are $\{ui : i \in A\}$ with A a non-empty subset of $\{1, \dots, n\}$ is
$$\pi(n)(1-q_\pi)^{\#A-1}q_\pi^{n-\#A}. \tag{4.16}$$
- The root is of type s.

Equation (4.16) gives no new information and follows from Lemma 4.10 respectively (4.15). Indeed, we can expand

$$\varphi((1-q_\pi)r + q_\pi l) = \sum_{n\geq 0}\pi(n)\sum_{m=0}^{n}\binom{n}{m}(1-q_\pi)^m q_\pi^{n-m}r^m l^{n-m}.$$

Subtracting the $m = 0$ term in order to have only $m \geq 1$. To get the probability that there are m survivors under the n children, we just observe the $r^m l^{n-m}$ term. Its coefficients are

$$\mathbb{P}(S = m, E = n-m) = \pi(n)\binom{n}{m}(1-q_\pi)^m q_\pi^{n-m}$$

and thus if $A \subset \{1, \dots, n\}$ is of size $\#A = m \geq 1$, we get exactly (4.16).
This probability indeed defines a probability measure as

$$\begin{aligned}\sum_{n\in\mathbb{N}^*} \sum_{\substack{A\subset\{1,\dots,n\}\\A\neq\emptyset}} \pi(n)(1-q_\pi)^{\#A-1} q_\pi^{n-\#A} &= \sum_{n\in\mathbb{N}^*} \sum_{k=1}^{n} \pi(n) \binom{n}{k} (1-q_\pi)^{k-1} q_\pi^{n-k} \\ &= \sum_{n\in\mathbb{N}^*} \pi(n) \frac{1-q_\pi^n}{1-q_\pi} \\ &= \frac{\varphi(1)-\varphi(q_\pi)}{1-q_\pi} = 1.\end{aligned}$$

If we want to describe a tree of the same structure as $\tilde{\mathbf{T}}^s$ but without the labels of types, we denote it by $\mathbf{T}^s$. So in the example (Figure 4.2), $\mathbf{T}^s$ would be the same tree with all nodes having the same color.

4.5 Kesten's tree

Next, another important type of random tree is introduced. Informally, Kesten's tree consists of one infinite spine and on each side there are multiple independent Galton Watson trees attached, see Figure 4.3 for an example of the structure. Kesten's tree will play a big role in Chapter 6.

Definition 4.12. *Kesten's tree* is an infinite random plane tree denoted by $\mathcal{K}$, where all individuals reproduce independently, but there are two different offspring distributions depending on the type of the node:

- normal: offspring distribution $\pi(k)$
- mutant: size based offspring distribution of π given by $\bar{\pi}(k) = \frac{1}{m_\pi} k\pi(k)$

The root of the tree $\mathcal{K}$ is a mutant particle. Each following individual is either normal or a mutant. Normal nodes produce only normal individuals according to π. Mutant individuals produce with respect to $\bar{\pi}$ and all children are normal except one chosen uniformly at random, this is mutant again.

Notice that in each generation, there is exactly one mutant node and there is always at least one offspring, since $\bar{\pi}(0) = 0$. We now want to take a closer look at the law of $\mathcal{K}$ and the relation between Galton Watson trees and Kesten's trees.

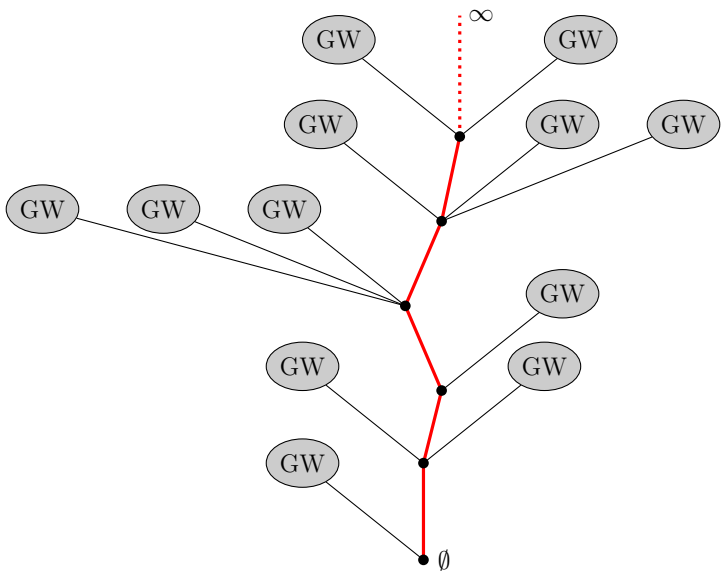

Figure 4.3: Example of a Kesten's tree with independent Galton Watson trees attached to the red infinite spine. The black dots represent mutant nodes.

Lemma 4.13. *Let* $\mathbf{T}$ *be a Galton Watson tree with offspring distribution* π *and* $\mathcal{K}$ *Kesten's tree corresponding to* π*. For all* $n \in \mathbb{N}$ *and any plane tree* $\mathtt{t}$ *of height* n *with* l *individuals at height* n*, we have*

$$\mathbb{P}(tr_n(\mathcal{K}) = \mathtt{t}) = \frac{l}{m_\pi^n} \prod_{\substack{u \in \mathtt{t} \\ H(u) \neq n}} \pi(k_u(\mathtt{t})) = \frac{l}{m_\pi^n} \mathbb{P}(tr_n(\mathbf{T}) = \mathtt{t}).$$

Proof. Suppose $v \in \mathtt{t}$ is the end of the spine of $\mathcal{K}$ at height n. In other words, v is the mutant individual at height n. Fix this v. Recall that A_u denotes the ancestors of u. We then have

$$\begin{aligned}
\mathbb{P}(tr_n(\mathcal{K}) = \mathtt{t} \mid \text{spine ends at } v) &= \prod_{\substack{u \in \mathtt{t} \setminus A_u \\ H(u) < n}} \pi(k_u(\mathtt{t})) \prod_{u \in A_u} \bar{\pi}(k_u(\mathtt{t})) \frac{1}{k_u(\mathtt{t})} \\
&= \prod_{\substack{u \in \mathtt{t} \setminus A_u \\ H(u) < n}} \pi(k_u(\mathtt{t})) \prod_{u \in A_u} \frac{\pi(k_u(\mathtt{t}))}{m_\pi} \\
&= \frac{1}{m_\pi^n} \prod_{\substack{u \in \mathtt{t} \\ H(u) \neq n}} \pi(k_u(\mathtt{t})) = \frac{1}{m_\pi^n} \mathbb{P}(tr_n(\mathbf{T}) = \mathtt{t}).
\end{aligned}$$

where the last equation follows from Lemma 4.2. By definition of $\mathcal{K}$, there exists only one mutant particle at height n. So summing over all possible v at height n, which are by definition exactly l, gives the result. □

Chapter 5

Conditioning of super-critical Galton Watson trees

Let us first have a closer look at super-critical Galton Watson trees $\mathbf{T}$. Recall that for the mean offspring m_π, it holds that $m_\pi > 1$. Thanks to Corollary 4.7, we have seen that super-critical Galton Watson trees both survive and get extinct with a positive probability. So we are able to condition on these two events. The goal in this chapter is to to identify the distribution of $\mathbf{T}$ under

- $\mathbb{P}(\cdot \mid \mathcal{H}(\mathbf{T}) < \infty)$, which describes conditioning on extinction and
- $\mathbb{P}(\cdot \mid \mathcal{H}(\mathbf{T}) = \infty)$, denoting conditioning on survival.

Recall the definition of the height of a tree, given in Equation (3.2):

$$\mathcal{H}(\mathtt{t}) = \sup\left\{H(u) : u \in \mathtt{t}\right\},$$

where $H(u)$ is the height of a node u.

5.1 Conditioning on extinction

We start by condition a Galton Watson tree on extinction. Recall that we denote this event as $\{\mathcal{H}(\mathbf{T}) < \infty\}$ and notice that extinction is only possible if $\pi(0) > 0$.

Lemma 5.1. *Let $\mathbf{T}$ be a super-critical Galton Watson tree with offspring π. If $\pi(0) > 0$, conditioning on extinction yields to a sub-critical Galton Watson tree $\tilde{\mathbf{T}}$ with offspring distribution*

$$\tilde{\pi}(k) = q_\pi^{k-1}\pi(k). \tag{5.1}$$

Mathematically, for all measurable functions $f : \mathbb{T}^f \to \mathbb{R}_+$,

$$\mathbb{E}_\pi[f(\mathbf{T}) \mid \mathcal{H}(\mathbf{T}) < \infty] = \mathbb{E}_{\tilde{\pi}}[f(\mathbf{T})].$$

Proof. Suppose that $\pi(0) > 0$. Start by remarking that the offspring distribution $\tilde{\pi}$ is sub-critical, since $\varphi_{\tilde{\pi}}(r) = \sum_{k\geq 0} \tilde{\pi}(k) r^k = \frac{1}{q_\pi} \sum_{k\geq 0} \pi(k)(q_\pi r)^k = \frac{\varphi(q_\pi r)}{q_\pi}$ implies that $\tilde{m}_\pi = \varphi'_{\tilde{\pi}}(1) = \varphi'(q_\pi) < 1$. By Corollary 4.7, we immediately get $\mathbb{P}_{\tilde{\pi}}(\mathcal{H}(\tilde{\mathbf{T}}) < \infty) = 1$. This means that $\tilde{\mathbf{T}}$ is almost surely extincted, so $\tilde{\mathbf{T}} \in \mathbb{T}^f$. For $\mathtt{t} \in \mathbb{T}^f$, we have

$$\begin{aligned} q_\pi \mathbb{P}_{\tilde{\pi}}(\tilde{\mathbf{T}} = \mathtt{t}) &= q_\pi \prod_{u\in\mathbf{T}} \tilde{\pi}(k_u(\mathtt{t})) \\ &= q_\pi \prod_{u\in\mathbf{T}} q_\pi^{k_u(\mathtt{t})-1} \pi(k_u(\mathtt{t})) \\ &= q_\pi^{1+\sum_{u\in\mathtt{t}}(k_u(\mathtt{t})-1)} \prod_{u\in\mathtt{t}} \pi(k_u(\mathtt{t})) \\ &= \mathbb{P}_\pi(\mathbf{T} = \mathtt{t}). \end{aligned} \tag{5.2}$$

