Bachelor Thesis

at

Heidelberg University

Fakultät für Mathematik und Informatik

Scaling Limits of Random Trees

From Itō excursion theory to a limit theorem towards the Brownian tree

Submitted by:	Tobias Schröder
Matriculation Number:	3397300
Supervisor:	Prof. Dr. Anna Wienhard
Date of submission:	June 28, 2019

Abstract

In this thesis we investigate the asymptotic behavior of large random Galton-Watson (GW) -trees. It is found that in the space of compact real trees equipped with Gromov-Hausdorff metric a limit in distribution can be defined. This limit is a real tree coded by the normalized excursion of Brownian motion. First, Itōs excursion measure of the Brownian motion is introduced. Using the Markov property of Brownian motion the law of the normalized Brownian excursion is specified. With this result and the methods of Le Gall and Miermont, [14], a limit theorem of the excursion of a simple random walk towards the normalized excursion is proven. An excursion of a simple random walk coincides with the distribution of a GW-tree's contour process. It is shown that similarly arbitrary continuous functions of the unit interval yield a real tree using a rerooting isometry. Random trees can then be seen as random variables taking values in the space of real compact trees endowed with the pointed Gromov-Hausdorff topology. With respect to this metric, the mapping from the space of excursions to the space of compact real trees turns out to be continuous. Using the continuous mapping theorem and the limit theorem of excursions it follows that the sequence of rescaled GW-trees converges in distribution towards Aldous' Continuum Random Tree (CRT, [1]).

Zusammenfassung

In dieser Arbeit wird das asymptotische Verhalten großer zufälliger Galton-Watson (GW) -Bäume untersucht. Es wird festgestellt, dass ein schwacher Grenzwert im Raum kompakter reeller Bäume mit Gromov-Hausdorff Abstand definiert werden kann und durch die normalisierte Exkursion einer Brownschen Bewegung beschrieben wird. Zunächst wird Itōs Exkursionsmaß der Brownschen Bewegung eingeführt. Die Markoveigenschaft der Brownschen Bewegung wird benutzt um das Gesetz der normalisierten Brownschen Exkursion zu spezifizieren. Mit diesem Resultat wird mit den Methoden von Le Gall und Miermont [15] ein Grenzwertsatz der Exkursionen von einfachen stochastischen Irrfahrten gegen die normalisierte Brownsche Exkursion gezeigt. Eine Exkursion einer stochastischen Irrfahrt beschreibt den Konturprozess zufälliger GW-Bäume. Es wird gezeigt, dass beliebige stetige Funktionen auf dem Einheitsintervall reelle Bäume kodieren indem eine Rerooting-Isometrie genutzt wird. Zufällige Bäume können dann als Zufallsvariablen mit Werten im Raum kompakter reeller Bäume aufgefasst werden, der mit der punktierten Gromov-Hausdorff Topologie ausgestattet ist. Bezüglich dieser Metrik ist die Abbildung vom Raum der Exkursionen in den Raum kompakter reeller Bäume stetig. Aus dem Grenzwertsatz des Konturprozesses von GW-Bäumen gegen die normalisierte Brownsche Exkursion und dem Continuous Mapping Theorem folgt, dass die Folge reskalierter GW-Bäume in Verteilung gegen Aldous' Continuum Random Tree (CRT, [1]) konvergiert.

Introduction

Limit theorems are one of the most used tools in probability theory. The central limit theorem, for example, emphasizes the central role of the normal distribution and is used frequently in applied statistics. A very similar result is obtained in the theory of stochastic processes for the scaling limit of a symmetric random walk and is called Donsker's theorem. In this case, the scaling limit is known as Brownian motion, which is a process with independent normally distributed increments. Similar to the normal distribution, the Brownian motion has a prominent role in probability theory and statistics due to its universal properties as a scaling limit.

In this thesis we will prove three further limit theorems. First we will show the local central limit theorem (Theorem 3.7), which states that the scaling limit of a simple random walk is locally normally distributed, and the convergence towards the normal distribution happens uniformly. From this we can establish a conditional version of Donsker's theorem: The rescaled excursions of a random walk converge to the later introduced normalized excursion of Brownian motion. Although this theorem reads very similarly to Donsker's theorem, new concepts have to be introduced to make this conclusion. The theory of this limit theorem is the topic of chapters three and four. The final aim of this thesis is the scaling limit of random Galton-Watson trees with geometric offspring distribution and rescaled graph distance as a sequence of compact metric spaces. We will identify Aldous' CRT [1] as the weak limit in the Gromov-Hausdorff topology. (This case even gives access to a larger class of graphs by the application of the Cori-Vauquelin-Schaeffer bijection, compare [15].)

The first chapter is devoted to a preliminary discussion of stochastic processes, in particular Brownian motion. We introduce the relevant types of convergence of random variables and their relation. Afterwards we discuss Markov processes and Brownian motion as the most important example of a continuous time Markov process.

We need a definition of a normalized Brownian excursion, an object that only appears with probability zero, and a well defined law on the subspace of normalized excursions. This can be done using the excursion theory of Itō, which we will discuss in the second chapter. Although this theory can be quite technical it is interesting in its own right and serves as a tool in other areas of probability theory as well. It allows us to condition Brownian motion on the P-null set of excursions by defining a Poisson point process attached to Brownian motion, following the general case proven by Itō [8]. Central to the discussion will be Theorem 2.15 which emphasises the Markovian properties of the excursion law and the connection to killed Brownian motion. We will conclude with a probability density of the finite dimensional distributions of the normalized Brownian excursion.

In the third chapter we will prove the limit theorem of the rescaled normalized excursions of the simple random walk towards the normalized excursion of Brownian motion. This requires to establish tightness of the sequence of probability laws and to prove that the finite dimensional distributions converge to the previously derived probability density. We follow the reasoning of

[15] and generalize the argument which shows the convergence of one dimensional distributions to the finite dimensional case. In this context we will also establish the local central limit theorem (Theorem 3.7) and combinatorial tools like the cycle lemma, which are interesting by themselves.

The second part of the thesis deals with random trees. In chapter four we introduce rooted plain trees and a random version, the Galton-Watson (GW) tree. Rooted trees with k edges can be uniquely identified with a contour function of length $2k$ or, equivalently, a Dyck word of length $2k$. We show that in the case of a random GW-tree with geometric offspring distribution this contour function becomes the excursion of a random walk. This will allow us to relate random trees to the previously investigated excursions.

In the final chapter we discuss compact real trees and the scaling limit of rescaled GW-trees. Similarly to plain trees, compact real trees are geodesic compact metric spaces without loops and can be coded by a continuous non-negative function which is zero at its endpoints, i.e. a normalized excursion. To prove this assertion we show a Rerooting Lemma (Lemma 5.3) which states that a real tree can be rerooted at an arbitrary point by an isometric mapping. Secondly we introduce a metric on the set of rooted real trees, the pointed Gromov-Hausdorff distance. The set of rooted real trees becomes a metric space and weak convergence can be defined as usual. By proving that the mapping from the set of excursions to the set of real trees is continuous, the scaling limit of GW-trees follows as a corollary.

Acknowledgements

On my search for a possible topic of my thesis I got intrigued when Prof. Dr. Anna Wienhard mentioned random geometry and in particular random trees. I want to thank her for giving me the opportunity to write a thesis on this topic and for giving me the freedom to choose my own perspective on it.

I want to thank M.Sc. Joseph Theo Meyer for all the support, encouragement and suggestions he gave me during the final stages of my thesis. His advice was incredibly helpful to improve formal points and proofs.

Contents

1. Preliminaries

This chapter gives an introduction to important concepts in the theory of stochastic processes and its most prominent example, Brownian motion. For general, often measure theoretic, statements we refer to Billingsley [4], more information on Brownian motion can be found in Rogers, Williams [12], [13] and Revuz, Yor [24].

1.1. On convergence types in probability theory

The theorems of the first two sections were stated like this in the lecture notes [16]. We introduce the convergence in distribution, convergence in probability and almost sure convergence and discuss the relation between them.

Definition 1.1 (Random variable). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(S, \mathcal{A})$ a measurable space. A random variable is a $\mathcal{F}/\mathcal{A}$-measurable function $X : \Omega \to S$.*

Definition 1.2 (Generated sigma-algebra). *For a collection $(G_i)_{i\in I}$ of subsets $G_i \subseteq \Omega$ we call the smallest σ-algebra that contains $(G_i)_{i\in I}$ the σ-algebra generated by $(G_i)_{i\in I}$ and write $\sigma((G_i)_{i\in I})$. Given a random variable $X : (\Omega, \mathcal{F}) \to (S, \mathcal{A})$, we define the sub-$\sigma$-algebra of $\mathcal{F}$ generated by its preimages, i.e.*

$$\sigma(X) := \sigma(\{X^{-1}(A) \mid A \in \mathcal{A}\}) \subseteq \mathcal{F} .$$

In the following we will discuss random variables (r.v.) where (S, d) is a metric space and $\mathcal{A}$ the Borel-σ-algebra. The Borel-σ-algebra is the σ-algebra generated by the sets that are open with respect to d. The key property of a random variable is its distribution, which is given by the image measure of P under X.

Definition 1.3 (Pushforward measure). *Given a random variable X on a probability space $(\Omega, \mathcal{F}, P)$ we define the pushforward or image measure on $\mathcal{A}$ as*

$$X_*P(A) := P(X^{-1}(A)) \text{ for all } A \in \mathcal{A} .$$

The following alternative more common notations exist: The pushforwards measure is sometimes called the law of X and is denoted as $\mathcal{L}(X)$. One also writes $P^X := X_*P$ and we often use the conventional notation

$$P(X \in A) := P(\{\omega \mid X(\omega) \in A\}) = P(X^{-1}(A)) .$$

Recall that a measurable $\mathbb{R}$-valued function is called integrable, iff the integral of $|f|$ exists. The image measure enjoys the following transformation rule for the integral:

Theorem 1.4. *Let $g : S \to \mathbb{R}$ be a Borel-measurable function and $X : (\Omega, \mathcal{F}, P) \to (S, \mathcal{A})$ a random variable. It then holds for $A \in \mathcal{A}$*

$$\int_A g \,\mathrm{d}P^X = \int_{X^{-1}(A)} g \circ X \,\mathrm{d}P$$

if one of the two integrals is defined.

A proof can be found in [23]. This holds not only for probability measure but for general measures μ. Depending on the context the following notation for the integral of a Borel function $f : S \to \mathbb{R}$ with respect to a measure μ on S is commonly used,

$$\mu[f] := \int_S f \mathrm{d}\mu \,.$$

The notation is especially relevant in the theory of Markov processes, where it is important to distinguish between expectations under various measures. An integral with respect to a probability measure is called expectation:

Definition 1.5 (Expectation)**.** *Let $X : \Omega \to \mathbb{R}$ be an integrable function. The expectation is defined as*

$$\mathrm{E}X = \int_\Omega X \mathrm{d}P \,.$$

From the transformation rule for $g = \mathrm{id}$ we can write the expectation in the more familiar form $\mathrm{E}X = \int_{\mathbb{R}} x \mathrm{d}P^X$.

Definition 1.6 (Almost surely)**.** *We say that an event $E \subseteq \Omega$ happens almost surely (a.s.) if there exists a nullset $N \in \mathcal{F}$, i.e. $P(N) = 0$ such that $N^C \subseteq E$. We write $X_n \xrightarrow{a.s.} X$ if a sequence $(X_n)_{n \in \mathbb{N}}$ of S-valued r.v. converges to $X : \Omega \to S$ almost surely.*

Almost sure convergence is equivalent to the pointwise convergence almost everywhere. Informally this definition means that X_n converges to X with probability one. The definition is stated like this since the set E may not be measurable. For example, we will later claim that Brownian motion is continuous almost surely, although this is not a measurable event.

Definition 1.7 (Convergence in probability)**.** *A sqeuence of $(S, \mathcal{A})$ valued r.v. $(X_n)_{n \in \mathbb{N}}$ converges to X in probability, denoted $X_n \xrightarrow{p} X$, iff*

$$P(d(X_n, X) > \varepsilon) \to 0 \,.$$

Convergence in probability is equivalent to the convergence in measure. Finally we also have:

Definition 1.8 (Convergence in distribution)**.** *A sequence of probability measures $(\mu_n)_{n \in \mathbb{N}}$ on $(S, \mathcal{A})$ converges weakly to a probability measure μ on $(S, \mathcal{A})$, denoted $\mu_n \rightharpoonup \mu$, iff*

$$\lim_{n \to \infty} \int_S f \mathrm{d}\mu_n = \int_S f \mathrm{d}\mu$$

for all $f \in C_b(S) = \{g : S \to \mathbb{R} \,|\, g \text{ bounded and continuous } \}$.
A sequence of $(S, \mathcal{A})$ valued r.v. $(X_n)_{n \in \mathbb{N}}$ converges to X in distribution, denoted $X_n \xrightarrow{d} X$, iff

$$X_{n*}P \rightharpoonup X_*P$$

It is called convergence in distribution, because the Helly-Bray theorem relates it in the case of real valued random variables to the pointwise convergence of distribution functions at all continuity points, [4]. In the final chapter of the thesis we need two more fundamental results. The first one is a generalization of Slutzky's theorem.

Theorem 1.9 (Slutzky)**.** *For $(S, \mathcal{A})$ valued sequences of random variables X_n, Y_n on some common probability space assume*

$$X_n \xrightarrow{d} X \qquad \text{and} \qquad d(X_n, Y_n) \xrightarrow{p} 0\,.$$

Then also

$$Y_n \xrightarrow{d} X$$

A proof can be found in [4, Theorem 3.1]. By Slutzky's theorem applied to $X_n = X$, convergence in probability implies convergence in distribution. Finally we have

Theorem 1.10 (Continuous mapping theorem)**.** *Let (S, d_S) and (T, d_T) be metric spaces, $g : S \to T$ a continuous mapping and X_n a sequence of $(S, \mathcal{A})$-valued random variables converging in distribution to a r.v. X. Then*

$$g(X_n) \xrightarrow{d} g(X)\,.$$

The theorem relies on the fact that for any $f \in C_b(T)$ the composition is continuous and bounded as well, $f \circ g \in C_b(S)$. This theorem is important to infer new limit laws from known ones.

1.2. Stochastic processes

We will now take a closer look at a collection of random variables indexed by "time" and how to fix the law of an uncountable family of random variables by Kolmogorov's extension theorem. An example of such a process is Brownian motion. From now on $(\Omega, \mathcal{F}, P)$ denotes a probability space.

Definition 1.11 (Real valued stochastic process)**.** *Let T be some index set and $(X_t)_{t\in T}$ be a family of random variables $X_t : \Omega \to \mathbb{R}^n$, where $\mathbb{R}^n$ is endowed with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$. The collection $X := (X_t)_{t\in T}$ is called stochastic process. $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is called state space of the process.*

The index set is usually one of these cases: In the time discrete case $T = \mathbb{N}_0$ and sometimes the notation $X_n := (X_n)_{n\in\mathbb{N}_0}$ is used. In the time continuous case $T = \mathbb{R}_{\geq 0}$ or $T = \mathbb{R}_{>0}$ in which case X_t can occasionally denote the entire process.

Definition 1.12. *For every $\omega \in \Omega$ the function $X(\omega) : T \to \mathbb{R}^n$ is called sample function or realization of X corresponding to ω. We denote the set of all possible functions $T \to \mathbb{R}^n$ by $(\mathbb{R}^n)^T$.*

The stochastic process should rather be viewed as a random variable in itself. For this consider the family of projection maps:

$$\pi_I^J : (\mathbb{R}^n)^J \to (\mathbb{R}^n)^I,\ (x_t)_{t\in J} \mapsto (x_t)_{t\in I}$$

for all $I \subseteq J \subseteq T$.

Definition 1.13. *We denote $\mathcal{B}(\mathbb{R}^n)^T$ as the smallest σ-algebra in which π_t^T is measurable for all $t \in T$. This σ-algebra is also called the σ-algebra of σ-cylinders.*

Details can be found in [12] II.25.

Lemma 1.14. *A map $f : (\Omega, \mathcal{F}) \to ((\mathbb{R}^n)^T, \mathcal{B}(\mathbb{R}^n)^T)$ is measurable if an only if $\pi_t^T(f)$ is measurable for all $t \in T$.*

Proof. The σ-algebra $\mathcal{B}(\mathbb{R}^n)^T$ is generated by the sets $(\pi_t^T)^{-1}(A)$ for $A \in \mathcal{B}(\mathbb{R}^n)$. Consequently we must check that

$$\{\omega \,|\, f(\omega) \in (\pi_t^T)^{-1}(A)\} \in \mathcal{F} \text{ for all } t \in T$$

but $\{\omega \,|\, f(\omega)(t) \in A\} \in \mathcal{F}$ by assumption. The converse direction follows directly from the definition, since the composition of measurable maps is measurable. □

Corollary 1.15. *A stochastic process $(X_t)_{t\in T}$ is a $(\mathbb{R}^{nT}, \mathcal{B}(\mathbb{R}^n)^T)$-valued random variable.*

Fixing the law of stochastic processes

The law of the stochastic process is, as always, given by X_*P. However, this is usually a quantity that can not be fixed easily. Usually one can only give a description of the distributions of $(X_{t_1}, \dots, X_{t_k})$ for $t_1, \dots, t_k \in T$, and thats also the information accessible from an experiment. As it turns out it is sufficient information to determine a unique probability law, if these informations are consistent in the following sense:

Definition 1.16 (Projective family). *A family of probability measures $(P_I)_{I\subseteq T \text{ finite}}$, P_I being a measure on $(\mathbb{R}^{nI}, \mathcal{B}(\mathbb{R}^n)^I)$ for a finite subset $I \subseteq T$, is called a projective family iff*

$$P_I(B) = P_J({\pi_I^J}^{-1}(B)) \text{ for all } \mathcal{B}(\mathbb{R}^n)^I,\ I \subseteq J \subseteq T \text{ finite}$$

A stochastic process naturally gives rise to a projective family.

Definition 1.17 (Finite dimensional distributions). *Let $(X_t)_{t\in T}$ be a stochastic process. The family $(P_I)_{I\subseteq T \text{ finite}}$ with $P_I := P^{(X_t)_{t\in I}}$ defines a projective family called the finite dimensional distributions (f.d.d.) of X.*

The finite dimensional distributions are now in fact enough to fix the law of a stochastic process.

Theorem 1.18 (Kolmogorov's extension theorem). *Let $E \subseteq \mathbb{R}^n$ be a connected closed subset. Given a projective family $(P_I)_{I\subseteq T \text{ finite}}$ there exists exactly one measure Q on $(E^T, \mathcal{B}(E^T))$ such that for all $A \in \mathcal{B}(E)^I$, $I \subseteq T$ finite, it holds*

$$P_I(A) = Q((\pi_I^T)^{-1}(A))\,.$$

In particular, given a stochastic process and its finite dimensional distributions there exists a unique law of the stochastic process whose projections are its finite dimensional distributions.

For a proof see [3, page 482]. This theorem will ensure the existence of Brownian motion later on. The fact that the description of a stochastic process only depends on its finite dimensional distributions, which are measures on $(\mathbb{R}^{nI}, \mathcal{B}^{nI})$, is important for many proofs on stochastic processes.

1.3. Independence, Conditional expectations and Filtrations

The time arrow of a stochastic process is modeled using filtrations. In our case these are especially important for the theory of Markov processes and mostly appear in conditional expectations. These terms will be defined in this section. For a source of definitions and statements see [12, II].

Definition 1.19 (Independence)**.** *We say that sub-σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \dots \subseteq \mathcal{F}$ are independent iff for all $G_i \in \mathcal{G}_i, i \in \mathbb{N}$ and for $i_1, i_2, \dots, i_n \in \mathbb{N}$*

$$P(G_{i_1} \cap G_{i_2} \cap \dots \cap G_{i_n}) = \prod_{k=1}^{n} P(G_{i_k}) .$$

Random variables are called independent, iff they generate independent sub-σ-algebras, i.e. for $X_1, X_2, \dots$ the σ-algebras $\sigma(X_1), \sigma(X_2), \dots$ are independent.

The first definitions are taken from [16] and given in similar way in [12]. The next definition is one of the most important ones for the second chapter of this text.

Definition 1.20 (Conditional expectation)**.** *Consider an integrable real-valued random variable X on a probability space $(\Omega, \mathcal{F}, P)$ and a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. A random variable $\mathrm{E}[X|\mathcal{G}]$ is called conditional expectation of X with respect to $\mathcal{G}$, iff*

1. $\mathrm{E}[X|\mathcal{G}]$ *is $\mathcal{G}$ measurable*
2. $\int_G \mathrm{E}[X|\mathcal{G}]\mathrm{d}P = \int_G X\mathrm{d}P$ *for all $G \in \mathcal{G}$*

If $\mathcal{G} = \sigma(Y)$ denotes the σ-algebra generated by a random variable Y we also write $\mathrm{E}[X|Y]$.

One can show that such an object exists and is unique P a.s.([3]). Conditional expectations are of great importance in the theory of stochastic processes. They give our best prediction of the expected values of X given that we possess the "information" contained in $\mathcal{G}$ about X. For example, if $\mathcal{G} = \{\emptyset, \Omega\}$, which means that we have no non trivial information about X, then $\mathrm{E}[X|\mathcal{G}] = \mathrm{E}[X]$. In the theory of Markov processes the following more general non trivial statements are needed.

Lemma 1.21. *Consider random variables X, Y and σ-algebra $\mathcal{G}$. We have*

1. *If X and $\mathcal{G}$ are independent, then* $\mathrm{E}[X|\mathcal{G}] = \mathrm{E}[X]$.
2. *$E[XY|\mathcal{G}] = X\mathrm{E}[Y|\mathcal{G}]$ if X is $\mathcal{G}$ measurable.*
3. *For $\mathcal{G}_1 \subseteq \mathcal{G}_2$ we have* $\mathrm{E}[\mathrm{E}[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathrm{E}[X|\mathcal{G}_1]$.

For proofs see [3].

Definition 1.22 (Conditional probability)**.** *Consider a random variable $X : (\Omega, \mathcal{F}) \to (S, \mathcal{A})$ and $A \in \mathcal{A}$. In this case $\mathbb{1}_A$ is a measurable function and we define*

$$P(X \in A \,|\, \mathcal{G}) = E[\mathbb{1}_A(X) \,|\, \mathcal{G}]$$

The gain of information with time is now conveniently described using filtrations.

Definition 1.23 (Filtration)**.** *A filtration is a family of σ-algebras $(\mathcal{F}_t)_{t\in T}$ with $\mathcal{F}_t \subseteq \mathcal{F}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t, s \in T$.*

Example 1.24. *Given a stochastic process $(X_t)_{t\in T}$ the filtration $\mathcal{F}_t^X := \sigma(X_s,\, s \leq t, s \in T)$ is called natural filtration.*

Definition 1.25 (Adapted process)**.** *A process $(X_t)_{t\in T}$ is called adapted to $(\mathcal{F}_t)_{t\in T}$, iff X_t is $\mathcal{F}_t$-measurable for every $t \in T$.*

Stopping times are real valued random variables which usually correspond to the time that some random event occurs.

Definition 1.26 (Stopping-time)**.** *A $(\mathcal{F}_t)_{t\geq 0}$-stopping time S is a real valued random variable such that*

$$\{S \leq t\} \in \mathcal{F}_t \text{ for all } t \in T\,.$$

This definition means that at any given time we posses enough information to tell whether the stopping time has passed already.

1.4. Markov processes

In this section we give a general definition of a homogeneous Markov process. For a Markov process the evolution after a fixed time is independent of the history of the process up to this time. When the process starts at some point in time and space, its evolution between any two times is given by a transition function. Transition functions are a special type of kernel, which we will define first:

Definition 1.27 (Kernel)**.** *Let $(E, \mathcal{E})$ be a measurable space. A kernel N on E is a map from $E \times \mathcal{E}$ into $\mathbb{R}_+ \cup \{\infty\}$ such that*

1. *For all $x \in E$ $N(x, \cdot)$ is a measure on $\mathcal{E}$*
2. *For all $A \in \mathcal{E}$ the map $N(\cdot, A)$ is $\mathcal{E}$-measurable*

For two kernels N, M we define a product

$$MN(x, A) := \int_E N(y, A) M(x, \mathrm{d}y)\,, \tag{1.1}$$

which again is a kernel, and on Borel measurable functions $f : \mathcal{E} \to \mathbb{R}$ we set

$$Nf(x) = \int_E f(y) N(x, \mathrm{d}y)$$

A kernel N is called transition probability, if $N(x, E) = 1$ for all $x \in E$.

Transition functions are a special family of kernels.

Definition 1.28 (Transition function)**.** *A homogeneous transition function on $(E, \mathcal{E})$ is a family $(P_t)_{t\in\mathbb{R}}$ of transition probabilities that satisfy the Chapman-Kolmogorov equations*

$$P_{t+s}(x, A) = P_t P_s(x, A)$$

where the product to the right hand side has to be understood in the sense of (1.1).

Intuitively the transition function $P_t(x, A)$ describes the probability that a stochastic process starting at x lies in A at time t. We are especially interested in processes that can "die". In this case a process starting at some point might not live long enough to hit anything at a later time, which means that in general our transition functions will fulfill $P_t(x, E) < 1$. On the other hand such processes may still fulfill the Markov property at long as the process lives. In order to include such cases explicitly in the definition we add a "coffin" state δ. This makes the definition of a Markov process more complicated than in the conventional case but simplifies technicalities. The following definitions are taken from Blumenthal and Getoor [18].

Definition 1.29 (Markov process). *Consider the space $E_\delta = E \cup \{\delta\}$ endowed with the σ-Algebra $\mathcal{E}_\delta = \sigma(\mathcal{E}, \{\delta\})$, where δ is often called coffin state of the process. Consider a tuple $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in T}, X, \theta_t, P^x)$ where*

1. *$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in T})$ is a sufficiently rich filtered measurable space.*
2. *X is an $(\mathcal{F}_t)$-adapted process with values in E_δ.*
3. *The time shift operators $\theta_t : \Omega \to \Omega$ fulfill*

$$X_t \circ \theta_s = X_{t+s}$$

   for all $s, t \in \mathbb{R}_{\geq 0}$.

4. *$(P^x)_{x\in E_\delta}$ is a family of probability measures on $(\Omega, \mathcal{F})$ with $P^\delta(X_0 = \delta) = 1$*

In this setup X (or respectively the tuple) is called homogeneous $(\mathcal{F}_t)_{t\geq 0}$-Markov process iff

1. *$x \mapsto P^x(X_t \in B)$ is measurable for each $t \in \mathbb{R}_{\geq 0}$ and $B \in \mathcal{E}$.*
2. *The Markov property holds in the sense that*

$$P^x(X_{t+s} \in B|\mathcal{F}_t) = P^{X_t}(X_s \in B)$$

   for all $x \in E_\delta, B \in \mathcal{E}_\delta$ and $t, s \in \mathbb{R}_{\geq 0}$.

The measures P^x should be understood as measures under which the Markov process starts at x, i.e. $X_0 = x$ P^x-a.s.. The Markov property means that the process at time $t+s$ is independent of the process up to time t. It only depends on the value of X at t and its evolution after this time. We will denote E^x as the expectation with respect to P^x. By constructing a canonical probability space we will be able to find a sufficiently rich space Ω. For the quantities we just defined it can be observed that

$$P_t(x, A) = P^x(X_t \in A)$$

defines a transition function. Compare with [18, Proposition 3.5]. This transition function is entirely determined by its values on $(E, \mathcal{E})$. We have for $A \in \mathcal{E}$ and $x \in E$:

$$P_t(x, A \cup \{\delta\}) = P_t(x, A) + P_t(x, \{\delta\}) = P_t(x, A) + 1 - P_t(x, E)$$

We define $Q_t(x, A) = P_t(x, A)$ for $A \in \mathcal{E}$ for the (possibly) submarkovian transition function restricted on $(E, \mathcal{E})$. More details about the last points can be found in [18]. The in some sense more general notion allows us to define killed Markov processes. We now relate this definition back to the more applicable ones.

Theorem 1.30. *A process X is a homogenous Markov process on a probability space $(\Omega, \mathcal{F}, P)$ with respect to its canonical filtration $\mathcal{F}_t^X$ with transition function (P_t) if one of the following equivalent conditions hold:*

1. *for any bounded Borel-measurable function $f : (E_\delta, \mathcal{E}_\delta) \to \mathbb{R}$ and $s < t$*

$$E[f(X_t)|\mathcal{F}_s^X] = P_{t-s}f(X_s) \quad P \text{ a.s.}$$

2. *for all $0 < t_1 < t_2 < \cdots < t_k$ and $f_i : E \to \mathbb{R}$, $i = 0, \ldots, k$ Borel-measurable and bounded it holds that there exists a Borel-measure ν such that*

$$\begin{aligned} E\left[\prod_{i=1}^{k} f_i(X_{t_i})\right] &= \int_E \nu(\mathrm{d}x) \int_E P_{t_1}(x, \mathrm{d}x_1) f_1(x_1) \int_E \cdots \int_E P_{t_k - t_{k-1}}(x_{k-1}, \mathrm{d}x_k) f_k(x_k) \\ &= E\left[\prod_{i=1}^{k-1} f_i(X_{t_i}) P_{t_k - t_{k-1}} f_k(X_{t_{k-1}})\right] \end{aligned}$$

   where $\nu(A) = P(X_0 \in A)$ is called the initial measure.

