Brownian continuum objects:
the excursion, tree, snake, map, web, net, and castle

J.M. Swart

February 17, 2022

Contents

Chapter 1

Topological prerequisites

1.1 Topological spaces

A *topological space* is a set $\mathcal{X}$ equipped with a collection $\mathcal{O}$ of subsets of $\mathcal{X}$ that are called *open* sets, such that

(i) If $(O_\gamma)_{\gamma\in\Gamma}$ is any collection of (possibly uncountably many) sets $O_\gamma \in \mathcal{O}$, then $\bigcup_{\gamma\in\Gamma} O_\gamma \in \mathcal{O}$.

(ii) If $O_1, O_2 \in \mathcal{O}$, then $O_1 \cap O_2 \in \mathcal{O}$.

(iii) $\emptyset, \mathcal{X} \in \mathcal{O}$.

Any such collection of sets is called a *topology*. It is fairly standard to also assume the *Hausdorff* property

(iv) For each $x_1, x_2 \in \mathcal{X}$, $x_1 \neq x_2$ $\exists O_1, O_2 \in \mathcal{O}$ s.t. $O_1 \cap O_2 = \emptyset$, $x_1 \in O_1$, $x_2 \in O_2$.

A set $V \subset \mathcal{X}$ is a *neighbourhood* of a point $x \in \mathcal{X}$ if $x \in O \subset V$ for some $O \in \mathcal{O}$. We let $\mathcal{V}_x$ denote the set of all neighbourhoods of x. A *fundamental system* of neighbourhoods of x is a set $\mathcal{V}'_x \subset \mathcal{V}_x$ such that

$$\forall V \in \mathcal{V}_x \ \exists V' \in \mathcal{V}'_x \text{ s.t. } V' \subset V.$$

For example, the set of all $O \in \mathcal{O}$ such that $x \in O$ is a fundamental system of neighbourhoods of x. A sequence of points $x_n \in \mathcal{X}$ converges to a limit x in a given topology $\mathcal{O}$ if for each $V \in \mathcal{V}_x$ there is an n such that $x_m \in V$ for all $m \geq n$. It suffices to check this condition for a fundamental system of neighbourhoods $\mathcal{V}'_x$. If the topology is Hausdorff, then limits are unique, i.e., $x_n \to x$ and $x_n \to x'$ implies $x = x'$.

If $(\mathcal{X}, \mathcal{O})$ is a topological space (with $\mathcal{O}$ the collection of open subsets of $\mathcal{X}$) and $\mathcal{X}' \subset \mathcal{X}$ is any subset of $\mathcal{X}$, then $\mathcal{X}'$ is also naturally equipped with a topology given by the collection of open subsets $\mathcal{O}' := \{O \cap \mathcal{X}' : O \in \mathcal{O}\}$. This topology is called the *induced* topology from $\mathcal{X}$. If $x_n, x \in \mathcal{X}'$, then $x_n \to x$ in the induced topology on $\mathcal{X}'$ if and only if $x_n \to x$ in $\mathcal{X}$.

A *basis* of a topology is a subset $\mathcal{O}' \subset \mathcal{O}$ such that each element of $\mathcal{O}$ can be written as the union of (possibly uncountably many) elements of $\mathcal{O}'$. Equivalently, this says that

$$\mathcal{O} = \{O \subset \mathcal{X} : \forall x \in O\ \exists O' \in \mathcal{O} \text{ s.t. } x \in O' \subset O\}.$$

If $\mathcal{O}'$ is a basis for $\mathcal{O}$, then $\mathcal{V}'_x := \{O \in \mathcal{O}' : x \in O\}$ is a fundamental system of neighbourhoods of x. A topology is *first countable* if every $x \in \mathcal{X}$ has a countable fundamental system of neighbourhoods. A topology is *second countable* if there exists a countable basis of the topology.

A set $C \subset \mathcal{X}$ is called *closed* if its complement is open. Because of property (i) in the definition of a topology, for each $A \subset \mathcal{X}$, the union of all open sets contained in A is itself an open set. We call this the *interior* of A, denoted as $\mathrm{int}(A) := \bigcup\{O : U \subset A,\ O \text{ open}\}$. Then clearly $\mathrm{int}(A)$ is the smallest open set contained in A. Similarly, by taking complements, for each set $A \subset \mathcal{X}$ there exists a smallest closed set containing A. We call this the *closure* of A, denoted as $\overline{A} := \bigcap\{C : C \supset A,\ C \text{ closed}\}$. If the topology is first countable, then

$$\overline{A} = \{x \in \mathcal{X} : \exists x_n \in \mathcal{X} \text{ s.t. } x_n \to x\}, \tag{1.1}$$

i.e., $\overline{A}$ is the set of all limits of sequences in A. A similar statement holds for general topological spaces if we replace sequences by the more general concept of a *net*, that we will not discuss here. Since a set is closed if and only if it coincides with its closure, it follows from (1.1) that in a first countable topological space, knowing all convergent sequences and their limits uniquely determines the closed sets and their complements, the open sets, and hence the whole topology.

A topological space is called *separable* if there exists a countable set $D \subset \mathcal{X}$ such that D is dense in $\mathcal{X}$, where we say that a set $D \subset \mathcal{X}$ is *dense* if its closure is $\mathcal{X}$, or equivalently, if every nonempty open subset of $\mathcal{X}$ has a nonempty intersection with D.

A *metric* on a set $\mathcal{X}$ is a function $d : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ such that for all $x, y, z \in \mathcal{X}$,

(i) $d(x, y) = d(y, x)$,

(ii) $d(x, z) \leq d(x, y) + d(y, z)$,

(iii) $d(x,y) = 0$ implies $x = y$.

A *metric space* is a space with a metric defined on it. If d is a metric on $\mathcal{X}$, and $B_\varepsilon(x) := \{y \in \mathcal{X} : d(x,y) < \varepsilon\}$ denotes the open ball around x of radius ε, then

$$\mathcal{O} := \big\{O \subset \mathcal{X} : \forall x \in O \ \exists \varepsilon > 0 \text{ s.t. } B_\varepsilon(x) \subset O\big\}$$

defines a Hausdorff topology on $\mathcal{X}$ such that convergence $x_n \to x$ in this topology is equivalent to $d(x_n, x) \to 0$. Note that the open balls form a basis for this topology. Since open balls of radius $1/n$ around a point x form a fundamental system of neighbourhoods, metric spaces are first countable. We say that the metric d *generates* the topology $\mathcal{O}$. If for a given topology $\mathcal{O}$ there exists a metric d that generates $\mathcal{O}$, then we say that the topological space $(\mathcal{X}, \mathcal{O})$ is *metrisable*. Such a metric, if it exist, can always be chosen such that it is bounded. For example, if d is any metric on $\mathcal{X}$, then $d'(x,y) := d(x,y) \wedge 1$ is a bounded metric that generates the same topology. A metrisable space is always first countable. It is second countable if and only if it is separable.

A sequence x_n in a metric space $(\mathcal{X}, d)$ is a *Cauchy sequence* if for all $\varepsilon > 0$ there is an n such that $d(x_k, x_l) \le \varepsilon$ for all $k, l \ge n$. A metric space is *complete* if every Cauchy sequence converges. Every metric space $(\mathcal{X}, d)$ has a *completion*, i.e., there exists a complete metric space $(\overline{\mathcal{X}}, \overline{d})$ such that $\mathcal{X} \subset \overline{\mathcal{X}}$ is dense and the metric on $\mathcal{X}$ is the *induced metric* from $\overline{\mathcal{X}}$, i.e., $d(x,y) = \overline{d}(x,y)$ for all $x, y \in \mathcal{X}$. Such a completion is unique up to isometries.

A *Polish space* is a separable topological space $(\mathcal{X}, \mathcal{O})$ such that there exists a metric d on $\mathcal{X}$ with the property that $(\mathcal{X}, d)$ is complete and d generates $\mathcal{O}$. *Warning:* there may be many different metrics on $\mathcal{X}$ that generate the same topology. It may even happen that $\mathcal{X}$ is not complete in some of these metrics, and complete in others (in which case $\mathcal{X}$ is still Polish). Example: $\mathbb{R}$ is separable and complete in the usual metric $d(x,y) = |x - y|$, and therefore $\mathbb{R}$ is a Polish space. But $d'(x,y) := |\arctan(x) - \arctan(y)|$ is another metric that generates the same topology, while $(\mathbb{R}, d')$ is not complete. (Indeed, the completion of $\mathbb{R}$ w.r.t. the metric d' is $[-\infty, \infty]$.)

1.2 Compactness

A subset K of a general topological space $\mathcal{X}$ (with collection of open sets $\mathcal{O}$) is called *compact* if every open cover has a finite subcover, i.e., if for any collection $(O_\gamma)_{\gamma \in \Gamma}$ of open subsets of $\mathcal{X}$ such that $\bigcup_{\gamma \in \Gamma} O_\gamma \supset K$, there exists a finite $\Delta \subset \Gamma$ such that $\bigcup_{\gamma \in \Delta} O_\gamma \supset K$. Using this definition, it is easy to see that the image of a compact set under a continuous function is again compact. Compact subsets of Hausdorff topological spaces are closed. A subset K of a metric space $\mathcal{X}$ is compact if and only if it is closed and *totally bounded*, which means that for every $\varepsilon > 0$ there exists a finite collection $\{B_\varepsilon(x_1), \ldots, B_\varepsilon(x_n)\}$ of open balls such that

$$B_\varepsilon(x_1) \cup \cdots \cup B_\varepsilon(x_n) \supset K.$$

From this, it is not hard to see that compact metrisabe spaces are always separable. If $(x_n)_{n\in\mathbb{N}}$ is a sequence and $m : \mathbb{N} \to \mathbb{N}$ is a function such that $m(n) \to \infty$ as $n \to \infty$, then setting $x'_n := x_{m(n)}$ $(n \in \mathbb{N})$ defines a new sequence. Such a sequence is called a *subsequence* of the original sequence. A *cluster point* of a sequence is a limit of a subsequence.

Theorem 1.1 (Bolzano-Weierstrass) *Let $\mathcal{X}$ be a metrisable space and let $K \subset \mathcal{X}$. Then K is compact if and only if every sequence in K has a subsequence that converges to a limit in K.*

The Bolzano-Weierstrass theorem also holds for second countable spaces. (Note that metrisable spaces need in general not be second countable, and conversely, not every second countable space is metrisable.) There is also a version of the Bolzano-Weierstrass theorem that holds in general topological spaces but in this case one has to replace sequences by the more general nets. A set A is *precompact* if its closure is compact. In metrisable spaces, this is equivalent to the statement that each sequence of points $x_n \in A$ has a convergent subsequence. Note that in this case we do not require that the limit is an element of A. The following simple lemma is often useful.

Lemma 1.2 (Convergence and compactness) *Let $\mathcal{X}$ be a metrisable space and let $x, x_n \in \mathcal{X}$. Then $x_n \to x$ if and only if the following two conditions are satisfied.*

(i) *The set $\{x_n : n \in \mathbb{N}\}$ is precompact.*

(ii) *For every subsequence $x_{n(m)}$ such that $x_{n(m)} \underset{m\to\infty}{\longrightarrow} x'$ for some $x' \in \mathcal{X}$, one has $x' = x$.*

If $(\mathcal{X}, \mathcal{O})$ is a topological space, then a *compactification* of $\mathcal{X}$ is a compact topological space $\overline{\mathcal{X}}$ such that $\mathcal{X}$ is a dense subset of $\overline{\mathcal{X}}$ and the topology on $\mathcal{X}$ is the induced topology from $\overline{\mathcal{X}}$. If $\overline{\mathcal{X}}$ is metrisable, then we say that $\overline{\mathcal{X}}$ is a *metrisable compactification* of $\mathcal{X}$. It turns out that each separable metrisable space $\mathcal{X}$ has a metrisable compactification [Cho69, Theorem 6.3].

A topological space $\mathcal{X}$ is called *locally compact* if for every $x \in \mathcal{X}$ there exists a compact neighbourhood of x. We cite the following proposition from [Eng89, Thms 3.3.8 and 3.3.9].

Proposition 1.3 (Compactification of locally compact spaces) *Let $\mathcal{X}$ be a metrisable topological space. Then the following statements are equivalent.*

(i) *$\mathcal{X}$ is locally compact and separable.*

(ii) *There exists a metrisable compactification $\overline{\mathcal{X}}$ of $\mathcal{X}$ such that $\mathcal{X}$ is an open subset of $\overline{\mathcal{X}}$.*

(iii) *For each metrisable compactification $\overline{\mathcal{X}}$ of $\mathcal{X}$, $\mathcal{X}$ is an open subset of $\overline{\mathcal{X}}$.*

We note that if $\mathcal{X}$ satisfies the equivalent conditions of Proposition 1.3, then it is possible to find a metrisable compactification $\overline{\mathcal{X}}$ of $\mathcal{X}$ such that $\overline{\mathcal{X}}\backslash\mathcal{X}$ consists of just one point, usually denoted by ∞. In this case, $\overline{\mathcal{X}} = \mathcal{X} \cup \{\infty\}$ is called the *one-point compactification* of $\mathcal{X}$. The open sets of $\mathcal{X} \cup \{\infty\}$ are all open sets of $\mathcal{X}$ plus all sets of the form $\{\infty\} \cup O$ where $\mathcal{X}\backslash O$ is a compact subset of $\mathcal{X}$.

A subset $A \subset \mathcal{X}$ of a topological space $\mathcal{X}$ is called a *G_δ-set* if A is a countable intersection of open sets (i.e., there exist $O_i \in \mathcal{O}$ such that $A = \bigcap_{i=1}^\infty O_i$. If $\mathcal{X}$ is metrisable, then every closed set $A \subset \mathcal{X}$ is a G_δ-set, since it is the intersection of the open sets $\{x \in \mathcal{X} : d(x, A) < 1/n\}$. The following result can be found in [Bou58, §6 No. 1, Theorem. 1]. See also [Oxt80, Thms 12.1 and 12.3].

Proposition 1.4 (Compactification of Polish spaces) *Let $\mathcal{X}$ be a metrisable topological space. Then the following statements are equivalent.*

(i) *$\mathcal{X}$ is Polish.*

(ii) *There exists a metrisable compactification $\overline{\mathcal{X}}$ of $\mathcal{X}$ such that $\mathcal{X}$ is a G_δ-subset of $\overline{\mathcal{X}}$.*

(iii) *For each metrisable compactification $\overline{\mathcal{X}}$ of $\mathcal{X}$, $\mathcal{X}$ is a G_δ-subset of $\overline{\mathcal{X}}$.*

Moreover, a subset $\mathcal{Y} \subset \mathcal{X}$ of a Polish space $\mathcal{X}$ is Polish in the induced topology if and only if $\mathcal{Y}$ is a G_δ-subset of $\mathcal{X}$.

We note that if $\overline{\mathcal{X}}$ is a compactification of a Polish space $\mathcal{X}$, equipped with a concrete metric, then $\overline{\mathcal{X}}$ is also the completion of $\mathcal{X}$ in this metric. Thus, unless $\mathcal{X}$ is itself compact, it will never be complete in such a metric (even though, by the definition of a Polish space, there exists metrics generating the same topology with respect to which $\mathcal{X}$ is complete).

1.3 Decomposition of measures

Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and let $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ be a Polish space, equipped with its Borel-σ-field. By definition, a *probability kernel* from $\mathcal{X}$ to $\mathcal{Y}$ is a measurable map

$$\mathcal{X} \ni x \mapsto K(x, \cdot\,) \in \mathcal{M}_1(\mathcal{Y}).$$

Since the Borel-σ-field on $\mathcal{M}_1(\mathcal{Y})$ is generated by the maps $\mu \mapsto \mu(A)$ with $A \in \mathcal{B}(\mathcal{Y})$, the measurability of K is equivalent to the statement that for each $A \in \mathcal{B}(\mathcal{Y})$, the function $K(\,\cdot\,, A)$ is a measurable real function on $\mathcal{X}$. More generally, if $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ is replaced by a general measurable space $(\mathcal{Y}, \mathcal{G})$, then we define a *probability kernel* from $\mathcal{X}$ to $\mathcal{Y}$ to be a map $K : \mathcal{X} \times \mathcal{G} \to [0, 1]$ such that $K(x, \cdot\,)$ is a probability measure on $\mathcal{Y}$ for each $x \in \mathcal{X}$ and $K(\,\cdot\,, A)$ is a measurable real function on $\mathcal{X}$ for each $A \in \mathcal{G}$. With these definitions, one has the following result.

Theorem 1.5 (Decomposition of probability measures) *Let $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$ be a measurable spaces. Let μ be a probability measure on $\mathcal{X}$ and let K be a probability kernel from $\mathcal{X}$ to $\mathcal{Y}$. Then there exists a unique probability measure ν on $\mathcal{X} \times \mathcal{Y}$, equipped with the product-σ-field, so that for any measurable function $f : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$,*

$$\int f(x, y)\nu\big(\mathrm{d}(x, y)\big) = \int \mu(\mathrm{d}x) \int K(x, \mathrm{d}y) f(x, y), \tag{1.2}$$

where on the right-hand side, $x \mapsto \int K(x, \mathrm{d}y) f(x, y)$ is a measurable function on $\mathcal{X}$ that is integrated against μ.

Assume that moreover, $(\mathcal{Y}, \mathcal{G}) = (\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ is a Polish space equipped with its Borel-σ-field. Then conversely for each probability measure ν on $\mathcal{X} \times \mathcal{Y}$, there exist a probability measure μ on $\mathcal{X}$ and probability kernel K from $\mathcal{X}$ to $\mathcal{Y}$ such that (1.2) holds. If (1.2) holds, then μ is the first marginal of ν. Moreover, (1.2) determines the kernel K a.s. uniquely, i.e., if K, K' are probability kernels so that (1.2) holds both for K and K', then there exists a set $N \in \mathcal{F}$ with $\mu(N) = 0$ such that $K(x, \cdot\,) = K'(x, \cdot\,)$ for all $x \in \mathcal{X} \backslash N$.

We note that by subtracting a constant, we see that (1.2) holds more generally for functions f that are bounded from below.

The deep part of Theorem 1.5 is the existence of K given ν;. this may fail in general if the Polish space $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ is replaced by an arbitrary measurable space $(\mathcal{Y}, \mathcal{G})$. Formally, we may define a 'measure' ρ on $\mathcal{X}$ with values in $\mathcal{M}_1(\mathcal{Y})$ by $\rho(A) := \nu(A \times \,\cdot\,)$ $(A \in \mathcal{F})$. Letting μ denote the first marginal of ν, we obviously have $\rho(A) = 0$ whenever $\mu(A) = 0$, i.e., formally $\rho \ll \mu$. Now (1.2) says that

$$\rho(A) = \int_A \mu(\mathrm{d}x) K(x,\,\cdot\,),$$

which we can formally read as saying that ρ has a density with respect to μ, which is given by the function $x \mapsto K(x,\,\cdot\,)$. Thus, Theorem 1.5 amounts to proving something like a Radon-Nikodym theorem for functions and measures with values in $\mathcal{M}_1(\mathcal{Y})$. In fact, if we are just interested in $K(\,\cdot\,,B)$ for one fixed $B \in \mathcal{B}(\mathcal{Y})$, then (1.2) says that

$$\nu(A \times B) = \int_A \mu(\mathrm{d}x) K(x, B) \qquad (A \in \mathcal{F}). \tag{1.3}$$

Since $\nu(\,\cdot\, \times B) \ll \mu$ (where μ is the first marginal of ν), the usual Radon-Nikodym now tells us that for this fixed B, there exists an a.s. unique function $K(\,\cdot\,,B)$ such that (1.3) holds. This argument does not tell us, however, whether for fixed x, we can choose $K(x,\,\cdot\,)$ such that it is a probability measure. Theorem 1.5 tells us that that such a *regular version* of K exists.