We used Lemma 4.2 in the first equation and then the definition of $\tilde{\pi}$. Further, exponential rules were used. For the last equation, remark that

$$1 + \sum_{u\in\mathtt{t}} (k_u(\mathtt{t}) - 1) = 1 + \sum_{u\in\mathtt{t}} k_u(\mathtt{t}) - \sum_{u\in\mathtt{t}} 1 = 1 + (\#\mathtt{t} - 1) - \#\mathtt{t} = 0,$$

where we used (3.1), and use again Lemma 4.2. If we look at any measurable function $f : \mathbb{T}^f \to \mathbb{R}_+$ and sum the previous equation over all $\mathtt{t} \in \mathbb{T}^f$, we get

$$\begin{aligned} \mathbb{E}_\pi[f(\mathbf{T})\mathbb{1}_{\{\mathbf{T}\in\mathbb{T}^f\}}] &= \sum_{\mathtt{t}\in\mathbb{T}^f} f(\mathtt{t}) \mathbb{P}_\pi(\mathbf{T} = \mathtt{t}) \\ &= \sum_{\mathtt{t}\in\mathbb{T}^f} f(\mathtt{t}) q_\pi \mathbb{P}_{\tilde{\pi}}(\tilde{\mathbf{T}} = \mathtt{t}) \\ &= q_\pi \mathbb{E}_{\tilde{\pi}}[f(\tilde{\mathbf{T}})\mathbb{1}_{\{\tilde{\mathbf{T}}\in\mathbb{T}^f\}}] \\ &= q_\pi \mathbb{E}_{\tilde{\pi}}[f(\tilde{\mathbf{T}})], \end{aligned} \tag{5.3}$$

where in the second equality (5.2) we used and for the last equality notice that $\tilde{\mathbf{T}} \in \mathbb{T}^f$, so $\mathbb{P}(\tilde{\mathbf{T}} \in \mathbb{T}^f) = 1$. Recall that the probability of extinction is according to Lemma 4.6 given by

$$\mathbb{P}(\mathbf{T} \in \mathbb{T}^f) = q_\pi. \tag{5.4}$$

By the definition of conditional expectation,

$$\begin{aligned} \mathbb{E}_\pi[f(\mathbf{T}) \mid \mathcal{H}(\mathbf{T}) < \infty] = \mathbb{E}_\pi[f(\mathbf{T}) \mid \mathbf{T} \in \mathbb{T}^f] &= \frac{\mathbb{E}_\pi[f(\mathbf{T})\mathbb{1}_{\{\mathbf{T}\in\mathbb{T}^f\}}]}{\mathbb{P}(\mathbf{T} \in \mathbb{T}^f)} \\ &= \frac{q_\pi \mathbb{E}_{\tilde{\pi}}[f(\tilde{\mathbf{T}})]}{q_\pi} = \mathbb{E}_{\tilde{\pi}}[f(\tilde{\mathbf{T}})], \end{aligned}$$

where the first part is just an equivalent notation and we used (5.3) and (5.4) in the third step. This concludes the proof. □

Remark that Equation (5.1) describes a probability distribution. Indeed, since

$$\sum_{n\geq 0}\tilde{\pi}(n)=\sum_{n\geq 0}\pi(n)q_\pi^{n-1}=\frac{1}{q_\pi}\sum_{n\geq 0}\pi(n)q_\pi^n=\frac{\varphi(q_\pi)}{q_\pi}=1,$$

where we used the definition of the distribution and the fact that q_π is a solution for $\varphi(r)=r$.

We want to link this result directly to branching processes.

Corollary 5.2. *A super-critical branching process $Z(\mathbf{T}):=(Z_n(\mathbf{T}))_{n\geq 0}$ conditioned on $\{T_0(Z(\mathbf{T}))<\infty\}$, with T_0 denoting the first time where $Z(\mathbf{T})$ reaches the value 0, is a sub-critical branching process with reproduction law $\tilde{\pi}$.*

Proof. Remark that conditioning a super-critical Galton Watson tree $\mathbf{T}$ on $\{\mathcal{H}(\mathbf{T})<\infty\}$ is equivalent to conditioning the corresponding branching process $Z(\mathbf{T})$ on the event $\{T_0(Z(\mathbf{T}))<\infty\}$, since saying that the tree is finite is the same as saying that the population reaches 0 at some point. By Lemma 5.1, conditioning $\mathbf{T}$ on extinction yields to a sub-critical Galton Watson tree $\tilde{\mathbf{T}}$ with offspring distribution $\tilde{\pi}$. As $Z(\mathbf{T})$ is the function that counts the generation size of a Galton Watson tree $\mathbf{T}$, it is strictly correlated to $\mathbf{T}$. Thus, by conditioning the super-critical process $Z(\mathbf{T})$ on the event $\{T_0(Z(\mathbf{T}))<\infty\}$, we obtain the to the tree $\tilde{\mathbf{T}}$ associated sub-critical branching process $Z(\tilde{\mathbf{T}})$ with distribution $\tilde{\pi}$. □

5.2 Conditioning on survival

Next, we investigate on conditioning super-critical Galton Watson trees on survival.

Lemma 5.3. *If $\mathbf{T}$ is a super-critical Galton Watson tree with offspring distribution π, then, conditionally on survival, $\mathbf{T}$ is distributed as a two-type Galton Watson tree $\mathbf{T}^s$, defined in Definition 4.11.*

Proof. If $\pi(0)=0$, then $q_\pi=0$ by Lemma 4.4 and $\mathbb{P}(\mathcal{H}(\mathbf{T})=\infty)=q_\pi$ by Lemma 4.6. Thus, $\mathbf{T}\overset{(d)}{=}\mathbf{T}^s$. Suppose now that $\pi(0)>0$, so also $q_\pi>0$ again by Lemma 4.4. We start by defining the set of all nodes at height h with type s as

$$S_h=\{u\in\mathbf{T}^s : H(u)=h, u \text{ has type } s \text{ in } \mathbf{T}^s\}.$$

Remark that $tr_h(\tilde{\mathbf{T}}^s)$ is characterized by $tr_h(\mathbf{T}^s)$ and S_h, since all ancestors of an individual of type s are also of type s and they have at least one child of the same type. So to prove the lemma, it suffices to show

$$\mathbb{P}(tr_h(\mathbf{T})=\mathtt{t}\mid\mathcal{H}=\infty)=\mathbb{P}(tr_h(\mathbf{T}^s)=\mathtt{t}), \tag{5.5}$$

since by Lemma 3.7 the law of a random tree is characterized by its truncations and by the remark before, we can just look at $tr_h(\mathbf{T}^s)$. Let $\mathtt{t} \in \mathbb{T}^f$ such that its height is h, i.e. $\mathcal{H}(\mathtt{t}) = h$. Let $A \subset \{u \in \mathtt{t} : H(u) = h\}$ with $A \neq \emptyset$. So A is just a set of some nodes of height h. We further define n to be the number of nodes at height h. Let $\mathtt{A}$ be the set of all ancestors of the nodes in A. Remark that $\mathtt{A}$ forms a tree. Define $\mathtt{A}^c = tr_{h-1} \setminus \mathtt{A}$. For $u \in \mathtt{A}$, denote the number of children of a u belonging to $A \cup \mathtt{A}$ as $k_u^s(\mathtt{t}, A)$. To get a better understanding and have a bit intuition for all those definitions, look at the example tree drawn in Figure 5.2 for height $h = 4$. All the blue nodes are an element of the set A, green denotes the tree formed by $\mathtt{A}$ and orange are all the nodes in $\mathtt{A}^c$. The gray nodes have no special meaning.

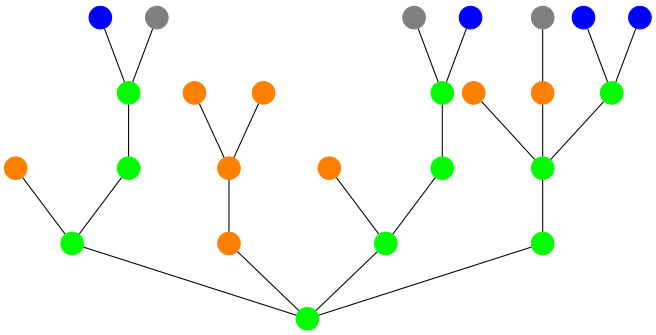

Figure 5.1: Example of a tree of height 4 illustrating the nodes of the sets A (blue), $\mathtt{A}$ (green) and $\mathtt{A}^c$ (orange).

With this setup, we have

$$\begin{aligned}
&\mathbb{P}(tr_h(\mathbf{T}^s) = \mathtt{t}, S_h = A) \\
&\quad = \prod_{u \in \mathtt{A}^c} \tilde{\pi}(k_u(\mathtt{t})) \prod_{u \in \mathtt{A}} \pi(k_u(t))(1-q_\pi)^{k_u^s(\mathtt{t},A)-1} q_\pi^{k_u(\mathtt{t})-k_u^s(\mathtt{t},A)} \\
&\quad = \prod_{u \in \mathtt{A}^c} q_\pi^{k_u(\mathtt{t})-1} \pi(k_u(\mathtt{t})) \prod_{u \in \mathtt{A}} \pi(k_u(t))(1-q_\pi)^{k_u^s(\mathtt{t},A)-1} q_\pi^{k_u(\mathtt{t})-k_u^s(\mathtt{t},A)} \\
&\quad = \left(\prod_{u \in \mathtt{A} \cup \mathtt{A}^c} \pi(k_u(\mathtt{t})) \right) \left(q_\pi^{\sum_{u \in \mathtt{A}^c}(k_u(\mathtt{t})-1)} (1-q_\pi)^{\sum_{u \in \mathtt{A}}(k_u^s(\mathtt{t},A)-1)} q_\pi^{\sum_{u \in \mathtt{A}}(k_u(\mathtt{t})-k_u^s(\mathtt{t},A))} \right) \\
&\quad = \left(\prod_{u \in \mathtt{A} \cup \mathtt{A}^c} \pi(k_u(\mathtt{t})) \right) \left(q_\pi^{\sum_{u \in \mathtt{A} \cup \mathtt{A}^c}(k_u(\mathtt{t})-1)} \left(\frac{1-q_\pi}{q_\pi} \right)^{\sum_{u \in \mathtt{A}}(k_u^s(\mathtt{t},A)-1)} \right). \qquad (5.6)
\end{aligned}$$

We first used Equation (4.16), then the definition of $\tilde{\pi}$ given in Equation (4.14) and for the following equations we did some algebra. Recall again that the tree $\mathtt{t}$ has height h and notice that the first part is just the probability that $tr_h(\mathbf{T}) = \mathtt{t}$. Let us have a

closer look at the two sums. Observe that $\mathtt{A} \cup \mathtt{A}^c = tr_{h-1}(\mathtt{t})$, what is also nicely visible in Figure 5.2, and that

$$\sum_{u \in tr_{h-1}(\mathtt{t})} k_u(\mathtt{t}) = \sum_{u \in \mathtt{t}} k_u(\mathtt{t}) = \#\mathtt{t} - 1,$$

since both sums count all nodes of the tree $\mathtt{t}$ and we used Equation (3.1) for the last equality. Further,

$$\sum_{u \in \mathtt{A} \cup \mathtt{A}^c} (k_u(\mathtt{t}) - 1) = \sum_{u \in tr_{h-1}(\mathtt{t})} (k_u(\mathtt{t}) - 1) = (\#\mathtt{t} - 1) - \#tr_{h-1}(\mathtt{t}) = n - 1,$$

where we used the equation before and observed that $\#\mathtt{t} - \#tr_{h-1}(\mathtt{t}) = n$, since we have n nodes at generation h and all the nodes at lower heights get subtracted. Similarly, we get

$$\sum_{u \in \mathtt{A}} k_u^s(\mathtt{t}, A) - 1 = \#A - 1$$

by applying the idea to the tree $\mathtt{A}$. Putting those into (5.6) gives

$$\begin{aligned}
\mathbb{P}(tr_h(\mathbf{T}^s) = \mathtt{t}, S_h = A) &= \mathbb{P}(tr_h(\mathbf{T}) = \mathtt{t}) q_\pi^{n-1} \left(\frac{1 - q_\pi}{q_\pi}\right)^{\#A - 1} \\
&= \frac{1}{1 - q_\pi} \mathbb{P}(tr_h(\mathbf{T}) = \mathtt{t})(1 - q_\pi)^{\#A} q_\pi^{n - \#A}. \qquad (5.7)
\end{aligned}$$