3. *If Z is a bounded random variable on $(\Omega, \mathcal{F}_\infty^X, P)$, then for every $t > 0$ we have*

$$E[Z \circ \theta_t \,|\, \mathcal{F}_t^X] = E_{X_t}[Z] \quad P \text{ -a.s.}$$

   where E_{X_t} is defined through 2. in the case of $\nu = \delta_{X_t}$

A proof can be found in [18]. The first property is the most known formulation of a Markov process. We will use it in the following chapter. The second property tells us that the law of the Markov process starting at x with probability ν is entirely defined by the transition function $(P_t)_{t\geq 0}$, since these fix the finite dimensional distributions. We can obtain the family of measures $(P^x)_{x\in E}$ from the second property by setting $\nu = \delta_x$, i.e. the unit mass at x. In this case all above expectations may also be indexed with an x to make this explicit. The third way of stating the property nicely admits the generalization to the strong Markov property. In general we are interested in the behaviour of Markov processes at stopping times, and we would like to have the Markov property also in this case.

Definition 1.31 (Strong Markov property)**.** *A process X on $(\Omega, \mathcal{F}_\infty^X, \mathcal{F}_t^X, P)$ is said to possess the strong Markov property, iff for all stopping times S on $(\Omega, \mathcal{F}_\infty^X, \mathcal{F}_t^X, P)$ and random variables Z on $(\Omega, \mathcal{F}_\infty^X, P)$:*

$$E[Z \circ \theta_S \,|\, \mathcal{F}_S^X] = E_{X_S}[Z] \quad P \text{ -as}. \tag{1.2}$$

Next we define killed Markov processes.

Definition 1.32. *Let $\zeta : (\Omega, \mathcal{F}, \mathcal{F}_t) \to \overline{\mathbb{R}}$ be a stopping time with respect to the filtration $\mathcal{F}_t$. We define for a Markov process X the killed Markov process as*

$$\hat{X}_t = \begin{cases} X_t & \text{for } t < \zeta \\ \delta & \text{for } t \geq \zeta \end{cases}$$

The random variable ζ is usually taken with a known distribution, for example an exponential distribution with parameter λ. In this case the parameter λ is also called death rate of the process. The random variable $\zeta(\omega) = \inf\{t \geq 0 \,|\, X_t(\omega) = \delta\}$ is called life time of the process. In the following we will have $\zeta = \inf\{t \geq 0 \,|\, X_t = 0\}$ for a real valued process X.

1.5. Brownian motion

In this section we want to introduce Brownian motion, a stochastic process with great importance for probability theory and statistics, and with similarities and connections to the scaling limit of random trees. We will heavily rely on its properties in chapter two, where we will infer the law of Brownian excursions from the transition function of Brownian motion. For the next definition compare with [16].

Definition 1.33 (Brownian motion)**.** *A stochastic process $B = (B_t)_{t\geq 0}$ of real valued random variables is called Brownian motion iff*

1. $B_0 = 0$ *a.s.*
2. *The increments $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_{k-1}}$ are independent for all $0 < t_1 < \cdots < t_k$ for $k \in \mathbb{N}$.*
3. *Increments follow a centered normal distribution with variance of the time increment, $\mathcal{L}(B_t - B_s) = N(0, t-s)$ for $0 \leq s < t$*
4. *$t \mapsto B_t$ is continuous almost surely.*

The first three properties fix the finite dimensional distributions of B. For details see [16]. By Kolmogorov's extension theorem there exists a unique probability measure P^B on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}^{[0,\infty)})$ such that the finite dimensional distributions coincide with above conditions. The difficult part is the fourth condition. Assume that a stochastic process X has law P^B. We can only guarantee continuity for a so called modification of X.

Definition 1.34 (Modifications)**.** *Two processes X, Y on the same probability space $(\Omega, \mathcal{F}, P)$ with identical state space and index set T are said to be modifications of each other iff*

$$P(X_t = Y_t) = 1 \quad \forall t \in T\,.$$

The continuity of Brownian motion is ensured by Kolmogorov's continuity criterion.

Theorem 1.35 (Kolmogorov's continuity criterion)**.** *Let X be a real valued process with index set $T \subseteq \mathbb{R}$ satisfying*

$$\mathrm{E}[|X_t - X_s|^\alpha] \leq C|t-s|^\beta$$

for all $s, t \in T$ and for $\alpha, C > 0, \beta > 1$. Then X has a modification Y of X which is almost surely continuous.

For a proof see [24] Chapter I, Section 2. One can show that a process that fulfills the first to third condition of Definition 1.33 also fulfills Kolmogorov's continuity criterion, and thus we can conclude:

Theorem 1.36. *There exists a modification of a process that fulfills Definition 1.33, 1. to 3. which is almost surely continuous. This modification is called Brownian motion.*

Proof. In the case of Brownian motion we can use the fourth moment of the normal distribution, $E[|B_t - B_s|^4] = 3(t-s)^2$ and the conditions of Kolmogorov's continuity criterion 1.35 are fulfilled. □

Since Brownian motion is almost surely continuous we can pass to the space $W := \mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R})$ as an underlying probability space. In the following we define Brownian motion as a random variable

$$B : (W, \mathcal{F}, P) \to (W, \mathcal{F}), \;\; w \mapsto B(w) = w\,,$$

where $\mathcal{F}$ is the Borel-σ-Algebra of uniform convergence and P is the so called Wiener measure, under which B is distributed like Brownian motion, i.e. B fulfills 1. to 3. in Definition 1.33. A detailed description of the construction of Wiener space, its σ-Algebra and the Wiener measure is given in the appendix. One important detail for the following chapter is the fact that the Wiener space contains all continuous functions, also those with $w(0) \neq 0$. However the Brownian motion starts at zero a.s.,

$$P(\{w \in W \,|\, w(0) = 0\}) = 1\,.$$

The advantage of not restricting the space to functions starting at zero lies in the existence of (time) shift operators.

Definition 1.37. *We define the for $t > 0$ the shift operator $\theta_t : W \to W$ such that*

$$B_s(\theta_t(w)) = B_{t+s}(w) = w(t+s) \text{ for all } s \in \mathbb{R}_{\geq 0}$$

With these we have

Proposition 1.38. *Brownian motion is a strong Markov process under P with transition function*

$$P_t(x, A) = \int_A p_t(x,y)\mathrm{d}y\,, \text{ with } p_t(x,y) = \frac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/2t}$$

In particular for P^x defined on W by

$$P^x(w(t_1) \in A_1, \ldots, w(t_k) \in A_k) = \int_{A_1} P_{t_1}(x, \mathrm{d}x_1) \int_{A_2} P_{t_2 - t_1}(x_1, \mathrm{d}x_2) \cdots \int_{A_k} P_{t_k - t_{k-1}}(x_{k-1}, \mathrm{d}x_k)$$

we have $P^x(B_0) = x$ a.s. and the process $(B_t)_{t \geq 0}$ under P^x is called the Brownian motion starting at x.

For a proof see [12, Theorem 12.1]. Due to the Markov property, Brownian motion has additional properties which are handy for explicit calculations.

Lemma 1.39 (Reflection principle of Brownian motion)**.** *Let T be an $\mathcal{F}_t^B$-stopping time, then the process*

$$\tilde{B}(t) = \begin{cases} B(t) \text{ for } t \leq T \\ 2B(T) - B(t) \text{ for } t > T \end{cases}$$

is a Brownian motion.

A proof can be found in [12, Theorem 13.1].

1.6. Excursions of Brownian motion and killed Brownian motion

In this section we define the killed Brownian motion and derive its distribution. This distribution is a crucial ingredient to specify the law of the normalized Brownian excursion. Before we define the killed process we fix some notation. Excursions of Brownian motion are special functions in Wiener space which leave the zero line for a positive but finite time and stay constant at zero once they hit zero again. The notion of an excursion will be delved in the next chapter.

Definition 1.40 (Space of excursions)**.** *We define the space of excursions of Brownian motion U to be the subspace*

$$U := \{w \in \mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R}) \,|\, 0 < \zeta(w) < \infty \text{ and } w(t) = 0 \text{ for } t \geq \zeta(w)\} \subseteq W$$

equipped with the metric

$$d_U(u_1, u_2) = \sup_{t \geq 0} |u_1(t) - u_2(t)| + |\zeta(u_1) - \zeta(u_2)|$$

for $u_1, u_2 \in U$.

We define $\mathcal{U} = \mathcal{B}(U)$ as the σ-algebra on U. U decomposes into positive (U_+) and negative (U_-) excursions and we all the corresponding σ-algebras $\mathcal{U}_+$ and $\mathcal{U}_-$. Additionally U decomposes into excursions of length r denoted as U_r. Due to Theorem B.4, the Borel-σ-algebra on U_r coincides with the σ-algebra induced by coordinate mappings. Again we do not fix the excursion to start at zero which enables us to define time shift operators θ_t on U. Furthermore we define $U_\delta = U \cup \delta$ and $\mathcal{U}_\delta = \sigma\{\mathcal{U}, \delta\}$, where δ is again an arbitrary "coffin state".

Corollary 1.41. *ζ is continous on U, in particular, ζ defines a random variable on U.*

Proof. This follows directly from the definition of d_U. □

Definition 1.42 (First Hitting Time)**.** *For any $x \in \mathbb{R}$ define*

$$T_x = \inf\{t > 0 \,|\, B_t = x\}$$

as the first hitting time of x. We always use the convention $\inf(\emptyset) = \infty$. These random variables are all stopping times ([24] I, Prop 4.5).

In the following we will consider a process which is killed once it is zero. In the special case of $x = 0$ we also write $\zeta := T_0$, which is a commonly used notation for the lifetime of a killed process.

Proposition 1.43 (Killed Brownian motion)**.** *We introduce the Brownian motion which is killed when it hits 0, which is a random variable on the Wiener space*

$$\hat{B} : W \to U_\delta,\, w \mapsto \left(t \mapsto \hat{B}_t(w) = \begin{cases} w(t) & \text{for} \quad t < \zeta(w) \\ \delta & \text{for} \quad t \geq \zeta(w) \end{cases}\right)$$

Under the measure P^x associated with the transition group of Brownian motion P_t, where P^x describes the law under which Brownian motion starts at x, $\hat{B}$ is a strong Markov process with submarkovian transition semi group $(Q_t)_{t \geq 0}$.

The semigroup Q_t on U_δ induces a measure Q_x on W, characterized by its finite dimensional distributions:

$$Q_x(w(t_1) \in A_1, \ldots, w(t_k) \in A_k) = \int_{A_1} Q_{t_1}(x, \mathrm{d}x_1) \int_{A_2} Q_{t_2 - t_1}(x_1, \mathrm{d}x_2) \cdots \int_{A_k} Q_{t_k - t_{k-1}}(x_{k-1}, \mathrm{d}x_k)$$

Using Q_x we in particular have by definition for $\Gamma \in \mathcal{U}$:

$$E[\mathbb{1}_\Gamma(\hat{B} \circ \theta_T) | \mathcal{F}_T] = E_{\hat{B}_T}[\mathbb{1}_\Gamma(\hat{B})] = Q_{B_T}(\Gamma) \tag{1.3}$$

which will be used later on.

Lemma 1.44 (Taboo transition density)**.** *The transition semi group of Brownian motion killed at* 0 *is given by* $Q_t(x, \mathrm{d}y) = q_t(x, y)\mathrm{d}y$ *with*

$$q_t(x, y) = p_t(x, y) - p_t(x, -y)$$

and

$$p_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/2t}$$

Proof. The following proof can be found in Rogers Vol. 1 I.13. [12]. The most important tool will be the reflection principle. We define the running supremum of Brownian motion given by $S_t = \sup\{B_s \,|\, s \leq t\}$. We then have for $a, y > 0$

$$P(S_t \geq a, B_t \leq a - y) = P(S_t \geq a,\, B_t \geq a + y) = P(B_t \geq a + y)\,,$$

where the reflection principle was applied at time T_a. Recall the law of Brownian motion starting at $x \in \mathbb{R}$, P^x. The common law is then for $a, x, y > 0$ given by

$$\begin{aligned} P^x(B_t \geq y,\, T_0 \leq t) &= P(B_t \geq y - x,\, T_{-x} \leq t) \\ &= P(B_t \leq x - y,\, T_x \leq t) \\ &= P(B_t \leq x - y,\, S_t \geq x) \\ &= P(B_t \geq x + y) = P^x(B_t \leq -y) \end{aligned}$$

This gives

$$\begin{aligned} P^x(B_t \geq y, T_0 > t) &= P^x(B_t \geq y) - P^x(B_t \geq y,\, T_0 \leq t) \\ &= P(B_t \geq y - x) - P(B_t \leq -y - x) \end{aligned}$$

Using $p_t(x, y) = \partial_y P(B_t \leq y - x)$ we obtain

$$q_t(x, y) = -\partial_y P^x(B_t \geq y,\, T_0 > t) = p_t(x, y) - p_t(x, -y)$$

Finally we know that by the definition of transition functions of a Markov process $Q_t(x, A) = P^x(\hat{B}_t \in A) = P^x(B_t \in A, T_0 > t)$ since the killed process coincides with Brownian motion until it hits zero, which proves the statement. □

1.7. Random walks

As an example of a time discrete Markov process which is important for this thesis we look at the simple random walk. It describes a process which after each step of unit time changes by ± 1 with equal probability.

Definition 1.45 (Random walk)**.** *Let $(X_i)_{i\in\mathbb{N}}$ be independent random variables homogeneously distributed variables on $\{-1,1\}$. We define the symmetric simple random walk $(S_n)_{n\in\mathbb{N}}$ as the partial sum process*

$$S_n = S_0 + \sum_{i=1}^{n} X_i$$

with $S_0 = 0$ P- almost surely. We denote P_l as the law under which the random walk starts at l, i.e. $S_0 = l$ P_l-a.s..

In the following we discuss some properties of random walks which we will makes use of later. The first two moments are given by

$$\mathrm{E}[S_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] = 0 \tag{1.4}$$

$$\mathrm{E}[S_n^2] = \sum_{i=1}^{n} \mathrm{E}[X_i^2] = \sum_{i=1}^{n} (1 \cdot 1/2 + 1 \cdot 1/2) = n$$

since all mixed terms $\mathrm{E}[X_i X_j]$ for $i \neq j$ vanish due to (1.4). The random walk has some particularly benefitial properties which are inherited by Brownian motion as a "limit" of a random walk.

Theorem 1.46 (Markov property)**.** *The random walk process obeys the Markov property, i.e. for $j_1 < \cdots < j_k < n$*

$$P(S_n = l | S_{j_1} = m_1, \ldots, S_{j_k} = m_k) = P(S_{n-j_k} = l - m_k)$$

Proof. We have

$$\begin{aligned} P(S_n = l | S_{j_1} = m_1, \ldots, S_{j_k} = m_k) &= P(S_{j_k} + \sum_{i=j_k+1}^{n} X_i = l | S_{j_1} = m_1, \ldots, S_{j_k} = m_k) \\ &= \frac{P(m_k + \sum_{i=j_k+1}^{n} X_i = l,\, S_{j_1} = m_1, \ldots, S_{j_k} = m_k)}{P(S_{j_1} = m_1, \ldots, S_{j_k} = m_k)} \\ &= P(\sum_{i=j_k+1}^{n} X_i = l - m_k) = P(S_{n-j_k} = l - m_k) \end{aligned}$$

where the third equality follows from $\sum_{i=j_k+1}^{n} X_i$ being independent of all $S_{j_1}, \ldots, S_{j_k}$. ☐

Lemma 1.47 (Reflection principle of the random walk)**.** *For $n \in \mathbb{N}$ it holds that $S_n \overset{d}{=} -S_n$, and more generally for $k \in \mathbb{N}$ and $n \geq k$*

$$P(S_k = l_1, S_n = l_1 + l_2) = P(S_k = l_1, S_n = l_1 - l_2)\,,$$

i.e. the process reflected at l_1 for $n \geq k$ has the same law as S_n.

Proof. By the definition of the random walk it is clear that $S_n \overset{d}{=} -S_n$. Consider without loss of generality $n > k$. It then holds due to the independence of the increments of the random walk:

$$\begin{aligned} P(S_k = l_1, S_n = l_1 + l_2) &= P(S_k = l_1)P(S_n - S_k = l_2) \\ &= P(S_k = l_1)P(S_n - S_k = -l_2) = P(S_k = l_1, S_n = l_1 - l_2) \end{aligned}$$

□

Lemma 1.48 (Time reversal)**.** *Let $a, b \in \mathbb{Z}$ and $k < n \in \mathbb{N}_0$. It holds that*

$$P(S_k = a,\ S_n = b) = P_b(S_{n-k} = a)\,,$$

where P_b is the probability measure such that $S_0 = b$ P_b-a.s..

Proof. For any set $x_1, x_2, \dots, x_{n-k} \in \{-1, 1\}$ with $a + \sum_{i=1}^{n-k} x_i = b$ it holds due to X_i being i.i.d. that

$$P(S_0 = 0, X_{k+1} = x_1, \dots, X_n = x_{n-k}) = P^b(S_0 = b, X_1 = -x_1, \dots, X_{n-k} = -x_{n-k})$$

and by summing over all such tupels we directly obtain the assertion. □

We can also consider the random walk as a process of continuous time. There are two ways to do this. We can either interpolate the points S_i, which yields a Lipschitz continuous process, or we define a piecewise constant process with increments of ± 1. More precisely we define, motivated by the following discussion in chapter three and five,

$$\begin{aligned} \tilde{s}_t^k &:= \frac{1}{\sqrt{2k}} S_{\lfloor 2kt \rfloor} = \frac{1}{\sqrt{2k}} \sum_{1 \le i \le \lfloor 2kt \rfloor} X_i \\ s_t^k &:= \frac{1}{\sqrt{2k}} \left(S_{\lfloor 2kt \rfloor} + (2kt - \lfloor 2kt \rfloor) X_{\lfloor 2kt \rfloor + 1} \right) . \end{aligned}$$

As it turns out there is no notable difference in the limit as $k \to \infty$.

$$P\left(\sup_{0 \le t \le 1} |\tilde{s}_t^k - s_t^k| > \varepsilon \right) = P\left(\sup_{0 \le t \le 1} \frac{2kt - \lfloor 2kt \rfloor}{\sqrt{2k}} > \varepsilon \right) \le P((2k)^{-1/2} > \varepsilon) \xrightarrow{k \to \infty} 0$$

2. Itō excursion theory

A point a is called recurrent, if it is visited by a stochastic process starting at a a.s. in finite time. For a Markov process and a recurrent point the time axis can be decomposed into the intervals between the visit times. The sample paths of the stochastic process on these intervals are then called excursions of the Markov process. The Japanese mathematician Kiyoshi Itō established in his paper [8] in 1972 that a probability theory of excursions attached to Markov processes can be described using the characteristic measure of a Poisson point process (PPP). In the special case of Brownian motion it is possible to deduce an explicit density of a normalized excursions which is the ultimate aim of this chapter.

2.1. Poisson point processes

A point process is a stochastic process which takes countably many values in the state space. Such a point process is called Poisson point process, if the number of points that appear in a chosen set and time are Poisson distributed. This kind of stochastic processes will be topic of the first section. The following description is inspired by [22] and [21] and also taken from [24].

Definition 2.1 (Point function). *Let $(U,\mathcal{U})$ be a measurable space with σ-algebra $\mathcal{U}$. We add a point to U: $U_\delta = U \cup \{\delta\}$ and name the corresponding σ algebra $\mathcal{U}_\delta = \sigma(\mathcal{U}, \{\delta\})$. A point function p is a measurable mapping which sends a countable set $D \subseteq (0,\infty)$ into U and the complement of D to δ:*

$$\begin{aligned} p : & D \to U, \ \ l \mapsto p(l) \in U \\ & D^c \to \{\delta\} \end{aligned}$$

The point function defines a counting measure $N_p(\mathrm{d}s, \mathrm{d}u)$ on $(0,\infty)\times\mathcal{U}$ with σ-field $\mathcal{B}(0,\infty)\otimes\mathcal{U}$ via

$$\begin{aligned} N_p((s,t] \times A) &= \sum_{l \in D,\, s<l\leq t} \mathbb{1}_A(p(l)), \quad \textit{for } t>0, A \in \mathcal{U} \\ &= \#\{s < l \leq t \,|\, p(l) \in A\} \end{aligned} \tag{2.1}$$

The counting measure simply counts the number of points in time which are mapped into a measurable subset of U. In principle, we would not need to restrict ourselves to rectangles $(s,t] \times A$ but it is sufficient in our case. A point process is a point function where the set D is taken at random.

Definition 2.2 (Point process). *A process $e = \{e_t, t > 0\}$ is said to be a point process on $(\Omega, \mathcal{F}, P)$ taking values in $(U_\delta, \mathcal{U}_\delta)$ iff*

1. *e is $\mathcal{B}((0,\infty)) \otimes \mathcal{F}\,/\,\mathcal{U}$- measurable*
2. *$D_\omega = \{l \,|\, e_l(\omega) \neq \delta\}$ is a.s. countable*

In other words $e.$ is a.s. a point function. We simplify the notation of the (random) counting measure and set

Definition 2.3. *For the counting measure defined as in* (2.1) *we denote for $\omega \in \Omega$*

$$N^A_{(s,t]}(\omega) := N_{e.(\omega)}((s,t] \times A)$$

and ommit in the notation s in the case of $s = 0$, i.e. $N^A_t := N^A_{(0,t]}$ and we may ommit A in the notation in the case of $A = U$. Similarly we define e^A as the point process into A with

$$e^A_t(\omega) = e_t(\omega) \text{ if } e_t(\omega) \in A$$
$$e^A_t(\omega) = \delta \text{ otherwise}$$

Definition 2.4. *A point process e is said to be σ discrete, if there exists a sequence $(U_n) \subset \mathcal{U}$, $U_n \uparrow U$ such that*

$$N^{U_n}_t < \infty \text{ a.s.}$$

for all $t \in \mathbb{R}_{>0}$ and $n \in \mathbb{N}$

We now give the definition of a Poisson Point process as in [24].

Definition 2.5 (Poisson Point process)**.** *A σ-discrete point process e is said to be $(\mathcal{F}_t)$-Poisson point process with respect to the filtration $(\mathcal{F}_t)$ if there exists a σ-finite measure n on $\mathcal{U}$ such that for any $s, t > 0$ and $\Gamma \in \mathcal{U}$ with $n(\Gamma) < \infty$ it holds*

$$P(N^\Gamma_{(s,s+t]} = k \,|\, \mathcal{F}_s) = \frac{t^k n(\Gamma)^k}{k!} \exp(-tn(\Gamma)) \tag{2.2}$$

The existence of such a process follows from Kolmogorovs extension theorem and the fact that (2.2) fixes the finite dimensional distributions of the process. The expectation of N^Γ_t is $tn(\Gamma)$, which follows from the Poisson law of N^Γ_t. The measure n is called characteristic measure or intensity measure of the Poisson point process.

2.2. Characteristic measure of a Poisson point process

The characteristic measure has great properties to study the probabilistic properties of the Poisson point process. The connection between the Poisson process and the characteristic measure is discussed in this section, most importantly by Lemma 2.8. We follow the description of [24] chapter XII.

Definition 2.6 (Characteristic measure)**.** *Let e be a $(\mathcal{F}_t)_{t>0}$ Poisson point process with counting measure N. The σ-finite measure n on $\mathcal{U}_\delta$ defined by*

$$n(\Gamma) = \frac{1}{t} E(N^\Gamma_t)$$

for $\Gamma \in \mathcal{U}$ with $E(N^\Gamma_t) < \infty$ and continued to $\mathcal{U}_\delta$ by $n(\{\delta\}) = 0$ is called characteristic measure of the Poisson point process.

Since N^Γ is Poisson-distributed, $n(\Gamma)$ is independent of t. This measure uniquely determines the point process and is the fundamental object in Itō excursion theory. To understand its importance we first observe that the first jump times are exponentially distributed.

Lemma 2.7. *Define for any $\Gamma \in \mathcal{U}$ with $n(\Gamma) < \infty$ the first jump time in Γ as*

$$T_\Gamma = \inf\{t > 0 \,|\, N_t^\Gamma = 1\} \,.$$

Then T_Γ is an exponentially distributed random variable with parameter $n(\Gamma)$.

Proof. Since N_t^Γ is $(\mathcal{F}_t)$-adapted, T_Γ is a $(\mathcal{F}_t)$-stopping time. We then have with (2.2)

$$P(T_\Gamma > t) = P(N_t^\Gamma = 0) = \exp(-n(\Gamma)t)$$

because of the Poisson distribution of N_t^Γ. This shows that T_Γ is exponentially distributed. □

The next lemma is crucial for the meaning of the characteristic measure. It allows us to define a probability measure on subsets of $\mathcal{U}$ on which n is finite:

Lemma 2.8. *For $\Lambda \in \mathcal{U}$ with $n(\Lambda) < \infty$ consider the first jump time $T_\Lambda = \inf\{t > 0 \,|\, N_t^\Lambda > 0\}$. In this case, T_Λ, e_{T_Λ} are independent random variables and for $\Gamma \in \mathcal{U}$ we have*

$$P(e_{T_\Lambda} \in \Gamma) = \frac{n(\Gamma \cap \Lambda)}{n(\Lambda)} \tag{2.3}$$

Proof. The statement and proof is taken from [24]. We distinguish three cases. If $\Gamma \cap \Lambda = \emptyset$ then any excursion e_{T_Λ} will not lie in Γ and we plainly have

$$P(e_{T_\Lambda} \in \Gamma) = 0 \,.$$

If $\Gamma \subseteq \Lambda$ we consider the Poisson processes $N_\cdot^\Gamma$ and $N_\cdot^{\Gamma^c \cap \Lambda}$. Let's denote the first jump times of N_t^Γ by T_Γ and of $N_t^{\Gamma^c\cap\Lambda}$ by $T_{\Gamma^c\cap\Lambda}$. Since $\Gamma \subseteq \Lambda$ we always have $T_\Lambda \le T_\Gamma$. On the other hand $e_{T_\Lambda} \in \Gamma$ implies $T_\Gamma \le T_\Lambda \le T_{\Gamma^c\cap\Lambda}$. Thus, using the exponential distribution of the first jump times we have:

$$\begin{aligned}
P(T_\Lambda > t, e_{T_\Lambda} \in \Gamma) &= P(t < T_\Gamma < T_{\Gamma^c\cap\Lambda}) \\
&= \int_t^\infty n(\Gamma)e^{-n(\Gamma)s} \int_s^\infty n(\Gamma^c \cap \Lambda)e^{-n(\Gamma^c\cap\Lambda)s'} \mathrm{d}s' \mathrm{d}s \\
&= \int_t^\infty n(\Gamma)e^{-n(\Gamma)s} e^{-n(\Gamma^c\cap\Lambda)s} \mathrm{d}s \\
&= \frac{n(\Gamma)}{n(\Gamma) + n(\Gamma^c \cap \Lambda)} e^{-n(\Gamma)t - n(\Gamma^c\cap\Lambda)t} = \frac{n(\Gamma)}{n(\Lambda)} e^{-n(\Lambda)t}
\end{aligned}$$

and we obtain the lemma from $n(\Gamma) + n(\Gamma^c \cap \Lambda) = n(\Lambda)$ from the σ-additivity of measures. Finally, in the general case we have

$$\begin{aligned}
P(t < T_\Lambda,\, e_{T_\Lambda} \in \Gamma) &= P(t < T_\Lambda,\, e_{T_\Lambda} \in \Gamma \cap \Lambda) + P(t < T_\Lambda,\, e_{T_\Lambda} \in \Gamma \setminus \Lambda) \\
&= \frac{n(\Gamma \cap \Lambda)}{n(\Lambda)} e^{-n(\Lambda)t} + 0
\end{aligned}$$

According to the previous lemma T_Λ is exponentially distributed with parameter $n(\Lambda)$, and thus T_Λ, e_{T_Λ} are independent random variables and e_{T_Λ} has the conjectured probability. □

This probability measure can be seen as a conditional version of the excursion measure, so

$$P(e_{T_\Lambda} \in \Gamma) =: n(\Gamma \,|\, \Lambda)$$

2.3. Local time and the space of excursions

The following section applies in very similar manner to general Markov processes, see [8]. However, we will focus on the case of Brownian motion, which is a Markov process with zero as one recurrent point. The canonical probability space for Brownian motion is the Wiener space $W = \mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R})$ of continuous functions on the non negative real line endowed with the Borel σ-algebra induced by uniform convergence and the Wiener measure P. We will view all random variables as random variables on W, in particular for $w \in W$ we write $B_t(w) = w(t)$ and have shift operators $\theta_t(w)(\cdot) = w(t + \cdot)$. We first define the length of the first excursion of Brownian motion,

$$\zeta(w) = \inf\{t > 0 \,|\, w(t) = 0\}\,.$$

Excursions are elements of W with $0 < \zeta(w) < \infty$ and $w(t) = 0\,\forall t > \zeta(w)$. Out aim in this section is the construction of a space of excursions as a subset of Wiener space and to define a meaningful measure on it. However, it holds that $\zeta(w) = 0$ a.s., so elements of the Wiener space with such properties are null sets with respect to the Wiener measure. We will use local time to decompose the time axis into its excursion intervals.

Definition 2.9 (Local time)**.** *The local time $(L_t^0)_{t\geq 0}$ of reflected Brownian motion $|B_t|$ at level 0 is a stochastic process defined by the approximation*

$$L_t^0 := \lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{[0,\varepsilon]}(|B_s|)\mathrm{d}s\,. \tag{2.4}$$

Compare [24, page 227]. The fact that this process is a.s. non trivial can be derived with the Itō integration formula, see for example [10]. This is not a trivial statement, as the set of zeroes of Brownian motion has very unintuitive properties.