Corollary 1.6 (Regular conditional probability) *Let Y be a random variable defined on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in some Polish space $\mathcal{Y}$, and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field. Then there exists a $\mathcal{M}_1(\mathcal{Y})$-valued random variable $\mathbb{P}[Y \in \cdot\,|\,\mathcal{G}]$, which is unique up to a $\mathcal{G}$-measurable set of probability zero, such that*

(i) $\mathbb{P}[Y \in \cdot\,|\,\mathcal{G}]$ *is $\mathcal{G}$-measurable.*

(ii) $\mathbb{E}\big[1_A \mathbb{P}[Y \in B\,|\,\mathcal{G}]\big] = \mathbb{E}[1_A 1_{\{Y \in B\}}]$ *for all $A \in \mathcal{G}$, $B \in \mathcal{B}(\mathcal{Y})$.*

Proof It is not hard to see that there exists a unique probability measure on $\omega \times \mathcal{Y}$, equipped with the σ-field $\mathcal{G} \otimes \mathcal{B}(\mathcal{Y})$, such that

$$\int \nu\big(\mathrm{d}(\omega, y)\big) f(\omega, y) := \int \mathbb{P}(\mathrm{d}\omega)\, f(\omega, Y(\omega))$$

for all $f : \omega \times \mathcal{Y} \to [0, \infty]$ that are measurable w.r.t. $\mathcal{G} \otimes \mathcal{B}(\mathcal{Y})$. Applying Theorem 1.5 to ν, we obtain a $\mathcal{G}$-measurable, $\mathcal{M}_1(\mathcal{Y})$-valued random variable $\mathbb{P}[Y \in \cdot\,|\,\mathcal{G}]$ (i.e., a probability kernel from $(\Omega, \mathcal{G})$ to $\mathcal{Y}$), unique up to a $\mathcal{G}$-measurable set of probability zero, such that

$$\int \mathbb{P}(\mathrm{d}\omega)\, f(\omega, Y(\omega)) = \int \mathbb{P}(\mathrm{d}\omega) \int \mathbb{P}[Y \in \mathrm{d}y\,|\,\mathcal{G}](\omega)\, f(\omega, y)$$

By the uniqueness theorem (applied to ν), to verify this equation, it suffices to check it for functions f of the form $f = 1_{A\times B}$ with $A \in \mathcal{G}$ and $B \in \mathcal{B}(\mathcal{Y})$. Thus, $\mathbb{P}[Y \in \cdot \,|\, \mathcal{G}]$ is a.s. uniquely determined by the requirement that

$$\begin{aligned}\mathbb{E}\big[1_A \mathbb{P}[Y \in B \,|\, \mathcal{G}]\big] &= \int \mathbb{P}(\mathrm{d}\omega)\mathbb{P}[Y \in \mathrm{d}y \,|\, \mathcal{G}](\omega)1_{A\times B}(\omega, y)\\ &= \nu(A \times B) = \mathbb{E}[1_A 1_{\{Y=B\}}].\end{aligned}$$

∎

1.4 Weak convergence

Let $\mathcal{X}$ be a Polish space. We let $\mathcal{B}(\mathcal{X})$ denote *Borel-σ-field* on $\mathcal{X}$, i.e., the σ-field generated by the open sets. We let $\mathcal{C}(\mathcal{X})$ denote the space of all continuous functions $f : \mathcal{X} \to \mathbb{R}$. We let $B_{\mathrm{b}}(\mathcal{X})$ denote the space of all bounded Borel-measurable real functions on $\mathcal{X}$ and we let $\mathcal{C}_{\mathrm{b}}(\mathcal{X}) := \mathcal{C}(\mathcal{X}) \cap B_{\mathrm{b}}(\mathcal{X})$ denote the space of all bounded continuous real functions on $\mathcal{X}$. We equip with $\mathcal{C}_{\mathrm{b}}(\mathcal{X})$ with the *supremumnorm*

$$\|f\|_\infty := \sup_{x\in\mathcal{X}} |f(x)|.$$

With this norm, $\mathcal{C}_{\mathrm{b}}(\mathcal{X})$ is a Banach space [Dud02, Theorem 2.4.9]. We let $\mathcal{M}(\mathcal{X})$ denote the space of all finite measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and write $\mathcal{M}_1(\mathcal{X})$ for the subspace of all probability measures. We cite the following well-known fact from [EK86, Theorems 3.1.7 and 3.3.1].

Proposition 1.7 (Weak convergence) *It is possible to equip $\mathcal{M}_1(\mathcal{X})$ with a metric d_{P} such that*

(i) $(\mathcal{M}_1(\mathcal{X}), d_{\mathrm{P}})$ *is a separable complete metric space,*

(ii) $d_{\mathrm{P}}(\mu_n, \mu) \to 0$ *if and only if* $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$ *for all* $f \in \mathcal{C}_{\mathrm{b}}(\mathcal{X})$.

The precise choice of d_{P} (there are several canonical ways to define such a metric) is not important to us. Since a metrisable topology is uniquely characterized by its convergent sequences, property (ii) uniquely characterizes the topology generated by d_{P} in terms of the topology on $\mathcal{X}$. (In particular, this topology does not depend on the choice of the complete metric d on $\mathcal{X}$ or the precise definition of the metric d_{P}.) We call this topology the *topology of weak convergence* and denote convergence in this topology as

$$\mu_n \Rightarrow \mu.$$

By property (i), the space $\mathcal{M}_1(\mathcal{X})$ equipped with the topology of weak convergence is a Polish space. One also has the following well-known characterization of weak convergence [EK86, Theorem 3.3.1].

Lemma 1.8 (Characterization with open and closed sets) *Let μ_n and μ be probability measures on a Polish space $\mathcal{X}$. Then the following statements are equivalent.*

(i) $\mu_n \Rightarrow \mu$.

(ii) $\limsup_{n\to\infty} \mu_n(C) \leq \mu(C)$ *for all closed* $C \subset \mathcal{X}$.

(iii) $\liminf_{n\to\infty} \mu_n(O) \geq \mu(O)$ *for all open* $O \subset \mathcal{X}$.

Exercise 1.9 (Measures concentrated on a subset) *Let $\mathcal{X}$ be a Polish space and let $\mathcal{X}' \subset \mathcal{X}$ be a G_δ-set, equipped with the induced topology. We naturally identify $\mathcal{M}_1(\mathcal{X}')$ with the subset of $\mathcal{M}_1(\mathcal{X})$ consisting of all $\mu \in \mathcal{M}_1(\mathcal{X})$ such that $\mu(\mathcal{X}') = 1$. Show that the topology on $\mathcal{M}_1(\mathcal{X}')$ coincides with the induced topology from its embedding in $\mathcal{M}_1(\mathcal{X})$. (Hint: Lemma 1.8.) Use this to conclude that $\mathcal{M}_1(\mathcal{X}')$ is a G_δ-subset of $\mathcal{M}_1(\mathcal{X})$. (Hint: Proposition 1.4).*

A very useful characterization of weak convergence in terms of coupling is given by the next theorem [EK86, Cor 3.1.6 and Thm 3.1.8].

Theorem 1.10 (Skorohod representation) *Let μ_n and μ be probability measures on a Polish space $\mathcal{X}$. Then $\mu_n \Rightarrow \mu$ if and only if it is possible to couple random variables X_n, X with laws μ_n, μ, respectively, in such a way that $X_n \to X$ a.s.*

A subset A of a topological space $\mathcal{X}$ is called *relatively compact* if its closure $\overline{A}$ is compact. The next result is known as Prohorov's theorem (see, e.g., [EK86, Theorem 3.2.2] or [Bil99, Theorems 5.1 and 5.2]).

Theorem 1.11 (Prohorov) *Let $\mathcal{X}$ be a Polish space. Let $\mathcal{M}_1(\mathcal{X})$ be equipped with the topology of weak convergence. Then a subset $\mathcal{C} \subset \mathcal{M}_1(\mathcal{X})$ is relatively compact if and only if $\mathcal{C}$ is* tight, *i.e.,*

$$\forall \varepsilon > 0 \ \exists K \subset \mathcal{X} \text{ compact, s.t. } \sup_{\mu \in \mathcal{C}} \mu(\mathcal{X} \backslash K) \leq \varepsilon.$$

1.5 Locally uniform convergence

Let E be a metric space and let $I \subset \mathbb{R}$ be a closed interval. We let $\mathcal{C}_I(E)$ denote the space of all continuous functions $w : [0,\infty) \to \mathbb{R}$.

Lemma 1.12 (Locally uniform convergence) *For $w_n, w \in \mathcal{C}_I(E)$, the following conditions are equivalent:*

(i) $\sup_{t\in C} d\big(w_n(t), w(t)\big) \underset{n\to\infty}{\longrightarrow} 0$ *for all compact* $C \subset I$,

(ii) $w_n(t_n) \underset{n\to\infty}{\longrightarrow} w(t)$ *for all* $t_n, t \in I$ *such that* $t_n \underset{n\to\infty}{\longrightarrow} t$.

Proof Assume (i) and let $t_n, t \in I$ satisfy $t_n \underset{n\to\infty}{\longrightarrow} t$. By Lemma 1.2 (i), there exists a compact set $C \subset I$ such that $t_n \in C$ for all n (and hence also $t \in C$). Then for each $\varepsilon > 0$, there exists an $N < \infty$ such that $d(w_n(t), w(t)) \leq \varepsilon$ for all $n \geq N$. Now

$$d\big(w_n(t_n), w(t)\big) \leq d\big(w_n(t_n), w(t_n)\big) + d\big(w(t_n), w(t)\big) \leq \varepsilon + d\big(w(t_n), w(t)\big)$$

for all $n \geq N$, and hence

$$\limsup_{n\to\infty} d\big(w_n(t_n), w(t)\big) \leq \varepsilon$$

by the continuity of w. Since $\varepsilon > 0$ is arbitrary, this shows that (i) implies (ii). On the other hand, if (i) fails for some compact $C \subset I$, then we can choose $t_n \in C$ and $\varepsilon > 0$ such that

$$d\big(w_n(t_n), w(t_n)\big) \geq \varepsilon \quad \forall n.$$

Since C is compact, by going to a subsequence, we can without loss of generality assume that $t_n \to t$ for some $t \in C$. Since

$$d\big(w_n(t_n), w(t)\big) \geq d\big(w_n(t_n), w(t_n)\big) - d\big(w(t_n), w(t)\big) \geq \varepsilon + d\big(w(t_n), w(t)\big),$$

using the continuity of w, we see that

$$\liminf_{n\to\infty} d\big(w_n(t_n), w(t)\big) \geq \varepsilon,$$

which contradicts (ii). ∎

There exists a metrisable topology on $\mathcal{C}_I(E)$ such that a $w_n \in \mathcal{C}_I(E)$ converges to a limit w if and only if the equivalent conditions of Lemma 1.12 are satisfied. Note that by (1.1) and the remarks below it, these conditions uniquely determine the topology. Note also that by condition (ii) of Lemma 1.12, the topology on $\mathcal{C}_I(E)$ depends only on the topology on E and not on the precise choice of the metric on E. A possible choice of a metric on $\mathcal{C}_I(E)$ is

$$\rho(v, w) := \sum_{n=1}^{\infty} 2^{-n} \sup_{t\in[0,n]} d\big(v(t), w(t)\big),$$

where d is a bounded metric that generates the topology on E. Such a metric can always be found: if d is any metric generating the topology on E, then $d'(x,y) := d(x,y) \wedge 1$ is a bounded metric that generates the same topology. Usually, we do not care about the precise choice of the metric on $\mathcal{C}_I(E)$; apart from ρ, there are many other possible choices. We call this the topology on $\mathcal{C}_I(E)$ the *topology of locally uniform convergence.*

1.6 The Hausdorff metric

Let (E,d) be a metric space, let $\mathcal{K}(E)$ be the space of all compact subsets of E and set $\mathcal{K}_+(E) := \{K \in \mathcal{K}(E) : K \neq \emptyset\}$. Then the *Hausdorff metric* $d_{\rm H}$ on $\mathcal{K}_+(E)$ is defined as

$$\begin{aligned} d_{\rm H}(K_1,K_2) &:= \sup_{x_1\in K_1}\inf_{x_2\in K_2} d(x_1,x_2) \vee \sup_{x_2\in K_2}\inf_{x_1\in K_1} d(x_1,x_2) \\ &= \sup_{x_1\in K_1} d(x_1,K_2) \vee \sup_{x_2\in K_2} d(x_2,K_1), \end{aligned} \tag{1.4}$$

where $d(x,A) := \inf_{y\in A} d(x,y)$ denotes the distance between a point $x \in E$ and a set $A \subset E$. The corresponding topology is naturally called the *Hausdorff topology*. Note the subtle difference between "the Hausdorff topology" (the topology generated by the Hausdorff metric) and "a Hausdorff topology" (any topology satisfying condition (iv) of Section 1.1). We extend this topology to $\mathcal{K}(E)$ by adding $\emptyset$ as an isolated point.

A good source for the Hausdorff topology is [SSS14, Appendix B], where one can find the proofs of all the lemmas in this section. The first lemma shows that the Hausdorff topology depends only on the topology on E, and not on the choice of the metric.

Lemma 1.13 (Convergence criterion) *Let $K_n, K \in \mathcal{K}_+(E)$ $(n \geq 1)$. Then $K_n \to K$ in the Hausdorff topology if and only if there exists a $C \in \mathcal{K}_+(E)$ such that $K_n \subset C$ for all $n \geq 1$ and*

$$\begin{aligned} K &= \{x \in E : \exists x_n \in K_n \text{ s.t. } x_n \to x\} \\ &= \{x \in E : \exists x_n \in K_n \text{ s.t. } x \text{ is a cluster point of } (x_n)_{n\in\mathbb{N}}\}. \end{aligned} \tag{1.5}$$

The following lemma shows that $\mathcal{K}(E)$ is Polish if E is.

Lemma 1.14 (Properties of the Hausdorff metric)

(a) *If (E,d) is separable, then so is $(\mathcal{K}_+(E), d_{\rm H})$.*

(b) *If (E,d) is complete, then so is $(\mathcal{K}_+(E), d_{\rm H})$.*

The following lemma shows in particular that $\mathcal{K}(E)$ is compact if E is compact.

Lemma 1.15 (Compactness in the Hausdorff topology) *A set $\mathcal{A} \subset \mathcal{K}(E)$ is precompact if and only if there exists a $C \in \mathcal{K}(E)$ such that $K \subset C$ for each $K \in \mathcal{A}$.*

The following lemma is useful when proving convergence of $\mathcal{K}(E)$-valued random variables.

Lemma 1.16 (Tightness criterion) *Assume that E is a Polish space and let K_n $(n \geq 1)$ be $\mathcal{K}(E)$-valued random variables. Then the collection of laws $\{\mathbb{P}[K_n \in \cdot\,] : n \geq 1\}$ is tight if and only if for each $\varepsilon > 0$ there exists a compact $C \subset E$ such that $\mathbb{P}[K_n \subset C] \geq 1 - \varepsilon$ for all $n \geq 1$.*

1.7 Squeezed space

Let (E, d) be a metric space, let $\{*\}$ be a set containing a single element called $*$, and let

$$\mathcal{R}(E) := \big(E \times \mathbb{R}\big) \cup \big\{(*, -\infty), (*, +\infty)\big\}. \tag{1.6}$$

We extend d to $E \cup \{*\}$ by setting $d(x, *) = d(*, x) := \infty$ if $x \neq *$ and $:= 0$ otherwise. Let $\overline{\mathbb{R}} := [-\infty, \infty]$ denote the usual two-point compactification of the real line. We fix a continuous function $\phi : \overline{\mathbb{R}} \to [0, \infty)$ such that $\phi(t) > 0$ for all $t \in \mathbb{R}$ and $\phi(\pm\infty) = 0$, we choose a metric $d_{\overline{\mathbb{R}}}$ that generates the topology on $\overline{\mathbb{R}}$, and we define $\rho : \mathcal{R}(E)^2 \to [0, \infty)$ by

$$\rho\big((x, s), (y, t)\big) := \big(\phi(s) \wedge \phi(t)\big)\big(d(x, y) \wedge 1\big) + \big|\phi(s) - \phi(t)\big| + d_{\overline{\mathbb{R}}}(s, t) \tag{1.7}$$

Lemma 1.17 (Metric on squeezed space) *The function ρ is a metric on $\mathcal{R}(E)$.*

Proof For brevity, we write $d'(x, y) := d(x, y) \wedge 1$. Then d' is a metric on E. The only nontrivial statement that we have to prove is the triangle inequality, and it suffices to prove this for the function

$$\rho'\big((x, s), (y, t)\big) := \big(\phi(s) \wedge \phi(t)\big) d'(x, y) + \big|\phi(s) - \phi(t)\big|.$$

We estimate

$$\rho'\big((x, s), (z, u)\big) \leq \big(\phi(s) \wedge \phi(u)\big)\big(d'(x, y) + d'(y, z)\big) + \big|\phi(s) - \phi(u)\big|. \tag{1.8}$$

If $\phi(t) \geq \phi(s) \wedge \phi(u)$, then $\phi(s) \wedge \phi(u)$ is less than $\phi(s) \wedge \phi(t)$ and also less than $\phi(t) \wedge \phi(u)$, so we can simply estimate the expression in (1.8) from above by

$$\big(\phi(s) \wedge \phi(t)\big)d'(x,y) + \big(\phi(t) \wedge \phi(u)\big)d'(y,z)\big) + \big|\phi(s) - \phi(t)\big| + \big|\phi(t) - \phi(u)\big|$$

and we are done. On the other hand, if $\phi(t) < \phi(s) \wedge \phi(u)$, then

$$\big|\phi(s) - \phi(t)\big| + \big|\phi(t) - \phi(u)\big| = \big|\phi(s) - \phi(u)\big| + 2\big(\phi(s) \wedge \phi(u) - \phi(t)\big).$$

Using the fact that $d' \leq 1$, we can now estimate the right-hand side of (1.8) from above by

$$\begin{aligned}
&\phi(t)\big(d'(x,y) + d'(y,z)\big) + 2\big(\phi(s) \wedge \phi(u) - \phi(t)\big) + \big|\phi(s) - \phi(u)\big| \\
&\quad = \big(\phi(s) \wedge \phi(t)\big)d'(x,y) + \big(\phi(t) \wedge \phi(u)\big)d'(y,z) \\
&\qquad + \big|\phi(s) - \phi(t)\big| + \big|\phi(t) - \phi(u)\big|,
\end{aligned}$$

and again we are done. ∎

The following lemma shows that the topology generated by the metric ρ depends only on the topology on E and not on the choice of the metric on E. Recall that by (1.1), a metrisable topology is uniquely characterised by its convergent sequences, so the topology on $\mathcal{R}(E)$ is uniquely characterised by the conditions (i) and (ii) below.

Lemma 1.18 (Topology on squeezed space) *A sequence $(x_n, t_n) \in \mathcal{R}(E)$ converges to a limit (x, t) in the metric ρ defined in (1.7) if and only if the following two conditions are satisfied:*

(i) *$t_n \to t$ in the topology on $\overline{\mathbb{R}}$,*

(ii) *if $t \in \mathbb{R}$, then $x_n \to x$ in the topology on E.*

Proof This is immediate from the definition of ρ. ∎

We can think of the space $\mathcal{R}(E)$ as being obtained from $E \times \overline{\mathbb{R}}$ by squeezing the sets $E \times \{\pm\}$ into the single points $(*, \pm)$. For this reason, we call $\mathcal{R}(E)$ the *squeezed space*. In the special case that $E = \overline{\mathbb{R}}$, we can make a picture of $\mathcal{R}(\overline{\mathbb{R}})$ by mapping $\overline{\mathbb{R}} \times \mathbb{R}$ into the closed unit disc using the function

$$(x,t) \mapsto \big(\sqrt{1 - \psi(t)^2}\,\psi(x), \psi(t)\big) \quad \text{with} \quad \psi(z) := \frac{z}{1 + |z|}$$

(with $\psi(\pm\infty) := \pm 1$), and mapping the points $(*, \pm\infty)$ to $(0, \pm 1)$. The following lemma shows that $\mathcal{R}(E)$ is a Polish space if E is Polish.

Lemma 1.19 (Properties of squeezed space)

(a) *If* (E, d) *is separable, then so is* $(\mathcal{R}(E), \rho)$.

(b) *If* (E, d) *is complete, then so is* $(\mathcal{R}(E), \rho)$.

Proof If D is a countable dense subset of (E, d), then $D \times \mathbb{Q}$ is a countable dense subset of $(\mathcal{R}(E), \rho)$, proving (a).