Since A was just an arbitrary subset, we sum Equation (5.7) over all possible A and get

$$\begin{aligned}
\mathbb{P}(tr_h(\mathbf{T}^s) = \mathtt{t}) &= \sum_A \frac{1}{1 - q_\pi} \mathbb{P}(tr_h(\mathbf{T}) = \mathtt{t})(1 - q_\pi)^{\#A} q_\pi^{n - \#A} \\
&= \frac{1}{1 - q_\pi} \mathbb{P}(tr_h(\mathbf{T}) = \mathtt{t}) \sum_{k=1}^{n} \binom{n}{k} (1 - q_\pi)^k q_\pi^{n-k} \\
&= \mathbb{P}(tr_h(\mathbf{T}) = \mathtt{t}) \frac{1 - q_\pi^n}{1 - q_\pi}. \qquad (5.8)
\end{aligned}$$

Looking at the left side of Equation (5.5), we can find by definition of conditional probability and using the branching property (4.9), we get

$$\begin{aligned}
\mathbb{P}(tr_h(\mathbf{T}) = \mathtt{t} \mid \mathcal{H} = \infty) &= \frac{\mathbb{P}(tr_h(\mathbf{T}) = \mathtt{t}) - \mathbb{P}(tr_h(\mathbf{T}) = \mathtt{t}, \mathcal{H} < \infty)}{1 - \mathbb{P}(\mathcal{H}(\mathbf{T} < \infty)} \\
&= \frac{\mathbb{P}(tr_h(\mathbf{T}) = \mathtt{t}) - \mathbb{P}(tr_h(\mathbf{T}) = \mathtt{t}, \mathcal{H}) q_\pi^n}{1 - q_\pi} \\
&= \mathbb{P}(tr_h(\mathbf{T}) = \mathtt{t}) \frac{1 - q_\pi^n}{1 - q_\pi}, \qquad (5.9)
\end{aligned}$$

Combining (5.8) and (5.9), we get the wanted equality and thus the lemma is proven. □

To recap the results of this section, a short summary. Looking at super-critical Galton Watson trees, we get the following:

- Conditioning on extinction gives a sub-critical Galton Watson tree with offspring distribution $\tilde{\pi}(k) = q_\pi^{k-1}\pi(k)$.
- Conditioning on survival yields to a two-type Galton Watson tree as given in Definition 4.11.

These results can directly be transferred to the super-critical branching process.

Chapter 6

Local limits of critical Galton Watson trees

We consider now critical Galton Watson trees. Recall that in this setting, the trees get extinct almost surely. So conditioning on extinction gives us no additional information. On the other hand, conditioning on survival is more interesting, but cannot be done in a usual way, as the probability to survive is zero. Hence, we need another approach. We will condition critical Galton Watson trees on "being large", for example by setting the height to be n, and then study the limit as $n \to \infty$. For this, we need local convergence of trees and therefore start the chapter with a reminder on weak convergence respectively convergence in distribution. With these tools, we can rigorously formulate and analyze critical Galton Watson trees, which are conditioned on "being large".

6.1 Preliminaries on weak convergence

Let us recall some definitions about weak convergence of probability measures. We work on a metric space $(S, \mathcal{S})$, where $\mathcal{S}$ denotes the Borel sigma-algebra on S. In particular, all those notions make sense on the space of trees $(\mathbb{T}, d_{\mathbb{T}}, \mathscr{B}(\mathbb{T}))$, since it is a Polish metric space according to Lemma 3.4 and we consider its Borel sigma-algebra $\mathscr{B}(\mathbb{T})$. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures and μ a probability measure on $(S, \mathcal{S})$.

Then, $(\mu_n)_{n\in\mathbb{N}}$ *converges weakly* to μ on the space S, denoted by $\mu_n \xrightarrow{(w)} \mu$, if for all bounded and continuous functions $f : S^d \to S$, it holds that

$$\int_S f d\mu_n \xrightarrow{n\to\infty} \int_S f d\mu.$$

If $(X_n)_{n\in\mathbb{N}}$ is a sequence of random variables with corresponding distributions $(\mu_n)_{n\in\mathbb{N}}$ and X a random variable with distribution μ, we say that $(X_n)_{n\in\mathbb{N}}$ *converges in distribution* to X if and only if $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ. We write $X_n \xrightarrow{(d)} X$. So in particular,

$$\mu_n \xrightarrow{(w)} \mu \iff X_n \xrightarrow{(d)} X.$$

We introduce different equivalent conditions for weak convergence of probability measures using Portmanteau's theorem. Recall that $A \in \mathcal{S}$ is a μ-continuous set if $\mu(\partial A) = 0$, where ∂A denotes the boundary of the set. Remark that $\partial A \in \mathcal{S}$, as A is closed.

Theorem 6.1. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures and μ a probability measure. The following are equivalent*

1. *The sequence converges weakly, i.e $\mu_n \xrightarrow{(w)} \mu$.*
2. *For all bounded and uniformly continuous functions f, $\int f d\mu_n \xrightarrow{n\to\infty} \int f d\mu$.*
3. *For all open sets G, $\liminf_n \mu_n(G) \geq \mu(G)$.*
4. *For all closed sets F, $\limsup_n \mu_n(F) \leq \mu(F)$.*
5. *For all μ-continuity sets A, $\mu_n(A) \to \mu(A)$.*

Proof. We want to demonstrate the individual implications.
1. $\Rightarrow$ 2.) Since uniformly continuous functions are continuous, this follows immediately.
2. $\Rightarrow$ 3.) Define a sequence of uniformly continuous functions given by

$$f_k(x) = \min\{1, k \cdot d(x, G^c)\}$$

with $d(x, G^c) = \inf_{y\in G^c} ||x - y||$ the distance from G^c. Remark that they are bounded, since $0 \leq f_k(x) \leq 1$, and the sequence is monotone increasing for every x. Further, $f_k(x) \xrightarrow{k\to\infty} \mathbb{1}_G(x)$. So the sequence is dominated by $\mathbb{1}_G(x)$. Then,

$$\begin{aligned}
\liminf_n \mu_n(G) &= \liminf_n \left(\int \mathbb{1}_G d\mu_n \right) \\
&\geq \sup_k \left(\liminf_n \left(\int f_k d\mu_n \right) \right) \\
&= \sup_k \left(\int f_k d\mu \right) = \lim_{k\to\infty} \left(\int f_k d\mu \right) \\
&= \int \mathbb{1}_G d\mu = \mu(G),
\end{aligned}$$

where the inequality follows from the fact that $f_k \leq \mathbb{1}_G$. Then, 2. is applied and further since the sequence is monotone increasing, the supremum is the same as taking the limit. By the monotone convergence theorem, we can take the limit inside and the result follows.
3. $\Leftrightarrow$ 4.) Let F be a closed set, then F^c is open. Since μ is a probability measure, we have

$$\mu(F) = 1 - \mu(F^c)$$

and the same for μ_n. Then,

$$\limsup_n \mu_n(F) = 1 - \liminf_n \mu_n(F^c) \leq 1 - \mu(F^c) = \mu(F),$$

where we first used an identity and then point 3. For the other implication, we can use similar arguments.
3.&4. $\Rightarrow$ 5.) Let f be a bounded and continuous function. By linearity, we can assume that $0 \leq f \leq K$ for some constant K. Then, by Fubini in the second step, we have

$$\begin{aligned}\int f d\mu &= \int \left(\int_0^k \mathbb{1}_{\{t \leq f\}} dt \right) d\mu \\ &= \int_0^k \left(\int \mathbb{1}_{\{t \leq f\}} d\mu \right) dt \\ &= \int_0^k \mu(\{t \leq f\}) dt.\end{aligned}$$

Similarly,

$$\int f d\mu_n = \int_0^k \mu_n(\{t \leq f\}) dt.$$

Notice that $\partial\{t \leq f\} \subseteq \{f = t\}$. Since f is continuous, $\mu(\{f = t\}) = 0$ and thus $\mu(\partial\{t \leq f\}) = 0$. We can apply 5. to have $\mu_n(\{t \leq f\}) \xrightarrow{n\to\infty} \mu(\{t \leq f\})$. By the dominated convergence theorem, it follows that

$$\int f d\mu_n = \int_0^k \mu_n(\{t \leq f\}) dt \xrightarrow{n\to\infty} \int_0^k \mu(\{t \leq f\}) dt = \int f d\mu.$$

This is exactly the definition of $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ. □

Not only Portmanteau's theorem but also the following lemma will help us characterize different conditions of weak convergence respectively convergence in distribution.

Lemma 6.2. *Suppose that $\mathscr{D}$ is a pi-system, S is separable and for all $x \in S$ and $\varepsilon > 0$, there exists an $D \in \mathscr{D}$ for which $x \in D^\circ \subset D \subset B(x, \varepsilon)$. If $\mu_n(D) \to \mu(D)$ for every $D \in \mathscr{D}$, then $\mu_n \xrightarrow{(w)} \mu$.*

Proof. By assumption, for each open set G, we have $x \in D_x^\circ \subset D_x \subset G$ for some $D_x \in \mathscr{D}$. Since S is a separable metric space, it is a Lindelöf space. By definition, every open cover has a countable subcover. So, there exists a countable subcollection $\{D_{x_i}^\circ\}$ of $\{D_x^\circ : x \in G\}$ that covers G, so we can write $G = \bigcup_i D_{x_i}^\circ$. Thus, each open set can be written as a countable union of sets in $\mathscr{D}$. Further, since $\mathscr{D}$ is a pi-system, the intersections of two elements is again in $\mathscr{D}$. So by assumption, we have also convergence for intersections. Using this and the inclusion-exclusion principle, we get

$$\begin{aligned}\mu_n\left(\bigcup_{0\le i\le r} D_i\right) &= \sum_i \mu_n(D_i) - \sum_{ij}\mu_n(D_i\cap D_j) + \sum_{ijk}\mu_n(D_i\cap D_j\cap D_k) + \dots \\ &\xrightarrow{n\to\infty} \sum_i \mu(D_i) - \sum_{ij}\mu(D_i\cap D_j) + \sum_{ijk}\mu(D_i\cap D_j\cap D_k) + \dots \\ &= \mu\left(\bigcup_{0\le i\le r} D_i\right).\end{aligned}$$

We have shown that each open set can be written as a countable union of sets in $\mathscr{D}$, so let $G = \bigcup_i D_i$, $D_i \in \mathscr{D}$, be an open set. Given $\varepsilon > 0$, choose an r such that

$$\mu(G) - \varepsilon < \mu\left(\bigcup_{i\le r} D_i\right).$$

Then,

$$\mu(G) - \varepsilon \le \mu\left(\bigcup_{i\le r} D_i\right) = \lim_{n\to\infty}\mu_n\left(\bigcup_{i\le r} D_i\right) \le \liminf_n \mu_n(G).$$

Since ε was arbitrary, $\mu(G) \le \liminf_n \mu_n(G)$. Thus, by Portmanteau's theorem, we have weak convergence, so $\mu_n \xrightarrow{(w)} \mu$. □

6.2 Local convergence of random trees

We can apply the knowledge of the preliminaries on weak convergence to the space of trees endowed with the local metric d_π, since by Lemma 3.4 it is a Polish metric space and we equipped it with its Borel sigma-algebra. In this context, convergence in distribution with respect to the local topology is also called *local convergence*.