Theorem 2.10. *The set of zeroes of Brownian motion $Z = \{t \,|\, B_t = 0\}$ is a.s. a closed set of Lebesgue measure zero without isolated points.*

For a proof of this statement see [24, Proposition 3.12]. Consequently it is hard to imagine how this set looks like, or how local time behaves for a given Brownian motion. Nevertheless this theorem allows us to define a meaningful rightinverse of local time.

Lemma 2.11. *The time change of local time*

$$\tau_l = \inf\{t \geq 0 \,|\, L_t^0 > l\}\,.$$

is a right continuous monotone function with existing left limits. In particular, the set of discontinuity times l is countable. We will denote the left limit of τ at l as

$$\tau_{l-} = \inf\{t \geq 0 \,|\, L_t^0 \geq l\} = \lim_{l'\uparrow l} \tau_{l'}\,.$$

I do not want to delve this discussion here but give an intuition why the set of discontinuity times is countable. Consider the random set $\mathcal{O}(w) = \bigcup_l (\tau_{l-}(w), \tau_l(w))$. The intervals are empty except at discontinuity times $D = \{l \,|\, \tau_l \neq \tau_{l-}\}$, where they correspond precisely to the excursion intervals (τ_{l-}, τ_l) for $l \in D$. Now one can show, as was done in [24] VI, proposition 2.5 that $Z(w) = \mathcal{O}(w)^C$ a.s.. $Z(w)$ is a closed set and thus $\mathcal{O}(w)$ is open in $\mathbb{R}$. This means that $\mathcal{O}(w)$ is a countable union of open disjoint intervals a.s.. This discussion implies that a Brownian motion can be decomposed into countably many excursions.

2.4. Itō excursion measure of Brownian motion

In the following we use the previously introduced notation

$$U := \{w \in \mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R}) \,|\, 0 < \zeta(w) < \infty \text{ and } w(t) = 0 \text{ for } t \geq \zeta(w)\} \subseteq W$$

and equip it for now with the σ-algebra generated by the coordinate mappings, denoted by $\mathcal{U}$. For later purposes we want to construct a probability law of the excursion of Brownian motion. As mentioned before the difficulty of ζ being zero almost surely arises. To solve this we aim to do the following. Whenever an excursion appears in the sample path of Brownian motion, we label it with a local discontinuity time $l \in D$. The map from such labels to the set of excursions defines a Poisson point processes. Consequently we can define an characteristic (excursion) measure which tells us the average number of excursions in $\Gamma \in \mathcal{U}$ that appear in Brownian motion. Remarkably the characteristic measure has exactly the probabilistic interpretation we seek, which is indicated by Theorem 2.15.

The excursion process of Brownian motion

Definition 2.12 (Excursion process)**.** *Let $D_w = \{l \in \mathbb{R}_{\geq 0} \,|\, \tau_l(w) \neq \tau_{l-}(w)\}$ be the random set of discontinuities of the time change of Brownian motion, which is a.s. countable. The excursion process of Brownian motion $(e_l, l \in \mathbb{R}_+)$ is a random point process $e : C(\mathbb{R}_+, \mathbb{R}) \times \mathbb{R}_{\geq 0} \to U_\delta$ with*

$$e(w) : \begin{cases} D_w \to U,\ l \mapsto e_l(w) \\ D_w^c \to \{\delta\} \end{cases} \tag{2.5}$$

where

$$e_l(w)(t) = \mathbb{1}_{[\tau_{l-}(w), \tau_l(w)]}(\tau_{l-}(w) + t) B_{\tau_{l-}(w)+t}(w) \tag{2.6}$$

We will write $u(t) = e_l(w)(t)$ for the excursion element, if $e_l(w) \in U$.

At each discontinuity time l of $\tau(w)$ the excursion process $e(w)$ "cuts" out the excursion of w which starts at time $\tau_{l-}(w)$. Note that the excursion process is indexed by local time (l), while the excursions themselves are functions in real time (t). As usual we denote the counting measure for $\Gamma \in \mathcal{U}$ by N^Γ.

Theorem 2.13 (Itō)**.** *The process e is a σ-discrete $(\mathcal{F}_t)$-Poisson point process*

The proof for this statement can be found in a much more general case in [8] section 6, using the fact that B_t is a strong Markov process. We will just sketch some points: To show that the mapping $(l, w) \mapsto e_l(w)$ is $\mathcal{B}((0, \infty)) \otimes \mathcal{F}$-measurable consider for arbitrary $r > 0$ and $e_l(w) \neq \delta$:

$$(l, w) \mapsto e_l(w)(r) \tag{2.7}$$

This mapping is measurable since τ_l, τ_{l-} and $B_{\tau_l + r}$ are measurable with respect to w. The σ-Algebra $\mathcal{U}$ is in fact also the smallest σ-Algebra in which all coordinate mappings are measurable. Hence (2.7) is already sufficient to prove measurability.
Secondly the set $D_w = \{l \in \mathbb{R}_+ \,|\, e_l(w) \neq \delta\}$ is a.s. countable. This is equivalent to say that the set of times l at which $\tau_l(w)$ is discontinuous is a.s. countable, which was motivated in Lemma

2.11.
For σ-discreteness consider the sets of excursions with length greater $1/n$ for $n \in \mathbb{N}$.

$$U_n = \left\{ u \in U \,|\, \zeta(u) > \frac{1}{n} \right\} < \infty \tag{2.8}$$

Then clearly

$$\bigcup_{i=1}^{\infty} U_i = U \tag{2.9}$$

and

$$N_l^{U_n}(w) = \sum_{s \leq l} \mathbb{1}_{U_n}(e_s(w)) < n\tau_l \tag{2.10}$$

since each excursion of w in U_n has at least length $1/n$. Note that the Poisson process N_l is indexed by local time, which corresponds to the real time less or equal τ_l.
It is a result of Itō [8] that N_t^Γ actually defines a Poisson process. This fact does not rely on any details of the law of Brownian motion, but only uses the fact that Brownian motion is a Markov process. See also [24, XII Def. 1.8].

Definition 2.14 (Itō excursion measure)**.** *The excursion measure of Brownian motion is the characteristic measure of the excursion point process, i.e. for the counting measure N induced by the excursion process of Brownian motion e, and for $\Gamma \in \mathcal{U}$ with $E[N_t^\Gamma] < \infty$ we define*

$$n(\Gamma) = \frac{1}{t} E[N_t^\Gamma] \,. \tag{2.11}$$

Since N_t^Γ is Poisson distributed, n is independent of t.

Markovian description of the excursion law

In this subsection we want to give an explicit integral formula for the finite dimensional distributions of a positive excursion under n. These will indicate a relation of the excursion process to the underlying Markov process. Let Q_x be the law of a Brownian motion killed at zero, starting at $x \in \mathbb{R}$. Furthermore define the so called entrance law of the excursion process as a Borel measure

$$\lambda_t(A) := n(u(t) \in A) = n(u(t) \in A, \zeta(u) > t)$$

for $A \in \mathcal{B}(\mathbb{R}_{>0})$. λ_t is called an entrance law of the transition semi group $(Q_t)_{t\geq 0}$ if it holds that $\int_{\mathbb{R}} \lambda_t(\mathrm{d}y) Q_s(y, A) = \lambda_{t+s}(A)$. The reason that we encounter an entrance law instead of a transition function and that excursion theory is difficult generally speaking lies in the short time behaviour of Brownian motion starting at zero. Because of this we investigate an excursion conditioned to live up to time t and let it run "freely" from there, governed by the transition function of killed Brownian motion. This will be the idea of the following reasoning.

Theorem 2.15 (Itō description of the excursion law)**.** *It holds for $A \in \mathcal{B}(\mathbb{R}_{>0})$ and $\Gamma \in \mathcal{U}_+$*

$$n(\zeta(u) > t \,, u(t) \in A \,, u \circ \theta_t \in \Gamma) = \int_A \lambda_t(\mathrm{d}x) Q_x(\Gamma) \tag{2.12}$$

In particular we have for $A_i \in \mathcal{B}(\mathbb{R}_{>0})$, $i = 1, \dots, k$ and $0 < t_1 < \cdots < t_k$:

$$n(u(t_1) \in A_1, \dots, u(t_k) \in A_k) = \int_{A_1} \lambda_{t_1}(\mathrm{d}x_1) \int_{A_2} Q_{t_2 - t_1}(x_1, \mathrm{d}x_2) \cdots \int_{A_k} Q_{t_k - t_{k-1}}(x_{k-1}, \mathrm{d}x_k) \tag{2.13}$$

Proof. Consider a measurable set $\Gamma \in \mathcal{U}_+$ and $A \in \mathcal{B}(\mathbb{R}_{>0})$. We have $0 < n(\zeta > t) < \infty$ for $t > 0$ and can use lemma 2.8 with the notations

$$\Lambda = \{u \,|\, \zeta(u) > t\}$$
$$l_\Lambda = \inf\{l \geq 0 \,|\, N_l^\Lambda = 1\}\,.$$

The first jump time was denoted by l_Λ to distinguish it as a local time from real times. We then have:

$$\frac{n(u(t) \in A,\, \zeta(u) > t,\, \theta_t(u) \in \Gamma)}{n(\zeta(u) > t)} = P(e_{l_\Lambda} \in \theta_t^{-1}(\Gamma), e_{l_\Lambda}(t) \in A) \tag{2.14}$$

where P denotes the Wiener measure. We can write above probability with real random time $T_\Lambda := \tau_{l_\Lambda -} + t$ as

$$P(e_{l_\Lambda} \in \theta_t^{-1}(\Gamma),\, e_{l_\Lambda}(t) \in A) = P(\{w \,|\, B_{T_\Lambda}(w) \in A,\; \hat{B} \circ \theta_{T_\Lambda}(w) \in \Gamma\})\,,$$

where $\hat{B}$ denotes the killed Brownian motion. Next we use the strong Markov property of $\hat{B}$. For this note that the time T_Λ is a $(\mathcal{F}_t)_{t\geq 0}$ stopping time, which makes $\mathbb{1}_A(B_{T_\Lambda})$ $\mathcal{F}_{T_\Lambda}$-measurable. This gives with the strong Markov property (1.3)

$$\begin{aligned} P(\{w \,|\, B_{T_\Lambda}(w) \in A,\; \hat{B} \circ \theta_{T_\Lambda}(w) \in \Gamma\}) &= E[\mathbb{1}_A(B_{T_\Lambda})\mathbb{1}_\Gamma(\hat{B} \circ \theta_{T_\Lambda})] \\ &= E[\mathbb{1}_A(B_{T_\Lambda})Q_{B_{T_\Lambda}}(\Gamma)] \\ &= \int_{B_{T_\Lambda}^{-1}(A)} P(dw)Q_{B_{T_\Lambda}(w)}(\Gamma) \\ &= \int_A \gamma_\Lambda(\mathrm{d}x)Q_x(\Gamma) \end{aligned} \tag{2.15}$$

with the measure $\gamma_\Lambda(\mathrm{d}x) = P(B_{T_\Lambda} \in \mathrm{d}x)$. Note that the set Λ only depends on t. Next consider the case $\Gamma = \{u(0) \in C\}$ for $C \in \mathcal{B}(\mathbb{R}_{>0})$. In this case $Q_x(\Gamma) = \mathbb{1}_C(x)$. We have by the definition of the entrance law and (2.15)

$$\begin{aligned} \lambda_t(C) &= n(\zeta(u) > t,\, u(t) \in \mathbb{R},\, u \in \theta_t^{-1}(\Gamma)) \\ &= n(\zeta(u) > t)\gamma_{\{\zeta(u)>t\}}(C) \end{aligned}$$

which finally yields

$$\begin{aligned} n(\zeta(u) > t\,, u(t) \in A,\, \theta_t(u) \in \Gamma) &= n(\zeta(u) > t) \int_A \gamma_\Lambda(\mathrm{d}x)Q_x(\Gamma) \\ &= \int_A \lambda_t(\mathrm{d}x)Q_x(\Gamma) \end{aligned}$$

The second assertion follows in the special case of $\Gamma = \{u \in U \,|\, u(t_2 - t_1) \in A_2, \ldots, u(t_k - t_1) \in A_k\}$, which is a measurable set as a union of cylindersets. In this case

$$\begin{aligned} n(u(t_1) \in A_1, \ldots, u(t_k) \in A_k) &= n(u(t_1) \in A_1, u \in \theta_{t_1}^{-1}(\Gamma)) \\ &= \int_{A_1} \lambda_{t_1}(\mathrm{d}x)Q_x(\Gamma) \\ &= \int_{A_1} \lambda_{t_1}(\mathrm{d}x) \int_{A_2} Q_{t_2-t_1}(x, \mathrm{d}x_2)\cdots \int_{A_k} Q_{t_k-t_1-t_{k-1}+t_1}(x_{k-1}, \mathrm{d}x_k) \end{aligned}$$

□

Implicitly we only considered positive excursions in all calculations. In many cases the restriction of n to U_+ is called n_+. Alternatively one can use $|B|$ as an underlying Markov process. In this case, for example in [15], the normalization of the characteristic measure n_+ changes by a factor of two to yield the same excursion measure as we just discussed.

Corollary 2.16. *For Borel-measurable functions $f_i : \mathbb{R} \to \mathbb{R}$ and times $t_1, \dots, t_k$, $i = 1, \dots, k$ it holds that*

$$\int_U n(\mathrm{d}u) \prod_{i=1}^{k+1} f_i(u(t_i)) = \int_U n(\mathrm{d}u) \prod_{i=1}^{k} f_i(u(t_i)) \int_{A_{k+1}} Q_{t_{k+1}-t_k}(x_k, \mathrm{d}x_{k+1}) f_{k+1}(x_{k+1}) \tag{2.16}$$

Proof. Using the transformation rule and the measurability of the projection mappings we have

$$\begin{aligned}
\int_U n(\mathrm{d}u) \prod_{i=1}^{k+1} f_i(u(t_i)) &= \int_{A_1} \cdots \int_{A_{k+1}} n(u(t_1) \in \mathrm{d}x_1, \theta_{t_1}(u)(t_i - t_1) \in \mathrm{d}x_i, i = 1 \dots k+1) \prod_{i=1}^{k+1} f_i(x_i) \\
&\overset{(2.13)}{=} \int_{A_1} \cdots \int_{A_{k+1}} \lambda_{t_1}(\mathrm{d}x_1) Q_{t_2-t_1}(x_1, \mathrm{d}x_2) \dots Q_{t_{k+1}-t_k}(x_k, \mathrm{d}x_{k+1}) \prod_{i=1}^{k+1} f_i(x_i) \\
&= \int_{A_1} \cdots \int_{A_k} \lambda_{t_1}(\mathrm{d}x_1) Q_{t_2-t_1}(x_1, \mathrm{d}x_2) \dots Q_{t_k-t_{k-1}}(x_{k-1}, \mathrm{d}x_k) \\
&\qquad \times \prod_{i=1}^{k} f_i(x_i) \int_{A_{k+1}} Q_{t_{k+1}-t_k}(x_k, \mathrm{d}x_{k+1}) f_{k+1}(x_{k+1}) \\
&= \int_{A_1} \cdots \int_{A_k} n(u(t_1) \in \mathrm{d}x_1, \theta_{t_1}(u)(t_i - t_1) \in \mathrm{d}x_i, i = 1 \dots n) \\
&\qquad \times \prod_{i=1}^{k} f_i(x_i) \int_{A_{k+1}} Q_{t_{k+1}-t_k}(x_k, \mathrm{d}x_{k+1}) f_{k+1}(x_{k+1}) \\
&= \int_U n(\mathrm{d}u) \prod_{i=1}^{k} f_i(u(t_i)) \int_{A_{k+1}} Q_{t_{k+1}-t_k}(x_k, \mathrm{d}x_{k+1}) f_{k+1}(x_{k+1})
\end{aligned}$$

□

In the next step we want to determine λ_t, for which we need the previous result.

Corollary 2.17. *λ_t has density*

$$l_t(y) = \frac{1}{\sqrt{2\pi t^3}} |y| \exp(-y^2/2t)$$

For the proof we will need to condition the excursion law n on random events. The following proof is taken from [24]. For this we introduce the following notation. The up and downcrossing times before and after t:

$$g_t = \sup\{s < t | B_s = 0\} \qquad d_t = \inf\{s > t | B_s = 0\}$$

as well as the age of an excursion

$$A_t = t - g_t \,.$$

We furthermore set $u_0(w) \in U$ as the unique elements which fulfills

$$u_0(t) = w(t) \text{ for } t < \zeta(w) \text{ and } u(t) = 0 \text{ for } t \geq \zeta(w)$$

and

$$u_s(w) = u_0(\theta_s(w)) \,.$$

With this we state

Proposition 2.18. *Let $T_A = \inf\{t > 0 \,|\, u(t) \in A\}$ be a $\mathcal{F}_t^0$ stopping time. Let f be a real valued Borel function on U. In this case we have*

$$n[f\mathbb{1}_{\zeta > T_A}] = E[f(u_{g_{T_A}})\mathbb{1}_{(0<g_{T_A}<T_A)}]n(\zeta > T_A)$$

Remember that $n[]$ denotes an integral with respect to n. The proof requires techniques which would go beyond the scope of this text. Our proposition is a special case of the proof of Proposition 3.5, Chapter XII in [24]. The proposition gives us the following insight: Not every excursion lives long enough to experience a non trivial event at a stopping time T. The excursion u_{g_T} on the other hand always fulfills $\zeta(u_{g_t}) > T$. We can thus interpret above formula as the definition of a excursion measure conditioned on the random event $\zeta(u) > T$. With this we now have the following for our excursion law.

Proof. of corollary 2.17
It is enough to compute the law on the sets $[y, \infty)$, $y > 0$, which generate the Borel σ-algebra $\mathcal{B}(\mathbb{R}_{>0})$. Define for $y > \varepsilon > 0$ the stopping time $T_\varepsilon = \inf\{t > 0 | u(t) > \varepsilon\}$. We then have

$$\begin{aligned}\lambda_t([y,\infty)) &= n(u(t) \in [y,\infty), T_\varepsilon < t)\\ &= n[\mathbb{1}_{\{T_\varepsilon<t\}}\mathbb{1}_{\{\zeta>T_\varepsilon\}}\mathbb{1}_{[y,\infty)}(u(t))]\\ &\overset{2.16}{=} n[\mathbb{1}_{\{T_\varepsilon<t\}}\mathbb{1}_{\{\zeta>T_\varepsilon\}}Q_{t-T_\varepsilon}(u(T_\varepsilon),[y,\infty))]\,,\end{aligned}$$

where it was used that 2.16 also holds for stopping times, [24]. Using $u(T_\varepsilon) = \varepsilon$ and the proposition just stated this is equal to

$$\lambda_t([y,\infty)) = E[\mathbb{1}_{\{\tilde{T}_\varepsilon<t\}}Q_{t-\tilde{T}_\varepsilon}(\varepsilon,[y,\infty))]]n(\zeta(u) > T_\varepsilon)$$

with $\tilde{T}_\varepsilon(w) = T_\varepsilon(u_{g_{T_\varepsilon}}(w))$ and the expectation being taken with respect to the Wiener measure. We furthermore use that $n(\zeta > T_\varepsilon) = \frac{1}{2\varepsilon}$. Again the proof can be found in [24] chapter XII Proposition 3.6. Then with the taboo transition density q_t associated with Q_t we have

$$\begin{aligned}\lambda_t([y,\infty)) &= E\left[\mathbb{1}_{\{\tilde{T}_\varepsilon<t\}}\int_y^\infty \frac{1}{\sqrt{2\pi(t-\tilde{T}_\varepsilon)}}\frac{e^{-(x-\varepsilon)^2/2(t-\tilde{T}_\varepsilon)} - e^{-(x+\varepsilon)^2/2(t-\tilde{T}_\varepsilon)}}{2\varepsilon}\mathrm{d}x\right]\\ &= \int_0^t \left(\frac{\Phi_{t-s}(y-\varepsilon) - \Phi_{t-s}(y+\varepsilon)}{2\varepsilon} + \Phi'_{t-s}(y) - \Phi'_{t-s}(y)\right)P(T_\varepsilon \in \mathrm{d}s)\end{aligned}$$

where $\Phi_t(y) = \int_y^\infty p_t(x,0)\mathrm{d}x$. We then have by Taylors formula applied in y for ε small enough:

$$\begin{aligned}\int_0^t \left|\frac{\Phi_{t-s}(y-\varepsilon) - \Phi_{t-s}(y+\varepsilon)}{2\varepsilon} + \Phi'_{t-s}(y)\right| P(\tilde{T}_\varepsilon \in \mathrm{d}s) &= \int_0^t \left|\frac{2\Phi^{(3)}_{t-s}(y)\varepsilon^2}{6} + o(\varepsilon^2)\right| P(\tilde{T}_\varepsilon \in \mathrm{d}s)\\ &\le 1 \cdot C\varepsilon^2 \xrightarrow{\varepsilon\to 0} 0\end{aligned}$$

for a constant $C > 0$. Note that $\sup_{s<t}\Phi^{(3)}_{t-s}(y) < C'$ since Φ is a Schwarz function in y and $1/(t-s)$. It is crucial that $y \neq 0$, where Φ_{t-s} diverges as $s \to t$. On the other hand $T_\varepsilon \to 0$ a.s. and thus also in distribution, such that

$$\int_0^t -\Phi'_{t-s}(y)P(T_\varepsilon \in \mathrm{d}s) \xrightarrow{\varepsilon\to 0} \int_0^t -\Phi'_{t-s}(y)\delta_0(\mathrm{d}s) = p_t(y,0)$$

Thus the density of the entrance law becomes

$$l_t(y) = -\partial_y\lambda_t([y,\infty) = -\partial_y p_t(y,0) = \frac{|y|}{\sqrt{2\pi t^3}}e^{-y^2/2t}$$

□

2.5. Normalized Brownian excursions

In the next step we want to define the law of the normalized Brownian excursion. In order to do this we need some more results of the theory of Brownian motion and local times. The set of excursions of length one is an n nullset, so we need to find a way on how to condition on this event.

Proposition 2.19 (Scaling properties of Brownian motion and local time)**.** *Define the scaling operator for* $\lambda > 0$

$$\Phi_\lambda : (w(t))_{t\geq 0} \mapsto (\lambda^{1/2} w(t/\lambda))_{t\geq 0}\,. \tag{2.17}$$

Then

$$B_t \overset{d}{=} \Phi_\lambda^{-1}(B_t) = \lambda^{-1/2} B_{\lambda t} \tag{2.18}$$

and

$$L_t \overset{d}{=} \lambda^{-1/2} L_{\lambda t} \tag{2.19}$$

A proof can be found in many books on this subject, for example [24]. The scaling of local time can be derived naively from the approximation formula (2.4) and the scaling of Brownian motion (2.18). These scaling properties now translate to the excursion measure.

Proposition 2.20 (Scaling properties of the excursion law)**.** *For the excursion law it holds*

$$n(\Gamma) = \lambda^{-1/2} n(\Phi_\lambda^{-1}(\Gamma)) \tag{2.20}$$

Proof. The idea of this proof was taken from [11]. An excursion of w starts at a real time $t \in \mathbb{R}_{\geq 0}$ iff the rescaled excursion of $\Phi_\lambda^{-1}(w)$ starts at t/λ, since $w(t) = w(\lambda t/\lambda) = \lambda^{1/2}\Phi_\lambda^{-1}(w)(t/\lambda)$. This gives for any $w \in W$:

$$\begin{aligned} N_l^\Gamma(w) &= \#\{t \in \mathbb{R}_{\geq 0} \,:\, L_t(w) \leq l,\, e_{L_t}(w) \in \Gamma\} \\ &= \#\{t \in \mathbb{R}_{\geq 0} \,:\, L_t(w) \leq l,\, e_{L_{t/\lambda}} \circ \Phi_\lambda^{-1}(w) \in \Phi_\lambda^{-1}(\Gamma)\} \end{aligned}$$

Due to the scaling property of local time and Brownian motion both applied to the event $L_t \leq l$ we now have the identity in distribution:

$$\begin{aligned} N_l^\Gamma &\overset{d}{=} \#\{t \in \mathbb{R}_{\geq 0} \,:\, \lambda^{1/2} L_{t/\lambda} \circ \Phi_\lambda^{-1} \leq t,\, e_{L_{t/\lambda}} \circ \Phi_\lambda^{-1} \in \Phi_\lambda^{-1}(\Gamma)\} \\ &= \#\{t' \in \mathbb{R}_{\geq 0} \,:\, L_{t'} \circ \Phi_\lambda^{-1} \leq \lambda^{-1/2} t,\, e_{L_{t'}} \circ \Phi_\lambda^{-1} \in \Phi_\lambda^{-1}(\Gamma)\} \\ &= N_{\lambda^{-1/2}t}^{\Phi_\lambda^{-1}(\Gamma)} \circ \Phi_\lambda^{-1} \overset{d}{=} N_{\lambda^{-1/2}t}^{\Phi_\lambda^{-1}(\Gamma)} \end{aligned}$$

where the last equality in distribution again holds due to the scaling property of Brownian motion. Hence it follows for the expectations:

$$n(\Gamma) = \frac{1}{t} E[N_t^\Gamma] = \frac{\lambda^{-1/2}}{\lambda^{-1/2} t} E[N_{\lambda^{-1/2}t}^{\Phi_\lambda^{-1}(\Gamma)}] = \lambda^{-1/2} n(\Phi_\lambda^{-1}(\Gamma))$$

□

We now prove the following corollary, which is exercise 2.13 in [24] chapter XII. We define the space of normalized excursions $U^1 = \{u \in \mathcal{C}([0,1], \mathbb{R}_{\geq 0}) \,|\, \zeta(u) = 1\}$. We endow U^1 with the Borel σ-algebra of uniform convergence $\mathcal{U}^1$. This is consistent with our previous choice of σ-algebra, see also Theorem B.4. The scaling operator defines a map

$$s_1 : U_+ \to U^1, u \mapsto s_1(u) = \Phi^{-1}_{\zeta(u)}(u) \tag{2.21}$$

which maps any positive excursion to an excursion of length one while preserving its shape. Unlike Φ the map is not injective.

Corollary 2.21 (Law of the normalized Brownian excursions)**.** *Let* $\Gamma \in \mathcal{U}^1$. *The law of normalized excursions is the probability measure*

$$n_{(1)}(\Gamma) = \frac{n(s_1^{-1}(\Gamma) \cap \{u \in U_+ : \zeta(u) \geq c\})}{n(\{u \in U_+ : \zeta(u) \geq c\})} \tag{2.22}$$

and is independent of $c \in \mathbb{R}_{>0}$.

Proof. The set $s_1^{-1}(\Gamma)$ is invariant under scaling, i.e. $\Phi_c^{-1}(s_1^{-1}(\Gamma))) = s_1^{-1}(\Gamma)$, since by definition

$$\begin{aligned} s_1(\Phi_c(u)) &= s_1(c^{1/2}u(c\,\cdot)) = \Phi^{-1}_{\zeta\circ\Phi_c(u)}(c^{1/2}u(c\,\cdot)) \\ &= \Phi^{-1}_{c\zeta(u)}(c^{1/2}u(c\,\cdot)) = c^{-1/2}\zeta(u)^{-1/2}c^{1/2}u\left(\frac{1}{c\zeta(u)}c\,\cdot\right) \\ &= \Phi^{-1}_{\zeta(u)}(u) = s_1(u) \end{aligned}$$

Thus

$$\begin{aligned} n(s_1^{-1}(\Gamma) \cap \{u \in U_+ : \zeta(u) \geq c\}) &= n(\Phi_c^{-1}(s_1^{-1}(\Gamma) \cap \{u \in U_+ : \zeta(u) \geq 1\})) \\ &= c^{1/2} n(s_1^{-1}(\Gamma) \cap \{u \in U_+ : \zeta(u) \geq 1\}) \end{aligned}$$

Furthermore it is true that $n(u \in U_+ : \zeta(u) \geq c) = (1/(2\pi c))^{1/2}$. For a proof see [24] chapter XII proposition 2.8. With this result

$$\begin{aligned} \frac{n(s_1^{-1}(\Gamma) \cap \{\in U_+ : \zeta(u) \geq c\})}{n(\{u \in U_+ : \zeta(u) \geq c\})} &= c^{1/2}\frac{n(s_1^{-1}(\Gamma) \cap \{u \in U_+ : \zeta(u) \geq 1\})}{n(\{u \in U_+ : \zeta(u) \geq c\})} \\ &= \frac{n(s_1^{-1}(\Gamma) \cap \{u \in U_+ : \zeta(u) \geq 1\})}{n(\{u \in U_+ : \zeta(u) \geq 1\})} \end{aligned}$$

□

Furthermore one sees directly that $n_{(1)}(U^1) = 1$, which makes the law of normalized excursions a probability law. We now have the tools at hands to determine the density of normalized excursions under $n_{(1)}$. Consider the the family $(\pi_r)_{r>0}$ where π_r is a probability measure on $U_+ \cap \{\zeta = r\}$ with finite dimensional distributions for $0 < t_1 < \cdots < t_n < r$

$$\begin{aligned} &\pi_r(u(t_1) \in A_1, \ldots, u(t_n) \in A_n \cap \{\zeta(u) = r\}) \\ &\quad = \int_{A_1} \cdots \int_{A_n} 2\sqrt{2\pi r^3} l_{t_1}(x_1) q_{t_2-t_1}(x_1, x_2) \ldots q_{t_n-t_{n-1}}(x_{n-1}, x_n) l_{r-t_1}(x_n) \mathrm{d}x_1 \ldots \mathrm{d}x_n \,. \end{aligned}$$

Theorem 2.22 (Excursion law decomposed w.r.t. length)**.** *For the excursion measure it holds*

$$\begin{aligned} n(\Gamma) &= \int_0^\infty \pi_r(\Gamma \cap \{\zeta = r\}) \frac{\mathrm{d}r}{2\sqrt{2\pi r^3}} \\ &= \int_0^\infty \pi_r(\Gamma \cap \{\zeta = r\}) n(\zeta \in \mathrm{d}r) \end{aligned}$$

Proof. The result follows by direct computation. Consider the sets $\Gamma = \bigcap_{i=1\dots n}\{u(t_i) \in A_i\}$, which generate the σ-algebra on U. Then the right hand side is by definition of the finite dimensional distributions

$$\begin{aligned} &\int_0^\infty \pi_r(\Gamma \cap \{\zeta = r\}) \frac{\mathrm{d}r}{2\sqrt{2\pi r^3}} \\ &\quad = \int_0^\infty \int_{A_1} \cdots \int_{A_n} 2\sqrt{2\pi r^3} l_{t_1}(x_1) q_{t_2-t_1}(x_1, x_2) \dots q_{t_n - t_{n-1}}(x_{n-1}, x_n) l_{r-t_1}(x_n) \mathrm{d}x_1 \dots \mathrm{d}x_n \frac{\mathrm{d}r}{2\sqrt{2\pi r^3}} \\ &\quad = \int_{A_1} \cdots \int_{A_n} l_{t_1}(x_1) q_{t_2-t_1}(x_1, x_2) \dots q_{t_n - t_{n-1}}(x_{n-1}, x_n) \int_0^\infty l_{r-t_1}(x_n) \mathrm{d}r \mathrm{d}x_1 \dots \mathrm{d}x_n \\ &\quad = \int_{A_1} \cdots \int_{A_n} l_{t_1}(x_1) q_{t_2-t_1}(x_1, x_2) \dots q_{t_n - t_{n-1}}(x_{n-1}, x_n) \mathrm{d}x_1 \dots \mathrm{d}x_n = n(\Gamma) \end{aligned}$$

where in the last but one equation it was used that l_t has integral one. □

The law π_r is also the law of a Bessel bridge of dimension three and of length r. We now come to the final conclusion of this chapter, which will be crucial for the next one.