To prove (b), let (x_n, t_n) be a Cauchy sequence in $(\mathcal{R}(E), \rho)$. Then by (1.7) t_n is a Cauchy sequence in $\overline{\mathbb{R}}$ and hence $t_n \to t$ for some $t \in \overline{\mathbb{R}}$. If $t \in \mathbb{R}$, then by (1.7) x_n is a Cauchy sequence in (E, d) so by the completeness of the latter, $x_n \to x$ for some $x \in E$. By Lemma 1.18, it follows that (x_n, t_n) converges, proving the completeness of $(\mathcal{R}(E), \rho)$. ∎

The following lemma identifies the compact subsets of $\mathcal{R}(E)$. In particular, the lemma shows that $\mathcal{R}(E)$ is compact if E is compact.

Lemma 1.20 (Compactness criterion) *A set* $A \subset \mathcal{R}(E)$ *is precompact if and only if for each* $T < \infty$, *there exists a compact set* $K \subset E$ *such that* $\{x \in E : (x, t) \in A,\ t \in [-T, T]\} \subset K$.

Proof Assume that $A \subset \mathcal{R}(E)$ has the property that for each $T < \infty$, there exists a compact set $K \subset E$ such that $\{x \in E : (x, t) \in A,\ t \in [-T, T]\} \subset K$. To show that A is precompact, we will show that each sequence $(x_n, t_n) \in A$ has a convergent subsequence. By the compactness of $\overline{\mathbb{R}}$, we can select a subsequence (x'_n, t'_n) such that $t'_n \to t$ for some $t \in \overline{\mathbb{R}}$. If $t = \pm\infty$, then by Lemma 1.18 $(x'_n, t'_n) \to (*, \pm\infty)$ and we are done. Otherwise, there exists a $T < \infty$ such that $t'_n \in [-T, T]$ for all n large enough. By assumption, there then exists a compact set $K \subset E$ such that $x'_n \in K$ for all n large enough, so we can select a further subsequence such that (x''_n, t''_n) converges to a limit $(x, t) \in E \times \mathbb{R}$.

Assume, on the other hand, that $A \subset \mathcal{R}(E)$ has the property that for some $T < \infty$, there does not exist a compact set $K \subset E$ such that $\{x \in E : (x, t) \in A,\ t \in [-T, T]\} \subset K$. Set

$$B := \big\{x \in E : (x, t) \in A \text{ for some } t \in [-T, T]\big\}$$

The closure of B cannot be compact, since this would contradict our assumption. It follows that there exists a sequence $x_n \in B$ that does not contain a convergent subsequence, and there exist $t_n \in [-T, T]$ such that $(x_n.t_n) \in A$. But then, in view of Lemma 1.18, the sequence (x_n, t_n) cannot contain a convergent subsequence either, proving that A is not precompact. ∎

1.8 Path space

Let E be a metrisable space and let $\mathcal{R}(E)$ be the squeezed space defined in Section 1.7. By definition, a *path* in E is a nonempty compact subset $\pi \subset \mathcal{R}(E)$ such that $\{x \in E : (x,t) \in \pi\}$ has at most one element for each given $t \in \overline{\mathbb{R}}$. The set $\overline{I}_\pi := \big\{t \in \overline{\mathbb{R}} : \exists x \in E \text{ s.t. } (x,t) \in \pi\big\}$ is called the *domain* of π and

$$\sigma_\pi := \inf \overline{I}_\pi \quad \text{and} \quad \tau_\pi := \sup \overline{I}_\pi \tag{1.9}$$

the *starting time* and *final time* of the path π. For each $t \in \overline{I}_\pi$, we let $\{\pi(t)\} := \{x \in E : (x,t) \in \pi\}$ denote the unique point $\pi(t) \in E$ such that $(\pi(t), t) \in \pi$. Then $t \mapsto \pi(t)$ is a function from $\overline{I}_\pi$ to E. We let $\Pi(E)$ denote the set of all paths in E and set $I_\pi := \overline{I}_\pi \cap \mathbb{R}$.

Lemma 1.21 (Path viewed as a function) *The domain I_π of a path $\pi \in \Pi(E)$ is a closed subset of $\overline{\mathbb{R}}$, and $t \mapsto \pi(t)$ is a continuous function from I_π to E. Conversely, if $I \subset \mathbb{R}$ is closed and $t \mapsto f(t)$ is a continuous function from I to E, then there exists a path $\pi \in \Pi(E)$ such that $I_\pi = I$ and $\pi(t) = f(t)$ $(t \in I)$.*

Proof We first show that for each $\pi \in \Pi(E)$, the function $I_\pi \ni t \mapsto \pi(t)$ is continuous. Assume that $t_n, t \in I_\pi$ and $t_n \to t$. Since π is compact, the sequence $(\pi(t_n), t_n)$ is precompact. Since $\pi(t)$ is the only element of $\{x \in E : (x,t) \in \pi\}$, each subsequence of the $(\pi(t_n), t_n)$ must converge to $(\pi(t), t)$. By Lemma 1.2, we conclude that $(\pi(t_n), t_n) \to (\pi(t), t)$. Since $t \in \mathbb{R}$, by Lemma 1.18, we conclude that $\pi(t_n) \to \pi(t)$, which shows that $I_\pi \ni t \mapsto \pi(t)$ is continuous on I as claimed.

Let $I \subset \mathbb{R}$ be closed and let $f : I \to E$ be continuous. If I is nonempty, then let $\overline{I}$ be the closure of I in $\overline{\mathbb{R}}$, and set $\overline{I} := \{\infty\}$ otherwise. Extend f to $\overline{I}$ by setting $f(t) := *$ if $t = \pm\infty$. Let $\pi := \big\{(f(t), t) : t \in \overline{I}\big\}$. It follows from Lemma 1.18 and the continuity of f that the map

$$\overline{I} \ni t \mapsto \big(f(t), t\big) \in \mathcal{R}(E) \tag{1.10}$$

is continuous. Since $\overline{I}$ is compact and since π is the image of $\overline{I}$ under the continuous map (1.10), we conclude that π is compact. Clearly, $\{x \in E : (x,t) \in \pi\}$ has precisely one element for $t \in \overline{I}$, and is empty for $t \notin \overline{I}$. This shows that $\pi \in \Pi(E)$. ∎

In view of Lemma 1.21, we often view a path $\pi \in \Pi(E)$ as a continuous function defined on a closed domain $I_\pi \subset \mathbb{R}$. The correspondence between paths and continuous functions is almost one-to-one. The only ambiguity arises when $-\infty$ and/or $+\infty$ are not elements of the closure of I_π, and we have the choice whether to include them in $\overline{I}_\pi$ or not. If I_π is nonempty, then it is natural to include $\pm\infty$ only when they are elements of the closure of I_π. With this convention, if $I \subset \mathbb{R}$ is a closed nonempty interval, then we identify the space $\mathcal{C}_I(E)$ defined in Section 1.5 with the set $\{\pi \in \Pi(E) : \overline{I}_\pi = \overline{I}\}$, where $\overline{I}$ denotes the closure of I in $\overline{\mathbb{R}}$.

Let $\mathcal{K}(\mathcal{R}(E))$ be the set of compact subsets of the squeezes space $\mathcal{R}(E)$. We equip $\mathcal{K}(\mathcal{R}(E))$ with the Hausdorff topology. We observe that $\Pi(E)$ is a subset of $\mathcal{K}(\mathcal{R}(E))$. We naturally equip $\Pi(E)$ with the induced topology from its embedding in $\mathcal{K}(\mathcal{R}(E))$.

Lemma 1.22 (Paths with a fixed domain) *Let $I \subset \mathbb{R}$ be a closed nonempty interval. The induced topology on $\mathcal{C}_I(E)$ from its embedding in $\Pi(E)$ is the topology of locally uniform convergence.*

Proof Assume that $\pi_n, \pi \in \mathcal{C}_I(E)$, viewed as functions, satisfy $\pi_n \to \pi$ locally uniformly. We need to show that viewed as compact subsets of $\mathcal{R}(E)$, the sets π_n, π satisfy $\pi_n \to \pi$ in the Hausdorff topology on $\mathcal{K}(\mathcal{R}(E))$. Let $\overline{I}$ denote the closure of I in $\overline{\mathbb{R}}$. By Lemma 1.13, we need to show that $\bigcup_n \pi_n$ is precompact and

$$\begin{aligned}
&\pi \subset \big\{(x,t) \in \mathcal{R}(E) : \exists t_n \in \overline{I} \text{ s.t. } \big(\pi_n(t_n), t_n\big) \to (x,t)\big\},\\
&\big\{(x,t) \in \mathcal{R}(E) : (x,t) \text{ is a cluster}\\
&\qquad\qquad \text{point of } \big(\pi_n(t_n), t_n\big) \text{ for some } t_n \in \overline{I}\big\} \subset \pi.
\end{aligned} \tag{1.11}$$

To see that $\bigcup_n \pi_n$ is precompact, we need to show that each sequence of the form $(\pi_{n(m)}(t_m), t_m)_{m\geq 1}$ has a convergent subsequence. If $n(m)$ infinitely often takes the same value n, then the claim is obvious from the compactness of π_n, so without loss of generality we may assume that $n(m) \to \infty$. Going to a subsequence if necessary, we may assume that $t_m \to t$ for some $t \in \overline{I}$. If $t = \pm\infty$, then the claim is again obvious so we may assume that $t \in I$. In this case Lemma 1.12 (ii) tells us that $\pi_{n(m)}(t_m) \to \pi(t)$ so we have found a convergent subsequence as required.

To prove the first inclusion in (1.11), let $(\pi(t), t) \in \pi$ and set $t_n := t$ for all n. If $t \in I$, then $\pi_n(t) \to \pi(t)$ since locally uniform convergence implies pointwise convergence, and if $t = \pm\infty$ then trivially $(*, t_n) \to (*, t)$. To prove the second inclusion, assume that $(\pi_{n(m)}(t_{n(m)}), t_{n(m)}) \to (x,t)$ as $m \to \infty$ for some $(x,t) \in \mathcal{R}(E)$, $t_n \in \overline{I}$, and $n(m) \to \infty$. If $t \in I$, then we can use Lemma 1.12 (ii) which tells us that $\pi_{n(m)}(t_{n(m)}) \to \pi(t)$ and hence $(x,t) = (\pi(t), t) \in \pi$. If $t = \pm\infty$, then trivially $x = *$ and $(*, t) \in \pi$.

Assume, conversely, that $\pi_n \to \pi$ in the Hausdorff topology on $\mathcal{K}(\mathcal{R}(E))$. We need to show that $\pi_n, \pi \in \mathcal{C}_I(E)$ and that $\pi_n \to \pi$ locally uniformly. Assume that $t_n, t \in I$ such that $t_n \to t$. By Lemma 1.12 (ii), it suffices to show that $\pi_n(t_n) \to \pi(t)$ for all such t_n, t. Equivalently, we may show that $(\pi_n(t_n), t_n) \to (\pi(t), t)$. By Lemma 1.2, it suffices to show that the set $\{(\pi_n(t_n), t_n) : n \in \mathbb{N}\}$ is precompact and $(\pi(t), t)$ is the only cluster point of the sequence $(\pi_n(t_n), t_n)$. By Lemma 1.13, there exists a compact set $C \subset \mathcal{R}(E)$ such that $\pi_n \subset C$ for all n, so $\{(\pi_n(t_n), t_n) : n \in \mathbb{N}\}$ is precompact as required. Let (x, t) be any cluster point. By Lemma 1.13 (ii), $(x, t) \in \pi$ and hence $x = \pi(t)$, which shows that $\pi_n(t_n) \to \pi(t)$ as required. ∎

Let $\pi \in \Pi(\mathbb{R}^d)$. Assume that $\overline{I}_\pi$ is the closure of I_π in $\overline{\mathbb{R}}$. Recall that σ_π and τ_π denote the starting time and final time of π. For each $t \in [\sigma_\pi, \tau_\pi] \cap \mathbb{R}$, let us write

$$\lfloor t \rfloor := \sup\{s \in I_\pi : s \leq t\} \quad \text{and} \quad \lceil t \rceil := \inf\{u \in I_\pi : t \leq u\}.$$

We define a *linearly interpolated* path $\hat{\pi}$ with domain $\overline{I}_{\hat{\pi}} := [\sigma_\pi, \tau_\pi]$ by $\hat{\pi}(t) := \pi(t)$ $(t \in \overline{I}_\pi)$ and

$$\hat{\pi}(t) := \frac{\lceil t \rceil - t}{\lceil t \rceil - \lfloor t \rfloor} \pi\big(\lfloor t \rfloor\big) + \frac{t - \lfloor t \rfloor}{\lceil t \rceil - \lfloor t \rfloor} \pi\big(\lceil t \rceil\big) \qquad (t \in [\sigma_\pi, \tau_\pi] \backslash \overline{I}_\pi).$$

It often happens that a sequence of functions $f_n : N \to \mathbb{R}^d$ converges, after a rescaling of time, to a continuous limit $f : [0, \infty) \to \mathbb{R}^d$. To formulate this properly, it is a common habit to linearly interpolate the functions f_n so that all functions are elements of the space $\mathcal{C}_{[0,\infty)}(\mathbb{R}^d)$. As the following exercise shows, when one uses the path space $\Pi(\mathbb{R}^d)$, no interpolation is necessary to formulate the result.

Exercise 1.23 (Convergence of interpolated paths) *Let $I \subset \mathbb{R}$ be a nonempty closed interval. Assume that $\pi \in \mathcal{C}_I(\mathbb{R}^d)$ and $\pi_n \in \Pi(\mathbb{R}^d)$. Show that $\pi_n \to \pi$ in the topology on $\Pi(\mathbb{R}^d)$ if and only if $\hat{\pi}_n \to \hat{\pi}$.*

Sometimes, when formulating convergence of a sequence of functions f_n to a limit f, one extrapolates with the aim of ensuring that all functions are defined on the same space. Let E be a metrisable space and for each $\pi \in \Pi(E)$, let π^+ denote the path with domain $\overline{I}_{\pi^+} := \overline{I}_\pi \cup [\tau_\pi, \infty]$ defined as $\pi^+(t) := \pi(t)$ if $t \in \overline{I}_\pi$ and

$$\pi^+(t) := \pi(\tau_\pi) \quad (\tau_\pi < t < \infty) \quad \text{and} \quad \pi^+(\infty) := *.$$

The next exercise shows that when one uses the path space $\Pi(\mathbb{R}^d)$, no extrapolation is necessary.

Exercise 1.24 (Convergence of extrapolated paths) *Let $\pi_n, \pi \in \Pi(E)$. Show that the following conditions are equivalent:*

(i) $\pi_n \to \pi$

(ii) $\pi_n^+ \to \pi^+$ *and* $\tau_{\pi_n} \to \tau_\pi$.

Our next proposition says that the space of paths in E is Polish provided E has this property.

Proposition 1.25 (Polish space) *If E is a Polish space, then so is $\Pi(E)$.*

The proof of Proposition 1.25 needs some preparations. Let d be a metric generating the topology on E and let $\pi \in \Pi(E)$. For each $\pi \in \Pi(E)$, $\delta > 0$ and $T < \infty$, we define

$$m_{T,\delta}(\pi) := \sup\big\{d\big(\pi(s),\pi(t)\big) : s,t \in I_\pi,\ -T \leq s \leq t \leq T,\ t-s \leq \delta\big\}. \tag{1.12}$$

The quantity $m_{T,\delta}(\pi)$ is called the *modulus of continuity* of the path π. More generally, for any compact subset $K \subset \mathcal{R}(E)$, we can define

$$m_{T,\delta}(K) := \sup\big\{d\big(x,y\big) : (x,s),(y,t) \in K,\ -T \leq s \leq t \leq T,\ t-s \leq \delta\big\},$$

which coincides with our previous definition if π is a path.

Lemma 1.26 (Characterisation of paths) *A compact subset $\pi \subset \mathcal{R}(E)$ is an element of the path space $\Pi(E)$ if and only if $\lim_{\delta\to 0} m_{T,\delta}(\pi) = 0$ for all $T < \infty$.*

Proof Assume that $\pi \in \mathcal{K}(\mathcal{R}(E))$ and $\limsup_{\delta\to 0} m_{T,\delta}(\pi) > 0$ for some $T < \infty$. Then we can find $(x_n,s_n),(y_n,t_n) \in \pi$ and $\delta > 0$ with $d(x_n,y_n) \geq \delta$, $-T \leq s_n \leq t_n \leq T$, and $t_n - s_n \leq 1/n$. Since π is compact, by going to a subsequence, we can assume that $(x_n,s_n) \to (x,s)$ and $(y_n,t_n) \to (y,t)$ for some $(x,s),(y,t) \in G$ with $d(x,y) \geq \delta > 0$, $-T \leq s \leq t \leq T$, and $t-s = 0$. This shows that $\pi \notin \Pi(E)$.

Conversely, if $\pi \notin \Pi(E)$, then there exist $(x,t),(y,t) \in \pi$ with $x \neq y$. Since $(*,\pm\infty)$ are the only points in $\mathcal{R}(E)$ with time coordinate $\pm\infty$ we must have $t \in \mathbb{R}$. But then $m_{\delta,T}(\pi) \geq d(x,y) > 0$ for all $T \geq |t|$, which shows that $\limsup_{\delta\to 0} m_{T,\delta}(\pi) > 0$ for some $T < \infty$. ∎

Proof of Proposition 1.25 If E is a Polish space, then by Lemma 1.19 so is $\mathcal{R}(E)$ and hence by Lemma 1.14 so is $\mathcal{K}(\mathcal{R}(E))$. For each $\varepsilon,\delta > 0$ and $T < \infty$, the set

$$A_{T,\varepsilon,\delta} := \{K \in \mathcal{K}(\mathcal{R}(E)) : m_{T,\delta}(K) \geq \varepsilon\}$$

is a closed subset of $\mathcal{K}(\mathcal{R}(E))$ and hence its complement $A^{\mathrm{c}}_{T,\varepsilon,\delta}$ is open. By Lemma 1.26,

$$\Pi(E) = \bigcap_{n,m}\bigcup_k A^{\mathrm{c}}_{n,1/m,1/k},$$

which is a countable intersection of open sets, i.e., a G_δ-set. ∎

A set $\mathcal{A} \subset \Pi(E)$ is called *equicontinuous* if

$$\lim_{\delta \to 0} \sup_{\pi \in \mathcal{A}} m_{T,\delta}(\pi) = 0 \qquad (T < \infty).$$

The following theorem identifies the compact subsets of $\Pi(E)$. Condition (ii) is called the *compact containment* condition. If $I \subset \mathbb{R}$ is a closed nonempty interval, then $\mathcal{C}_I(E)$ is a closed subset of Π and hence the following theorem can also be used to identify the precompact subsets of $\mathcal{C}_I(E)$. In this context, the result is known as the *Arzela-Ascoli theorem*. Note that while the definition of equicontinuity depends (at least a priori) on the choice of the metric d on E, whether a set $\mathcal{A} \subset \Pi(E)$ is precompact only depends on the topology on E, so when verufying conditions (i) and (ii) below, we are free to choose any metric d that generates the topology on E.

Theorem 1.27 (Arzela-Ascoli) *A set $\mathcal{A} \subset \Pi(E)$ is precompact if and only if*

(i) *$\mathcal{A}$ is equicontinuous,*

(ii) *for each $T < \infty$, there exists a compact set $K \subset E$ such that $\big\{\pi(t) : \pi \in \mathcal{A},\ t \in I_\pi \cap [-T,T]\big\} \subset K$.*

Proof Let $\overline{\mathcal{A}}$ denote the closure of $\mathcal{A}$, viewed as a subset of the space $\mathcal{K}(\mathcal{R}(E))$, equipped with the Hausdorff topology. Then $\mathcal{A}$ is a precompact subset of $\Pi(E)$ if and only if $\overline{\mathcal{A}}$ is a compact subset of $\mathcal{K}(\mathcal{R}(E))$ and $\overline{\mathcal{A}} \subset \Pi(E)$. By Lemma 1.20, $\overline{\mathcal{A}}$ is a compact subset of $\mathcal{K}(\mathcal{R}(E))$ if and only if condition (ii) holds. To complete the proof, it suffices to show that assuming that (ii) holds, one has $\overline{\mathcal{A}} \subset \Pi(E)$ if and only if (i) holds.

