Anton Bovier

Stochastic Processes

Lecture, Summer term 2013, Bonn

November 17, 2013

Contents

Chapter 1
A review of measure theory

In this first chapter I review the main concepts of measure theory that we will need. I will not give proofs in most cases. Those familiar with my W-Theorie 1 lecture will find that most of this material was covered there, except that here we will take a somewhat more abstract point of view, replacing the space of real numbers by arbitrary metric spaces. One will see, however, that this implies very few changes. For more details, there is a wealth of references on measure theory. See e.g. [2, 15, 11, 8, 5, 1].

1.1 Probability spaces

A *space*, Ω, is an arbitrary non-empty set. Elements of a space Ω will be denoted by ω. If $A \subset \Omega$ is a subset of Ω, we denote by $\mathbb{1}_A$ the indicator function of the set A, i.e.

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \in A^c \equiv \Omega \backslash A. \end{cases} \tag{1.1.1}$$

Definition 1.1. Let Ω be a space. A family $\mathfrak{A} \equiv \{A_\lambda\}_{\lambda \in I}$, $A_\lambda \subset \Omega$, with I an arbitrary set, is called a *class* of Ω. A non-empty class of Ω is called an *algebra*, if:

(i) $\Omega \in \mathfrak{A}$.
(ii) For all $A \in \mathfrak{A}$, $A^c \in \mathfrak{A}$.
(iii) For all $A, B \in \mathfrak{A}$, $A \cup B \in \mathfrak{A}$.

If $\mathfrak{A}$ is an algebra, and moreover

(iv) $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$, whenever for all $n \in \mathbb{N}$, $A_n \in \mathfrak{A}$,

then $\mathfrak{A}$ is called a σ-*algebra*.

Definition 1.2. A space, Ω, together with a σ-algebra, $\mathfrak{F}$, of subsets of Ω is called a *measurable space*, $(\Omega, \mathfrak{F})$.

Definition 1.3. Let $(\Omega, \mathfrak{F})$ be a measurable space. A map $\mu : \mathfrak{F} \to [0, \infty]$ from $\mathfrak{F}$ the non-negative real numbers (and infinity) is called a (positive) *measure*, if

(i) $\mu(\emptyset) = 0$.
(ii)For any countable family $\{A_n\}_{n\in\mathbb{N}}$ of mutually disjoint elements of $\mathfrak{F}$,

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n). \tag{1.1.2}$$

A measure, μ, is called *finite*, if $\mu(\Omega) < \infty$. A measure is called σ-finite, if there exists a countable class, Ω_n, of subsets of Ω, such that $\Omega = \bigcup_{n\in\mathbb{N}} \Omega_n$, such that, for all $n \in \mathbb{N}$, $\mu(\Omega_n) < \infty$.
A triple, $(\Omega, \mathfrak{F}, \mu)$, is called a *measure space*.

Definition 1.4. Let $(\Omega, \mathfrak{F})$ be a measurable space. A positive measure, $\mathbb{P}$, on $(\Omega, \mathfrak{F})$ that satisfies $\mathbb{P}[\Omega] = 1$ is called a *probability measure*. A triple $(\Omega, \mathfrak{F}, \mathbb{P})$, where Ω is a set, $\mathfrak{F}$ a σ-algebra of subsets of Ω, and $\mathbb{P}$ a probability measure on $(\Omega, \mathfrak{F})$, is called a *probability space*.

Probability spaces provide the scenery where probability theory takes place. The set of sceneries is huge, since we have so far not made any restriction on the allowable spaces Ω. In most instances, we will, however, want to stay on reasonable grounds. Fortunately, where is a quite canonical setting where everything we ever want to do can be constructed. This is the realm where Ω is a topological space and $\mathfrak{F} = \mathfrak{B}(\Omega)$ is the Borel-σ-algebra of Ω.

We recall the definition of a topological space.

Definition 1.5. A space, E, is called a *topological space*, if for every point $p \in E$ there exists a collection, $\mathscr{U}_p$, of subsets of E, called (open) *neighborhoods*, with the following properties:

(i) For every point, p, $\mathscr{U}_p \neq \emptyset$.
(ii)Every neighborhood of p contains p.
(iii)If $U_1, U_2 \in \mathscr{U}_p$, then there exists $U_3 \in \mathscr{U}_p$ such that $U_3 \subset U_1 \cap U_2$.
(iv)If $U \in \mathscr{U}_p$ and $q \in U$, then there exists $V \in \mathscr{U}_q$ such that $V \subset U$.

Recall that in a topological space, one can define the notions such *open sets* and *closed sets*; open sets have the property that any of its points has a neighborhood that is contained in the set, and closed sets are the complements of open sets. Note that the empty set is also considered as an open set by default. Since the entire space E is also open, the empty set is, however, also a closed set.

Definition 1.6. Two topological spaces are considered *equivalent*, or have the same *topology*, if they contain the same open sets. In particular, given two sets of collections of neighborhoods, $\mathscr{U}_p$, and $\mathscr{V}_p$, on a space E, then they generate the same topology[1], if for any $p \in E$, and any $U \in \mathscr{U}_p$, there exists $V \in \mathscr{V}_p$ such that $V \subset U$ and for any $V \in \mathscr{V}_p$ there exists $U \in \mathscr{U}_p$ such that $U \subset V$.

Definition 1.7. A topological space, E, is called:

(i) *Hausdorff*, if any two distinct points in E have disjoint neighborhoods.
(ii)*separable*, if there exists a countable subset, $E_0 \subset E$ whose closure[2] is E.

Definition 1.8. Let E be a topological space. The *Borel-σ-algebra*, $\mathfrak{B}(E)$ of E is the smallest σ-Algebra that contains all open sets of E.

The point behind the notion of the Borel-σ-algebra is that it is big enough to satisfy our needs, but small enough to ensure that it is possible to construct measures on it. Larger σ-algebras, such as the power set on any uncountable topological space, do not usually allow to define measures with nice properties on them.

One says that the Borel-σ-algebra is *generated* by the open sets of E. This notion will be used quite frequently. We say in general that a class, $\mathfrak{A}$, of a space Ω generates a σ-algebra, $\sigma(\mathfrak{A})$, defined as the smallest σ-algebra that contains $\mathfrak{A}$,

$$\sigma(\mathfrak{A}) \equiv \bigcap_{\substack{\mathfrak{F} \supset \mathfrak{A} \\ \mathfrak{F} \text{ is } \sigma-\text{algebra}}} \mathfrak{F}.$$

Even more structure appears if we work on so-called *metric spaces*.

Definition 1.9. Let E be a set. A map, $\rho : M \times M \to [0,\infty]$, is called a *metric*, if

(i) $\rho(x,y) = 0$, if and only if $x = y$;
(ii)$\rho(x,y) = \rho(y,x)$;
(iii)$\rho(x,z) \leq \rho(x,y) + \rho(y,z)$, for all $x,y,z \in E$.

The set $B_r(x) \equiv \{y \in E : \rho(x,y) < r\}$ is called the (open) ball of radius r.

The set of neighborhoods obtained from the open balls associated to a metric, ρ, is called the *metric topology*. A topological space endowed with a metric and its metric topology is called a *metric space*.

A sequence $x_n \in E$, $n \in \mathbb{N}$ is called a *Cauchy sequence*, if for any $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$, such that for all $n,m \geq n_0$, $\rho(x_n,x_m) < \varepsilon$. A metric space, E, is called *complete*, if any Cauchy sequence in E converges.

A related concept is that of a *normed space*.

Definition 1.10. Let E be a vector space. A map $\|\cdot\| : E \to \mathbb{R}_+$ is called a *norm*, if

(i) For all $x \in E$, $\|x\| \geq 0$, and $\|x\| = 0$ iff $x = 0$;
(ii)for all $x \in E$ and $\alpha \in \mathbb{R}$, $\|\alpha x\| = |\alpha|\|x\|$;
(iii)for any $x,y \in E$, $\|x+y\| \leq \|x\| + \|y\|$;

---

[1] One says that the collections of neighborhoods $B = \{\mathscr{U}_p, p \in E\}$ *generate* a topology $\mathscr{T}$, or it is a *base* for a topology $\mathscr{T}$, if every open set in $\mathscr{T}$ can be written as a union of elements of B

[2] The closure of a subset, A, of a topological space is the intersection of all closed subsets containing A.

A vector space equipped with a norm is called a *normed (vector) space*.

Defining $\rho(x,y) \equiv \|x-y\|$ yields a metric, so every normed space can be turned into a metric space. A normed vector space that is a complete metric space with respect to this norm is called a *Banach space*.

A further useful specialisation is the restriction to so called Polish spaces.

Definition 1.11. A topological space E is called *Polish* if it is *separable*, and *completely metrisable spaces*. A completely metrisable space is a space that it homeomorphic to a complete metric space.

That is, a Polish space is essentially a complete, separable metric space up to the fact that the metric may not have been fixed. Recall that $\mathbb{R}^d$ is a Polish space, and so is $\mathbb{R}^{\mathbb{N}}$ when equipped with the product topology.

Note that in many cases, different families of sets generate the same σ-algebra. For instance, if E is not only a topological space, but a metric space with the topology given by the metric topology, then the set of open balls generates the Borel-σ-algebra $\mathfrak{B}(E)$. But also the set of closed balls will generate $\mathfrak{B}(E)$. If E is the real line, the half-lines also generate the Borel-σ-algebra.

An advantage in Ω being a Polish space lies in the fact that one can choose as a generator of the Borel-σ-algebra a *countable* collection of sets. For example, in the case of the real line, the Borel-σ-algebra is already generated by the half-lines $(-\infty,q]$, with $q \in \mathbb{Q}$ (just observe that if x is any real number, there exists a sequence $q_n \downarrow x$, and the set $\bigcap_{n\in\mathbb{N}}(-\infty,q_n] = (-\infty,x]$ is also contained in the σ-algebra generated by these half-lines.

A related, but more general class of spaces are sometimes useful. These are called *Lousin spaces.* These are spaces that are homeomorphic to a Borel subset of a compact metric space.

Two notions of special types of classes are very useful in this context.

Definition 1.12. Let Ω be a space. A class of Ω, $\mathscr{T}$, is called a Π-system, if $\mathscr{T}$ is closed under finite intersections; a class, $\mathfrak{G}$, is called a λ-system, if

(i) $\Omega \in \mathfrak{G}$,
(ii) If $A,B \in \mathfrak{G}$, and $A \supset B$, then $A \setminus B \in \mathfrak{G}$,
(iii) If $A_n \in \mathfrak{G}$ and $A_n \subset A_{n+1}$, then $\lim_{n\to\infty} A_n \in \mathfrak{G}$.

The following useful observation is called Dynkin's theorem.

Theorem 1.13. *If $\mathscr{T}$ is a Π-system and $\mathfrak{G}$ is a λ-system, then $\mathfrak{G} \supset \mathscr{T}$ implies that $\mathfrak{G}$ contains the smallest σ-algebra containing $\mathscr{T}$.*

The most useful application of Dynkin's theorem is the observation that, if two probability measures are equal on a Π-system that generates the σ-algebra, then they are equal on the σ-algebra (since the set on which the two measures coincide forms a λ-system containing $\mathscr{T}$).

Examples.

The general setup allows allows to treat many important examples on the same footing.

Countable spaces. If Ω is a countable space, the natural topology is the *discrete topology*. Here the set of neighborhoods of a point p is just the set $\{p\}$. Clearly this is a topology, and all sets are open and closed with respect to it. The Borel-σ-algebra consists of the power set of Ω. Countable spaces equipped with the *discrete metric* defined by $\rho(x,y) = \mathbb{1}_{x\neq y}$ is a metric space.

Euclidean space. $\mathbb{R}^d$ equipped with the Euclidean metric $\rho(x,y) \equiv \|x-y\|$ is a metric space. Choosing as sets of neighborhoods the set of all open balls, $B_r(p) \equiv \{x \in R^d : \|x-p\| < r\}$ turns this into a topological space. The corresponding Borel-σ-algebra is the smallest σ-algebra containing all these balls.

Note that, since on $\mathbb{R}^d$ the Euclidean norm and the sup-norm are equivalent, the Borel-σ-algebra is also generated by open (or closed) rectangles.

Infinite product spaces. If E is a topological space, then the infinite Cartesian product space, E^∞, can also be turned into a topological space through the *product topology*. Here the set of neighborhoods of a point $p \equiv (p_1, p_2, p_3, \dots)$ is given by the collection of sets

$$U_{p_1} \times U_{p_2} \times U_{p_k} \times E \times E \times \dots, \tag{1.1.3}$$

where $k \in \mathbb{N}$, and $U_{p_i} \in \mathscr{U}_{p_i}$. If $\mathfrak{B}(E)$ is the Borel-σ-algebra of E, then the Borel-σ-algebra of E^∞ is the product σ-algebra, $\mathfrak{B}(E^\infty) = \mathfrak{B}(E)^{\otimes\infty}$, i.e. the σ-algebra that is generated by the family of sets $A_1 \times \cdots \times A_k \times E \times \dots$, $k \in \mathbb{N}$, $A_i \in \mathfrak{B}(E)$ (where of course it also suffices to choose the sets $E \times \cdots \times E \times A_k \times E \times \dots$, $k \in \mathbb{N}$, and A_k running through a generator of $\mathfrak{B}(E)$).

If E is a metric space, then one can also turn E^∞ into a metric space, such that the associated metric topology is equivalent to the product topology. This is done, e.g. by setting

$$\rho_{E^\infty}(p,q) \equiv \sum_{n=1}^{\infty} 2^{-n} \frac{\rho_E(p_n,q_n)}{1+\rho_E(p_n,q_n)}. \tag{1.1.4}$$

Note that this implies that, if E is a Polish space, then the infinite product space E^∞ equipped with the product topology is also Polish.

Infinite product spaces will be the scenario to discuss stochastic processes with discrete time, the main topic of this course.

Function spaces. Important examples of metric spaces are normed function spaces, such as the space of bounded, real-valued functions on $\mathbb{R}$ (or subsets $I \subset \mathbb{R}$), equipped with the supremum norm

$$\|f-g\|_\infty \equiv \sup_{t\in I} |f(t)-g(t)|. \tag{1.1.5}$$

In the case when I is infinite (e.g. $I = \mathbb{R}_+$), we will often use a weaker topology that “ignores infinity”, called the topology of “uniform convergence on finite subsets”.

It can be metrized with a norm

$$\|f-g\| \equiv \sum_{n=1}^{\infty} 2^{-n} \frac{\sup_{0\leq t\leq n} |f(t)-g(t)|}{1+\sup_{0\leq t\leq n} |f(t)-g(t)|}. \tag{1.1.6}$$

We will begin to deal with such examples in the later parts of this course, when we introduce Gaussian random processes with continuous time.

Spaces of measures. Another space we are often encountering in probability theory is that of *measures* on a Borel-σ-algebra. There are various ways to introduce topologies on spaces of measures, but a very common one is the so-called *weak topology*. Let E be the topological space in question, and $C_0(E,\mathbb{R})$ the space of real-valued, bounded, and continuous functions on E. We denote by $\mathscr{M}_+(E,\mathfrak{B}(E))$ the set of all positive measures on $(E,\mathfrak{B}(E))$. One can then define neighborhoods of a measure μ of the form

$$B_{\varepsilon,k,f_1,\dots,f_k}(\mu) \equiv \left\{ \nu \in \mathscr{M}_+(E,\mathfrak{B}(E)) : \max_{i=1}^{k} |\mu(f_i)-\nu(f_i)| < \varepsilon \right\}, \tag{1.1.7}$$

where $\varepsilon > 0$, $k \in \mathbb{N}$, and $f_i \in C_0(E,\mathbb{R})$.

If E is a Polish space, then the weak topology can also be derived from a suitably defined metric.

1.2 Construction of measures

The problem of the construction of measures in the general context of topological spaces is not entirely trivial. This is due to the richness of a Borel-σ-algebra and the hidden subtlety associated with the requirement of σ-additivity. The general strategy is to construct a "measure" first on a simpler set, an algebra or a semi-algebra, and then to use a powerful theorem ensuring the unique extendibility to the σ-algebra.

To do this we first define the notion of a σ-additive set-function.

Definition 1.14. Let $\mathfrak{A}$ be a class of subset of some set Ω. A function $\nu : \mathfrak{A} \to [0,\infty]$ is called a positive, σ-additive (or countably additive) set-function, if

(i) $\nu(\emptyset) = 0$,
(ii)for any sequence, A_k, $k \in \mathbb{N}$, of mutually disjoint elements of $\mathfrak{A}$ such that $\bigcup_{k\in\mathbb{N}} A_k \in \mathfrak{A}$,

$$\nu\left(\bigcup_{k\in\mathbb{N}} A_k\right) = \sum_{k\in\mathbb{N}} \nu(A_k). \tag{1.2.1}$$

The aim of this section is to prove the following version of Carathéodory's theorem.

Theorem 1.15 (Carathéodory's theorem). *Let Ω be a set and let $\mathscr{S}$ be an algebra on Ω. Let μ_0 be a countably additive function $\mathscr{S} \to [0,\infty]$. Then there exists a* *measure, μ, on $(\Omega, \sigma(\mathscr{S}))$, such that $\mu = \mu_0$ on $\mathscr{S}$. If μ_0 is σ-finite, then μ is unique.*

Proof. We begin by defining the notion of an *outer measure*.

Definition 1.16. Let Ω be a set. A map $\mu^* : \mathscr{P}(\Omega) \to [0,\infty]$ is called an outer measure if,

(i) $\mu^*(\emptyset) = 0$;
(ii)If $A \subset B$, then $\mu^*(A) \leq \mu^*(B)$ (increasing);
(iii)for any sequence $A_n \in \mathscr{P}(\Omega)$, $n \in \mathbb{N}$,

$$\mu^*\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mu^*(A_n) \tag{1.2.2}$$

(σ-sub-additivity).

Note that an outer measure is far less constraint than a measure; this is why it can be defined on any set, not just on σ-algebras.
Example. If $(\Omega, \mathfrak{F}, \mu)$ is a measure space, we can define an extension of μ that will be an outer measure on $\mathscr{P}(\Omega)$ as follows: For any $D \subset \Omega$, let

$$\mu^*(D) \equiv \inf\{\mu(F) : F \in \mathfrak{F}; F \supset D\}. \tag{1.2.3}$$

This is of course not how we want to proceed when constructing a measure. Rather, we will construct an outer measure from a σ-additive function on an algebra (that is also a Π-system), and then use this to construct a measure.

Next we define the notion of μ^*-measurability of sets.

Definition 1.17. A subset $B \subset \Omega$ is called μ^*-measurable, if, for all subsets $A \subset \Omega$,

$$\mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c). \tag{1.2.4}$$

The set of μ^*-measurable sets is called $\mathscr{M}(\mu^*)$.

Theorem 1.18.

(i) $\mathscr{M}(\mu^*)$ is a σ-algebra that contains all subsets $B \subset \Omega$ such that $\mu^*(B) = 0$.
(ii)The restriction of μ^* to $\mathscr{M}(\mu^*)$ is a measure.

Proof. Note first that in general, by sub-additivity,

$$\mu^*(A) \leq \mu^*(A \cap B) + \mu^*(A \cap B^c). \tag{1.2.5}$$

If $\mu^*(B) = 0$, we have also that

$$\mu^*(A) \geq \mu^*(A \cap B^c) = \mu^*(A \cap B) + \mu^*(A \cap B^c). \tag{1.2.6}$$

Thus, $\mathscr{M}(\mu^*)$ contains all sets B with $\mu^*(B) = 0$. This implies in particular that $\emptyset \in \mathscr{M}(\mu^*)$. Also, by the symmetry of the definition, $\mathscr{M}(\mu^*)$ contains all its sets together with their complements. Thus the only non-trivial thing to show (i) is the stability under countable unions. Let B_1, B_2 be in $\mathscr{M}(\mu^*)$. Then

$$\begin{aligned}\mu^*(A\cap(B_1\cup B_2)) &= \mu^*(A\cap(B_1\cup B_2)\cap B_1)+\mu^*(A\cap(B_1\cup B_2)\cap B_1^c)\\ &= \mu^*(A\cap B_1)+\mu^*(A\cap B_2\cap B_1^c),\end{aligned} \tag{1.2.7}$$

where we used that $B_1 \in \mathscr{M}(\mu^*)$ for the first equality. Then

$$\begin{aligned}&\mu^*(A\cap(B_1\cup B_2))+\mu^*(A\cap(B_1\cup B_2)^c)\\ &= \mu^*(A\cap B_1)+\mu^*(A\cap B_2\cap B_1^c)+\mu^*(A\cap B_1^c\cap B_2^c)\\ &= \mu^*(A\cap B_1)+\mu^*(A\cap B_1^c)=\mu^*(A).\end{aligned} \tag{1.2.8}$$

Thus $B_1\cup B_2 \in \mathscr{M}(\mu^*)$. This implies that $\mathscr{M}(\mu^*)$ is closed under finite union. Since it is also closed under passage to the complement, it is closed under finite intersection. Thus it is enough to show that countable unions of pairwise disjoint sets, $B_k \in \mathscr{M}(\mu^*)$, $k\in\mathbb{N}$, are in $\mathscr{M}(\mu^*)$. To show this, we show that, for all $m\in\mathbb{N}$,

$$\mu(A)=\sum_{n=1}^m \mu^*(A\cap B_n)+\mu^*\left(A\cap\bigcap_{n=1}^m B_n^c\right). \tag{1.2.9}$$

This holds for $m=1$ by definition, and if it holds for m, then

$$\begin{aligned}\mu^*\left(A\cap\bigcap_{n=1}^m B_n^c\right) &= \mu^*\left(A\cap\bigcap_{n=1}^m B_n^c\cap B_{m+1}\right)+\mu^*\left(A\cap\bigcap_{n=1}^{m+1} B_n^c\right)\\ &= \mu^*\left(A\cap B_{m+1}\right)+\mu^*\left(A\cap\bigcap_{n=1}^{m+1} B_n^c\right),\end{aligned}$$

so, inserting this into (1.2.9), it holds for $m+1$. Hence, by induction, it is true for all $m\in\mathbb{N}$.

From (1.2.9) we deduce further that

$$\mu(A)\geq\sum_{n=1}^m \mu^*(A\cap B_n)+\mu^*\left(A\cap\bigcap_{n=1}^\infty B_n^c\right). \tag{1.2.10}$$

Now we let m tend to infinity, and use sub-additivity:

$$\begin{aligned}\mu(A) &\geq \sum_{n=1}^\infty \mu^*(A\cap B_n)+\mu^*\left(A\cap\bigcap_{n=1}^\infty B_n^c\right)\\ &\geq \mu^*\left(A\cap\left(\bigcup_{n=1}^\infty B_n\right)\right)+\mu^*\left(A\cap\bigcap_{n=1}^\infty B_n^c\right).\end{aligned} \tag{1.2.11}$$

Since the converse inequality follows by sub-additivity, equality holds in (1.2.11) and thus the union $\bigcup_{n=1}^\infty B_n \in \mathscr{M}(\mu^*)$.

It remains to prove that μ^* restricted to $\mathscr{M}(\mu^*)$ is a measure. We know already that $\mu^*(\emptyset) = 0$. Let now B_n be disjoint as above. Let us choose in the first line of (1.2.11) $A = \bigcup_{n=1}^{\infty} B_n$. This gives

$$\mu^*\left(\bigcup_{n=1}^{\infty} B_n\right) \geq \sum_{n=1}^{\infty} \mu^*(B_n). \tag{1.2.12}$$

Since the converse inequality holds by sub-additivity, equality holds and the result is proven. $\square$

The preceding theorem provides a clear strategy for proving Carathéodory's theorem. All we need is to prescribe a σ-additive function, μ, on the algebra. Then construct an outer measure μ^*. This can be done in the following way: If $\mathscr{S}$ is an algebra, set

$$\mu^*(D) = \inf\{\mu(A) : A \in \mathscr{S}; A \supset D\} \tag{1.2.13}$$

One needs to show that this is sub-additive and defines an outer measure. Once this is done, it remains to show that $\mathscr{M}(\mu^*)$ contains $\sigma(\mathscr{S})$. This is done by showing that it contains $\mathscr{S}$, since $\mathscr{M}(\mu^*)$ is a σ-algebra.

Let us now conclude our proof by carrying out these steps.

Lemma 1.19. *Let $\mathscr{S}$ be an algebra, μ a σ-additive function on $\mathscr{S}$, and μ^* defined by (1.2.13). Then μ^* is an outer measure.*

Proof. First, note that the first two conditions for μ^* to be an outer measure are trivially satisfied. To prove sub-additivity, let A_n, $n \in \mathbb{N}$ be a family of subsets of Ω. For each n, we can choose sets $F_n \in \mathscr{S}$, such that $A_n \subset F_n$, and $\mu(F_n) \leq \mu^*(A_n) + \varepsilon 2^{-n}$, for any $\varepsilon > 0$ (since $\mu^*(A_n) = \inf\{\mu(F) : F \in \mathscr{S}; F \supset A_n\}$. Then, since $\bigcup_n F_n \supset \bigcup_n A_n$, and μ is σ-additive,

$$\mu^*\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \mu\left(\bigcup_{n\in\mathbb{N}} F_n\right) \leq \sum_{n\in\mathbb{N}} \mu(F_n) \leq \sum_{n\in\mathbb{N}} \mu^*(A_n) + 2\varepsilon, \tag{1.2.14}$$

which proves the claim since $\varepsilon > 0$ is arbitrary. $\square$

Lemma 1.20. *Let μ^* be the outer measure defined by (1.2.13). Let $\mathscr{M}(\mu^*)$ be the σ-algebra of μ^*-measurable sets. Then $\sigma(\mathscr{S}) \subset \mathscr{M}(\mu^*)$.*

Proof. We must show that $\mathscr{M}(\mu^*)$ contains a family that generates $\sigma(\mathscr{S})$. In fact, we will show that it contains all the elements of the algebra $\mathscr{S}$. To see this, let $A \subset \Omega$ be arbitrary. Then (if $\mu^*(A) < \infty$), for any $\varepsilon > 0$, there is a set $F \in \mathscr{S}$, such that $A \subset F$ and $\mu^*(A) \geq \mu(F) - \varepsilon$. But then, for $B \in \mathscr{S}$,

$$\mu^*(A \cap B) \leq \mu(F \cap B) \tag{1.2.15}$$

and also

$$\mu^*(A \cap B^c) \leq \mu(F \cap B^c) \tag{1.2.16}$$

But the two sets on the right-hand sides are disjoint and in $\mathscr{S}$. Thus

$$\mu^*(A\cap B)+\mu^*(A\cap B^c)\leq \mu(F\cap B)+\mu(F\cap B^c)=\mu(F)\leq \mu^*(A)+\varepsilon. \quad (1.2.17)$$

This proves

$$\mu^*(A)\geq \mu^*(A\cap B)+\mu^*(A\cap B^c) \quad (1.2.18)$$

and since the opposite inequality follows by sub-additivity, $B\in \mathscr{M}(\mu^*)$. □

Thus we have in fact constructed an outer measure that is a measure on $\sigma(\mathscr{S})$ and that extends μ on $\mathscr{S}$. The uniqueness of the extension in the finite case follows from Dynkin's theorem. Assume that there are two extensions, μ and ν that coincide on $\mathscr{S}$. One verifies easily that the class of sets where $\mu(B)=\nu(B)$ is a λ-system which contains the Π-system $\mathscr{S}$; by Dynkin's theorem this λ-system must be $\sigma(\mathscr{S})$. Finally, if μ is σ-finite, one uses the following standard argument (that allows to carry many results from finite to σ-finite measures): By σ-finiteness, there exists a sequence of increasing sets, $\Omega_n\uparrow\Omega$, with $\mu(\Omega_n)<\infty$. Then the measure $\mu_n\equiv\mu\mathbb{1}_{\Omega_n}\uparrow\mu$. So if there are two extensions of a given σ-additive set-function, then their restrictions to all Ω_n are finite measures and must coincide. But then so must their limits. This concludes the proof of Carathéodory's theorem. □

Remark. Carathéodory's theorem should appear rather striking at first by its generality. It makes no assumptions on the nature of the space Ω whatsoever. Does this mean that the construction of a measure is in general trivial? The answer is of course no, but Caratherodory's theorem separates clearly the topological aspects form the algebraic aspects of measure theory. Namely, it shows that in a concrete situation, to construct a measure one needs to construct a σ-additive set-function on an algebra that contains a Π-system that will generate the desired σ-algebra. The proof of Carathéodory's theorem shows that the extension to a measure is essentially a matter of algebra and completely general. We will see later how topological aspects enter into the construction of additive set-functions, and why aspects like separability and metric topologies become relevant.

Remark. The σ-algebra $\mathscr{M}(\mu^*)$ is in general not equal to the σ-algebra generated by $\mathscr{S}$. In particular, we have seen that $\mathscr{M}(\mu^*)$ contains all sets of μ^*-measure zero, all of which need not be in $\sigma(\mathscr{S})$. This observation suggests to consider in general extensions of a given σ-algebra with respect to a measure that ensures that all sets of measure zero are measurable. Let $(\Omega,\mathfrak{F},\mu)$ be a measure space. Define the outer measure, μ^*, as in (1.2.3), and define the inner measure, μ_*, as

$$\mu_*(D)\equiv\sup\{\mu(F):F\in\mathfrak{F};F\subset D\}. \quad (1.2.19)$$

Then

$$\mathscr{M}(\mu)\equiv\{A\subset\Omega:\mu_*(A)=\mu^*(A)\}. \quad (1.2.20)$$

One can easily check that $\mathscr{M}(\mu)$ is a σ-algebra that contains $\mathfrak{F}$ and all sets of outer measure zero.

Terminology. A measure, μ, defined on a Borel-σ-algebra $\mathfrak{F} = \mathfrak{B}(\Omega)$ is sometimes called a *Borel measure*. The measure space $(\Omega, \mathscr{M}(\mu), \mu)$ is called the *completion* of $(\Omega, \mathfrak{F}, \mu)$.

It is a nice feature of null-sets that not only can they be added, but they can also be gotten rid off. This is the content of the next lemma.