Corollary 2.23 (Density of normalized excursions)**.** *The normalized excursions* $\mathbb{e}_t \in U^1$ *under* $n_{(1)}$ *have the density* π_1*.*

Proof. We first use that

$$n(\{\zeta \geq c\} \cap U_+) = \sqrt{\frac{1}{2\pi c}}$$

which can be either taken from [24] Chapter XII Proposition 2.8 or compute directly from theorem 2.22, where, however, this result has been used already in the derivation. Thus for $\Gamma \in \mathcal{U}_+$

$$\begin{aligned} n_{(1)}(\Gamma) &\overset{(2.22)}{=} \int_c^\infty \pi_r(s_1^{-1}(\Gamma) \cap \{\zeta = r\}) \sqrt{2\pi c} \frac{\mathrm{d}r}{2\sqrt{2\pi r^3}} \\ &= \int_c^\infty \pi_r(s_1^{-1}(\Gamma) \cap \{\zeta = r\}) \frac{1}{2}\sqrt{\frac{c}{r^3}} \mathrm{d}r\,. \end{aligned}$$

This, of course, must hold independently from c such that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}c} n_{(1)}(\Gamma) &= 0 = -\pi_c(s_1^{-1}(\Gamma) \cap \{\zeta = c\})/(2c) + \int_c^\infty \pi_r(s_1^{-1}(\Gamma) \cap \{\zeta = r\}) \frac{1}{4}\sqrt{\frac{1}{cr^3}} \\ &= -\frac{1}{2c}\pi_c(s_1^{-1}(\Gamma) \cap \{\zeta = c\}) + \frac{1}{2c} n_{(1)}(\Gamma) \end{aligned}$$

In particular, in the case of $c = 1$ we get $n_{(1)} = \pi_1$. □

3. Scaling limit of a random walk's excursion

Recall the linearly interpolated continuous process of a random walk S_n

$$S_{nt} = S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1} \, .$$

for $r \in [0,1]$ and $n \in \mathbb{N}$. Again the first time the simple random walk reaches -1 is denoted by $T = \inf\{n \geq 0 : S_n = -1\}$. The excursion of the random walk with length $2k$is defined as a stochastic process $C_k : \Omega \to \mathcal{C}([0,1])$ which has the same distributed as the random walk conditioned on $T = 2k+1$, i.e.

$$\mathcal{L}((C_k(2kt))_{t\in[0,1]}) = \mathcal{L}((S_{2kt})_{t\in[0,1]} | T = 2k+1) \, .$$

With these notations we state the following limit theorem

Theorem 3.1 (Limit Theorem of Excursions)**.** *In the space $\mathcal{C}([0,1])$ equipped with the topology of uniform convergence and corresponding Borel σ-algebra, the normalized excursions of a random walk converge in distribution to the normalized excursion of Brownian motion:*

$$\left(\frac{1}{\sqrt{2k}} C_k(2kt) \right)_{0\leq t\leq 1} \xrightarrow[k\to\infty]{d} (\mathbb{e}_t)_{0\leq t\leq 1}$$

In this chapter we will proof Theorem 3.1. In the first section we give a criterion for weak convergence, and develop tools for this task in the second section. The third and fourth section prove the tightness of the sequence of probability measures and that their finite dimensional distributions converge. Together these imply the weak convergence of probability measures.

3.1. Weak convergence

We want to discuss methods to establish the weak convergence of measures. Due to Kolmogorov's extension theorem we can assume that we need the following:

Definition 3.2 (Convergence of finite dimensional distributions)**.** *Let $(X^k)_{k\in\mathbb{N}}$ be a sequence of $\mathbb{R}$-valued stochastic processes with index set T. We say that X^k converges to a stochastic process X in finite dimensional distributions, $X^k \xrightarrow{f.d.} X$, iff for every $(t_1, \dots, t_p) \subseteq T$ it holds*

$$(X^k_{t_1}, \dots, X^k_{t_p}) \xrightarrow{d} (X_{t_1}, \dots, X_{t_p})$$

The convergence of finite dimensional distributions is necessary, but not sufficient. We additionally need the relatively compactness of the sequence of probability measures. Due to the theorem of Prokhorov a sequence of probability measures is relatively compact, if and only if it is a uniformly tight set of measures.

Definition 3.3 (Uniform tightness)**.** *The collection of probability measures $\mathcal{P} = \{P^1, P^2, \dots\}$ is called uniformly tight, iff for every $\varepsilon > 0$ there exists a compact set K_ε such that*

$$P^k(K_\varepsilon) \geq 1 - \varepsilon \quad \forall k \in \mathbb{N}$$

It is sufficient to check uniform tightness only for $k \geq k_0 \in \mathbb{N}$ because any finite collection of probability measures on a compact set $T \subseteq \mathbb{R}$ is uniformly tight. See for example [16]. This is true in general for so called Polish spaces. In the case of continuous processes we can now make use of the following criterion.

Theorem 3.4 (Kolmogorov's criterion of weak convergence)**.** *Let $(X^k)_{k\in\mathbb{N}}$ be a sequence of continuous real valued stochastic processes on a compact set $T \subseteq \mathbb{R}$. It is equivalent:*

1. *The family $\{P^{X^1}, P^{X^2}, \dots\}$ is uniformly tight.*
2. *There exist positive constants α, β, γ and a $k_0 \in \mathbb{N}$ such that for all $s, t \in T$ and every $k \geq k_0$*
$$E[|X_s^k - X_t^k|^\alpha] \leq \beta |s - t|^{\gamma+1} \,. \tag{3.1}$$

In particular we have that X^k converges weakly to a real-valued continuous stochastic process X in $\mathcal{C}(T, \mathbb{R})$, if $X^k \xrightarrow{f.d.} X$ and (3.1) holds.

For a proof of this theorem see [24] Chapter XIII Section 1.

3.2. Tools

Before we come to the proof of tightness and convergence of finite dimensional distributions we need two more statements, which help us to control the distribution of the excursions of random walks. Recall that P_l denotes the law under which the random walk starts at l, i.e. $S_0 = l$ P_l-a.s.

$$P_l(S_n = m) = P(S_n + l = m) \,.$$

The hitting time theorem relates the distribution of the first hitting time to the distribution of the simple random walk.

Lemma 3.5 (Hitting time theorem)**.** *Consider the first hitting time $T = \inf\{n \geq 0 : S_n = -1\} < \infty$ a.s.. For every $l \in \mathbb{N}_0$ and every integer $n \geq 1$ we have*

$$P_l(T = n) = \frac{l+1}{n} P_l(S_n = -1) \tag{3.2}$$

We will prove this theorem using the cycle lemma of Dvoretzky and Motzkin [7] in a slightly varied formulation. The lemma has some additional combinatorial applications, for example, to compute the Catalans numbers which are mentioned in the next chapter.

Lemma 3.6 (Cycle Lemma)**.** *Consider a sequence $(x_1, \dots, x_n)$ of integers $x_i \in \{-1, 1\}$ $\forall i = 1 \dots n$ with $\sum_{i=1}^n x_i = -l - 1$. Let the sequence consist of p positive elements and q negative elements. Then there exist exactly $q - p = l + 1$ cyclic permutations π of $(x_1, \dots, x_n)$ such that*

$$\sum_{i=1}^k x_{\pi(i)} \geq -l \quad \text{for all } k \in \{1, \dots, n-1\} \tag{3.3}$$

Proof. Write the sequence as a sequence of ones and minus ones $(s_1, \ldots, s_{p+q})_{\text{cyc}}$ arranged in cyclic order, i.e. s_1 follows s_n and is followed by s_2 etc.. Any cyclic permutation corresponds to a point in the cycle at which the summation in (3.3) starts. We want to find these starting points. Whenever a 1 is followed by a -1 along the ordering of the sequence the pair can be considered as zero, as when summing these two numbers the sum can not decrease below $-l$. Continuing in this manner all ones are eliminated by a following -1. For this reason the procedure yields uniquely determined negative elements $s^* = (s_{i_j})_j = (s_{i_1}, \ldots s_{i_{q-p}})$. For every element $t \in s^*$ there exists exactly one cyclic permutation π such that $\pi(s_n) = t$. By construction it then holds

$$\sum_{i=1}^{k} x_{\pi(i)} \geq -l \quad \forall k \in \{1, \ldots, n-1\}$$

since less than $l+1$ elements of s^* are summed and the $2p$ other values always yield a non negative sum. Conversely, if $s_\pi(n) \notin s^*$, then it is either $s_{\pi(n)} = 1$ or there exists a $m \in \mathbb{N}$ such that $\sum_{j=0}^{m'} s_{\pi(n-2m+j)} \geq 0$ for all $m' \leq 2m$. In both cases there exists a $k \leq n-1$ such that

$$\sum_{i=1}^{k} x_{\pi(i)} = -l - 1\,.$$

□

The cycle lemma can be used as an elegant tool to prove the hitting time theorem, which we will do now.

Proof. (Hitting time theorem) Under P_l we can write $S_n = \sum_{i=1}^n X_i + l$ with independent homogeneously distributed $X_i \in \{-1, 1\}$. Note that the random walk starts at l. We analyze possible values of the vector $(X_1, X_2, \ldots, X_n)$ in the case that $S_n = -1$, which means that $\sum_{i=1}^n X_i = -l - 1$. There exist n cyclic permutations of this sequence, and any cyclic permutation is a bijection on the set of possible paths from $S_0 = l$ to $S_n = -1$. From the cycle lemma it follows that exactly $l+1$ cyclic permutations correspond to paths staying nonnegative, hence $l+1$ out of n permutations are favorable. Thus

$$P_l(T = n) = \frac{l+1}{n} P_l(S_n = -1)\,.$$

This formula is a special case of more general statements, see for example in [17] Section 5.1. □

Next we note that the conjectured limit contains Gaussian functions. We need uniform approximations of such normal distributions by random walks. This is discussed in the following. We will make use of the o-notation, which is described in Appendix B. We now prove two local central limit theorems. Set $X_i' = (X_i + 1)/2 \in \{0, 1\}$ and define the random walk

$$Z_n = \sum_{i=1}^{n} X_i' = \frac{S_n + n}{2} \in \mathbb{N}_0\,. \tag{3.4}$$

In this situation we have the following local central limit theorem:

Lemma 3.7 (Local central limit theorem)**.** *For the random walk Z_n the following holds:*

$$\sup_{m\in\mathbb{N}} \left| \sqrt{n} P(Z_n = m) - \frac{2}{\sqrt{2\pi}} e^{-(m-n/2)^2/(n/2)} \right| \xrightarrow{n\to\infty} 0 \tag{3.5}$$

Proof. For $n \in \mathbb{N}, m \in \mathbb{N}_0$

$$P(Z_n = m) = E[\mathbb{1}_{Z_n=m}]$$

is always positive. Since $\{e^{ikx}, k \in \mathbb{Z}\}$ is an orthonormal set with respect to the $L_2([-\pi,\pi])$ inner product we can rewrite a Kronecker-δ as

$$\delta_{Z_n,m} = \mathbb{1}_{Z_n=m} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{itZ_n} e^{-itm} \mathrm{d}t\,.$$

Applying the expectation we have

$$E\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{itZ_n} e^{-itm} \mathrm{d}t\right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} E\left[e^{itZ_n}\right] e^{-itm} \mathrm{d}t\,.$$

The expectation can be interchanged with the integral sign since $0 \leq Z_n \leq n$, hence the left hand side is a finite weighted sum of integrals. All integrals exist and the integration can be interchanged with the sum. Now we write the exponential as a product $\exp(itZ_n) = \prod_{i=0}^{n} \exp(itX_i')$ and use the independence of X_i' which gives:

$$E\left[e^{itZ_n}\right] = \prod_{i=0}^{n} E\left[\exp(itX_i')\right] = e^{itn\mu} \varphi(t)^n$$

where $\mu = E[X_i'] = 1/2$ and $\varphi(t) = E\left[\exp(it(X_i' - \mu))\right]$. Up to this point the reasoning was given by Terence Tao [20]. We now proceed in a more direct way. We plug in $\mu = 1/2$. The characteristic function $\varphi(t)$ can be simply computed using yet again that the expectation corresponds to a summation of two absolutely convergent series

$$\begin{aligned}\varphi(t) &= E\left[\exp(it(X_i' - 1/2))\right] = E\left[\sum_{k=0}^{\infty} \frac{(it(X_i' - 1/2))^k}{k!}\right] \\ &= \sum_{k=0}^{\infty} \frac{(it)^k}{k!} E((X_i' - 1/2)^k)\,.\end{aligned}$$

At this point we have $E((X_i' - 1/2)^{2k-1}) = 0$ and $E((X_i' - 1/2)^2 k) = (1/2)^{2k}$ for $k \in \mathbb{N}$ since

$$E((X_i' - 1/2)^k) = \frac{1}{2}^k \frac{1}{2} + \left(-\frac{1}{2}\right)^k \frac{1}{2}$$

and thus

$$\varphi(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} E((X_i' - 1/2)^k) = \sum_{k=0}^{\infty} \frac{(it/2)^{2k}}{(2k)!} = \cos(t/2)\,.$$

We now turn back to the integral and make a change of variables by defining x by $t = x/\sqrt{n}$. Thus we have

$$\begin{aligned}\sqrt{n} P(Z_n = m) &= \frac{\sqrt{n}}{2\pi} \int_{-\pi}^{\pi} e^{itn/2} \varphi^n(t) e^{-itm} \\ &= \frac{1}{2\pi} \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \cos^n(x/(2\sqrt{n})) e^{-ix(m/\sqrt{n} - \sqrt{n}/2)} \mathrm{d}x\,.\end{aligned} \tag{3.6}$$

We now use the following Fourier identity of a Gaussian function for $\sigma > 0$ and $\omega \in \mathbb{R}$:

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-\frac{1}{2}t^2\sigma^2} e^{-i\omega t}\mathrm{d}t = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\omega^2}{2\sigma^2}}$$

The trick is to choose ω such that the integrand is in phase with (3.6). We read of (3.6) and (3.5) that

$$\begin{aligned} \omega &= m/\sqrt{n} - \sqrt{n}/2 \\ \sigma &= 1/2 \end{aligned}$$

and consequently

$$\frac{2}{\sqrt{2\pi}} e^{-(m-n/2)^2/(n/2)} = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\frac{1}{2}(1/2)^2x^2} e^{-ix(m/\sqrt{n}-\sqrt{n}/2)}\mathrm{d}x \,. \tag{3.7}$$

Note that $\sigma = \sqrt{\mathrm{Var}(X')} = 1/2$ as one would expect, which is also how one can generalize this statement to other random walks. The identity (3.7) then reduces our problem (3.5) to show uniform convergence to zero of

$$\begin{aligned} &\left|\sqrt{n}P(Z_n = m) - \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\frac{1}{2}(1/2)^2x^2} e^{-ix(m/\sqrt{n}-\sqrt{n}/2)}\mathrm{d}x\right| \\ &\qquad \le \frac{1}{2\pi}\int_{\mathbb{R}} \left|\cos^n(x/2\sqrt{n})\mathbb{1}_{(-\pi\sqrt{n},\pi\sqrt{n})}(x) - e^{-x^2/8}\right| \,, \end{aligned}$$

where the right hand side is independent of m! By the dominating convergence theorem it is sufficient to prove pointwise convergence and find a integrable dominating function of the interesting sequence

$$x \mapsto f_n(x) = \cos^n(x/2\sqrt{n})\mathbb{1}_{(-\pi\sqrt{n},\pi\sqrt{n})}(x) \,.$$

The pointwise convergence can be shown by noting that for $|x| < \pi\sqrt{n}$, a series expansion of the cos and $a_n \in o(1/n^2)$ we have

$$\cos(x/2\sqrt{n})^n = \left(1 - \frac{x^2}{8n} + a_n(x)\right)^n \xrightarrow{n\to\infty} e^{-x^2/8} \,. \tag{3.8}$$

The higher order terms a_n can be neglected in the limit, which can be shown by using the binomial theorem, and the rest is then the standard sequence converging to the exponential function. Next we want to find a dominating integrable function of $\cos(x/(2\sqrt{n}))^n$. To do this we rearrange (3.8)

$$a_n(x) = \frac{x^2}{8n}\left(\frac{\cos(x/2\sqrt{n})}{x^2/8n} + 1 - \frac{1}{x^2/8n}\right) \,.$$

The function $t \mapsto f(t) = \frac{\cos(t)}{t^2} + 1 - \frac{1}{t^2}$ is bounded, has its minimum at $t = 0$, and increases in $|t|$ on the interval $(-\pi/2, \pi/2)$. This can be checked by computing its derivative. On the other hand

$$\cos^n(x/2\sqrt{n})\chi_{[-\pi\sqrt{n},\pi\sqrt{n}]}(x)$$

vanishes at $x > \pi\sqrt{n}$, hence in any nontrivial case $t = x/2\sqrt{n} < \pi/2$. This means that

$$0 < a_n < \frac{x^2}{8n}\left(1 - \left(\frac{2}{\pi}\right)^2\right) = \frac{x^2}{8n}\delta$$

with $\delta < 1$ independent of x and n. Finally

$$\cos(x/2\sqrt{n})^n < \left(1 - \frac{x^2}{8n} + \frac{x^2}{8n}\delta\right)^n = \exp\left(n \log\left[1 - \frac{x^2}{8n}(1-\delta)\right]\right)$$
$$\leq \exp\left(n\frac{-x^2}{8n}(1-\delta)\right) = \exp\left(-\frac{x^2}{8}(1-\delta)\right)$$

which is an integrable dominating function independent of n. This shows that

$$\frac{1}{2\pi}\int_{\mathbb{R}} \left|\cos^n(x/2\sqrt{n})\mathbb{1}_{(-\pi\sqrt{n},\pi\sqrt{n})}(x) - e^{-x^2/8}\right| \xrightarrow{n\to\infty} 0\,,$$

which implies the theorem. □

As a corollary we get the following continuous version which will be needed a lot afterwards.

Corollary 3.8 (Local central limit theorem)**.** *For any $s_0 > 0$ and numbers $r_1, r_2, r_3 \in \mathbb{R}_{\geq 0}$ we have*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}, s\geq s_0} \left|\sqrt{n+r_1}P(S_{\lfloor ns+r_2\rfloor} \in \{\lfloor x\sqrt{n}+r_3\rfloor, \lfloor x\sqrt{n}+r_3\rfloor+1\}) - \frac{2}{\sqrt{2\pi s}}e^{-x^2/2s}\right| = 0$$

Proof. Transforming back we have $X_i = 2X_i' - 1$ and $S_n = 2Z_n - n$. This means that

$$P(S_n \in \{m, m+1\}) = P\left(Z_n = \left\lfloor\frac{m+n+1}{2}\right\rfloor\right)\,,$$

To simplify notation we introduce

$$a(n,m) := \left\lfloor\frac{m+n+1}{2}\right\rfloor$$
$$f(n,m) := (m-n/2)^2/(n/2)$$

and now write for $s_0 > 0$ and $x_0 > 0$:

$$\left|\sqrt{n+r_1}P\Big(Z_{\lfloor ns+r_2\rfloor} = a(\lfloor ns+r_2\rfloor, \lfloor x\sqrt{n}+r_3\rfloor)\Big) - \frac{2}{\sqrt{2\pi s}}e^{-x^2/2s}\right|$$
$$\leq \left|\sqrt{n+r_1} - \sqrt{\lfloor ns+r_2\rfloor/s}\right| P\Big(Z_{\lfloor ns+r_2\rfloor} = a(\lfloor ns+r_2\rfloor, \lfloor x\sqrt{n}+r_3\rfloor)\Big)$$
$$+ \frac{1}{\sqrt{s}}\left|\sqrt{\lfloor ns+r_2\rfloor}\, P\Big(Z_{\lfloor ns+r_2\rfloor} = a(\lfloor ns+r_2\rfloor, \lfloor x\sqrt{n}+r_3\rfloor)\Big) - \frac{2}{\sqrt{2\pi}}e^{-f(\lfloor ns+r_2\rfloor, a(\lfloor ns+r_2\rfloor, \lfloor x\sqrt{n}+r_3\rfloor))}\right|$$
$$+ \frac{1}{\sqrt{s}}\left|\frac{2}{\sqrt{2\pi}}e^{-f(\lfloor ns+r_2\rfloor, a(\lfloor ns+r_2\rfloor, \lfloor x\sqrt{n}+r_3\rfloor))} - \frac{2}{\sqrt{2\pi}}e^{-x^2/2s}\right| = A + B + C$$

Let $\varepsilon > 0$. In the case of the first term there exists due to Lemma 3.7 a $\kappa \geq 0$ and $K \in \mathbb{N}$ such that for all $n \geq K$:

$$A \leq (r_1 + r_2/s_0 + 1/s_0)P\Big(Z_{\lfloor ns+r_2\rfloor} = a(\lfloor ns+r_2\rfloor, \lfloor x\sqrt{n}+r_3\rfloor)\Big) \tag{3.9}$$
$$\leq \kappa\frac{2(r_1 + r_2/s_0 + 1/s_0)}{\sqrt{2\pi n}} < \varepsilon$$

uniformly in s and x. Also the second term follows directly from lemma 3.7 as

$$B \leq \sup_{m\in\mathbb{N}} \frac{1}{\sqrt{s_0}} \left| \sqrt{\lfloor ns+r_2 \rfloor}\, P\Big(Z_{\lfloor ns+r_2 \rfloor} = m\Big) - \frac{2}{\sqrt{2\pi}} e^{-f(\lfloor ns+r_2 \rfloor, m)} \right| < \varepsilon \tag{3.10}$$

for $\lfloor ns+r_2 \rfloor \in \mathbb{N}$ large enough. To treat term C we first truncate for an $x_0 > 0 \in \mathbb{R}$. Now take $n \in \mathbb{N}$ large enough, such that $ns_0 > 1$. It then exists an $K(x_0, s_0)$ such that

$$\begin{aligned}
&f(\lfloor ns+r_2 \rfloor, a(\lfloor ns+r_2 \rfloor, \lfloor x\sqrt{n}+r_3 \rfloor)) \\
&= \left(\left\lfloor \frac{\lfloor x\sqrt{n}+r_3 \rfloor + \lfloor ns+r_2 \rfloor + 1}{2} \right\rfloor - \frac{1}{2}\lfloor ns+r_2 \rfloor \right)^2 \cdot \left(\frac{\lfloor ns+r_2 \rfloor}{2} \right)^{-1} \\
&\leq \left(\left| \frac{x\sqrt{n}+r_3+ns+r_2}{2} - \frac{ns+r_2}{2} \right| + 3 \right)^2 \frac{2}{ns+r_2-1} \\
&\leq \frac{x^2 n + 2|x|\sqrt{n}(r_3+3) + (r_3+3)^2}{2ns-2} \\
&< \frac{x^2}{2s}\frac{n}{n-1/s_0} + \frac{2x_0\sqrt{n}(r_3+3)+(r_3+3)^2}{2ns_0-2} \\
&< \frac{x^2}{2s}(1+\varepsilon)
\end{aligned}$$

for all $n \geq K(x_0, s_0)$ and all $s > s_0$, $-x_0 \leq x \leq x_0$. By the same arguments there exists for any $\varepsilon > 0$ an integer $K'(x_0, s_0)$ such that for all $n \geq K'(x_0, s_0)$

$$\begin{aligned}
&f(\lfloor ns+r_2 \rfloor, a(\lfloor ns+r_2 \rfloor, \lfloor x\sqrt{n}+r_3 \rfloor)) \\
&\geq \left(\left| \frac{x\sqrt{n}+r_3+ns+r_2+1}{2} - \frac{ns+r_2}{2} \right| - 2 \right)^2 \frac{2}{ns+r_2} \\
&\geq \frac{x^2}{2s}\frac{n}{n+r_2/s_0} - \frac{2(x_0\sqrt{n}+r_3+r_2+1)}{2ns_0} > \frac{x^2}{2s}(1-\varepsilon)
\end{aligned}$$

and all $s > s_0$, $-x_0 \leq x \leq x_0$. Together we obtain:

$$\sup_{s\geq s_0, |x|\leq x_0} \left| f(\lfloor ns+r_2 \rfloor, a(\lfloor ns+r_2 \rfloor, \lfloor x\sqrt{n}+r_3 \rfloor)) - \frac{x^2}{2s} \right| \xrightarrow{n\to\infty} 0$$

uniformly. Secondly we observe that $\exp(-f(\lfloor ns+r_2 \rfloor, a(\lfloor ns+r_2 \rfloor, \lfloor x\sqrt{n}+r_3 \rfloor))) \xrightarrow{x\to\infty} 0$ for all $n \in \mathbb{N}$ and $\exp(-x^2/2s) \xrightarrow{x\to\infty} 0$. Choose $x_0 > 0$ such that for all $x > x_0$ and all $n \in \mathbb{N}$:

$$\sup_{s\geq s_0, |x|>x_0} \left| e^{-f(\lfloor ns+r_2 \rfloor, a(\lfloor ns+r_2 \rfloor, \lfloor x\sqrt{n}+r_3 \rfloor))} - e^{-x^2/2s} \right| < \varepsilon$$

Due to the Lipschitz continuity of $t \to e^{-t}$ we can then choose a $K(x_0, s_0) \in \mathbb{N}$ such that for all $n \geq K$

$$\begin{aligned}
C \leq &\frac{2}{\sqrt{2\pi s_0}} \Bigg(\sup_{s\geq s_0, |x|>x_0} \left| e^{-f(\lfloor ns+r_2 \rfloor, a(\lfloor ns+r_2 \rfloor, \lfloor x\sqrt{n}+r_3 \rfloor))} - e^{-x^2/2s} \right| \\
&+ \sup_{s\geq s_0, |x|\leq x_0} \left| f(\lfloor ns+r_2 \rfloor, a(\lfloor ns+r_2 \rfloor, \lfloor x\sqrt{n}+r_3 \rfloor)) - \frac{x^2}{2s} \right| \Bigg) < 2\varepsilon
\end{aligned} \tag{3.11}$$

So with (3.9), (3.10) and (3.11) we obtain the stated uniform convergence. □

3.3. Tightness

We first come to the proof of tightness. We want to prove that the laws of

$$\left(\frac{1}{\sqrt{2k}}C_k(2kt)\right)_{0\leq t\leq 1}$$

are tight. To do this we use Kolmogorov's criterion for tightness, 3.4. We want to show

Proposition 3.9. *There exist positive constants α, β, γ and an integer $k_0 \in \mathbb{N}$ such that for all $s, t \in T$ and every $k \geq k_0$*

$$E[|(C_k(2ks) - C_k(2kt))/\sqrt{2k}|^\alpha] \leq \beta|s-t|^{\gamma+1}\,.$$

Consider a fixed $k \in \mathbb{N}$. Define $c_k(t) := (2k)^{-1/2}C_k(2kt)$ for all $t \in [0,1]$. By the Rerooting Lemma 5.3, which is proven later independent of this result, we can define

$$c'_k(t) = c_k(s) + c_k(\overline{s+t}) - 2m_{c_k}(s, \overline{s+t})\,.$$

where $\overline{s+t} = s+t-\lfloor s+t\rfloor$ and $m_{c_k}(s,t) = \inf_{r\in[s\wedge t, s\vee t]} c_k(r)$. For notations compare with Lemma 5.3. It is shown there that such a mapping defines a bijection on the set of continuous functions $g \in C([0,1])$ with $g(0) = g(1) = 0$. In our case we can check from the definition that $C'_k(i) := \sqrt{2k}c'_k(i/2k)$ defines again an excursion of the random walk, and thus the mapping $C_k \mapsto C'_k$ is a bijection on the set of excursions with length $2k$. This bijection now yields that c_k and c'_k have the same law, and thus

$$E[c_k(s) + c_k(\overline{s+t}) - 2m_{c_k}(s, \overline{s+t})] = E[c'_k(t)] = E[c_k(t)]\,.$$

To apply the weak convergence criterion we have a closer look at

$$\begin{aligned} E[|c_k(t) - c_k(s)|^4] &= E[|c_k(t) + c_k(s) - 2c_k(s)|^4] \\ &\leq E[|c_k(t) + c_k(s) - 2m_{c_k}(s,t)|^4] = E[|c_k(t-s)|^4]\,. \end{aligned}$$

Consider at first $t = j/2k, s = i/2k$ for natural numbers $0 \leq i < j \leq 2k$. We will use the following:

Lemma 3.10. *For the simple random walk $(S_i)_{i\in\mathbb{N}}$ and first hitting time of minus one, T, it holds*

$$P(S_i = l\,|\,T = 2k+1) = \frac{2(2k+1)(l+1)^2}{(2k+1-i)(i+1)}\frac{P_l(S_{i+1} = -1)P_l(S_{2k+1-i} = -1)}{P(S_{2k+1} = -1)}\,, \tag{3.12}$$

where again P_l denotes the law under which $S_0 = l$ a.s..