We first show that (i) implies $\overline{\mathcal{A}} \subset \Pi(E)$. Assume that $\pi_n \in \mathcal{A}$ converge in the Hausdorff topology to a compact subset $\pi \subset \mathcal{R}(E)$. To show that $\pi \in \Pi(E)$, will apply Lemma 1.26. If $(x,s),(y,t) \in \pi$, then by Lemma 1.13, there exist $(x_n,s_n),(y_n,t_n) \in \pi_n$ such that $(x_n,s_n) \to (x,s)$ and $(y_n,t_n) \to (y,t)$. If $s,t \in [-T,T]$ and $|t-s| \leq \delta$, then for n large enough we have $s_n,t_n \in [-T-1,T+1]$ and $|t_n - s_n| \leq 2\delta$. Since $d(x_n,y_n) \to d(x,y)$, it follows that

$$\limsup_{\delta \to 0} m_{T,\delta}(\pi) \leq \limsup_{\delta \to 0} \sup_n m_{T+1,2\delta}(\pi_n) = 0 \qquad (\delta > 0,\ T < \infty),$$

which by Lemma 1.26 implies that $\pi \in \Pi(E)$.

Assume now that (ii) holds but (i) fails. Then there exist $T < \infty$ and $\varepsilon > 0$ such that for each $n \geq 1$, we can find $\pi_n \in \mathcal{A}$ with $m_{T,1/n}(\pi_n) \geq \varepsilon$. This means that there exist $-T \leq s_n \leq t_n \leq T$ such that $d(\pi_n(s_n), \pi_n(t_n)) \geq \varepsilon$ and $t_n - s_n \leq 1/n$. By (ii), $\overline{\mathcal{A}}$ is a compact subset of $\mathcal{K}(\mathcal{R}(E))$, so by going a subsequence we may assume that $\pi_n \to \pi \in \mathcal{K}(\mathcal{R}(E))$. By going to a further subsequence, we may assume that $s_n \to s$ and $t_n \to t$ for some $s, t \in [-T, T]$. But then $s = t$ since $t_n - s_n \leq 1/n$. Let $x_n := \pi_n(s_n)$ and $y_n := \pi_n(t_n)$. By (ii), we can select a further subsequence such that $x_n \to x$ and $y_n \to y$ for some x, y with $d(x, y) \geq \varepsilon$. By Lemma 1.13, we have $(x, t), (y, t) \in \pi$ which shows that $\pi \notin \Pi(E)$ and hence $\overline{\mathcal{A}}$ is not a subset of $\Pi(E)$. ∎

For real-valued paths, the compact containment condition of Theorem 1.27 can be relaxed.

Theorem 1.28 (Arzela-Ascoli - real version) *Assume that* $\pi_n \in \Pi(\mathbb{R})$ *satisfy:*

(i) $\{\pi_n : n \in \mathbb{N}\}$ *is equicontinuous,*

(ii) *there exist* $t_n \in I_{\pi_n}$ *such that* $\sup_n |t_n| < \infty$ *and a compact set* $K \subset \mathbb{R}$ *such that* $\pi_n(t_n) \in K$ *for all* n.

Then $\{\pi_n : n \in \mathbb{N}\}$ *is a precompact subset of* $\Pi(\mathbb{R})$.

Proof For any set $A \subset \mathbb{R}$ and $r \geq 0$, we write $A^r := \{x \in \mathbb{R} : \inf_{y \in A} |x - y| \leq r\}$. Then A^r is a closed subset of $\mathbb{R}$. If A is compact, then so is A^r.

To prove the claim of the theorem, it suffices to check condition (ii) of Theorem 1.27. It suffices to check this for T sufficiently large, so without loss of generality, we can assume that $t_n \in [-T, T]$ for all n. Fix $\varepsilon > 0$. By equicontinuity, we can choose $\delta > 0$ such that $|\pi_n(s) - \pi_n(t)| \leq \varepsilon$ for all n and $s, t \in I_{\pi_n} \cap [-T, T]$. Let K be the compact set from condition (ii) above. Then $\pi_n(t) \in K^\varepsilon$ for all $t \in I_{\pi_n}$ such that $|t - t_n| \leq \delta$, and by induction, for each $k \geq 1$, we obtain that $\pi_n(t) \in K^{k\varepsilon}$ for all $t \in I_{\pi_n}$ such that $|t - t_n| \leq k\delta$. Choosing k large enough such that $\delta k \geq 2T$, we see that $\{\pi_n(t) : n \in \mathbb{N},\ t \in [-T, T] \cap I_{\pi_n}\} \subset K^{k\delta}$, so condition (ii) of Theorem 1.27 is satisfied. ∎

Chapter 2

The excursion

2.1 Scaling limit of simple random walk

Let $(X_k)_{k\geq 1}$ be i.i.d. and uniformly distributed on $\{-1,+1\}$, and let

$$S_n := \sum_{k=1}^{n} X_k \qquad (n \geq 0),$$

with naturally $S_0 := 0$. Then $(S_n)_{n\geq 0}$ is a one-dimensional nearest-neighbour random walk. It will be convenient to interpolate linearly. We set

$$S_t := \big(\lceil t\rceil - t\big)S_{\lfloor t\rfloor} + \big(t - \lfloor t\rfloor\big)S_{\lceil t\rceil} \qquad (t \geq 0).$$

Then $S = (S_t)_{t\geq 0}$ is a random variable taking in the space

$$\mathcal{C}_0 := \big\{f \in \mathcal{C}_{[0,\infty)}(\mathbb{R}) : f_0 = 0\big\}. \tag{2.1}$$

Donsker's invariance principle says that S, diffusively rescaled, converges to Brownian motion. To formulate this properly, for $\lambda > 0$, let $\theta_\lambda : \mathbb{R}^2 \to \mathbb{R}^2$ be the *diffusive scaling* map defined as

$$\theta_\lambda(x,t) := (\lambda x, \lambda^2 t) \qquad \big((x,t) \in \mathbb{R}^2\big). \tag{2.2}$$

which we extend to a (clearly unique) continuous map $\theta_\lambda : \mathcal{R}(\overline{\mathbb{R}}) \to \mathcal{R}(\overline{\mathbb{R}})$. For any subset $A \subset \mathcal{R}(\overline{\mathbb{R}})$, we let $\theta_\lambda A$ denote the image of A under θ_λ. In particular, we can apply this to S, which we can view as an element of the path space $\Pi(\mathbb{R})$ and hence as a compact subset of $\mathcal{R}(\mathbb{R})$. For each $\varepsilon > 0$, the *diffusively rescaled* path

$$S^\varepsilon := \theta_\varepsilon S \tag{2.3}$$

is then the random variable taking values in the space $\mathcal{C}_0$ defined as

$$S^\varepsilon_t := \varepsilon S_{\varepsilon^{-2}t} \qquad (t \in \varepsilon\mathbb{N}).$$

The following fact is well-known. Below, we naturally identify the path $(B_t)_{t\geq 0}$ of a Brownian motion with an element of the path space $\Pi(\mathbb{R})$.

Theorem 2.1 (Donsker's invariance principle) *One has*

$$\mathbb{P}\big[(S^\varepsilon_t)_{t\geq 0} \in \cdot\,\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[(B_t)_{t\geq 0} \in \cdot\,\big], \tag{2.4}$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion and $\Rightarrow$ denotes weak convergence of probability measures on $\mathcal{C}_0$, equipped with the topology of locally uniform convergence.

As we have seen in Exercise 1.23, to formulate Theorem 2.1, it was in fact not necessary to interpolate linearly. Instead, we can also view S as an element of the path space $\Pi(\mathbb{R})$ with domain $I_S = \mathbb{N}$ and then formulate Theorem 2.1, as weak convergence in law of random variables with values in $\Pi(\mathbb{R})$. However, in what follows, the linear interpolation will turn out to be convenient for other purposes.

Note that combining Donsker's invariance principle with Skorohod's representation theorem (Theorem 1.10), one obtains that if ε_n are positive constants tending to zero, then the random variables $(S^{\varepsilon_n}_t)_{t\geq 0}$ for different values of n can be coupled to a Brownian motion $(B_t)_{t\geq 0}$ in such a way that

$$\sup_{t\in[0,T]} \big|S^{\varepsilon_n}_t - B_t\big| \underset{n\to\infty}{\longrightarrow} 0 \quad \text{a.s.} \quad \forall T < \infty. \tag{2.5}$$

We conclude this section with a well-known fact.

Lemma 2.2 (Brownian scaling) *If $B = (B_t)_{t\geq 0}$ is a Brownian motion and $\lambda > 0$, then the process $B^\lambda := \theta_\lambda B$ is equally distributed with B.*

Proof This is of course well-known, but it is interesting to observe that it actually follows from Theorem 2.1. Indeed, the latter says that if $\varepsilon_n > 0$ converge to zero, then the processes $\theta_{\varepsilon_n} S$ converge weakly in law to B. Since the map $\Pi(\mathbb{R}) \ni \pi \mapsto \theta_\lambda \pi \in \Pi(\mathbb{R})$ is continuous, it follows that the processes $\theta_{\lambda\varepsilon_n} S$ converge weakly in law to B^λ. On the other hand, since $\varepsilon'_n := \lambda\varepsilon_n$ are positive constants tending to zero, Theorem 2.1 also tells us that the processes $\theta_{\lambda\varepsilon_n} S$ converge weakly in law to B, so B^λ and B must be equal in law. This proof reveals a general fact: a probability law that arises as the scaling limit of other probability laws must itself be scale invariant. ∎

2.2 Brownian local time

Recall from (2.1) that $\mathcal{C}_0$ is the space of continuous functions $f:[0,\infty)\to\mathbb{R}$ that satisfy $f_0=0$. We let

$$m_t(f):=\inf_{0\le s\le t} f_s \qquad \big(t\ge 0,\ f\in\mathcal{C}_0\big) \tag{2.6}$$

denote the *running minimum* of the function f. We will be interested in

$$g_t:=f_t-m_t(f) \qquad (t\ge 0).$$

We observe that $g_t\ge 0$ and $h_t:=-m_t(f)$ is a nondecreasing function that increases only at times when $g_t=0$. The following lemma says that these properties characterise g and h uniquely. Note that if $h\in\mathcal{C}_0$ is nondecreasing, then it is the distribution function of a measure on $[0,\infty)$, which we denote by $\mathrm{d}h$. Condition (iii) below says that this measure is concentrated on the set $\{t\in[0,\infty):g(t)=0\}$. This makes precise the intuitive concept that h increases only at times when $g_t=0$.

Lemma 2.3 (Skorohod reflection) *For each $f\in\mathcal{C}_0$, there exist unique functions $g,h\in\mathcal{C}_0$ such that*

(i) $g_t=f_t+h_t$ $(t\ge 0)$,

(ii) $g\ge 0$ *and h is nondecreasing,*

(iii) $\displaystyle\int_0^\infty 1_{\{g(t)>0\}}\mathrm{d}h(t)=0.$

These functions are given by

$$g_t=f_t-m_t(f) \quad \textit{and} \quad h_t=-m_t(f) \qquad (t\ge 0). \tag{2.7}$$

Proof (sketch) It is not hard to check that if we define g and h by (2.7), then (i)–(iii) are satisfied. To prove uniqueness, it suffices to show that if g,h and g',h' both solve (i)–(iii), then $g'\le g$. Imagine that $g'_t>g_t$ for some $t>0$. Let $s:=\sup\{u\in[0,t]:g'_u=g_u\}$. Then $g'_u>g_u$ for all $s<u\le t$. By (i) we have $h'_s=h_s$. Now

$$g'_t-g_t=\big(f_t+h'_t\big)-\big(f_t+h_t\big)=h'_t-h_t. \tag{2.8}$$

By (ii) we have $g\ge 0$ and hence $g'_u>g_u\ge 0$ for all $s<u\le t$, which by (iii) implies that $h'_t=h'_s$. On the other hand, by (ii) h is nondecreasing and

hence $h_t \geq h_s$. It follows that the right-hand side of (2.8) is $\leq h'_s - h_s = 0$, which contradicts $g'_t > g_t$. ∎

We will especially be interested in the case that the function f from Lemma 2.3 is Brownian motion. In this case, the function g is reflected Brownian motion, and h is its local time at the origin. To explain this in a bit more detail, we need to take a small detour.

If $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion, then we can define a stochastic process $(\ell_t)_{t\geq 0}$ taking values in the space $\mathcal{M}(\mathbb{R}^d)$ of finite measures on $\mathbb{R}^d$ by

$$\int_{\mathbb{R}^d} \ell_t(\mathrm{d}x)\, f(x) := \int_0^t \mathrm{d}t\, f(B_t) \qquad \big(t \geq 0,\ f \in B_{\mathrm{b}}(\mathbb{R}^d)\big).$$

The random measure ℓ_t is called the *occupation local measure* of the Brownian motion $(B_t)_{t\geq 0}$. In particular

$$\ell_t(A) = \int_0^t \mathrm{d}t\, 1_A(B_t) \qquad \big(A \in \mathcal{B}(\mathbb{R}^d)\big)$$

is the amount of time the Brownian motion has spent inside a measurable set A up to time t. In one dimension, it is well-known that ℓ_t has a density with respect to the Lebesgue measure. The following theorem is originally due to Trotter. The process $(L_t)_{t\geq 0}$ below is called *Brownian local time.*

Theorem 2.4 (Brownian local time) *Let $(B_t)_{t\geq 0}$ be a one-dimensional Brownian motion. Then almost surely, there exists a random continuous function $L : [0,\infty) \times \mathbb{R} \to [0,\infty)$ such that*

$$\int_{\mathbb{R}^d} \mathrm{d}x\, L_t(x) f(x) = \int_0^t \mathrm{d}t\, f(B_t) \qquad \big(t \geq 0,\ f \in B_{\mathrm{b}}(\mathbb{R}^d)\big).$$

Modern proofs of Theorem 2.4 are based on *Tanaka's formula*, which says that

$$|B_t| = \int_0^t \mathrm{sgn}(B_s)\, \mathrm{d}B_s + L_t(0) \qquad (t \geq 0), \tag{2.9}$$

where the integral is an Itô stochastic integral. Tanaka's formula can be used as a definition of Brownian local time, for which one then proves the properties described in Theorem 2.4. For details, we refer to [McK69, Mey76, RW87]. In fact, in the remainder of this chapter, we will mostly work with Tanaka's formula as the definition of $L_t(0)$ and do not really need its interpretation as local time in the sense of Theorem 2.4.

Proposition 2.5 (Reflected Brownian motion) *Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion and let $(L_t(0))_{t\geq 0}$ be its local time at* 0. *Let $W = (W_t)_{t\geq 0}$ be another standard Brownian motion and let*

$$A_t := W_t - m_t(W) \quad \textit{and} \quad L_t := -m_t(W) \qquad (t \geq 0). \tag{2.10}$$

Then

$$\mathbb{P}\big[\big(|B_t|, L_t(0)\big)_{t\geq 0} \in \cdot\,\big] = \mathbb{P}\big[(A_t, L_t)_{t\geq 0} \in \cdot\,\big].$$

Proof (sketch) Let $(B_t)_{t\geq 0}$ be a Brownian motion and let

$$W_t := -\int_0^t \mathrm{sgn}(B_s)\,\mathrm{d}B_s \qquad (t \geq 0).$$

It is not hard to show that $W = (W_t)_{t\geq 0}$ is a Brownian motion. We will show that $A_t = |B_t|$ and $L_t = L_t(0)$ $(t \geq 0)$. We apply Lemma 2.3. Tanaka's formula (2.9) says that $|B_t| = L_t(0) - W_t$ $(t \geq 0)$. Clearly $|B_t|$ is nonnegative and $L_t(0)$ is nondecreasing and increases only when $|B_t| = 0$. For the details, we refer to [KS91, Thm 3.6.17]. ∎

2.3 Scaling limit of reflected random walk

Let S be the simple random walk defined in Section 2.1 and let $(R_t, K_t)_{t\geq 0}$ be defined by

$$R_t := S_t - m_t(S) \quad \text{and} \quad K_t := -m_t(S) \qquad (t \geq 0). \tag{2.11}$$

It is easy to see that $(R_t, K_t)_{t\in[0,\infty)}$ is the linear interpolation of the discrete time process $(R_n, K_n)_{n\in\mathbb{N}}$. Moreover, $(R_n)_{n\in\mathbb{N}}$ is a Markov chain with state space $\mathbb{N}$ and transition probabilities

$$P(x, y) = \mathbb{P}\big[R_n = y \,\big|\, R_{n-1} = x\big] \qquad (x, y \in \mathbb{N})$$

given by

$$\left.\begin{array}{l} P(x, x+1) = \frac{1}{2}, \\ P(x, x-1) = \frac{1}{2}, \end{array}\right\} \quad (x > 0) \quad \text{and} \quad \begin{array}{l} P(0, 1) = \frac{1}{2}, \\ P(0, 0) = \frac{1}{2}. \end{array}$$

In words, in each step, the process R_n jumps up or down by one with equal probabilities, except when this would result in a negative value, in which case the chain stays in 0. We call $(R_n)_{n\geq 0}$ a random walk with *reflection* at zero. It is now easy to see that the process K from (2.11) is given by

$$K_n = \sum_{k=1}^{n} 1_{\{R_{k-1} = R_k = 0\}} \qquad (n \geq 0).$$

Informally, K_n counts the number of times the chain $(R_n)_{n\geq 0}$ has "attempted to jump below zero", but was reflected. The following theorem says that the process (R, K) has a diffusive scaling limit.

Theorem 2.6 (Scaling limit of reflected random walk) *Let (R, K) be defined in (2.11) and for each $\varepsilon > 0$, let $(R^\varepsilon, K^\varepsilon)$ denote the diffusively rescaled process*

$$(R^\varepsilon_t, K^\varepsilon_t) := \big(\varepsilon R_{\varepsilon^{-2}t}, \varepsilon K_{\varepsilon^{-2}t}\big) \qquad (t \geq 0). \tag{2.12}$$

Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion and let $(L_t(0))_{t\geq 0}$ be its local time at 0. Then

$$\mathbb{P}\big[(R^\varepsilon_t, K^\varepsilon_t)_{t\geq 0} \in \cdot\,\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[\big(|B_t|, L_t(0)\big)_{t\geq 0} \in \cdot\,\big],$$

where $\Rightarrow$ denotes weak convergence of probability measures on $\mathcal{C}_{[0,\infty)}(\mathbb{R}^2)$, equipped with the topology of locally uniform convergence.

Proof We observe that

$$R^\varepsilon_t := S^\varepsilon_t - m_t(S^\varepsilon) \quad \text{and} \quad K^\varepsilon_t := -m_t(S^\varepsilon) \qquad (t \geq 0),$$

where S^ε is the diffusively rescaled random walk defined in (2.3). It is straightforward to check that the map

$$\mathcal{C}_0 \ni f \mapsto (g, h) \in \mathcal{C}_{[0,\infty)}(\mathbb{R}^2)$$

defined in (2.7) is continuous with respect to the topology of locally uniform convergence. Therefore, Theorem 2.1 implies that

$$\mathbb{P}\big[(R^\varepsilon_t, K^\varepsilon_t)_{t\geq 0} \in \cdot\,\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[(A_t, L_t)_{t\geq 0} \in \cdot\,\big],$$

where $(A_t, L_t)_{t\geq 0}$ is the reflected Brownian motion defined in (2.10). The claim now follows from Proposition 2.5. ∎

Theorem 2.6 yields the following useful consequence.