Lemma 1.21. *Let $(\Omega, \mathfrak{F}, \mu)$ be a probability space and assume that $G \subset \Omega$ is such that $\mu^*(G) = 1$. Then for any $A \in \mathfrak{F}$, $\mu^*(G \cap A) = \mu(A)$ and if $\mathfrak{G} \equiv \mathfrak{F} \cap G$ (that is the set of all subsets of G of the form $G \cap A$, $A \in \mathfrak{F}$), then $(G, \mathfrak{G}, \mu^*)$ is a probability space.*

Proof. Exercise. □

Lebesgue measure. The prime example of the construction of a measure using Carathéodory's theorem ist the Lebesgue measure on $\mathbb{R}$. Consider the algebra, $\mathscr{S}$, of all sets that can be written as finite unions of semi-open, disjoint intervals of the form $(a,b]$, and $(a,+\infty)$, $a \in \mathbb{R} \cup \{-\infty\}$, $b \in \mathbb{R}$. Clearly, the function λ, defined by

$$\lambda\left(\bigcup_i (a_i, b_i]\right) = \sum_i (b_i - a_i) \tag{1.2.21}$$

provides a countably additive set-function (this needs a proof!!). Then we know that this can be extended to $\sigma(\mathscr{S}) = \mathfrak{B}(\Omega)$; more precisely, one actually constructs a measure on the σ-algebra $\mathscr{M}(\lambda^*)$, and strictly speaking it is this measure on the complete measure space $(\mathbb{R}, \mathscr{M}(\lambda), \lambda)$ that is called the Lebesgue measure.

Of course the same construction can be carried out on any finite non-empty interval, $I \subset \mathbb{R}$; the corresponding measures are finite and thus unique. It is easy to see that λ as a measure on $\mathbb{R}$ is σ-finite and hence also unique.

The construction carries over, with obvious modificatons, to $\mathbb{R}^d$: just replace half-open intervals by half-open rectangles. The key is that we have a natural notion of volume for the elementary objects, and that this provides a σ-additive function an the corresponding algebra.

On topological spaces, one can ask for a number of continuity related properties of measures that occasionally will come very handy.

Definition 1.22. Let Ω be a Hausdorff space and $\mathfrak{B}(\Omega)$ the corresponding Borel-σ-algebra. A measure, μ, on $(\Omega, \mathfrak{F} = \mathfrak{B}(\Omega))$, is called:

(0)*Borel measure*, if for any compact set[3], $C \in \mathfrak{F}$, $\mu(C) < \infty$;
(i) *inner regular* or *tight*, if, for all $B \in \mathfrak{F}$, $\mu(B) = \sup_{C \subset B} \mu(C)$, where the supremum is over all compact sets contained in B;
(ii)*outer regular*, if for all $B \in \mathfrak{F}$, $\mu(B) = \inf_{O \supset B} \mu(O)$, where the infimumum is over all open sets containing B.

[3] For Hausdorff spaces it holds also that compact sets are closed. Closed sets in a compact topological space are compact.

(iii) *locally finite*, if for any point $p \in \Omega$ there exists a neighborhood $\mathscr{U}_p$ such that $\mu(\mathscr{U}_p) < \infty$.
(iv) *Radon measure*, if it is inner regular and locally finite.

A very important result is that on a compact metrisable spaces[4], all probability measures are inner regular. The following result will be used in the construction of stochastic processes in Section 3.2.

Theorem 1.23. *Let Ω be a (Hausdorff) compact metrisable space and let $\mathbb{P}$ be a probability measure on $(\Omega, \mathfrak{B}(\Omega))$. Then $\mathbb{P}$ is inner regular.*

Proof. Let $\mathfrak{A}$ be the class of elements, B, of $\mathfrak{B}(\Omega)$, such that, for all $\varepsilon > 0$, there exists a compact set, $K \subset B$, and an open set, $G \supset B$, such that $\mathbb{P}(B \backslash K) < \varepsilon$ and $\mathbb{P}(G \backslash B) < \varepsilon$.

Step 1: Show that $\mathfrak{A}$ *is an algebra*. First, if $B \in \mathfrak{A}$, then its complement, B^c, will also be in $\mathfrak{A}$ (for G^c is closed, $B^c \supset G^c$, and $B^c \backslash G^c = G \backslash B$, and vice versa). Next, if $B_1, B_2 \in \mathfrak{A}$, then there are $K_i \subset B_i$ and $G_i \supset B_i$, such that $\mathbb{P}(B_i \backslash K_i) < \varepsilon/2$ and $\mathbb{P}(G_i \backslash B_i) < \varepsilon/2$. Then $K = K_1 \cup K_2$ and $G = G_1 \cup G_2$ are the desired sets for $B = B_1 \cup B_2$. Thus $\mathfrak{A}$ is an algebra.

Step 2: Show that $\mathfrak{A}$ *is a* σ*-algebra.* Now let B_n be an increasing sequence of elements of $\mathfrak{A}$ such that $\bigcup_{n\in\mathbb{N}} B_n = B$. We choose sets K_n and G_n as before, but with $\varepsilon/2$ replaced by $2^{-n-1}\varepsilon$. Then there exists a $N < \infty$ such that $\mathbb{P}\left(B \backslash \bigcup_{n=1}^N K_n\right) < \varepsilon$. Indeed, $\mathbb{P}\left(B \backslash \bigcup_{n\in\mathbb{N}} K_n\right) < \varepsilon/2$, while $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} K_n \backslash \bigcup_{n=1}^N K_n\right) < \varepsilon/2$ for N large enough. Therefore, there exists a compact set $K \equiv \bigcup_{n=1}^N K_n$ such that $\mathbb{P}(B \backslash K) < \varepsilon$. The same construction works for the corresponding open sets, and so $B \in \mathfrak{A}$. Thus $\mathfrak{A}$ is a σ-algebra.

Step 3: Show that $\mathfrak{B}(\Omega) = \mathfrak{A}$. We need to verify that any compact set $K \in \mathfrak{B}(\Omega)$ is in $\mathfrak{A}$. Since Ω is metrisable, there exists a metric, ρ, such that the topology of Ω is equivalent to the metric topology. If K is a closed and thus compact subset of Ω, then K is the intersection of a sequence of open sets $G_n \equiv \{\omega \in \Omega : \rho(\omega, K) < 1/n\}$,

$$K = \bigcap_{n\in\mathbb{N}} G_n. \tag{1.2.22}$$

Since $G_n \downarrow K$ and $\mathbb{P}$ is finite, it follows that $\mathbb{P}[G_n] \downarrow \mathbb{P}[K]$, because $G_n \backslash K \downarrow \emptyset$ and by σ-additivity $\mathbb{P}[G_n \backslash K] \downarrow 0$. This means that $K \in \mathfrak{A}$. Thus $\mathfrak{A}$ is a σ-algebra that contains all closed sets, and since $\mathfrak{B}(\Omega)$ is the smallest σ-algebra that contains all closed sets, then $\mathfrak{B}(\Omega) \subset \mathfrak{A}$. By definition $\mathfrak{A} \subset \mathfrak{B}(\Omega)$, thus $\mathfrak{B}(\Omega) = \mathfrak{A}$.

Now for any $B \in \mathfrak{B}(\Omega)$ and $K \subset B$ compact, $\mathbb{P}(B) = \mathbb{P}(K) + \mathbb{P}(B \backslash K)$. But since, for any $B \in \mathfrak{A}$, and for any $\varepsilon > 0$, by definition, there exists K such that $\mathbb{P}(B \backslash K) < \varepsilon$. Thus $\sup\{\mathbb{P}(K) : K \subset B\} = \mathbb{P}(B)$, so $\mathbb{P}$ is inner regular. $\square$

Remark. Note that the proof shows that $\mathbb{P}$ is also outer regular. Measures that are both inner and outer regular are sometimes called regular.

[4] A topological space E is compact if every open cover of E has a finite subcover. In other words, if E is the union of a family of open sets, there is a finite subfamily whose union is E.

1.3 Random variables

Definition 1.24. Let $(\Omega,\mathfrak{F})$ and $(E,\mathfrak{G})$ be two measurable spaces. A map $f:\Omega\to E$ is called measurable from $(\Omega,\mathfrak{F})$ to $(E,\mathfrak{G})$, if, for all $A\in\mathfrak{G}$, $f^{-1}(A)\equiv\{\omega\in\Omega: f(\omega)\in A\}\in\mathfrak{F}$.

The notion of measurability implies that a measurable map is capable of transporting a measure from one space to another. Namely, if $(\Omega,\mathfrak{F},\mathbb{P})$ is a probability space, and f is a measurable map from $(\Omega,\mathfrak{F})$ to $(E,\mathfrak{G})$, then

$$\mathbb{P}_f \equiv \mathbb{P}\circ f^{-1}$$

defines a probability measure on $(E,\mathfrak{G})$, called the *induced measure*. Namely, for any $B\in\mathfrak{G}$, by definition

$$\mathbb{P}_f(B)=\mathbb{P}\left(f^{-1}(B)\right)$$

is well defined, since $f^{-1}(B)\in\mathfrak{F}$.

The standard notion of a *random variable* refers to a measurable function from some measurable space to the space $(\mathbb{R},\mathfrak{B}(\mathbb{R}))$. We will generally extend this notion and call any measurable map from a measurable space $(\Omega,\mathfrak{F})$ to a measurable space $(E,\mathfrak{B}(E))$, where E is a topological, respectively metric space, a *E-valued random variable* or a E-valued *Borel function*. Our privileged picture is then that we have an unspecified, so called *abstract probability space* $(\Omega,\mathfrak{F},\mathbb{P})$ on which all kinds of random variables, be it, reals, infinite sequences, functions, or measures, are defined, possibly simultaneously.

An important notion is then that of the *σ-algebra generated by random variables*.

Definition 1.25. Let $(\Omega,\mathfrak{F})$ be a measurable space, and let $(E,\mathfrak{B}(E))$ be a topological space equipped with its Borel-σ-algebra. Let f be an E-valued random variable. We say that $\sigma(f)$ is the smallest σ-algebra such that f is measurable from $(\Omega,\sigma(f))$ to $(E,\mathfrak{B}(E))$.

Note that $\sigma(f)$ depends on the set of values f takes. E.g., if f is real valued, but takes only finitely many values, the σ-algebra generated by f has just finitely many elements. If f is the constant function, then $\sigma(f)=\{\Omega,\emptyset\}$, the trivial σ-algebra. This notion is particularly useful, if several random variables are defined on the same probability space.

Dynkin's lemma has a sometimes useful analogue for so-called monotone classes of functions.

Theorem 1.26 (Monotone class theorem). *Let $\mathscr{H}$ be a class of bounded functions on Ω to $\mathbb{R}$. Assume that*

(i) $\mathscr{H}$ is a vector space over $\mathbb{R}$,
(ii) $1\in\mathscr{H}$,
(iii) if $f_n\geq 0$ are in $\mathscr{H}$, and $f_n\uparrow f$, where f is bounded, then $f\in\mathscr{H}$.

If $\mathscr{H}$ contains the indicator functions of every element of a Π-system $\mathscr{S}$, then $\mathscr{H}$ contains any bounded $\sigma(\mathscr{S})$-measurable function.

Proof. Let $\mathscr{D}$ be the class of subsets D of Ω such that $\mathbb{1}_D \in \mathscr{H}$. Then $\mathscr{D}$ is a λ-system. Since by hypothesis $\mathscr{D}$ contains $\mathscr{S}$, by Dynkin's theorem, $\mathscr{D}$ contains the σ-algebra generated by $\mathscr{S}$. Now let f be a $\sigma(\mathscr{S})$-measurable function s.t. $0 \leq f \leq K < \infty$ for some constant K. Set

$$D(n,i) \equiv \{\omega \in \Omega : i2^{-n} \leq f(\omega) < (i+1)2^{-n}\}, \tag{1.3.1}$$

and set

$$f_n(\omega) \equiv \sum_{i=0}^{K2^n} i2^{-n} \mathbb{1}_{D(n,i)}(\omega). \tag{1.3.2}$$

Every $D(n,i)$ is $\sigma(\mathscr{S})$-measurable, and so $\mathbb{1}_{D(n,i)} \in \mathscr{H}$, and so by (i), $f_n \in \mathscr{H}$. Since $f_n \uparrow f$, $f \in \mathscr{H}$.

To conclude, we take a general $\sigma(\mathscr{S})$-measurable function and decompose it into the positive and negative part and treat each part as before. □

An important property of measurable functions is that the space of measurable functions if closed under limit procedures.

Lemma 1.27. *Let f_n, $n \in \mathbb{N}$, be real valued random variables. Then the functions*

$$f^+ \equiv \limsup_{n\to\infty} f_n \quad \text{and} \quad f^- \equiv \liminf_{n\to\infty} f_n \tag{1.3.3}$$

are measurable. In particular, if the $f_n \to f$ pointwise, than f is measurable.

The proof is left as an exercise.

If Ω is a topological space, we have the natural class of continuous functions from Ω to $\mathbb{R}$. It is easy to see that all continuous functions are measurable if Ω and $\mathbb{R}$ are equipped with their Borel σ-algebras. Thus, all functions that are pointwise limits of continuous functions are measurable, etc..

Remark. Instead of introducing the Borel-σ-algebra, one could go a different path and introduce what is called the *Baire*-σ-algebra. Here one proceeds form the idea that on a topological space one naturally has the notion of continuous functions. One certainly will want all of these to be measurable functions, but certainly one will want more: any pointwise limit of a continuous function should be measurable, as well as limits of sequences of such functions. In this way one arrives at a class of functions, called *Baire*-functions, that is defined as the smallest class of functions that is closed under pointwise limits and that contains the continuous functions. One can then define the *Baire*-σ-algebra as the smallest σ-algebra that makes all Baire-functions measurable. It is in general true that the Borel-σ-algebra contains the Baire-σ-algebra, but in general they are not the same. However, on most spaces we will consider (Polish spaces), the two concepts coincide.

1.4 Integrals

We will now recall the notion of an integral of a measurable function (respectively expectation value of random variables).

To do this one first introduces the notion of *simple functions*:

Definition 1.28. A function, $g : \Omega \to \mathbb{R}$, is called simple if it takes only finitely many values, i.e. if there are numbers, $w_1, \dots, w_k$, and a partition of Ω, $A_i \in \mathfrak{F}$ with $\bigcup_{i=1}^k A_i = \Omega$, such that $A_i = \{\omega \in \mathfrak{F} : g(\omega) = w_i\}$. Then we can write

$$g(\omega) = \sum_{i=1}^{k} w_i \mathbb{1}_{A_i}(\omega).$$

The space of simple measurable functions is denoted by $\mathscr{E}_+$.

It is obvious what the integral of a simple function should be.

Definition 1.29. Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and $g = \sum_{i=1}^k w_i \mathbb{1}_{A_i}$ a simple function. Then

$$\int_\Omega g d\mu = \sum_{i=1}^{k} w_i \mu(A_i). \tag{1.4.1}$$

The integral of a general measurable function is defined by approximation with simple functions.

Definition 1.30.

(i) Let f be non-negative and measurable. Then

$$\int_\Omega f d\mu \equiv \sup_{g \leq f, g \in \mathscr{E}_+} \int_\Omega g d\mu \tag{1.4.2}$$

Note that the value of the integral is in $\mathbb{R} \cup \{+\infty\}$.

(ii)If f is measurable, set

$$f(\omega) = \mathbb{1}_{f(\omega) \geq 0} f(\omega) + \mathbb{1}_{f(\omega) < 0} f(\omega) \equiv f_+(\omega) - f_-(\omega)$$

If either $\int_\Omega f_+(\omega) < \infty$ or $-\int_\Omega f_-(\omega) d\mu < \infty$, define

$$\int_\Omega f d\mu \equiv \int_\Omega f_+(\omega) d\mu - \int_\Omega f_-(\omega) d\mu \tag{1.4.3}$$

(iii)We call a function f integrable or *absolutely integrable*, if

$$\int_\Omega |f| d\mu < \infty.$$

We state the key properties of the integral without proof.

The most fundamental property is the monotone convergence theorem, which to a large extent justifies the (otherwise strange) definition above.

Theorem 1.31. *Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and f a real valued non-negative measurable function. Let $f_1 \leq f_2 \leq \cdots \leq f$ be a monotone increasing sequence of non-negative measurable functions that converge pointwise to f. Then*

$$\int_\Omega f d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu \tag{1.4.4}$$

The monotone convergence theorem allows to provide an "explicit" construction of the integral as originally used by Lebesgue as a definition.

Lemma 1.32. *Let f be a non-negative measurable function. Then*

$$\int_\Omega f d\mu \equiv \lim_{n\to\infty} \left[\sum_{k=0}^{n2^n-1} 2^{-n} k \mu(\{\omega : 2^{-n}k \leq f(\omega) < 2^{-n}(k+1)\}) \right. \\ \left. + n\mu(\{\omega : f(\omega) \geq n\}) \right] \tag{1.4.5}$$

The following lemma is known as *Fatou's lemma*:

Lemma 1.33. *Let f_n be a sequence of measurable non-negative functions. Then*

$$\int_\Omega \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n d\mu. \tag{1.4.6}$$

Equally central is *Lebesgue's dominated convergence theorem*:

Theorem 1.34. *Let f_n be a sequence of absolutely integrable functions, and let f be a measurable function such that*

$$\lim_{n\to\infty} f_n(\omega) = f(\omega) \quad \text{for } \mu\text{-almost all } \omega.$$

Let $g \geq 0$ be a positive function such that $\int_\Omega g d\mu < \infty$ and

$$|f_n(\omega)| \leq g(\omega) \quad \text{for } \mu\text{-almost all } \omega.$$

Then f is absolutely integrable with respect to μ and

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu. \tag{1.4.7}$$

In the case when we are dealing with integrals with respect to a probability measure, there exists a very useful improvement of the dominated convergence theorem that leads us to the important notion of *uniform integrability*.

Let us first make the following observation.

Lemma 1.35. *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and let X be an integrable real valued random variables on this space. Then, for any $\varepsilon > 0$, there exists $K < \infty$, such that*

$$\mathbb{E}\left(|X| \mathbb{1}_{|X|>K}\right) < \varepsilon. \tag{1.4.8}$$

Proof. This is a direct consequence from the monotone convergence theorem. We leave the details to the reader. $\square$

When dealing with families of random variables, one problem is that this property will in general not hold uniformly. A nice situation occurs if it does:

Definition 1.36. Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space. A class, $\mathfrak{C}$, of real valued random variables is called *uniformly integrable*, if, for any $\varepsilon > 0$, there exists $K < \infty$, such that, for all $X \in \mathfrak{C}$,

$$\mathbb{E}\left(|X|\mathbb{1}_{|X|>K}\right) < \varepsilon. \tag{1.4.9}$$

Note that, in particular, if $\mathfrak{C}$ is uniformly integrable, then there exists a constant, $C < \infty$, such that, for all $X \in \mathfrak{C}$, $\mathbb{E}(|X|) \leq C$.

Remark. The simplest example of a class of random variables that is not uniformly integrable is given as follows. Take X_n such that

$$\mathbb{P}(X_n = 1) = 1 - 1/n \quad \text{and} \quad \mathbb{P}(X_n = n) = 1/n. \tag{1.4.10}$$

Clearly, for any K, $\lim_{n\to\infty} \mathbb{E}(|X_n|\mathbb{1}_{|X_n|>K}) = 1$. One should always keep this example in mind when reflecting upon uniform integrability.

Note that on the other hand the class of functions, Y_n, with

$$\mathbb{P}(Y_n = 1) = 1 - 1/n \quad \text{and} \quad \mathbb{P}(Y_n = \sqrt{n}) = 1/n \tag{1.4.11}$$

is uniformly integrable.

Theorem 1.37 (Uniform integrability). *Let X_n, $n \in \mathbb{N}$ and X be integrable random variables on some probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. Then $\lim_{n\to\infty} \mathbb{E}|X_n - X| = 0$, if and only if*

(i) $X_n \to X$ in probability, and
(ii)the family $X_n, n \in \mathbb{N}$ is uniformly integrable.

Proof. We show the "if" part. Define

$$\phi_K(x) \equiv \begin{cases} K, & \text{if } x > K, \\ x, & \text{if } |x| \leq K, \\ -K, & \text{if } x < -K. \end{cases} \tag{1.4.12}$$

We have obviously from the uniform integrability that

$$\mathbb{E}(|\phi_K(X_n) - X_n|) \leq \varepsilon, \tag{1.4.13}$$

for $n \geq 0$ (where for convenience we set $X \equiv X_0$). Moreover, since $|\phi_K(x) - \phi_K(y)| \leq |x - y|$, (i) implies that $\phi_K(X_n) \to \phi_K(X)$ in probability. Since, moreover, $\phi_K(X_n)$ is bounded, we may choose n_0 such that, for $n \geq n_0$, $\mathbb{P}(|\phi_K(X_n) - \phi_K(X)| > \delta) \leq \varepsilon/K$. Then

$$\mathbb{E}(|\phi_K(X_n) - \phi_K(X)|) \leq \delta + 2\varepsilon$$

and so $\lim_{n\to\infty} \mathbb{E}(|\phi_K(X_n) - \phi_K(X)|) = 0$. In view of the fact that (1.4.13) holds for any ε, it follows that $\mathbb{E}(|X_n - X|) \to 0$.

Let us now show the converse ("only") direction. If $\mathbb{E}(|X_n - X|) \to 0$, then by Chebychev's inequality, $\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}(|X_n - X|)}{\varepsilon} \to 0$, so $X_n \to X$ in probability. Moreover, X is absolutely integrable.

Now write $X_n = (X_n - X) + X$ and use that, by the triangle inequality,

$$\mathbb{E}(|X_n|) \leq \mathbb{E}(|X|) + \mathbb{E}(|X_n - X|). \tag{1.4.14}$$

For any $\varepsilon > 0$, there exists n_0 such that, for all $n \geq n_0$, $\mathbb{E}(|X_n - X|) < \varepsilon$. Since all X_i and X are integrable, there exists K such that, for all $n \leq n_0$, $\mathbb{E}(|X_n| \mathbb{1}_{|X_n|>K}) < \varepsilon$. Hence

$$\mathbb{E}(|X_n| \mathbb{1}_{|X_n|>2K}) \leq \begin{cases} \varepsilon & \text{if } n \leq n_0, \\ \mathbb{E}(|X| \mathbb{1}_{|X_n|>2K}) + \varepsilon & \text{if } n > n_0. \end{cases} \tag{1.4.15}$$

Finally we use that, for $n > n_0$,

$$\begin{aligned} \mathbb{E}(|X| \mathbb{1}_{|X_n|>2K}) &\leq \mathbb{E}(|X| \mathbb{1}_{|X|>2K-|X-X_n|}) \\ &\leq \mathbb{E}(|X| \mathbb{1}_{|X|>K}) + \mathbb{E}(|X| \mathbb{1}_{|X|\leq K} \mathbb{1}_{|X-X_n|>K}) \\ &\leq \varepsilon + K\mathbb{P}(|X - X_n| > K) \leq 2\varepsilon. \end{aligned} \tag{1.4.16}$$

This concludes the proof. $\square$

The importance of this result lies in the fact that in probability theory, we are very often dealing with functions that are not really bounded, and where Lebesgue's theorem is not immediately applicable either. Uniform integrability is the best possible condition for convergence of the integrals. Note that the simple example (1.4.11) of a uniformly integrable family given above furnishes a nice example where $\mathbb{E}(|X_n - X|) \to 0$, but where Lebesgue's dominated convergence theorem cannot be applied.

Exercise: *Use the previous criterion to prove Lebesgue's dominated convergence theorem in the case of probability measures.*

1.5 $\mathscr{L}^p$ and L^p spaces

I will only rather briefly summarize some frequently used notions concerning spaces of integrable functions. Given a measure space, $(\Omega, \mathfrak{F}, \mu)$, one defines, for $p \in [1, \infty]$ and measurable functions, f,

$$\|f\|_{p,\mu} \equiv \|f\|_p \equiv (\mathbb{E}|f|^p)^{1/p} = \left(\int_\Omega |f|^p d\mu \right)^{1/p}. \tag{1.5.1}$$

The set of functions, f, such that $\|f\|_{p,\mu} < \infty$ is denoted by $\mathscr{L}^p(\Omega, \mathfrak{F}, \mu) \equiv \mathscr{L}^p$.

There are two crucial inequalities.

Lemma 1.38 (Minkowski inequality). *For $f,g \in \mathscr{L}^p$,*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p, \tag{1.5.2}$$

Lemma 1.39 (Hölder inequality). *For measurable functions f,g and $p,q \in [1,\infty]$ are such that $\frac{1}{p}+\frac{1}{q}=1$, then*

$$|\mathbb{E}(fg)| \leq \|f\|_p \|g\|_q, \tag{1.5.3}$$

Both inequalities follow from one of the most important inequalities in integration theory, Jensen's inequality.

Theorem 1.40 (Jensen's inequality). *Let $(\Omega,\mathfrak{F},\mu)$ be a probability space, let X be an absolutely integrable random variable, and let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function. Then, for any $c \in \mathbb{R}$,*

$$\mathbb{E}\varphi(X - \mathbb{E}X + c) \geq \varphi(c), \tag{1.5.4}$$

and in particular

$$\mathbb{E}\varphi(X) \geq \varphi(\mathbb{E}X). \tag{1.5.5}$$

Proof. If φ is convex, then for any y there is a straight line below φ that touches φ at $(y,\varphi(y))$, i.e. there exists $m \in \mathbb{R}$ such that $\varphi(x) \geq \varphi(y) + (x-y)m$. Choosing $x = X - \mathbb{E}X + c$ and $y = c$ and taking expectations on both sides yields (1.5.4). □

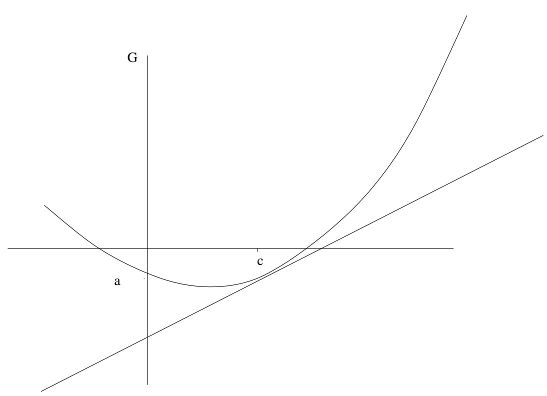


Fig. 1.1 Convex funktion

Exercise: *Prove the Hölder inequalties (for $p > 1$) using Jensen's inequality.*

Since Minkowski's inequality is really a triangle inequality and linearity is trivial, we would be inclined to think that $\|\cdot\|_p$ is a norm and $\mathscr{L}^p$ is a normed space. In fact, the only problem is that $\|f\|_p = 0$ does not imply $f = 0$, since f maybe non-zero on sets of μ-measure zero. Therefore to define a normed space, one considers equivalence classes of functions in $\mathscr{L}^p$ by calling two functions, f, f' *equivalent*, if $f - f'$ is non-zero only on set of measure zero. The space of these equivalence classes is called $L^p \equiv L^p(\Omega, \mathfrak{F}, \mu)$.

The following fact about L^p spaces will be useful to know.

Lemma 1.41. *The spaces $L^p(\Omega, \mathfrak{F}; \mu)$ are Banach spaces (i.e. complete normed vector space).*

Proof. The by now only non-trivial fact that needs to be proven is the completeness of L^p. Let $f_i \in L^p$, $i \in \mathbb{N}$ be a Cauchy sequence. Then there are $n_k \in \mathbb{N}$, such that, for all $i, j \geq n_k$, $\|f_i - f_j\|_p \leq 2^{-k-k/p}$. Set $g_k \equiv f_{n_k}$ and

$$F \equiv \sum_{k \in \mathbb{N}} 2^{kp} |g_k - g_{k+1}|^p. \tag{1.5.6}$$

Then

$$\mathbb{E}(F) = \sum_{k \in \mathbb{N}} 2^{kp} \mathbb{E}(|g_k - g_{k+1}|^p) = \sum_{k \in \mathbb{N}} 2^{kp} \|g_k - g_{k+1}\|_p^p \leq 1. \tag{1.5.7}$$

Therefore, F is integrable and hence finite except possibly on a set of measure zero. It follows that for all $\omega \in \Omega$ s.t. $F(\omega)$ is finite, $|g_k(\omega) - g_{k+1}(\omega)| \leq 2^{-k} F(\omega)^{1/p}$. It follows further, using telescopic expansion and the triangle inequality, that $g_k(\omega)$ is a Cauchy sequence of real numbers, and hence convergent. Set $f(\omega) = \lim_{k \to \infty} g_k(\omega)$. For the ω in the null-set where $F(x) = +\infty$, we set $f(\omega) = 0$. It follows readily that

$$\mathbb{E}(|g_k - f|^p) \to 0, \tag{1.5.8}$$

and using once more the Cauchy property of f_n, that

$$\mathbb{E}(|f_n - f|^p) \to 0. \tag{1.5.9}$$

$\square$

The case $p = 2$ is particularly nice, in that the space L^2 is not only a Banach space, but a Hilbert space. The point here is that the Hölder inequality, applied for the case $p = 2$, yields

$$\mathbb{E}(fg) \leq \sqrt{\mathbb{E}(f^2)\mathbb{E}(g^2)} = \|f\|_2 \|g\|_2. \tag{1.5.10}$$

This means that on L^2, there exists a quadratic form $(\cdot, \cdot)_\mu$,

$$(f, g)_\mu \equiv \int_{\mathbb{R}} fg d\mu \equiv \mathbb{E}(fg) \tag{1.5.11}$$

which has the properties of a *scalar product*. The L^2-norm being the derived norm, $\|f\|_2 = \sqrt{(f, g)_\mu}$. Although somehow L^2 spaces are not the most natural settings for probability, it is sometimes quite convenient to exploit this additional structure.

1.6 Fubini's theorem

An always important tool for the computation of integral on product spaces is Fubini's theorem. We consider first the case of non-negative functions.