Definition 6.3. A sequence of random trees $(\mathbf{t}_n)_{n\in\mathbb{N}}$ with distributions $(\mu_n)_{n\in\mathbb{N}}$ *converges in distribution* with respect to d_π (respectively *converges locally*) towards a random tree $\mathbf{t}$ with distribution μ if for all bounded and continuous functions $f : \mathbb{T} \to \mathbb{T}$

$$\lim_{n\to\infty} \int_{\mathbb{T}} f(\mathtt{t}) d\mu_n = \int_{\mathbb{T}} f(\mathtt{t}) d\mu, \tag{6.1}$$

denoted by $\mathbf{t}_n \xrightarrow{(d)} \mathbf{t}$.

This definition is equivalent to

$$\mu_n(tr_k(\mathbf{t}) = tr_k(\mathtt{t})) \xrightarrow{n\to\infty} \mu(tr_k(\mathbf{t}) = tr_k(\mathtt{t})). \tag{6.2}$$

Indeed, suppose (6.1) holds and remark that

$$\{\mathbf{t} : tr_k(\mathbf{t}) = tr_k(\mathtt{t})\} = \{\mathbf{t} : d_{\mathbb{T}}(\mathtt{t}, \mathbf{t}) \leq 2^{-k}\} = B(\mathtt{t}, 2^{-k}).$$

By Lemma 3.2, this set is both open and closed. So the boundary, which is defined as the closure without the interior, is

$$\begin{aligned}\partial\{tr_k(\mathbf{t}) = tr_k(\mathtt{t})\} = \partial B(\mathtt{t}, 2^{-k}) &= \overline{B(\mathtt{t}, 2^{-k})} \setminus B(\mathtt{t}, 2^{-k})^\circ \\ &= B(\mathtt{t}, 2^{-k}) \setminus B(\mathtt{t}, 2^{-k}) = \emptyset.\end{aligned}$$

Thus, $\mu(\partial\{tr_k(\mathbf{t}) = tr_k(\mathtt{t})\}) = 0$ and so this set is μ-continuous. So by the last point of Portmanteau's theorem 6.1, we get (6.2). For the other direction, suppose (6.2) holds. Remark that $\mathscr{C} = \big\{\emptyset\big\} \cup \big\{B(\mathtt{t}, 2^{-n}), \mathtt{t} \in \mathbb{T}, n \in \mathbb{N}\big\}$ is a pi-system by Lemma 3.5 and $\mathbb{T}$ is separable as it is Polish by Lemma 3.4. Further, since each $C \in \mathscr{C}$ is either the empty set or an open ball itself, for all $\mathtt{t} \in \mathbb{T}$ and for all $\varepsilon > 0$ there exists $C \in \mathscr{C}$ such that $\mathtt{t} \in C^\circ \subset C \subset B(\mathtt{t}, \varepsilon)$. Thus, all conditions of Lemma 6.2 are fulfilled and we can apply it to get weak convergence respectively local convergence.

To find a key characterization for local convergence, we introduce grafted trees. Recall that $\mathscr{L}_0(\mathtt{t})$ denotes the set of leaves of a tree $\mathtt{t} \in \mathbb{T}$. A grafted tree is made of two trees $\mathtt{t}$ and $\mathtt{t}'$. Choose one leaf x of the tree $\mathtt{t}$, detach it and attach the root of the tree $\mathtt{t}'$ at this position. So in the resulting tree, x is not a leaf anymore, it is replaced by the root of $\mathtt{t}'$ and has now children. For a visualization, look at the example in Figure 6.1. Mathematically, we have the following definition of grafted trees.

Definition 6.4. Let $t, t' \in \mathbb{T}$ be two trees and $x \in \mathscr{L}_0(\mathtt{t})$, so x is a leaf of the tree $\mathtt{t}$. The tree we get when we graft $\mathtt{t}'$ on the leaf x is

$$\mathtt{t} \circledast_x \mathtt{t}' = \{u \in \mathtt{t}\} \cup \{xu, u \in \mathtt{t}'\}.$$

We call it *grafted tree*. The set of all trees of this form is denoted by

$$\mathbb{T}(\mathtt{t}, x) = \{\mathtt{t} \circledast_x \mathtt{t}' : \mathtt{t}' \in \mathbb{T}\}.$$

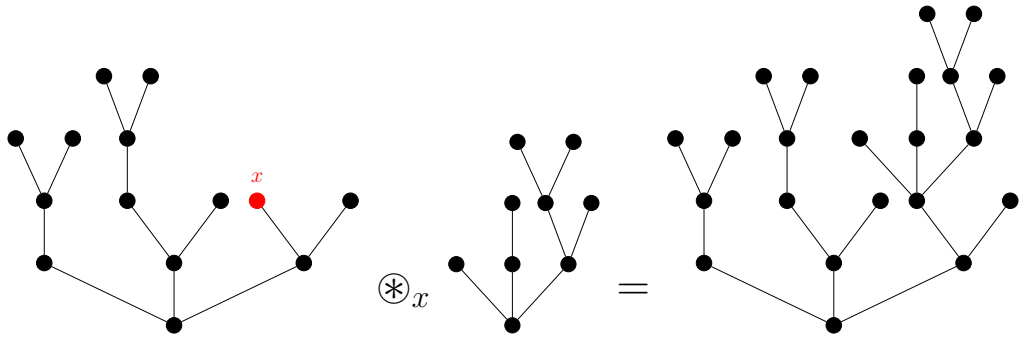

Figure 6.1: Example of a grafted tree with x the red node.

Now, let us connect this to Galton Watson trees. Let $\mathbf{T}$ be a Galton Watson tree with offspring distribution π. The probability that $\mathbf{T}$ is of the form of a grafted tree is

$$\begin{aligned}\mathbb{P}(\mathbf{T} = \mathtt{t} \circledast_x \mathtt{t}') &= \prod_{u \in \mathtt{t} \circledast_x \mathtt{t}'} \pi(k_u(\mathtt{t} \circledast_x \mathtt{t}')) = \prod_{u \in \mathtt{t} \setminus \{x\}} \pi(k_u(\mathtt{t})) \cdot \prod_{u \in \mathtt{t}'} \pi(k_u(\mathtt{t}')) \\ &= \frac{\prod_{u \in \mathtt{t}} \pi(k_u(\mathtt{t}))}{\pi(k_x(\mathtt{t}))} \cdot \prod_{u \in \mathtt{t}'} \pi(k_u(\mathtt{t}')) \\ &= \frac{1}{\pi(0)} \mathbb{P}(\mathbf{T} = \mathtt{t}) \mathbb{P}(\mathbf{T} = \mathtt{t}'), \end{aligned} \tag{6.3}$$

where in the first and last equation Lemma 4.2 was used. For the final step, remark that x has no children, so $k_x(\mathtt{t}) = 0$ and thus $\pi(k_x(\mathtt{t})) = \pi(0)$. Also,

$$\mathbb{P}(\mathbf{T} = \mathtt{t}) = \mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x), k_x(\mathbf{T}) = 0) = \pi(0) \cdot \mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x)) \tag{6.4}$$

In Lemma 4.13, we have already seen a connection between Galton Watson trees and Kesten's trees. We now find another including the set of grafted trees.

Lemma 6.5. *Let $\mathbf{T}$ be a Galton Watson tree with offspring distribution π such that $m_\pi > 0$ is finite and let $\mathcal{K}$ be Kesten's tree associated with π. Then, for all $\mathtt{t} \in \mathbb{T}^f$ and $x \in \mathscr{L}_0(\mathtt{t})$ we have*

$$\mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x)) = \frac{1}{m_\pi^{H(x)}} \mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x)).$$

Proof. Let $\mathtt{t} \in \mathbb{T}^f$ and $x \in \mathscr{L}_0(\mathtt{t})$ a leaf. Suppose $\mathcal{K} \in \mathbb{T}(\mathtt{t}, x)$. Recall that $\mathcal{K}$ is infinite by definition. Since on the other hand $\mathtt{t}$ is finite, x has to be a mutant node, so a part

of the infinite spine. Recall that A_x describes the ancestors of the node x. So,

$$\begin{aligned}
\mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x)) &= \prod_{u \in \mathtt{t} \setminus \{A_x \cup \{x\}\}} \pi(k_u(\mathtt{t})) \prod_{u \in A_x} \frac{\pi(k_u(t))}{m_\pi} \\
&= \prod_{u \in A_x} \frac{1}{m_\pi} \left(\prod_{u \in \mathtt{t}} \pi(k_u(\mathtt{t})) \right) \frac{1}{\pi(k_x(\mathtt{t}))} \\
&= \prod_{u \in A_x} \frac{1}{m_\pi} \cdot \mathbb{P}(\mathbf{T} = \mathtt{t}) \cdot \frac{1}{\pi(0)} \\
&= \prod_{u \in A_x} \frac{1}{m_\pi} \pi(0) \cdot \mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x)) \cdot \frac{1}{\pi(0)} \\
&= \frac{1}{m_\pi^{H(x)}} \cdot \mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x)),
\end{aligned}$$

where we used Lemma 4.2, the fact that x has no children, so $k_x(t) = 0$ and Equation (6.4). In the end, since the set of all ancestors of x has cardinality $H(x)$, the product gives exactly $\frac{1}{m_\pi^{H(x)}}$. □

In the particular case of a critical offspring distribution π, we have $m_\pi = 1$ and thus

$$\mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x)) = \mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x)).$$

So locally, Kesten's tree looks like an ordinary critical Galton Watson tree. The difference is mainly visible at infinity.

Recall that $\mathbb{T}^f$ describes the set of all finite trees and $\mathbb{T}^1$ denotes the set of trees with exactly one infinite spine. We introduced all this new notations and sets to get to the key characterization of local convergence.

Theorem 6.6. *Let $(\mathbf{t}_n)_{n \in \mathbb{N}}$ be a sequence of random trees and $\mathbf{t}$ a random tree. All of them be an element of $\mathbb{T}^f \cup \mathbb{T}^1$. Then, $\mathbf{t}_n \xrightarrow{(d)} \mathbf{t}$ if for all $\mathtt{t} \in \mathbb{T}^f$ and for all $x \in \mathscr{L}_0(\mathtt{t})$, we have*

1. $\lim_{n \to \infty} \mathbb{P}(\mathbf{t}_n = \mathtt{t}) = \mathbb{P}(\mathbf{t} = \mathtt{t})$ *and*
2. $\liminf_{n \to \infty} \mathbb{P}(\mathbf{t}_n \in \mathbb{T}(\mathtt{t}, x)) \geq \mathbb{P}(\mathbf{t}_n \in \mathbb{T}(\mathtt{t}, x))$.