Proof. An outline of the following steps was given by [15]. We start with some manipulations of the law of a conditioned random walk: Consider for $i \in \{1, \ldots, 2k\}$ and $l \in \mathbb{N}_0$

$$P(S_i = l\,|\,T = 2k+1) = \frac{P(S_i = l, T = 2k+1)}{P(T = 2k+1)}\,.$$

Instead of conditioning on $T = 2k+1$ we can use $S_i = l$ as condition which yields

$$\begin{aligned}
P(S_i = l, T = 2k+1) &= P(S_i = l, T = 2k+1, T > i) \\
&= P(T = 2k+1 \,|\, S_i = l, T > i) P(S_i = l, T > i) \\
&= P_l(T = 2k+1-i) P(S_i = l, T > i) \,.
\end{aligned}$$

In the last equality the Markov property of the random walk was used. A random walk conditioned on reaching l at i coincides with the law of a random walk starting at l. We observe that

$$P(S_i = l, T > i) = P(S_i = l, S_j \geq 0 \text{ for } j \leq i)$$

corresponds to all random walks starting at zero, reaching l at i and with arbitrary values past i. By time reversal we can view this as a random walk starting at l and reaching 0 at i. Thus

$$\begin{aligned}
P(S_i = l, S_j \geq 0 \text{ for } j \leq i) &= P_l(S_i = 0, S_j \geq 0 \text{ for } j \leq i) \\
&= P_l(S_i = 0, S_j \geq 0 \text{ for } j \leq i, S_{i+1} = -1) \\
&\quad + P_l(S_i = 0, S_j \geq 0 \text{ for } j \leq i, S_{i+1} = 1) \\
&= 2P_l(S_i = 0, S_j \geq 0 \text{ for } j \leq i, S_{i+1} = -1) = 2P_l(T = i+1)
\end{aligned} \tag{3.13}$$

With these computations we finally find

$$P(S_i = l \,|\, T = 2k+1) = \frac{2P_l(T = i+1)P_l(T = 2k+1-i)}{P(T = 2k+1)}$$

We now use the hitting time theorem (Lemma 3.5) which yields

$$P(S_i = l \,|\, T = 2k+1) = \frac{2(2k+1)(l+1)^2}{(2k+1-i)(i+1)} \frac{P_l(S_{i+1} = -1)P_l(S_{2k+1-i} = -1)}{P(S_{2k+1} = -1)} \,. \tag{3.14}$$

□

This equation will be important again later on as well. Assume i, l have same parity, which means in our case that i, l are either both odd or both even, and $i \neq 0$, since this case is always trivial. We have

$$P(c(i/2k) = l/\sqrt{2k}) = P(C_k(i) = l) = P(S_i = l \,|\, T = 2k+1)$$

By assumption $2k+1-i$ and $-l-1$ have same parity, such that

$$P_l(S_{2k+1-i} = -1) = P(S_{2k+1-i} \in \{-l-1, -l\})$$

We apply corollary 3.8 and obtain:

$$\sup_{l \in \mathbb{Z}} \left| \sqrt{2k+1-i}\,P(S_{2k+1-i} \in \{-l-1, -l\}) - \frac{2}{\sqrt{2\pi}} \right| \xrightarrow{n \to \infty} 0$$

$$\left| \sqrt{2k+1}\,P(S_{2k+1} = -1) - \frac{2}{\sqrt{2\pi}} \right| \xrightarrow{n \to \infty} 0$$

from which we can see that there for any constant $\kappa_1 > 0$ there exists a $K \in \mathbb{N}$ independent of $l \in \mathbb{Z}$ such that

$$P(S_{2k+1-i} \in \{-l-1, -l\}) \leq \frac{2/\sqrt{2\pi} + \kappa_1}{\sqrt{k}} \quad \text{and} \quad P(S_{2k+1} = -1) \geq \frac{2/\sqrt{2\pi} - \kappa_1}{\sqrt{k}}$$

for all $k \geq K$. Further we may assume $i \leq k$, since $C_k(2k-i)$ has the same distribution as $C_k(i)$. Thus we get the estimate

$$P(C_k(i) = l) \leq \frac{2(2k+1)(l+1)^2}{(2k+1-i)(i+1)} \frac{(2/\sqrt{2\pi}+\kappa_1)/\sqrt{2k}}{(2/\sqrt{2\pi}-\kappa_1)/\sqrt{2k}} P_l(S_{i+1} = -1) < \kappa_3 \frac{(l+1)^2}{(i+1)} P(S_{i+1} = l+1) \tag{3.15}$$

with $\kappa_3 = 4(2/\sqrt{2\pi}+\kappa_1)/(2/\sqrt{2\pi}-\kappa_1)$. Hence we need to estimate the scaling of

$$\begin{aligned} \mathrm{E}[|c(t-s)|^4] &= \frac{1}{2k} \sum_{l=-\infty}^{\infty} l^4 P(C(j-i) = l) \\ &\leq \frac{1}{2k} \sum_{l=-\infty}^{\infty} (l+1)^4 \kappa_3 \frac{(l+1)^2}{(j-i+1)} P(S_{j-i+1} = l+1) = \frac{1}{2k} \frac{\kappa_3}{j-i+1} \mathrm{E}[S_{j-i+1}^6] \,. \end{aligned}$$

The sixth moment of a random walk can be computed explicitly. We need to compute the expectation of

$$\mathrm{E}[S_{i+1}^6] = \mathrm{E}\left[\left(\sum_{m=1}^{i+1} X_m\right)^6\right]$$

which is a sum of expressions $\mathrm{E}[X_m^6] = \mathrm{E}[X_{m_1}^4 X_{m_2}^2] = \mathrm{E}[X_{m_1}^2 X_{m_2}^2 X_{m_3}^2] = 1$ due to the independence of random variables X_m and $\mathrm{E}[X_m] = 0$. There are now

$$i+1+\binom{i+1}{2}\binom{4}{2}\binom{2}{2}+\binom{i+1}{3}\binom{6}{3} \leq \kappa_4 (i+1)^3$$

such terms, $\kappa_4 > 0$, where the scaling to the third power is due to the third term. This gives

$$\mathrm{E}[|c(t-s)|^4] \leq \kappa_3\kappa_4 \frac{(j-i+1)^2}{2k} = \kappa_3\kappa_4((t-s)^2 + o(1)) \,.$$

Finally in the case of general $t < s \in [0,1]$ we have

$$\begin{aligned} \mathrm{E}[|c(t-s)|^4] &\leq \frac{1}{2k} \mathrm{E}[|C(\lfloor 2kt - 2ks \rfloor)|^4] \wedge \mathrm{E}[|C(\lceil 2kt - 2ks \rceil)|^4] \\ &\leq \kappa_3\kappa_4(t-s)^2 + o(1) \end{aligned}$$

Adjusting the constant prefactor once again to absorb the terms of order $o(1)$ this now shows Proposition 3.9.

3.4. Convergence of finite dimensional distributions

To analyze the convergence of finite dimensional distributions, remember the following notations. The density of the entrance law of the excursion process

$$l_t = \frac{|x|}{\sqrt{2\pi t^3}} \exp\left(-\frac{x^2}{2t}\right) ,$$

the Gaussian or Brownian transition density

$$p_t(x,y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y-x)^2}{2t}\right) ,$$

as well as the transition density of Brownian motion killed when it hits zero.

$$q_t(x,y) = p_t(x,y) - p_t(x,-y)$$

Note that this notation differs slightly from [15], as we used the notation as in the literature, e.g. [24]. We realized in the introduction that in $k \in \mathbb{N}$ it holds

$$\frac{1}{\sqrt{2k}} S_{2kt} - \frac{1}{\sqrt{2k}} S_{\lfloor 2kt \rfloor} \xrightarrow{p} 0\,.$$

Due to the theorem of Slutzky 1.9 the convergence of finite dimensional distributions can then be stated as

Proposition 3.11 (Convergence of finite dimensional distributions)**.** *Consider a discretization of the excursion interval* $0 < t_1 < t_2 < \cdots < t_p < 1$ *for* $p \in \mathbb{N}$*. The finite dimensional distributions converge, i.e.*

$$\frac{1}{\sqrt{2k}} \left(S_{\lfloor 2kt_1 \rfloor}, \ldots, S_{\lfloor 2kt_p \rfloor} \,|\, T = 2k+1 \right) \xrightarrow[k\to\infty]{d} (\mathbb{e}(t_1), \ldots, \mathbb{e}(t_p))$$

where for measurable sets $A_1, \ldots A_p \in \mathcal{B}(\mathbb{R}_{\geq 0})$

$$\begin{aligned} n_{(1)}(\mathbb{e}(t_1) \in A_1, \ldots, \mathbb{e}(t_p) \in A_p) = \; & \int_{A_1} 2\sqrt{2\pi}\, l_{t_1}(x_1) \int_{A_2} q_{t_2 - t_1}(x_1, x_2) \int_{A_3} \cdots \\ & \cdots \int_{A_p} q_{t_p - t_{p-1}}(x_{p-1}, x_p) l_{1-t_p}(x_p) \,\mathrm{d}x_1 \ldots \mathrm{d}x_p \end{aligned}$$

It is sufficient to check the convergence of probability measures on closed boxes only. This will simplify the proof significantly. Consider boxes $V := A_1 \times A_2 \times \cdots \times A_p$ with $A_i = [a_i, b_i] \subseteq \mathbb{R}_{\geq 0}$ for $i = 1, \ldots, p$. These sets form a $\cap$ stable system which generates $\mathcal{B}(\mathbb{R}^p_{\geq 0}) = \mathcal{B}(\mathbb{R}_{\geq 0})^p$. (Sometimes also called generating π-system.) It is sufficient to check the the convergence only on these sets because of the separability of $\mathbb{R}_{\geq 0}$, compare with Billingsley Theorem 2.3 [4]. We then take a subdivision of the boxes into smaller and smaller cubes, i.e. for $j = 1, 2, \ldots, p$:

$$a_j \leq \frac{\lfloor \sqrt{2k} a_j \rfloor + 1}{\sqrt{2k}} < \frac{\lfloor \sqrt{2k} a_j \rfloor + 2}{\sqrt{2k}} < \cdots < \frac{\lfloor \sqrt{2k} b_j \rfloor}{\sqrt{2k}} \leq b_j \text{ of } A_j = [a_j, b_j]$$

and name the interior points for $j = 1, \ldots, p$ as $x_i^{(j)} = (\lfloor \sqrt{2k} a_j \rfloor + i)/\sqrt{2k}$ for $i \leq n_j = \lfloor \sqrt{2k} b_j \rfloor - \lfloor \sqrt{2k} a_j \rfloor$. Next we use (3.15), which tells us that there exists a $\kappa > 0$ independent of x and t such that

$$\begin{aligned} & P\left(\frac{1}{\sqrt{2k}} S_{\lfloor 2kt \rfloor} \in \{x_i^{(j)}, x_{i+1}^{(j)}\} \,|\, T = 2k+1 \right) \qquad (3.16) \\ & \leq \kappa \frac{(\lfloor \sqrt{2k} a_j \rfloor + i)^2}{\lfloor 2kt \rfloor} P(S_{\lfloor 2kt \rfloor + 1} \in \{\lfloor \sqrt{2k} a_j \rfloor + i, \lfloor \sqrt{2k} a_j \rfloor + i + 1\}) \\ & \overset{\text{Cor. 3.8}}{\leq} \kappa\kappa_2 \frac{2}{\sqrt{4\pi t k}} e^{-a_j^2/2t} \leq \frac{\kappa_3(t)}{\sqrt{k}} \end{aligned}$$

where $\kappa_2, \kappa_3 > 0$. With this notation the probability becomes

$$\begin{aligned}
&P\left(\frac{1}{\sqrt{2k}}S_{\lfloor 2kt_1\rfloor} \in A_1, \dots, \frac{1}{\sqrt{2k}}S_{\lfloor 2kt_p\rfloor} \in A_p \,\middle|\, T = 2k+1\right) \\
&= \sum_{i_1=1}^{n_1} P\left(\frac{1}{\sqrt{2k}}S_{\lfloor 2kt_1\rfloor} = x^{(1)}_{i_1}, \frac{1}{\sqrt{2k}}S_{\lfloor 2kt_2\rfloor} \in A_2 \dots, \frac{1}{\sqrt{2k}}S_{\lfloor 2kt_p\rfloor} \in A_p \,\middle|\, T = 2k+1\right) \\
&= \frac{1}{2}\cdot\left[\sum_{i_1=2}^{n_1-1} P\left(\frac{1}{\sqrt{2k}}S_{\lfloor 2kt_1\rfloor} \in \{x^{(1)}_{i_1}, x^{(1)}_{i_1+1}\}, \dots, \frac{1}{\sqrt{2k}}S_{\lfloor 2kt_p\rfloor} \in A_p \,\middle|\, T = 2k+1\right)\right. \\
&\qquad \left. + P\left(\frac{1}{\sqrt{2k}}S_{\lfloor 2kt_1\rfloor} \in \{x^{(1)}_1, x^{(1)}_n\} \dots, \frac{1}{\sqrt{2k}}S_{\lfloor 2kt_p\rfloor} = x^{(p)}_{i_p} \,\middle|\, T = 2k+1\right)\right] \\
&\quad \vdots \text{ proceeding inductively} \\
&= \frac{1}{2^p}\sum_{i_1=2}^{n_1-1}\cdots\sum_{i_p=2}^{n_p-1} P\left(\frac{1}{\sqrt{2k}}S_{\lfloor 2kt_1\rfloor} \in \{x^{(1)}_{i_1}, x^{(1)}_{i_1+1}\}, \dots, \frac{1}{\sqrt{2k}}S_{\lfloor 2kt_p\rfloor} \in \{x^{(p)}_{i_p}, x^{(p)}_{i_p+1}\} \,|\, T = 2k+1\right) \\
&+ \text{Boundary terms}
\end{aligned}$$

where the additional terms vanish since

$$\text{Boundary terms} \le \sum_{j=1}^{p} \frac{\kappa_3(t_j)}{\sqrt{2k}} \xrightarrow{k\to\infty} 0\,,$$

because of (3.16). From now on we use the shorthand notation

$$f_{t_1,\dots,t_p}(x_1,\dots,x_p) = 2\sqrt{2\pi}\, l_{t_1}(x_1) q_{t_2-t_1}(x_1,x_2)\dots q_{t_p-t_{p-1}}(x_{p-1},x_p) l_{1-t_p}(x_p)\,.$$

for the density of the excursion measure with respect to the Lebesgue measure. We can similarly treat the density:

$$\begin{aligned}
\int_{A_1\times\dots\times A_p} f_{t_1,\dots,t_p}\,\mathrm{d}\lambda^p &= \int_{A_2\times\dots\times A_p}\left[\int_{a_1}^{x^{(1)}_1} f_{t_1,\dots,t_p}\mathrm{d}\lambda + \sum_{i_1=2}^{n_1-1}\int_{x^{(1)}_{i_1}}^{x^{(1)}_{i_1+1}} f_{t_1,\dots,t_p}\mathrm{d}\lambda + \int_{x^{(1)}_{n_1}}^{b} f_{t_1,\dots,t_p}\mathrm{d}\lambda\right]\mathrm{d}\lambda^{p-1} \\
&= \int_{A_2\times\dots\times A_p}\sum_{i_1=2}^{n_1-1}\frac{1}{\sqrt{2k}} f_{t_1,\dots,t_p}(x^{(1)}_{i_1}, x_2,\dots,x_p)\mathrm{d}\lambda^{p-1} + o(1) \\
&\ \ \vdots \\
&= \frac{1}{\sqrt{2k}^p}\sum_{i_1=2}^{n_1-1}\cdots\sum_{i_p=2}^{n_p-1} f_{t_1,\dots,t_p}(x^{(1)}_{i_1},\dots,x^{(p)}_{i_p}) + o(1)
\end{aligned}$$

where it was used that $f_{t_1,\dots,t_p}$ is continuous and $x^{(j)}_{i_j+1} - x^{(j)}_{i_j} = \sqrt{2k}^{-1}$. From the preceding discussion we see

$$\begin{aligned}
&\Bigg|2^p P\left(\frac{1}{\sqrt{2k}} S_{\lfloor 2kt_1 \rfloor} \in A_1, \dots, \frac{1}{\sqrt{2k}} S_{\lfloor 2kt_p \rfloor} \in A_p \,|\, T = 2k+1\right) \\
&- 2^p \int_V f_{t_1,\dots,t_p}(x_1,\dots,x_p) \mathrm{d}x_1 \dots \mathrm{d}x_p \Bigg| \\
\le& \sum_{i_1=2}^{n_1-1} \cdots \sum_{i_p=2}^{n_p-1} \Bigg| P\left(\frac{1}{\sqrt{2k}} S_{\lfloor 2kt_1 \rfloor} \in \{x^{(1)}_{i_1}, x^{(1)}_{i_1+1}\}, \dots, \frac{1}{\sqrt{2k}} S_{\lfloor 2kt_p \rfloor} \in \{x^{(p)}_{i_p}, x^{(p)}_{i_p+1}\} \,|\, T = 2k+1\right) \\
&- \frac{2^p}{\sqrt{2k}^p} f_{t_1,\dots,t_p}(x^{(1)}_{i_1}, \dots, x^{(p)}_{i_p}) \Bigg| + o(1) \\
\le& \sup_{x \in V} \Bigg| \sqrt{2k}^p P\Big(\frac{1}{\sqrt{2k}} S_{\lfloor 2kt_1 \rfloor} \in \{x^{(1)}, x^{(1)} + 2k^{-1/2}\}, \dots \\
&\dots, \frac{1}{\sqrt{2k}} S_{\lfloor 2kt_1 \rfloor} \in \{x^{(p)}, x^{(p)} + 2k^{-1/2}\} \,|\, T = 2k+1\Big) - 2^p f_{t_1,\dots,t_p}(x^{(1)}, \dots, x^{(p)}) \Bigg| + o(1)
\end{aligned}$$

The result follows if we show that in the argument $(x^{(1)}, \dots, x^{(p)}) \in \mathbb{R}^p$ uniformly on V. Consequently we state the following theorem which is equivalent to the convergence of finite dimensional distributions:

Theorem 3.12. *Let $p \in \mathbb{N}$ be arbitrary and $V \subseteq \mathbb{R}^p$ be a box $V = A_1 \times \dots \times A_p$, where $A_i = [a_i, b_i]$. It holds*

$$\begin{aligned}
&\sup_{x \in V} \Bigg| \sqrt{2k}^p P\Big(\frac{1}{\sqrt{2k}} S_{\lfloor 2kt_1 \rfloor} \in \{x^{(1)}, x^{(1)} + 2k^{-1/2}\}, \dots \\
&\dots, \frac{1}{\sqrt{2k}} S_{\lfloor 2kt_1 \rfloor} \in \{x^{(p)}, x^{(p)} + 2k^{-1/2}\} \,|\, T = 2k+1\Big) - 2^p f_{t_1,\dots,t_p}(x^{(1)}, \dots, x^{(p)}) \Bigg| \xrightarrow{k \to \infty} 0
\end{aligned}$$

Proof. In the following we generalize the argument given by [15] to the arbitrary finite dimensional case. Take $n \in \mathbb{N}$ and integers $0 < i_1 < \dots < i_n < 2k$. We first do some general manipulations of the expressions we want to take the limit of. By the definition of the conditional probability:

$$P(S_{i_1} = l_1, \dots, S_{i_n} = l_n \,|\, T = 2k+1) = \frac{P(S_{i_1} = l_1, \dots, S_{i_n} = l_n, T = 2k+1)}{P(T = 2k+1)} .$$

Now we have:

$$\begin{aligned}
P(S_{i_1} = l_1, \dots, S_{i_n} = l_n, T = 2k+1) &= P(T = 2k+1 | S_{i_1} = l_1, \dots, S_{i_n} = l_n, T > i_n) \\
&\quad \times P(S_{i_1} = l_1, \dots, S_{i_n} = l_n, T > i_n) \\
&= P_{l_n}(T = 2k+1-i_n) P(S_{i_1} = l_1, \dots, S_{i_n} = l_n, T > i_n) ,
\end{aligned}$$

where the Markov property was used in the second line. For the second factor we observe:

$$\begin{aligned}
P(S_{i_1} = l_1, \dots, S_{i_n} = l_n, T > i_n) &= P(S_{i_n} = l_n, T > i_n | S_{i_1} = l_1, \dots, S_{i_{n-1}} = l_{n-1}, T > i_{n-1}) \\
&\quad \times P(S_{i_1} = l_1, \dots, S_{i_{n-1}} = l_{n-1}, T > i_{n-1}) \\
&= P_{l_{n-1}}(S_{i_n - i_{n-1}} = l_n, T > i_n - i_{n-1}) \\
&\quad \times P(S_{i_1} = l_1, \dots, S_{i_{n-1}} = l_{n-1}, T > i_{n-1})
\end{aligned}$$

Doing the same calculations recursively for the last factor we end up with

$$P(S_{i_1} = l_1, \dots, S_{i_n} = l_n, T > i_n) = \prod_{j=1}^{n-1} \left[P_{l_{n-j}}(S_{i_{n-j+1}-i_{n-j}} = l_{n-j+1}, T > i_{n-j+1} - i_{n-j})\right] \\ \times P(S_{i_1} = l_1, T > i_1)$$

We already showed in (3.13) that

$$P(S_{i_1} = l_1, T > i_1) = 2P_{l_1}(T = i_1 + 1)\,, \tag{3.17}$$

which we can use for the last factor. Let us now focus on the terms in the product. Here we can use the reflection principle of the random walk to simplify further. For this we use the same tricks that were used in the derivation of the Taboo transition function. First consider for $j \in \{1, \dots, n-1\}$:

$$\begin{aligned} &P_{l_{n-j}}(S_{i_{n-j+1}-i_{n-j}} = l_{n-j+1}, T < i_{n-j+1} - i_{n-j}) \\ &\quad = P(S_{i_{n-j+1}-i_{n-j}} = l_{n-j+1} - l_{n-j},\, T_{-l_{n-j}} < i_{n-j+1} - i_{n-j}) \\ &\quad = P(S_{i_{n-j+1}-i_{n-j}} = l_{n-j} - l_{n-j+1},\, T_{l_{n-j}} < i_{n-j+1} - i_{n-j}) \\ &\quad = P(S_{i_{n-j+1}-i_{n-j}} = l_{n-j} - l_{n-j+1},\, \max_{k \le i_{n-j+1}-i_{n-j}} S_k \ge l_{n-j}) \\ &\quad = P(S_{i_{n-j+1}-i_{n-j}} = l_{n-j} + l_{n-j+1},\, \max_{k \le i_{n-j+1}-i_{n-j}} S_k \ge l_{n-j}) = P_{-l_{n-j+1}}P(S_{i_{n-j+1}-i_{n-j}} = l_{n-j}) \end{aligned}$$

This allows us to write

$$\begin{aligned} &P_{l_{n-j}}(S_{i_{n-j+1}-i_{n-j}} = l_{n-j+1}, T > i_{n-j+1} - i_{n-j}) \\ &\quad = P_{l_{n-j}}(S_{i_{n-j+1}-i_{n-j}} = l_{n-j+1}) - P_{-l_{n-j+1}}(S_{i_{n-j+1}-i_{n-j}} = l_{n-j})\,. \end{aligned}$$

Finally we use the hitting Time Theorem (3.5) to rewrite everything in terms of the random walk

$$\begin{aligned} P_{l_1}(T = i_1 + 1) &= \frac{l_1 + 1}{i_1 + 1} P_{l_1}(S_{i_1+1} = -1) \\ P_{l_n}(T = 2k + 1 - i_n) &= \frac{l_n + 1}{2k + 1 - i_n} P_{l_n}(S_{2k+1-i_n} = -1)\,. \end{aligned}$$

Altogether we can infer

$$\begin{aligned} &P(S_{i_1} = l_1, \dots, S_{i_n} = l_n \,|\, T = 2k+1) \\ &= \frac{2(2k+1)(l_1+1)(l_n+1)}{(2k+1-i_n)(i_1+1)} \frac{2P_{l_1}(S_{i_1+1} = -1)P_{l_n}(S_{2k+1-i_n} = -1)}{P(S_{2k+1} = -1)} \\ &\times \prod_{j=1}^{n-1} [P(S_{i_{n-j+1}-i_{n-j}} = l_{n-j+1} - l_{n-j}) - P(S_{i_{n-j+1}-i_{n-j}} = l_{n-j} + l_{n-j+1})] \end{aligned}$$

Take now a discretization of the normalized excursion intervall $0 < t_1 < \dots < t_p < 1$ We will employ the notation

$$p(n, m) = P(S_n \in \{m, m+1\})\,.$$

Define then for $\alpha \in \{0,1\}^p$:

$$\begin{aligned} \tilde{p}_k^\alpha(t_i, x^{(i)}, i = 1 \dots p) :=& \sqrt{2k}^p P\Big(S_{\lfloor 2kt_1 \rfloor} = \lfloor \sqrt{2k}x^{(1)} + \alpha(1) \rfloor, \dots \\ &\dots, S_{\lfloor 2kt_p \rfloor} = \lfloor \sqrt{2k}x^{(p)} + \alpha(p) \rfloor \,\Big|\, T = 2k+1\Big) \end{aligned} \tag{3.18}$$

and the index set $K(\alpha) = \{k \in \mathbb{N} \,|\, \tilde{p}_k^\alpha(t_i, x^{(i)}, i = 1 \dots p) > 0\} \subseteq N$. $K(\alpha) = \{k_1, k_2, \dots\}$ is countably infinite, since $P(S_n = l) > 0$ if $l + n$ is even (We will speak of same parity of n and l). We then have, writing all terms out

$$\begin{aligned}\tilde{p}_k^\alpha(t_i, x^{(i)}, i = 1 \dots p) = \sqrt{2k}^p \; & \frac{2(2k+1)(\lfloor x^{(p)}\sqrt{2k} + \alpha(p)\rfloor + 1)(\lfloor x^{(1)}\sqrt{2k} + \alpha(1)\rfloor + 1)}{(2k + 1 - \lfloor 2kt_p\rfloor)(\lfloor 2kt_1\rfloor + 1)} \\ & \times \frac{2p(\lfloor 2kt_1 + 1\rfloor, \lfloor x^{(1)}\sqrt{2k} + \alpha(1)\rfloor + 1)p(\lfloor 2kt_p\rfloor, \lfloor x^{(p)}\sqrt{2k} + \alpha(p)\rfloor + 1)}{P(S_{2k+1} = -1)} \\ & \times \prod_{j=1}^{p-1} \Bigg[p\left(\lfloor 2k(t_{j+1} - t_j)\rfloor, \lfloor \sqrt{2k}x^{(j+1)}\rfloor + \alpha(j+1) - \lfloor \sqrt{2k}x^{(j)}\rfloor - \alpha(j)\right) \\ & \quad - p\left(\lfloor 2k(t_{j+1} - t_j)\rfloor, \lfloor \sqrt{2k}x^{(j+1)}\rfloor + \alpha(j+1) + \lfloor \sqrt{2k}x^{(j)}\rfloor + \alpha(j)\right) \Bigg]\end{aligned}$$

Let $(k_l)_{l\in\mathbb{N}} \subseteq K(\alpha)$ be the diverging list of elements in $K(\alpha)$. Using Corollary 3.8 again we have directly with $s = t_{j+1} - t_j$ and $x = x^{(j+1)} \pm x^{(j)}$:

$$\begin{aligned}& \sqrt{2k_l}p\left(\lfloor 2k_l(t_{j+1} - t_j)\rfloor, \lfloor \sqrt{2k_l}x^{(j+1)}\rfloor \pm \lfloor \sqrt{2k_l}x^{(j)}\rfloor + \alpha(j+1) \pm \alpha(j)\right) \\ & \xrightarrow{l\to\infty} 2p_{t_{j+1}-t_j}(x^{(j+1)}, \pm x^{(j)})\end{aligned} \tag{3.19}$$

for fixed α. Analogously we have

$$\begin{aligned} 2\sqrt{2k_l}p(\lfloor 2k_lt_1 + 1\rfloor, \lfloor x^{(1)}\sqrt{2k_l} + \alpha(1)\rfloor + 1) &\xrightarrow{l\to\infty} 2p_{t_1}(x^{(1)}) \\ \sqrt{2k_l}p(\lfloor 2k_lt_p\rfloor, \lfloor x^{(p)}\sqrt{2k_l} + \alpha(p)\rfloor + 1) &\xrightarrow{l\to\infty} p_{t_p}(x^{(p)})\,. \end{aligned} \tag{3.20}$$