Theorem 1.42 (Fubini-Tonnelli). *Let $(\Omega_1, \mathfrak{F}_1, \mu_1)$, and $(\Omega_2, \mathfrak{F}_2, \mu_2)$ be two measure spaces, and let f be a real-valued, non-negative measurable function on $(\Omega_1 \times \Omega_2, \mathfrak{F}_1 \otimes \mathfrak{F}_2)$. Then the two functions*

$$h(x) \equiv \int_{\Omega_2} f(x,y)\mu_2(dy) \quad \textit{and} \quad g(y) \equiv \int_{\Omega_1} f(x,y)\mu_1(dx)$$

are measurable with respect to $\mathfrak{F}_1$ resp. $\mathfrak{F}_2$, and

$$\int_{\Omega_1 \times \Omega_2} f d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} h d\mu_1 = \int_{\Omega_2} g d\mu_2 \tag{1.6.1}$$

Now we turn to the general case.

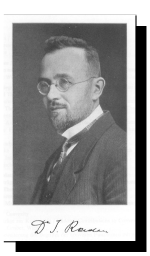
Theorem 1.43 (Fubini-Lebesgue). *Let $f : (\Omega_1 \times \Omega_2, \mathfrak{F}_1 \otimes \mathfrak{F}_2) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ be absolutely integrable with respect to the product measure $\mu_1 \otimes \mu_2$. Then*

(i) For μ_1-almost all x, $f(x,y)$ is absolutely integrable with respect to μ_2, and vice versa.
(ii)The functions $h(x) = \int_{\Omega_2} f(x,y)\mu_2(dy)$ and $g(y) = \int_{\Omega_1} f(x,y)\mu_1(dx)$, are well-defined except possibly on a set of measure zero with respect to the measures μ_1, resp. μ_2, and absolutely integrable with respect to these same measures.
(iii)The equation

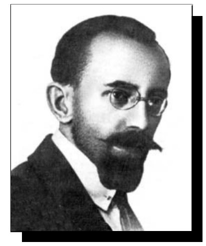
$$\int_{\Omega_1 \times \Omega_2} f d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} h(x)\mu_1(dx) = \int_{\Omega_2} g(y)\mu_2(dy) \tag{1.6.2}$$

holds.

1.7 Densities, Radon-Nikodým derivatives



In Probability 1 we have encountered the notion of a probability density. In fact, we had constructed the Lebesgue-Stieltjes measure on $\mathbb{R}$ by prescribing a distribution function, F, (i.e. a non-decreasing, right-continuous function) in term of which any interval $(a,b]$ had measure $\mu((a,b]) = F(b) - F(a)$. In the special case when there was a positive function f, such that for all $a < b$, $F(b) - F(a) = \int_a^b f(x)dx$, where dx indicates the standard Lebesgue measure, we called f the density of μ and said that μ is *absolutely continuous* with respect to Lebesgue measure.



We now want to generalise these notions to the general context of positive measures. In particular, we want to be able to say when two measures are absolutely continuous with respect to each other, and define the corresponding relative densities.

First we notice that it is rather easy to modify a given measure μ on a measurable space $(\Omega, \mathfrak{F})$ with the help of a measurable function f. To do so, we set, for any $A \in \mathfrak{F}$,

$$\mu_f(A) \equiv \int_A f d\mu. \tag{1.7.1}$$

Exercise: *Show that if f is measurable and integrable, but not necessarily non-negative, μ_f, defined as in (1.7.1), defines an additive set-function. Show that, if $f \geq 0$, μ_f is indeed a measure on $(\Omega, \mathfrak{F})$.*

We see that in the case when μ is the Lebesgue measures, μ_f is the absolutely continuous measure with density f. In the general case, we have that, if $\mu(\mathcal{O}) = 0$, then it is also true that $\mu_f(\mathcal{O}) = 0$. The latter property will define the notion of absolute continuity between general measures.

Definition 1.44. Let μ, ν be two measures on a measurable space $(\Omega, \mathfrak{F})$.

- We say that ν is *absolutely continuous* with respect to μ, or $\nu \ll \mu$, if and only if, all μ-null sets, $\mathcal{O}$ (i.e. all sets $\mathcal{O}$ with $\mu(\mathcal{O}) = 0$), are ν-null sets.
- We say that two measures, μ, ν, are equivalent if $\mu \ll \nu$ and $\nu \ll \mu$.
- We say that a measure ν is *singular* with respect to μ, or $\nu \perp \mu$, if there exists a set $\mathcal{O} \in \mathfrak{F}$ such that $\mu(\mathcal{O}) = 0$ and $\nu(\mathcal{O}^c) = 0$.

It is important to keep in mind that the notion of absolute continuity is not symmetric.

The following important theorem, called the *Radon-Nikodým theorem*, asserts that relative absolute continuity is equivalent to the existence of a density.

Theorem 1.45. *Let μ, ν be two σ-finite measures on a measurable space $(\Omega, \mathfrak{F})$. Then the following two statements are equivalent:*

(i) $\nu \ll \mu$.
(ii)There exists a non-negative measurable function, f, such that $\nu = \mu_f$.

Moreover, f is unique up to μ-null sets.

Definition 1.46. If $\nu \ll \mu$, then a positive measurable function f such that $\nu = \mu_f$ is called the *Radon-Nikodým derivative* of ν with respect to μ, denoted

$$f = \frac{d\nu}{d\mu}. \tag{1.7.2}$$

Proof. Note that the implication $(ii) \Rightarrow (i)$ is obvious from the definition. The other direction is more tricky.

We consider for simplicity the case when μ, ν are finite measures. The extension to σ-finite measures can then easily be carried through by using suitable partitions of Ω.

We need a few concepts and auxiliary results. The first is the notion of the *essential supremum*.

Definition 1.47. Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and T an arbitrary non-empty set. The *essential supremum*, $g \equiv \mathrm{esup}_{t\in T}\, g_t$, of a class, $\{g_t, t \in T\}$, of measurable functions $g_t : \Omega \to [-\infty, +\infty]$ (with respect to μ), is defined by the properties

(i) g is measurable;
(ii)$g \geq g_t$, μ-almost everywhere, for each $t \in T$;
(iii)for any h that satisfies (i) and (ii), $h \geq g$, μ- a.e.

Note that by definition, if there are two g that satisfy this definition, then they are μ-a.e. equal. Not that the essential supremum depends on μ only trough its null-sets.

The first fact we need to establish is that the essential supremum is always equal to the supremum over a countable set.

Lemma 1.48. *Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space with μ a σ-finite measure. Let $\{g_t, t \in T\}$ be a non-empty class of real measurable functions. Then there exists a countable subset $T_0 \subset T$, such that*

$$\sup_{t\in T_0} g_t = \mathrm{esup}_{t\in T}\, g_t. \tag{1.7.3}$$

Proof. It is enough to consider the case when μ is finite. Moreover, we may restrict ourselves to the case when $|g_t| < C$, for all $t \in T$ (e.g. by passing from g_t to $\tanh^{-1}(g_t)$, which is monotone and preserves all properties of the definition). Let $\mathscr{S}$ denote the class of all countable subsets of T. Set

$$\alpha \equiv \sup_{I\in\mathscr{S}} \mathbb{E}\left(\sup_{t\in I} g_t\right). \tag{1.7.4}$$

Now let $I_n \in \mathscr{S}$ be a sequence such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left(\sup_{t\in I_n} g_t\right) = \alpha, \tag{1.7.5}$$

and set $T_0 = \bigcup_{n\in\mathbb{N}} I_n$. Of course, T_0 is countable and $\alpha = \mathbb{E}\left(\sup_{t\in T_0} g_t\right)$. The function $g \equiv \sup_{t\in T_0} g_t$ is measurable, since it is the supremum over a countable set of measurable functions. To see that it also satisfies (ii), assume that there exists $t \in T$, such that $g_t > g$ on a set of positive measure. Then for this t, $\mathbb{E}(\max(g, g_t)) > \mathbb{E}(g) = \alpha$. On the other hand, $T_0 \cup \{t\}$ is a countable subset of T, and so by definition of α, $\mathbb{E}(\max(g, g_t)) \leq \alpha$, which yields a contradiction. Thus (ii) holds. To show (iii), assume that there exists h satisfying (i) and (ii). By (ii), $h \geq g_t$, a.e., for each $t \in T$, and thus also $h \geq \sup_{t\in T_0} g_t$, a.e., since a countable union of null-sets is a null set. Thus g satisfies property (iii), too. Therefore, $g = \mathrm{esup}_{t\in T} g_t$. $\square$

The notion of essential supremum is used in the next lemma, which is the major step in the proof of the Radon-Nikodým theorem.

Lemma 1.49. *Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space, with μ a σ-finite measure, and let ν be another σ-finite measure on $(\Omega, \mathfrak{F})$. Let $\mathscr{H}$ be the family of all measurable functions, $h \geq 0$, such that, for all $A \in \mathfrak{F}$, $\int_A h d\mu \leq \nu(A)$. Then, for all $A \in \mathfrak{F}$,*

$$\nu(A) = \psi(A) + \int_A g d\mu, \tag{1.7.6}$$

where ψ is a measure that is singular with respect to μ and

$$g = \mathrm{esup}_{h\in\mathscr{H}} h \tag{1.7.7}$$

with respect to μ.

Proof. We again assume μ, ν to be finite, and leave the extension to σ-finite measures as an easy exercise. We also exclude the trivial case of $\mu = 0$. From Lemma 1.48 we know that there exists a sequence of functions $h_n \in \mathscr{H}$, such that $g = \sup_{n\in\mathbb{N}} h_n$. Let us first note that if $h_1, h_2 \in \mathscr{H}$, then so is $h \equiv \max(h_1, h_2)$. To see this, note that the disjoint sets

$$A_1 \equiv \{\omega \in A : h_1(\omega) \geq h_2(\omega)\}, \quad A_2 \equiv \{\omega \in A : h_2(\omega) > h_1(\omega)\} \tag{1.7.8}$$

are measurable and $A_1 \cup A_2 = A$. But

$$\int_A h d\mu = \int_{A_1} h_1 d\mu + \int_{A_2} h_2 d\mu \leq \nu(A_1) + \nu(A_2) = \nu(A), \tag{1.7.9}$$

which implies $h \in \mathscr{H}$. We may therefore assume the sequence h_n ordered such that $h_n \leq h_{n+1}$, for all $n \geq 1$. Then $g = \lim_{n\to\infty} h_n$, and by monotone convergence, for all $A \in \mathfrak{F}$,

$$\int_A g d\mu = \lim_{n\to\infty} \int_A h_n d\mu \leq \nu(A). \tag{1.7.10}$$

As a consequence, ψ defined by (1.7.6) satisfies $\psi(A) \geq 0$, for all $A \in \mathfrak{F}$. Moreover, trivially $\psi(\emptyset) = 0$, as both ν and $g d\mu$ are measures, ψ defined as their difference is σ-additive. Thus ψ is a measure.

It remains to show that ψ is singular with respect to μ. To this end we construct a set of zero ψ-measure whose complement has zero μ-measure. Of course, this can only be done through a delicate limiting procedure. To begin we define collections of sets whose ψ-measure is much smaller than their μ-measure. More precisely, for $n \in \mathbb{N}$ and $A \in \mathfrak{F}$ with $\mu(A) > 0$, let

$$\mathscr{D}_n(A) \equiv \left\{B \in \mathfrak{F} : B \subset A, \psi(B) < n^{-1}\mu(B)\right\}. \tag{1.7.11}$$

The key fact is that any set A of positive μ measure contains such subsets, i.e. $D_n(A) \neq \emptyset$ whenever $\mu(A) \neq 0$. This is proven by contradiction: assume that $D_n(A) = \emptyset$. Then set $h_0 = n^{-1}\mathbb{1}_A$. For all $B \in \mathfrak{F}$ one has that

$$\int_B h_0 d\mu = n^{-1}\mu(A \cap B) \leq \psi(A \cap B) \leq \psi(B) = \nu(B) - \int_B g d\mu. \tag{1.7.12}$$

But then $\int_B (h_0 + g) d\mu \leq \nu(B)$, for all $B \in \mathfrak{F}$, so that $g + h_0 \in \mathscr{H}$, which contradicts the fact that $g = \operatorname{esup}_{h \in \mathscr{H}} h$, since $h_0 > 0$ on a set of positive μ-measure.

Since *any* set of positive μ-measure contains ψ-tiny subsets, one may expect that a set of full μ-measure is ψ-tiny. Below we show this by successively collecting all the μ mass in such sets.

We can now choose $B_{1,n} \in \mathscr{D}_n(\Omega)$ with the property that

$$\mu(B_{1,n}) \geq \frac{1}{2} \sup \{\mu(B) : B \in \mathscr{D}_n(\Omega)\} \equiv \alpha_{1,n}. \tag{1.7.13}$$

Morally, $B_{1,n}$ is our first attempt to pick up as much μ-mass as we can from the ψ-tiny sets. If we were lucky, and $\mu(B^c_{1,n}) = 0$, then we stop the procedure. Otherwise, we continue by picking up as much mass as we can from what was left, i.e. we choose $B_{2,n} \in \mathscr{D}_n(B^c_{1,n})$ with

$$\mu(B_{2,n}) \geq \frac{1}{2} \sup \left\{\mu(B) : B \in \mathscr{D}_n(B^c_{1,n})\right\} \equiv \alpha_{2,n}. \tag{1.7.14}$$

If $\mu\left((B_{2,n} \cup B_{1,n})^c\right) = 0$, we are happy and stop. Otherwise, we continue and choose $B_{3,n} \in \mathscr{D}_n\left((B_{1,n} \cup B_{2,n})^c\right)$ with

$$\mu(B_{3,n}) \geq \frac{1}{2} \sup \left\{\mu(B) : B \in \mathscr{D}_n(B^c_{1,n} \cap B^c_{2,n})\right\} \equiv \alpha_{3,n}, \tag{1.7.15}$$

and so on. If the process stops at some k_n-th step, set $B_{j,n} = \emptyset$ for $j > k_n$.

It is obvious from the definition that $B_{j,n} \in \mathscr{D}_n(\Omega)$, if $B_{j,n} \neq \emptyset$. Since $\mathscr{D}_n(\Omega)$ is closed under countable disjoint unions (both ψ and μ being measures), also $M_n \equiv \bigcup_{j=1}^{\infty} B_{j,n} \in \mathscr{D}_n(\Omega)$. We want to show that $\mu(M^c_n) = 0$, that is we have picked up all the mass eventually. To do this, note again that, if $\mu(M^c_n) > 0$, then there exists $D \in \mathscr{D}_n(M^c_n)$ with $\mu(D) > 0$.

On the other hand, for any $m \in \mathbb{N}$,

$$\begin{aligned} 2\alpha_{m,n} &= \sup \left\{ \mu(B) : B \in \mathscr{D}_n \left(\bigcap_{j=1}^{m-1} B^c_{j,n} \right) \right\} \\ &\geq \sup\{\mu(B) : B \in \mathscr{D}_n(M^c_n)\} \geq \mu(D). \end{aligned} \tag{1.7.16}$$

Thus, if $\mu(D) > 0$, then there exists some $\alpha > 0$, such that $\mu(B_{m,n}) \geq \alpha_{m,n} = \alpha$, for all m. Since all $B_{j,n}$ are disjoint, this would imply that $\mu(M_n) = \infty$, which contradicts the assumption that μ is a finite measure. Thus we conclude that $\mu(M^c_n) = 0$, and so $\psi(M_n) < n^{-1}\mu(M_n) = n^{-1}\mu(\Omega)$. Therefore,

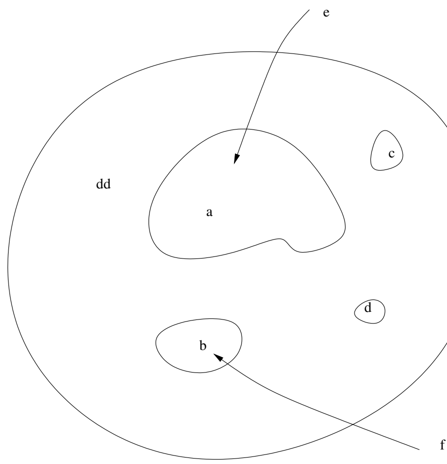


Fig. 1.2 Construction of the sets $B_{i,n}$

$$\psi\left(\bigcap_{n=1}^{\infty} M_n\right) \leq \lim_{n=1}^{\infty} \psi\left(M_n\right) = 0, \tag{1.7.17}$$

$$\mu\left(\left(\bigcap_{n=1}^{\infty} M_n\right)^c\right) = \mu\left(\bigcup_{n=1} \infty M_n^c\right) \leq \sum_{n=1}^{\infty} \mu\left(M_n^c\right) = 0.$$

This proves that ψ is singular with respect to μ. $\square$

As the first consequence of this lemma, we state the famous Lebesgue decomposition theorem.

Theorem 1.50. *If μ, ν are σ-finite measures on a measurable space $(\Omega, \mathfrak{F})$, then there exist two uniquely determined measures, ν_c, ν_s, such that $\nu = \nu_s + \nu_c$, where ν_c is absolutely continuous with respect to μ and ν_s is singular with respect to μ.*

Proof. Lemma 1.49 provides the existence of two measures ν_s and ν_c with the desired properties. To prove the uniqueness of this decomposition, assume that there are $\tilde{\nu}_s, \tilde{\nu}_c$ with the same properties. Since the measures $\nu_s, \tilde{\nu}_s$ are carried on sets of zero μ-mass, they can only be different if there exists a set $A \in \mathfrak{F}$ with $\mu(A) = 0$ and $\nu_s(A) \neq \tilde{\nu}_s(A) > 0$. But then $\nu_c(A) \neq \tilde{\nu}_c(A)$ as well, while by absolute continuity, $\nu_c(A) = \tilde{\nu}_c(A) = 0$. Thus $\nu_s = \tilde{\nu}_s$ and consequently $\nu_c = \tilde{\nu}_c$. $\square$

The Radon-Nikodým theorem is now immediate: Assume that ν is absolutely continuous with respect to μ. The decomposition (1.7.6) applied to μ-null sets A then implies that for all these sets, $\psi(A) = 0$. But ψ is singular with respect to μ, so there should be a μ-null set, A, for which $\psi(A^c) = 0$. But since for all such A, $\psi(A) = 0$, it follows that $\psi(\Omega) = \psi(A) + \psi(A^c) = 0$, and so ψ is the zero-measure.

All that remains is to assert that the Radon-Nikodým derivative is unique a.e.. To do this, assume that there exists another measurable function, g^*, such that

$$\nu(A) = \int_A g^* d\mu. \tag{1.7.18}$$

Now define the measurable set $A = \{\omega : C > g^* > g > -C\}$. Then, by assumption,

$$\int_A g^* d\mu = \nu(A) = \int_A g d\mu. \tag{1.7.19}$$

But since on A $g^* > g$, this can only hold if $\mu(A) = 0$, for all $C < \infty$. Thus, $\mu(g^* > g) = 0$. In the same way one shows that $\mu(g^* < g) = 0$, implying that g and g^* differ at most on sets of measure zero.

□

Remark. We have said (and seen in the proof), that the Radon-Nikodým derivative is defined modulo null-sets (w.r.t. μ). This is completely natural. Note that if μ and ν are equivalent, then $0 < \frac{d\nu}{d\mu} < \infty$ almost everywhere, and $\frac{d\nu}{d\mu} = \frac{1}{\frac{d\mu}{d\nu}}$.

The following property of the Radon-Nikodým derivative will be needed later.

Lemma 1.51. *Let μ, ν be σ-finite measures on $(\Omega, \mathfrak{F})$, and let $\nu \ll \mu$. If X is $\mathfrak{F}$-measurable and ν-integrable, then, for any $A \in \mathfrak{F}$,*

$$\int_A X d\nu = \int_A X \frac{d\nu}{d\mu} d\mu. \tag{1.7.20}$$

Proof. We may assume that μ is finite and X non-negative. Appealing to the monotone convergence theorem, it is also enough to consider bounded X (otherwise, approximate and pass to the limit on both sides). Let $\mathscr{H}$ be the class of all bounded non-negative $\mathfrak{F}$-measurable functions for which (1.7.20) is true. Then $\mathscr{H}$ satisfies the hypothesis of Theorem 1.26: clearly, (i) $\mathscr{H}$ is a vector space, (ii) the function 1 is contained in $\mathscr{H}$ be definition of the Radon-Nikodým derivative, and the property (1.7.20) is stable under monotone convergence by the monotone convergence theorem. Also, $\mathscr{H}$ contains the indicator functions of all elements of $\mathfrak{F}$. Then the assertion of Theorem 1.26 implies that $\mathscr{H}$ contains all bounded $\mathfrak{F}$-measurable function, as claimed. □

Chapter 2
Conditional expectations and conditional probabilities

In this chapter we will generalise the notion of conditional expectations and conditional probabilities from elementary probability theory considerably. In elementary probability, we could condition only on *events* of positive probability. This notion is too restrictive, as we have seen in the context of Markov processes, where this limited us to consider discrete state spaces. The new notions we will introduce is conditioning on σ-algebras. In this section we follow largely the presentation in Chow and Teicher [4] where much further material can be found.

2.1 Conditional expectations

Definition 2.1. Consider a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. Let $\mathfrak{G} \subset \mathfrak{F}$ be sub-σ-algebra of $\mathfrak{F}$. Let X be a random variable, i.e. a $\mathfrak{F}$-measurable (real-valued) function on Ω such that $|\mathbb{E}X| \leq \infty$. We say that a function Y is a conditional expectation of X given $\mathfrak{G}$, written $Y = \mathbb{E}(X|\mathfrak{G})$, if

(i) Y is $\mathfrak{G}$-measurable, and
(ii)For all $A \in \mathfrak{G}$,

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}. \tag{2.1.1}$$

Remark. If two functions Y, Y' both satisfy the conditions of a conditional expectation, then they can differ only on sets of probability zero, i.e. $\mathbb{P}(Y = Y') = 1$. One calls such different realizations of a conditional expectation *versions*.

Remark. The condition $|\mathbb{E}X| \leq \infty$ means that $\mathbb{E}X$ is well-defined, in the sense that $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$ and either $\mathbb{E}X^+ < \infty$ or $\mathbb{E}X^- < \infty$. It is the weakest possible under which a definition of conditional expectation can make sense. Existence of conditional expectations can be established under just this condition (see [4]), however, we will in the sequel only treat the simple case when X is absolutely integrable, $\mathbb{E}(|X|) < \infty$.

Intuitively, this notion of conditional expectation can be seen as "integrating" the random variable partially, i.e. with respect to all degrees of freedom that do not affect the σ-algebra $\mathfrak{G}$. A trivial example would be the case where $\Omega = \mathbb{R}^2$, and $\mathfrak{G}$ is the σ-algebra of events that depend only on the first coordinate, say x. Then the conditional expectation of a function $f(x,y)$ is just the integral with respect of the variables y (recall the construction of the integral in Fubini's theorem), modulo renormalisation. What is left is, of course, a function that depends only on x, and that also satisfies property (ii). The advantage of the notion of a conditional expectation given a σ-algebra is that it largely generalises this concept.

Before we discuss the existence of conditional expectations with respect to σ-algebras, we want to discuss the relation to the more elementary notion of conditional expectations with respect to sets. Recall that if $A \in \mathfrak{F}$ has positive mass, $\mathbb{P}(A) > 0$, we can define the conditional expectation, given A, as

$$\mathbb{E}(X|A) = \frac{\int_A X d\mathbb{P}}{\mathbb{P}(A)}. \tag{2.1.2}$$

Recall that when we were studying Markov chains, we wanted to define conditional expectations of the form

$$\mathbb{E}(f(X_{n+1})|X_n = x). \tag{2.1.3}$$

In the case of finite state spaces, we could do this using the definition (2.1.2), because we could without loss generality assume that $\mathbb{P}(X_n = x)$ was strictly positive. In the case of continuous state space, the canonical situation would be that $\mathbb{P}(X_n = x) = 0$, for any $x \in S$, and so the definition (2.1.2) is not applicable. It is to overcome this difficulty that we introduce our new notion of conditional expectation given a σ-algebra.

Let us now see how these two concepts connect. To this end, we define

$$Y(\omega) \equiv \sum_{x \in \mathscr{S}} \mathbb{E}(f(X_{n+1})|X_n = x) \mathbb{1}_{X_n(\omega)=x}.$$

Clearly, this is a $\sigma(X_n)$-measurable function and for any $A \in \sigma(X_n)$

$$\begin{aligned}
\mathbb{E}Y\mathbb{1}_A &= \sum_{x \in \mathscr{S}} \mathbb{E}\left(\mathbb{1}_{X_n(\omega)\in A} \mathbb{1}_{X_n(\omega)=x} \mathbb{E}(f(X_{n+1})|X_n = x)\right) \\
&= \sum_{x \in X_n(A)} \mathbb{P}(X_n = x)\mathbb{E}(f(X_{n+1})|X_n = x) \\
&= \mathbb{E}\left(\sum_{x \in X_n(A)} f(X_{n+1})\mathbb{1}_{X_n(\omega)=x}\right) \\
&= \mathbb{E}\left(\mathbb{1}_A f(X_{n+1})\right),
\end{aligned} \tag{2.1.4}$$

and thus Y is the conditional expectation of $f(X_{n+1})$ given the σ-algebra $\sigma(X_n)$. Note also that we want to think of the $\sigma(X_n)$-measurable function $Y(\omega)$ as a function of the *value* of the random variable X_n, since it depends on ω only through this value.

In many cases that we will encounter, the σ-algebra, $\mathfrak{G}$, with respect to which we are conditioning is the σ-algebra, $\sigma(Y)$, generated by some other random variable, Y. In that case we will often write

$$\mathbb{E}(X|\sigma(Y)) \equiv \mathbb{E}(X|Y) \tag{2.1.5}$$

and call this *the conditional expectation of X given Y*. We may than also think of is a function of the value of the random variable Y.

As we can see, the difficulty associated with constructing conditional expectations in the general case relates to making sense of expressions of the form $0/0$. The key to the construction of conditional expectations in the general case will use the concept of the Radon-Nikodým derivative.

Theorem 2.2. *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space, let X be a random variable such that $\mathbb{E}(|X|) < \infty$, and let $\mathfrak{G} \subset \mathfrak{F}$ be a sub-σ-algebra of $\mathfrak{F}$. Then*

(i) there exists a $\mathfrak{G}$-measurable function, $\mathbb{E}(X|\mathfrak{G})$, unique up to sets of measure zero, the conditional expectation *of X given $\mathfrak{G}$, such that for all $A \in \mathfrak{G}$,*

$$\int_A \mathbb{E}(X|\mathfrak{G}) d\mathbb{P} = \int_A X d\mathbb{P}. \tag{2.1.6}$$

(ii)If X is absolutely integrable and Z is an absolutely integrable, $\mathfrak{G}$-measurable random variable such that, for some Π-System $\mathscr{D}$ with $\sigma(\mathscr{D}) = \mathfrak{G}$,

$$\mathbb{E}(Z) = \mathbb{E}(X), \quad \textit{and} \quad \int_A Z d\mathbb{P} = \int_A X d\mathbb{P} \textit{ for all } A \in \mathscr{D}, \tag{2.1.7}$$

then $Z = \mathbb{E}(X|\mathfrak{G})$ almost everywhere.

Proof. We begin by proving (i). Define the set functions $\lambda, \lambda^+, \lambda^-$ as

$$\lambda^{\pm}(A) \equiv \int_A X^{\pm} d\mathbb{P}, \quad \lambda \equiv \lambda^+ - \lambda^- \tag{2.1.8}$$

Now we can consider the restriction of λ to $\mathfrak{G}$, denoted by $\lambda_{\mathfrak{G}}$, and the restriction of $\mathbb{P}$ to $\mathfrak{G}$, $\mathbb{P}_{\mathfrak{G}}$. Clearly, $\lambda^{\pm}$ are absolutely continuous with respect to $\mathbb{P}$, and their restrictions to $\mathfrak{G}$, $\lambda^{\pm}_{\mathfrak{G}}$, are absolutely continuous with respect to the restriction of $\mathbb{P}$ to $\mathfrak{G}$, $\mathbb{P}_{\mathfrak{G}}$. But since X is assumed to be absolutely integrable with respect to $\mathbb{P}$ and $\mathbb{P}$ is a probability measure, it follows that also $\lambda^{\pm}_{\mathfrak{G}}$ are finite measures. Therefore, the Radon-Nikodým theorem 1.45 implies that there exist $\mathfrak{G}$-measurable functions, $Y^{\pm} = \frac{d\lambda^{\pm}_{\mathfrak{G}}}{d\mathbb{P}_{\mathfrak{G}}}$, such that, for all $A \in \mathfrak{G}$,

$$\int_A Y^{\pm} d\mathbb{P} = \lambda^{\pm}(A) = \int_A X^{\pm} d\mathbb{P}, \tag{2.1.9}$$

and hence $Y = \frac{d\lambda_{\mathfrak{G}}}{d\mathbb{P}_{\mathfrak{G}}} \equiv Y^+ - Y^-$, such that

$$\int_A Y d\mathbb{P} = \lambda(A) = \int_A X d\mathbb{P}. \tag{2.1.10}$$

Thus, Y has the properties of a conditional expectation and we may set $\mathbb{E}(X|\mathfrak{G}) = Y = \frac{d\lambda_{\mathfrak{G}}}{d\mathbb{P}_{\mathfrak{G}}}$. Note that Y is unique up to sets of measure zero. Finally, to show that the conditional measure is unique in the same sense, assume that there is a function Y' satisfying the conditions of the conditional expectation that differs from Y on a set of positive measure. Then one may set $A^{\pm} = \{\omega : \pm(Y'(\omega) - Y(\omega)) > 0\}$, and at least one of these sets, say A^+, has positive measure. Then

$$\int_{A^+} X d\mathbb{P} = \int_{A^+} Y' d\mathbb{P} > \int_{A^+} Y d\mathbb{P} = \int_{A^+} X d\mathbb{P}, \tag{2.1.11}$$

which is impossible. This proves uniqueness and hence (i) is established.