To prove this, we need two definitions and state two preliminary lemmas. Define the set

$$\mathcal{F} = \{\{\mathtt{t}\}, \mathtt{t} \in \mathbb{T}^f\} \cup \{\mathbb{T}(\mathtt{t}, x), \mathtt{t} \in \mathbb{T}^f, x \in \mathscr{L}_0(\mathtt{t})\} \cup \{\emptyset\}$$

and label the *trace of* $\mathcal{F}$ on $\mathbb{T}^f \cup \mathbb{T}^1$ as

$$\mathcal{F}' = \{A \cap (\mathbb{T}^f \cup \mathbb{T}^1) : A \in \mathcal{F}\}.$$

Lemma 6.7. *$\mathcal{F}$ is a pi-system and all elements of it are open.*

Proof. We aim to show that $\mathcal{F}$ is stable under intersections and is non-empty. Then it is a pi-system. It is non-empty because $\emptyset \in \mathcal{F}$. Further, for all $\mathtt{t}_1, \mathtt{t}_2 \in \mathbb{T}^f$ and for all $x_1 \in \mathscr{L}_0(\mathtt{t}_1), x_2 \in \mathscr{L}_0(\mathtt{t}_2)$, we have

$$\mathbb{T}(\mathtt{t}_1, x_1) \cap \mathbb{T}(\mathtt{t}_2, x_2) = \begin{cases} \mathbb{T}(\mathtt{t}_1, x_1) & \text{if } \mathtt{t}_1 = \mathtt{t}_2 \circledast_{x_2} \mathtt{t}_1', x_1 \wedge x_2 = x_2 \\ \mathbb{T}(\mathtt{t}_2, x_2) & \text{if } \mathtt{t}_2 = \mathtt{t}_1 \circledast_{x_1} \mathtt{t}_2', x_1 \wedge x_2 = x_1 \\ \{\mathtt{t}_1 \cup \mathtt{t}_2\} & \text{if} \mathtt{t}_1 = \mathtt{t} \circledast_{x_2} \mathtt{t}_1' \text{or } \mathtt{t}_2 = \mathtt{t} \circledast_{x_1} \mathtt{t}_2', x_1 \neq x_2 \\ \emptyset & \text{otherwise} \end{cases} \tag{6.5}$$

Looking at the first case, remark that $\mathtt{t}_1$ is included in the tree $\mathbb{T}(\mathtt{t}_2, x_2)$. Since $x_1 \wedge x_2 = x_2$, we must have that $x_1 \in \mathtt{t}_1'$. Grafting all trees on $\mathtt{t}_2$ at x_2 includes all trees obtained by grafting on $\mathtt{t}_1$ at x_1. Thus, we have $\mathbb{T}(\mathtt{t}_1, x_1) \subseteq \mathbb{T}(\mathtt{t}_2, x_2)$ and the intersection is the smaller set, which is $\mathbb{T}(\mathtt{t}_1, x_1)$. For the second case, just switch the labels. To find the third case, look at Figure 6.2 for some intuition. First, $\mathbb{T}(\mathtt{t}_1, x_1)$ is drawn and the second tree represents $\mathbb{T}(\mathtt{t}_2, x_2)$, where the sides with the multiple colorful trees are just a representation of trees, but there every finite tree can be added. The intersection of those two sets is exactly the tree, where the left part is determined by this part of $\mathbb{T}(\mathtt{t}_1, x_1)$ and the right part by $\mathbb{T}(\mathtt{t}_2, x_2)$. This gives exactly $\{\mathtt{t}_1 \cup \mathtt{t}_2\}$. So $\mathcal{F}$ is a pi-system. To prove that all elements are open, notice that we have already seen in Lemma 3.2 that sets of the form $\{\mathtt{t}\}$ are open. Consider $\mathbb{T}(\mathtt{t}, x)$. For a fixed leaf $x \in \mathscr{L}_0(\mathtt{t})$ and for all $\mathtt{t}' \in \mathbb{T}(\mathtt{t}, x)$ we have

$$B(\mathtt{t}', 2^{-\mathcal{H}(t)-1}) \subset \mathbb{T}(\mathtt{t}, x).$$

So for each $\mathtt{t}' \in \mathbb{T}(\mathtt{t}, x)$ there exists an open neighborhood that is entirely contained in $\mathbb{T}(\mathtt{t}, x)$, so $\mathbb{T}(\mathtt{t}, x)$ is open. As $\emptyset$ is also open, all sets of $\mathcal{F}$ are open.

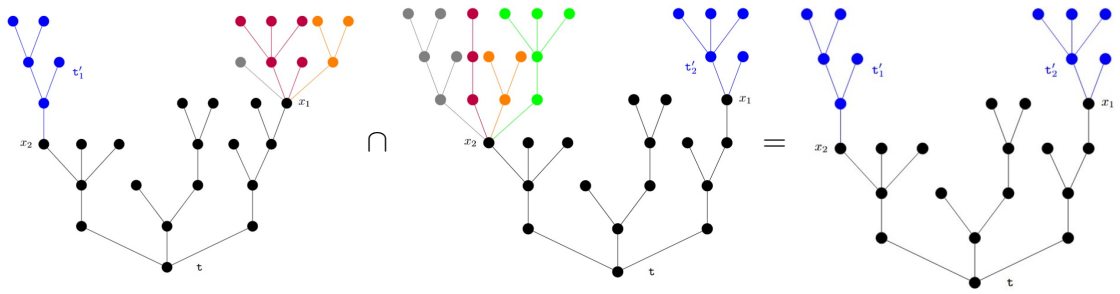

Figure 6.2: Third case: $\mathbb{T}(\mathtt{t}_1, x_1) \cap \mathbb{T}(\mathtt{t}_2, x_2) = \{\mathtt{t}_1 \cup \mathtt{t}_2\}$.

□

Lemma 6.8. *$\mathcal{F}'$ is an open neighborhood system in $\mathbb{T}^f \cup \mathbb{T}^1$.*

Proof. Since all elements of $\mathcal{F}$ are open, we need to show that for all $\mathtt{t} \in \mathbb{T}^f \cup \mathbb{T}^1$ and $\varepsilon > 0$, there exists an element U of $\mathcal{F}$, which is a subset of $B(\mathtt{t}, \varepsilon)$ and contains $\mathtt{t}$. If $\mathtt{t} \in \mathbb{T}^f$, we can just take the set to be $U = \{\mathtt{t}\}$. If $\mathtt{t} \in \mathbb{T}^1$, then by definition it has exactly one infinite spine, denote it as $(u_n)_{n\in\mathbb{N}^*}$ such that $\bar{u}_n = (u_1, \dots u_n) \in \mathtt{t}$ for all $n \in \mathbb{N}^*$. Fix now $n \in \mathbb{N}^*$ such that $2^{-n} < \varepsilon$ and let $\mathtt{t}'$ be the tree obtained by cutting the infinite spine at this fixed height n. Then $\bar{u}_n \in \mathtt{t}'$ for the fix n and take $U = \mathbb{T}(\mathtt{t}', \bar{u}_n)$, which is an element of $\mathcal{F}$. Then, $\mathtt{t} \in U \subseteq B(\mathtt{t}, \varepsilon)$. Therefore, $\mathcal{F}'$ is an open neighborhood system. □

Now we are prepared to prove Theorem 6.6.

Proof. Since $\mathbb{T}^f$ is countable, $\mathcal{F}$ is countable and therefore also $\mathcal{F}'$. By this observation and Lemma 6.8, any open set $U \in \mathbb{T}^f \cup \mathbb{T}^1$ can be written as a countable union of elements of $\mathcal{F}'$, so $U = \bigcup_{i\in\mathbb{N}} A_i$, where $A_i \in \mathcal{F}'$. Further, for all $\varepsilon > 0$ there exists $N \in \mathbb{N}^*$ such that

$$\mathbb{P}(\mathbf{t} \in U) \leq \varepsilon + \mathbb{P}\left(\mathbf{t} \in \bigcup_{1\leq i\leq N} A_i\right).$$

Without loss of generality, we can find open sets such that no set is contained in another. For finite unions, we have the inclusion-exclusion formula

$$\mathbb{P}\left(\mathbf{t} \in \bigcup_{1\leq i\leq N} A_i\right) = \sum_{i=1}^{N} \mathbb{P}(\mathbf{t} \in A_i) - \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{P}(\mathbf{t} \in A_i \cap A_j) + \dots \tag{6.6}$$

By Equation (6.5), the intersection of two such open sets is either empty or a singleton. So

$$\mathbb{P}(A_i \cap \dots \cap A_j) = \begin{cases} 0 \\ \mathbb{P}(\mathbf{t} = \mathtt{t}_k) & \text{for some } \mathtt{t}_k \end{cases}$$

and

$$\mathbb{P}(A_i) = \begin{cases} \mathbb{P}(\mathbf{t} \in \mathbb{T}(\mathtt{t}_j, x_j)) & \text{for some } \mathtt{t}_j, x_j, \\ \mathbb{P}(\mathbf{t} = \mathtt{t}_k) & \text{for some } \mathtt{t}_k \end{cases}$$

Putting those expressions into the inclusion-exclusion formula (6.6), we obtain for some n_1 and n_2

$$\mathbb{P}\left(\mathbf{t} \in \bigcup_{1\leq i\leq N} A_i\right) = \sum_{j=1}^{n_1} \mathbb{P}(\mathbf{t} \in \mathbb{T}(\mathtt{t}_j, x_j)) + \sum_{k=1}^{n_2} \alpha_k \mathbb{P}(\mathbf{t} = \mathtt{t}_k). \tag{6.7}$$

Now assume that the two conditions from the theorem hold. Then,

$$\begin{aligned}
\liminf_{n\to\infty} \mathbb{P}(\mathbf{t}_n \in U) &\geq \liminf_{n\to\infty} \mathbb{P}(\mathbf{t}_n \in \bigcup_{1\leq i\leq N} A_i) \\
&= \liminf_{n\to\infty} \left(\sum_{1\leq j\leq n_1} \mathbb{P}(\mathbf{t}_n \in \mathbb{T}(\mathtt{t}_j, x_j)) + \sum_{1\leq l\leq n_2} \alpha_l \mathbb{P}(\mathbf{t}_n = \mathtt{t}_l) \right) \\
&\geq \sum_{1\leq j\leq n_1} \mathbb{P}(\mathbf{t} \in \mathbb{T}(\mathtt{t}_j, x_j)) + \sum_{1\leq l\leq n_2} \alpha_l \mathbb{P}(\mathbf{t} = \mathtt{t}_l) \\
&= \mathbb{P}(\mathbf{t} \in \bigcup_{1\leq i\leq N} A_i) \\
&\geq \mathbb{P}(\mathbf{t} \in U) - \varepsilon,
\end{aligned}$$

where we first used the fact that $\bigcup_{1\leq i\leq N} A_i \subseteq U$ and then Equation (6.7). The next inequality follows by the two conditions, then again Equation (6.7) and for the last step (6.2). Since ε was arbitrary, we get

$$\liminf_{n\to\infty} \mathbb{P}(\mathbf{t}_n \in U) = \mathbb{P}(\mathbf{t} \in U)$$

for U an open set. By the third point of the Portmanteau theorem (Theorem 6.1), we conclude that $(\mathbf{t}_n)_{n\geq 0}$ converges in distribution to $\mathbf{t}$.