All these convergences are uniform on V by Corollary 3.8. Additionally we have for the first terms

$$\frac{2(2k+1)(\lfloor x^{(p)}\sqrt{2k} + \alpha(p)\rfloor + 1)(\lfloor x^{(1)}\sqrt{2k} + \alpha(1)\rfloor + 1)}{(2k + 1 - \lfloor 2kt_p\rfloor)(\lfloor 2kt_1\rfloor + 1)} \xrightarrow{k\to\infty} \frac{2x^{(p)}x^{(1)}}{(1 - t_p)t_1} \tag{3.21}$$

because for large k the fraction is determined by the terms proportional to k. Also this convergence is happening uniformly due to the boundedness of V and similar arguments as in the ones leading to (3.11). Finally

$$\sqrt{2k}P(S_{2k+1} = -1) \xrightarrow{k\to\infty} \frac{2}{\sqrt{2\pi}}\,. \tag{3.22}$$

independent of $x \in V$. Note that the factors of $\sqrt{2k}^p$ work out precisely, since $p - 1$ factors are needed for the terms in the big product, one each for the limits (3.19), 2 more factors go into the limit of (3.20), while one factor is gained since (3.22) appears in the denominator. In total we can conclude that for any α:

$$\begin{aligned}\tilde{p}_{k_l}^\alpha(t_i, x^{(i)}, i = 1 \dots p) &\xrightarrow{l\to\infty} \prod_{j=1}^{p-1} \left[2p_{t_{j+1}-t_j}(x^{(j+1)}, x^{(j)}) - 2p_{t_{j+1}-t_j}(x^{(j+1)}, -x^{(j)})\right] \\ &\quad \times 2p_{t_1}(x^{(1)})p_{t_p}(x^{(p)})\frac{2x^{(p)}x^{(1)}}{(1 - t_p)t_1}\frac{\sqrt{2\pi}}{2} \\ &= 2^p\sqrt{2\pi}l_{t_1}(x^{(1)})\prod_{j=1}^{p-1} q_{t_{j+1}-t_j}(x^{(j+1)}, x^{(j)})l_{t_p}(x^{(p)}) = 2^p f_{t_1,\dots,t_p}(x^{(1)}, \dots, x^{(p)})\end{aligned} \tag{3.23}$$

uniformly on V. Now note that for each $k \in \mathbb{N}$ and each vector $(x^{(1)}, \dots, x^{(p)})$ there exists exactly one $\alpha \in \{0,1\}^p$ such that

$$\begin{aligned}\tilde{p}_k(t_i, x^{(i)}, i = 1 \dots p) &:= \sqrt{2k}^p P\Big(S_{\lfloor 2kt_1 \rfloor} \in \{\lfloor \sqrt{2k} x^{(1)} \rfloor, \lfloor \sqrt{2k} x^{(1)} \rfloor + 1\}, \dots \\ &\qquad \dots, S_{\lfloor 2kt_p \rfloor} \in \{\lfloor \sqrt{2k} x^{(p)} \rfloor, \lfloor \sqrt{2k} x^{(p)} \rfloor + 1\} \Big| T = 2k+1\Big) \\ &= \sqrt{2k}^p P\left(S_{\lfloor 2kt_1 \rfloor} = \lfloor \sqrt{2k} x^{(1)} + \alpha(1) \rfloor, \dots, S_{\lfloor 2kt_p \rfloor} = \lfloor \sqrt{2k} x^{(p)} + \alpha(p) \rfloor \,\Big|\, T = 2k+1\right)\end{aligned}$$

Consider now the full sequence $(\tilde{p}_k(t_i, x^{(i)}, i = 1 \dots p))_{k \in \mathbb{N}}$. For every subsequence $(\tilde{p}_{k_l})_{l \in \mathbb{N}} \subseteq (\tilde{p}_k)_{k \in \mathbb{N}}$ there exists at least one $\alpha \in \{0,1\}^p$ such that there is a subsequence $(\tilde{p}^{\alpha}_{k_{l_m}})_{m \in \mathbb{N}} \subseteq (\tilde{p}_{k_l})_{l \in \mathbb{N}}$ which converges to the claimed limit as in (3.23). Due to the subsequence principle also $\tilde{p}_k$ must converge to that limit. This proves the convergence of finite dimensional distributions. □

4. Galton-Watson trees

The final goal of this thesis is to determine the scaling limit of Galton-Watson trees. In this chapter we will define finite rooted ordered trees and what we mean by a random tree. Main purpose of this chapter is to establish a mapping between plane trees and non negative excursions. This will allow us to use our previously developed theory of excursions of the random walk and Brownian motion. This chapter is mostly taken from [15].

4.1. Plane trees

We first define finite rooted ordered trees which we will just call plane trees. We define the set of finite sequences of positive integers as

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

where $\mathbb{N}$ denotes the set of positive natural numbers $\mathbb{N} = \{1, 2, \dots\}$ and by convention $\mathbb{N}^0 = \emptyset$. An element of $\mathcal{U}$ later represents a node of the tree. For a sequence $u = (u_1, \dots, u_n) \in \mathcal{U}$ the number $|u| = n$ is called the generation of u. It describes the height of the node u in the tree. The product of two elements in $\mathcal{U}$ is their concatenation: For $u = (u_1, \dots, u_n)$, $v = (v_1, \dots, v_m) \in \mathcal{U}$ we define

$$uv = (u_1, u_2, \dots, u_n, v_1, v_2, \dots, v_m)\,.$$

We further define a projection

$$\pi : \mathcal{U} \to \mathcal{U}, \ \ (u_1, u_2, \dots, u_n) \mapsto (u_1, u_2, \dots, u_{n-1})$$

with $\pi(i) := \emptyset$ and $\pi(\emptyset) = \emptyset$. $\pi(u)$ will later represent the parental node of u. With these definitions a tree can be defined as follows:

Definition 4.1 (Plane Tree)**.** *A plane tree τ is a finite subset of $\mathcal{U}$ such that*

1. *The tree has a root: $\emptyset \in \tau$*
2. *For any node in the tree, the tree contains the entire geneology, so $\pi|^{\tau} : \tau \to \tau$ is a welldefined mapping.*
3. *For any node $u \in \tau$ there exists a number of children $k_u(\tau) \in \mathbb{N}_0$ such that $uj \in \tau$ iff $1 \leq j \leq k_u(\tau)$, $j \in \mathbb{N}$.*

We further define the set of all finite plane trees as $\boldsymbol{A}$ and the set of trees with k edges as $\boldsymbol{A}_k$.

We additionally use the notations $\#\tau$ for the cardinality of τ, which is the number of nodes, and $|\tau| = \#\tau - 1$ for the number of edges. A tree can also be defined recursively, which will become useful to extend the definition to random trees.

Lemma 4.2. *Define natural numbers indexed by $\mathcal{U}$: $(k_u)_{u\in\mathcal{U}} \in \mathbb{N}_0$. We then define a set τ recursively by*

1. $\emptyset \in \tau$
2. $u = (u_1, \ldots, u_n) \in \tau$ *if and only if* $\pi(u) \in \tau$ *and* $u_n \leq k_{(u_1,\ldots,u_{n-1})}$

If τ is finite, then τ is a tree.

Proof. Obviously $\emptyset \in \tau$. Furthermore $\pi(u) = (u_1, \ldots, u_{n-1}) \in \tau$ by definition. Lastly condition three is fulfilled since for $u \in \tau$ and $v = uj$ we have that $v \in \tau$ iff $j \leq k_u$. □

This lemma is helpful to promote the tree to a random variable later on. A tree can always be associated with its contour function or so called Dyck path. Imagine traveling the tree at one edge per time unit from left to right. This will trace out the contour function. It can be formalized as follows.

Definition 4.3 (Contour process)**.** *Let τ be a tree. Define a mapping $\Phi_\tau : \{0, 1, \ldots, 2|\tau|+1\} \to \tau$ recursively. We set $\Phi_\tau(0) := \emptyset$ as the root of the tree and $\Phi_\tau(1) = (1)$. For $i > 1$ we define*

$$D_i = \{u \in \tau \,|\, u \notin \{\Phi_\tau(0), \ldots, \Phi_\tau(i-2)\},\, \pi(u) = \Phi_\tau(i-1)\}$$

as the set of children of $\pi(u)$, which do not lie in the image of $\{0, 1, \ldots, i-2\}$ under Φ_τ. We set

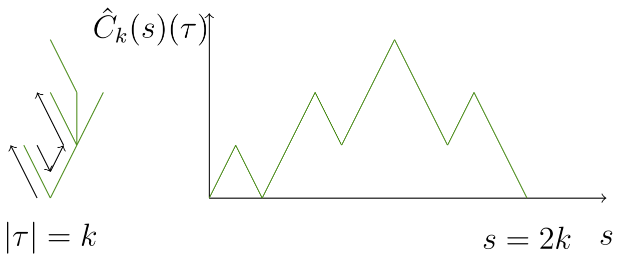
$$\Phi_\tau(i) = \begin{cases} \pi(\Phi_\tau(i-1)) \text{ if } D_i = \emptyset \\ \Phi_\tau(i-1)n' \text{ else, where } n' = \min\{n \in \mathbb{N} \,|\, \Phi_\tau(i-1)n \in D_i\} \end{cases}$$

with the convention $\pi(\emptyset) = \emptyset$. The contour function is defined by the generation of this process, i.e. for $\tau \in \boldsymbol{A}$ and $i \in \{0, 1, \ldots, 2|\tau|\}$ we set

$$\hat{C}_{|\tau|}(i)(\tau) = |\Phi_\tau(i)|\,,$$

which defines a mapping from trees to excursions, $\Psi : \tau \mapsto \hat{C}_{|\tau|}(\cdot)(\tau)$ and we write $\boldsymbol{C} = \Psi(\boldsymbol{A})$, $\boldsymbol{C}_k = \Psi(\boldsymbol{A}_k)$.

By definition, after $i = |\tau|$ every edge is visited exactly once and it holds $\hat{C}_{|\tau|}(0)(\tau) = \hat{C}_{|\tau|}(|\tau|)(\tau) = 0$.



A mapping constructed this way from $\boldsymbol{A}_k$ into the set of all contour functions of length $2k$ is a bijection. It should be noted that this relies on the fact that plane trees have a fixed ordering of vertices from left to right. For our purpose we only need that any Dyck path of length $2k$ of this type yields a tree with k edges. Additionally the correspondence of Dyck paths and trees tells us that the number of trees with size $|\tau| = k$ is given by the kth Catalan number:

Corollary 4.4.

$$\#\boldsymbol{A}_k = \frac{1}{k+1}\binom{2k}{k}$$

Proof. In order to show this we again use the cycle lemma. We need to count the number of non-negative walks with length $2k$. Consider $x_i \in \{-1, 1\}$ for $i \in \{1, 2, \ldots, 2k+1\}$ with

$$\sum_{i=1}^{2k+1} x_i = -1 \tag{4.1}$$

and define for $j = 1, \ldots 2k$

$$\tilde{C}_k(j) := \sum_{i=1}^{j} x_i\,. \tag{4.2}$$

The sum of the first j elements with $j \in \{0, 1, 2, \ldots, 2k\}$ stays non-negative if and only if $\hat{C}_k \in \boldsymbol{C}_k$ where the function $\hat{C}_k : [0, 2k] \to \mathbb{R}$ is constructed by linear interpolation between $\tilde{C}_k(j)$ and $\tilde{C}_k(j+1)$ for $j = 0, 2, \ldots, 2k-1$. In particular $\hat{C}_k(2k) = 0$. If in general a sequence $(x_1, \ldots, x_{2k+1})$ fulfills equation 4.1, then k out of $2k+1$ elements must be equal to one. Due to the cycle lemma there exists exactly one out of $2k+1$ cyclic permutations π such that $(x_{\pi(1)}, \ldots, x_{\pi(2k+1)})$ fulfills equation 4.2. Hence the number of Dyck paths of lengths $2k$ is

$$\#\boldsymbol{A}_k = \#\boldsymbol{C}_k = \frac{1}{2k+1}\binom{2k+1}{k} = \frac{1}{k+1}\binom{2k}{k}$$

□

4.2. Galton-Watson trees

In this section we introduce trees whose branching is described by a Galton-Watson process. Main result of this section is the fact that Galton-Watson trees are distributed uniformly on $\boldsymbol{A_k}$. Galton-Watson processes were introduced to model the inheritance and possible extinction of surnames. A Galton-Watson process can be defined as follows:

Definition 4.5. *A Galton- Watson process is a stochastic process $(Z_n)_{n\in\mathbb{Z}_{\geq 0}}$ which is defined recursively by $Z_0 = 1$ and*

$$Z_{n+1} = \sum_{j=1}^{Z_n} K_j^n$$

where $\{K_j^n\}$ are a set of independent identically distributed natural valued random variables, which describe the number of "children" of individual j of the nth generation.

Since we want to define finite random trees the following lemma is helpful

Lemma 4.6. *If the law of K_j^n (also called offspring distribution) is critical or subcritical, i.e. $\mathbb{E}[K_1^1] \leq 1$ then the process will go extinct with probability one:*

$$\lim_{n\to\infty} P(Z_n = 0) = 1$$

A proof can be found in [19]. In this section we want to promote τ to a random variable by first promoting the number k_u to a random variable K_u. As we have seen in Lemma 4.2 we need to choose these random variables such that the resulting tree is finite. This can be ensured by choosing a critical or subcritical probability measure μ on $\mathbb{Z}_{\geq 0}$ of the number of children, i.e.

$$\sum_{k=0}^{\infty} k\mu(k) \leq 1\,.$$

This is made precise in the following theorem.

Theorem 4.7. *Let $(K_u, u \in \mathcal{U})$ be a collection of independent random variables with critical or subcritical law μ and define θ_μ recursively by*

1. $\emptyset \in \theta_\mu$
2. $u = (u_1, \ldots, u_n) \in \theta_\mu$ *if and only if* $\pi(u) \in \theta_\mu$ *and* $u_n \leq K_{(u_1,\ldots,u_{n-1})}$

In this case $(Z_n,\, n \geq 0)$ defined by $Z_n = \#\{u \in \theta_\mu \,|\, |u| = n\}$ is a Galton- Watson process with offspring distribution μ and initial value $Z_0 = 1$. In particular θ_μ is a.s. a finite plane tree.

Proof. Enumerate all sequences in θ_μ with length n: $u_j, j = 1, 2, \ldots, Z_n$. Then then the total number of nodes of the $n+1$st generation is given by

$$Z_{n+1} = \sum_{j=1}^{Z_n} K_{u_j}$$

Thus $(Z_n,\, n \geq 0)$ is a Galton-Watson process with offspring distribution μ and in the critical or subcritical case we have that the process goes extinct almost surely. Thus, with probability one, θ_μ is a tree. □

By this theorem, θ_μ becomes a random variable as well. Next we want to understand the probabilistic structure of the contour function. For this we introduce the notation of a shifted tree. If τ is a tree and $u \in \tau$, the subtree of τ with root u is denoted by

$$T_u(\tau) = \{v \in \mathcal{U} \,|\, uv \in \tau\}$$

Furthermore we have the following explicit formula for the law of θ_μ:

Proposition 4.8. *For every $\tau \in \boldsymbol{A}$ we have*

$$P(\theta_\mu = \tau) = \prod_{u\in\tau} \mu(k_u(\tau))$$

If the offspring distribution is geometric, i.e.

$$\mu_0(k) = 2^{-k-1}\,,$$

then θ_{μ_0} is a critical Galton-Watson tree and

$$P(\theta_{\mu_0} = \tau) = 2^{-2|\tau|-1}$$

Proof.

$$P(\theta_\mu = \tau) = \prod_{u\in\tau} P(K_u = k_u(\tau)) = \prod_{u\in\tau} \mu(k_u(\tau))$$

We can then compute under $\mu = \mu_0$

$$P(\theta_{\mu_0} = \tau) = \prod_{u\in\tau} 2^{-k_u(\tau)-1} = 2^{\sum_{u\in\tau}(-k_u(\tau)-1)}$$
$$= 2^{-|\tau|-\#\tau} = 2^{-2|\tau|-1}$$

□

In conclusion the measure $P(\theta_{\mu_0} \in \mathrm{d}\tau \,|\, |\theta_{\mu_0}| = k)$ is uniform on $\boldsymbol{A}_k$ and we can also write θ_k for a random tree uniformly distributed on $\boldsymbol{A}_k$.

4.3. Galton-Watson trees and random walks

In this section we see from Theorem 4.9 that the distribution of Galton-Watson trees with geometric offspring distribution yields a contour function distributed according to the excursion of a random walk. Recall the definition of the excursion of a random walk with length $2k$

$$\mathcal{L}((C_k(2kt))_{t\in[0,1]}) = \mathcal{L}((S_{2kt})_{t\in[0,1]}|T = 2k+1)\,.$$

where S is the simple random walk and T the first hitting time of minus one. We have the following identity in distribution:

Theorem 4.9. *Let θ be a $\boldsymbol{A}$-valued random variable, i.e. a random tree, S a simple random walk and θ_S a tree coded by S_n for $n \leq T-1$. Then θ is a Galton-Watson tree with geometric offspring distribution μ_0 if and only if $\theta \overset{d}{=} \theta_S$ or equivalently due to $T = 2|\theta_S| + 1$ iff $\hat{C}_{|\theta|}(\cdot)(\theta) \overset{d}{=} C_{|\theta|}$.*

Proof. Define θ_S to be the tree coded by the first excursion of a simple random walk S, which is a.s. finite. We are going to prove that

$$P(k_u(\theta_S) = m \,|\, u \in \theta_S) = P(k_\emptyset(T_u(\theta_S)) = m|\, u \in \theta_S) = P(k_\emptyset(\theta_S)) = \mu_0(m) \tag{4.3}$$

which shows that $\theta_S \overset{d}{=} \theta$. Let's define the upcrossing times of the random walk from 0 to 1:

$$U_1 = \inf\{n \geq 0 \,|\, S_n = 1\}, \quad V_1 = \inf\{n \geq U_1 \,|\, S_n = 0\}$$
$$U_j = \inf\{n \geq V_{j-1} \,|\, S_n = 1\}, \quad V_j = \inf\{n \geq U_j \,|\, S_n = 0\} \text{ for } j > 0$$

Using these upcrossing times we can count the number of children of the root. For this we consider the random variable:

$$K = \sup\{j \,|\, U_j \leq T\}$$

where by convention $\sup\emptyset = 0$. In this case we have $K = k_\emptyset(\theta_S)$. Whenever a new subtree of θ_S starts, the root and the next descendant of the root must be traveled, which corresponds to the upcrossing times U_j. We now proof that the random walk generates the correct offspring distribution at the root of the tree, i.e.

$$P(K = m) = \frac{1}{2}P(K = m-1) = (1/2)^{m+1} = \mu_0(m)\,.$$

To do this we have

$$\begin{aligned} P(K > m) &= P(S_{V_m+1} = 1, S_i \geq 0 \; \forall i \leq V_m) \\ &= P(S_{V_m+1} = -1, S_i \geq 0 \; \forall i \leq V_m) = P(K = m) \,, \end{aligned}$$

where it was used that S_{V_m+1} is independent of its past due to the strong Markov property. The same holds also in the conditional case:

$$\begin{aligned} P(K > m-1 \,|\, K \geq m-1) &= P(S_{V_{m-1}+1} = 1 \,|\, S_i \geq 0 \; \forall i \leq V_{m-1}) \\ &= P(S_{V_{m-1}+1} = 1 \,|\, S_i \geq 0 \; \forall i \leq V_{m-1}) = P(K = m-1 \,|\, K \geq m-1) \,. \end{aligned}$$

Together this shows the induction step.

$$\begin{aligned} 2P(K = m) &= P(K = m) + P(K > m) = P(K > m-1) \\ &= P(K > m-1 \,|\, K \geq m-1) P(K \geq m-1) \\ &= P(K = m-1 \,|\, K \geq m-1) P(K \geq m-1) = P(K = m-1) \,. \end{aligned}$$

Finally it holds $P(K = 0) = P(T < U_1) = P(S_1 = -1) = 1/2$. This proves that

$$P(k_\emptyset(\theta_S) = m) = P(K = m) = \mu_0(m) \,.$$

Next we generalize above equation to other nodes of the tree. For this we use the function Φ_{θ_S} defined in the last subsection and let

$$i = \inf\{j \in \mathbb{N} \,|\, \Phi_{\theta_S}(j) = u\} \,.$$

be the time at which u is traveled first and $n = |u|$ be the generation of u. Define

$$S'_m = S_{i+m} - n \text{ for } 0 \leq m \leq T'$$

where $T' = \inf\{j \in \mathbb{N} \,|\, S'_j = -1\}$. Clearly S'_m is the excursion of a random walk and it holds due to the Markov property of random walks

$$\begin{aligned} P(k_u(\theta) = m \,|\, u \in \theta) &= P(\#\{0 < l < T' \,:\, S_{i+l} = n\} = m \,|\, i < T) \\ &= P(\#\{0 < l < T' \,:\, S'_l = 0\} = m) = P(k_\emptyset(\theta_S) = m) \,. \end{aligned}$$

Hence a tree coded by a random walk is a μ_0 Galton-Watson tree. On the other hand due to the bijection between Galton-Watson trees and contour functions, the contour function of any Galton-Watson tree must be the excursion of a random walk by above proof. □

5. The Brownian tree as a scaling limit in the Gromov-Hausdorff sense

In this chapter we want to put our previous observations together in order to establish a scaling limit of a sequence of Galton-Watson trees. When taking the limit, the number of vertices of the random trees grows towards infinity. Consequently we need to introduce a larger space than the space of plane trees, the space of real trees. In the following section we describe trees as a special kind of metric space. A distance between two metric spaces is given by the Gromov-Hausdorff distance which is introduced in section two. With these ingredients we can proof the desired limit theorem in section three of this chapter.

5.1. Real trees coded by normalized excursions

We take the definition of real trees from [5] and take the coding of trees by excursions from [14]. We start with the definition of a real compact tree.

Definition 5.1 (Real tree). *Consider a compact metric space $(\mathcal{T}, d)$. $(\mathcal{T}, d)$ is called a real tree if for any $\sigma, \sigma' \in \mathcal{T}$:*

1. *There exists an unique isometry $f_{\sigma,\sigma'} : [0, d(\sigma, \sigma')] \to \mathcal{T}$ such that*

$$f_{\sigma,\sigma'}(0) = \sigma \text{ and } f_{\sigma,\sigma'}(d(\sigma, \sigma')) = \sigma'. \tag{5.1}$$

   We denote the segment between σ and σ' as $[\sigma, \sigma'] := \mathrm{im}(f_{\sigma,\sigma'})$.

2. *For $\eta \in \mathcal{T}$ with $[\sigma, \eta] \cap [\eta, \sigma'] = \{\eta\}$ it follows that*

$$[\sigma, \eta] \cup [\eta, \sigma'] = [\sigma, \sigma'] . \tag{5.2}$$

The tree is rooted if it has a distinguished point $\rho = \rho(\mathcal{T})$. We define the space of rooted real trees as $\mathbb{T}_*$.

The first property means that any two points can be connected by a unique geodesic. The second property implies that there exist no loops in the space. If a loop existed in $\mathcal{T}$, one could choose two different points which satisfy (5.2), giving rise to two different segments $[\sigma, \sigma']$ in contradiction to the first property. The equivalence of this definition with other possible definitions is discussed in [5]. Similar to the contour functions in the previous section, a real tree can be coded by a continuous function. We first define the space of normalized excursions with trivial endpoints:

$$U_0^1 = \{g \in \mathcal{C}([0, 1], \mathbb{R}_{\geq 0}) \mid \ g(0) = g(1) = 0\} \subseteq U^1$$

Proposition 5.2. *Consider a continuous function $g \in U_0^1$ with $g \neq 0$. Define for $s, t \in [0,1]$ the minimum of the function between s and t.*

$$m_g(s,t) = \inf_{r \in [s \wedge t, s \vee t]} g(r)\,. \tag{5.3}$$

We then define the following distance function on $[0,1]$:

$$d_g(s,t) = g(s) + g(t) - 2m_g(s,t)\,. \tag{5.4}$$

The function d_g defines a pseudometric on $[0,1]$, i.e. d_g is symmetric, fulfills the triangle inequality and $d_g(s,s) = 0$ for all $s \in [0,1]$.

Proof. The distance function is symmetric and by definition $d_g(s,s) = 2g(s) - 2g(s) = 0$. We verify the triangle inequality:

$$\begin{aligned} d_g(s,t) &= g(s) + g(t) - 2m_g(s,t) \\ &= g(s) + g(r) + g(t) + g(r) - 2m_g(s,t) - 2g(r) \end{aligned}$$

Now if $s < r < t$ then either $m_g(s,t) = m_g(s,r)$ or $m_g(s,t) = m_g(r,t)$. Assume without loss of generality the first case. Then

$$-2m_g(s,t) - 2g(r) \leq -2m_g(s,t) - 2m_g(r,t) = -2m_g(s,r) - 2m_g(r,t)\,. \tag{5.5}$$

The case $m_g(s,t) = m_g(r,t)$ follows the same argument. If on the other hand $s < t < r$ then by definition $m_g(s,r) \leq m_g(s,t)$ and we have again the case (5.5), and if $r < s < t$, then $m_g(r,t) \leq m_g(s,t)$, and we again can argue as before. Thus in all possible cases we have

$$\begin{aligned} d_g(s,t) &= g(s) + g(r) + g(t) + g(r) - 2m_g(s,t) - 2g(r) \\ &\leq g(s) + g(r) - 2m_g(s,r) + g(t) + g(r) - 2m_g(r,t) = d_g(s,r) + d_g(r,t)\,. \end{aligned}$$

□

We define the relation $s \sim_g t$ iff $d_g(s,t) = 0$. It is symmetric due to the symmetry of d_g, reflexive because $d_g(s,s) = g(s) + g(s) - 2g(s) = 0$ and transitive due to the triangle inequality and the non negativity of d_g. As usual we can now define the canonical projection

$$p_g : [0,1] \to [0,1]/\sim_g\,,\quad s \mapsto [s]$$

and the metric space induced by the pseudometric:

$$\begin{aligned} \mathcal{T}_g &:= [0,1]/\sim_g \\ d_g^*(p_g(s), p_g(t)) &:= d_g(s,t) \end{aligned}$$

This metric space $(\mathcal{T}_g, d_g^*)$ now defines a compact real tree. The tree becomes a rooted tree by setting $\rho = p_g(0)$. Since $d_g^* \circ p_g = d_g$ only depends on the equivalence class we omit the asterik in the notation and plainly write d_g for d_g^* on the set of equivalence classes. The positive definiteness of the metric is obtained from the definition and all other properties are inherited. The space $\mathcal{T}_g$ is compact and connected as an image of the compact interval $[0,1]$ under the continuous mapping p_g. We now need to check that $(\mathcal{T}_g, d_g)$ in fact defines a real tree. To prove this statement we need an interesting theorem which holds on metric spaces defined as above. Its intuitive meaning is that we can always reroot the "tree" to an arbitrary other point by a natural isometry.

Rerooting

Lemma 5.3 (Rerooting Lemma). *Let $s_0 \in [0,1)$ and define for $r \in \mathbb{R}_{\geq 0}$ $\bar{r} = r - \lfloor r \rfloor$ or equivalently the projection of r onto the fractional part $\bar{r} \in [0,1)$. We define $g'[s_0] \in U_0^1$ by*

$$g'[s_0](s) = d_g(s_0, \overline{s_0 + s}) \quad \textit{for } s \in [0,1].$$

In this case $(\mathcal{T}_{g'[s_0]}, d_{g'[s_0]})$ is a welldefined metric space and there exists a unique isometry $R[s_0] : (\mathcal{T}_{g'[s_0]}, d_{g'[s_0]}) \to (\mathcal{T}_g, d_g)$ with

$$R[s_0](p_{g'[s_0]}(s)) = p_g(\overline{s_0 + s})\,.$$

For every $s_0 \in [0,1)$ the mapping $\Phi[s_0] : U_0^1 \to U_0^1$, $g \mapsto g'[s_0]$ is a bijection with inverse $\Phi[1 - s_0]$.

Proof. We start by proving that for $s, t \in [0,1]$

$$d_{g'[s_0]}(s,t) = d_g(\overline{s_0 + s}, \overline{s_0 + t})\,. \tag{5.6}$$

We can assume without loss of generality $s < t$. We will distinguish several cases of $m_{g'[s_0]}$. From now on we fix $s_0 \in [0,1)$ and write g' for $g'[s_0]$.