Lemma 2.7 (Scale invariance) *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion and let $(L_t(0))_{t\geq 0}$ is its local time at* 0*. Then*

$$\mathbb{P}\big[\big(|B_t|, L_t(0)\big)_{t\geq 0} \in \cdot\,\big] = \mathbb{P}\big[\big(\lambda|B_{\lambda^{-2}t}|, \lambda L_{\lambda^{-2}t}(0)\big)_{t\geq 0} \in \cdot\,\big] \qquad (\lambda > 0).$$

Proof The proof is very similar to the proof of Lemma 2.2. As we observed there, a probability law that arises as the scaling limit of other probability laws must itself be scale invariant. In the present setting, we can make this general principle precise as follows. Fix $\lambda > 0$ and let ε_n be positive constants tending to zero. By (2.12),

$$(R^{\lambda\varepsilon}_t, K^{\lambda\varepsilon}_t) = (\lambda R^{\varepsilon}_{\lambda^{-2}t}, \lambda K^{\varepsilon}_{\lambda^{-2}t}) \qquad (t \geq 0),$$

so Theorem 2.6 tells us that

$$\mathbb{P}\big[(R^{\lambda\varepsilon}_t, K^{\lambda\varepsilon}_t)_{t\geq 0} \in \cdot\,\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[\big(\lambda|B_{\lambda^{-2}t}|, \lambda L_{\lambda^{-2}t}(0)\big)_{t\geq 0} \in \cdot\,\big].$$

However, $\lambda\varepsilon_n$ are positive constants tending to zero, so Theorem 2.6 also tells us that

$$\mathbb{P}\big[(R^{\lambda\varepsilon}_t, K^{\lambda\varepsilon}_t)_{t\geq 0} \in \cdot\,\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[\big(|B_t|, L_t(0)\big)_{t\geq 0} \in \cdot\,\big].$$

∎

2.4 Excursion decomposition

We will be interested in the theory of Brownian excursions. Our exposition is loosely inspired by [Rog89]. Recall from (2.1) that $\mathcal{C}_0$ is the space of continuous functions $f : [0,\infty) \to \mathbb{R}$ that satisfy $f_0 = 0$. We let

$$\begin{aligned}\mathcal{R}_0 := \big\{(g,h) : g,h \in \mathcal{C}_0,\ g \geq 0,\ h \text{ is nondecreasing},\\ \text{and} \quad \textstyle\int_0^\infty 1_{\{g(t)>0\}}\mathrm{d}h(t) = 0\big\},\end{aligned} \tag{2.13}$$

denote the set of pairs of functions (g,h) that satisfy conditions (ii) and (iii) of Lemma 2.3. We view $\mathcal{R}_0$ as a subset of $\mathcal{C}_{[0,\infty)}(\mathbb{R}^2)$ and equip it with the topology of locally uniform convergence. In Lemma 2.3, we have seen that setting $g_t := f_t - m_t(f)$ and $h_t := -m_t(f)$ defines a bijection $f \mapsto (g,h)$ from $\mathcal{C}_0$ to $\mathcal{R}_0$.

We now want to go one step further, and decompose the function g in excursions away from zero. Recall that σ_π and τ_π, defined in (1.9), denote the starting time and final time of a path π. We define a *space of excursions* by

$$\mathcal{E} := \big\{\pi \in \Pi([0,\infty)) : \sigma_\pi = 0,\ 0 \leq \tau_\pi < \infty,\ \pi(0) = \pi(\tau_\pi) = 0\big\}. \tag{2.14}$$

We call the final time τ_π of an excursion $\pi \in \mathcal{E}$ the *duration* of π. We observe that $\mathcal{F} := \{\pi : \pi(0) = \pi(\tau_\pi) = 0,\ \sigma_\pi = 0\}$ is a closed subset of $\Pi([0,\infty))$ and $\mathcal{E}$ is an open subset of $\mathcal{F}$, so using Proposition 1.4 we see that $\mathcal{E}$ is a Polish space. We set

$$\mathring{\mathcal{E}} := \big\{\pi \in \mathcal{E} : \tau_\pi > 0,\ \pi(t) > 0\ \forall 0 < t < \tau_\pi\big\}. \tag{2.15}$$

We call elements of $\mathring{\mathcal{E}}$ *proper excursions.*

Let $h \in \mathcal{C}_0$ be nondecreasing. By definition, a *plateau* of h is an open interval $\iota = (\iota^-, \iota^+)$ with $0 \leq \iota^- < \iota^+ < \infty$ such that $h_{\iota^-} = h_{\iota^+}$, and no strictly larger open subinterval of $[0,\infty)$ has this property. We set

$$\mathcal{I}(h) := \big\{\iota : \iota \text{ is a plateau of } h\big\}. \tag{2.16}$$

For brevity, we write

$$h_\iota := h_{\iota^-} = h_{\iota^+} \qquad (\iota \in \mathcal{I}(h)).$$

For each $(g,h) \in \mathcal{R}_0$ and $\iota \in \mathcal{I}(h)$, setting

$$\tau_\iota := \iota^+ - \iota^- \quad \text{and} \quad \pi^{g,\iota}_t := g_{t-\iota^-} \quad (0 \leq t \leq \tau_\iota)$$

defines an excursion $\pi^{g,\iota} \in \mathcal{E}$ with duration τ_ι. Given a function $f \in \mathcal{C}_0$ and functions $(g,h) \in \mathcal{R}_0$ defined as in (2.7), we set

$$\Xi(f) := \big\{\big(h_\iota, \pi^{g,\iota}\big) : \iota \in \mathcal{I}(h)\big\}. \tag{2.17}$$

We will especially be interested in the case that f is a (diffusively rescaled) simple random walk, or Brownian motion. In this case, g is a reflected random walk or Brownian motion and h is its reflection local time at the origin. The set Ξ records all excursions of the reflected random walk or Brownian motion away from the origin together with the reflection local time when such an excursion happens.

It follows from the way we have defined plateaus that $h(\iota) \neq h(\iota')$ for each $\iota, \iota' \in \mathcal{I}(h)$ with $\iota \neq \iota'$. We use this to define a function $s \mapsto E^s$ from $[0,\infty)$ to $\mathcal{E}$ by

$$E^s = \begin{cases} \pi & \text{if } (s,\pi) \in \Xi \text{ for some } \pi \in \mathcal{E}, \\ o & \text{otherwise,} \end{cases} \tag{2.18}$$

where $o \in \mathcal{E}$ denotes the *trivial excursion* of duration $\tau_o := 0$.

The excursion set $\Xi(S)$ of the simple random walk S of Section 2.1 is easy to understand. Let $(R, K) = (R_n, K_n)_{n\geq 0}$ by the reflected random walk defined in (2.11). We define inductively $\iota^-(0) := 0$ and

$$\left.\begin{aligned} \iota^+(k) &:= \inf\{i \geq \iota^-(k) : K_{i+1} > K_i\}, \\ \iota^-(k+1) &:= \iota^+(k) + 1 \end{aligned}\right\} \qquad (k \geq 0).$$

Then the set of plateaus of the function K is

$$\mathcal{I}(K) = \big\{\big(\iota^-(k), \iota^+(k)\big) : k \geq 0,\ \iota^-(k) < \iota^+(k)\big\},$$

and the excursion set $\Xi(S)$ is given by

$$\Xi(S) = \big\{(k, E^k) : k \in \mathbb{N},\ E^k \neq o\big\}, \tag{2.19}$$

where E^k, defined as in (2.18), is the excursion that belongs to the plateau $(\iota^-(k), \iota^+(k))$ if $\iota^-(k) < \iota^+(k)$, and the trivial excursion o if $\iota^-(k) = \iota^+(k)$. For reflected random walk, not all excursions are proper excursions, since it may happen that $R_i = 0$ for some $\iota^-(k) < i < \iota^+(k)$. Since the process "starts anew" after each increase of K, it is easy to see that:

$$\text{The } \mathcal{E}\text{-valued random variables } (E^k)^{k\in\mathbb{N}} \text{ are i.i.d.} \tag{2.20}$$

For Brownian motion, the situation is more complex, since we can no longer enumerate the excursions by the time at which they occur. Nevertheless, something similar to the i.i.d. property of (2.20) still holds. The following theorem is due to Itô [Ito71]. The σ-finite measure ν below is called the *excursion measure*.

Theorem 2.8 (Poisson set of excursions) *There exists a σ-finite measure ν on $\mathcal{E}$ such that the set Ξ is a Poisson point set with intensity measure $\ell\otimes\nu$, where ℓ is the Lebesgue measure on $[0, \infty)$. The measure ν is concentrated on $\mathring{\mathcal{E}}$.*

As a preparation for the proof of Theorem 2.8, we make the following observation.

Lemma 2.9 (Only proper excursions) *The excursion set $\Xi(B)$ of a Brownian motion B is concentrated on the set of proper excursions $\mathring{\mathcal{E}}$.*

Proof To show that Ξ is concentrated on $\mathring{\mathcal{E}}$, one has to show that $L_t(0)$ increases each time B_t hits zero. By Proposition 2.5, one may equivalently show that if a Brownian motion $W = (W_t)_{t\geq 0}$ is started at some initial state $W_0 = x > 0$ and $\tau_0 := \inf\{t \geq 0 : W_t = 0\}$, then W immediately crosses the time axis, i.e., $\inf\{t \geq 0 : W_t < 0\} = \tau_0$. By the strong Markov property, it suffices to show that Brownian motion started in zero immediately crosses the time axis, which is well-known. ∎

Proof of Theorem 2.8 (crude sketch) In Section 2.6 we will give a proof of Theorem 2.8 based on finite approximation. Traditionally, there is a tendency to view such proofs as ugly.[1] Whether that is a good philosophy is questionable. Here, we sketch the outline of a classical proof using stochastic analysis.

The idea is to show that for each measurable $A \subset \mathcal{E}$, the process

$$N_s(A) := \Xi\big([0,s] \times A\big) \qquad (s \geq 0)$$

is stationary with independent increments, and moreover, if $A_1, \ldots, A_n$ are disjoint, then the processes $(N_s(A_1))_{s\geq 0}, \ldots, (N_s(A_n))_{s\geq 0}$ are independent. For each deterministic $s \geq 0$, the random time

$$\rho_s := \inf\big\{t \geq 0 : L_t(0) \geq s\big\}$$

is a stopping time for the Markov process $(|B_t|, L_t(0))_{t\geq 0}$. Using the strong Markov property for the stopping time ρ_s, one obtains that

$$\big(|B_{\rho_s+t}|, L_{\rho_s+t}(0) - L_{\rho_s}(0)\big)_{t\geq 0} \text{ is independent of } \big(|B_t|, L_t(0)\big)_{0\leq t\leq \rho_s},$$

and equally distributed with the original process $(|B_t|, L_t(0))_{t\geq 0}$. Using this, one obtains that for any $0 \leq s_1 \leq s_2$, the increment $N_{s_2}(A) - N_{s_1}(A)$ is independent of the restriction of Ξ to $[0, s_1] \times \mathcal{E}$ and equally distributed with $N_{s_2 - s_1}(A)$, i.e., the process $(N_s(A))_{s\geq 0}$ is stationary with independent increments as claimed. Using also that it is a pure jump process one can now apply abstract results to conclude that Ξ must be a Poisson point set with the claimed properties. ∎

It is possible to "invert" the decomposition into excursions and reconstruct a reflected random walk or reflected Brownian motion from the set Ξ defined in (2.17). The construction is slightly different in the discrete and continuous cases. For the reflected random walk $(R_t, K_t)_{t\geq 0}$, we set

$$\begin{aligned}
\rho_s &:= s + \sum_{(u,\pi)\in\Xi:\, u<s} \tau_\pi && (s \geq 0),\\
K_t &:= \sup\{s \geq 0 : \rho_s \leq t\} && (t \geq 0),\\
R_t &:= E^{K_t}_{t-\rho_{K_t}} && (t \geq 0).
\end{aligned} \tag{2.21}$$

[1]For example, it seems the main reason, apart from some minor inaccuracies, why the original proof of the Jordan curve theorem was not widely accepted, was that it used discrete approximation.

For the reflected Brownian motion $\big(|B_t|, L_t(0)\big)_{t\geq 0}$, we set

$$\begin{aligned} \rho_s &:= \sum_{(u,\pi)\in\Xi:\,u<s} \tau_\pi && (s \geq 0),\\ L_t(0) &:= \sup\{s \geq 0 : \rho_s \leq t\} && (t \geq 0),\\ |B_t| &:= E^{(L_t)}_{t-\rho_{L_t}} && (n \geq 0). \end{aligned} \tag{2.22}$$

The only difference between these formulas is in the definition of the function $(\rho_s)_{s\geq 0}$, which is the inverse of the reflection local times $(K_t)_{t\geq 0}$ and $(L_t(0))_{t\geq 0}$, respectively. In the discrete case, compared to the continuum case, we have to add a term $+s$ to to the definition of ρ_s. This has to do with the fact that K_t increases at speed one during the times when R_t is zero, while $L_t(0)$ increases at infinite speed during the times when $|B_t|$ is zero.

Formula (2.22) shows how to construct the absolute value of Brownian motion, i.e., the process $(|B_t|)_{t\geq 0}$, together with the local time at the origin of $(B_t)_{t\geq 0}$, from a Poisson set of excursions. In a similar way, one can also construct the Brownian motion $(B_t)_{t\geq 0}$ itself (instead of its absolute value). The idea is to assign signs to the excursions that are i.i.d. and uniformly distributed on $\{-1, +1\}$. It is also interesting to consider signs that are i.i.d. but not uniformly distributed on $\{-1, +1\}$. In this case, one obtains a Markov process known as *skew Brownian motion*.

The following proposition is a consequence of Brownian scaling. As before, we view paths as compact subsets of $\mathcal{R}(\mathbb{R})$ and we let $\theta_\lambda\pi$ denote the image of π under the diffusive scaling map θ_λ defined in (2.2). In this way, in (2.23) below, we naturally view θ_λ as a map from $\mathcal{E}$ to $\mathcal{E}$.

Proposition 2.10 (Diffusive scaling) *The excursion measure ν from Theorem 2.8 satisfies*

$$\nu \circ \theta_\lambda^{-1} = \lambda\nu \qquad (\lambda > 0). \tag{2.23}$$

Proof Let $(B_t)_{t\geq 0}$ be a standard Brownian motion and let $(L_t(0))_{t\geq 0}$ be its local time at 0. Fix $\lambda > 0$ and set

$$B^\lambda_t := \lambda B_{\lambda^{-2}t} \quad \text{and} \quad L^\lambda_t(0) := \lambda L_{\lambda^{-2}t}(0) \qquad (t \geq 0).$$

By Lemma 2.7, the processes $(|B_t|, L_t(0))_{t\geq 0}$ and $(|B^\lambda_t|, L^\lambda_t(0))_{t\geq 0}$ are equally distributed. Define Ξ and Ξ^λ as in (2.17) in terms of $(|B_t|, L_t(0))_{t\geq 0}$ and $(|B^\lambda_t|, L^\lambda_t(0))_{t\geq 0}$, respectively. Then

$$\Xi^\lambda = \big\{(\lambda s, \theta_\lambda\pi) : (s, \pi) \in \Xi\big\}.$$

Since both Ξ and Ξ^λ are Poisson point sets on $[0,\infty)\times\mathcal{E}$ with intensity measure $\ell\otimes\nu$, we see that the measure $\ell\otimes\nu$ is equal to its image under the map

$$(s,\pi)\mapsto(\lambda s,\theta_\lambda\pi).$$

In particular, for any measurable $A\subset\mathcal{E}$, we have

$$\begin{aligned}\lambda\nu(A)&=\ell\otimes\nu\big([0,\lambda]\times A\big)\\&=\ell\otimes\nu\big([0,1]\times\theta_\lambda^{-1}(A)\big)=\nu\circ\theta_\lambda^{-1}(A).\end{aligned}$$

∎

2.5 Standard excursions

We continue our study of the excursion measure ν from Theorem 2.8. We let

$$\mathcal{H}_r:=\big\{\pi\in\mathring{\mathcal{E}}:\sup_{0\le t\le\tau_\pi}\pi(t)\ge r\big\}\qquad(r\ge0)\tag{2.24}$$

denote the set of proper excursions that have height at least r. The next lemma determines $\nu(\mathcal{H}_r)$.

Lemma 2.11 (Height of the excursion) *Let ν be the excursion measure from Theorem 2.8. Then*

$$\nu(\mathcal{H}_r)=r^{-1}\qquad(r>0).\tag{2.25}$$

Proof Let $(B_t)_{t\ge0}$ be a standard Brownian motion and let $(L_t(0))_{t\ge0}$ is its local time at 0. Let

$$\sigma_r:=L_{\tau_r}\quad\text{with}\quad\tau_r:=\inf\big\{t\ge0:|B_t|=r\big\}.$$

Then

$$\sigma_r:=\inf\big\{s\ge0:\Xi\cap([0,s]\times\mathcal{H}_r)\neq\emptyset\big\}.$$

By Theorem 2.8, σ_r is exponentially distributed with mean $1/\nu(\mathcal{H}_r)$. By Tanaka's formula (2.9), $|B_t|-L_t(0)$ is a martingale. By optional stopping, it follows that

$$\mathbb{E}\big[|B_{\tau_r\wedge t}|\big]=\mathbb{E}\big[L_{\tau_r\wedge t}(0)\big]\qquad(t\ge0).$$

Letting $t\to\infty$, using the fact that $L_{\tau_r\wedge t}(0)$ increases to $L_{\tau_r}=\sigma_r$, and using dominated convergence for the left-hand side, together with $|B_{\tau_r}|=r$, we obtain that $\mathbb{E}[\sigma_r]=r$ and hence $\nu(\mathcal{H}_r)=1/r$. ∎

Since $0<\nu(\mathcal{H}_r)<\infty$ for each $r>0$, we can define a conditional probability laws $\nu(\,\cdot\,|\mathcal{H}_r)$ on $\mathcal{E}$ by the usual formula

$$\nu(A|\mathcal{H}_r):=\frac{\nu(A\cap\mathcal{H}_r)}{\nu(\mathcal{H}_r)}\qquad\big(A\in\mathcal{B}(\mathcal{E})\big).$$

For each excursion $E \in \mathcal{H}_r$, we let

$$\sigma_{E,r} := \inf\{t \geq 0 : E_t = r\} \tag{2.26}$$

denote the first time the excursion E reaches the height r. As before, τ_E denotes the duration of E.

Lemma 2.12 (Conditional excursion law) *For each $r > 0$, under the conditional law $\nu(\,\cdot\,|\mathcal{H}_r)$, the process*

$$(E_{\sigma_{E,r}+t})_{0 \leq t \leq \tau_R - \sigma_{E,r}}$$

is distributed as a Brownian motion started at r and stopped at the first time it hits zero.

Proof (sketch) Let $(B_t)_{t \geq 0}$ be a standard Brownian motion. Let

$$\begin{gathered}\sigma_r^1 := \inf\{t \geq 0 : |B_t| = r\}, \quad \sigma_r^2 := \inf\{t \geq \sigma_r^1 : |B_t| = 0\},\\ \text{and} \quad \sigma_r^0 := \sup\{t < \sigma_r^1 : |B_t| = 0\},\end{gathered}$$

and let $E \in \mathring{\mathcal{E}}$ be the excursion with duration $\tau_E := \sigma_r^2 - \sigma_r^0$ defined by

$$E_t := B_{\sigma_r^0 + t} \qquad (0 \leq t \leq \tau_E).$$

Then E_t is the first excursion in the Poisson point set Ξ of Theorem 2.8 that has height $\geq r$. Using the strong Markov property of Poisson point sets, one sees that E is distributed according to the conditional law $\nu(\,\cdot\,|\mathcal{H}_r)$. Using the strong Markov property of Brownian motion, one sees that

$$(B_{\sigma_r^1 + t})_{0 \leq t \leq \sigma_r^2 - \sigma_r^1}$$

is distributed as a Brownian motion started at r and stopped at the first time it hits zero. ∎

We let

$$\mathcal{D}_t := \big\{\pi \in \mathring{\mathcal{E}} : \tau_\pi > t\big\} \qquad (t \geq 0)$$

denote the set of excursions that have duration at least t. The next lemma determines $\nu(\mathcal{D}_t)$.