Now suppose that $\mathbf{t}_n \xrightarrow{(d)} \mathbf{t}$. We want to show that the two conditions are fulfilled. Remark that $\{\mathtt{t}\}$ is open and closed by Lemma 3.2. Then, by the third point of Portmanteau's theorem 6.1, we have

$$\limsup_{n\to\infty} \mathbb{P}(\mathbf{t}_n = \mathtt{t}) = \limsup_{n\to\infty} \mathbb{P}(\mathbf{t}_n \in \{\mathtt{t}\}) \leq \ \mathbb{P}(\mathbf{t} \in \{\mathtt{t}\}) = \mathbb{P}(\mathbf{t} = \mathtt{t})$$

and by the fourth point

$$\liminf_{n\to\infty} \mathbb{P}(\mathbf{t}_n = \mathtt{t}) = \liminf_{n\to\infty} \mathbb{P}(\mathbf{t}_n \in \{\mathtt{t}\}) \geq \ \mathbb{P}(\mathbf{t} \in \{\mathtt{t}\}) = \mathbb{P}(\mathbf{t} = \mathtt{t})$$

Together,

$$\lim_{n\to\infty} \mathbb{P}(\mathbf{t}_n = \mathtt{t}) = \mathbb{P}(\mathbf{t} = \mathtt{t}),$$

what is exactly the first assumption we wanted to show. By Lemma 6.7, $\mathbb{T}(\mathtt{t}, x)$ is open. Directly from Portmanteau, we have

$$\liminf_{n\to\infty} \mathbb{P}(\mathbf{t}_n \in \mathbb{T}(\mathtt{t}, x)) = \mathbb{P}(\mathbf{t} \in \mathbb{T}(\mathtt{t}, x)),$$

so also the second assumption holds. □

Having a characterization for local convergence, we can derive a general theorem, which shows that critical Galton Watson trees conditioned on "being large" converge to Kesten's tree under suitable assumptions.

To find this result, we introduce a functional $A : \mathbb{T}^f \to \mathbb{N}$ with $\{\mathtt{t} : A(\mathtt{t}) > n\} \neq \emptyset$ for all $n \in \mathbb{N}^*$. We will impose three different assumptions on A depending on the theorem and context: For all $\mathtt{t} \in \mathbb{T}^f$ and all $x \in \mathscr{L}_0(\mathtt{t})$, there exists a $n_0 \in \mathbb{N}^*$ (and for the Additivity property a $D(\mathtt{t}, x) \geq 0$) such that for all $\mathtt{t}' \in \mathbb{T}^f$ with $A(\mathtt{t} \circledast_x \mathtt{t}') \geq n_0$:

- $A(\mathtt{t} \circledast_x \mathtt{t}') \geq A(\mathtt{t}')$ (*Monotonicity property*)
- $A(\mathtt{t} \circledast_x \mathtt{t}') = A(\mathtt{t}') + D(\mathtt{t}, x)$ (*Additivity property*)
- $A(\mathtt{t} \circledast_x \mathtt{t}') = A(\mathtt{t}')$ (*Identity property*)

The conditions are ordered from weaker to stronger. Remark that the Identity property is a special case of the Additivity property with $D(\mathtt{t}, x) = 0$ and the Additivity property is a particular case of the Monotonicity property, since $A(\mathtt{t} \circledast_x \mathtt{t}') \geq A(\mathtt{t}')$ means that there exists a k such that $A(\mathtt{t}') + k = A(\mathtt{t} \circledast_x \mathtt{t}')$, so $k = D(\mathtt{t}, x)$.

Some examples are

- the height $A(\mathtt{t}) = \mathcal{H}(\mathtt{t})$, which fulfills the Additivity property with $n_0 = \mathcal{H}(\mathtt{t})$ and $D(\mathtt{t}, x) = H(x) \geq 0$,
- the cardinality $A(\mathtt{t}) = \#\mathtt{t}$ satisfying the Monotonicity and Additivity property for $n_0 = 0$, $D(\mathtt{t}, x) = \#\mathtt{t} - 1 \geq 0$ and
- the maximal degree $A(\mathtt{t}) = M(\mathtt{t}) = \max\{k_u(\mathtt{t}), u \in \mathtt{t}\}$ fulfilling the Identity property with $n_0 = M(\mathtt{t}) + 1$.

In Section 6.4, we will use them and prove that they indeed fulfill the stated properties.

6.3 Convergence towards Kesten's tree

As already mentioned earlier, critical Galton Watson trees get extinct almost surely. Hence, conditioning on extinction does not change the law of the tree. So we are only interested in conditioning on survival. But the probability of survival is zero, thus we cannot condition critical Galton Watson trees naively on survival, as it is undefined. An approach is to condition on "being large" and then study the limit, which will mimic conditioning on survival. Concretely, we condition on the events

$$\mathbb{A}_n = \{A(\mathbf{T}) \in [n, n + n_1)\},$$

$n_1 \in \mathbb{N}^* \cup \{\infty\}$. This also includes the special cases $\{A(\mathbf{T}) \geq n\}$ and $\{A(\mathbf{T}) = n\}$ by setting $n_1 = \infty$ respectively $n_1 = 1$.

From the next theorem, many results will be deduced later, including the famous Kesten's theorem.

Theorem 6.9. *Let $\mathbf{T}$ be a critical Galton Watson tree with offspring distribution π such that $\pi(1) < 1$. Let $\mathbf{T}_n$ be a random tree distributed according to $\mathbf{T}$ conditionally on $\mathbb{A}_n$ for a fixed n_1 and $\mathbb{P}(\mathbb{A}_n) > 0$ for n large enough. Denote by $\mathcal{K}$ Kesten's tree associated to π. Suppose one of the following conditions is fulfilled*

1. *$n_1 = \infty$ and A satisfies the Monotonicity property.*
2. *$n_1 \in \mathbb{N}^* \cup \{\infty\}$ and A satisfies the Identity property.*
3. *$n_1 \in \mathbb{N}^* \cup \{\infty\}$, A satisfies the Additivity property and*

$$\limsup_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} \leq 1. \tag{6.8}$$

Then $\mathbf{T}_n \xrightarrow[n\to\infty]{(d)} \mathcal{K}$, so $\mathbf{T}_n$ converges locally to $\mathcal{K}$.

Proof. Notice that the condition $\pi(1) < 1$ ensures that the Galton Watson tree has branching and so the conditioning is non-trivial. We consider critical trees, they die almost surely according to Lemma 4.6 and therefore $\mathbf{T}_n \in \mathbb{T}^f$. Since by definition Kesten's tree has exactly one infinite spine, $\mathcal{K} \in \mathbb{T}^1$. So we are in the setting of Theorem 6.6 and can use this to prove convergence in distribution. We just need to verify the two conditions. Let $\mathtt{t} \in \mathbb{T}^f$ and recall the definition $\mathbb{A}_n = \{A(\mathbb{T}) \in [n, n + n_1)\}$, where $n_1 \in \mathbb{N}^* \cup \{\infty\}$. Starting to show the first condition, notice that

$$\mathbb{P}(\mathbf{T}_n = \mathtt{t}) = \frac{\mathbb{P}(\mathbf{T}_n = \mathtt{t}, \mathbb{A}_n)}{\mathbb{P}(\mathbb{A}_n)} \leq \frac{1}{\mathbb{P}(\mathbb{A}_n)} \mathbb{1}_{\{A(\mathtt{t}) \in [n, n+n_1)\}}.$$

As $\mathtt{t} \in \mathbb{T}^f$, there exists an $N \in \mathbb{N}^*$ such that $N > A(\mathtt{t})$ and thus for all $n \geq N$, $\mathbb{1}_{\{A(\mathtt{t}) \in [n, n+n_1)\}} = 0$. This implies that $\lim_{n\to\infty} \mathbb{P}(\mathbf{T}_n = \mathtt{t}) \leq 0$ and since probabilities are positive, it is equal to zero. On the other hand, since $\mathcal{K}$ is infinite almost surely by definition, $\mathbb{P}(\mathcal{K} = \mathtt{t}) = 0$. Hence,

$$\lim_{n\to\infty} \mathbb{P}(\mathbf{T}_n = \mathtt{t}) = 0 = \mathbb{P}(\mathcal{K} = \mathtt{t}),$$

which is exactly the first condition. To find the second condition, let still $\mathtt{t} \in \mathbb{T}^f$ and in addition fix a leaf $x \in \mathscr{L}_0(\mathtt{t})$. Since $\mathbf{T}$ is finite almost surely, there exist finite many $\mathtt{t}'$ to graft on $\mathtt{t}$ to get $\mathbf{T}$ and we can sum over all possibilities

$$\mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x), \mathbb{A}_n) = \sum_{\mathtt{t}' \in \mathbb{T}^f} \mathbb{P}(\mathbf{T} = \mathtt{t} \circledast_x \mathtt{t}') \mathbb{1}_{\{A(\mathtt{t} \circledast_x \mathtt{t}') \in [n, n+n_1)\}} \tag{6.9}$$

Further, using (6.3) and (6.4), we find

$$\begin{aligned}\mathbb{P}(\mathbf{T} = \mathtt{t} \circledast_x \mathtt{t}') &= \frac{1}{\pi(0)} \cdot \mathbb{P}(\mathbf{T} = \mathtt{t})\mathbb{P}(\mathbf{T} = \mathtt{t}') \\ &= \frac{1}{\pi(0)} \cdot \pi(0)\mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x)))\mathbb{P}(\mathbf{T} = \mathtt{t}') \\ &= m_\pi^{H(x)}\mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x))\mathbb{P}(\mathbf{T} = \mathtt{t}') \qquad (6.10)\end{aligned}$$

Put (6.10) into (6.9) and recall that $m_\pi = 1$ in the critical case, so

$$\mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x), \mathbb{A}_n) = \mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x)) \sum_{\mathtt{t}' \in \mathbb{T}^f} \mathbb{P}(\mathbf{T} = \mathtt{t}')\mathbb{1}_{\{A(\mathtt{t} \circledast_x \mathtt{t}') \in [n, n+n_1)\}}.$$

The theorem has tree cases, we will now inspect each of them.

1. Assume that $n_1 = \infty$ and A satisfies the Monotonicity property. Then,

$$\begin{aligned}\mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x), \mathbb{A}_n) &= \sum_{\mathtt{t}' \in \mathbb{T}^f} \mathbb{P}(\mathbf{T} = \mathtt{t} \circledast_x \mathtt{t}')\mathbb{1}_{\{A(\mathtt{t} \circledast_x \mathtt{t}') \in [n, \infty)\}} \\ &\geq \sum_{\mathtt{t}' \in \mathbb{T}^f} \mathbb{P}(\mathbf{T} = \mathtt{t} \circledast_x \mathtt{t}')\mathbb{1}_{\{n \leq A(\mathtt{t}')\}},\end{aligned}$$

   where in the inequality the Monotonicity property was used. By (6.10),

$$\begin{aligned}\mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x), \mathbb{A}_n) &\geq \mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x)) \sum_{\mathtt{t}' \in \mathbb{T}^f} \mathbb{P}(\mathbf{T} = \mathtt{t}')\mathbb{1}_{\{n \leq A(\mathtt{t}')\}} \\ &= \mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x))\mathbb{P}(\mathbb{A}_n).\end{aligned}$$

   Thus,

$$\begin{aligned}\mathbb{P}(\mathbf{T}_n \in \mathbb{T}(\mathtt{t}, x)) &= \mathbb{P}(\mathbb{T} \in \mathbb{T}(\mathtt{t}, x) \mid \mathbb{A}_n) \\ &= \frac{\mathbb{P}(\mathbf{T} = \mathtt{t}, \mathbb{A}_n)}{\mathbb{P}(\mathbb{A}_n)} \\ &\geq \mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x))\end{aligned}$$

   for all n. So especially, if we take the limes inferior, we still have the inequality. This is exactly the second condition of Theorem 6.6.