First case, $s, t \in [0, 1 - s_0]$: It then holds $g(\overline{s_0 + s}) = g(s_0 + s)$. We have two possibilities.
i) Assume that the infimum is attained before $s_0 + s$,

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

It always holds for $r \in [s,t]$ that $m_g(s_0, s_0 + r) \leq m_g(s_0, s_0 + s)$ such that

$$m_g(s_0, s_0 + r) = m_g(s_0, s_0 + s) \tag{5.7}$$

and thus

$$\begin{aligned} m_{g'}(s,t) &= g(s_0) + \inf_{r \in [s \wedge t, s \vee t]} [g(s_0 + r) - 2m_g(s_0, s_0 + r)] \qquad (5.8)\\ &\overset{(5.7)}{=} g(s_0) + \inf_{r \in [s \wedge t, s \vee t]} g(s_0 + r) - 2m_g(s_0, s_0 + s) \\ &\overset{\text{Def}}{=} g(s_0) + m_g(s_0 + s, s_0 + t) - 2m_g(s_0, s_0 + s) \\ &\overset{(5.7)}{=} g(s_0) + m_g(s_0 + s, s_0 + t) - m_g(s_0, s_0 + t) - m_g(s_0, s_0 + s) \end{aligned}$$

where in the last line our assumption was used again that the minimum on the intervall $[s_0, s_0+t]$ is alsoready attained in $[s_0, s_0 + s]$. We can then compute:

$$\begin{aligned} d_{g'}(s,t) =& 2g(s_0) + g(s_0 + s) + g(s_0 + t) - 2m_g(s_0, s_0 + s) - 2m_g(s_0, s_0 + t) \\ &- 2g(s_0) + 2m_g(s_0, s_0 + s) + 2m_g(s_0, s_0 + t) - 2m_g(s_0 + s, s_0 + t) \\ \overset{(5.8)}{=}& g(s_0 + s) + g(s_0 + t) - 2m_g(s_0 + s, s_0 + t) = d_g(s_0 + s, s_0 + t) \end{aligned}$$

ii) Now consider the other case, so $m_g(s_0 + s, s_0 + t) < m_g(s_0, s_0 + s)$. Because g is continuous and $m_g(s_0 + s, s_0 + t) < m_g(s_0, s_0 + s)$ there exists with the intermediate value theorem:

$$\tilde{r} = \inf\{r \in [s,t] \,|\, g(s_0 + r) \leq m_g(s_0, s_0 + s)\} \tag{5.9}$$

We then have for $r \in [\tilde{r}, t]$ due to $\tilde{r} \leq r$

$$g(s_0 + r) - 2m_g(s_0, s_0 + r) \geq -m_g(s_0, s_0 + r) \geq -m_g(s_0, s_0 + \tilde{r}) = -m_g(s_0, s_0 + s)\,. \tag{5.10}$$

Thus we see that the infimum of the first term in (5.10) is attained in $[s, \tilde{r}]$. From the definition (5.9) on the other hand we see that $g(s_0 + r) \geq g(s_0 + \tilde{r}) = m_g(s_0, s_0 + s) = m_g(s_0, s_0 + r)$ for $r \in [s, \tilde{r}]$. We then have:

$$\begin{aligned} \inf_{r\in[s,t]}[g(s_0 + r) - 2m_g(s_0, s_0 + r)] &\overset{(5.10)}{=} \inf_{r\in[s,\tilde{r}]}[g(s_0 + r) - 2m_g(s_0, s_0 + r)] \\ &\overset{(5.9)}{=} \inf_{r\in[s,\tilde{r}]}[g(s_0 + r)] - 2m_g(s_0, s_0 + s) \\ &\overset{(5.9)}{=} -m_g(s_0, s_0 + s) \end{aligned} \tag{5.11}$$

Finally again using that the minimum on $[s_0, s_0 + t]$ is attained in $[s_0 + s, s_0 + t]$:

$$\begin{aligned} d_{g'}(s,t) =& g(s_0 + s) + g(s_0 + t) - 2m_g(s_0, s_0 + s) - 2m_g(s_0, s_0 + t) \overset{(5.11)}{+} 2m_g(s_0, s_0 + s) \\ =& g(s_0 + s) + g(s_0 + t) - 2m_g(s_0, s_0 + t) = g(s_0 + s) + g(s_0 + t) - 2m_g(s_0 + s, s_0 + t) \\ =& d_g(s_0 + s, s_0 + t) \end{aligned}$$

Second Case, $s, t \in [1 - s_0, 1]$: In this case we have $\overline{s_0 + t} = t - (1 - s_0) = 1 - (1 - t + 1 - s_0)$. The complicate way of writing this operation gives the advantage to relate the second case to the first one. We define the function which runs g in reverse as $\tilde{g}$, i.e. $\tilde{g}(x) = g(1 - x)$ for $x \in [0, 1]$ as well as $\tilde{s}_0 = 1 - s_0$ and

$$\tilde{g}'[\tilde{s}_0](x) = \tilde{g}(\tilde{s}_0) + \tilde{g}(\overline{\tilde{s}_0 + x}) - 2m_{\tilde{g}}(\tilde{s}_0, \overline{\tilde{s}_0 + x})$$

In this case

$$\tilde{g}'[\tilde{s}_0](x) = g'[s_0](1 - x)$$

and $1 - s, 1 - t \in [0, 1 - \tilde{s}_0)$. The first case applies and we get

$$\begin{aligned} d_{g'[s_0]}(s,t) &= d_{\tilde{g}'[\tilde{s}_0]}(1 - s, 1 - t) = d_{\tilde{g}}(\tilde{s}_0 + 1 - s, \tilde{s}_0 + 1 - t) \\ &= d_g(s - (1 - s_0), t - (1 - s_0)) = d_g(\overline{s_0 + s}, \overline{s_0 + t})\,. \end{aligned}$$

Third Case, $s \in (0, 1 - s_0)$, $t \in (1 - s_0, 1)$: The equation $d_g(s,t) = d_{g'}(\overline{s_0 + s}, \overline{s_0 + t})$ is then equivalent to

$$\begin{aligned} m_g(\overline{s_0 + s}, s_0 + s) + m_g(s_0, \overline{s_0 + t}) + \inf_{s\leq r\leq t}[g(\overline{s_0 + r}) - 2m_g(s_0, \overline{s_0 + r})] \\ = m_g(\overline{s_0 + s}, \overline{s_0 + t})\,. \end{aligned} \tag{5.12}$$

Note that in this case $\overline{s_0 + s} = s_0 + s \in (s_0, 1)$ and $\overline{s_0 + t} = s_0 + t - 1 \in (0, s_0)$. Thus we can write

$$I = [\,\overline{s_0 + s}, \overline{s_0 + t}\,] = [s_0 + t - 1, s_0] \cup [s_0, s_0 + s] = I_1 + I_2$$

where I_1 and I_2 denote the corresponding intervals. Since I is compact and g continuous, the minimum is attained at some point and lies either in I_1 or I_2. Assume is is attained in I_1. In this case (5.12) reduces to

$$m_g(s_0, s_0 + s) + \inf_{s\leq r\leq t}[g(\overline{s_0 + r}) - 2m_g(s_0, \overline{s_0 + r})] = 0\,.$$

For $s \le r \le 1 - s_0$ we have

$$m_g(s_0, s_0 + r) \le m_g(s_0, s_0 + s)$$

and for $1 - s_0 \le r \le t$

$$m_g(s_0, \overline{s_0 + r}) \le m_g(s_0 + t - 1, s_0) \le m_g(s_0, s_0 + s) \tag{5.13}$$

where the last inequality follows from our assumption that the minimum in I is attained in I_1. Thus by using (5.13) for one term of $m_g(s_0, \overline{s_0 + r})$:

$$\begin{aligned} m_g(s_0, s_0 + s) + &\inf_{s \le r \le t} [g(\overline{s_0 + r}) - 2m_g(s_0, \overline{s_0 + r})] \\ &\ge \inf_{s \le r \le t} [g(\overline{s_0 + r}) - m_g(s_0, \overline{s_0 + r})] \ge 0\,. \end{aligned}$$

On the other hand we have that $g(s_0 + \cdot)$ reaches the values $g(s_0 + 1 - s_0) = 0$ and $g(s_0 + s)$ on $[s, t]$. Due to the continuity of g, $0 \le m_g(s_0, s_0 + s) \le g(s_0 + s)$ and the intermediate value theorem there exists in particular a $\tilde{s} \in [s, t]$ such that $g(s_0 + \tilde{s}) = m_g(s_0, s_0 + \tilde{s}) = m_g(s_0, s_0 + s)$ and hence

$$\begin{aligned} m_g(s_0, s_0 + s) + &\inf_{s \le r \le t} [g(\overline{s_0 + r}) - 2m_g(s_0, \overline{s_0 + r})] \\ &\le m_g(s_0, s_0 + s) + [g(\overline{s_0 + \tilde{s}}) - 2m_g(s_0, \overline{s_0 + \tilde{s}})] = 0 \end{aligned}$$

This now yields the claimed result. The case in which the minimum is attained in I_2 is symmetric to this one.

Φ **defines a bijection**: This follows from (5.6). We have by definition $g(t) = d_g(0, t)$. Appling the given mapping to g'_{s_0} yields:

$$\begin{aligned} (g'[s_0])'[1 - s_0](t) &= d_{g''[1-s_0]}(0, t) = d_{g'[s_0]}(1 - s_0, \overline{1 - s_0 + t}) \\ &= d_g(\overline{s_0 + 1 - s_0}, \overline{s_0 + \overline{1 - s_0 + t}}) = d_g(0, t) = g(t) \end{aligned}$$

$R[s_0]$ **is an isometry:** Since d_g is a welldefined function on $\mathcal{T}_g$ we have

$$d^*_{g'}(p_{g'}(s), p_{g'}(t)) = d_{g'}(s, t) = d_g(\overline{s_0 + s}, \overline{s_0 + t}) = d^*_g(R(p_{g'}(s)), R(p_{g'}(t)))\,.$$

□

Theorem 5.4. *The metric space $(\mathcal{T}_g, d_g)$ with root $\rho = p_g(0)$ is a real rooted tree in above sense.*

Proof. A sketch of this proof was given by [14]. **Step 1**, construction of the isometry $f_{\sigma,\sigma'}$: Consider two points of the tree $\sigma, \sigma' \in \mathcal{T}_g$. We define partial order on $\mathcal{T}_g$ by

$$\sigma \preceq \sigma' \Leftrightarrow d_g(\sigma, \sigma') = d_g(\rho, \sigma') - d_g(\rho, \sigma)\,.$$

Intuitively σ lies along the branch that connects the root with σ' and in that sense is an ancestor of σ'. We can check that $\preceq$ defines a partial order on $\mathcal{T}_g$. The relation is clearly reflexive, $\sigma \preceq \sigma$. If further $\sigma \preceq \sigma'$ and $\sigma' \preceq \sigma$ then

$$d_g(\sigma, \sigma') = d_g(\rho, \sigma') - d_g(\rho, \sigma) = -d_g(\sigma, \sigma')$$

and thus $d_g(\sigma, \sigma') = 0$. Transitivity follows from the triangle inequality. For $\sigma \preceq \eta$ and $\eta \preceq \sigma'$ it is

$$\begin{aligned} d_g(\sigma, \sigma') &\leq d_g(\sigma, \eta) + d_g(\eta, \sigma') = d_g(\rho, \sigma') - d_g(\rho, \sigma) \\ &\leq d_g(\rho, \sigma) + d_g(\sigma, \sigma') - d_g(\rho, \sigma) = d_g(\sigma, \sigma') \end{aligned}$$

Hence $d_g(\sigma, \sigma') = d_g(\rho, \sigma') - d_g(\rho, \sigma)$ and thus $\sigma \preceq \sigma'$. We will make use of the following crucial characterization of ancestry: For $\sigma = p_g(s) \preceq \sigma' = p_g(t)$ it holds

$$\begin{aligned} d_g(\sigma, \sigma') = g(s) + g(t) - 2m_g(s,t) = d_g(\rho, \sigma') - d_g(\rho, \sigma) = g(t) - g(s) \\ \Rightarrow \boldsymbol{g(s) = m_g(s,t)} \end{aligned} \tag{5.14}$$

We want to construct the isometry $f_{\sigma,\sigma'} : [0, d_g(\sigma, \sigma')] \to \mathcal{T}_g$ with $f(0) = \sigma$ and $f(d_g(\sigma, \sigma')) = \sigma'$. Let $s_0 \in p_g^{-1}(\sigma)$. We apply the rerooting lemma, which for given σ, σ' returns a metric space $(\mathcal{T}_{g'[s_0]}, d_{g'[s_0]})$ and an isometry $R[s_0] : \mathcal{T}_{g'[s_0]} \to \mathcal{T}_g$ with $R[s_0](\rho) = \sigma$. We further define $\tau = R[s_0]^{-1}(\sigma')$ and search without loss of generality for isometries $f : [0, d_{g'[s_0]}(\rho, \tau)] \to \mathcal{T}_{g'[s_0]}$ with $f(0) = \rho$ and $f(d_{g'}(\rho, \tau)]) = \tau$. If we find f we can then set

$$f_{\sigma,\sigma'} := R[s_0] \circ f\,,$$

which fulfills all requirements. We fix s_0 and write g' for $g'[s_0]$. Next define for $s \in p_{g'}^{-1}(\tau)$ and $a \in [0, d_{g'}(\rho, \tau)]$:

$$s_a = \inf\{r \in [0, s] \,|\, m_{g'}(r, s) = a\}\,.$$

This element exists because $a \leq d_{g'}(\rho, \tau) = g'(t)$. It holds that

$$g'(s_a) = m_{g'}(s_a, s) = a\,.$$

If this was not true, i.e. $g'(s_a) > a$, there would exist due to the intermediate value theorem an $\tilde{r} < s_a$ with $g'(\tilde{r}) = a$, in contradiction to the definition of s_a. On the other hand $g'(s_a) \geq m_{g'}(s_a, s)$ by definition. We can then construct the isometry as:

$$a \mapsto f(a) = p_{g'}(s_a)\,.$$

By its definition, $p_{g'}(s_a) \preceq \tau$, so intuitively f maps the distance of a point and the root to an ancestor of τ with that distance. f indeed defines an isometry: Clearly $f(0) = p_{g'}(0) = \rho$ and $f(d_{g'}(\rho, \sigma))) = p_{g'}(s) = \tau$. Furthermore we have for $0 \leq a \leq b \leq d_{g'}(\rho, \tau)$

$$d_{g'}(f(a), f(b)) = g'(s_a) + g'(s_b) - 2m_{g'}(s_a, s_b)$$

Since

$$\begin{aligned} m_{g'}(s_a, s) = a \leq b = m_{g'}(s_b, s) \\ \Rightarrow s_a \leq s_b \end{aligned}$$

we have

$$m_{g'}(s_a, s_b) \leq m_{g'}(s_b, s)$$

such that

$$m_{g'}(s_a, s_b) = m_{g'}(s_a, s) = a\,.$$

We can conclude that f is an isometry, because

$$d_{g'}(f(a), f(b)) = g'(s_a) + g'(s_b) - 2m_{g'}(s_a, s_b) = a + b - 2a = b - a\,.$$

The uniqueness can be proven the following way: Suppose another isometry $\tilde{f}$ with $\tilde{f}(0) = \rho$ and $\tilde{f}(d_{g'}(\rho,\tau)) = \tau$ exists. Then

$$\begin{aligned} d_{g'}(\tau, \tilde{f}(a)) &= d_{g'}(\tilde{f}(d_g(\rho,\tau), \tilde{f}(a))) \\ &= d_{g'}(\rho,\tau) - a = d_{g'}(\rho,\tau) - d_{g'}(\rho, \tilde{f}(a)) \end{aligned}$$

which shows that any isometry only maps to ancestors of τ, i.e. $\tilde{f}(a) \preceq \tau$. Next we write this point of the tree as $\tilde{f}(a) = p_{g'}(t)$ for $t \in [0, d_{g'}(\rho,\tau)]$. Then we have, as was noted in (5.14), $a = g'(t) = m_{g'}(t,s))$ and hence

$$d_{g'}(f(a), \tilde{f}(a)) = 2a - 2m_{g'}(s_a, t) = 2a - 2a = 0$$

Now we define the branch that connects the two points $\sigma, \sigma' \in \mathcal{T}_g$ as

$$[\sigma,\sigma'] := \{\tau \in \mathcal{T}_g \,|\, d_g(\sigma,\sigma') = d_g(\sigma,\tau) + d_g(\tau,\sigma')\}\,.$$

Note that for any $\eta \in [\rho,\sigma]$ it holds that

$$d(\eta,\sigma) = d(\rho,\sigma) - d(\rho,\eta)\,, \tag{5.15}$$

and thus $\eta \preceq \sigma$. This shows that the image of f is indeed the segment $[\rho,\tau]$. Applying $R[s_0]$ we see that $R[s_0] \circ f([0, d_g(\sigma,\sigma')]) = [\sigma,\sigma']$ which justifies our notation for a segment $[\sigma,\sigma']$.

Step 2 Let $\sigma, \sigma', \eta \in \mathcal{T}_g$ such that

$$[\sigma,\eta] \cap [\sigma',\eta] = \{\eta\}\,. \tag{5.16}$$

We claim that $[\sigma,\eta] \cup [\sigma',\eta] = [\sigma,\sigma']$. We write $p_g(s) = \sigma, p_g(s') = \sigma'$. The Rerooting Lemma yields an isometry which maps segments to segments and we can assume without loss of generality that $\eta = \rho$ is the root of the tree. Furthermore we can assume $s \leq s'$. We get from (5.16) that

$$g(t) = m_g(s,t) = m_g(s',t) \;\Leftrightarrow\; g(t) = 0\,, \tag{5.17}$$

where it was used that $p_g(t) = \tau \in [\rho,\sigma]$ if and only if $\tau \preceq \sigma$, which implies for the values of g the characterization (5.14). Furthermore $p_g(t) = \rho$ iff $g(t) = 0$. Choose now $\tau \in [\sigma,\sigma']$ and choose a representant of $p_g(t) = \tau$. $\tau \in [\sigma,\sigma']$ is equivalent to

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

After all cancelations we are left with $m_g(s,s') \overset{!}{=} m_g(s',t) + m_g(s,t) - g(t)$. We first of all have

$$m_g(s,s') \geq \min\{m_g(s',t), m_g(s,t)\}\,. \tag{5.18}$$

On the other hand there must exist an $a \in [s,s']$ in which the minimum on that interval is attained, such that with (5.16)

$$g(a) = m_g(s,a) = m_g(a,s') \;\Rightarrow\; g(a) = m_g(s,s') = 0\,.$$

Consequently either $m_g(s,t)$ or $m_g(s',t)$ must vanish. We can conclude from (5.17) and (5.18) that $\tau \in [\sigma, \sigma']$ if and only if

$$g(t) = \max \begin{cases} m_g(s,t) & \Leftrightarrow \tau \preceq \sigma \\ m_g(s',t) & \Leftrightarrow \tau \preceq \sigma' \end{cases}$$

and thus $\tau \in [\sigma, \sigma']$ if and only if $\tau \in [\rho, \sigma] \cup [\rho, \sigma']$. Together with the rerooting to other points, the first step and the previously mentioned compactness this shows that $(\mathcal{T}_g, d_g)$ is a compact real tree.

□

5.2. Gromov-Hausdorff metric

As we just saw we can understand a real tree as a compact metric space. In the next step we want to translate the convergence of excursions of the random walk to the normalized excursion of Brownian motion to trees. The convergence described before then takes place in a space of compact real trees. This space must be employed with a metric, the Gromov-Hausdorff metric. This part follows [6] and [9]. The Hausdorff distance compares closed subsets of some metric space and assigns a distance to them.

Definition 5.5 (Hausdorff distance)**.** *Consider a metric space (Z, d) and a closed subset $C \subseteq Z$. We define the r neighbourhood of C with respect to d as*

$$U_r(C) := \{p \in Z \,|\, \mathrm{dist}(p, C) < r\} \,.$$

For closed subsets $A, B \subseteq Z$ the Hausdorff distance in (Z, d) is defined as

$$d_H^d(A, B) := \inf\{r > 0 \,|\, A \subseteq U_r(B),\ B \subseteq U_r(A)\} \,.$$

Theorem 5.6. *The set $\mathfrak{M}(Z)$ of compact subsets of Z together with d_H^d is a metric space.*

Proof. The compactness ensures the finiteness of d_H^d. The Hausdorff distance d_H^d is non negative and symmetric, which is evident from its definition. For the triangle inequality choose $A, B, C \in \mathfrak{M}(Z)$. Define $r_1 := d_H^d(A, B)$ and $r_2 := d_H^d(B, C)$. Let $\varepsilon > 0$ be arbitrary. For any $a \in A$ there exists by definition a $b \in B$ and for this b an element $c \in C$ such that

$$d(a, c) \leq d(a, b) + d(b, c) < r_1 + r_2 + 2\varepsilon \,,$$

because the triangle inequality holds in (Z, d). Hence for all $\varepsilon > 0$ $A \subseteq U_{r_1+r_2+\varepsilon}(C)$. By exchanging the roles of A and C it is $C \subseteq U_{r_1+r_2+\varepsilon}(A)$ and we obtain

$$d_H^d(A, C) \leq r_1 + r_2 + 2\varepsilon = d_H^d(A, B) + d_H^d(B, C) + 2\varepsilon \,.$$

Since ε was arbitrary this shows the triangle inequality. To see that the distance is positive definite we assume that A, B are closed sets and $A \neq B$. In this case there exists an element $b \in B$ with $r := \mathrm{dist}(b, A) > 0$. Such an element exists because $A^c \neq \emptyset$ is open. This implies $b \notin U_r(A)$ and thus $d_H(A, B) \geq r > 0$. □

Later we will compare trees as metric space together with an distinguished point, the root. The Hausdorff distance can easily be extended to pointed subsets of (Z, d).

Definition 5.7 (Pointed Hausdorff distance)**.** *Let (A, a) and (B, b) be two pointed closed subsets of (Z, d). We define the pointed Hausdorff distance*

$$d^d_{H*}((A, a), (B, b)) := d^d_H(A, B) + d(a, b)\,, \tag{5.19}$$

which is a metric as a sum of metrics.

The Gromov-Hausdorff distance measures the distance between metric spaces. The idea behind it is to isometrically embed metric spaces into a "larger" metric space and to compare their distance there with help of the Hausdorff distance. One then takes the infimum over all such embeddings.

Definition 5.8 (Pointed Gromov-Hausdorff distance)**.** *Let (X, ρ_x, d_X) and (Y, ρ_y, d_Y) be pointed compact metric spaces. Consider a metric space (Z, d) and isometries $f_X : X \to X' \subseteq Z$, $f_Y : Y \to Y' \subseteq Z$ and denote this as a tupel as (Z, d, f_X, f_Y). The pointed Gromov-Hausdorff distance is then defined by*

$$d_{GH}((X, \rho_x, d_X), (Y, \rho_y, d_Y)) := \inf_{(Z,d,f_X,f_Y)} d^d_{H*}((f_X(X), f_X(\rho_x)), (f_Y(Y), f_Y(\rho_y))) \tag{5.20}$$

The compactness of metric spaces ensures that their images under isometries are compact (since these are in particular continuous), which means that the Hausdorff metric is always finite. Before we check that such a definition can be made and that the Gromov-Hausdorff distance actually defines a metric on some appropriate space, we first give an alternative description. So far the metric space (Z, d) appears to be arbitrary, which makes it hard to work with. As it turns out we can always choose the disjoint union $X \sqcup Y := X \times \{1\} \cup Y \times \{2\}$ for Z with some appropriate metric. By the disjoint union we can make any two metric spaces disjoint by redefining $X' = X \times \{1\}$, $\rho'_x = (\rho_x, 1)$ and for $x_1, x_2 \in X$ $d_{X'}((x_1, 1), (x_2, 1)) = d_X(x_1, x_2)$. Thus in the following we will just assume that X, Y being disjoint and write X for the embedding of X into $X \sqcup Y$. Y is treated the same way. The way to define a metric on $X \sqcup Y$ is to first define points in X and Y that correspond to each other.

Definition 5.9 (Correspondence and distortion)**.** *A correspondence is a subset or relation $\mathfrak{R} \subseteq X \times Y$ such that for any $x \in X$ there exists a $y \in Y$ such that $(x, y) \in \mathfrak{R}$ and for every $y \in Y$ there exists an element $x \in X$ with $(x, y) \in \mathfrak{R}$.*
The distortion of a correspondence $\mathfrak{R}$ is defined as

$$\text{dis}\mathfrak{R} := \sup\{|d_X(x, x') - d_Y(y, y')| \,:\, (x, y), (x', y') \in \mathfrak{R}\} \tag{5.21}$$

With this we are able to define a metric on $X \sqcup Y$ and we will also need correspondences later again.

Definition 5.10. *Let (X, ρ_x, d_X) and (Y, ρ_y, d_Y) be pointed compact metric spaces. A metric d on the disjoint union $X \sqcup Y$ is called admissible, if $d|_{X \times X} = d_X$ and $d|_{Y \times Y} = d_Y$ and such a metric can be always constructed.*

Proof. Let $\mathfrak{R}$ be a correspondence with positive distortion on $X \times Y$ and define

$$r = \text{dis}\mathfrak{R} \tag{5.22}$$

Next define the admissible metric on $X \times Y$ as

$$d(x_1, y_1) = \frac{r}{2} + \inf\{d_X(x_1, x') + d_Y(y', y_1) \,|\, (x', y') \in \mathfrak{R}\} =: d(y_1, x_1)\,.$$

The metric is by definition symmetric and positive definite. Further we have for $x_1, x_2 \in X$:

$$\begin{aligned} d(x_1, x_2) &\leq \inf\{d_X(x_1, x') + d_X(x', x'') + d_X(x'', x_2) \,|\, x', x'' \in X\} \\ &\overset{(5.22)}{\leq} \inf\{d_X(x_1, x') + d_Y(y', y'') + r + d_X(x'', x_2) \,|\, (x', y'), (x'', y'') \in \mathfrak{R}\} \\ &\leq r + \inf\{d_X(x_1, x') + d_Y(y', y_1) \,|\, (x', y') \in \mathfrak{R}\} + \inf\{d_Y(y_1, y'') + d_X(x'', x_2) \,|\, (x'', y'') \in \mathfrak{R}\} \\ &= d(x_1, y_1) + d(y_1, x_2) \end{aligned}$$

And for $x_1 \in X, y_1 \in Y$:

$$\begin{aligned} d(x_1, y_1) &\leq \frac{r}{2} + \inf\{d_X(x_1, x_2) + d_X(x_2, x') + d_Y(y', y_1) \,|\, (x', y') \in \mathfrak{R}\} \\ &= d(x_1, x_2) + d(x_2, y_1)\,. \end{aligned}$$

The other inequalities follow exchanging roles of X and Y. □

We now get the following alternative description of d_{GH}:

Lemma 5.11. *The Gromov-Hausdorff distance is given by*

$$d_{GH}((X, d_X, \rho_x), (Y, d_Y, \rho_y)) = \inf\{d^d_{H*}((X, \rho_x), (Y, \rho_y)) \,|\, d \text{ admissible on } X \sqcup Y\} \tag{5.23}$$

Proof. To define an admissible metric on $X \sqcup Y$ we only need to fix the values on $X \times Y$. For any metric space (Z, d_Z) and isometric embeddings f_X, f_Y we can define an admissible metric for $x \in X, y \in Y$ and $\varepsilon > 0$ by

$$d_{X \sqcup Y}(x, y) = \varepsilon + d_Z(f_X(x), f_Y(y))\,.$$

This defines a metric as an inherited property from Z, since it holds for $x_1, x_2 \in X, y_1 \in Y$

$$d_{X \sqcup Y}(x_1, x_2) = d_Z(f_X(x_1), f_X(x_2)) \leq d_Z(f_X(x_1), f_Y(y_1)) + d(f_Y(y_1), f_X(x_2)) + 2\varepsilon\,,$$

which proves the triangle inequality after similar arguments in the other cases, and the positive definiteness holds by definition. Let $r = d^{d_Z}_{H*}(f(X, \rho_x), f(Y, \rho_y))$. For every $x \in X$ there exists a $y \in Y$ such that

$$d_Z(f_X(x), f_Y(y)) = d^{X \sqcup Y}(x, y) - \varepsilon < r + \varepsilon\,.$$

Conversely there exists a $x \in X$ for every $y \in Y$ such that $d_{X \sqcup Y}(x, y) < r + 2\varepsilon$. Thus the Hausdorff distance turns out to be

$$d^{d_{X \sqcup Y}}_{H*}((X, \rho_x), (Y, \rho_y)) < d^{d_Z}_{H*}(f(X, \rho_x), f(Y, \rho_y)) + 2\varepsilon$$

By taking the infimum over all possible tupels of metric spaces and isometries, (Z, d_Z, f_X, f_Y) and the corresponding admissable metrics, the same inequality also holds for the Gromov-Hausdorff distance. As ε was arbitrary it follows

$$\inf\{d^d_{H*}((X, \rho_x), (Y, \rho_y)) \,|\, d \text{ admissible on } X \sqcup Y\} \leq d_{GH}((X, \rho_x, d_X), (Y, \rho_y, d_Y))$$

On the other hand, for any admissible metric d, $(X \sqcup Y, d)$ is a metric space and the canonical embeddings $\iota_X : X \to X \sqcup Y, x \mapsto (x, 1)$ and ι_Y are isometries. Thus, taking the infimum,

$$\inf\{d^d_{H*}((X, \rho_x), (Y, \rho_y)) \,|\, d \text{ admissible on } X \sqcup Y\} \geq d_{GH}((X, \rho_x, d_X), (Y, \rho_y, d_Y))$$

□

Theorem 5.12. *The Gromov-Hausdorff distance fulfills on the class of compact metric spaces the triangle inequality, is symmetric and non negative.*

Proof. Symmetry and non negativity of d_{GH} are again clear from the definition. We want to verify the triangle inequality. Let X, Y, Z be compact pointed metric spaces. By definition, for any $\varepsilon > 0$ there exist admissible metrics d_{XY} and d_{YZ} such that

$$d_{H*}^{d_{XY}}((X,\rho_x),(Y,\rho_y)) < d_{GH}((X,\rho_x,d_X),(Y,\rho_y,d_Y)) + \varepsilon$$
$$d_{H*}^{d_{YZ}}((Y,\rho_y),(Z,\rho_z)) < d_{GH}((Y,\rho_y,d_Y),(Z,\rho_z,d_Z)) + \varepsilon$$

Using these define

$$d_{XZ}(x_1,z_1) = \inf_{y_1\in Y}(d_{XY}(x_1,y_1) + d_{YZ}(y_1,z_1))\,.$$

Again non negativity and symmetry are obvious from the definition. For the triangle inequality we have

$$\begin{aligned} d_{XZ}(x_1,z_1) + d_{XZ}(z_1,x_2) &= \inf_{y_1\in Y}\{d_{XY}(x_1,y_1) + d_{YZ}(y_1,z_1)\} + \inf_{y_2\in Y}\{d_{XY}(x_2,y_2) + d_{YZ}(y_2,z_1)\} \\ &\geq \inf_{y_1\in Y}\inf_{y_2\in Y}\{d_{XY}(x_1,y_1) + d_Y(y_1,y_2) + d_{XY}(x_2,y_2)\} \\ &\geq \inf_{y_1\in Y}\inf_{y_2\in Y}\{d_{XY}(x_1,y_2) + d_{XY}(x_2,y_2)\} \geq d_X(x_1,x_2)\,. \end{aligned}$$

and correspondingly for reversed roles of Z and X. The other cases can be treated in a similar way. This gives three admissible metrics d_{XY}, d_{YZ} and d_{XZ}. These together define an admissible metric d_{XYZ} on $X \sqcup Y \sqcup Z$, where each of them is defined on the corresponding disjoint union, while on X, Y, Z we still keep the metrics d_X, d_Y, d_Z. For example, for $x \in X, y \in Y, z \in Z$ we have

$$\begin{aligned} d_{XYZ}(x,z) + d_{XYZ}(z,y) &= \inf_{y'\in Y}\{d_{XY}(x,y') + d_{YZ}(y',z)) + d_{YZ}(z,y)\} \\ &\geq \inf_{y'\in Y}\{d_{XY}(x,y') + d_{XY}(y',y)\} \geq d_{XY}(x,y) \end{aligned}$$

We can now use the triangle inequality of the usual Hausdorff convergence. We have

$$\begin{aligned} d_{GH}((X,\rho_x,d_X),(Z,\rho_z,d_Z)) &\leq d_{H*}^{d_{XYZ}}((X,\rho_x),(Z,\rho_z)) \\ &\leq d_{H*}^{d_{XYZ}}((X,\rho_x),(Y,\rho_y)) + d_{H*}^{d_{XYZ}}((Y,\rho_y),(Z,\rho_z)) \\ &= d_{H*}^{d_{XY}}((X,\rho_x),(Y,\rho_y)) + d_{H*}^{d_{YZ}}((Y,\rho_y),(Z,\rho_z)) \\ &< d_{GH}((X,\rho_x,d_X),(Y,\rho_y,d_Y)) + d_{GH}((Y,\rho_y,d_Y),(Z,\rho_z,d_Z)) + 2\varepsilon \end{aligned} \tag{5.24}$$

Since ε was chosen arbitrarily we obtain the triangle inequality. □

Finally we want positive definiteness of the pseudometric. For this we show that $d_{GH}(X,Y) = 0$ if and only if X and Y are isometric. Two pointed metric spaces (X,ρ_x,d_X) and (Y,ρ_y,d_Y) are called isometric iff there exists an isometry $f : (X,d_X) \to (Y,d_Y)$ with $f(\rho_x) = \rho_y$.