Lemma 2.13 (Duration of the excursion) *Let ν be the excursion measure from Theorem 2.8. Then*

$$\nu(\mathcal{D}_t) = \frac{2}{\sqrt{2\pi}} t^{-1/2} \qquad (t > 0). \tag{2.27}$$

Proof of Lemma 2.13 We define $\mathcal{H}_r$ as in (2.24) and for each $E \in \mathcal{H}_r$ we define $\sigma_{E,r}$ as in (2.26). As before, τ_E denotes the duration of E. For each $t > 0$, we set

$$\mathcal{H}_{r,t} := \big\{E \in \mathcal{H}_r : \tau_E \geq \sigma_{E,r} + t\big\},$$

i.e., these are all excursions that reach the height r and after that live for at least time t. Lemma 2.12 implies that

$$\nu(\mathcal{H}_{r,t}) = \nu(\mathcal{H}_r)\mathbb{P}\big[r + B_s > 0\ \forall 0 \leq s \leq t\big],$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion. It is a consequence of the reflection principle that

$$\mathbb{P}\big[r + B_s > 0\ \forall 0 \leq s \leq t\big] = \mathbb{P}\big[|B_t| \leq r\big] = \int_{-r}^{r} \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2}\mathrm{d}x.$$

Using Lemma 2.11, which tells us that $\nu(\mathcal{H}_r) = r^{-1}$, it follows that

$$\nu(\mathcal{H}_{r,t}) = r^{-1}\Big[\frac{2rt^{-1/2}}{\sqrt{2\pi}} + O(r^2)\Big] \quad \text{as } r \to 0.$$

Letting $r \to 0$, using the fact that $\mathcal{H}_{r,t}$ increases to $\mathcal{D}_t$, the claim follows. ∎

We let

$$\mathcal{E}_1 := \big\{\pi \in \mathcal{E} : \tau_\pi = 1\big\} \tag{2.28}$$

denote the space of excursions of duration one and set $\mathring{\mathcal{E}}_1 := \mathcal{E}_1 \cap \mathring{\mathcal{E}}$. A random variable whose law is the probability measure ν_1 from Proposition 2.14 below is called a *standard Brownian excursion*.

Proposition 2.14 (Decomposition of the excursion measure) *Let ρ be the measure on $(0,\infty)$ defined as*

$$\rho(\mathrm{d}t) := \frac{1}{\sqrt{2\pi}} t^{-3/2}\mathrm{d}t. \tag{2.29}$$

There exists a probability measure ν_1 on $\mathring{\mathcal{E}}_1$ such that the excursion measure from Theorem 2.8 is the image of the measure $\rho \otimes \nu_1$ under the map

$$(0,\infty) \times \mathcal{E}_1 \ni (t, f) \mapsto \theta_t f \in \mathcal{E}. \tag{2.30}$$

Proof (sketch) The map in (2.30) is invertible. Its inverse is the map

$$\mathcal{E} \ni \pi \mapsto \big(\tau_\pi, \theta_{1/\sqrt{\tau_\pi}}\pi\big) \in (0,\infty) \otimes \mathcal{E}_1,$$

where as before τ_π denotes the duration of an excursion $\pi \in \mathcal{E}$. Let μ be the image of the excursion measure ν under this inverse map. Then Proposition 2.10 implies that

$$\mu \circ \psi_\lambda^{-1} = \lambda\mu \quad \text{where} \quad \psi_\lambda(s, f) := (\lambda^2 s, f) \qquad (\lambda, s > 0,\ f \in \mathcal{E}_1). \quad (2.31)$$

Using the fact that by Lemma 2.13, $\nu(\mathcal{D}_1)$ is finite, it follows that we can decompose μ as

$$\mu(\mathrm{d}\lambda, \mathrm{d}f) = \rho(\mathrm{d}\lambda)P(\lambda, \mathrm{d}f)$$

for some probability kernel P (compare Theorem 1.5). By (2.31), $P(\lambda, \cdot\,)$ does not depend on λ, so in fact $\mu = \rho \otimes \nu_1$ for some probability measure ν_1 on $\mathring{\mathcal{E}}_1$. The scaling relation (2.31) moreover implies that

$$\rho\big([\lambda^{-2}t, \infty)\big) = \lambda\rho\big([t, \infty)\big) \qquad (\lambda, t > 0),$$

which shows that there exists a constant $c > 0$ such that

$$\rho\big([t, \infty)\big) = ct^{-1/2} \qquad (\lambda > 0).$$

The correct formula for the constant c follows from Lemma 2.13. ∎

2.6 Scaling limits of excursions

In this section, we give a proof of Theorem 2.8 based on finite approximation. As a side result, we obtain that if S^ε are the diffusively rescaled simple random walks defined in (2.3) and let B is a standard Brownian motion, then the excursion sets $\Xi(S^\varepsilon)$ defined in (2.17) converge in an appropriate sense to $\Xi(B)$.

We first need a few definitions. By definition, a *local* subset of the set of excursions $\mathcal{E}$ is a measurable set $A \subset \mathcal{E}$ such that $o \notin \overline{A}$, where $\overline{A}$ denotes the closure of A and o denotes the trivial excursion of duration $\tau_o := 0$. Similarly, a *local* subset of $[0, \infty) \times \mathcal{E}$ is a measurable set $B \subset [0, \infty) \times \mathcal{E}$ such that

$$B \subset [0, S] \times A \quad \text{for some } S < \infty \text{ and local } A \subset \mathcal{E}.$$

We say that a measure ν on $\mathcal{E}$ is *locally finite*[2] if $\nu(A) < \infty$ for all local $A \subset \mathcal{E}$. Similarly, we say that a measure μ on $[0, \infty) \times \mathcal{E}$ is *locally finite* if

[2]We use this term in an unusual sense here. More usually, a measure μ on a locally compact space $\mathcal{X}$ is called *locally finite* if $\mu(K) < \infty$ for each compact $K \subset \mathcal{X}$. The space $\mathcal{E}$ is, however, not locally compact, so such a definition would not make much sense in our setting.

$\mu(B) < \infty$ for all local $B \subset [0,\infty) \times \mathcal{E}$. When $\mathcal{X} = \mathcal{E}$ or $= [0,\infty) \times \mathcal{E}$, we let $\mathcal{M}_{\rm loc}(\mathcal{X})$ denote the space of locally finite measures on $\mathcal{X}$.

The *support* of a measurable real function f defined on a topological space $\mathcal{X}$ is the set

$$\mathrm{supp}(f) := \overline{\{x \in \mathcal{X} : f(x) \neq 0\}},$$

where the overbar denotes closure. In the special case that $\mathcal{X} = \mathcal{E}$ or $= [0,\infty) \times \mathcal{E}$, we define a *local* function on $\mathcal{X}$ to be a measurable function $f : \mathcal{X} \to \mathbb{R}$ such that $\mathrm{supp}(f)$ is a local subset of $\mathcal{X}$. We let $\mathcal{C}_{\rm loc}(\mathcal{X})$ denote the space of all local bounded continuous functions on $\mathcal{X}$ and we let $\mathcal{M}_{\rm loc}(\mathcal{X})$ denote the space of all locally finite measures on $\mathcal{X}$.

Let $\mathcal{X}$ be a Polish space and let $(x_i)_{i\in I}$ be a countable collection of points in $\mathcal{X}$. Then

$$\xi := \sum_{i\in I} \delta_{x_i} \tag{2.32}$$

defines a counting measure on $\mathcal{X}$. In particular, if Ξ is a countable subset of $\mathcal{X}$, then Ξ defines a counting measure by

$$\xi_\Xi := \sum_{x\in\Xi} \delta_x. \tag{2.33}$$

Note that ξ_Ξ is *simple*, in the sense that $\xi_\Xi(\{x\}) \leq 1$ for all $x \in \mathcal{X}$. In general, counting measures of the form (2.32) need not be simple, since it may happen that $x_i = x_j$ for some $i \neq j$. We often tacitly identify countable subsets of $\mathcal{X}$ with their associated counting measures. In particular, we say that a countable subset Ξ of $\mathcal{X} = \mathcal{E}$ or $= [0,\infty) \times \mathcal{E}$ is *locally finite* if ξ_Ξ has this property.

Let $\phi : \mathcal{X} \to [0,1]$ be measurable and let $(\chi_i)_{i\in I}$ be independent *Bernoulli* random variables (i.e., variables with values in $\{0,1\}$) with $\mathbb{P}[\chi_i = 1] = \phi(x_i)$. Then the random counting measure

$$\xi' := \sum_{i\in I} \chi_i \delta_{x_i}$$

is called a *ϕ-thinning* of ξ. In the special case that $\mathcal{X}$ is either $\mathcal{E}$ or $[0,\infty)\times\mathcal{E}$, we let $\mathcal{N}_{\rm loc}(\mathcal{X})$ denote the space of all locally finite counting measures on $\mathcal{X}$. Then

$$K_\phi(\xi,\,\cdot\,) := \mathbb{P}\big[\xi' \in \cdot\,\big]$$

defines a probability kernel on $\mathcal{N}_{\rm loc}(\mathcal{X})$. Generalising our earlier definition of a thinning, when ξ and ξ' are random locally finite counting measures on $\mathcal{X}$, then we say that ξ' is a *ϕ-thinning* of ξ if

$$\mathbb{P}\big[\xi' \in \cdot \,\big|\, \xi\big] = K_\phi(\xi,\,\cdot\,).$$

For any counting measure ξ of the form (2.32) and measurable $\phi : \mathcal{X} \to [0,1]$, we introduce the notation

$$\phi^\xi := \prod_{i\in I} \phi(x_i) = e^{\int \xi(\mathrm{d}x)\,\log\phi(x)},$$

with the convention that $e^{-\infty} := 0$. If ξ' is a ϕ-thinning of ξ, then it is easy to see that

$$\mathbb{P}\big[\xi' = 0\big] = \mathbb{E}\big[(1-\phi)^\xi\big].$$

We say that μ on a measurable space $\mathcal{X}$ is *nonatomic* if $\mu(\{x\}) = 0$ for all $x \in \mathcal{X}$. Recall that a counting measure ξ is called *simple* if $\xi(\{x\}) \leq 1$ for all $x \in \mathcal{X}$. We need the following result.[3]

Theorem 2.15 (Poisson counting measure) *Let μ be a locally finite measure on $[0,\infty) \times \mathcal{E}$. Then there exists a random locally finite counting measure ξ on $[0,\infty) \times \mathcal{E}$ such that*

$$\mathbb{E}\big[(1-\phi)^\xi\big] = e^{-\int \mu(\mathrm{d}x)\phi(x)} \tag{2.34}$$

for each measurable $\phi : [0,\infty) \times \mathcal{E} \to [0,1]$. The law of ξ is uniquely determined by the requirement that (2.34) holds for all local continuous ϕ. If $B_1, \ldots, B_n$ are disjoint measurable local subsets of $[0,\infty) \times \mathcal{E}$, then

$$\xi(B_1), \ldots, \xi(B_n) \quad \text{are Poisson distributed with mean} \quad \mu(B_1), \ldots, \mu(B_n).$$

If μ is nonatomic, then ξ is almost surely simple.

Formula (2.34) has an interpretation in terms of thinning. Let $\phi\mu$ denote the measure μ weighted with the density ϕ. If ξ' is a ϕ-thinning of ξ, then ξ' is a Poisson counting measure with intensity measure $\phi\mu$. In particular, if $\int \phi\,\mathrm{d}\mu < \infty$, then the number of points of ξ' is Poisson distributed with mean $\int \phi\,\mathrm{d}\mu < \infty$, and hence $\mathbb{P}[\xi' = 0] = \exp(-\int \phi\,\mathrm{d}\mu)$, which is formula (2.34).

We now turn our attention to the proof of Theorem 2.8. We will use discrete approximation. Let S be the simple random walk from Section 2.1, let S^ε be the diffusively rescaled random walk from (2.3), and let B be standard Brownian motion. We recall from (2.19) that the excursion set of S is given by

$$\Xi(S) = \big\{(k, E^k) : k \in \mathbb{N},\ E^k \neq o\big\},$$

[3] This is largely standard, but many sources such as [Kal97, Chapter 10] treat only locally compact spaces. Oir definition of local finiteness is also nonstandard.

where $(E^k)^{k\in\mathbb{N}}$ are the i.i.d. excursions from (2.20). It follows that

$$\Xi(S^\varepsilon) = \big\{(\varepsilon k, \theta_\varepsilon E^k) : k \in \mathbb{N},\ E^k \neq o\big\}. \tag{2.35}$$

Note that in view of (2.12), we have to rescale the reflection local time k by a factor ε and not by ε^2. We will prove Theorem 2.8 together with the following theorem, which describes the tail of the law of E^0, i.e., in the small probabilities of very large excursions.

Theorem 2.16 (Tail of the excursion law) *Let ν be the excursion measure from Theorem 2.8. One has*

$$\varepsilon^{-1}\mathbb{E}\big[g(\theta_\varepsilon E^0)\big] \underset{\varepsilon\to 0}{\longrightarrow} \int_{\mathcal{E}} g(\pi)\,\nu(\mathrm{d}\pi) \tag{2.36}$$

for each $g \in \mathcal{C}_{\mathrm{loc}}(\mathcal{E})$.

The proof of Theorem 2.8 depends on two technical results, the proofs of which will be postponed till the next section. Recall from (2.1) that $\mathcal{C}_0 := \big\{f \in \mathcal{C}_{[0,\infty)}(\mathbb{R}) : f_0 = 0\big\}$.

Lemma 2.17 (Locally finite excursion set) *For each $f \in \mathcal{C}_0$ such that $\liminf_{t\to\infty} f_t = -\infty$, the set $\Xi(f)$ defined in (2.17) is locally finite.*

We have seen in (2.5) that it is possible to couple diffusively rescaled random walks S^{ε_n} and a Brownian motion B such that almost surely $S^{\varepsilon_n} \to B$ locally uniformly. The following theorem says that then also the associated excursion sets converge. In (2.37) below, we identify the countable sets $\Xi(S^{\varepsilon_n})$ and $\Xi(B)$ with their associated counting measures as in (2.33).

Theorem 2.18 (Scaling limit of excursion sets) *Let ε_n be positive constants tending to zero, let S^{ε_n} be the diffusively rescaled simple random walk defined in (2.3) and let B be a standard Brownian motion. Assume that these random variables are coupled as in (2.5). Then the excursion sets $\Xi(S^{\varepsilon_n})$ and $\Xi(B)$ defined in (2.17) almost surely satisfy*

$$(1-\phi)^{\Xi(S^{\varepsilon_n})} \underset{n\to\infty}{\longrightarrow} (1-\phi)^{\Xi(B)} \tag{2.37}$$

for all local continuous $\phi : [0,\infty) \times \mathcal{E} \to [0,1]$.

Proof of Theorems 2.8 and 2.16 Let ε_n be positive constants tending to zero. We fix a local continuous function $g : \mathcal{E} \to [0,1]$ and a continuous compactly supported function $h : [0,\infty) \to [0,1]$. Then $\phi(s,\pi) := h(s)g(\pi)$

defines a local continuous function $\phi : [0,\infty) \times \mathcal{E} \to [0,1]$. Applying Theorem 2.18, using bounded pointwise convergence to interchange the integral and the limit, we see that

$$\mathbb{E}\big[(1-\phi)^{\Xi(S^{\varepsilon_n})}\big] \underset{n\to\infty}{\longrightarrow} \mathbb{E}\big[(1-\phi)^{\Xi(B)}\big]. \tag{2.38}$$

By (2.20) and (2.35), we can rewrite the left-hand side as

$$\mathbb{E}\big[(1-\phi)^{\Xi(S^{\varepsilon_n})}\big] = \prod_{k=0}^{\infty}\Big(1 - h(\varepsilon_n k)\mathbb{E}\big[g(\theta_{\varepsilon_n}E^0)\big]\Big).$$

By going to a subsequence, we can assume that

$$G_n := \varepsilon_n^{-1}\mathbb{E}\big[g(\theta_{\varepsilon_n}E^0)\big] \underset{n\to\infty}{\longrightarrow} G \in [0,\infty].$$

We claim that then

$$\mathbb{E}\big[(1-\phi)^{\Xi(S^{\varepsilon_n})}\big] \underset{n\to\infty}{\longrightarrow} e^{-G\int_0^\infty h(t)\mathrm{d}t}.$$

The claim is trivial if $h = 0$, so we assume $h \neq 0$ without loss of generality. We use the concavity of the logarithm and Riemman sum approximation of the integral to estimate

$$\begin{aligned}\log\mathbb{E}\big[(1-\phi)^{\Xi(S^{\varepsilon_n})}\big] &= \sum_{k=0}^{\infty}\log\Big(1-\varepsilon_n G_n h(\varepsilon_n k)\Big)\\ &\leq -G_n\varepsilon_n\sum_{k=0}^{\infty}h(\varepsilon_n k) \underset{n\to\infty}{\longrightarrow} -G\int_0^\infty h(t)\mathrm{d}t.\end{aligned}$$

This already proves the statement when $G = \infty$, so it suffices to prove the other inequality under the assumption that $G < \infty$. Then $\varepsilon_n G_n \to 0$ while $h \leq 1$, so

$$\log\Big(1-\varepsilon_n G_n h(\varepsilon_n k)\Big) = -\varepsilon_n G_n h(\varepsilon_n k) + O(\varepsilon_n^2).$$

Since h is compactly supported, only $O(\varepsilon_n^{-1})$ terms in the sum are nonzero, so the claim follows easily. Using (2.38), we now see that the limit G has to be the same for each subsequence, so for each local continuous function $g : \mathcal{E} \to [0,1]$, there exists a constant $\nu(g) \in [0,\infty]$ such that

$$\varepsilon_n^{-1}\mathbb{E}\big[g(\theta_{\varepsilon_n}E^0)\big] \underset{n\to\infty}{\longrightarrow} \nu(g). \tag{2.39}$$

Formula (2.38) moreover tells us that for any local continuous $g : \mathcal{E} \to [0,1]$ and continuous compactly supported $h : [0,\infty) \to [0,1]$,

$$\mathbb{E}\Big[\prod_{(s,\pi)\in\Xi(B)}\big(1-h(s)g(\pi)\big)\Big] = e^{-\nu(g)\int_0^\infty h(t)\mathrm{d}t}.$$

By Lemma 2.17, the set $\Xi(B)$ is a.s. locally finite, so only finitely many factors in the product are different from one. If $h \leq \frac{1}{2}$, then the product is with positive probability positive, which proves that $\nu(g) < \infty$ for each local continuous $g : \mathcal{E} \to [0,1]$. Combining this with (2.39), we see that there must exist a locally finite measure ν on $\mathcal{E}$ such that

$$\nu(g) = \int_{\mathcal{E}} \nu(\mathrm{d}\pi)\, g(\pi).$$

To complete the proof, it suffices to prove that $\Xi(B)$ is a Poisson point set with intensity measure $\ell \otimes \nu$. By Theorem 2.15, it suffices to show that

$$\mathbb{E}\big[(1-\phi)^{\Xi(B)}\big] = e^{-\int_0^\infty \mathrm{d}s\, \int_{\mathcal{E}} \nu(\mathrm{d}\pi)\phi(s,\pi)}$$

for each local continuous $\phi : [0,\infty) \times \mathcal{E} \to [0,1]$. Our arguments so far already show that this is true for ϕ of the form[4] $\phi(s,\pi) = h(s)g(\pi)$ with local continuous $g : \mathcal{E} \to [0,1]$ and continuous compactly supported $h : [0,\infty) \to [0,1]$. We again use (2.38) and setting $g_s(\pi) := \phi(s,\pi)$, we write

$$\log \mathbb{E}\big[(1-\phi)^{\Xi(S^{\varepsilon_n})}\big] = \sum_{k=0}^{\infty} \log\Big(1 - \mathbb{E}\big[g_{\varepsilon_n k}(\theta_{\varepsilon_n} E^0)\big]\Big),$$

where we can estimate

$$\log\Big(1 - \mathbb{E}\big[g_{\varepsilon_n k}(\theta_{\varepsilon_n} E^0)\big]\Big) = -\varepsilon_n \int \nu(\mathrm{d}\pi) g_{\varepsilon_n k}(\pi) + O(\varepsilon_n^2).$$

The claim then follows from Riemann sum approximation to the integral. ∎

2.7 Limits of excursion sets

In this section we provide the proofs of Lemma 2.17 and Theorem 2.18, which are still missing.

Proof of Lemma 2.17 and Theorem 2.18 The main work is the proof of Theorem 2.18. We will obtain Lemma 2.17 as a side result. If the map $\mathcal{C}_0 \ni f \mapsto \Xi(f)$ were continuous with respect to the sort of convergence we are considering, then the statement of Theorem 2.18 would be trivial. This is not true, but we will show that if B is a Brownian motion, then the map

[4] I actually do not know if this is already enough to conclude that $\Xi(B)$ is a Poisson point set with intensity measure $\ell \otimes \nu$.

$f \mapsto \Xi(f)$ is almost surely continuous in the point $B \in \mathcal{C}_0$, which is all we need.