To prove (ii), set

$$\mathfrak{A} \equiv \left\{ A \in \mathfrak{F} : \int_A Z d\mathbb{P} = \int_A X d\mathbb{P} \right\}. \tag{2.1.12}$$

then $\Omega \in \mathfrak{A}$, and $\mathscr{D} \subset \mathfrak{A}$, by assumption. Also, $\mathfrak{A}$ is a λ-system, and so by Dynkin's theorem, $\mathfrak{A} \supset \sigma(\mathscr{D}) = \mathfrak{G}$, and so Z is the desired conditional expectation. □

2.2 Elementary properties of conditional expectations

Conditional expectations share most of the properties of ordinary expectations. The following is a list of elementary properties:

Lemma 2.3. *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and let $\mathfrak{G} \subset \mathfrak{F}$ be a sub-σ-algebra. Then:*

(i) If X is $\mathfrak{G}$-measurable, then $\mathbb{E}(X|\mathfrak{G}) = X$, a.s.;
(ii)The map $X \to \mathbb{E}(X|\mathfrak{G})$ is linear;
(iii)$\mathbb{E}[\mathbb{E}(X|\mathfrak{G})] = \mathbb{E}(X)$;
(iv)If $\mathfrak{B} \subset \mathfrak{G}$ is a σ-algebra, then $\mathbb{E}[\mathbb{E}(X|\mathfrak{G})|\mathfrak{B}] = \mathbb{E}(X|\mathfrak{B})$, a.s..
(v)$|\mathbb{E}(X|\mathfrak{G})| \leq \mathbb{E}(|X||\mathfrak{G})$, a.s.;
(vi)If $X \leq Y$, then $\mathbb{E}(X|\mathfrak{G}) \leq \mathbb{E}(Y|\mathfrak{G})$, a.s.;

Proof. Left as an exercise! □

The following theorem summarises the most important properties of conditional expectations with regard to limits.

Theorem 2.4. *Let X_n, $n \in \mathbb{N}$ and Y be absolutely integrable random variables on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, and let $\mathfrak{G} \subset \mathfrak{F}$ be a sub-σ-algebra. Then*

(i) If $Y \leq X_n \uparrow X$ a.s., then $\mathbb{E}(X_n|\mathfrak{G}) \uparrow \mathbb{E}(X|\mathfrak{G})$ a.s..
(ii)If $Y \leq X_n$ a.s., then

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n|\mathfrak{G}\right) \leq \liminf_{n\to\infty} \mathbb{E}\left(X_n|\mathfrak{G}\right). \tag{2.2.1}$$

(iii)If $X_n \to X$ a.s., and $|X_n| \leq |Y|$, for all n, then

$$\lim_{n\to\infty} \mathbb{E}(X_n|\mathfrak{G}) = \mathbb{E}(X|\mathfrak{G}) \text{ a.s..}$$

Of course, these are just the analogs of the three basic convergence theorems for ordinary expectations. We leave the proofs as exercises.

A useful, but not unexpected, property is the following lemma.

Lemma 2.5. *Let X be integrable and let Y be bounded and $\mathfrak{G}$-measurable. Then*

$$\mathbb{E}(XY|\mathfrak{G}) = Y\mathbb{E}(X|\mathfrak{G}), \text{ a.s.} \tag{2.2.2}$$

Proof. We may assume that X,Y are non-negative; otherwise decompose them into positive and negative parts and use linearity of the conditional expectation. Moreover, it is enough to consider bounded random variables; otherwise, consider increasing sequences of bounded random variables that converge to them and use the monotone convergence theorem.

Define, for any $A \in \mathfrak{F}$,

$$\nu(A) \equiv \int_A XY d\mathbb{P}, \quad \mu(A) \equiv \int_A X d\mathbb{P}. \tag{2.2.3}$$

Both μ and ν are finite measures that are absolutely continuous with respect to $\mathbb{P}$. Then

$$\frac{d\nu_{\mathfrak{G}}}{d\mathbb{P}_{\mathfrak{G}}} = \mathbb{E}(XY|\mathfrak{G}), \quad \frac{d\mu_{\mathfrak{G}}}{d\mathbb{P}_{\mathfrak{G}}} = \mathbb{E}(X|\mathfrak{G}), \quad \frac{d\mu}{d\mathbb{P}} = X. \tag{2.2.4}$$

Then, using Lemma 1.51, for any $A \in \mathfrak{G}$,

$$\int_A Y d\mu_{\mathfrak{G}} = \int_A Y \frac{d\mu_{\mathfrak{G}}}{d\mathbb{P}_{\mathfrak{G}}} d\mathbb{P}_{\mathfrak{G}} = \int_A Y\mathbb{E}(X|\mathfrak{G}) d\mathbb{P}_{\mathfrak{G}}, \tag{2.2.5}$$

whereas for any $A \in \mathfrak{F}$,

$$\int_A Y d\mu = \int_A Y \frac{d\mu}{d\mathbb{P}} d\mathbb{P} = \int_A YX d\mathbb{P}. \tag{2.2.6}$$

Specializing the second equality to the case when $A \in \mathfrak{G}$, we find that for those A,

$$\int_A Y\mathbb{E}(X|\mathfrak{G}) d\mathbb{P} = \int_A YX d\mathbb{P}. \tag{2.2.7}$$

Now $Z \equiv Y\mathbb{E}(X|\mathfrak{G})$ is $\mathfrak{G}$-measurable, and (2.2.7) is precisely the defining property for Z to be the conditional expectation of XY. This concludes the proof. □

There should be a natural connection between independence and conditional expectation, as it was the case for the elementary notion of conditional expectation. Here it is.

Theorem 2.6. *Two σ-algebras, $\mathfrak{G}_1, \mathfrak{G}_2$, are independent, if and only if, for all $\mathfrak{G}_2$-measurable integrable random variables, X,*

$$\mathbb{E}(X|\mathfrak{G}_1) = \mathbb{E}(X) \quad \text{a.s.} \tag{2.2.8}$$

Note that in the theorem we can replace “for all integrable $\mathfrak{G}_2$ measurable random variable” by “for all random variables of the form $X = \mathbb{1}_B$, $B \in \mathfrak{G}_2$”.

Proof. Assume first that $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are independent. Let $A \in \mathfrak{G}_1$ and X be $\mathfrak{G}_2$-measurable. The random variables $\mathbb{1}_A$ and X are independent, thus

$$\mathbb{E}(\mathbb{1}_A X) = \mathbb{E}(\mathbb{1}_A)\mathbb{E}(X) = \mathbb{E}(\mathbb{1}_A \mathbb{E}(X))$$

and from the definition of conditional expectation

$$\mathbb{E}[\mathbb{1}_A \mathbb{E}(X|\mathfrak{G}_1)] = \mathbb{E}(\mathbb{1}_A X)$$

for all $A \in \mathfrak{G}_1$. Thus (2.2.8) holds.

Now assume that (2.2.8) holds. Choose $X = \mathbb{1}_B$, $B \in \mathfrak{G}_2$. Then

$$\mathbb{E}(\mathbb{1}_B|\mathfrak{G}_1) = \mathbb{E}(\mathbb{1}_B) = \mathbb{P}(B).$$

Then, for all $A \in \mathfrak{G}_1$,

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{E}(\mathbb{1}_A \mathbb{1}_B) = \mathbb{E}[\mathbb{E}(\mathbb{1}_A \mathbb{1}_B|\mathfrak{G}_1)] \\ &= \mathbb{E}[\mathbb{E}(\mathbb{1}_B|\mathfrak{G}_1)\mathbb{1}_A] = \mathbb{E}(\mathbb{P}(B)\mathbb{1}_A) = \mathbb{P}(A)\mathbb{P}(B).\end{aligned}$$

Thus $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are independent. □

2.3 The case of random variables with absolutely continuous distributions

Let us consider some cases where conditional expectations can be computed more “explicitly”. For this, consider two random variables, X, Y, with values in $\mathbb{R}^m$ and $\mathbb{R}^n$ (in the sequel, nothing but notation changes if the assume $n = m = 1$, so we will do this). We assume that the joint distribution of X and Y is absolutely continuous with respect to Lebesgue’s measure with density $p(x,y)$. That is, for any function $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}_+$,

$$\mathbb{E}(f(X,Y)) = \int f(x,y)p(x,y)dxdy.$$

The (marginal) density of the random variable Y is then

$$q(y) = \int p(x,y)dx$$

(where we should modify the density to be zero, when $\int p(x,y)dx = \infty$. This can be done because this can be true only on a set of Lebesgue measure zero). Let us note first that the set where $q(y)=0$ has mesure zero: indeed,

$$\int\int \mathbb{1}_{q(y)=0}p(x,y)dxdy = \int \mathbb{1}_{q(y)=0}q(y)dy = 0.$$

Let now $h:\mathbb{R}^m \to \mathbb{R}_+$ be a measurable function. We want to compute $\mathbb{E}(h(X)|Y)$. To do this, take a measurable function $g:\mathbb{R}^n \to \mathbb{R}_+$. Then

$$\begin{aligned}
\mathbb{E}(h(X)g(Y)) &= \int h(x)g(y)p(x,y)dxdy \qquad (2.3.1)\\
&= \int \left(\int h(x)p(x,y)dx\right) g(y)dy\\
&= \int \left(\frac{\int h(x)p(x,y)dx}{q(y)}\right) g(y)q(y)\mathbb{1}_{q(y)>0}dy\\
&\equiv \int \phi(y)g(y)q(y)\mathbb{1}_{q(y)>0}dy\\
&= \mathbb{E}(\phi(Y)g(Y)),
\end{aligned}$$

where we were allowed to introduce the indicator function $\mathbb{1}_{q(y)>0}$ because as we have seen, the complementray set has measure zero.

From this calculation we can derive the following

Proposition 2.7. *With the notation above, let $\nu(y,dx)$ be the measure on $\mathbb{R}^m$ defined by*

$$\nu(y,dx) \equiv \begin{cases} \frac{p(x,y)}{q(y)}dx, & \text{if } q(y)>0,\\ \delta_0(dx), & \text{if } q(y)=0. \end{cases} \qquad (2.3.2)$$

Then for any measurable[1] *function $h:\mathbb{R}^m \to \mathbb{R}_+$,*

$$\mathbb{E}(h(X)|Y)(\omega) = \int h(x)\nu(Y(\omega),dx). \qquad (2.3.3)$$

Proof. It is obvious that the right-hand side of Equation (2.3.3) is measurable with respect to $\sigma(Y)$. Verifying the second defining property of the conditional expectation amounts to repeating the compuations in Eq. (2.3.1). □

Definition 2.8. The function $\frac{p(x,y)}{q(y)}$ as a function of x is called the *conditional density* of X given $Y=y$.

What is particular here is that we can represent it as an expectation with respect to an explicitely given probability measure. We see that in this context, we are formally quite close to the discrete case and the intuitive notion of conditional expectations.

[1] One can show that the statement holds true for any measurable and integrable $h:\mathbb{R}^m \to \mathbb{R}$.

2.4 The special case of L^2-random variables

Conditional expectations have a particularly nice interpretation in the case when the random variable X is square-integrable, i.e. if $X \in L^2(\Omega, \mathfrak{F}, \mathbb{P})$ (since for the moment we think of conditional expectations as equivalence classes modulo sets of measure zero, we may consider X as an element of L^2 rather than $\mathscr{L}^2$). We will identify that space $L^2(\Omega, \mathfrak{G}, \mathbb{P})$ as the subspace of $L^2(\Omega, \mathfrak{F}, \mathbb{P})$ for which at least one representative of each equivalence class is $\mathfrak{G}$-measurable.

Theorem 2.9. *If $X \in L^2(\Omega, \mathfrak{F}, \mathbb{P})$, then $\mathbb{E}(X|\mathfrak{G})$ is the orthogonal projection of X on $L^2(\Omega, \mathfrak{G}, \mathbb{P})$.*

Proof. The Jensen-inequality applied to the conditional expectation yields that $\mathbb{E}(X^2|\mathfrak{G}) \geq \mathbb{E}(X|\mathfrak{G})^2$, and hence $\mathbb{E}[\mathbb{E}(X|\mathfrak{G})^2] \leq \mathbb{E}[\mathbb{E}(X^2|\mathfrak{G})] = \mathbb{E}(X^2) < \infty$, so that $\mathbb{E}(X|\mathfrak{G}) \in L^2(\Omega, \mathfrak{G}, \mathbb{P})$. Moreover, for any bounded, $\mathfrak{G}$-measurable function Z,

$$\mathbb{E}[Z(X - \mathbb{E}(X|\mathfrak{G}))] = \mathbb{E}(ZX) - \mathbb{E}[Z\,\mathbb{E}(X|\mathfrak{G})] = \mathbb{E}(ZX) - \mathbb{E}[\mathbb{E}(ZX|\mathfrak{G})] = 0. \quad (2.4.1)$$

Thus, $X - \mathbb{E}(X|\mathfrak{G})$ is orthogonal to all bounded $\mathfrak{G}$-measurable random variables, and using that these form a dense set in $L^2(\Omega, \mathfrak{G}, \mathbb{P})$, it is orthogonal to $L^2(\Omega, \mathfrak{G}, \mathbb{P})$. This proves the theorem. $\square$

Note that this interpretation of the conditional expectation can be used to *define* the conditional expectation for L^2-random variables.

2.5 Conditional probabilities and conditional probability measures

From conditional expectations we now want to construct conditional probability measures. These seems quite straightforward, but there are some non-trivial technicalities that arise from the version business of conditional expectations.

As before we consider a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ and a sub-σ-algebra $\mathfrak{G}$. For any $A \in \mathfrak{F}$, we can define

$$\mathbb{P}(A|\mathfrak{G}) \equiv \mathbb{E}(\mathbb{1}_A|\mathfrak{G}), \quad (2.5.1)$$

and call it the *conditional probability of A given* $\mathfrak{G}$. It is a $\mathfrak{G}$-measurable function that satisfies

$$\int_B \mathbb{P}(A|\mathfrak{G}) d\mathbb{P} = \mathbb{E}\left(\int_B \mathbb{1}_A d\mathbb{P} \,\middle|\, \mathfrak{G} \right) = \mathbb{E}(\mathbb{P}(A \cap B)|\mathfrak{G}) = \mathbb{P}(A \cap B),$$

for any $B \in \mathfrak{F}$.

It clearly inherits from the conditional expectation the following properties:

(i) $0 \leq \mathbb{P}(A|\mathfrak{G}) \leq 1$, a.s.;

(ii) $\mathbb{P}(A|\mathfrak{G}) = 0$, a.s., if and only if $\mathbb{P}(A) = 0$; also $\mathbb{P}(A|\mathfrak{G}) = 1$, a.s., if and only if $\mathbb{P}(A) = 1$;
(iii) If $A_n \in \mathfrak{F}$, $n \in \mathbb{N}$, are disjoint sets, then

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n \,\middle|\, \mathfrak{G}\right) = \sum_{n\in\mathbb{N}} \mathbb{P}\left(A_n|\mathfrak{G}\right), \text{a.s.}; \tag{2.5.2}$$

(iv) If $A_n \in \mathfrak{F}$, such that $\lim_{n\to\infty} A_n = A$, then

$$\lim_{n\to\infty} \mathbb{P}(A_n|\mathfrak{G}) = \mathbb{P}(A|\mathfrak{G}), \text{a.s.}. \tag{2.5.3}$$

These observations bring us close to thinking that conditional probabilities can be thought of as $\mathfrak{G}$-measurable functions taking values in the probability measures, at least for almost all ω. The problem, however, is that the requirement of σ-additivity which seems to be satisfied due to (iii) is in fact problematic: (iii) says, that, for any sequence A_n, there exists a set of measure one, such that, for all ω in this set,

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n \,\middle|\, \mathfrak{G}\right)(\omega) = \sum_{n\in\mathbb{N}} \mathbb{P}\left(A_n|\mathfrak{G}\right)(\omega). \tag{2.5.4}$$

However, this set may depend on the sequence, and since that space is not countable, it is unclear whether there exists a set of full measure on which (2.5.4) holds for all sequences of sets.

These considerations lead to the definition of so-called *regular conditional probabilities*.

Definition 2.10. Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and let $\mathfrak{G}$ be a sub-σ-algebra. A *regular conditional probability measure* or *regular conditional probability* on $\mathfrak{F}$ given $\mathfrak{G}$ is a function, $P(\omega, A)$, defined for all $A \in \mathfrak{F}$ and all $\omega \in \Omega$, such that

(i) for each $\omega \in \Omega$, $P(\omega, \cdot)$ is a probability measure on $(\Omega, \mathfrak{F})$;
(ii) for each $A \in \mathfrak{F}$, $P(\cdot, A)$ is a $\mathfrak{G}$-measurable function coinciding with the conditional probability $\mathbb{P}(A|\mathfrak{G})$ almost everywhere.

The point is that, if we have a regular conditional probability, then we can express conditional expectations as expectations with respect normal probability measures.

Theorem 2.11. *With the notation form above, if $P_\omega(A) \equiv P(\omega, A)$ is a regular conditional probability on $\mathfrak{F}$ given $\mathfrak{G}$, then for a $\mathfrak{F}$-measurable integrable random variable, X,*

$$\mathbb{E}(X|\mathfrak{G})(\omega) = \int X dP_\omega \quad \text{a.s.} \tag{2.5.5}$$

Proof. As often, we may assume X positive. The proof then goes through the monotone class theorem (Theorem 1.26), quite similar to the proof of Theorem 1.51. One defines the class of functions where (2.5.5) holds, verifies that it satisfies the hypothesis of the monotone class theorem and notices that it is true for all indicator functions of sets in $\mathfrak{F}$. $\square$

The question remains whether and when regular conditional probabilities exist. An example of a regular conditional probability measure (on the measure space $(\mathbb{R}^n \times \mathbb{R}^m, \mathfrak{B}((\mathbb{R}^n \times \mathbb{R}^m), p(x,y)dxdy)$ is the measure ν from Proposition 2.7. It is easy to check that this has all the required properties properties, in particular it exitst for every y.

A central result for us is the existence in the case when Ω is a Polish space.

Theorem 2.12. *Let $(\Omega, \mathfrak{B}(\Omega), \mathbb{P})$ be a probability space where Ω is a Polish space and $\mathfrak{B}(\Omega)$ is the Borel-σ-algebra. Let $\mathfrak{G} \subset \mathfrak{B}(\Omega)$ be a sub-σ-algebra. Then there exists a regular conditional probability $P(\omega, A)$ given $\mathfrak{G}$.*

We will not give the proof of this theorem here.

Chapter 3
Stochastic processes

We are finally ready to come to the main topic of this course, stochastic processes. In this chapter we give the basic definitions, prove the fundamental theorem of Kolmogorov, and discuss some examples.

3.1 Definition of stochastic processes

There are various equivalent ways in which stochastic processes can be defined, and it will be useful to always keep them in mind.

The traditional definition.

The standard way to define stochastic processes is as follows. We begin with an abstract probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. Next we need a measurable space $(S, \mathfrak{B})$ (which in almost all cases will be a Polish space together with its Borel σ-algebra). The space S is called the *state space*. Next, we need a set I, called the *index set*. Then a stochastic process with state space S and index set I is a collection of $(S, \mathfrak{B})$-valued *random variables*, $\{X_t, t \in I\}$ defined on $(\Omega, \mathfrak{F}, \mathbb{P})$.

We call such a stochastic process also a stochastic process *indexed by* I. The term stochastic process is often reserved to the cases when I is either $\mathbb{N}, \mathbb{Z}, \mathbb{R}_+$, or $\mathbb{R}$. The index set is then interpreted as a *time parameter*. Depending on whether the index set is discrete or continuous, one refers to stochastic processes with *discrete* or *continuous* time. However, there is also an extensive theory of stochastic processes indexed by more complicated sets, such as $\mathbb{R}^d$, $\mathbb{Z}^d$, etc.. Often these are also referred to as *stochastic fields*. We will mostly be concerned with the standard case of one-dimensional index sets, but I will give examples of the more general case below.

From the point of view of mappings, we have the picture that for any $t \in I$, there is a measurable map,

$$X_t : \Omega \to S,$$

whose inverse maps $\mathfrak{B}$ into $\mathfrak{F}$.

For this to work, we do want, of course, $\mathfrak{F}$ to be so rich that it makes all functions $X_t, t \in I$ measurable. We denote this σ-algebra by

$$\sigma(X_t; t \in I). \tag{3.1.1}$$

An example of a stochastic process with discrete time are families of independent random variables.

Sample paths.

Given a stochastic process as defined above, we can take a different perspective and view, for each $\omega \in \Omega$, $X(\omega)$ as a map from I to S,

$$\begin{aligned} X(\omega) : I &\to S \\ t &\mapsto X_t(\omega) \end{aligned}$$

We call such a function a *sample path* of X, or a *realisation* of X. Clearly here we want to see the stochastic process as a random variable with values in the space of functions,

$$\begin{aligned} X : \Omega &\to S^I \\ \omega &\mapsto X(\omega), \end{aligned}$$

where we view S^I as the space of functions $I \to S$. To complete this image, we need to endow S^I with a σ-algebra, $\mathfrak{B}^I$. How should we choose the σ-algebra on S^I? Our picture will be that X maps $(\Omega, \mathfrak{F})$ to $(\mathscr{S}^I, \mathfrak{B}^I)$. If this map is measurable, then the marginals $X_t : \Omega \to S$ should be measurable. This will be the case if the projection maps $\pi_t : S^I \mapsto S$ that map a function $x \in S^I$ to its value at time t, $\pi_t(x) = x_t$, are measurable from $\mathfrak{B}^I$ to $\mathfrak{B}$.

Lemma 3.1. *Let $\mathfrak{B}^I$ be the smallest σ-algebra that contains all subsets of S^I of the form*

$$C(A,t) \equiv \left\{ x \in S^I : x_t \in A \right\}. \tag{3.1.2}$$

with $A \in \mathfrak{B}$, $t \in I$. Then $\mathfrak{B}^I$ is the smallest σ-algebra such that all the maps $\pi_t : S^I \to S$ that map $x \mapsto x_t$, are measurable. Then $\sigma(X_t, t \in I) \subset \mathfrak{F}$ is the smallest σ-algebra such that the map $X : \Omega \mapsto S^I$ is measurable from $(\Omega, \sigma(X_t, t \in I))$ to $(S^I, \mathfrak{B}^I)$.

Proof. We first show that all π_t are measurable from

$$\sigma(C(A,t), A \in \mathfrak{B}, t \in I) \to \mathfrak{B}. \tag{3.1.3}$$

To do this, let $A \in \mathfrak{B}$, and chose $t \in I$. Then

$$\pi_t^{-1}(A) = C(A,t).$$

Thus each π_t is measurable. On the other hand, assume that there is some t and some A such that $C(A,t) \notin \mathfrak{B}^I$. Then clearly $\pi_t^{-1}(A) \notin \mathfrak{B}^I$, and then π_t is not measurable! So all $C(A,t)$ must be contained, but none more have to.

Finally, $x_t(\omega) = \pi_t(X(\omega)$, so if X is measurable from $\mathfrak{F}$ to $\mathfrak{B}^I$ and π_t from $\mathfrak{B}^I$ to $\mathfrak{B}$, then the composition is measurable from $\mathfrak{F}$ to $\mathfrak{B}$. The fact that $\sigma(X_t, t \in I)$ is minimal follows as before. □

Definition 3.2. If $J \subset I$ is finite, and $B \in \mathfrak{B}^J$, we call a set

$$C(B,J) \equiv \{x \in S^I : x_J \equiv \{x_t, t \in J\} \in B\} \tag{3.1.4}$$

a *cylinder set* or more precisely *finite dimensional cylinder sets*. If B is of the form $B = \times_{t\in J} A_t$, $A_t \in \mathfrak{B}$, we call such a set a *special cylinder*.

It is clear that $\mathfrak{B}^I$ contains all finite dimensional cylinder sets, but of course it contains much more. We call $\mathfrak{B}^I$ the product σ-algebra, or the algebra generated by the cylinder sets.

It is easy to check that the special cylinders form a Π-system, and the cylinders form an algebra; both generate $\mathfrak{B}^I$.

Lemma 3.3. *The map $X : \Omega \to S^I$ is measurable from $\mathfrak{F} \to \mathfrak{B}^I$ if and only if, for each t, X_t is measurable from $\mathfrak{F} \to \mathfrak{B}$.*

Proof. Since a map, X, is measurable from a σ-algebra $\mathfrak{F} \to \mathfrak{B}^I$, if $X^{-1}(C) \in \mathfrak{F}$ for all C in a class that generates $\mathfrak{B}$, to check measurablility it is enough to consider C of the form $C(A,t)$. But

$$X^{-1}(C(A,t)) = \{\omega \in \Omega : X_t(\omega) \in A\},$$

which is in $\mathfrak{F}$ whenever X_t is measurable. To prove the converse implication is equally trivial. □

Thus we see that the choice of the σ-algebra $\mathfrak{B}^I$ is just the right one to make the two points of view on stochastic processes equivalent from the point of view of measurablility.

The law of a stochastic process.

Once we view X as a map from Ω to the S-valued functions on I, we can define the probability distribution induced by $\mathbb{P}$ on the space $(S^I, \mathfrak{B}^I)$,

$$\mu_X \equiv \mathbb{P} \circ X^{-1} \tag{3.1.5}$$

on $(S^I, \mathfrak{B}^I)$ as the distribution of the random variable X.

Canonical process.

Given a stochastic process with law μ, one can of course realise this process on the probability space $(S^I, \mathfrak{B}^I, \mu)$. In that case the random variable X is the trivial map

$$\begin{aligned} X : S^I &\to S^I \\ x &\mapsto X(x) = x. \end{aligned}$$

The viewpoint of the canonical process is, however, not terribly helpful, since more often than not, we want to keep a much richer probability space on which many other random objects can be defined.

3.2 Construction of stochastic processes; Kolmogorov's theorem

The construction of a stochastic process may appear rather formidable, but we may draw encouragement from the fact that we have introduced a rather coarse σ-algebra on the space S^I. The most fundamental observation is that stochastic processes are determined by their observation on just finitely many points in time. We first make this important notion precise.

For any $J \subset I$, we will denote by π_J the canonical projection from S^I to S^J, i.e. $\pi_J X \in S^J$, such that, for all $t \in J$, $(\pi_J X)_t = X_t$. Naturally, we can define the distributions

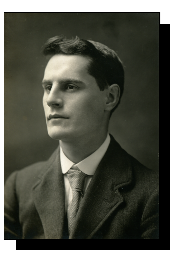
$$\mu_X^J \equiv \mathbb{P} \circ (\pi_J X)^{-1}$$

on S^J.

Definition 3.4. Let $F(I)$ denote the set of all finite, non-empty subsets of I. Then the collection of probability measures

$$\left\{ \mu_X^J : J \in F(I) \right\} \tag{3.2.1}$$

is called the collection of finite dimensional distributions[1] of X.



Note that the finite dimensional distributions determine μ_X on the algebra of finite dimensional cylinder sets. Hence, by Dynkin's theorem, they determine the distribution on the σ-algebra $\mathfrak{B}^I$. This is nice. What is nicer, is that one can also go the other way and *construct* the law of a stochastic process from specified finite dimensional distributions. This will be the content of the fundamental theorem of Daniell and Kolmogorov.

[1] Alternative appellation are "finite dimensional marginal distributions" or "finite dimensional marginals".

Theorem 3.5. *Let S be a compact metrisable space, and let $\mathfrak{B} \equiv \mathfrak{B}(S)$ be its Borel-σ-algebra. Let I be a set. Suppose that, for each $J \in F(I)$, there exists a probability measure, μ^J, on $(S^J, \mathfrak{B}^J)$, such that for any $J_1 \subset J_2 \in F(I)$,*

$$\mu^{J_1} = \mu^{J_2} \circ \pi_{J_1}^{-1}, \tag{3.2.2}$$

where π_{J_1} denotes the canonical projection from $S^{J_2} \to S^{J_1}$. Then there exists *a* unique *measure, μ, on $(S^I, \mathfrak{B}^I)$, such that, for all $J \in F(I)$,*

$$\mu \circ \pi_J^{-1} = \mu^J. \tag{3.2.3}$$

Proof. It will not come as a surprise that we will use Carathéodory's theorem to prove our result. To do this, we have to construct a σ-additive set function on an algebra that generates the σ-algebra $\mathfrak{B}^I$. Of course, this algebra will be the algebra of all finite-dimensional cylinder events. It is rather easy to see what this set function will have to be. Namely, if B is a finite dimensional cylinder, then there exists $J \in F(I)$, and $A_J \in \mathfrak{B}^J$, such that $B = A_J \times S^{I\setminus J}$ (we call in such a case J the *base* of the cylinder). Then we can define

$$\mu_0(B) = \mu^J(A_J). \tag{3.2.4}$$

Clearly $\mu_0(\emptyset) = 0$, and μ_0 is finitely additive: if B_1, B_2 are disjoint finite dimensional cylinders with basis J_i, then we can write B_i, $i = 1,2$, in the form $A_i \times S^{I\setminus J}$, where $J = J_1 \cup J_2$, and $A_i \in \mathfrak{B}^J$ are disjoint. Then it is clear that

$$\mu_0(B_1 \cup B_2) = \mu^J(A_1 \cup A_2) = \mu^J(A_1) + \mu^J(A_2) = \mu_0(B_1) + \mu_0(B_2) \tag{3.2.5}$$

where the consistency relations (3.2.2) were used in the last step. The usual way to prove σ-additivity is to use the fact that an additive set-function, μ_0, is σ-additive if and only if for any sequence $G_n \downarrow \emptyset$, $\mu(G_n) \downarrow 0$.

Therefore, the proof will be finished once we establish the following lemma.

Lemma 3.6. *Let B_n, $n \in \mathbb{N}$ be a sequence of cylinder sets such that $B_n \supset B_{n+1}$ for all n. If there exists an $\varepsilon > 0$, such that for all $n \in \mathbb{N}$, $\mu_0(B_n) \geq 2\varepsilon$, then* $\lim_{n\to\infty} B_n \neq \emptyset$.