2. Assume $n \in \mathbb{N}^* \cup \{\infty\}$ and A satisfies the Identity property. Then, for all $n \geq n_0$, by using the Identity property,

$$\sum_{\mathtt{t}' \in \mathbb{T}^f} \mathbb{P}(\mathbf{T} = \mathtt{t}')\mathbb{1}_{\{A(\mathtt{t} \circledast_x \mathtt{t}') \in [n, n+n_1)\}} = \sum_{\mathtt{t}' \in \mathbb{T}^f} \mathbb{P}(\mathbf{T} = \mathtt{t}')\mathbb{1}_{\{A(\mathtt{t}') \in [n, n+n_1)\}} = \mathbb{P}(\mathbb{A}_n).$$

Again, by Equation (6.10), we get

$$\mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x)) = \mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x))\mathbb{P}(\mathbb{A}_n).$$

Similarly as in the first case, just with equality instead of inequality, we get the second condition.

3. Assume $n \in \mathbb{N}^* \cup \{\infty\}$, A satisfies the Additivity property and Equation (6.8) holds. By (6.3) and the Additivity property,

$$\begin{aligned}\mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x), \mathbb{A}_n) &= \mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x)) \sum_{\mathtt{t}' \in \mathbb{T}^f} \mathbb{P}(\mathbf{T} = \mathtt{t}')\mathbb{1}_{\{A(\mathtt{t}')+D(\mathtt{t},x) \in [n, n+n_1)\}} \\ &= \mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x))\mathbb{P}(\mathbb{A}_n \in [n - D(\mathtt{t}, x), n + n_1 - D(\mathtt{t}, x))) \\ &= \mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x))\mathbb{P}(\mathbb{A}_{n-D(\mathtt{t},x)})\end{aligned}$$

Then,

$$\mathbb{P}(\mathbf{T}_n \in \mathbb{T}(\mathtt{t}, x)) = \frac{\mathbb{P}(\mathbf{T} \in \mathbb{T}(\mathtt{t}, x), \mathbb{A}_n)}{\mathbb{P}(\mathbb{A}_n)} = \mathbb{P}(\mathcal{K} \in \mathbb{T}(\mathtt{t}, x))\frac{\mathbb{P}(\mathbb{A}_{n-D(\mathtt{t},x)})}{\mathbb{P}(\mathbb{A}_n)}.$$

As the fraction is smaller or equal to one by (6.8), we again have the second property.

So we have shown that in all cases the two properties of Theorem 6.6 are fulfilled and therefore we can apply it and get local convergence. □

From now on will focus on the cases where n_1 is finite for all of the proofs. Most of the statements also work for $n_1 = \infty$. For this, consider looking at the mixture argument in [AD20].

6.4 Kesten's theorem and other applications

Using Theorem 6.9, we can look at many applications. We will condition critical Galton Watson trees on the height, the cardinality and the maximal degree of the nodes and then study the resulting limit.

6.4.1 Kesten's theorem (height)

We first focus on the height of a tree and therefore set the functional to be

$$A(\mathtt{t}) = \mathcal{H}(\mathtt{t}).$$

It satisfies the Additivity property with $n_1 = \mathcal{H}(\mathtt{t})$ and $D(\mathtt{t}, x) = H(x) \geq 0$. Indeed, let $n_0 = H(t)$, so by assumption $\mathcal{H}(\mathtt{t} \circledast_x \mathtt{t}') \geq \mathcal{H}(\mathtt{t})$. Further, we find that

$$\mathcal{H}(\mathtt{t} \circledast_x \mathtt{t}') = \max\{H(u), u \in \mathtt{t} \circledast_x \mathtt{t}'\} = \max\{\mathcal{H}(\mathtt{t}), \mathcal{H}(\mathtt{t}') + H(x)\}.$$

So

$$\max\{\mathcal{H}(\mathtt{t}), \mathcal{H}(\mathtt{t}') + H(x)\} \geq \mathcal{H}(\mathtt{t}).$$

This implies that

$$\mathcal{H}(\mathtt{t}') + H(x) \geq \mathcal{H}(\mathtt{t}).$$

Therefore,

$$\mathcal{H}(\mathtt{t} \circledast_x \mathtt{t}') = \mathcal{H}(\mathtt{t}') + H(x), \tag{6.11}$$

which is exactly the Additivity property for $D(\mathtt{t}, x) = H(x)$.

Lemma 6.10. *Consider a critical Galton Watson tree with offspring distribution π such that $\pi(1) < 1$. Let $n_1 \in \mathbb{N}^* \cup \{\infty\}$. We then have*

$$\lim_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = 1.$$

Proof. Assume $n_1 = 1$, so we look at events of the form $\mathbb{A}_n = \{\mathcal{H}(\mathbf{T}) = n\}$. The other cases can be deduced using mixture arguments. Recall Z_n, which describes the corresponding Galton Watson process, and the recursive generating functions labeled as φ_n. We have seen that we can look at it recursively: $\varphi_{n+1}(r) = \varphi(\varphi_n(r))$. Notice that

$$\mathbb{A}_n = \{Z_{n+1} = 0\} \setminus \{Z_n = 0\}$$

Then, using this and the definition of φ_n,

$$\mathbb{P}(\mathbb{A}_n) = \mathbb{P}(Z_{n+1} = 0) - \mathbb{P}(Z_n = 0) = \varphi_{n+1}(0) - \varphi_n(0).$$

So the fraction is

$$\frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = \frac{\varphi_{n+2}(0) - \varphi_{n+1}(0)}{\varphi_{n+1}(0) - \varphi_n(0)}. \tag{6.12}$$

To calculate this expression, let us have a closer look at $\varphi_{n+2}(r)$. Using some algebra and Taylor approximation,

$$\begin{aligned}
\varphi_{n+2}(0) &= \varphi(\varphi_{n+1}) \\
&= \varphi(\varphi_n(0) + (\varphi_{n+1}(0) - \varphi_n(0)) \\
&= \varphi_{n+1}(0) + (\varphi_{n+1}(0) - \varphi_n(0))\varphi'(\varphi_n(0)) + o(\varphi_{n+1}(0) - \varphi_n(0)).
\end{aligned}$$

Putting this expression back into (6.12), we have

$$\frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = \varphi'(\varphi_n(0)) + o(\varphi_{n+1}(0) - \varphi_n(0)).$$

Next, we analyze the behavior for $n \to \infty$. First, recall from Lemma 4.4 that

$$\lim_{n\to\infty} \varphi_n(0) = q_\pi = 1.$$

It holds that $q_\pi = 1$ as we work in the critical case. Further, the little-o term vanishes in the limit. Thus,

$$\lim_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = \varphi'(1) = m_\pi = 1$$

by (4.7) and in the critical case $m_\pi = 1$. □

Kesten's theorem is then a corollary of Theorem 6.9 using the previous lemma. Since the result is well-known, we will still denote it as a theorem.

Theorem 6.11. *Let* $\mathbf{T}$ *be a critical Galton Watson tree with offspring distribution* π *such that* $\pi(1) < 1$ *and* $\mathbf{T}_n$ *a random tree distributed as* $\mathbf{T}$ *conditionally on* $\{\mathcal{H}(t) \geq n\}$. *Denote* $\mathcal{K}$ *to be Kesten's tree associated with* π. *Then,*

$$\mathbf{T}_n \xrightarrow[n\to\infty]{(d)} \mathcal{K}.$$

Proof. By Lemma 6.10, $\lim_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = 1$, which implies $\limsup_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = 1$. In addition, $\mathcal{H}(\mathtt{t})$ satisfies the Additivity property as seen at the beginning of this section, see (6.11). So by the third point of Theorem 6.9, we have $\mathbf{T}_n \xrightarrow[n\to\infty]{(d)} \mathcal{K}$. □

6.4.2 Cardinality

Now suppose that

$$A(\mathtt{t}) = \#\mathtt{t},$$

so we look at the cardinality of the tree, which describes the total number of nodes. It satisfies the Additivity property with $n_0 = 0$ and $D(\mathtt{t}, x) = \#\mathtt{t} - 1$. To show this, take into account that

$$\#\{\mathtt{t} \circledast_x \mathtt{t}'\} = \#\mathtt{t} + \#\mathtt{t}' - 1, \tag{6.13}$$

since the cardinality of the grafted tree is just the sum of the two individual trees reduced by one to avoid counting the node x twice. Setting $D(\mathtt{t}, x) = \#\mathtt{t} - 1$, this is exactly the Additivity property.

Next, we introduce p as the *period* of the offspring distribution π defined as

$$p = \max\{k : \mathrm{supp}(\pi) \subseteq k\mathbb{N}\} = \gcd(\mathrm{supp}(\pi)),$$

where $\mathrm{supp}(\pi) = \{k \in \mathbb{N} : \pi(k) > 0\}$ describes the set of all $k \in \mathbb{N}$ such that the probability to have exactly k offspring is positive. If $p = 1$, π is called *aperiodic*.

Lemma 6.12. *Let* $\mathbf{T}$ *be a critical Galton Watson tree with offspring distribution* π *satisfying* $\pi(1) < 1$ *and with period* p. *Set* $\mathbb{A}_n = \{\#\mathbf{T} \in [n, n + n_1)\}$ *with* $n_1 \in \{p, \infty\}$. *Then,*

$$\lim_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = 1.$$

To prove this lemma, recall Dwass formula and the Strong Ratio Theorem.

Let $\mathbf{T}$ be a Galton Watson tree with offspring distribution π and let $(\mathbf{T}_k)_{k\in\mathbb{N}^*}$ be independent Galton Watson trees distributed as $\mathbf{T}$. Let $(X_k)_{k\in\mathbb{N}^*}$ be independent random variables distributed according to π. For $k \in \mathbb{N}^*$ and $n \geq k$, we have

$$\mathbb{P}(\#\mathbf{T}_1 + \cdots + \#\mathbf{T}_k = n) = \frac{k}{n}\mathbb{P}(X_1 + \cdots + X_n = n - k). \tag{6.14}$$

This formula is known under the name *Dwass formula*.

Let Y be an integrable aperiodic random variable with values in $\mathbb{Z}$ such that it is centered ($\mathbb{E}[Y] = 0$) and $\mathbb{P}(Y = 0) < 1$. Consider the random walk $(S_n)_{n\in\mathbb{N}}$ defined by $S_0 = 0$ and $S_n = \sum_{k=1}^n Y_k$. Then the random walk is recurrent, which means that it returns to its starting point almost surely. Define the periodicity of S as

$$p = \gcd\{n \geq 0 : \mathbb{P}(S_n = 0) > 0\}.$$

For S a recurrent aperiodic random walk, the *Strong Ratio Theorem* states that for $l \in \mathbb{Z}$,

$$\lim_{n\to\infty} \frac{\mathbb{P}(S_n = l)}{\mathbb{P}(S_n = 0)} = \lim_{n\to\infty} \frac{\mathbb{P}(S_n = 0)}{\mathbb{P}(S_{n+1} = 0)} = 1.$$

Further, if S is p-periodic, then for all $n \in \{1, \ldots, p\}$, there exists $j_n \in \mathbb{Z}$ and $k_n \in \mathbb{N}^*$ such that for all $k \geq k_n$

$$\mathbb{P}(S_{kp+n} = j_n) > 0$$

and the Strong Ratio Theorem can be modified to: For $l \in \mathbb{Z}$ and $n \in \{1, \ldots, p\}$,

$$\lim_{k\to\infty} \frac{\mathbb{P}(S_{kp+n} = lp + j_n)}{\mathbb{P}(S_{kp} = 0)} = 1. \tag{6.15}$$

With this, we are now able to prove Lemma 6.12.