We will prove the following statement. Assume that $f_n, f \in \mathcal{C}_0$ satisfy $f_n \to f$ locally uniformly, that $\liminf_{t\to\infty} f(t) = -\infty$, and that $\Xi(f)$ is concentrated on $[0,\infty) \times \mathring{\mathcal{E}}$. Then

$$(1-\phi)^{\Xi(f_n)} \underset{n\to\infty}{\longrightarrow} (1-\phi)^{\Xi(f)} \tag{2.40}$$

for all local continuous $\phi : [0,\infty)\times\mathcal{E} \to [0,1]$. Note that this is a deterministic statement: the only way randomness enters our proof is in the fact that if B is a Brownian motion, then almost surely $\liminf_{t\to\infty} B_t = -\infty$ and $\Xi(B)$ is concentrated on $[0,\infty) \times \mathring{\mathcal{E}}$, which follows from Theorem 2.8.

Assume, therefore, that $f_n, f \in \mathcal{C}_0$ satisfy $f_n \to f$ locally uniformly, that $\liminf_{t\to\infty} f(t) = -\infty$, and that $\Xi(f)$ is concentrated on $[0,\infty)\times\mathring{\mathcal{E}}$. Let (g_n, h_n) be defined in terms of f_n as in (2.7) and let (g, h) be similarly defined in terms of f. Let $\mathcal{I}(h)$ denote the set of plateaus of h, i.e., an open intervals of maximal length on which h is constant, and let $\mathcal{I}(h_n)$ be the plateaus of h_n.

Let $\iota = (\iota^-, \iota^+) \in \mathcal{I}(h)$. Our assumption that $\Xi(f)$ is concentrated on $[0,\infty) \times \mathring{\mathcal{E}}$ means that the function g is strictly positive on ι. The locally uniform convergence $g_n \to g$ then implies that for each $\varepsilon > 0$, the function g_n must be strictly positive on $(\iota^- + \varepsilon, \iota^+ - \varepsilon)$ for all n large enough. Since h_n increases only at times when g_n is zero, this then implies that h_n must be constant on $(\iota^- + \varepsilon, \iota^+ - \varepsilon)$.

On the other hand, since ι is a maximal interval on which h is constant, $h(t) < h(\iota^-)$ for all $t < \iota^-$ and $h(\iota^+) < h(t)$ for all $\iota^+ < t$. The locally uniform convergence $h_n \to h$ then implies that for each $\varepsilon > 0$, the function h_n is not constant on $(\iota^- - \varepsilon, \iota^+ + \varepsilon)$ for all n large enough. These arguments show that for each plateau $\iota \in \mathcal{I}(h)$ and for each $\varepsilon \le \varepsilon_0 := (\iota^+ - \iota^-)/3$, there exists an $m(\varepsilon)$ such that for all $n \ge m(\varepsilon)$, there exists a (clearly unique) plateau $\jmath \in \mathcal{I}(h_n)$ with $|\jmath^\pm - \iota^\pm| \le \varepsilon$. For $n \ge m(\varepsilon_0)$, we let $\phi_n(\iota) := \jmath$ denote this plateau, and we define $\phi_n(\iota)$ in an abritrary way for the remaining values of n. Then clearly the left and right boundaries of the plateau $\phi_n(\iota)$ satisfy

$$\phi_n(\iota)^\pm \underset{n\to\infty}{\longrightarrow} \iota^\pm. \tag{2.41}$$

Let $(s,\pi) \in \Xi(f)$ denote the excursion of g corresponding to the plateau ι, and let $\psi_n(s,\pi) \in \Xi(f_n)$ denote the excursion of g_n corresponding to the plateau $\phi_n(\iota)$. Using the fact that $g_n \to g$ and $h_n \to h$ locally uniformly, we see that

$$\psi_n(s,\pi) \underset{n\to\infty}{\longrightarrow} (s,\pi) \tag{2.42}$$

in the topology on $[0,\infty)\times\mathcal{E}$.

For each $S<\infty$ and $\delta>0$, let us set

$$\mathcal{I}_{S,\delta}(h):=\big\{\iota\in\mathcal{I}(h):h(\iota^\pm)<S,\ \iota^+-\iota^->\delta\big\},$$

and define $\mathcal{I}_{S,\delta}(h_n)$ similarly. We claim that for large n, the map ϕ_n is a bijection from $\mathcal{I}_{S,\delta}(h)$ to $\mathcal{I}_{S,\delta}(h_n)$ and hence ψ_n is a bijection from $\Xi_{S,\delta}(f)$ to $\Xi_{S,\delta}(f_n)$. Let $T:=\sup\{t:h(t)<S\}$ and $T_n:=\sup\{t:h_n(t)<S\}$. Then $T<\infty$ by the assumption that $\liminf_{t\to\infty}f(t)=-\infty$ and $T_n\to T$ by the fact that $h_n\to h$ locally uniformly. Since all plateaus $\iota\in\mathcal{I}_{S,\delta}$ are contained in $[0,T]$, the set $\mathcal{I}_{S,\delta}(h)$ can contain at most T/δ elements and is therefore finite. It follows from (2.41) and (2.42) that for large enough n, the map ϕ_n maps the space $\mathcal{I}_{S,\delta}(h)$ into $\mathcal{I}_{S,\delta}(h_n)$. It follows immediately from our definition of ϕ_n that this map is also one-to-one for n large enough.

To see that it is moreover surjective for n large enough, assume that conversely, for infinitely many values of n, there exists a $\jmath_n\in\mathcal{I}_{S,\delta}(h_n)$ that is not the image under ϕ_n of some $\iota\in\mathcal{I}_{S,\delta}(h)$. Since $\jmath_n\subset[0,T_n]$, by going to a subsequence, we can assume that $\jmath_n^\pm\to\jmath^\pm$ for some interval $\jmath$. But then h has to be constant on $\jmath$, which implies that $\jmath\subset\iota$ for some $\iota\in\mathcal{I}(h)$. But this implies that $\jmath_n$ has nonempty intersection with $\phi_n(\iota)$ for all n large enough, which leads to a contradiction.

For $S<\infty$ and $\delta,\varepsilon>0$, let us set

$$\begin{aligned}\Xi_{S,\delta}(f)&:=\big\{(s,\pi)\in\Xi(f):s<S,\ \tau_\pi>\delta\big\},\\ \Xi^\varepsilon_S(f)&:=\big\{(s,\pi)\in\Xi(f):s<S,\ \sup_{0\le t\le\tau_\pi}\pi(t)>\varepsilon\big\}.\end{aligned}$$

There is a one-to-one correspondence between $\mathcal{I}_{S,\delta}(f)$ and $\Xi_{S,\delta}(f)$. We have just proved that the former is finite[5] for each S and δ, and hence the same is true for the latter. We claim that

$$\forall\varepsilon>0\ \exists\delta>0\text{ s.t. }\Xi^\varepsilon_S(f)\subset\Xi_{S,\delta}(f)\qquad(S<\infty).\tag{2.43}$$

As before, let $T:=\sup\{t:h(t)<S\}$. Let $m_{T,\delta}(f)$ be the modulus of continuity defined in (1.12). If there exists a $(s,\pi)\in\Xi_{S,\varepsilon}$ such that $(s,\pi)\notin\Xi_{S,\delta}(f)$, then $m_{T,\delta}(f)>\varepsilon$. Now (2.43) follows from the fact that by Lemma 1.26, for each $\varepsilon>0$, there exists a $\delta>0$ such that $m_{T,\delta}(f)\le\varepsilon$. By the same argument, using the equicontinuity of the functions f_n, which follows from

[5]This part of the argument used that $\liminf_{t\to\infty}f_t=-\infty$ and hence $T:=\sup\{t:h(t)<S\}$ is finite, but it did not need the assumption that $\Xi(f)$ is concentrated om $[0,\infty)\times\mathring{\mathcal{E}}$.

the fact that $f_n \to f$ and The Arzela-Ascoli theorem (Theorem 1.27), we see that

$$\forall \varepsilon > 0\ \exists \delta > 0 \text{ s.t. } \Xi^{\varepsilon}_{S}(f_n) \subset \Xi_{S,\delta}(f_n) \qquad (n \geq 1,\ S < \infty). \tag{2.44}$$

We are finally ready to prove (2.40). Fix a local continuous function $\phi : [0, \infty) \times \mathcal{E} \to [0, 1]$, and let $A := \mathrm{supp}(\phi)$ be its support. We claim that there exist $\delta, \varepsilon > 0$ such that

$$\forall \pi \in A\ \tau_\pi > \delta \text{ or } \sup_{0 \leq t \leq \tau_\pi} \pi_t > \varepsilon. \tag{2.45}$$

Indeed, if (2.45) does not hold, then there exist $\pi_n \in A$ with $\tau_\pi \leq n^{-1}$ and $\sup_{0 \leq t \leq \tau_\pi} \pi_t \leq n^{-1}$. But then $\pi_n \to o$, the trivial excursion, which contradicts the fact that A is closed with $o \notin A$. Using (2.43) and (2.44), we see that there exists a $\delta' > 0$ such that

$$\begin{aligned} \big\{(s, \pi) \in \Xi(f) : s \leq S,\ \pi \in A\big\} &\subset \Xi_{S,\delta'}(f), \\ \big\{(s, \pi) \in \Xi(f_n) : s \leq S,\ \pi \in A\big\} &\subset \Xi_{S,\delta'}(f_n) \qquad (n \geq 1). \end{aligned}$$

Using (2.42) and the fact that for large n, the map ψ_n is a bijection from the finite set $\Xi_{S,\delta'}(f)$ to $\Xi_{S,\delta'}(f_n)$, we see that (2.40) holds. This completes the proof of Theorem 2.18.

Along the way, we have established that if $f \in \mathcal{C}_0$ satisfies $\liminf_{t \to \infty} f_t = -\infty$, then the set

$$\big\{(s, \pi) \in \Xi(f) : s \leq S,\ \pi \in A\big\}$$

is finite for each $S < \infty$ and $A \subset \mathcal{E}$ that is closed with $o \notin A$, proving Lemma 2.17. ∎

2.8 Large random walk excursions

Let $(R, K) = (R_n, K_n)_{n \geq 0}$ be the reflected random walk from Section 2.3 and let $\tau := \inf\{n \geq 0 : K_n < K_{n+1}\}$. Note that since $R_0 = 0 = R_\tau$ and up to time τ, the reflected random walk R steps up or down in each time step, τ is almost surely an even number. The path of the walk up to time τ

$$E^0 = (R_n)_{0 \leq n \leq \tau}$$

is the first of the i.i.d. excursions $(E^k)^{k \in \mathbb{N}}$ of (R, K) discussed in (2.20). In Theorem 2.16, we have seen that the Brownian excursion measure ν describes the tail of the law of E^0, i.e., the small probabilities of very large excursions. In the present section, our aim is to prove the following theorem.

Theorem 2.19 (Scaling limit of large excursion) *Let* $\varepsilon_n := 1/\sqrt{2n}$. *Then*

$$\mathbb{P}\big[\theta_{\varepsilon_n} E^0 \in \cdot \,\big|\, \tau_0 = 2n\big] \underset{n\to\infty}{\Longrightarrow} \nu_1, \tag{2.46}$$

where ν_1 *is the law of the standard Brownian excursion, defined in Proposition 2.14.*

Despite its apparent simplicity, the proof of Theorem 2.19 is quite tricky and we will not completely prove it in this section. We will get quite close, however, and indicate what needs to be done to complete the proof. We want to use excursion theory to prove Theorem 2.19. This may seem natural, but apparently a proof using this approach has been published only fairly recently in [LeG10, Thm 6.1]. That paper is concerned with a class of discrete excursions that is more general than the one we consider, but also a bit different so that Theorem 2.19 is not formally included in [LeG10, Thm 6.1] although it is very similar.

The proof of Theorem 2.19 needs some preparations. We say that a measure ρ on $(0,\infty)$ is *locally finite* if $\rho([s,S]) < \infty$ for all $0 < s < S < \infty$. We say that a sequence of locally finite measures ρ_n on $(0,\infty)$ converges *vaguely* to a limit ρ if

$$\int_0^\infty \rho_n(\mathrm{d}t)\, h(t) \underset{n\to\infty}{\longrightarrow} \int_0^\infty \rho(\mathrm{d}t)\, h(t)$$

for all continuous compactly supported $h : (0,\infty) \to \mathbb{R}$. We postpone the proof of the following simple lemma till later.

Lemma 2.20 (Integrals along paths) *Let* $\pi_n, \pi \in \Pi(\mathbb{R})$ *be paths such that* $I_\pi = [0,\infty)$ *and* $I_{\pi_n} \subset [0,\infty)$ *for all* n. *Let* ρ_n, ρ *be locally finite measures on* $(0,\infty)$ *such that* ρ_n *is concentrated on* I_{π_n} *for each* n. *Assume that* $\pi_n \to \pi$ *in the topology on path space* $\Pi(\mathbb{R})$ *and that the* ρ_n *converge vaguely to* ρ. *Then*

$$\int_0^\infty \rho_n(\mathrm{d}t) h(t) \pi_n(t) \underset{n\to\infty}{\longrightarrow} \int_0^\infty \rho(\mathrm{d}t) h(t) \pi(t)$$

for each continuous compactly supported $h : (0,\infty) \to \mathbb{R}$.

For each $m \in 2\mathbb{N} := \{2n : n \in \mathbb{N}\}$, we let μ_m denote the conditional law

$$\mu_m := \mathbb{P}\big[\theta_{1/\sqrt{m}} E^0 \in \cdot \,\big|\tau_0 = m\big]. \tag{2.47}$$

For any bounded continuous funtion $g : \mathcal{E}_1 \to \mathbb{R}$, we write

$$\langle \mu_m, g \rangle := \int_{\mathcal{E}_1} \mu_m(\mathrm{d}\pi)\, g(\pi).$$

We will need the following technical result, that we will not prove in this chapter.

Proposition 2.21 (Equicontinuity of conditional laws) *Let $g : \mathcal{E}_1 \to \mathbb{R}$ be bounded and continuous. For each $\delta \in (0,1]$, let $\pi^\delta \in \Pi(\mathbb{R})$ be the path defined by*

$$I_{\pi^\delta} := 2\delta^2\mathbb{N} \cap [1,\infty) \quad \textit{and} \quad \pi^\delta(t) := \langle \mu_{\delta^{-2}t}, g\rangle \quad (t \in I_{\pi^\delta}). \tag{2.48}$$

Then the paths $\{\pi^\delta : \delta \in (0,1]\}$ are equicontinuous.

Proof of Theorem 2.19 Let $g : \mathcal{E}_1 \to \mathbb{R}$ be bounded and continuous and let $h : (0,\infty) \to \mathbb{R}$ be continuous and compactly supported. Define $f : \mathcal{E} \to \mathbb{R}$ by

$$f(\pi) := h(\tau_\pi) g(\theta_{1/\sqrt{\tau_\pi}}\pi) \qquad (\pi \neq o),$$

with $f(o) := 0$. Then f is bounded and continuous with $o \notin \mathrm{supp}(f)$, so Theorem 2.16 tells us that

$$\delta^{-1}\mathbb{E}\big[f(\theta_\delta E^0)\big] \underset{\delta\to 0}{\longrightarrow} \int_{\mathcal{E}} \nu(\mathrm{d}\pi)\, f(\pi). \tag{2.49}$$

By Proposition 2.14, we can rewrite the right-hand side of (2.49) as

$$\int_{\mathcal{E}} \nu(\mathrm{d}\pi)\, f(\pi) = \langle \nu_1, g\rangle \int_0^\infty \rho(\mathrm{d}t) h(t), \tag{2.50}$$

where ρ is the measure in (2.29). We rewrite the left-hand side of (2.49) as

$$\begin{aligned}
&\delta^{-1}\mathbb{E}\big[f(\theta_\delta E^0)\big]\\
&\quad = \delta^{-1}\sum_{m\in 2\mathbb{N}} \mathbb{P}\big[\tau_0 = m\big] h\big(\delta^2 m\big)\mathbb{E}\big[g(\theta_{1/\sqrt{m}}\, E^0)\,\big|\,\tau_0 = m\big]\\
&\quad = \delta^{-1}\sum_{m\in 2\mathbb{N}} \mathbb{P}\big[\tau_0 = m\big] h\big(\delta^2 m\big)\langle \mu_m, g\rangle,
\end{aligned}$$

where $\langle \mu_m, g\rangle$ denotes the integral of g with respect to the measure μ_m defined in (2.47). Using the definition

$$\rho_\delta := \delta^{-1}\sum_{n=1}^\infty \mathbb{P}\big[\tau_0 = 2n\big]\delta_{2\delta^2 n},$$

we can rewrite (2.49) as

$$\int_0^\infty \rho_\delta(\mathrm{d}t) h(t)\langle \mu_{\delta^{-2}t}, g\rangle \underset{\delta\to 0}{\longrightarrow} \langle \nu_1, g\rangle \int_0^\infty \rho(\mathrm{d}t) h(t). \tag{2.51}$$

Assume that $\delta_n \in (0,1]$ satisfy $\delta_n \to 0$. Applying (2.51) with g the function that is constantly one and general h, we see that the measures ρ_{δ_n} converge vaguely to ρ as $n \to \infty$. Let

$$t_n := \inf\big(2\delta_n^2\mathbb{N} \cap [1,\infty)\big)$$

and let $\pi_n \in \Pi(\mathbb{R})$ be the path defined by

$$I_{\pi_n} := 2\delta_n^2\mathbb{N} \quad \text{and} \quad \pi_n(t) := \begin{cases} \langle \mu_{\delta^{-2}t}, g\rangle & \text{if } t \geq 1, \\ \langle \mu_{\delta^{-2}t_n}, g\rangle & \text{if } t < 1. \end{cases}$$

By Proposition 2.21, the paths π_n are equicontinuous. Since g is bounded and $\mu_{\delta^{-2}t_n}$ is a probability measure, there exists a compact set $C \subset \mathbb{R}$ such that $\pi_n(t) \in C$ for all n and $t \in I_{\pi_n}$. Therefore, by the Arzela-Ascoli theorem (Theorem 1.27), $\{\pi_n : n \in \mathbb{N}\}$ is a precompact subset of $\Pi(\mathbb{R})$. As a consequence, by Lemma 1.2, to show that the paths π_n converge in the topology on $\Pi(\mathbb{R})$ to a limit π, it suffices to show that all cluster points of the sequence π_n are the same.

Assume that a subsequence $\pi_{n(m)}$ converges as $m \to \infty$ to a limit $\pi \in \Pi(\mathbb{R})$. Then clearly $I_\pi = [0, \infty)$. By Lemma 2.20,

$$\int_0^\infty \rho_{\delta_n}(\mathrm{d}t)h(t)\pi_n(t) \underset{n\to\infty}{\longrightarrow} \int_0^\infty \rho(\mathrm{d}t)h(t)\pi(t) \tag{2.52}$$

for each continuous compactly supported $h : (0, \infty) \to \mathbb{R}$. Since the paths π_n are constant on $[0, 1]$, their limit π must have the same property. If $h : (0, \infty) \to \mathbb{R}$ is continuous and compactly supported with $\mathrm{supp}(h) \subset [1, \infty)$, then combining (2.51) with (2.52) we see that

$$\int_0^\infty \rho(\mathrm{d}t)h(t)\pi(t) = \langle \nu_1, g\rangle \int_0^\infty \rho(\mathrm{d}t)h(t).$$

Since $\pi : [0, \infty) \to \mathbb{R}$ is a continuous function, the measure ρ in (2.29) has a density with respect to the Lebesgue measure, and h is arbitrary, we conclude that $\pi(t) = \langle \nu_1, g\rangle$ for all $t \geq 1$. Since π is constant on $[0, 1]$, this equality extends to $t \geq 0$. This proves that the only cluster point of the sequence π_n is the constant path

$$\pi(t) = \langle \nu_1, g\rangle \qquad (t \geq 0),$$

and hence by Lemma 1.2 $\pi_n \to \pi$ in the topology on $\Pi(\mathbb{R})$. This clearly implies (2.46), so the proof is complete. ∎

We conclude this chapter by providing the proof of Lemma 2.20.