Proof. If B_n satisfies the assumptions of the lemma, then there exists a sequence $J_n \in F(I)$ and $A_n \in \mathfrak{B}^{J_n}$, such that $B_n = A_n \times S^{I\setminus J_n}$, $J_n \subset J_{n+1}$ and

$$\mu_0(B_n) = \mu_{J_n}(A_n).$$

It will be enough to assume that $J_n = \{1, \dots, n\}$. Since μ^{J_n} is a probability measure on the compact metrisable space S^{J_n}, Theorem 1.23 implies that, for any $\varepsilon > 0$, there exists a compact subset, $K_n \subset A_n$, such that

$$\mu^{J_n}(K_n) \geq \mu^{J_n}(A_n) - 2^{-n}\varepsilon,$$

or, with $H_n = K_n \times S^{I\setminus J_n}$,

$$\mu_0(H_n) \geq \mu_0(B_n) - 2^{-n}\varepsilon. \tag{3.2.6}$$

Now, under the hypothesis of the lemma, for all $n \in \mathbb{N}$,

$$\mu_0(H_1 \cap \cdots \cap H_n) \geq \mu_0(B_1 \cap \cdots \cap B_n) - \sum_{i=1}^{n} \mu_0(B_i \backslash H_i) \geq 2\varepsilon - \varepsilon \sum_{i=1}^{\infty} 2^{-i} = \varepsilon. \tag{3.2.7}$$

In particular, for any n, $H_1 \cap \cdots \cap H_n \neq \emptyset$. Now let $x_n \in H_1 \cap \cdots \cap H_n$, and hence $\pi_{J_k} x_n \in K_1 \cap \cdots \cap K_k$, for any $k \leq n$. By compactness of this set, there exist a subsequence, n_i, such that $\lim_{i \to \infty} \pi_{J_k} x_{n_i} \in \bigcap_{j=1}^{k} K_j$.

Taking subsequently sub-subsequences[2], we can construct a sequence in such a way that $\pi_{J_k} x_{n_i} \to x^k \in \bigcap_{j=1}^{k} K_j$ for all k. Clearly, $\pi_{J_\ell} x^k = x^\ell$, for all $\ell \leq k$. Then there exist an $x \in S^I$ whose projections are equal to these limits for all k and hence $x \in \cap_{j=1}^{k} B_j$ for all k, hence $x \in \cap_{n=1}^{\infty} B_n$ and so $\bigcap_{n \in \mathbb{N}} B_n \neq \emptyset$. But this is the claim of the lemma. $\square$

So we are done: μ_0 is σ-additive on the algebra of finite dimensional cylinders, and so there exists a unique probability measure on the σ-algebra $\mathfrak{B}^I$ with the advertised properties. $\square$

Remark. Note that we have used the assumption on the space S only to ensure that the measures μ^J, for $J \in F(I)$, are all inner regular. Thus we can replace the assertion of the theorem by:

Theorem 3.7. *Let S be a topological space, and let $\mathfrak{B} = \mathfrak{B}(S)$ be its Borel-σ-algebra. Let I be a set. Suppose that, for each $J \in F(I)$, there exists an inner regular probability measure, μ^J, on $(S^J, \mathfrak{B}^J)$, such that for any $J_1 \subset J_2 \in F(I)$,*

$$\mu^{J_1} = \mu^{J_2} \circ \pi_{J_1}^{-1}, \tag{3.2.8}$$

where π_{J_1} denotes the canonical projection from $S^{J_2} \to S^{J_1}$. Then there exists *a* unique *measure, μ, on $(S^I, \mathfrak{B}^I)$, such that, for all $J \in F(I)$,*

$$\mu \circ \pi_J^{-1} = \mu^J. \tag{3.2.9}$$

Finally, one can show that the assumption that Ω be compact and metrisable in Theorem 1.23 can be replaced by assuming that Ω be Polish[3]. In fact by inspecting the proof one sees that if we replace the requirement "compact" by "closed", then the compactness requirement on Ω is no longer needed. Thus all what remains to be seen is that the closed sets $K = K_\varepsilon$ that approximate B well from within can be chosen bounded on a separable metric space. But this follows for instance since on a metric space, $\mathbb{P}(B_n(x)) \uparrow 1$, where $B_n(x)$ is the closed metric ball of radius n

[2] This is possible because the subsequences for $k+1$ is a sub-subsequence for k, due to $\cap_{j=1}^{k} K_j \supset \cap_{j=1}^{k+1} K_j$. Indeed, denote $(n_{i,k})_{i \geq 1}$ the subsequence for k. Applying the Cantor diagonal procedure, i.e., taking $(n_{k,k})_{k \geq 1}$ provides the desired subsequence.

[3] The notations here are as in Theorem 1.23.

around x and using instead of K_ε simply $\tilde{K}_\varepsilon = K_{\varepsilon/2} \cap B_{n_\varepsilon}(0)$ with n_ε chosen such that $\mathbb{P}(B_{n_\varepsilon}(0)) \geq 1 - \varepsilon/2$, so that $\mathbb{P}(B \setminus \tilde{K}_\varepsilon) < \varepsilon$ and $\tilde{K}_\varepsilon$ is a closed and bounded set.

Remark. Note that we have seen no need to distinguish cases according to the nature of the set I.

3.3 Examples of stochastic processes

The Kolmogorov-Daniell theorem goes a long way in helping to construct stochastic processes. However, one should not be deceived: prescribing a *consistent family of finite dimensional distributions* (i.e. distributions satisfying (3.2.3)) is by no means an easy task and in practise we want to have a simpler way of describing a stochastic process of our liking.

In this section I discuss some of the most important classes of examples without going into too much detail.

3.3.1 Independent random variables

We have of course already encountered independent random variables in the first course of probability. We can now formulate the existence of independent random variables in full generality and with full rigour.

Theorem 3.8. *Let I be a set and let, for each $t \in I$, μ_t be a probability measure on $(S, \mathfrak{B}(S))$, where S is a polish space. Then there exists a unique probability measure, μ, on $(S^I, \mathfrak{B}^I)$, such that, for $J \in F(I)$, and $A_t \in \mathfrak{B}$,*

$$\mu\left(\bigcap_{t \in J} \pi_t^{-1}(A_t)\right) = \prod_{t \in J} \mu_t(A_t). \tag{3.3.1}$$

Proof. Under the hypothesis that S is polish, the proof is direct from the Kolmogorov-Daniell theorem. Note that this hypothesis is not, however, necessary. $\square$

Remark. Note that we don't assume I to be countable. In the case when I is uncountable, such a collection of random variables is sometimes called *white noise*. This is, however, a rather unpleasant object. When we discuss seriously the issue of stochastic processes with continuous time, we will see that we always will want additional properties of sample paths that the theorem above does not provide.

Independent random variables are a major building block for more interesting stochastic processes. We have already encountered sums of independent random variables. Other interesting processes are e.g. maxima of independent random variables: If $X_i, i \in \mathbb{N}$ are independent random variables, define

$$M_n = \max_{1\leq k\leq n} X_k. \tag{3.3.2}$$

The study of such maxima is an interesting topic in itself.

Of course one can look at many more functions of independent random variables.

3.3.2 Gaussian processes

Gaussian processes are one of the most important class of stochastic process that can be defined with the help of densities. Let us proceed in two steps.

First, we consider finite dimensional Gaussian vectors. Let $n \in \mathbb{N}$ be fixed, and let C be a real symmetric *positive definite* $n \times n$ matrix. We denote by C^{-1} its inverse. Define the *Gaussian density*,

$$f_C(x_1,\dots,x_n) \equiv \frac{1}{(2\pi)^{n/2}\sqrt{\det C}} \exp\left(-\frac{1}{2}(x, C^{-1}x)\right). \tag{3.3.3}$$

You see that the necessity of having C positive derives from the fact that we want this density to be integrable with respect to the n-dimensional Lebesgue measure.

Definition 3.9. A family of n real random variables is called jointly Gaussian with mean zero and covariance C, if and only if their distribution is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^n$ with density given by f_C.

Remark. In this section I will always consider only Gaussian random variables with mean zero. The corresponding expressions in the general case can be recovered by simple computations.

The definition of Gaussian vectors is no problem. The question is, can we define Gaussian processes? From what we have learned, it will be crucial to be able to define compatible families of finite dimensional distributions.

The following result will be important.

Lemma 3.10. *Let $X_1,\dots,X_n$ be random variables whose joint distribution is Gaussian with density covariance matrix C and mean zero.*

(i) For any $k,\ell \in \{1,\dots,n\}$,

$$\mathbb{E}(X_k X_\ell) = C_{k,\ell}. \tag{3.3.4}$$

(ii)If $J \subset \{1,\dots,n\}$ with $|J| = m$, then the random variables $X_\ell, \ell \in J$ are jointly Gaussian with covariance given by the $m \times m$-matrix C^J with elements $C^J_{k,\ell} = C_{k,\ell}$, if $k,\ell \in J$.

Proof. For technical reasons it is very convenient to compute the moment generating function, or the Laplace transform, of our jointly Gaussian vector. We define, for $u \in \mathbb{C}^n$,

$$\phi_C(u) \equiv \mathbb{E}\left(e^{(u,X)}\right) \equiv \mathbb{E}\left(e^{\sum_{i=1}^n u_i X_i}\right) = \int d^n x f_C(x_1,\dots,x_n) e^{\sum_{i=1}^n u_i x_i}. \tag{3.3.5}$$

It is easy to see that this integral is always finite. Its computation involves a nice trick, that is well worth remembering! To understand it, recall that a real positive matrix can always be written in the form $C = A^t A$, where A^t denotes the transpose of A, and A is itself invertible. Then likewise $C^{-1} = A^{-1}(A^t)^{-1}$. For simplicity we write $B = (A^t)^{-1}$.

$$\begin{aligned}
\phi_C(u) &= \frac{1}{(2\pi)^{n/2}\sqrt{\det C}} \int d^n x \exp\left(-\frac{1}{2}(x, C^{-1}x) + (u,x)\right) \qquad (3.3.6)\\
&= \frac{1}{(2\pi)^{n/2}\sqrt{\det C}} \int d^n x \exp\left(-\frac{1}{2}(Bx, Bx) + (u,x)\right)\\
&= \frac{1}{(2\pi)^{n/2}\sqrt{\det C}} \int d^n x \exp\left(-\frac{1}{2}(Bx - Au, Bx - Au) + \frac{1}{2}(Au, Au)\right)\\
&= \frac{\exp\left(-\frac{1}{2}(u, Cu)\right)}{(2\pi)^{n/2}\sqrt{\det C}} \int d^n x \exp\left(-\frac{1}{2}(x - Cu, C^{-1}(x - Cu))\right)\\
&= \exp\left(-\frac{1}{2}(u, Cu)\right),
\end{aligned}$$

where in the last line we used that the domain of integration in the integral is invariant under translation.

Now it is easy to compute the correlation function. Clearly,

$$\mathbb{E}(X_k X_\ell) = \left.\frac{d^2 \phi_C(u)}{du_k du_\ell}\right|_{u=0} = C_{k,\ell}.$$

This establishes (i). (ii) is now quite simple. To compute the Laplace transform of the vector $X_\ell, \ell \in J$, we just need to set $u_i = u_i^J$ for $i \in J$ and $u_i = 0$ for $i \notin J$. The result is precisely the Laplace transform of a Gaussian vector with covariance C^J. Since the Laplace transform determines the distribution uniquely, (ii) follows. □

This result if very encouraging for the prospect of defining Gaussian vectors. If we can specify an infinite dimensional positive matrix, C then all its finite dimensional sub-matrices, C^J, $J \in F(\mathbb{N})$, are positive and the ensuing family of finite dimensional distributions are Gaussian distributions that do satisfy the consistency requirements of Kolmogorov's theorem! The result is:

Theorem 3.11. *Let C be a symmetric positive quadratic form on $\mathbb{R}^{\mathbb{N}}$. Then there exists a unique Gaussian process with index set $\mathbb{N}$, state space $\mathbb{R}$, such that, for all finite $J \subset \mathbb{N}$, the marginal distributions are $|J|$-dimensional Gaussian vectors with covariance C^J.*

Thus the trick is to construct positive quadratic forms. Of course is easy to guess a few by going the other way, and using independent Gaussian random variables as building blocks. For example, consider X_n, $n \in \mathbb{N}$ to be independent, Gaussian random variables with mean zero and variance σ_n^2. Set $Z_n \equiv \sum_{k=1}^n X_k$. Then

$$C_{n,m} \equiv \mathbb{E}(Z_n Z_m) = \mathbb{E}\left(\sum_{i=1}^n \sum_{j=1}^m X_i X_j\right) = \sum_{i=1}^{m\wedge n} \mathbb{E}(X_i^2) = \sum_{i=1}^{m\wedge n} \sigma_i^2.$$

Thus the quadratic form $C_{n,m} = \sum_{i=1}^{m\wedge n} \sigma_i^2$ is apparently positive. In fact, if $u \in \mathbb{R}^{\mathbb{N}}$,

$$\begin{aligned}(u, Cu) &= \sum_{n,m\in\mathbb{N}} u_n u_m \sum_{i=1}^{m\wedge n} \sigma_i^2 = \sum_{i\in\mathbb{N}} \sigma_i^2 \sum_{m\geq i} u_m \sum_{n\geq i} u_n \\ &= \sum_{i\in\mathbb{N}} \sigma_i^2 \left(\sum_{m\geq i} u_m\right)^2 \geq 0\end{aligned}$$

and it is equal to zero if and only if $u = 0$.

Now we have seen that in the construction of stochastic processes, the fact to have discrete time did not appear (so far) to be much of an advantage. Thus the above example may make us courageous to attempt to define a Gaussian process on $\mathbb{R}_+$. To this end, define a function $C : \mathbb{R}_+ \times \mathbb{R}_+ \mapsto \mathbb{R}_+$ by

$$C(t,s) \equiv t \wedge s. \tag{3.3.7}$$

What we have to check is that, for any $J \in F(\mathbb{R}_+)$, the restriction of C to a quadratic form on $\mathbb{R}^J$ is positive. But indeed,

$$\sum_{t,s\in J} u_t u_s (t\wedge s) = \sum_{t,s\in J} u_t u_s \int_0^{(t\wedge s)} 1\, dr = \int_0^\infty dr \left(\sum_{t\in J, t\geq r} u_t\right)^2 \geq 0.$$

Thus all finite dimensional distributions exist as Gaussian vectors, and the compatibility conditions are trivially satisfied. Therefore there exists a Gaussian process on $\mathbb{R}_+$ with this covariance. This process is called "Brownian motion". *Note, however, that this constructs the process only in the product topology, which does not yet yield any nice path properties.* We will later see that this process can actually be constructed on the space of continuous functions, and this object will then more properly called Brownian motion.

Exercise. Let $X_k, k \in \mathbb{N}$, be independent Gaussian random variables with mean zero and variance $\sigma^2 = 1$. Define, for $n \in \mathbb{N}$, and $t \in [0,1]$,

$$Z_n(t) \equiv \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} X_k,$$

where $[\cdot]$ denotes the largest integer smaller than $\cdot$. Show that

(i) $Z_n(t)$ is a stochastic process with indexset $[0,1]$ and state space $\mathbb{R}$.
(ii)Compute the covariance, C_n, of Z_n and show that for any $I \in F([0,1])$, $C_n^I \to C^I$, where $C(s,t) = s \wedge t$.

(iii) Show that the finite dimensional distributions of the processes Z_n converge, as $n \to \infty$, to those of the "Brownian motion" defined above.
(iv) Show that the results $(i)-(iii)$ remain true if instead of requiring that the X_k are Gaussian we just assume that their variance equals to 1.

Note that to prove (iv), you need to prove the multi-dimensional analogue of the central limit theorem. This requires, however, little more than an adaptation of the notation from the standard CLT in dimension one.

3.3.3 Markov processes

Gaussian processes were build from independent random variables using densities. Another important way to construct non-trivial processes uses conditional probabilities. Markov processes are the most prominent examples. In the case of Markov processes we really think of the index set, $\mathbb{N}_0$ or $\mathbb{R}_+$, as time. The process X_t then shall have two properties: (1) it should be *causal*, i.e. we want an expression for the law of X_t given the σ-algebra $\mathfrak{F}_{t_-} \equiv \sigma(X_s, s < t)$, (2) we want this law to be *forgetful of the past*: if we know the position (value; we will think mostly of a Markov process as a "particle" moving around in S) of X at some time $s < t$, then the law of X_t should be independent of the positions of $X_{s'}$ with $s' < s$. In a way, Markov processes are meant be the stochastic analogues of deterministic evolution (differential equations).

To set such a process up, let us consider the (much simpler) case of discrete time, i.e. $I = \mathbb{N}_0$ (we always want zero in our index set). The main building block for a Markov chain is then the so called (one-step) transition kernel, $\mathscr{P} : \mathbb{N}_0 \times S \times \mathfrak{B} \to [0,1]$, with the following properties:

(i) For each $t \in \mathbb{N}_0$ and $x \in S$, $\mathscr{P}_t(x, \cdot)$ is a probability measure on $(S, \mathfrak{B})$.
(ii) For each $A \in \mathfrak{B}$, and $t \in \mathbb{N}_0$, $\mathscr{P}_t(\cdot, A)$ is a $\mathfrak{B}$-measurable function on S.

Then, a stochastic process X with state space S and index set $\mathbb{N}_0$ is a discrete time Markov process with law $\mathbb{P}$, if, for all $A \in \mathfrak{B}$, $t \in \mathbb{N}$,

$$\mathbb{P}(X_t \in A | \mathfrak{F}_{t-1})(\omega) = \mathscr{P}_{t-1}(X_{t-1}(\omega), A), \quad \mathbb{P}-\text{a.s.} \tag{3.3.8}$$

The remarkable thing is that this requirement fixes the law $\mathbb{P}$ up to one more probability measure on $(S, \mathfrak{B})$, the so-called *initial distribution*, P_0.

Theorem 3.12. *Let $(S, \mathfrak{B})$ be a Polish space, let $\mathscr{P}$ be a transition kernel and P_0 an probability measure on $(S, \mathfrak{B})$. Then there exists a unique stochastic process satisfying (3.3.8) and $\mathbb{P}(X_0 \in A) = P_0(A)$, for all A.*

Proof. In view of the Kolmogorov-Daniell theorem, we have to show that our requirements fix all finite dimensional distributions, and that these satisfy the compatibility conditions. This is more a problem of notation than anything else. We will need to be able to derive formulas for

$$\mathbb{P}\left(X_{t_n} \in A_n, \dots, X_{t_1} \in A_1\right).$$

To get started, we consider

$$\mathbb{P}\left(X_t \in A | \mathfrak{F}_s\right),$$

for $s < t$. To do this, we use that by the elementary properties of conditional expectations (we drop the a.s. that applies to all equations relating to conditional expectations).

$$\begin{aligned} \mathbb{P}\left(X_t \in A | \mathfrak{F}_s\right) &= \mathbb{E}\left[\mathbb{P}\left(X_t \in A | \mathfrak{F}_{t-1}\right) | \mathfrak{F}_s\right] \\ &= \mathbb{E}\left[\mathscr{P}_{t-1}(X_{t-1}(\omega), A) | \mathfrak{F}_s\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathscr{P}_{t-1}(X_{t-1}(\omega), A) | \mathfrak{F}_{t-2}\right] | \mathfrak{F}_s\right] \end{aligned} \tag{3.3.9}$$

where we used $\mathfrak{F}_s \subset \mathfrak{F}_{s'}$ for all $s < s'$. Further,

$$\begin{aligned} (3.3.9) &= \mathbb{E}\left[\mathbb{E}\left[\int \mathscr{P}_{t-1}(x_{t-1}, A)\mathscr{P}_{t-2}(X_{t-2}(\omega), dx_{t-1}) \Big| \mathfrak{F}_{t-2}\right] \Big| \mathfrak{F}_s\right] \\ &= \mathbb{E}\Bigg[\int \mathscr{P}_{t-1}(x_{t-1}, A)\mathscr{P}_{t-2}(x_{t-2}, dx_{t-1}) \cdots \\ &\qquad \cdots \mathscr{P}_{s+1}(x_{s+1}, dx_{s+2})\mathscr{P}_s(X_s(\omega), dx_{s+1}) \Big| \mathfrak{F}_s\Bigg] \\ &= \int \mathscr{P}_{t-1}(x_{t-1}, A)\mathscr{P}_{t-2}(x_{t-2}, dx_{t-1}) \cdots \mathscr{P}_s(X_s(\omega), dx_{s+1}) \end{aligned}$$

since X_s is $\mathfrak{F}_s$ measurable. We will set

$$P_{s,t}(x, A) \equiv \int \mathscr{P}_{t-1}(x_{t-1}, A)\mathscr{P}_{t-2}(x_{t-2}, dx_{t-1}) \cdots \mathscr{P}_s(x, dx_{s+1}) \tag{3.3.10}$$

and call $P_{s,t}$ the transition kernel from time s to time t. With this object defines, we can now proceed to more complicated expressions:

$$\begin{aligned} &\mathbb{P}\left(X_{t_n} \in A_n, \dots, X_{t_1} \in A_1\right) \\ &= \mathbb{E}\left[\mathbb{P}(X_{t_n} \in A_n | \mathfrak{F}_{t_{n-1}})\mathbb{1}_{A_{n-1}}(X_{t_{n-1}}) \cdots \mathbb{1}_{A_1}(X_{t_1})\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[P_{t_{n-1},t_n}(X_{t_{n-1}}(\omega), A_n) | \mathfrak{F}_{t_{n-1}}\right] \mathbb{1}_{A_{n-1}}(X_{t_{n-1}}) \cdots \mathbb{1}_{A_1}(X_{t_1})\right] \\ &= \mathbb{E}\Bigg[\mathbb{E}\left[\int_{A_{n-1}} P_{t_{n-1},t_n}(x_{n-1}, A_n) P_{t_{n-2},t_{n-1}}(X_{t_{n-2}}(\omega), dx_{n-1}) \Big| \mathfrak{F}_{t_{n-2}}\right] \\ &\qquad \times \mathbb{1}_{A_{n-2}}(X_{t_{n-2}}) \cdots \mathbb{1}_{A_1}(X_{t_1})\Bigg] \\ &= \int_{A_{n-1}} P_{t_{n-1},t_n}(x_{n-1}, A_n) \int_{A_{n-2}} P_{t_{n-2},t_{n-1}}(x_{n-2}, dx_{n-1}) \\ &\qquad \cdots \int_{A_1} P_{t_1,t_2}(x_1, dx_2) \int_S P_{0,t_1}(x_0, dx_1) P_0(dx_0). \end{aligned} \tag{3.3.11}$$

Thus, we have the desired expression of the marginal distributions in terms of the transition kernel P and the initial distribution P_0. The compatibility relations follow from the following obvious, but important property of the transition kernels.

Lemma 3.13. *The transition kernels $P_{s,t}$ satisfy the* Chapman-Kolmogorov *equations*

$$P_{s,t}(x,A) = \int P_{r,t}(y,A)P_{s,r}(x,dy) \tag{3.3.12}$$

for any $s < r < t$.

Proof. This is obvious from the definition. □

The proof of the compatibility relations is now also obvious; if some of the A_i are equal to S, we can use (3.3.12) and recover the expressions for the lower dimensional marginals. □

Exercise. Consider the Brownian motion process from the last sub-section. Show that this process is Markov in the sense that all finite dimensional distributions satisfy the Markov property.
Hint: Let $J = t_n > t_{n-1} > \cdots > t_1$. Show that the family of random variables $Y_n \equiv X_{t_n} - X_{t_{n-1}}, X_{t_{n-1}}, X_{t_{n-2}}, \ldots, X_{t_1}$ are jointly Gaussian and that Y_n is independent of the σ-algebra generated by $X_{t_{n-1}}, \ldots, X_{t_1}$.

3.3.4 Gibbs measures

As an aside, I will briefly explain another important way to construct stochastic processes with the help of conditional expectations and densities, that is central in *statistical mechanics*. It is particularly useful in the setting where I is not an ordered set, the most prominent example is $I = \mathbb{Z}^d$.

In order not to introduce too much notation, I will stick to a simple example, the so-called *Ising-model*. In this case, $S = \{-1, 1\}$. The main object is family of functions, $H_\Lambda : S^{\mathbb{Z}^d} \to \mathbb{R}$, called *Hamiltonians*, that are defined for every finite $\Lambda \subset \mathbb{Z}^d$, and are given by

$$H_\Lambda(X) = -\sum_{i,j: i \vee j \in \Lambda} X_i X_j J_{ij}. \tag{3.3.13}$$

Using this function, we will construct a family of probability kernels, μ_Λ, that have the following properties:

(i) For each $y \in S^{\mathbb{Z}^d}$, $\mu_\Lambda(\cdot, \tau)$ is a probability measure on $S^{\mathbb{Z}^d}$;
(ii)For each $A \in \mathfrak{B}^{\mathbb{Z}^d}$, $\mu_\Lambda(A, \cdot)$ is a $\mathfrak{F}_{\Lambda^c}$-measurable function, where $\mathfrak{F}_{\Lambda^c} = \sigma(X_i, i \in \Lambda^c)$;
(iii)For any pair of volumes, Λ, Λ', with $\Lambda \subset \Lambda'$, and any $A \in \mathfrak{B}^{\mathbb{Z}^d}$,

$$\int \mu_\Lambda(z,A)\mu_{\Lambda'}(x,dz) = \mu_{\Lambda'}(x,A). \tag{3.3.14}$$

We will indeed give an explicit formula for μ_Λ:

$$\mu_\Lambda(A,y) = \frac{\sum_{x_i, i\in\Lambda} \mathbb{1}_{(x_\Lambda, y_{\Lambda^c})\in A} e^{-\beta H_\Lambda((x_\Lambda, y_{\Lambda^c}))}}{\sum_{x_i, i\in\Lambda} e^{-\beta H_\Lambda((x_\Lambda, y_{\Lambda^c}))}}. \tag{3.3.15}$$

It is easily checked that this expression indeed defines a kernel with properties (i) and (ii). An expression of this type is called a *local Gibbs specification*.

Now we see that the properties of these kernels are reminiscent of those of regular conditional probabilities.

One defines the notion of a Gibbs measure as follows:

Definition 3.14. A probability measure on $S^{\mathbb{Z}^d}$ is called a Gibbs measures, if and only if, for any finite $\Lambda \subset \mathbb{Z}^d$, the kernel μ_Λ is a regular conditional probabilty for μ given $\mathfrak{F}_{\Lambda^c}$.

More specifically, if the kernel is the Gibbs specification (3.3.15), it will be called a Gibbs measures for the d-dimensional Ising model at temperature β^{-1}.

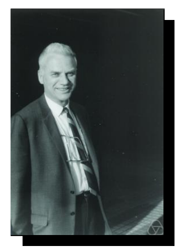
One can prove that such Gibbs measures exist; for this one shows that any accumulation point of a sequence $\mu_{\Lambda_n}(\cdot,x)$, where $\Lambda_n \uparrow \mathbb{Z}^d$ is any increasing sequence of volumes that converges to $\mathbb{Z}^d$ (in the sense, that, for any finite Λ, there exists n_0, such that, for all $n \geq n_0$, $\Lambda \subset \Lambda_n$), will be a Gibbs measure. This is relatively straightforward, by writing equation (3.3.14) for a sequence of volumes $\Lambda_n \uparrow \mathbb{Z}^d$:

$$\int \mu_\Lambda(z,A)\mu_{\Lambda_n}(x,dz) = \mu_{\Lambda_n}(x,A).$$

If μ_{Λ_n} conveges weakly to some measure μ, then the right-hand side converges to $\mu(A)$. The left-hand side will converge to $\int \mu_\Lambda(z,A)\mu(x,dz)$, since one can easily see that $\mu_\Lambda(z,A)$ is a continuous function, if A is a cylinder event (in fact, in our example, it is a local function on a discrete space). But then μ satisfies the desired properties of a Gibbs measure.

The existence of accumulation points is then guaranteed by the fact that $S^{\mathbb{Z}^d}$ is compact (Tychonov, since $S = \{-1,1\}$ is compact), and that the set of probability measures over a compact space is compact. What makes this setting interesting is that there is no general uniqueness result. In fact, if $d \geq 2$, and $\beta > \beta_c$, for a certain β_c, then it is known that there exists more than one Gibbs measure. This mathematical fact is connected to the physical phenomenon of a so-called *phase transition*, and this is what makes the study of Gibbs measures so interesting. For deeper material on Gibbs measures see [3, 13, 6].

Chapter 4
Martingales



In this chapter we introduce the fundamental concept of *martingales*, which will keep playing a central rôle in our investigation of stochastic processes. Martingales are “truly random” stochastic processes, in the sense that their observation in the past does not allow for useful prediction of the future. By useful we mean here that no gambling strategies can be devised that would allow for systematic gains. In thi chapter we will always assume that random variables take values in $\mathbb{R}$, unless specified otherwise. The treatment of martingales follows largely the book of Rogers and Williams [12], with the exception of a section on the central limit theorem, which is inspired by Billingsley's presentation[2].

4.1 Definitions

We begin by formally introducing the notion of a filtration of a σ-algebra that we have already briefly encountered in the context of Markov processes. We remain in the context of discrete index sets.

Definition 4.1. Let $(\Omega, \mathfrak{F})$ be a measurable space. A family of sub-σ-algebras, $\{\mathfrak{F}_n, n \in \mathbb{N}_0\}$ of $\mathfrak{F}$ that satisfies

$$\mathfrak{F}_0 \subset \mathfrak{F}_1 \subset \cdots \subset \cdots \mathfrak{F}_\infty \equiv \sigma\left(\bigcup_{n\in\mathbb{N}_0} \mathfrak{F}_n\right) \subset \mathfrak{F}, \tag{4.1.1}$$

is called a *filtration* of the σ-algebra $\mathfrak{F}$. We call a quadruple $(\Omega, \mathfrak{F}, \mathbb{P}, \{\mathfrak{F}_n, n \in \mathbb{N}_0\})$ a filtered (probability) space.

In this chapter we will henceforth always assume that we are given a filtered space.

Filtrations and stochastic processes are closely linked. We will see that this goes in two ways.

Definition 4.2. A stochastic process, $\{X_n, n \in \mathbb{N}_0\}$, is called *adapted* to the filtration $\{\mathfrak{F}_n, n \in \mathbb{N}_0\}$, if, for every n, X_n is $\mathfrak{F}_n$-measurable.