Proof. Assume that $n_1 = p$. We then have $\mathbb{A}_n = \{\#\mathbf{T} \in [n, n+p)\}$ by definition. Hence, we want to show

$$\lim_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = \lim_{n\to\infty} \frac{\mathbb{P}(\#\mathbf{T} \in [n+1, n+1+p))}{\mathbb{P}(\#\mathbf{T} \in [n, n+p))} = 1.$$

Notice that $\#\mathbf{T} \in 1 + p\mathbb{N}$ almost surely by definition of p. Let $(S_n)_{n\in\mathbb{N}}$ be a random walk corresponding to the Galton Watson tree $\mathbf{T}$. It is defined as $S_0 = 0$ and $S_n = \sum_{k=1}^{n}(X_k - 1)$, where X_k are random variables distributed according to $\pi(k)$. S has period p. By Dwass formula (6.14) and the definition of the random walk, we have

$$\begin{aligned}\mathbb{P}(\#\mathbf{T} = kp+1) &= \frac{1}{kp+1}\mathbb{P}(X_1 + \cdots + X_{kp+1} = kp+1-1)\\ &= \frac{1}{kp+1}\mathbb{P}(S_{kp+1} + kp + 1 = kp)\\ &= \frac{1}{kp+1}\mathbb{P}(S_{kp+1} = -1)\end{aligned}$$

and hence, for $k \in \lfloor \frac{n-1}{d} \rfloor$,

$$\mathbb{P}(\mathbb{A}_n) = \mathbb{P}(A(\mathbb{T}) \in [n, n+p)) = \mathbb{P}(\#\mathbf{T} = kp+1) = \frac{1}{kp+1}\mathbb{P}(S_{kp+1} = -1). \tag{6.16}$$

Then,

$$\lim_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = \lim_{k\to\infty} \frac{kp+1}{(k+1)p+1} \frac{\mathbb{P}(S_{(k+1)p+1} = -1)}{\mathbb{P}(S_{kp+1} = -1)}.$$

Remark that for $k \to \infty$, the prefactor goes to one. Further, we can split the expression into

$$\frac{\mathbb{P}(S_{(k+1)p+1} = -1)}{\mathbb{P}(S_{kp+1} = -1)} = \frac{\mathbb{P}(S_{(k+1)p+1} = -1)}{\mathbb{P}(S_{(k+1)p} = 0)} \cdot \frac{\mathbb{P}(S_{(k+1)p} = 0)}{\mathbb{P}(S_{kp} = 0)} \cdot \frac{\mathbb{P}(S_{kp} = 0)}{\mathbb{P}(S_{kp+1} = -1)}.$$

We can look at each part independently and use the Strong Ratio Theorem for all of them. For the first term, we set $n = 1$ and have $j_1 = -1$, $l = 0$. Since by (6.16) we have

$$\mathbb{P}(S_{kp+1} = -1) = (kp+1) \cdot \mathbb{P}(\mathbb{A}_n) > 0,$$

the prerequisite of the Strong Ratio Theorem is fulfilled. Therefore, by (6.15), this fraction goes to one as $k \to \infty$. For the second term, set $n = 0, j_0 = 0$ and $l = 0$. As S is $p-$periodic, $\mathbb{P}(S_{kp} = 0) > 0$ and therefore we can again apply the Strong Ratio

Theorem and this fraction is one in the limit. The last term is similar as the first one, just the reciprocal value, which again gives one. Therefore,

$$\lim_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = 1 \cdot 1 \cdot 1 = 1.$$

□

Using this lemma and the important Theorem 6.9, we can also prove convergence to Kesten's tree. Remark that we condition on $\{\#\mathbf{T} = 1 + pn\}$ to avoid dividing by zero.

Corollary 6.13. *Let $\mathbf{T}$ be a critical Galton Watson tree with offspring distribution π and $\pi(1) < 1$. Let p be the period of $\mathbf{T}$ and $\mathbf{T}_n$ a random tree distributed as $\mathbf{T}$ conditionally on $\{\#\mathbf{T} = 1 + pn\}$ for large n. Denote $\mathcal{K}$ to be Kesten's tree associated with π. Then,*

$$\mathbf{T}_n \xrightarrow[n\to\infty]{(d)} \mathcal{K}.$$

Proof. We have seen that the cardinality fulfills the Additivity property for $n_0 = 0$ and $D(\mathtt{t}, x) = \#\mathtt{t} - 1$, see (6.11).By Lemma 6.12, $\lim_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = 1$, which implies $\limsup_{n\to\infty} \frac{\mathbb{P}(\mathbb{A}_{n+1})}{\mathbb{P}(\mathbb{A}_n)} = 1$. And applying the third point of Theorem 6.9 gives local convergence. □

6.4.3 Maximal degree

Lastly, consider the maximal degree, so

$$A(\mathtt{t}) = M(\mathtt{t}) = \max\{k_u(\mathtt{t}), u \in \mathtt{t}\}.$$

Remark that

$$M(\mathtt{t} \circledast_x \mathtt{t}') = \max\{M(\mathtt{t}), M(\mathtt{t}')\},$$

as we can just look at the two trees individually and pick the maximal degree over both, since grafting the trees only leads to a change for the node x, where we graft. As $k_x(\mathtt{t}) = 0$ before grafting, it was certainly not the maximum. And the degree that node x had in the tree $\mathtt{t}'$ is still included as $k_x(\mathtt{t} \circledast_x \mathtt{t}') = k_x(\mathtt{t}')$. If we set $n_0 = M(\mathtt{t}) + 1$, then $M(\mathtt{t} \circledast_x \mathtt{t}') \geq M(\mathtt{t}) + 1$, from which we deduce that $M(\mathtt{t} \circledast_x \mathtt{t}') < M(\mathtt{t})$ as they are integers. So the maximum is $M(\mathtt{t}')$ and thus

$$M(\mathtt{t} \circledast_x \mathtt{t}') = \max\{M(\mathtt{t}), M(\mathtt{t}')\} = M(\mathtt{t}'). \tag{6.17}$$

Hence, $M(\mathtt{t})$ fulfills the Identity property.

Corollary 6.14. *Let* $\mathbf{T}$ *be a critical Galton Watson tree with offspring distribution* π *fulfilling* $\pi(1) < 1$ *and unbounded support. Let* $\mathbf{T}_n$ *be a random tree distributed as* $\mathbf{T}$ *conditionally on* $\{M(\mathbf{T}) \geq n\}$*. Then,*

$$\mathbf{T}_n \xrightarrow[n\to\infty]{(d)} \mathcal{K}.$$

Further, if we let $\mathbf{T}_n$ *be a random tree distributed as* $\mathbf{T}$ *conditionally on* $\{M(\mathbf{T}) = n\}$*. Then, along the subsequence* $\{n, \pi(n) > 0\}$*, we have*

$$\mathbf{T}_n \xrightarrow[n\to\infty]{(d)} \mathcal{K}.$$

Proof. $M(\mathtt{t})$ fulfills the Identity property, as seen in (6.17). Further, $\mathbb{P}(\mathbb{A}_n) > 0$ as we assume unbounded support, so there for sure exists an n such that the probability is not zero. Thus, the first part follows directly from the second point Theorem 6.9 with $n_1 = \infty$. For $\{M(\mathbf{T}) = n\}$, we do not immediately have that the probability is bigger than zero. So we restrict it to the subsequence $\{n, \pi(n) > 0\}$. There, $\mathbb{P}(\mathbb{A}_n) = \{M(\mathbf{T}) = n\} = \pi(n) > 0$. Then, again by the second part of Theorem 6.9, using $n_1 = 1$, we get result. □

To conclude Chapter 6, we recall that conditioning critical Galton Watson trees directly on survival cannot be done, as the probability is zero. Instead, we work with convergence of random trees. We used the functional A, conditioned the trees on the events $\mathbb{A}_n$ studied the limit as $n \to \infty$. As a result, for suitable assumptions, critical Galton Watson trees conditioned on $\mathbb{A}_n$ converge locally to Kesten's tree as $n \to \infty$.

Chapter 7

Conclusion

In this thesis, the focus was to study Galton Watson trees under conditioning on different events related to survival and extinction. We grouped the trees into three categories, depending on their mean number of children m_π: super-critical if $m_\pi > 1$, critical if $m_\pi = 1$ and sub-critical if $m_\pi < 1$. We focused on the first two cases. Only for super-critical Galton-Watson trees both survival and extinction happen with positive probability. This discovery already answers the opening question of the introduction: Your surname survives only if the underlying branching process is super-critical. But it was not the main focus of the thesis. The goal was to understand how Galton Watson trees behave under different types of conditioning. The fact that for super-critical Galton Watson trees both survival and extinction happen with positive probability allowed us to condition directly on these events. We found that conditioning super-critical Galton Watson trees on survival leads to a two-type Galton Watson tree and conditioning on extinction yields to a sub-critical Galton Watson tree. In the critical case, conditioning naively is no longer possible as survival occurs with probability zero. Instead, we introduced local convergence of trees. We defined a functional A, conditioned critical Galton Watson trees on events of the form $\mathbb{A}_n = \{A(\mathbf{T}) \in [n, n + n_1)\}$ for some $n_1 \in \mathbb{N}^* \cup \{\infty\}$ and looked at the resulting limit of the conditioned trees as $n \to \infty$. With grafted trees, we found suitable conditions for A ensuring that conditioning on $\mathbb{A}_n$ is meaningful. Under these assumptions, critical Galton Watson trees conditioned on $\mathbb{A}_n$ converge locally to Kesten's tree, that is a random tree with exactly one infinite spine and independent Galton Watson trees attached to it. We considered different choices for A: the height, the cardinality and the maximal degree. Always leading to the same result, conditioned critical Galton Watson trees converge locally to Kesten's tree. Several further functionals could be analyzed, such as the number of leaves, the number of nodes with given out degree and more. For this, see [AD20], where also the sub-critical case ($m_\pi < 1$) is treated.

Bibliography

[AD13] R. Abraham and J.-F. Delmas. *Local Limits of Conditioned Galton Watson Trees I: The Infinite Spine Case*. 2013. URL: https://arxiv.org/pdf/1304.4035v4.

[AD20] R. Abraham and J.-F. Delmas. *An introduction to Bienaymé-Galton-Watson trees and their local limits*. 2020. URL: https://arxiv.org/abs/1506.05571.

[Bar17] Blaszczyszyn Bartolomiej. *Lecture Notes on Random Geometric Models - Random Graphs, Point Processes and Stochastic Geometry*. Doctoral. Japan. 2017. URL: https://inria.hal.science/cel-01654766v2.

[Bil99] Patrick Billingsley. *Convergence of Probability Measures*. 1999.

[C24] N. C. *Random walks and graphs: a few topics*. Lecture notes, University of Orsay. 2024.

[Duq24] Thomas Duquesne. *Processus markoviens 1*. 2024.

[Gal22] J.-F. Le Gall. *Measure Theory, Probability and Stochastic Processes*. Springer, 2022. URL: https://doi.org/10.1007/978-3-031-14205-5.

Some AI tools (Microsoft Copilot and ChatGPT) were used as a help for the adjustment of tikz trees, finding errors in the codes, having a better understanding of some statements and proofs and improving the clarity of them. All parts have been checked by me and I take full responsibility for the content.