Proof Lemma 2.20 We claim that

$$\sup_{t\in[0,T]\cap I_{\pi_n}} \big|\pi_n(t) - \pi(t)\big| \underset{n\to\infty}{\longrightarrow} 0 \qquad (T < \infty).$$

This can be proved directly by the same sort of arguments as used in the last paragraph of the proof of Lemma 1.22. Alternatively we can extend π_n to

$[0,\infty)$ by linear interpolation and constant extrapolation and use Exercises 1.23 and 1.24 to see that the extended paths also converge to π in the topology on $\Pi(\mathbb{R})$. Then Lemma 1.22 implies that the extended paths converge locally uniformly to π, which implies the claim.

Choosing T large enough such that $\mathrm{supp}(h) \subset [0,T]$ and setting

$$\varepsilon_n := \sup_{t\in[0,T]\cap I_{\pi_n}} \big|\pi_n(t) - \pi(t)\big|,$$

we can now estimate

$$\begin{aligned}
&\Big|\int_0^\infty \rho_n(\mathrm{d}t)h(t)\pi_n(t) - \int_0^\infty \rho(\mathrm{d}t)h(t)\pi(t)\Big| \\
&\quad \leq \varepsilon_n \int_0^\infty \rho_n(\mathrm{d}t)\big|h(t)\big| + \Big|\int_0^\infty \rho_n(\mathrm{d}t)h(t)\pi(t) - \int_0^\infty \rho(\mathrm{d}t)h(t)\pi(t)\Big|.
\end{aligned} \tag{2.53}$$

Here the second term on the right-hand side tends to zero since $t \mapsto h(t)\pi(t)$ is a continuous compactly supported function and $\rho_n \to \rho$ vaguely. Since $t \mapsto |h(t)|$ is also a continuous compactly supported function, we moreover have that

$$\int_0^\infty \rho_n(\mathrm{d}t)\big|h(t)\big| \underset{n\to\infty}{\longrightarrow} \int_0^\infty \rho(\mathrm{d}t)\big|h(t)\big|,$$

which shows in particular that

$$\limsup_{n\to\infty} \int_0^\infty \rho_n(\mathrm{d}t)\big|h(t)\big| < \infty$$

and hence the first term on the right-hand side of (2.53) tends to zero. ∎

Chapter 3

The tree

3.1 Graphs

By definition, a *graph* is a pair $G = (V, E)$ where V is a set and E is another set whose elements are subsets of V containing precisely two elements. A *finite* graph is a graph for which V (and hence also E) are finite. Elements of V are called *vertices* and elements of E are called *edges*. Two vertices v, w are called *adjacent* if $\{v, w\} \in E$. The number of vertices w that are adjacent to v is called the *degree* of the vertex v. A *graph isomorphism* between two graphs $G = (V, E)$ and $G' = (V', E')$ is a bijection $\psi : V \to V'$ such that $\{\psi(v), \psi(w)\} \in E'$ if and only if $\{v, w\} \in E$. If such an isomorphism exists, the graphs are called *isomorphic*. A *subgraph* of $G = (V, E)$ is a graph $G' = (V', E')$ such that $V' \subset V$ and $W' \subset W$.

Two vertices $v, w \in V$ are *disconnected* if there exists a subset $W \subset V$ such that $v \in V \backslash W$, $w \in W$, and $\{v', w'\} \notin E$ for all $v' \in V \backslash W$ and $w' \in W$. Two vertices that are not disconnected are called *connected*. We write $v \leftrightsquigarrow w$ if v is connected to w. It is easy to see that $\leftrightsquigarrow$ is an equivalence relation on V. The equivalence classes are called the *connected components* of G.

A *cycle* is a nonempty finite connected graph in which each vertex has degree precisely two. A *tree* is a nonempty connected graph G that does not contain cycles, i.e., there exists no subgraph G' of G that is a cycle. A *path* is a finite tree in which each vertex has degree at most two.

If $G = (V, E)$ is a path, then we can enumerate the elements of V as $V = \{v_0, \ldots, v_n\}$ with $n \geq 0$ and $v_k \neq v_l$ for all $0 \leq k < l \leq n$, in such a way that $E = \big\{\{v_{k-1}, v_k\} : 1 \leq k \leq n\big\}$. The integer n is called the *length* of the path and v_0 and v_n are called its *endvertices*. If $G = (V, E)$ is an arbitrary graph and $v, w \in V$, then a path *connecting* v and w is a subgraph G' of G such that G' is a path and v and w (which may coincide) are its endvertices.

A *walk* in a graph is an ordered sequence $(v_0, \ldots, v_n)$ of vertices with $n \geq 0$ such that $\{v_{k-1}, v_k\} \in E$ for all $1 \leq k \leq n$. Note that contrary to paths, walks can pass more than once through the same vertex. We call n the *length* and we call v_0 and v_n its *endvertices*.

The *graph distance* $d(v, w)$ between two vertices $v, w \in V$ is the length of the shortest walk connecting v and w if v and w are connected, and $d(v, w) := \infty$ if there does not exist a walk connecting v and w. One can check that d is a metric on V and $d(v, w) < \infty$ if and only if v and w are connected. Each walk of length $d(v, w)$ connecting v and w is actually a path. A graph $G = (V, E)$ is a tree if and only if for each $v, w \in V$, there exists a unique path connecting v and w.

3.2 Random trees

A *rooted* tree is a tree $T = (V, E)$ with one specially marked vertex $\varnothing \in V$, which is called the *root*. Two rooted trees $T = (V, E)$ and $T' = (V', E')$ are called *isomorphic* if there exists a graph isomorphism $\psi : V \to V'$ that preserves the root, i.e., $\psi(\varnothing) = \varnothing$. In a rooted tree, for each $\{v, w\} \in E$, either $d(0, v) = d(0, w) - 1$ or $d(0, v) = d(0, w) + 1$. In the first case, we say that w is a *offspring* of v and in the second case, we say that w is the *parent* of v. Note that parents are unique. Vertices without children are called *leaves*. When we make a picture of T, we draw the root at the bottom and we draw the children of a vertex above the vertex.[1] The children, together with all their children and their children, recursively, are called the *descendants* of a vertex. Similarly, the parent, the parent of the parent, and so on are collectively called the *ancestors* of a vertex.

A *labeled* tree is a triple $T = (V, E, l)$ where (V, E) is a tree and $l : V \to L$ is a one-to-one map that assigns to each vertex $v \in V$ a unique label $l(v) \in L$, where L is some fixed *label set*. Two labeled trees $T = (V, E, l)$ and $T' = (V', E', l)$ are called *isomorphic* if there exists a graph isomorphism $\psi : V \to V'$ that preserves the labels, i.e., $l'(\psi(v)) = l(v)$ for all $v \in V$. The simplest way to attach labels to the vertices of a tree is to simply enumerate them. A tree with vertex set of the form $V = \{0, \ldots, n\}$ is called a *cladogram*. We set

$$\mathcal{T}_n := \big\{T : T \text{ is a cladogram with } n + 1 \text{ vertices}\big\}. \tag{3.1}$$

Naturally, we can view cladograms as rooted trees, with 0 playing the role of the root. *Cayley's formula* says that $|\mathcal{T}_n| = n^{n-1}$.

[1] This is a difference between mathematics and computer science. In computer science, the root of a tree sits at the top and the leaves at the bottom of the tree.

Another natural way of attaching labels to the vertices of a rooted tree is as follows. Let $\mathbb{T}$ denote the space of all finite words $\mathbf{i} = i_1 \cdots i_n$ ($n \in \mathbb{N}$) made up from the alphabet $\mathbb{N}_+ = \{1, 2, \ldots\}$. We denote the length of a word $\mathbf{i} = i_1 \cdots i_n$ by $|\mathbf{i}| := n$ and let $\varnothing$ denote the word of length zero. We define the concatenation $\mathbf{ij}$ of two words $\mathbf{i}, \mathbf{j} \in \mathbb{T}$ with $\mathbf{i} = i_1 \cdots i_m$ and $\mathbf{j} = j_1 \cdots j_n$ by $\mathbf{ij} := i_1 \cdots i_m j_1 \cdots j_n$. A *plane tree* is a nonempty subset $\mathbb{U} \subset \mathbb{T}$ with the following properties:

(i) if $i_1 \cdots i_n \in \mathbb{U}$ and $n \geq 1$, then $i_1 \cdots i_{n-1} \in \mathbb{U}$,

(ii) if $i_1 \cdots i_n \in \mathbb{U}$ and $i_n > 1$, then $i_1 \cdots i_{n-1}(i_m - 1) \in \mathbb{U}$.

For each word $\mathbf{i} = i_1 \cdots i_n \in \mathbb{T}$ with length $n \geq 1$, we write $\overleftarrow{\mathbf{i}} := i_1 \cdots i_{n-1}$. Then condition (i) says that $\overleftarrow{\mathbf{i}} \in \mathbb{U}$ for all $\mathbf{i} \in \mathbb{U} \backslash \{\varnothing\}$. Note that (i) implies that $\varnothing \in \mathbb{U}$. We view $\mathbb{U}$ as a rooted tree tree with root $\varnothing$ and set of edges

$$E := \big\{ \{\overleftarrow{\mathbf{i}}, \mathbf{i}\} : \mathbf{i} \in \mathbb{U} \backslash \{\varnothing\} \big\}. \tag{3.2}$$

Because of condition (ii), for each $\mathbf{i} \in \mathbb{U}$, there is a $\kappa_{\mathbf{i}} \in \mathbb{N}$ such that

$$\mathbf{i}j \in \mathbb{U} \quad \text{if and only if} \quad 1 \leq j \leq \kappa_{\mathbf{i}}. \tag{3.3}$$

When we make a picture of $\mathbb{U}$, above each vertex $\mathbf{i}$, we draw its children $\mathbf{i}1, \ldots, \mathbf{i}\kappa_{\mathbf{i}}$ ordered from left to right. Note that in general, when we draw a rooted tree, there is no prescribed order in which to draw the children of a vertex. Therefore, there are different ways of drawing the same rooted tree in the plane. There is (essentially) only one way of drawing a plane tree in the plane, which explains their name. We set

$$\mathcal{U}_n := \big\{ \mathbb{U} : \mathbb{U} \text{ is a plane tree with } n+1 \text{ vertices} \big\}. \tag{3.4}$$

We will be interested in random rooted trees. A natural way to create a random rooted tree with $n+1$ vertices is to first pick a cladogram at random according to the uniform distribution on the set $\mathcal{T}_n$ of all cladograms with $n+1$ vertices, and then forget about all labels except for the label 0 of the root. Another way is to choose a random plane tree with $n+1$ vertices according to the uniform distribution on $\mathcal{U}_n$, and then again forget about all labels except for the label 0 of the root. It is easy to check (for example for $n = 4$) that these two procedures are not equivalent, i.e., they lead to different distributions on the set of all (non-isomorphic) rooted trees with $n+1$ vertices.

Branching processes also provide a natural way to construct random trees. Let $\rho = (\rho_k)_{k \geq 0}$ be a probability distribution on $\mathbb{N}$, and let $(\kappa_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ be i.i.d. with common law ρ. Then

$$\mathbb{U} := \big\{ i_1 \cdots i_n \in \mathbb{T} : i_k \leq \kappa_{i_1 \cdots i_{k-1}} \ \forall 1 \leq k \leq n \big\} \tag{3.5}$$

defines a random plane tree. We call this the *Galton-Watson tree* with *offspring distribution* $\rho = (\rho_k)_{k\geq 0}$. If $\mathbb{U}$ is such a Galton-Watson tree, then setting

$$X_n := \big|\{\mathbf{i} \in \mathbb{U} : |\mathbf{i}| = n\}\big| \qquad (n \geq 0) \tag{3.6}$$

defines *Galton-Watson branching process* $(X_n)_{n\geq 0}$ with initial state $X_0 = 1$. Let

$$\langle\rho\rangle := \sum_{k=0}^{\infty} \rho_k k \tag{3.7}$$

denote the mean of the offspring distribution. A branching process is called *subcritical* if $\langle\rho\rangle < 1$, *critical* if $\langle\rho\rangle = 1$, and *supercritical* if $\langle\rho\rangle > 1$. It is well-known that, excluding the trivial case that $\rho_1 = 1$, a Galton-Watson tree $\mathbb{U}$ is a.s. finite if and only if the branching process is subcritical or critical.

There is a convenient way of coding plane trees in terms of random walk excursions. By definition, a *contour function* (also called *Dyck path* of length $2n$ is a function $f : \{0, \ldots, 2n\} \to \mathbb{N}$ such that

$$f(0) = f(2n) = 0 \quad \text{and} \quad \big|f(k) - f(k-1)\big| = 1 \forall (0 < k \leq 2n). \tag{3.8}$$

We set

$$\mathcal{D}_n := \big\{f : f \text{ is a contour function of length } 2n\big\}. \tag{3.9}$$

By definition, a *discrete interval* is a set of the form

$$[l : r] := \{l, \ldots, r\} = \{k \in \mathbb{Z} : l \leq k \leq r\} \tag{3.10}$$

with $l, r \in \mathbb{Z}$. Fix $f \in \mathcal{D}_n$. We extend f to $[-1, 2n+1]$ by setting $f(-1) = f(2n+1) := -1$. For each height $h \in \mathbb{N}$, we let

$$\mathcal{I}_h := \big\{[l : r] : f(k) \geq h \forall l \leq k \leq r,\ f(l-1) = h - 1 = f(r+1)\big\} \tag{3.11}$$

be the set of all maximal discrete intervals on which $f \geq h$. We view the set

$$V := \big\{(I, h) : h \geq 0,\ I \in \mathcal{I}_h\big\} \tag{3.12}$$

as the vertex set of a tree $T(f) = (V, E)$ with edge set

$$E := \big\{\{(I, h), (J, h+1)\} : I \in \mathcal{I}_h,\ J \in \mathcal{I}_{h+1},\ J \subset I\big\}. \tag{3.13}$$

We equip $T(f) = (V, E)$ with the structure of a rooted tree by setting $\varnothing := ([0 : 2n], 0)$ the root. The children $(J_1, h+1), \ldots, (J_\kappa, h+1)$ of a vertex $(I, h) \in V$ are naturally ordered from left to right, so we can naturally equip

$T(f)$ with the structure of a plane tree. Let $\mathbb{U}(f)$ denote the resulting plane tree. It is not hard to see (picture!) that the map

$$\mathcal{D}_n \ni f \mapsto \mathbb{U}(f) \in \mathcal{U}_n \tag{3.14}$$

is a bijection, i.e., for each plane tree $\mathbb{U}$ with $n+1$ vertices there exists a a unique contour function f of length $2n$ such that $\mathbb{U} = \mathbb{U}(f)$. We call f the *contour function* of $\mathbb{U}$.

Lemma 3.1 (Geometric offspring distribution) *Let $0 < p < 1$ and let $\mathbb{U}$ be a Galton-Watson tree with offspring distribution $\rho_k = p^k(1-p)$ $(k \geq 0)$. Then*

$$\mathbb{P}\big[\mathbb{U} \in \cdot \,\big|\, |\mathbb{U}| = n+1\big] \tag{3.15}$$

is the uniform distribution on $\mathcal{U}_n$.

Proof Let $X = (X_k)_{k\geq 0}$ be a random walk on $\mathbb{Z}$ with $X_0 = 0$ and transition probabilities

$$\mathbb{P}\big[X_{n+1} = x+1 \,\big|\, X_n = x\big] = p, \quad \mathbb{P}\big[X_{n+1} = x-1 \,\big|\, X_n = x\big] = 1 = p. \tag{3.16}$$

Define a random variable N with values in $\mathbb{N} \cup \{\infty\}$ by

$$2N + 1 := \inf\big\{k \geq 1 : X_k = -1\big\}. \tag{3.17}$$

On the event that $N < \infty$, let F be the random element of the space $\mathcal{D}_N$ of contour functions of length $2N$ defined by

$$F(k) := X_k \qquad (0 \leq k \leq 2N). \tag{3.18}$$

Then

$$\mathbb{P}\big[(N, F) = (n, f)\big] = p^n(1-p)^{n+1} \qquad (n \geq 0,\ f \in \mathcal{D}_n), \tag{3.19}$$

where we have used that up to the time $2N+1$, the random walk X makes N upward jumps and $N+1$ downward jumps.

Now let $\mathbb{U}$ be a Galton-Watson tree with offspring distribution $\rho_k = p^k(1-p)$ $(k \geq 0)$, and let $U \in \mathcal{U}_n$ be a fixed plane tree U with $n+1$ vertices. Let $\kappa_{\mathbf{i}}$ denote the number of children of $\mathbf{i} \in U$. Then using the fact that $\sum_{\mathbf{i}\in U} \kappa_{\mathbf{i}} = n$, we see that

$$\mathbb{P}\big[(N, \mathbb{U}) = (n, U)\big] = \prod_{\mathbf{i}\in U} p^{\kappa_{\mathbf{i}}}(1-p) = p^n(1-p)^{n+1} \qquad (n \geq 0,\ U \in \mathcal{U}_n). \tag{3.20}$$

Comparing this with our previous formula, we see that the conditional law of $\mathbb{U}$ given that $\mathbb{U}$ is finite is equal to the conditional law of $\mathbb{U}(F)$ given that the random variable N defined in (3.17) is finite.

Since the right-hand sides of (3.19) and (3.20) depend only on n and not on f or U, respectively, we see that the conditional law of $\mathbb{U}(F)$ given $N = n$ is the uniform distribution on $\mathcal{U}_n$. ∎

For cladograms, a similar result to Lemma 3.1 is known to hold if the geometric distribution is replaced by a Poisson distribution with mean one.

For each $n \geq 0$, let $\mathbb{U}_n$ be a random plane tree with $n+1$ vertices, chosen according to the uniform distribution on $\mathcal{U}_n$. We will be interested in the shape of the tree $\mathbb{U}_n$ when n is large. In Section **??** below, we will prove that the trees $\mathbb{U}_n$, properly rescaled, converge in distribution to a continuum random tree whose contour function is the standard Brownian excursion. To formulate this properly, in the next sections, we start studying continuum trees.

Bibliography

[BBI01] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry.* Graduate Studies in Mathematics 33. AMS, Providence, 2001.

[Bil99] P. Billingsley. *Convergence of Probability Measures. 2nd ed.* Wiley, New York, 1999.

[Bou58] N. Bourbaki. *Éléments de Mathématique. VIII. Part. 1: Les Structures Fondamentales de l'Analyse. Livre III: Topologie Générale. Chap. 9: Utilisation des Nombres Réels en Topologie Générale. 2iéme éd.* Actualités Scientifiques et Industrielles 1045. Hermann & Cie, Paris, 1958.

[Cho69] G. Choquet. *Lectures on Analysis. Volume I. Integration and Topological Vector Spaces.* Benjamin, London, 1969.

[Dud02] R.M. Dudley. *Real Analysis and Probability.* Reprint of the 1989 edition. Camebridge University Press, Camebridge, 2002.

[EK86] S.N. Ethier and T.G. Kurtz. *Markov Processes; Characterization and Convergence.* John Wiley & Sons, New York, 1986.

[Eng89] R. Engelking. *General Topology.* Heldermann, Berlin, 1989.

[Ito71] K. Itô. Poisson point processes attached to Markov processes. *Proc. Sixth Berkeley Sympos. Math. Statist. Probab.* 3 (1971), 225–240.

[Kal97] O. Kallenberg. *Foundations of modern probability.* Springer, New York, 1997.

[KS91] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus. 2nd ed.* Graduate Texts in Mathematics 113. Springer, New York, 1991.

[LeG10] J.-F. Le Gall. Itô's excursion theory and random trees. *Stochastic Processes Appl.* 120(5) (2010), 721–749.

[McK69] H.P. McKean. *Stochastic integrals.* Academic Press, New York, 1969.

[Mey76] P.A. Meyer. Un cours sur lest intégrales stochastiques. Pages 245–400 in: *Sém. de Probabilités X*, Lecture Notes in Math. 511. Springer, Berlin, 1976.

[Rog89] L.C.G. Rogers. A guided tour through excursions. *Bull. Lond. Math. Soc.* 21(4) (1989), 305–341.

[RW87] L.C.G. Rogers and D. Williams. *Diffusions, Markoc processes, and martingales*, Vol. 2. Wiley, Chichester, 1987.

[SSS14] E. Schertzer, R. Sun, and J.M. Swart. Stochastic flows in the Brownian web and net. *Mem. Am. Math. Soc.* Vol. 227 (2014), Nr. 1065.

[Oxt80] J.C. Oxtoby. *Measure and Category. Second Edition.* Springer, New York, 1980.

Index