Now the other direction:

Definition 4.3. Let $\{X_n, n \in \mathbb{N}_0\}$ be a stochastic process on $(\Omega, \mathfrak{F}, \mathbb{P})$. The natural filtration, $\{\mathscr{W}_n, n \in \mathbb{N}_0\}$ with respect to X is the smallest filtration such that X is adapted to it, that is,

$$\mathscr{W}_n = \sigma(X_0, \dots, X_n). \tag{4.1.2}$$

We see that the basic idea of the natural filtration is that functions of the process that are measurable with respect to $\mathscr{W}_n$ depend only on the observations of the process up to time n.

We now define martingales.

Definition 4.4. A stochastic process, X, on a filtered space is called a *martingale*, if and only if the following hold:

(i) The process X is adapted to the filtration $\{\mathfrak{F}_n, n \in \mathbb{N}_0\}$;
(ii)For all $n \in \mathbb{N}_0$, $\mathbb{E}(|X_n|) < \infty$;
(iii)For all $n \in \mathbb{N}$,

$$\mathbb{E}(X_n | \mathfrak{F}_{n-1}) = X_{n-1}, \text{ a.s..} \tag{4.1.3}$$

If (i) and (ii) hold, but instead to (iii), it holds $\mathbb{E}(X_n|\mathfrak{F}_{n-1}) \geq X_{n-1}$, respectively $\mathbb{E}(X_n|\mathfrak{F}_{n-1}) \leq X_{n-1}$, then the process X is called a *sub-martingale*, respectively a *super-martingale*.

In particular, for a martingale it holds $\mathbb{E}(X_n) = \mathbb{E}(X_{n-1})$, for a sub-martingale $\mathbb{E}(X_n) \geq \mathbb{E}(X_{n-1})$, finally, for a super-martingale $\mathbb{E}(X_n) \leq \mathbb{E}(X_{n-1})$.

It is clear that the property (iii) is what makes martingales special: intuitively, it means that the best guess for what X_n could be, knowing what happened up to time $n-1$ is simply X_{n-1}. No prediction on the direction of change is possible.

We will now head for the fundamental theorem concerning the impossibility of winning systems in games build on martingales.

To put us into the gambling mood, we think of the increments of the process, $Y_n \equiv X_n - X_{n-1}$, as the result of (not necessarily independent) games (Examples: (i) Coin tosses, or (ii) the daily increase of the price of a stock). We are allowed to bet on the outcome in the following way: at each moment in time, $n-1$, we choose a number $C_n \in \mathbb{R}$. Then our wealth will increase by the amount $C_n Y_n$, i.e. the wealth process, W_n is given by $W_n = \sum_{k=1}^n C_n Y_n$ (Example: (i) in the coin toss, case, choose $C_n > 0$ means to bet on head ($= \{Y_n = +1\}$) an amount C_n, and $C_n < 0$ means to bet on the outcome tails ($= \{Y_n = -1\}$) the amount $-C_n$; (ii) in the stock case, C_n represents the amount of stock an investor decides to hold at time $n-1$ up to time n (here negative values can be realised by short-selling).

The choice of the C_n is done knowing the process up to time $n-1$. This justifies the following definition.

Definition 4.5. A stochastic process $\{C_n, n \in \mathbb{N}\}$ is called *previsible*[1], if, for all $n \in \mathbb{N}$, C_n is $\mathfrak{F}_{n-1}$-measurable.

Given an adapted stochastic process, X and a previsible process C, we can define the wealth process

$$W_n \equiv \sum_{k=1}^{n} C_k (X_k - X_{k-1}) \equiv (C \bullet X)_n. \tag{4.1.4}$$

Definition 4.6. The process $C \bullet X$ is called the *martingale transform* of X by C or the *discrete stochastic integral* of C with respect to X.

Now we can formulate the general "no-system" theorem for martingales:

Theorem 4.7. *Let $(\Omega, \mathfrak{F}, \mathbb{P}, \{\mathfrak{F}_n, n \in \mathbb{N}\})$ be a filtered space.*

(i) Let C be a bounded non-negative previsible process such that there exists $K < \infty$, such that, for all n, and all $\omega \in \Omega$, $|C_n(\omega)| \leq K$. Let X be a super-martingale. Then $C \bullet X$ is a super-martingale that vanishes for $n = 0$.

(ii) Let C be a bounded previsible process (boundedness as above) and X be a martingale. Then $C \bullet X$ is a martingale that vanishes at zero.

(iii) Both in (i) and (ii), the condition of boundedness can be replaced by $C_n \in \mathscr{L}^2$, if also $X_n \in \mathscr{L}^2$.

Remark. In terms of gambling, (i) says that, if the underlying process has a tendency to fall, then playing against the trend ("investing in a falling stock") leads to a wealth process that tends to fall. On the other hand, (ii) says that, if the underlying process X is a martingale, then no matter what strategy you use, the wealth process has mean zero.

Proof. (i) and (ii). To check integrability it is trivial. We also have that $W_n - W_{n-1} = C_n(X_n - X_{n-1})$. Then

$$\mathbb{E}(W_n - W_{n-1} | \mathfrak{F}_{n-1}) = C_n \mathbb{E}(X_n - X_{n-1} | \mathfrak{F}_{n-1}), \tag{4.1.5}$$

by Lemma 2.5. If X is a martingale, the conditional expectation on the right is zero, so $\mathbb{E}(W_n - W_{n-1} | \mathfrak{F}_{n-1}) = 0$, and W is a martingale. If X is a super-martingale, the conditional expectation is non-positive and this remains true for the product, if C_n is non-negative. This proves (i) and (ii).

To prove (iii), we just need to show that under the hypothesis of (iii), Eq. (4.1.5) still holds. But first, by the Cauchy-Schwartz inequality, $C_n(X_n - X_{n-1})$ is absolutely integrable. Next, chose since the bounded functions are dense in L^2, take a sequence of bounded functions C_n^k that converge to C_n. Then

$$\mathbb{E}(W_n - W_{n-1} | \mathfrak{F}_{n-1}) = C_n^k \mathbb{E}(X_n - X_{n-1} | \mathfrak{F}_{n-1}) + \mathbb{E}((C_n - C_n^k)(X_n - X_{n-1}) | \mathfrak{F}_{n-1}). \tag{4.1.6}$$

Again by Cauchy-Schwartz, the second term tends to zero as $k \uparrow \infty$, while the first tends to $C_n \mathbb{E}(X_n - X_{n-1} | \mathfrak{F}_{n-1})$, almost surely. □

---

[1] The teminology previsible refers to the fact that C_n can be foreseen from the information available at time $n-1$.

The quantities $Y_n = X_n - X_{n-1}$ are called *martingale differences*. A sequence $S_n \equiv \sum_{k=1}^n Y_n$ where $\mathbb{E}(Y_n|\mathfrak{F}_{n-1}) = 0$ is called a *martingale difference sequence*. If Y_n are square integrable, then the variance of a martingale difference sequence satisfies

$$\mathbb{E}(S_n^2) = \sum_{k=1}^n \mathbb{E}(Y_k^2). \tag{4.1.7}$$

Some examples.

A canonical way to construct a martingale is to take any random variable, X, on a filtered probability space, $(\Omega, \mathfrak{F}, \mathscr{P}, \{\mathfrak{F}_n, n \in \mathbb{N}_0\})$, and to define

$$X_n \equiv \mathbb{E}(X|\mathfrak{F}_n).$$

Then, by the properties of conditional expectation,

$$\mathbb{E}(X_n|\mathfrak{F}_{n-1}) = \mathbb{E}[\mathbb{E}(X|\mathfrak{F}_n)|\mathfrak{F}_{n-1}] = \mathbb{E}(X|\mathfrak{F}_{n-1}) = X_{n-1}, \text{a.s.}$$

In this case, we should expect that $\lim_{n\to\infty} X_n = X$, a.s..

Another example is a Markov chain whose transition kernel has the property that

$$\int xP(y,dx) = y.$$

In particular, sums of iid random variables with mean zero are martingales.

4.2 Upcrossings and convergence

Consider an interval $[a,b]$. We want to count the number of times a process crosses this interval from below.

Definition 4.8. Let $a < b \in \mathbb{R}$ and let X_s be a stochastic process with values in $\mathbb{R}$. We say that an upcrossing of $[a,b]$ occurs between times s and t, if

(i) $X_s < a, X_t > b$,
(ii for all r such that $s < r < t, X_r \in [a,b]$.

We denote by $U_N(X,[a,b])(\omega)$ the number of uprossings in the time interval $[0,N]$.

We will now consider a (obviously) previsible process constructed as follows:

$$C_1 = \mathbb{1}_{X_0<a}; \quad C_n = \mathbb{1}_{C_{n-1}=1}\mathbb{1}_{X_{n-1}\leq b} + \mathbb{1}_{C_{n-1}=0}\mathbb{1}_{X_{n-1}<a}, \text{ for } n \geq 2. \tag{4.2.1}$$

This process represents a “winning” strategy: wait until the process (say, price of ....) drops below a. Buy the stock, and hold it until its price exceeds b; sell, wait until the price drops below a, and so on. Our wealth process is $W = C \bullet X$.

Now each time there is an upcrossing of $[a,b]$ we win at least $(b-a)$. Thus, at time N, we have

$$W_N \geq (b-a)U_N(X,[a,b]) - |a-X_N|\mathbb{1}_{X_N<a}, \tag{4.2.2}$$

where the last term count is the maximum loss that we could have incurred if we are invested at time N and the price is below a.

Naive intuition would suggest that in the long run, the first term must win. Our theorem above says that this is false, if we are in a fair or disadvantageous game (that is, in practice, always).

Theorem 4.9 (Doob's upcrossing lemma). *Let X be a super-martingale.. Then Then for any $a<b\in\mathbb{R}$,*

$$(b-a)\,\mathbb{E}(U_N(X,[a,b])) \leq \mathbb{E}\left(|a-X_N|\mathbb{1}_{X_N<a}\right). \tag{4.2.3}$$

Proof. The process C defined in (4.2.1) is a bounded, non-negative previsible process. Therefore (i) of Theorem 4.7 implies that $W\equiv C\bullet X$ is super-martingale with $W_0=0$. Therefore $0\geq \mathbb{E}W_N$ and taking the expectation of (4.2.2) gives (4.2.3). □

The result has the following, quite remarkable consequence:

Corollary 4.10. *Let X_n be a $\mathscr{L}^1$-bounded super-martingale, i.e.* $\sup_n \mathbb{E}|X_n|<\infty$. *Define $U_\infty(X,[a,b]) = \lim_{n\to\infty} U_n(X,[a,b])$ for any interval $[a,b]$. Then*

$$(b-a)\,\mathbb{E}(U_\infty(X,[a,b])) \leq a + \sup_n \mathbb{E}(|X_n|) < \infty. \tag{4.2.4}$$

In particular, $\mathbb{P}(U_\infty(X,[a,b])=\infty)=0$.

Proof. Exercise! □

Remark. We will say in general that a stochastic process X_n is bounded in $\mathscr{L}^p$, if $\sup_n \mathbb{E}|X_n|^p<\infty$. Note that this requirement is *strictly stronger* than just asking that for all n, $\mathbb{E}|X_n|^p<\infty$.

This is quite impressive: a (super) martingale that is $\mathscr{L}^1$-bounded cannot cross any interval infinitely often. The next result is even more striking, and in fact one of the most important results about martingales.

Theorem 4.11 (Doob's super-martingale convergence theorem). *Let X_n be a $\mathscr{L}^1$-bounded super-martingale. Then, almost surely, $X_\infty \equiv \lim_{n\to\infty} X_n$ exists and is a finite random variable.*

Proof. Define

$$\begin{aligned}
\Lambda &\equiv \{\omega : X_n(\omega)\,\text{does not converge to a limit in}\,[-\infty,+\infty]\} \qquad (4.2.5)\\
&= \{\omega : \limsup_{n\to\infty} X_n(\omega) > \liminf_{n\to\infty} X_n(\omega)\}\\
&= \bigcup_{a<b\in\mathbb{Q}} \{\omega : \limsup_{n\to\infty} X_n(\omega) > b > a > \liminf_{n\to\infty} X_n(\omega)\} \equiv \bigcup_{a<b\in\mathbb{Q}} \Lambda_{a,b}.
\end{aligned}$$

But

$$\Lambda_{a,b} \subset \{\omega : U_\infty(X,[a,b])(\omega) = \infty\}. \tag{4.2.6}$$

Therefore, by Corollary 4.10, $\mathbb{P}(\Lambda_{a,b}) = 0$, and thus also $\mathbb{P}(\bigcup_{a<b\in\mathbb{Q}} \Lambda_{a,b}) = 0$, since countable unions of null-sets are null-sets.

Thus the limit of X_n exists in $[-\infty,\infty]$ with probability one. It remains to show that it is finite. To do this, we use Fatou's lemma:

$$\mathbb{E}(|X_\infty|) = \mathbb{E}(\liminf_{n\to\infty} |X_n|) \leq \liminf_{n\to\infty} \mathbb{E}(|X_n|) \leq \sup_{n\in\mathbb{N}_0} \mathbb{E}(|X_n|) < \infty. \tag{4.2.7}$$

So X_∞ is almost surely finite. □

Doob's convergence theorem implies that positive super-martingale always converge a.s.. This is because the super-martingale property ensures in this case that $\mathbb{E}(|X_n|) = \mathbb{E}(X_n) \leq \mathbb{E}(X_0)$, so the uniform boundedness in $\mathscr{L}^1$ is always guaranteed.

Our next result gives a sharp criterion for convergence that brings to light the importance of the notion of the uniform integrability.

Theorem 4.12. *Let X be a $\mathscr{L}^1$-bounded super-martingale, so that, by Theorem 4.11 $X_\infty \equiv \lim_{n\to\infty} X_n$ exists a.s.. Then $X_n \to X_\infty$ in $\mathscr{L}^1$, if and only if the sequence $\{X_n, n \in \mathbb{N}_0\}$ is uniformly integrable. Then, for $n \in \mathbb{N}_0$,*

$$\mathbb{E}(X_\infty|\mathfrak{F}_n) \leq X_n, \text{ a.s.} \tag{4.2.8}$$

with equality holding if X is a martingale.

Proof. The first statement follows from Theorem 1.37. For $m \geq n$, $\mathbb{E}(X_m|\mathfrak{F}_n) \leq X_n$ a.s.. We let m tend to infinity and use $\mathscr{L}^1$-convergence to obtain $\lim_{m\to\infty} E(X_m|\mathfrak{F}_n) = \mathbb{E}(\lim_{m\to\infty} X_m|\mathfrak{F}_n) = \mathbb{E}(X_\infty|\mathfrak{F}_n)$, we obtain (4.2.8). □

A martingale with the property that there exists integrable X_∞ such that $X_n = \mathbb{E}(X_\infty|\mathfrak{F}_n)$ is called a closed martingale. The same applies to super (sub) martingales upon appropriate modification of the equality relation. The preceding theorem thus says in particular that (sup,super) martingales that are uniformly integrable and converge a.s. are closed.

The martingales of our first example are by definition closed. The next result implies that such martingales converge almost surely and in $\mathscr{L}^1$. To show this, we need yet another result of Doob that implies the uniform integrability of conditional expectations.

Theorem 4.13. *Let X be an absolutely integrable random variable on some probability space $(\Omega,\mathfrak{F},\mathbb{P})$. Then the family*

$$\{\mathbb{E}(X|\mathfrak{G}) : \mathfrak{G} \text{ is a sub-}\sigma\text{-algebra of } \mathfrak{F}\} \tag{4.2.9}$$

is uniformly integrable.

Proof. Since X is absolutely integrable, for any $\varepsilon > 0$, we can find $\delta > 0$ such that, if $F \in \mathfrak{F}$ with $\mathbb{P}(F) < \delta$, then $\mathbb{E}(|X|\mathbb{1}_F) < \varepsilon$. Let such ε and δ be given. Choose K such that $K^{-1}\mathbb{E}(|X|) < \delta$. Let now $\mathfrak{G} \subset \mathfrak{F}$ be a σ-algebra, and let Y be a version of $\mathbb{E}(X|\mathfrak{G})$. Then Jensen's inequality for conditional expectations implies that

$$|Y| \leq \mathbb{E}(|X|\,|\mathfrak{G}), \text{ a.s.}$$

By Chebychev inequality we have, $K\mathbb{P}(|Y| > K) \leq \mathbb{E}(|Y|) \leq \mathbb{E}(|X|)$. Thus $\mathbb{P}(|Y| > K) < \delta$. Moreover, since the event $\{|Y| > K\} \in \mathfrak{G}$, we can argue that

$$\begin{aligned}\mathbb{E}(|Y|\mathbb{1}_{|Y|>K}) &\leq \mathbb{E}[\mathbb{1}_{|Y|>K}\mathbb{E}(|X||\mathfrak{G})] = \mathbb{E}[\mathbb{E}(|X|\mathbb{1}_{|Y|>K}|\mathfrak{G})] \\ &= \mathbb{E}(|X|\mathbb{1}_{|Y|>K})) < \varepsilon,\end{aligned}$$

where in the last step we have set $F = \{|Y| > K\}$. This is the uniform integrability property we want to prove. $\square$

Theorem 4.14. *Let ξ be an absolutely integrable random variable on a filtered probability space $(\Omega, \mathfrak{F}, \mathbb{P}, \{\mathfrak{F}_n, n \in \mathbb{N}_0\})$. Define $X_n \equiv \mathbb{E}(\xi|\mathfrak{F}_n)$, a.s.. Then X_n is a uniformly integrable martingale and*

$$X_n \to X_\infty = \mathbb{E}(\xi|\mathfrak{F}_\infty), \tag{4.2.10}$$

almost surely and in $\mathscr{L}^1$.

Proof. X_n is a $\mathscr{L}^1$-bounded martingale by the properties of conditional expectations. The proceedings Theorem 4.13 implies that X_n is uniformly integrable. Thus X_n converges almost surely and in $\mathscr{L}^1$. We have to show the last equality in (4.2.10). For any n, and any $F \in \mathfrak{F}_n$,

$$\mathbb{E}[\mathbb{1}_F\mathbb{E}(\xi|\mathfrak{F}_\infty)] = \mathbb{E}[\mathbb{E}[\mathbb{E}(\mathbb{1}_F\xi|\mathfrak{F}_n)]|\mathfrak{F}_\infty]] = \mathbb{E}(\mathbb{1}_F X_n).$$

But for all $m > n$,

$$\mathbb{E}[\mathbb{1}_F X_n] = \mathbb{E}[\mathbb{1}_F X_m],$$

and so

$$\mathbb{E}[\mathbb{1}_F X_n] = \lim_{m\uparrow\infty}\mathbb{E}[\mathbb{1}_F X_m] = \mathbb{E}[\mathbb{1}_F X_\infty]$$

since X_m converges in $\mathscr{L}^1$. Thus $\mathbb{E}[\mathbb{1}_F\mathbb{E}(\xi|\mathfrak{F}_\infty)] = \mathbb{E}(\mathbb{1}_F X_\infty)$ for any F in the π-system $\bigcup_{n\in\mathbb{N}_0}\mathfrak{F}_n$ that generates the σ-algebra $\mathfrak{F}_\infty$. But this means that $\mathbb{E}(\xi|\mathfrak{F}_\infty) = X_\infty$ almost surely. $\square$

Note that, when $\mathfrak{F} = \mathfrak{F}_\infty$, the theorem says that $\mathbb{E}(\xi|\mathfrak{F}_n) \to \xi$.
An application of this result is Kolmogorov's $0-1$ law.

Theorem 4.15 (Kolmogorov's $0-1$ law). *Let $X_n, n \in \mathbb{N}$ be a sequence of independent random variables. Define $\mathscr{T}_n = \sigma(X_{n+1}, X_{n+2}, \dots)$ and $\mathscr{T} \equiv \bigcap_{n\in\mathbb{N}}\mathscr{T}_n$. Then, $\mathbb{P}(F) \in \{0,1\}$ if $F \in \mathscr{T}$.*

Proof. Let $\mathfrak{F}_n \equiv \sigma(X_1,\dots,X_n)$, $F \in \mathscr{T}$ and set $\eta = \mathbb{1}_F$. Since η is bounded and $\mathfrak{F}_\infty$-measurable, the preceding theorem tells us that

$$\eta = \mathbb{E}(\eta|\mathfrak{F}_\infty) = \lim_{n\to\infty} \mathbb{E}(\eta|\mathfrak{F}_n), \text{ a.s.}$$

Now η is $\mathscr{T}_n$-measurable for each n and hence independent of $\mathfrak{F}_n$. Thus, for any n

$$\mathbb{E}(\eta|\mathfrak{F}_n) = \mathbb{E}(\eta) = \mathbb{P}(F), \text{ a.s.}$$

and so $\eta = \mathbb{P}(F)$, a.s.. But η takes only the values 0 and 1, being an indicator function. Thus $\mathbb{P}(F) \in \{0,1\}$, proving the theorem. □

The next theorem relates to filtrations to the infinite past. It is called the Lévy-Doob downward theorem. It is somehow an inverted version of the upward theorem.

Theorem 4.16. *Let $(\Omega,\mathfrak{F},\mathbb{P})$ be a probability space, and let $\{\mathfrak{G}_{-n}, n \in \mathbb{N}\}$ be a collection of sub-σ-algebras of $\mathfrak{F}$ such that, for all $n \in \mathbb{N}$,*

$$\mathfrak{G}_{-\infty} \equiv \bigcap_{k\in\mathbb{N}} \mathfrak{G}_{-k} \subset \cdots \subset \mathfrak{G}_{-n-1} \subset \mathfrak{G}_{-n} \subset \cdots \subset \mathfrak{G}_{-1}. \tag{4.2.11}$$

Let $\{X_{-n}, n \in \mathbb{N}\}$ be a super-martingale relative to $\{\mathfrak{G}_{-n}, n \in \mathbb{N}\}$, i.e.

$$\mathbb{E}(X_{-n}|\mathfrak{G}_{-m}) \le X_{-m}, \text{ a.s.}$$

for $m \ge n$. Assume that $\sup_{n\ge1} \mathbb{E}(X_{-n}) < \infty$. Then the process X is uniformly integrable and the limit

$$X_{-\infty} = \lim_{n\to\infty} X_{-n}$$

exists a.s. and in $\mathscr{L}^1$. Moreover,

$$\mathbb{E}(X_{-n}|\mathfrak{G}_{-\infty}) \le X_{-\infty}, \text{ a.s.}$$

with equality in the martingale case.

Remark. Note that the limit we are considering here is really quite different form the one in the previous convergence theorems. We are really looking backward in time: as n tends to infinity, X_{-n} is measurable with respect to smaller and smaller σ-algebras, contrary to the usual X_n, that depend on more information. Therefore, while a convergent martingale X_n can converge to a constant only if the entire sequence is a constant, but usually is a random variable, a convergent X_{-n} has a much better chance to converge to a real constant. We will see shortly why this can be used to prove things like the strong law of large numbers.

Proof. The nice thing about the upcrossing theorem is that it also provides a proof of the convergence of X_{-n}. In fact, just as before, if $\mathbb{E}(|X_{-1}|)$ is bounded, it follows that the number of upcrossings of any $[a,b]$ by the process X_{-n} is a.s. finite.

Therefore the limit exists in $[-\infty,\infty]$. The finiteness then follows since the condition $\sup_{n\in\mathbb{N}}\mathbb{E}(X_{-n})<\infty$, and the super-martingale property imply that $\infty>\mathbb{E}(X_{-\infty})\geq\mathbb{E}(X_{-1})>-\infty$. This implies that $X_{-\infty}$ is finite almost surely.

Thus we just need to prove uniform integrability to obtain convergence in $\mathscr{L}^1$. Now we know that $\mathbb{E}(X_{-n})$ is monotone increasing, and $\lim_{n\to\infty}\mathbb{E}(X_{-n})<\infty$. Thus, for any $\varepsilon>0$, there is $k\in\mathbb{N}$, such that

$$0\leq\mathbb{E}(X_{-n})-\mathbb{E}(X_{-k})\leq\varepsilon/2, \tag{4.2.12}$$

for all $n\geq k$. Now, for such n,k, and $\lambda>0$,

$$\begin{aligned}\mathbb{E}(|X_{-n}|\mathbb{1}_{|X_{-n}|>\lambda}) &= -\mathbb{E}(X_{-n}\mathbb{1}_{X_{-n}<-\lambda})+\mathbb{E}(X_{-n})-\mathbb{E}(X_{-n}\mathbb{1}_{X_{-n}\leq\lambda})\\ &\leq -\mathbb{E}(X_{-k}\mathbb{1}_{X_{-n}<-\lambda})+\mathbb{E}(X_{-n})-\mathbb{E}(X_{-k}\mathbb{1}_{X_{-n}\leq\lambda})\end{aligned}$$

where we used the super-martingale property to replace n by k. Next we can replace $\mathbb{E}(X_{-n})$ by $\mathbb{E}(X_{-k})$ with an error of at most $\varepsilon/2$, after which the right-hand side reproduces the left hand one with n replaced by k in the first place, i.e.

$$\mathbb{E}(|X_{-n}|\mathbb{1}_{|X_{-n}|>\lambda})\leq\mathbb{E}(|X_{-k}|\mathbb{1}_{|X_{-n}|>\lambda})+\varepsilon/2. \tag{4.2.13}$$

Since X_{-k} is absolutely integrable, there exists $\delta>0$ such that for all F with

$$\mathbb{P}(F)<\delta\Rightarrow\mathbb{E}(|X_{-k}|\mathbb{1}_F)<\varepsilon/2. \tag{4.2.14}$$

But $\mathbb{P}(|X_{-n}|>\lambda)\leq\lambda^{-1}\mathbb{E}(|X_{-n}|)$. To control $\mathbb{E}(|X_{-n}|)$, let us set $X^-\equiv\max(-X,0)$, and write

$$\mathbb{E}(|X_{-n}|)=\mathbb{E}(X_{-n})+2\mathbb{E}(X_{-n}^-).$$

But X^- is a sub-martingale, and so

$$\mathbb{E}(|X_{-n}|)\leq\sup_{n\in\mathbb{N}}\mathbb{E}(X_{-n})+2\mathbb{E}(X_{-1}^-). \tag{4.2.15}$$

Thus we can choose $K<\infty$ such that

$$\begin{aligned}\mathbb{P}(|X_{-n}|>K)\leq\delta,&\ \text{if } n\geq k,\\ \mathbb{E}(|X_{-j}|\mathbb{1}_{|X_{-j}|>K})<\varepsilon,&\ \text{if } j<k,\end{aligned} \tag{4.2.16}$$

(for the second we just use the integrability for the finitely many values of i; for the first we use the uniform bound (4.2.15)). Then the first inequalities imply that $\mathbb{E}(|X_{-n}|\mathbb{1}_{|X_{-n}|>K})\leq\varepsilon$ for $n\geq k$ via (4.2.13) and the implication (4.2.14). This proves the uniform integrability. □

As an application we give a new proof of Kolmogorov's law of large numbers.

Theorem 4.17 (Kolmogorov's law of large numbers). *Let $X_n, n\in\mathbb{N}$ be iid random variables with $\mathbb{E}(|X_n|)<\infty$. Let $\mu=\mathbb{E}(X_n)$. Set $S_n\equiv\sum_{i=1}^n X_i$. Then*

$$n^{-1}S_n\to\mu, \tag{4.2.17}$$

a.s. *and in* $\mathscr{L}^1$.

Proof. Define $\mathfrak{G}_{-n} = \sigma(S_n, S_{n+1}, \dots,)$. Then, for $n \geq 1$,

$$\mathbb{E}(X_1|\mathfrak{G}_{-n}) = \mathbb{E}(X_2|\mathfrak{G}_{-n}) = \cdots = \mathbb{E}(X_n|\mathfrak{G}_{-n}). \tag{4.2.18}$$

The reason for these equalities is simply that knowing something about the sums S_n, S_{n+1}, etc. effects the expectation of the X_k, $k \leq n$ all in the same way: we could simply re-label the first indices without changing anything. Then, by linearity

$$\mathbb{E}(X_1|\mathfrak{G}_{-n}) = (n-1)^{-1}\mathbb{E}(S_{n-1}|\mathfrak{G}_{-n}) = n^{-1}\mathbb{E}(S_n|\mathfrak{G}_{-n}) = n^{-1}S_n, \text{ a.s.} \tag{4.2.19}$$

where we used the fact that S_n is $\mathfrak{G}_{-n}$ measurable. Thus, $L_{-n} \equiv n^{-1}S_n$ is a martingale with respect to the filtration $\{\mathfrak{G}_{-n}, n \in \mathbb{N}\}$. Thus, by the preceding theorem $L \equiv \lim_{n\to\infty} L_{-n}$ exists a.s. and in $\mathscr{L}^1$.

But clearly we also have, for any finite k, that $L = \lim_{n\to\infty} n^{-1}(X_{k+1} + \cdots + X_{n+k})$, which means that L is measurable with respect to $\mathscr{T}_k$, for any k. Now Kolmogorov's zero-one law implies that, for any c, $\mathbb{P}(L \leq c) \in \{0,1\}$. Since as a function of c this is monotone and right-continuous, there must be exactly one c_0, such that $\mathbb{P}(L = c) = 1$ for all $c \geq c_0$ and $\mathbb{P}(L = c) = 0$ for all $c < c_0$. Then $\mathbb{E}(L) = c_0$. But $\mathbb{E}(L_{-n}) = \mu$, for all n, so $c_0 = \mu$. $\square$

The proof above shows some of the power of martingales!

4.3 Inequalities

In this section we derive some fundamental inequalities for martingales. One of the most useful ones is the following *maximum inequality*.