Theorem 5.13. *For compact metric spaces (X,ρ_x,d_X) and (Y,ρ_y,d_Y) it holds that*

$$d_{GH}((X,\rho_x,d_X),(Y,\rho_y,d_Y)) = 0$$

if and only if the two spaces are isometric.

Proof. If (X, ρ_x, d_X) and (Y, ρ_y, d_Y) are isometric there exists a map f such that $f(X, \rho_x, d_X) = (Y, \rho_y, d_Y)$. By the definition of the Hausdorff distance, $d_{H*}^{d_Y}(f(X, \rho_x), (Y, \rho_y)) = 0$, and thus in particular the Gromov-Hausdorff distance vanishes.
We now prove the converse. Assume that $d_{GH}((X, \rho_x, d_X), (Y, \rho_y, d_Y)) = 0$ and (X, ρ_x, d_X), (Y, ρ_y, d_Y) are not isometric. Since the Gromov-Hausdorff distance vanishes there exists a sequence of admissible metrics d_n on $X \sqcup Y$ such that $d_{H*}^{d_n}((X, \rho_x), (Y, \rho_y)) < 1/n$. The spaces X and Y are compact and thus, in particular, seperable. Let $X' = \{x^0 = \rho_x, x^1, x^2, \dots\}$ be a countable dense subset of X and $y^0 = \rho_y$. By definition $d_n(x^0, y^0) < 1/n$. There exists a sequence $(y_n^1)_{n\in\mathbb{N}}$ such that $d_n(x^1, y_n^1) < 1/n$ for all $n \in \mathbb{N}$. Y is compact, so there exists a subsequence $(y_{n_m}^1)_{m\in\mathbb{N}}$ which converges to some value $y^1 \in Y$. Using the triangle inequality:

$$d_{n_m}(x^1, y^1) \le d_{n_m}(x^1, y_{n_m}^1) + d_{n_m}(y_{n_m}^1, y^1) \xrightarrow{m\to\infty} 0$$

So by passing to the subsequence of d_n and iteration of the process for $i = 2, 3, \dots$ we obtain by a diagonal argument a sequence of admissible metrics $(d_{n_m})_{m\in\mathbb{N}}$ on $X \sqcup Y$ and a sequence $\{y^0, y^1, \dots\}$ such that $d_{n_m}(x^i, y^i) \xrightarrow{m\to\infty} 0$ for all $i \in \mathbb{N}$. We now define the mapping $f : X' \to Y$, $x^i \mapsto y^i$. We then have, using the triangle inequality and $d_{n_m}|_X = d_X, d_{n_m}|_Y = d_Y$, since d_{n_m} are admissible metrics,

$$\begin{aligned} &d_Y(f(x^i), f(x^j)) - d_X(x^i, x^j) = d_{n_m}(f(x^i), f(x^j)) - d_{n_m}(x^i, x^j) \\ &\le d_{n_m}(f(x^i), x^i) + d_{n_m}(x^i, x^j) + d_{n_m}(x^j, f(x^j)) - d_{n_m}(x^i, x^j) \xrightarrow{m\to\infty} 0\,. \end{aligned}$$

The left hand side is independent of m, so it has to be less or equal to zero. The same argument holds for $d_X(x^i, x^j) - d_Y(f(x^i), f(x^j))$. Thus we constructed an isometry $f : X' \to Y$. Since f is an isometry it is in particular uniformly continuous and can be uniquely extended to $X = \overline{X'}$. This is again an isometry, since for $a, b \in X$ and arbitrary $\varepsilon > 0$ there exist due to the density of X' and continuity of f elements $x^i, x^j \in X'$ such that

$$d_Y(f(a), f(b)) \le d_Y(f(a), f(x^i)) + d_X(x^i, x^j) + d_Y(f(x^j), f(b)) \le d_X(x^i, x^j) + 2\varepsilon \le d_X(a, b) + 4\varepsilon\,.$$

As ε was arbitrary it follows that $f : X \to Y$ is an isometry. By exchanging roles of X and Y we can also construct an isometry from Y to X. Thus X, Y are isometric. □

Define the equivalence relation: $(X, \rho_x, d_X) \sim (Y, \rho_y, d_Y)$ iff X, Y are isometric pointed metric spaces. By the preceeding discussion we now see that the set $(\mathbb{T}_*/\sim, d_{GH})$ of compact real rooted trees up to isometry becomes a metric space.

5.3. Convergence of random trees in the Gromov-Hausdorff metric

We now want to apply this newly gained knowledge to the case where we consider the convergence of random Galton-Watson trees as convergence of compact pointed metric spaces. For this we need yet another characterization of the Gromov-Hausdorff convergence. Recall the definition of the distortion (5.21). Lemma 5.11 has the following important corollary.

Corollary 5.14. *The Gromov-Hausdorff metric is given by*

$$d_{GH}((X, \rho_x, d_X), (Y, \rho_y, d_Y)) = \frac{1}{2} \inf_{\mathfrak{R}(X,Y), (\rho_x, \rho_y)\in\mathfrak{R}} \mathrm{dis}\mathfrak{R}$$

where the infimum runs over all possible correspondences $\mathfrak{R}(X,Y) \subseteq X \times Y$ that contain the pair of roots.

Proof. Let first $d_{GH}((X,\rho_x,d_X),(Y,\rho_y,d_Y)) = r/2$. By definition it must exist for every $\varepsilon > 0$ an admissible metric d on $X \sqcup Y$ such that $d_H^d((X,\rho_x),(Y,\rho_y)) < r/2 + \varepsilon$. Since this means that the distance of no point in X to Y is larger or equal to $r/2+\varepsilon$ and vice versa, this also implies that for all $x \in X$ there exist $y \in Y$ such that $d(x,y) < r/2+\varepsilon$ and for all $y \in Y$ there exist $x \in X$ such that $d(x,y) < r/2+\varepsilon$. This implies that the relation

$$\mathfrak{R} = \{(x,y) \in X \times Y \,|\, d(x,y) < r/2 + \varepsilon\}$$

is a correspondence. Note that due to $d_H^d((X,\rho_x),(Y,\rho_y)) < r/2+\varepsilon$ also $(\rho_x,\rho_y) \in \mathfrak{R}$. Then for all $(x,y),(x',y') \in \mathfrak{R}$ by the triangle inequality

$$|d(x,x') - d(y,y')| \leq d(x,y) + d(y',x') < r + 2\varepsilon.$$

and thus, taking the supremum with respect to all pairs $(x,y),(x',y') \in \mathfrak{R}$ we get $\mathrm{dis}\mathfrak{R} \leq r+\varepsilon$. By letting ε go to zero we infer

$$\inf_{\mathfrak{R}(X,Y),(\rho_x,\rho_y)\in\mathfrak{R}} \mathrm{dis}\mathfrak{R} \leq 2d_{GH}((X,\rho_x,d_X),(Y,\rho_y,d_Y))\,.$$

To proof the converse inequality take any correspondence $\mathfrak{R}$ between (X,ρ_x,d_X) and (Y,ρ_y,d_Y) with $(\rho_x,\rho_y) \in \mathfrak{R}$ and define $r = \mathrm{dis}\mathfrak{R}$. We can assume $r \neq 0$. If $r = 0$ the spaces (X,ρ_x,d_X) and (Y,ρ_y,d_Y) are isometric. Because of Theorem 5.13 the Gromov-Hausdorff distance vanishes as well and the claimed equality is plainly satisfied. For $r \neq 0$ we define the admissible metric on $X \sqcup Y$

$$d(x_1,y_1) = \frac{r}{2} + \inf\{d_X(x_1,x') + d_Y(y',y_1) \,|\, (x',y') \in \mathfrak{R}\} =: d(y_1,x_1)$$

for $x_1 \in X, y_1 \in Y$. It was already shown in Definition 5.10 that this indeed defines a metric. Whenever (x,y) are in correspondence we have $d(x,y) = r/2$. Thus for any $x \in X$ there exists a $y \in Y$ with $d(x,y) = r/2$ and for any $y \in Y$ there exists a $x \in X$ with $d(x,y) = r/2$. Thus

$$d_{GH}((X,\rho_x,d_X),(Y,\rho_y,d_Y)) \leq d_{H*}^d((X,\rho_x),(Y,\rho_y)) \leq \frac{r}{2}\,,$$

which proves our claim. □

This theorem has the following crucial consequence.

Theorem 5.15. *For any g, g' continuous with $g(0) = g(1) = g'(0) = g'(1)$ we have*

$$d_{GH}((\mathcal{T}_g,d_g,\rho),(\mathcal{T}_{g'},d_{g'},\rho')) \leq 2\|g-g'\|_\infty\,. \tag{5.25}$$

In particular the map $\boldsymbol{\mathcal{T}} : (U_0^1, \|\cdot\|_\infty) \to (\mathbb{T}_*/\sim, d_{GH})$, $g \mapsto [(\mathcal{T}_g,\rho_g,d_g)]$, where $[(\mathcal{T}_g,\rho_g,d_g)]$ denotes the equivalence class of trees isometric to $(\mathcal{T}_g,\rho_g,d_g)$, is continuous with respect to the topology of uniform convergence and Gromov-Hausdorff topology.

Proof. For g, g' we define the following correspondence:

$$\mathfrak{R} = \{(a, a') \in \mathcal{T}_g \times \mathcal{T}_{g'} \mid \exists\, t \in [0,1] \text{ s.t. } p_g(t) = a, p_{g'}(t) = a'\}\,.$$

We then have for $s, t \in [0,1]$ which project to $(a, a'), (b, b') \in \mathfrak{R}$

$$\begin{aligned}|d_g(a,b) - d_{g'}(a',b')| &= |g(s) + g(t) - 2m_g(s,t) - g'(s) - g'(t) + 2m_{g'}(s,t)| \\ &\leq |g(s) - g'(s)| + |g(t) - g'(t)| + 2|m_g(s,t) - m_{g'}(s,t)| \\ &\leq 4\|g - g'\|_\infty\end{aligned}$$

By taking the supremum we obtain $\mathrm{dis}\mathfrak{R} \leq 4\|g - g'\|_\infty$. □

This now yields our final theorem

Theorem 5.16 (Convergence of random Galton-Watson trees towards the CRT)**.** *Let θ_k be a homogeneously distributed random variable on $\boldsymbol{A}_k$ and d_{gr} the graph distance. Then the sequence of rescaled random trees converges as a sequence of metric spaces in the union of compact real trees and plane trees $(\mathbb{T}_* \cup \boldsymbol{A}/\sim, d_{GH})$ in distribution to the Brownian tree*

$$\left(\theta_k, \emptyset, \frac{1}{\sqrt{2k}} d_{\mathrm{gr}}\right) \xrightarrow{d} (\mathcal{T}_{\mathbb{e}}, \rho_{\mathbb{e}}, d_{\mathbb{e}})\,. \tag{5.26}$$

Proof. We know that the contour function of the random tree, $C_k(\theta)$, is the excursion of a random walk of length $2k$. Define for $t \in [0,1]$

$$c_k(t)(\theta_k) = \frac{1}{\sqrt{2k}} C_k(2kt)(\theta_k)\,.$$

We already know from chapter three that the rescaled excursions converge to the Brownian normalized excursion

$$c_k(\theta_k) \xrightarrow[k\to\infty]{d} \mathbb{e}\,.$$

Since $\boldsymbol{\mathcal{T}} : U_0^1 \to \mathbb{T}_*/\sim$ is continuous it follows from the Continuous Mapping Theorem 1.10 that in $(\mathbb{T}_*/\sim, d_{GH})$

$$(\mathcal{T}_{c_k(\theta_k)}, \rho_{c_k(\theta_k)}, d_{c_k(\theta_k)}) \xrightarrow[k\to\infty]{d} (\mathcal{T}_{\mathbb{e}}, \rho_{\mathbb{e}}, d_{\mathbb{e}})\,.$$

The only difference between the metric spaces $\left(\theta_k, \emptyset, 1/\sqrt{2k} d_{\mathrm{gr}}\right)$ and $(\mathcal{T}_{c_k(\theta_k)}, \rho_{c_k(\theta_k)}, d_{c_k(\theta_k)})$ is the metric which, however, vanishes as $k \to \infty$. To see this we recall the mapping in Definition 4.3, $\Phi : \{0, 1, \ldots, 2k\} \to \theta_k$ and define an embedding $\theta_k \to \mathcal{T}_{c_k(\theta_k)}$, $\Phi(i) \mapsto p_{c_k(\theta_k)}(i/2k)$ for suitable $i \in \{0, 1, \ldots, 2k\}$. We then define the correspondence

$$\mathfrak{R}' := \{(u, \sigma) \in \theta_k \times \mathcal{T}_{c_k(\theta_k)} : d_{c_k(\theta_k)}(u, \sigma) \leq 1/\sqrt{2k}\}\,.$$

This in fact defines a correspondence: The graph distance is given for $u, v \in \theta_k$ by

$$\begin{aligned}d_{\mathrm{gr}}(u,v) = |u| + |v| - 2|u \wedge v| &= C_k(\theta_k)(i) + C_k(\theta_k)(j) - 2m_{C_k(\theta_k)}(i,j) \\ &= \sqrt{2k} d_{c_k(\theta_k)}(i/2k, j/2k)\end{aligned}$$

where $i, j \in \{0, 1, \ldots, 2k\}$ are chosen such that with the mapping in Definition 4.3 $\Phi_{\theta_k}(i) = u, \Phi_{\theta_k}(j) = v$ and where $u \wedge v$ denotes the last common ancestor of u and v. Due to the Lipschitz continuity of $C_k(\theta_k)$ we now have for every $t \in [0,1]$:

$$d_{c_k(\theta_k)}(t, \lfloor 2kt \rfloor/2k) \leq 1/\sqrt{2k}\ |t - \lfloor 2kt \rfloor/2k| \leq 1/\sqrt{2k}.$$

Consequently we can find for every $t \in [0,1]$ an $i \leq 2k$ such that $p_{c_k(\theta_k}(t)$ corresponds to $p_{c_k(\theta_k})(i)$ and the converse is trivial. We then have for $(u,\sigma) = (\Phi_{\theta_k}(i), p_{c_k(\theta_k)}(s)) \in \mathfrak{R}'$ and $(v,\sigma') = (\Phi_{\theta_k}(j), p_{c_k(\theta_k)}(t)) \in \mathfrak{R}'$:

$$|1/\sqrt{2k}d_{\mathrm{gr}}(u,v) - d_{c_k(\theta_k)}(\sigma,\sigma')| \leq d_{c_k(\theta_k)}(u,\sigma) + \mathrm{d}_{c_k(\theta_k)}(v,\sigma') \leq 2/\sqrt{2k}$$

Using again the distortion description of the Gromov-Hausdorff distance we see that

$$d_{GH}\left(\left(\mathcal{T}_{c_k(\theta_k)}, \rho_{c_k(\theta_k)}, d_{c_k(\theta_k)}\right), \left(\theta_k, \emptyset, 1/\sqrt{2k}d_{\mathrm{gr}}\right)\right) \leq \frac{1}{\sqrt{2k}} \xrightarrow{k\to\infty} 0\,.$$

In particular this convergence must hold in probability, and using Slutzky's theorem we finally have shown that

$$\left(\theta_k,\, \emptyset,\, \frac{1}{\sqrt{2k}}d_{\mathrm{gr}}\right) \xrightarrow{d} (\mathcal{T}_{\mathbb{e}}, \rho_{\mathbb{e}}, d_{\mathbb{e}})\,.$$

□

A. Notations

Measure theory

For a metric space (S, d) we denote the Borel σ-algebra generated by the open sets in X by $\mathcal{B}(S)$. For the integral of a Borel-measureable function $f : S \to \mathbb{R}$ with respect to a measure μ in $\mathcal{B}(S)$ we may write

$$\mu[f] := \int_S f \mathrm{d}\mu = \int_S f(x)\mu(\mathrm{d}x)\,.$$

For a random variable $X : (\Omega, \mathcal{F}, P) \to S$ we have the following alternative notations for the image measure $P(X^{-1}(\cdot))$ on $\mathcal{B}(S)$:

$$X_*P,\;\; P^X,\;\; \mathcal{L}(X),\;\; P(X \in \cdot)$$

The third notation reads as the law of X and the last notation as the probability of X being in A. We may also write the integral with respect to the image measure as

$$\int_S f(x)P(X \in \mathrm{d}x) := \int_S f(x)\mathrm{d}P^X(x)\,.$$

$P(A|B)$ denotes the probability of A conditioned B.

Excursions

For a stochastic process on a sufficiently rich probability space the shift operators are denoted as $\theta_t : \Omega \to \Omega,\;\; X_s(\theta_t(\omega)) = X_{t+s}(\omega)$. The law under which X starts at x is denoted by P^x, i.e. $X_0 = x$ P^x-a.s.. The corresponding expectation is denoted by E_x.
For a Markov process X the stochastic process $\hat{X}$ denotes the killed process. The life time of the process is denoted by ζ, in our case usually $\zeta := \inf\{t > 0 : X_t = 0\}$. The space of excursions is defined as

$$U := \{w \in \mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R}) \,|\, 0 < \zeta(w) < \infty \text{ and } w(t) = 0 \text{ for } t \geq \zeta(w)\} \subseteq W$$

equipped with the metric

$$d_U(u_1, u_2) = \sup_{t \geq 0} |u_1(t) - u_2(t)| + |\zeta(u_1) - \zeta(u_2)|\,,$$

U^1 denotes the space of normalized excursions, U^1_0 the space of normalized excursions with trivial endpoints. We use by convention that $\inf \emptyset = \infty$.

o-Notation and rounding

We denote the rounding operations for $t \in \mathbb{R}$ by $\lfloor t \rfloor := \max\{n \in \mathbb{Z} : n \leq t\}$ and $\lceil t \rceil := \min\{n \in \mathbb{Z} : n \geq t\}$. We say for real valued sequences $(a_k)_{k\in\mathbb{N}}$, $(b_k)_{k\in\mathbb{N}}$ that

$$a_k \in o(b_k) \Leftrightarrow \lim_{k\to\infty} a_k/b_k = 0\,.$$

To simplify the notation we write $a_k = b_k + o(c_k)$ iff there exists a sequence $(c'_k)_{k\in\mathbb{N}} \in o(c_k)$ such that $a_k = b_k + c'_k$. It for example holds for $t \in (0,1)$ and $i \in \mathbb{N}$ that

$$\lfloor 2kt \rfloor + i = 2kt(1 + o(1))\,.$$

One only needs to be careful with the equal sign, since it is not a transitive relation anymore and has to be interpreted in above sense. Similarly we write $a_k \in \mathcal{O}(b_k)$ iff $\limsup_{k\to\infty} a_k/b_k < \infty$.

Trees

We denote the minimum of $s, t \in \mathbb{R}$ as $s \wedge t := \min\{s,t\}$ and the maximum as $s \vee t := \max\{s,t\}$. With this motivation we denote in a real tree $\mathcal{T}$ $\sigma \wedge \sigma'$ as the last common ancestor of σ, σ' and in a plain tree $\tau \in \boldsymbol{A}$ $u \wedge v$ as the last common ancestor of u and v.

Distances

We denote the distance of $x \in Z$ for a metric space (Z,d) to a set $C \subseteq Z$ as $\mathrm{dist}(x, C) := \inf\{d(x,y) : y \in C\}$.

B. Wiener space

In this chapter we want to give a short description how Wiener space can be constructed, as in this thesis Brownian motion is always viewed as a random variable on the space of real valued continuous functions on the positive halfline. For this we make use of Kolmogorov's extension theorem to exchange the probability space.

Definition B.1 (Equivalent processes)**.** *Let $(\Omega, \mathcal{F}, P)$ and $(\Omega', \mathcal{F}', P')$ be two probability spaces and $X = (X_t)_{t \in T}$ a stochastic process on Ω, $Y = (Y_t)_{t \in T}$ a stochastic process on Ω', both with values in $(\mathbb{R}^{nT}, \mathcal{B}^{nT})$. X, Y are said to be equivalent or versions of each other iff they have the same finite dimensional distributions.*

Definition B.2 (Canonical process)**.** *Define a mapping $\varphi : \Omega \to \mathbb{R}^T$ by*

$$\varphi(\omega) = X_\cdot(\omega)$$

and coordinate mappings

$$Y_t(w) := w(t)$$

for any $w \in \mathbb{R}^T$. Then $Y_t \circ \varphi$ is measurable for each $t \in T$, and the image measure of P under φ denoted by $\varphi^* P = P \circ \varphi^{-1}$ fulfills for $A_i \in \mathcal{B}$

$$P(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n) = \varphi^* P(Y_{t_1} \in A_1, \ldots, Y_{t_n} \in A_n)\,.$$

The process Y is an equivalent version of X and Y is called the canonical process equivalent to X.

We first pass to the canonical version equivalent to Brownian motion, i.e. we define a probability space $(\mathbb{R}^{[0,\infty)}, \mathcal{B}^{[0,\infty)}, P^B)$ and a stochastic process B with $B_t(w) = w(t)$ for $w \in \mathbb{R}^{[0,\infty)}$, which is distributed according to the law $P^B = B_* P$. It would be natural to change our probability space to the continuous functions $W := \mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R})$. However, this set is not measurable in $(\mathbb{R}^{[0,\infty)}, \mathcal{B}^{[0,\infty)})$, [24]. Instead we define the trace σ-algebra

$$\mathcal{A} := \mathcal{B}^{[0,\infty)} \cap \mathcal{C}(\mathbb{R}_{\geq 0}, \mathbb{R})\,.$$

On $\mathcal{A}$ we define a new measure Q as $Q(A) = P^B(\bar{A})$ for $A \in \mathcal{A}$ and $\bar{A} \in \mathcal{B}^{[0,\infty)}$ with $\bar{A} \cap W = A$.

Lemma B.3. *Q is a welldefined probability measure on $\mathcal{A}$*

Proof. To check that Q is welldefined assume we choose for $A \in \mathcal{A}$ measurable sets $\bar{A}_1, \bar{A}_2 \in \mathcal{B}^{[0,\infty)}$ with $A = \bar{A}_i \cap W$. We have

$$(\bar{A}_1 \setminus \bar{A}_2 \cup \bar{A}_2 \setminus \bar{A}_1) \cap W = \emptyset\,.$$

Since B is continuous almost surely there exists a set N with $P^B(N) = 0$ such that $N^C \subseteq W \subseteq (\bar{A}_1 \setminus \bar{A}_2 \cup \bar{A}_2 \setminus \bar{A}_1)^C$. Due to the monotony and subadditivity of the measure we can conclude

$$1 = P^B(N^C) \leq P(\bar{A}_1 \setminus \bar{A}_2 \cup \bar{A}_2 \setminus \bar{A}_1)^C) \leq 1$$

and thus for the complementary events

$$P^B(\bar{A}_1 \setminus \bar{A}_2) = P^B(\bar{A}_2 \setminus \bar{A}_1) = 0\,.$$

This implies the unambiguity of Q:

$$P^B(\bar{A}_1) = P^B(\bar{A}^1 \cap \bar{A}^2) + P^B(\bar{A}^1 \setminus \bar{A}^2 = P^B(\bar{A}^1 \cap \bar{A}^2) + P^B(\bar{A}^2 \setminus \bar{A}^1) = P^B(\bar{A}_2)$$

The other properties are now clear from the definition. □

In the following we will just write P instead of Q and define the Wiener-space $(W, \mathcal{A}, P)$ as probability space for Brownian motion. In this paragraph we followed a concise description of the construction of Wiener space in [2] and [24]. The new σ-algebra $\mathcal{A}$ is very convenient, as it is the Borel σ-algebra induced by uniform convergence on compact subsets.

Theorem B.4. *The Borel-σ-algebra on W induced by the topology of uniform convergence coincides with the trace-σ-algebra $\mathcal{A}$.*

A proof can be found in [24] Chapter XIII, 1.2.

Bibliography

[1] D. Aldous. The continuum random tree. i. *Ann. Probab.*, 19(1):1–28, 01 1991.

[2] C. Bayer. Brownian motion and ito calculus. *WK-summer camp 2006*, 2006.

[3] P. Billingsley. *Probability and Measure*. John Wiley&Sons, 3 edition, 1995.

[4] P. Billingsley. *Convergence of Probability Measures*. John Wiley&Sons, 2 edition, 1999.

[5] I. Chiswell. *Introduction to [lambda]-trees*. World Scientific, 2001.

[6] S. I. Dmitri Burago, Yuri Burago. A course in metric geometry, 2001.

[7] A. Dvoretzky and T. Motzkin. A problem of arrangements. *Duke Math. J.*, 14(2):305–313, 06 1947.

[8] K. Itô. Poisson point processes attached to markov processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Probability Theory*, pages 225–239, Berkeley, Calif., 1972. University of California Press.

[9] D. Jansen. Notes on pointed gromov-hausdorff convergence. 03 2017.

[10] R. W. K.L. Chung. *Introduction to Stochastic Integration*. Birkhäuser, 2nd edition, 1990.

[11] J. Kühn. *Limit theorems for random excursions*. PhD thesis, Freiburg University, 2013.

[12] D. L.C.G. Rogers. *Diffusions, Markov Processes, and Martingales*, volume 1. 2 edition, 2000.

[13] D. L.C.G. Rogers. *Diffusions, Markov Processes, and Martingales*, volume 2. 2 edition, 2000.

[14] J.-F. Le Gall. Random trees and applications. *Probab. Surveys*, 2:245–311, 2005.

[15] J.-F. Le Gall and G. Miermont. Scaling limits of random trees and planar maps. 01 2011.

[16] E. Mammen. Probability theory ii, lecture notes, 2018.

[17] J. Pitman. Combinatorial stochastic processes. 2002.

[18] R. G. R.M. Blumenthal. *Markov Processes and Potential Theory*, volume 29. Elsevier, 1968.

[19] S. Roch. Branching processes, 2015.

[20] T. Tao. 275a, notes 5: Variants of the central limit theorem, 2015.

[21] S. Watanabe. Ito's theory of excursion point processes and its developments. *Elsevier*, 2009.

[22] W. Werner. Poisson point processes, excursions and stable processes in two-dimensional structures. *Elsevier*, 2010.

[23] J. Yeh. *Real Analysis: Theory of Measure and Integration.* World Scientific, 2006.

[24] M. Yor and D. Revuz. *Continuous Martingales and Brownian motion.* Springer, 3rd edition, 1999.

Erklärung

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Ort, Datum | Unterschrift