Theorem 4.18 (Sub-martingale maximum inequality). *Let Z be a non-negative sub-martingale. Then, for $c > 0$, and $n \in \mathbb{N}$,*

$$c\mathbb{P}\left(\max_{k\leq n} Z_k \geq c\right) \leq \mathbb{E}\left(Z_n \mathbb{1}_{\{\max_{k\leq n} Z_k \geq c\}}\right) \leq \mathbb{E}(Z_n). \tag{4.3.1}$$

Remark. You may recall a similar result for sums of iid random variables as Kolmogorov's inequality. The estimate is extremely powerful, since it gives the same estimate for the probability of the maximum to exceed c as Chebychev's inequality would give for just the endpoint!

Proof. Define the sequence of disjoint events $F_0 \equiv \{Z_0 \geq c\}$,

$$F_k \equiv \bigcap_{\ell<k} \{Z_\ell < c\} \cap \{Z_k \geq c\} = \{\omega : \min(\ell \leq n : X_\ell \geq c) = k\}. \tag{4.3.2}$$

Then

$$F \equiv \left\{ \sup_{k \le n} Z_k \ge c \right\} = \bigcup_{k=0}^{n} F_k. \tag{4.3.3}$$

Clearly, the events $F_k \in \mathfrak{F}_k$. Moreover, on F_k we know that $Z_k \ge c$. Thus

$$\mathbb{E}(Z_n \mathbb{1}_{F_k}) \ge \mathbb{E}(Z_k \mathbb{1}_{F_k}) \ge c\mathbb{P}(F_k) \tag{4.3.4}$$

for all $k \le n$. Here the first inequality used of course the sub-martingale property of Z. Thus

$$\mathbb{E}(Z_n \mathbb{1}_F) = \sum_{k=0}^{n} \mathbb{E}(Z_n \mathbb{1}_{F_k}) \ge c \sum_{k=0}^{n} \mathbb{P}(F_k) = c\mathbb{P}(F). \tag{4.3.5}$$

This implies the assertion of the theorem. □

This implies the following corollary.

Corollary 4.19. *Let M be a martingale and $f : \mathbb{R} \to [0,\infty)$ a positive function that is convex and , increasing on $\mathbb{R}_+$. Then, for any $c > 0$,*

$$\mathbb{P}\left(\sup_{k \le n} M_k > c \right) \le \frac{\mathbb{E}(f(M_n))}{f(c)}. \tag{4.3.6}$$

Proof. Note that if M_n is a martingale and f a convex function such that $\mathbb{E} f(M_n) < \infty$, then $f(M_n)$ is a sub-martingale. Namely, convexity of f implies that there is a constant k such that $f(M_n) - f(M_{n-1}) \ge k(M_n - M_{n-1})$. Therefore

$$\begin{aligned} &\mathbb{E}\left(f(M_n)|\mathfrak{F}_{n-1}\right) \ge \mathbb{E}\left(f(M_{n-1})|\mathfrak{F}_{n-1}\right) + c\mathbb{E}\left(M_n - M_{n-1}|\mathfrak{F}_{n-1}\right) \\ &f(M_{n-1}), \text{ a.s..} \end{aligned} \tag{4.3.7}$$

Since f is increasing, $\mathbb{P}(\max_{k \le n} M_n > c) = \mathbb{P}\max_{k \le n} f(M_k) > f(c))$. Using Theorem 4.18 for the positive sub-martingale $f(M_n)$ yields the assertion of the corollary. □

This allows to obtain many useful inequalities from the one of Theorem 4.18! In particular, Kolmogorov's inequality follows by choosing $f(X) = X^2$. Other useful choices are the exponential function, $f(x) = \exp(\lambda x)$, for $\lambda > 0$.

Our next target is Doob's $\mathscr{L}^p$ inequality. The next lemma is a first step in this direction.

Lemma 4.20. *Let X and Y be non-negative random variables such that, for all $c > 0$,*

$$c\,\mathbb{P}(X \ge c) \le \mathbb{E}(Y \mathbb{1}_{X \ge c}). \tag{4.3.8}$$

Then, for $p > 1$ and $q^{-1} = 1 - p^{-1}$,

$$\|X\|_p \le q\|Y\|_p. \tag{4.3.9}$$

Proof. By our hypothesis, it holds that

$$L \equiv \int_0^\infty pc^{p-1}\mathbb{P}(X \geq c)dc \leq \int_0^\infty pc^{p-2}\mathbb{E}(Y\mathbb{1}_{X\geq c})dc \equiv R.$$

Using Fubini's theorem for non-negative integrands, we can write

$$\begin{aligned} L &= \int_0^\infty pc^{p-1}\left(\int_\Omega \mathbb{1}_{X(\omega)\geq c}\mathbb{P}(d\omega)\right)dc \\ &= \int_\Omega \left(\int_0^{X(\omega)} pc^{p-1}dc\right)\mathbb{P}(d\omega) = \int_\Omega X(\omega)^p\mathbb{P}(d\omega) = \mathbb{E}(X^p). \end{aligned}$$

Starting from the right-hand side, we can perform the same calculation, and derive that

$$R = q\mathbb{E}(X^{p-1}Y) \leq q\|Y\|_p\|X^{p-1}\|_q,$$

where the second inequality is just Hölder's inequality. Then

$$\mathbb{E}(X^p) \leq q\|Y\|_p\|X^{p-1}\|_q. \tag{4.3.10}$$

Assume that $\|X\|_q$ is finite. Clearly, $(p-1)q = p$, and so

$$\|X^{p-1}\|_q = \left(\mathbb{E}X^{q(p-1)}\right)^{1/q} = (\mathbb{E}X^p)^{1/q}.$$

Therefore (4.3.10) reads

$$\|X\|_p^p \leq q\|Y\|_p\|X\|_p^{p/q},$$

or $\|X\|_p \leq q\|Y\|_p$, as claimed. If $\|X\|_p = \infty$, one derives the inequality first for $X \wedge n$, and then uses monotone convergence. This proves the lemma. □

We can now formulate Doob's $\mathscr{L}^p$-inequality.

Theorem 4.21 (Doob's $\mathscr{L}^p$-inequality). *Let $p > 1$ and $q^{-1} = 1 - p^{-1}$. Let Z be a non-negative sub-martingale bounded in $\mathscr{L}^p$, and define*

$$Z^* \equiv \sup_{k\in\mathbb{N}_0} Z_k. \tag{4.3.11}$$

Then $Z^* \in \mathscr{L}^p$, and

$$\|Z^*\|_p \leq q \sup_{n\in\mathbb{N}_0} \|Z_n\|_p. \tag{4.3.12}$$

The limit, $Z_\infty \equiv \lim_{n\to\infty} Z_n$, exists a.s. and in $\mathscr{L}^p$, and

$$\|Z_\infty\|_p = \sup_{n\in\mathbb{N}_0}\|Z_n\|_p = \lim_{n\to\infty}\|Z_n\|_p. \tag{4.3.13}$$

If Z is of the form $Z = |M|$, where M is a martingale bounded in $\mathscr{L}^p$, then $M_\infty \equiv \lim_{n\to\infty} M_n$ exists a.s. and in $\mathscr{L}^p$, and $Z_\infty = |M_\infty|$, a.s..

Proof. Define $Z_n^* \equiv \sup_{k\leq n} Z_k$. Theorem 4.18 and Lemma 4.20 imply that

$$\|Z_n^*\|_p \le q\|Z_n\|_p \le q \sup_{k\le n} \|Z_k\|_p.$$

Using the monotone convergence theorem, we get (4.3.12). Now $-Z$ is a super-martingale bounded in $\mathscr{L}^p$, and hence in $\mathscr{L}^1$, it follows that Z_∞ exists a.s.. But

$$|Z_\infty - Z_n|^p \le (\max(Z_\infty, Z_n))^p \le (Z^*)^p \in \mathscr{L}^1,$$

so that, by Lebesgue's dominated convergence theorem, $\mathbb{E}(|Z_\infty - Z_n|^p) \to 0$, i.e. $X_n \to X_\infty$ in $\mathscr{L}^p$.

The last assertion in (4.3.13) follows since by Jensen's inequality and the sub-martingale property

$$\mathbb{E}(Z_n^p) = \mathbb{E}\left(\mathbb{E}(Z_n^p|\mathfrak{F}_{n-1})\right) \ge \mathbb{E}\left(\mathbb{E}(Z_n|\mathfrak{F}_{n-1})^p\right) \ge \mathbb{E}(Z_{n-1}^p),$$

and so $\|Z_n\|_p$ is a non-decreasing sequence. The remaining assertions are straight-forward. □

4.4 Doob decomposition

One of the games when dealing with stochastic processes is to "extract the martingale part". There are several such decompositions, but the following *Doob decomposition* is very important and its continuous time analogue will be fundamental for the theory of stochastic integration.

Theorem 4.22 (Doob decomposition).

(i) Let $\{X_n, n \in \mathbb{N}_0\}$ be an adapted process on a filtered space $(\Omega, \mathfrak{F}, \mathbb{P}, \{\mathfrak{F}_n, n \in \mathbb{N}_0\})$ with $X_n \in \mathscr{L}^1$ for all n. Then X can be written in the form[2]

$$X = X_0 + M + A, \tag{4.4.1}$$

where M is a martingale with $M_0 = 0$ and A is a previsible process with $A_0 = 0$. This decomposition is unique modulo indistinguishability, *i.e. if for some other M', A', $X = X_0 + M' + A'$, then*

$$\mathbb{P}(M_n = M_n', A_n = A_n', \forall n \in \mathbb{N}) = 1.$$

(ii) The process X is a sub-martingale, if and only if A is an increasing process in the sense that

$$\mathbb{P}(A_n \le A_{n+1}, \forall n \in \mathbb{N}) = 1.$$

Proof. The proof is unsurprisingly very easy. All we need to do is to derive explicit formulae for M and A. Now assume that a decomposition of the claimed form exists.

[2] To make sure that there is no confusion about notation: the following equation is to be understood in the sense that $X_0 = X_0$, and for $n \ge 1$, $X_n = X_0 + M_n + A_n$.

Then

$$\begin{aligned}\mathbb{E}((X_n - X_{n-1})|\mathfrak{F}_{n-1}) &= \mathbb{E}((M_n - M_{n-1})|\mathfrak{F}_{n-1}) + \mathbb{E}((A_n - A_{n-1})|\mathfrak{F}_{n-1}) \\ &= 0 + A_n - A_{n-1} \end{aligned} \tag{4.4.2}$$

by the martingale and predictability properties. Therefore

$$A_n = \sum_{k=1}^{n} \mathbb{E}((X_k - X_{k-1})|\mathfrak{F}_{k-1}), \text{ a.s.} \tag{4.4.3}$$

So now just *define* A_n by (4.4.3), and M_n by $M_n \equiv X_n - X_0 - A_n$. Clearly M is then a martingale, and A is by construction predictable. To see uniqueness, we use $M_n - M'_n = A'_n - A_n$ and applying the conditional expectation with respect to $\mathfrak{F}_{n-1}$ we have $M_{n-1} - M'_{n-1} = A'_n - A_n$ a.s.. Then, by $M_0 = M'_0 = 0$ follows $A'_1 = A_1$ a.s., from which $M_1 = M'_1$ a.s. and so on. This ends the proof of (i). The assertion of (ii) is obvious from (4.4.2). $\square$

An immediate application of the decomposition theorem is a maximum inequality without positivity assumption.

Lemma 4.23. *If X is either a sub-martingale or a super-martingale then, for $n \in \mathbb{N}$ and $c > 0$,*

$$c\mathbb{P}\left(\sup_{k\le n} |X_k| \ge 3c\right) \le 4\mathbb{E}(|X_0|) + 3\mathbb{E}(|X_n|). \tag{4.4.4}$$

Proof. We consider the case when X is a sub-martingale, the case of the super-martingale is identical by passing to $-X$. Then there is a Doob decomposition

$$X = X_0 + M + A$$

with A an increasing process. Thus

$$\sup_{k\le n} |X_k| \le |X_0| + \sup_{k\le n} |M_k| + \sup_{k\le n} |A_k| = |X_0| + \sup_{k\le n} |M_k| + A_n.$$

Note that $|M|$ is a non-negative sub-martingale, so for the supremum of $|M_k|$ we can use Theorem (4.18). We use the simple observation that, if $x + y + z > 3c$, then at least one of the x, y, z must exceed c. Thus,

$$\begin{aligned} c\mathbb{P}\left(\sup_{k\le n} |X_k| \ge 3c\right) &\le c\mathbb{P}(|X_0| \ge c) + c\mathbb{P}\left(\sup_{k\le n} |M_k| \ge c\right) + c\mathbb{P}(A_n \ge c) \\ &\le \mathbb{E}(|X_0|) + \mathbb{E}(|M_n|) + \mathbb{E}(A_n) \end{aligned} \tag{4.4.5}$$

Now

$$\mathbb{E}(|M_n|) = \mathbb{E}(|X_n - X_0 - A_n|) \le \mathbb{E}(|X_n|) + \mathbb{E}(|X_0|) + \mathbb{E}(A_n)$$

and

$$\mathbb{E}(A_n) = \mathbb{E}(X_n - X_0 - M_n) = \mathbb{E}(X_n - X_0) \leq \mathbb{E}(|X_n|) + \mathbb{E}(|X_0|).$$

Inserting these two bounds into (4.4.5) gives the claimed inequality. $\square$

The Doob decomposition gives rise to two important derived processes associated to a martingale, M, the *bracket*, $\langle M \rangle$, and $[M]$.

Definition 4.24. Let M be a martingale in $\mathscr{L}^2$ with $M_0 = 0$. Then M^2 is a submartingale with Doob decomposition

$$M^2 = N + \langle M \rangle,$$

where N is a martingale that vanishes at zero and $\langle M \rangle$ is a previsible process that vanishes at zero. The process $\langle M \rangle$ is called the *bracket* of M.

Note that boundedness in $\mathscr{L}^1$ of $\langle M \rangle$ is equivalent to boundedness in $\mathscr{L}^2$ of M. From the formulas associated with the Doob decomposition, we derive that

$$\langle M \rangle_n - \langle M \rangle_{n-1} = \mathbb{E}((M_n^2 - M_{n-1}^2)|\mathfrak{F}_{n-1}) = \mathbb{E}((M_n - M_{n-1})^2|\mathfrak{F}_{n-1}). \tag{4.4.6}$$

Definition 4.25. Let M be as before. We define

$$[M]_n \equiv \sum_{k=1}^{n} (M_k - M_{k-1})^2. \tag{4.4.7}$$

Lemma 4.26. *If M is as before,*

$$M^2 - [M] \equiv V = (C \bullet M), \tag{4.4.8}$$

where $C_n \equiv 2M_{n-1}$. V is a martingale. If M is bounded in $\mathscr{L}^2$, then V is bounded in $\mathscr{L}^1$.

Proof. Exercise! $\square$

4.5 A discrete time Itô formula.

We will now give in some way a justification of the name "discrete stochastic integral" for the martingale tranform. We consider a martingale M zero in zero and a function $F : \mathbb{R} \to \mathbb{R}$. We want to consider the process $F(M_T)$ and ask whether we can represent $F(M_T) - F(M_0)$ as a "stochastic integral. Since we have called $C \bullet M$ a stochastic integral, we might expect that this formula could simply read $F(M_T) = (F' \bullet X)_T + F(M_0)$, as in the usual fundamental theorem of calculus, but this will not turn out to be the case in general.

Let us consider the situation when the increments of M_t are getting very small; the idea here is that the spacings between consecutive times are really small. So we introduce parameter $\varepsilon > 0$ that will later tend to zero, while we think that $T =$

$\varepsilon^{-1}C$. We also assume that $\mathbb{E}(M_t - M_{t-1})^2 = O(\varepsilon)$. To see why this may reasonable think of $M_t \equiv B_{t/T}$ with B Brownian motion, where $\mathbb{E}(B_{t/T} - B_{(t-1)T})^2 = 1/T = \varepsilon$. Assuming that F is a smooth function, we can expand $F(M_t)$ in a Taylor series:

$$\begin{aligned} F(M_t) = {} & F(M_{t-1}) + (M_t - M_{t-1})F'(M_{t-1}) \\ & + \frac{1}{2}F''(M_{t-1})(M_t - M_{t-1})^2 + O\left((M_t - M_{t-1})^3\right) \end{aligned} \tag{4.5.1}$$

where we assume that

$$\mathbb{E}\left[O\left((M_t - M_{t-1})^3\right)\right] \leq K\varepsilon^{3/2},$$

and therefore $T\,\mathbb{E}\left[O\left((M_T - M_{T-1})^3\right)\right] \leq K\varepsilon^{1/2} \downarrow 0$, so that as $\varepsilon \downarrow 0$ these error terms will be negligible. Now we may iterate this procedure to obtain

$$\begin{aligned} F(M_T) = {} & F(M_0) + \sum_{t=1}^{T} F'(M_{t-1})(M_t - M_{t-1}) \\ & + \frac{1}{2}\sum_{t=1}^{T} F''(M_{t-1})(M_t - M_{t-1})^2 + O(\varepsilon^{1/2}). \end{aligned} \tag{4.5.2}$$

This expression looks almost like the Doob decomposition of the process $F(M_t)$, except that the last term is not exactly predictable. In fact, from the Doob decomposition, we would instead expect a predictable term of the form

$$\sum_{t=1}^{T} F''(M_{t-1})\mathbb{E}\left[(M_t - M_{t-1})^2|\mathfrak{F}_{t-1}\right]. \tag{4.5.3}$$

However, under reasonable assumptions (on F and on the behavior of the increments of the martingale M), the martingale

$$\Delta_T \equiv \sum_{t=1}^{T} F''(M_{t-1})\left((M_t - M_{t-1})^2 - \mathbb{E}\left[(M_t - M_{t-1})^2|\mathfrak{F}_{t-1}\right]\right)$$

satisfies $\mathbb{E}(\Delta_T^2) = O(\varepsilon)$, and is therefore negligible in our approximation. This implies the discrete version of *Itô's formula*:

$$\begin{aligned} F(M_T) = {} & F(M_0) + \sum_{t=1}^{T} F'(M_{t-1})(M_t - M_{t-1}) \\ & + \frac{1}{2}\sum_{t=1}^{T} F''(M_{t-1})\mathbb{E}\left[(M_t - M_{t-1})^2|\mathfrak{F}_{t-1}\right] + O(\varepsilon^{1/2}). \end{aligned} \tag{4.5.4}$$

4.6 Central limit theorem for martingales

One important further result for martingales concerns *central limit theorems*. There are various different formulations of such theorems. We will present one which emphasises the rôle of the bracket.

Theorem 4.27 (Central limit theorem). *Let M be a martingale with $M_0 = 0$. Set $s_n^2 \equiv \sum_{i=1}^n \mathbb{E}(M_i - M_{i-1})^2 = \mathbb{E}([M]_n)$. Assume that, as $n \to \infty$, $s_n^{-2} \max_{k\leq n} \mathbb{E}(M_k - M_{k-1})^2 \downarrow 0$, and, for all $\varepsilon > 0$,*

$$s_n^{-2} \sum_{k=1}^n \mathbb{E}\left[(M_k - M_{k-1})^2 \mathbb{1}_{|M_k - M_{k-1}|>\varepsilon s_n} \middle| \mathfrak{F}_{k-1}\right] \downarrow 0, \text{ a.s.} \tag{4.6.1}$$

If moreoveer $\langle M\rangle_n / s_n^2 \to 1$ in probability, then

$$s_n^{-1} M_n \to \mathscr{N}(0,1) \tag{4.6.2}$$

in distribution.

Remark. Condition (4.6.1) is called the *conditional Lindeberg condition*. In the case when $M_n = S_n = \sum_{i=1}^n X_i$ with independent centered random variables X_i, (4.6.1) reduces to the usual Lindeberg condition

$$s_n^{-2} \sum_{k=1}^n \mathbb{E}\left[X_k^2 \mathbb{1}_{|X_k|>\varepsilon s_n}\right] \downarrow 0. \tag{4.6.3}$$

Moreover, in this case $\mathbb{E}([M]_n) = \langle M\rangle_n$, and so condition $\langle M\rangle_n/s_n^2 \to 1$ is trivially verified (it is equal to one for all n). Thus the above theorem implies the usual CLT for sums of independent random variables under the weakest possible conditions.

Interestingly, the conditions for the CLT for the martingale include a law of large numbers for the bracket of the martingale. This is worth keeping in mind.

Proof. To simplify notation we set $\tilde{M}_k \equiv M_k/s_n$. Then the assumptions of the theorem read:

$$\begin{aligned} &\max_{k\leq n} \mathbb{E}((\tilde{M}_k - \tilde{M}_{k-1})^2) \to 0, \\ &\sum_{k=1}^n \mathbb{E}\left((\tilde{M}_k - \tilde{M}_{k-1})^2 \mathbb{1}_{|\tilde{M}_k - \tilde{M}_{k-1}|>\varepsilon} \middle| \mathfrak{F}_{k-1}\right) \to 0, \\ &\langle \tilde{M}\rangle_n \to 1, \text{ in probability,} \end{aligned} \tag{4.6.4}$$

as $n \to \infty$. We have to prove that $\tilde{M}_n \to \mathscr{N}(0,1)$. This holds if and only if, for all $u \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{E}(e^{iu\tilde{M}_n}) = e^{-u^2/2}. \tag{4.6.5}$$

Let use set $\tilde{X}_k \equiv \tilde{M}_k - \tilde{M}_{k-1}$. Then, it holds

$$
\begin{aligned}
\langle \tilde{M} \rangle_n &= \sum_{k=1}^{n} \mathbb{E}(\tilde{X}_k^2 | \mathfrak{F}_{k-1}) = \langle \tilde{M} \rangle_{n-1} + \mathbb{E}(\tilde{X}_n^2 | \mathfrak{F}_{n-1}), \\
\tilde{M}_n &= \sum_{k=1}^{n} \tilde{X}_k = \tilde{M}_{n-1} + \tilde{X}_n.
\end{aligned} \tag{4.6.6}
$$

Things are a little tricky, and the following decomposition is quite helpful:

$$
\begin{aligned}
&\left| \mathbb{E} \left[e^{iu\tilde{M}_n} - e^{-\frac{u^2}{2}} \right] \right| \\
=&\left| \mathbb{E} \left[e^{iu\tilde{M}_n} \left(1 - e^{\frac{u^2}{2} \langle \tilde{M} \rangle_n} e^{-\frac{u^2}{2}} \right) + e^{-\frac{u^2}{2}} \left(e^{iu\tilde{M}_n} e^{\frac{u^2}{2} \langle \tilde{M} \rangle_n} - 1 \right) \right] \right| \\
\leq&\mathbb{E} \left[\left| 1 - e^{\frac{u^2}{2} \langle \tilde{M} \rangle_n} e^{-\frac{u^2}{2}} \right| \right] + \left| \mathbb{E} \left[e^{iu\tilde{M}_n} e^{\frac{u^2}{2} \langle \tilde{M} \rangle_n} - 1 \right] \right| \\
\leq&\mathbb{E} \left[\left| 1 - e^{\frac{u^2}{2} \langle \tilde{M} \rangle_n} e^{-\frac{u^2}{2}} \right| \right] + \sum_{k=1}^{n} \left| \mathbb{E} \left[e^{iu\tilde{M}_k} e^{\frac{u^2}{2} \langle \tilde{M} \rangle_k} - e^{iu\tilde{M}_{k-1}} e^{\frac{u^2}{2} \langle \tilde{M} \rangle_{k-1}} \right] \right|
\end{aligned} \tag{4.6.7}
$$

Now we show that the result holds under the assumption

$$
\langle \tilde{M} \rangle_n \leq C \tag{4.6.8}
$$

for some finite constant C. In a second step we will show how to remove this assumption. First, notice that the assumption that $\langle \tilde{M} \rangle_n \to 1$ in probability implies that

$$
\mathbb{E} \left[\left| 1 - e^{\frac{u^2}{2} \langle \tilde{M} \rangle_n} e^{-\frac{u^2}{2}} \right| \right] \to 0, \text{ as } n \to \infty. \tag{4.6.9}
$$

Thus we need to deal with the second term in (4.6.7). Using (4.6.6), we get

$$
\begin{aligned}
&\mathbb{E} \left[e^{iu\tilde{M}_k} e^{\frac{u^2}{2} \langle \tilde{M} \rangle_k} - e^{iu\tilde{M}_{k-1}} e^{\frac{u^2}{2} \langle \tilde{M} \rangle_{k-1}} \right] \\
=&\mathbb{E} \left[e^{iu\tilde{M}_{k-1}} e^{\frac{u^2}{2} \langle \tilde{M} \rangle_{k-1}} \left(e^{iu\tilde{X}_k + \frac{u^2}{2} \mathbb{E}(\tilde{X}_k^2 | \mathfrak{F}_{k-1})} - 1 \right) \right] \\
=&\mathbb{E} \left[e^{iu\tilde{M}_{k-1}} e^{\frac{u^2}{2} \langle \tilde{M} \rangle_{k-1}} \mathbb{E} \left(e^{iu\tilde{X}_k + \frac{u^2}{2} \mathbb{E}(\tilde{X}_k^2 | \mathfrak{F}_{k-1})} - 1 \middle| \mathfrak{F}_{k-1} \right) \right].
\end{aligned} \tag{4.6.10}
$$

This implies that

$$
\begin{aligned}
&\left| \mathbb{E} \left[e^{iu\tilde{M}_k} e^{\frac{u^2}{2} \langle \tilde{M} \rangle_k} - e^{iu\tilde{M}_{k-1}} e^{\frac{u^2}{2} \langle \tilde{M} \rangle_{k-1}} \right] \right| \\
&\leq e^{C \frac{u^2}{2}} \mathbb{E} \left[\left| \mathbb{E} \left(e^{iu\tilde{X}_k + \frac{u^2}{2} \mathbb{E}(\tilde{X}_k^2 | \mathfrak{F}_{k-1})} - 1 \middle| \mathfrak{F}_{k-1} \right) \right| \right]
\end{aligned} \tag{4.6.11}
$$

To simplify the notation, set $\sigma_k^2 \equiv \mathbb{E}(\tilde{X}_k^2 | \mathfrak{F}_{k-1})$. To bound $\mathbb{E} \left(e^{iu\tilde{X}_k + \frac{u^2}{2} \sigma_k^2} - 1 \middle| \mathfrak{F}_{k-1} \right)$, we use the following elementary estimates:

$$e^{ix} = 1 + ix - x^2/2 + R_1(x), \text{ with } |R_1(x)| \le \min(x^2, |x|^3), \tag{4.6.12}$$
$$e^{x^2/2} = 1 + x^2/2 + R_2(x), \text{ with } |R_2(x)| \le x^4 e^{x^2/2}. \tag{4.6.13}$$

With this we get

$$\begin{aligned}
&\mathbb{E}\left(e^{iu\tilde{X}_k + \frac{u^2}{2}\sigma_k^2} - 1 \middle| \mathfrak{F}_{k-1}\right) \\
&= \mathbb{E}\left(\left[1 + iu\tilde{X}_k - \frac{u^2}{2}\tilde{X}_k^2 + R_1(u\tilde{X}_k)\right]\left[1 + \frac{u^2}{2}\sigma_k^2 + R_2(u\sigma_k)\right] - 1 \middle| \mathfrak{F}_{k-1}\right).
\end{aligned}$$

Since σ_k is $\mathfrak{F}_{k-1}$-mesurable, the second bracket can be taken out of the conditional expectation. Also, $\mathbb{E}(\tilde{X}_k|\mathfrak{F}_{k-1}) = 0$ since $\tilde{M}$ is a martingale. Since $\mathbb{E}(\tilde{X}_k^2|\mathfrak{F}_{k-1}) = \sigma_k^2$, so that

$$\begin{aligned}
\mathbb{E}\left(e^{iu\tilde{X}_k + \frac{u^2}{2}\sigma_k^2} - 1 \middle| \mathfrak{F}_{k-1}\right) &= \left(1 + \frac{u^2}{2}\sigma_k^2 + R_2(u\sigma_k)\right)\mathbb{E}\left(R_1(u\tilde{X}_k)\middle|\mathfrak{F}_{k-1}\right) \\
&\quad + \left(1 - \frac{u^2}{2}\sigma_k^2\right)R_2(u\sigma_k) - \frac{u^4}{4}\sigma_k^4.
\end{aligned} \tag{4.6.14}$$

We use the following bounds:

(i) $\langle\tilde{M}\rangle_n = \sum_{k=1}^n \sigma_k^2 \le C$. In particular, σ_k^2 is both bounded and summable.
(ii) $\sigma_k^2 = \mathbb{E}(\tilde{X}_k^2|\mathfrak{F}_{k-1}) \le \varepsilon^2 + \mathbb{E}(\tilde{X}_k^2 \mathbb{1}_{|\tilde{X}_k|>\varepsilon}|\mathfrak{F}_{k-1})$. This is nice, because the second term is controlled by the Lindeberg condition.
(iii) $|\mathbb{E}(R_1(u\tilde{X}_k)|\mathfrak{F}_{k-1})| \le \varepsilon|u|^3\sigma_k^2 + u^2\mathbb{E}(\tilde{X}_k^2 \mathbb{1}_{|\tilde{X}_k|>\varepsilon}|\mathfrak{F}_{k-1})$. This holds by computing the conditional expectation given $\mathfrak{F}_{k-1}$ of both sides of the inequality

$$\begin{aligned}
\min\{u^2\tilde{X}_k^2, |u|^3|\tilde{X}_k|^3\} &= \min\{u^2\tilde{X}_k^2, |u|^3|\tilde{X}_k|^3\}\left(\mathbb{1}_{|\tilde{X}_k|\le\varepsilon} + \mathbb{1}_{|\tilde{X}_k|>\varepsilon}\right) \\
&\le u^3|\tilde{X}_k|^3\mathbb{1}_{|\tilde{X}_k|\le\varepsilon} + u^2|\tilde{X}_k|^2\mathbb{1}_{|\tilde{X}_k|>\varepsilon} \\
&\le \varepsilon|u|^3\tilde{X}_k^2 + u^2\tilde{X}_k^2\mathbb{1}_{|\tilde{X}_k|>\varepsilon}.
\end{aligned} \tag{4.6.15}$$

(iv) $|R_2(u\sigma_k)| \le e^{\frac{u^2}{2}C}u^4\sigma_k^4 \le e^{\frac{u^2}{2}C}u^4C^2$.

Using these estimates, we get

$$\begin{aligned}
|(4.6.14)| &\le \left(1 + \frac{u^2}{2}C + e^{\frac{u^2}{2}C}u^4C^2\right)\left(\varepsilon|u|^3\sigma_k^2 + u^2\mathbb{E}(\tilde{X}_k^2\mathbb{1}_{|\tilde{X}_k|>\varepsilon}|\mathfrak{F}_{k-1})\right) \\
&\quad + \left(1 + \frac{u^2}{2}C\right)e^{\frac{u^2}{2}C}u^4\left(\sigma_k^2\varepsilon^2 + C\mathbb{E}(\tilde{X}_k^2\mathbb{1}_{|\tilde{X}_k|>\varepsilon}|\mathfrak{F}_{k-1})\right) \\
&\quad + \frac{u^4}{4}\left(\sigma_k^2\varepsilon^2 + C\mathbb{E}(\tilde{X}_k^2\mathbb{1}_{|\tilde{X}_k|>\varepsilon}|\mathfrak{F}_{k-1})\right) \\
&\le K(u)\left(\sigma_k^2\varepsilon^2 + \mathbb{E}(\tilde{X}_k^2\mathbb{1}_{|\tilde{X}_k|>\varepsilon}|\mathfrak{F}_{k-1})\right),
\end{aligned} \tag{4.6.16}$$

for some constant $K(u) < \infty$. But

$$\sum_{k=1}^{n}\left(\sigma_k^2\varepsilon^2 + \mathbb{E}(\tilde{X}_k^2 \mathbb{1}_{|\tilde{X}_k|>\varepsilon}|\mathfrak{F}_{k-1})\right) \leq C\varepsilon^2 + \sum_{k=1}^{n}\mathbb{E}(\tilde{X}_k^2 \mathbb{1}_{|\tilde{X}_k|>\varepsilon}|\mathfrak{F}_{k-1}), \tag{4.6.17}$$

there by the Lindeberg condition the second term tends to zero for any $\varepsilon > 0$. Thus the limit as $n \uparrow \infty$ of the second trem in Eq. (4.6.7) is bounded by a constant times ε^2, for any $\varepsilon > 0$, that is it is equal to zero, as desired. This proves the CLT under the assumption (4.6.8).

To conclude, let us show that we can remove Assumption (4.6.8). Define

$$A_m \equiv \left\{\omega \in \Omega : \langle \tilde{M} \rangle_m \equiv \sum_{k=1}^{m} \mathbb{E}\left(\tilde{X}_k^2|\mathfrak{F}_{k-1}\right) \leq C\right\}. \tag{4.6.18}$$

Of course, for $m \leq n$, $A_n \subset A_m$, and so $\mathbb{P}(A_n) \leq \mathbb{P}(A_m)$. Moreover, by assumption $\langle \tilde{m} \rangle_n \to 1$ and so $\lim_{n\to\infty}\mathbb{P}(A_n) = 1$. Notice that $\sum_{k=1}^{m}\mathbb{E}\left(\tilde{X}_k^2|\mathfrak{F}_{k-1}\right)$ is $\mathfrak{F}_{m-1}$ measurable, and hence so is $\mathbb{1}_{A_m}$. Thus, if we set $Z_m \equiv \tilde{X}_m\mathbb{1}_{A_m}$, it holds that $\mathbb{E}(Z_m|\mathfrak{F}_{m-1}) = 0$, for all $m \leq n$. Therefore the variables $\{Z_m, m \leq n\}$, for fixed n, form a martingale difference sequence. Since $|Z_m| \leq |\tilde{X}_m|$, all the properties used in the calculations above carry over to the Z_m. Therefore, repeating the calculations above with $\tilde{M}_n$ replaced by $\widehat{M}_n \equiv \sum_{m=1}^{n} Z_m$, we find that

$$\lim_{n\to\infty}\mathbb{E}\left(e^{iu\widehat{M}_n}\right) = e^{-u^2/2}. \tag{4.6.19}$$

SInce on A_m it holds that $\tilde{M}_m = Z_m$ and since $A_n \subset A_m$, it is true that on A_n, we have that $\widetilde{M}_n = \tilde{M}_n$. Therefore,

$$\begin{aligned}
\lim_{n\to\infty}\mathbb{E}\left(e^{iu\tilde{M}_n}\right) &= \lim_{n\to\infty}\mathbb{E}\left(e^{iu\tilde{M}_n}\mathbb{1}_{A_n}\right) + \lim_{n\to\infty}\mathbb{E}\left(e^{iu\tilde{M}_n}\mathbb{1}_{A_n^c}\right) \qquad (4.6.20)\\
&= \lim_{n\to\infty}\mathbb{E}\left(e^{iu\widehat{M}_n}\mathbb{1}_{A_n}\right) + 0\\
&= \lim_{n\to\infty}\mathbb{E}\left(e^{iu\widehat{M}_n}\right) - \lim_{n\to\infty}\mathbb{E}\left(e^{iu\widehat{M}_n}\mathbb{1}_{A_n^c}\right)\\
&= e^{-u^2/2}.
\end{aligned}$$

This concludes the proof of the theorem. □

Very similar computations like those presented above play an important rôle in what is called the *concentration of measure phenomenon*. Without going into too many details, let me briefly describe this. The setting one is considering is the following. We have n independent, identically distributed random variables, $X_1, \dots, X_n$, assumed to have mean zero, variance one, and to satisfy, e.g. $\mathbb{E}(e^{uX_i}) < \infty$, for all $u \in \mathbb{R}$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function that satisfies

$$\sup_{k=1}^{n}\sup_{x\in\mathbb{R}^n}\left|\frac{\partial f}{\partial x_k}(x_1, \dots, x_n)\right| \leq 1.$$

Set $F \equiv f(X_1,\dots,X_n)$. Then one can show that for some constant, $C>0$,

$$\mathbb{P}\left(|F-\mathbb{E}(F)|>\rho n\right) \leq 2e^{-\frac{n\rho^2}{2C}}. \tag{4.6.21}$$

The proof relies on the exponential Markov inequality, that states that

$$\mathbb{P}\left(F-\mathbb{E}(F)>n\rho\right) \leq \inf_{t\geq 0} e^{-tn\rho}\mathbb{E}(e^{t(F-\mathbb{E}(F))}).$$

The trick is to bound the Laplace transform by

$$\mathbb{E}(e^{t(F-\mathbb{E}(F))}) \leq e^{t^2 nC/2}.$$

(and not, as one might worry, of order $\exp(n^2)$!!).

To do this, one writes $F-\mathbb{E}(F)$ as a martingale difference sequence with respect to the filtration generated by the random variables X_i:

$$F-\mathbb{E}(F) = \sum_{k=1}^{n}\left(\mathbb{E}(F|\mathfrak{F}_k)-\mathbb{E}(F|\mathfrak{F}_{k-1})\right). \tag{4.6.22}$$

The computations one has to do are quite similar to those we have perfored in the proof of the central limit theorem. There is one small trick that is useful to use: Set $F^u \equiv f(X_1,\dots,uX_k,X_{k+1},\dots X_n)$. Then

$$F-F^0 = \int_0^1 du\frac{d}{du}F^u = \int_0^1 duX_k\frac{\partial}{\partial x_k}f(X_1,\dots,uX_k,X_{k+1},\dots X_n)$$

and

$$\begin{aligned}\mathbb{E}(F|\mathfrak{F}_k)-\mathbb{E}(F|\mathfrak{F}_{k-1}) &= \int_0^1 du\left(\mathbb{E}\left[\frac{d}{du}F^u\Big|\mathfrak{F}_k\right]-\mathbb{E}\left[\frac{d}{du}F^u\Big|\mathfrak{F}_{k-1}\right]\right)\\ &\equiv \mathbb{E}(Z_k|\mathfrak{F}_k)-\mathbb{E}(Z_k|\mathfrak{F}_{k-1}),\end{aligned} \tag{4.6.23}$$

where $|Z_k|\leq|X_k|$. Hence

$$\begin{aligned}&\mathbb{E}\left(e^{\lambda(\mathbb{E}(F|\mathfrak{F}_k)-\mathbb{E}(F|\mathfrak{F}_{k-1}))}-1-\lambda\left(\mathbb{E}(F|\mathfrak{F}_k)-\mathbb{E}(F|\mathfrak{F}_{k-1})\right)\Big|\mathfrak{F}_{k-1}\right)\\ &\leq \lambda^2\mathbb{E}\left(\left(\mathbb{E}(F|\mathfrak{F}_k)-\mathbb{E}(F|\mathfrak{F}_{k-1})\right)^2 e^{\lambda|\mathbb{E}(Z_k|\mathfrak{F}_k)-\mathbb{E}(Z_k|\mathfrak{F}_{k-1})|}\Big|\mathfrak{F}_{k-1}\right)\\ &\leq \lambda^2 C\end{aligned} \tag{4.6.24}$$

by the assumption on the law of X_k. We leave the remaining details of the calculation as an exercise. For more on concentration of measure, see e.g. [9, 10].

4.7 Stopping times, optional stopping

In a stochastic process we often want to consider random times that are determined by the occurrence of a particular event. If this event depends only on what happens "in the past", we call it a *stopping time*. Stopping times are nice, since we can determine there occurrence as we observe the process; hence, if we are only interested in them, we can stop the process at this moment, hence the name.

Definition 4.28. A map $T : \Omega \to \mathbb{N}_0 \cup \{+\infty\}$ is called a stopping time (with respect to a filtration $\{\mathfrak{F}_n, n \in \mathbb{N}_0\}$), if, for all $n \in \mathbb{N}_0 \cup \{+\infty\}$,

$$\{T = n\} \in \mathfrak{F}_n. \tag{4.7.1}$$

Example. The most important examples of stopping times are hitting time. Let X be an adapted process, and let $B \in \mathfrak{B}$. Define

$$\tau_B \equiv \inf\{t > 0 : X_t \in B\}.$$

Then τ_B is a stopping time. To see this, note that, if $n \in \mathbb{N}$.

$$\{\tau_B = n\} = \{\omega : X_n(\omega) \in B, X_k(\omega) \notin B, \forall 0 < k < n\}.$$

This event is manifestly in $\mathfrak{F}_n$. The event $\{\tau_B = \infty\}$ occurs if $\{X_n \notin B, \forall n \in \mathbb{N}\} \subset \mathfrak{F}_\infty$.

In principle all stopping times can be realised as first hitting times of some process. To do so, define

$$X_{[T,\infty)}(n,\omega) = \begin{cases} 1, & \text{if } n \geq T(\omega), \\ 0, & \text{otherwise.} \end{cases}$$

This process is adapted, and $T = \tau_1$.

It is sometimes very convenient to have the notion of a σ-algebra of events that take place before a stopping time.

Definition 4.29. The pre-T-σ-algebra, $\mathfrak{F}_T$, is the set of events $F \subset \mathfrak{F}$, such that, for all $n \in \mathbb{N}_0 \cup \{+\infty\}$,

$$F \cap \{T \leq n\} \in \mathfrak{F}_n. \tag{4.7.2}$$

Pre-T-σ-algebras will play an important rôle in formulation the strong Markov property.

There are some useful elementary facts associated with this concept.

Lemma 4.30. *Let S, T be stopping times. Then:*

(i) If X is an adapted process, then X_T is $\mathfrak{F}_T$-measurable.
(ii)If $S < T$, then $\mathfrak{F}_S \subset \mathfrak{F}_T$.
(iii)$\mathfrak{F}_{T\wedge S} = \mathfrak{F}_T \cap \mathfrak{F}_S$.
(iv)If $F \in \mathfrak{F}_{S\vee T}$, then $F \cap \{S \leq T\} \in \mathfrak{F}_T$.
(v)$\mathfrak{F}_{S\vee T} = \sigma(\mathfrak{F}_T, \mathfrak{F}_S)$.

Proof. Exercise. □

We now return to our gambling mode. We consider a super-martingale X and we want to play a strategy, C, that depends of a stopping time, T: say, we keep one unit of stock until the random time T. Then

$$C_n \equiv C_n^T \equiv \mathbb{1}_{n \leq T}.$$

Note that C^T is a previsible process. Namely,

$$\{C_n^T = 0\} = \{T \leq n-1\} \in \mathfrak{F}_{n-1},$$

and since C_n^T only takes the two values $0,1$, this suffices to show that $C_n^T \in \mathfrak{F}_{n-1}$. The wealth process associated to this strategy is then

$$(C^T \bullet X)_n = X_{T\wedge n} - X_0.$$

Definition 4.31. We define define the *stopped process* X^T, via

$$X_n^T(\omega) \equiv X_{T(\omega)\wedge n}(\omega).$$

With this definition we have (for our choice of C)

$$C^T \bullet X = X^T - X_0.$$

Theorem 4.32. *(i) If X is a super-martingale and T is a stopping time, then the stopped process, X^T, is a super-martingale. In particular, for all $n \in \mathbb{N}$,*

$$\mathbb{E}(X_{T\wedge n}) \leq \mathbb{E}(X_0). \tag{4.7.3}$$

(ii)If X is a martingale and T is a stopping time, then X^T is a martingale. In particular

$$\mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_0). \tag{4.7.4}$$

Proof. It follows directly from Theorem 4.7(i) because C^T is positive and bounded. □

This theorem is disappointing news for whose who might have hoped to reach a certain gain by playing until they have won a preset sum of money, and stopping then. In a martingale setting, the sure gain that will occur *if* this stopping time is reached before time n is offset by the expected loss, if the target has not yet been reached.

Note, however, that the theorem does not assert that $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$ (see example below). The following theorem, called Doob's Optional Stopping Theorem, gives conditions under which even that holds.

Theorem 4.33 (Doob's Optional Stopping Theorem).

(i) Let T be a stopping time, and let X be a super-martingale. Then, X_T is integrable and

$$\mathbb{E}(X_T) \leq \mathbb{E}(X_0), \tag{4.7.5}$$

if one of the following conditions holds:

(a)T is bounded (i.e. there exists $N \in \mathbb{N}$, s.t. $T(\omega) \leq N \, \forall \omega \in \Omega$);
(b)X is bounded, and T is a.s. finite;
(c)$\mathbb{E}(T) < \infty$, and, for some $K < \infty$,

$$|X_n(\omega) - X_{n-1}(\omega)| \leq K, \tag{4.7.6}$$

for all $n \in \mathbb{N}, \omega \in \Omega$.

(ii)If X is a martingale and one of the conditions (a)-(c) holds, then $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

Remark. This theorem may look strange, and contradict the "no strategy" idea: take a simple random walk, S_n, (i.e. a series of fair games, and define a stopping time $T = \inf\{n : S_n = 10\}$. Then clearly $\mathbb{E}(X_T) = X_T = 10 \neq \mathbb{E}(X_0) = 0$! So we conclude, using (c), that $\mathbb{E}(T) = +\infty$. In fact, the "sure" gain if we achieve our goal is offset by the fact that on average, it takes infinitely long to reach it (of course, most games will end quickly, but chances are that some may take very very long!).

Proof. We already know that $\mathbb{E}(X_{T \wedge n}) - \mathbb{E}(X_0) \leq 0$ for all $n \in \mathbb{N}$. Consider case (a). Then we know that $T \wedge N = T$, and so $\mathbb{E}(X_T) = \mathbb{E}(X_{T \wedge N}) \leq \mathbb{E}(X_0)$, as claimed.

In case (b), we start from $\mathbb{E}(X_{T \wedge n}) - \mathbb{E}(X_0) \leq 0$ and let $n \to \infty$. Since T is almost surely finite, $\lim_{n\to\infty} X_{T \wedge n} = X_T$, a.s., and since X_n is uniformly bounded,

$$\lim_{n\to\infty} \mathbb{E}(X_{T \wedge n}) = \mathbb{E}(\lim_{n\to\infty} X_{T \wedge n}) = \mathbb{E}(X_T),$$

which implies the result.

In the last case, (c), we observe that

$$|X_{T \wedge n} - X_0| = \left| \sum_{k=1}^{T \wedge n} (X_k - X_{k-1}) \right| \leq KT,$$

and by assumption $\mathbb{E}(KT) < \infty$. Thus, we can again take the limit $n \to \infty$ and use Lebesgue's dominated convergence theorem to justify that the inequality survives.

Finally, to justify (ii), use that if X is a martingale, then both X and $-X$ are super-martingales. The ensuing two inequalities imply the desired equality. □

Case (c) in the above theorem is certainly the most frequent situation one may hope to be in. For this it is good to know how to show that $\mathbb{E}(T) < \infty$, if that is the case. The following lemma states that this is always the case, whenever, eventually, the probability that the event leading to T is reasonably big.

Lemma 4.34. *Suppose that T is a stopping time and that there exists $N \in \mathbb{N}$ and $\varepsilon > 0$, such that, for all $n \in \mathbb{N}$,*

$$\mathbb{P}(T \leq n+N|\mathfrak{F}_n) > \varepsilon, \text{ a.s.} \tag{4.7.7}$$

Then $\mathbb{E}(T) < \infty$.

Proof. Consider $\mathbb{P}(T > kN)$. Clearly we can write

$$\begin{aligned}
\mathbb{P}(T > kN) &= \mathbb{E}\left(\mathbb{1}_{T>(k-1)N}\mathbb{1}_{T>kN}\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{T>(k-1)N}\mathbb{1}_{T>kN}|\mathfrak{F}_{(k-1)N}\right)\right) \\
&= \mathbb{E}\left(\mathbb{1}_{T>(k-1)N}\mathbb{E}\left(\mathbb{1}_{T>kN}|\mathfrak{F}_{(k-1)N}\right)\right) \\
&\leq (1-\varepsilon)\mathbb{E}\left(\mathbb{1}_{T>(k-1)N}\right) \\
&\leq (1-\varepsilon)^k,
\end{aligned} \tag{4.7.8}$$

by iteration. The exponential decay of the probability implies the finiteness of the expectation of T immediately. □

Finally we state Doob's super-martingale inequalities for non-negative super-martingales.

Theorem 4.35. *Let X be non-negative super-martingale and T a stopping time. Then*

$$\mathbb{E}(X_T) \leq \mathbb{E}(X_0). \tag{4.7.9}$$

Moreover, for any $c > 0$,

$$c\,\mathbb{P}\left(\sup_k X_k > c\right) \leq \mathbb{E}(X_0). \tag{4.7.10}$$

Proof. We know that $\mathbb{E}(X_{T\wedge n}) \leq \mathbb{E}(X_0)$. Using Fatou's lemma allows to pass to the limit $n \to \infty$. For (4.7.10), set $T = \inf\{n : X_n > c\}$. Then, $\mathbb{E}(X_0) \geq \mathbb{E}(X_T) \geq c\mathbb{P}(X_T \geq c) \geq c\mathbb{P}(\sup_k X_k > c)$ because $\sup_k X_k > c$ implies $X_T \geq c$. □

Chapter 5
Markov processes

We have seen the definition and construction of discrete time Markov chains already in Chapter 3. Markov chains are among the most important stochastic processes that are used to model real live phenomena that involve disorder. This is because the construction of these processes is very much adapted to our thinking about such processes. Moreover, Markov processes can be very easily implemented in numerical algorithms. This allows to numerically simulate even very complicated systems. We will always imagine a Markov process as a "particle" moving around in state space; mind, however, that these "particles" can represent all kinds of very complicated things, once we allow the state space to be sufficiently general. In this section, S will always be a complete separable metric space.

5.1 Markov processes with stationary transition probabilities

In general, we call a stochastic process whose index set supports the action of a group (or semi-group) *stationary* (with respect to the action of this (semi) group, if all finite dimensional distributions are invariant under the simultaneous shift of all time-indices. Specifically, if our index sets, I, are $\mathbb{R}_+$ or $\mathbb{Z}$, resp. $\mathbb{N}$, then a stochastic process is stationary if for all $\ell \in \mathbb{N}$, $s_1, \dots, s_\ell \in I$, all $A_1, \dots, A_\ell \in \mathfrak{B}$, and all $t \in I$,

$$\mathbb{P}\left[X_{s_1} \in A_1, \dots, X_{s_\ell} \in A_\ell\right] = \mathbb{P}\left[X_{s_1+t} \in A_1, \dots, X_{s_\ell+t} \in A_\ell\right]. \tag{5.1.1}$$

We can express this also as follows: Define the shift θ, for any $t \in I$, as $(X \circ \theta_t)_s \equiv X_{t+s}$. Then X is stationary, if and only if, for all $t \in I$, the processes X and $X \circ \theta_t$ have the same finite dimensional distributions.

In the case of Markov processes, a necessary (but not sufficient) condition for stationarity is the stationarity of the transitions kernels. Recall that we have defined the one-step transition kernel $\mathscr{P}_t$ of a Markov process in Section 3.3.

Definition 5.1. A Markov process with discrete time $\mathbb{N}_0$ and state space S is said to have *stationary transition probabilities (kernels)*, if it's one step transition kernel,

$\mathscr{P}_t$, is independent of t, i.e. there exists a probability kernel, $P(x,A)$, s.t.

$$\mathscr{P}_t(x,A) = P(x,A), \tag{5.1.2}$$

for all $t \in \mathbb{N}$, $x \in S$, and $A \in \mathfrak{B}$.

Remark. With the notation $P_{t,s}$ for the transitions kernel from time s to time t, we could alternatively state that a Markov process has *stationary transition probabilities (kernels)*, if there exists a family of transition kernels $P_t(x,A)$, s.t.

$$P_{s,t}(x,A) = P_{t-s}(x,A), \tag{5.1.3}$$

for all $s < t \in \mathbb{N}$, $x \in S$, and $A \in \mathfrak{B}$. Note that there is a potential conflict of notation between $\mathscr{P}_t$ and P_t which should not be confused.

A key concept for Markov chains with stationary transition kernels is the notion of an *invariant* distribution.

Definition 5.2. Let P be the transition kernel of a Markov chain with stationary transition kernels. Then a probability measure, π, on $(S,\mathfrak{B})$ is called an *invariant (probability) distribution*, if

$$\int \pi(dx)P(x,A) = \pi(A), \tag{5.1.4}$$

for all $A \in \mathfrak{B}$. More generally, a positive, σ-finite measure, π, satisfying (5.1.4), is called an *invariant measure*.

Lemma 5.3. *A Markov chain with stationary probability kernels and initial distribution $P_0 = \pi$ is a stationary stochastic process, if and only if π is an invariant probability distribution.*

Proof. Exercise. $\square$

In the case when the state space, S, is finite, we have seen that there is always at least one invariant measure, which then can be chosen to be a probability measure. In the case of general state spaces, while there still will always be an invariant measure (through a generalisation of the Perron-Frobenius theorem to the operator setting), there appears a new issue, namely whether there is an invariant measure that is finite, viz. whether there exists a invariant probability distribution.

5.2 The strong Markov property

The setting of Markov processes is very much suitable for the application of the notions of stopping times introduced in the last section. In fact, one of the very important properties of Markov processes is the fact that we can split expectations between past and future also at random times.

Theorem 5.4. *Let X be a Markov process with stationary transition kernels. Let $\mathfrak{F}_n = \sigma(X_0,\dots,X_n)$ be the natural filtration, and let T be a stopping time. Let F and G be $\mathfrak{F}$-measurable functions, and let F in addition be measurable with respect to the pre-T-σ-algebra $\mathfrak{F}_T$. Then*

$$\mathbb{E}\left[\mathbb{1}_{T<\infty} F\, G\circ\theta_T|\mathfrak{F}_0\right] = \mathbb{E}\left[\mathbb{1}_{T<\infty} F\, \mathbb{E}'\left[G|\mathfrak{F}'_0\right](X_T)\big|\mathfrak{F}_0\right] \tag{5.2.1}$$

where $\mathbb{E}'$ and $\mathfrak{F}'$ refers to an independent copy, X', of the Markov chain X.

Remark. If this looks fancy, just think of G as a function of the Markov process, i.e. $G = G(X_{i_1},\dots,X_{i_k})$, and $F = F(X_T, X_{T-1},\dots,X_0)$. Then the statement of the theorem says that

$$\begin{aligned}&\mathbb{E}\left[\mathbb{1}_{T<\infty} F(X_T, X_{T-1},\dots.X_0)\, G(X_{T+i_1},\dots,X_{T+i_k})|\mathfrak{F}_0\right]\\ &= \mathbb{E}\left[\mathbb{1}_{T<\infty} F(X_T, X_{T-1},\dots.X_0)\, \mathbb{E}'\left[G(X'_{i_1},\dots,X'_{i_k})|\mathfrak{F}'_0\right](X_T)\big|\mathfrak{F}_0\right]\end{aligned} \tag{5.2.2}$$

Proof. We have

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_{T<\infty} F\, G\circ\theta_T|\mathfrak{F}_0\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{T<\infty} F\, G\circ\theta_T|\mathfrak{F}_T\right]|\mathfrak{F}_0\right]\\ &= \mathbb{E}\left[\mathbb{1}_{T<\infty} F\, \mathbb{E}\left[G\circ\theta_T|\mathfrak{F}_T\right]|\mathfrak{F}_0\right].\end{aligned} \tag{5.2.3}$$

Now $\mathbb{E}\left[G\circ\theta_T|\mathfrak{F}_T\right]$ depends only on X_T and by stationarity is equal to $\mathbb{E}'\left[G|\mathfrak{F}'_0\right](X_T)$, which yields the claim of the theorem. □

5.3 Markov processes and martingales

We now want to develop some theory that will be more important and more difficult in the continuous time case. First we want to see how the transition kernels can be seen as operators acting on spaces of measures respectively spaces of function.

If μ is a σ-finite measure on S, and P is a Markov transition kernel, we define the measure μP as

$$\mu P(A) \equiv \int_S P(x,A)\, d\mu(x), \tag{5.3.1}$$

and similarly, for the t-step transition kernel, P_t,

$$\mu P_t(A) \equiv \int_S P_t(x,A)\, d\mu(x). \tag{5.3.2}$$

By the Markov property, we have of course the

$$\mu P_t(A) = \mu P^t(A). \tag{5.3.3}$$

The action on measures has of course the following natural interpretation in terms of the process: if $\mathbb{P}(X_0\in A) = \mu(A)$, then

$$\mathbb{P}(X_t \in A) = \mu P_t(A). \tag{5.3.4}$$

Alternatively, if f is a bounded, measurable function on S, we define

$$(Pf)(x) \equiv \int_S f(y)P(x,dy), \tag{5.3.5}$$

and

$$(P_t f)(x) \equiv \int_S f(y)P_t(x,dy), \tag{5.3.6}$$

where again

$$P_t f = P^t f. \tag{5.3.7}$$

We say that P_t is a semi-group acting on the space of measures, respectively on the space of bounded measurable functions. The interpretation of the action on functions is given as follows.

Lemma 5.5. *Let P_t be a Markov semi-group acting on bounded measurable functions f. Then*

$$(P_t f)(x) = \mathbb{E}\left(f(X_t)|\mathfrak{F}_0\right)(x) \equiv \mathbb{E}_x f(X_t). \tag{5.3.8}$$

Proof. We only need to show this for $t = 1$. Then, by definition,

$$\mathbb{E}_x(f(X_1)) = \int_S f(y)\mathbb{P}(X_1 \in dy|\mathfrak{F}_0)(x) = \int_S f(y)P(x,dy) = (Pf)(x).$$

□

Notice that, by telescopic expansion, we have the elementary formula

$$P_t f - f = \sum_{s=0}^{t-1} P_s(P - \mathbb{1})f = \sum_{s=0}^{t-1} P_s L f, \tag{5.3.9}$$

where we call $L \equiv P - \mathbb{1}$ the (discrete) generator of our Markov process (this formula will have a complete analogon in the continuous-time case).

An interesting consequence is the following observation:

Lemma 5.6 (Discrete time martingale problem). *Let L be the generator of a Markov process, X_t, and let f be a bounded measurable function. Then*

$$M_t \equiv f(X_t) - f(X_0) - \sum_{s=0}^{t-1} Lf(X_s) \tag{5.3.10}$$

is a martingale.

Proof. Let $t, r \geq 0$. Then

$$\begin{aligned}
\mathbb{E}(M_{t+r}|\mathfrak{F}_t) &= \mathbb{E}(f(X_{t+r})|\mathfrak{F}_t) - \mathbb{E}(f(X_0)|\mathfrak{F}_t) - \sum_{s=0}^{t+r-1} \mathbb{E}(Lf(X_s)|\mathfrak{F}_t) \\
&= P^r f(X_t) - f(X_t) + f(X_t) - f(X_0) \\
&\quad - \sum_{s=t}^{t+r-1} \mathbb{E}(Lf(X_s)|\mathfrak{F}_t) - \sum_{s=0}^{t-1} \mathbb{E}(Lf(X_s)|\mathfrak{F}_t) \\
&= f(X_t) - f(X_0) - \sum_{s=0}^{t-1} Lf(X_s) \\
&\quad + P^r f(X_t) - f(X_t) - \sum_{s=0}^{r-1} P^r Lf(X_t) \\
&= M_t + 0. \qquad (5.3.11)
\end{aligned}$$

This proves the lemma. □

Remark. (5.3.10) is of course the Doob decomposition of the process $f(X_t)$, since $\sum_{s=0}^{t-1} Lf(X_s)$ is a previsible process. One may check that this can be obtained directly using the formula (4.4.3) [Exercise!].

What is important about this observation is that it gives rise to a characterisation of Markov processes that will be extremely useful in the continuous time setting.

Namely, one can ask whether the requirement that M_t be a martingale given a family of pairs (f, Lf) characterises fully a Markov process.

Theorem 5.7. *Let X be a discrete time stochastic process on a filtered space such that X is adapted. Then X is a Markov process with transition kernel $P = \mathbb{1} + L$, if and only if, for all bounded measurable functions, f, the expression on the right-hand side of (5.3.10) is a martingale.*

Proof. Lemma 5.6 already provides the "only if" part, so it remains to show the "if" part.

First, if we assume that X is a Markov process, setting $r = 1$ and $t = 0$ above and taking conditional expectations given $\mathfrak{F}_0$, we see from Lemma 5.5 that $\mathbb{E}(f(X_1)) = f(X_0) + (Lf)(X_0)$, implying that the transition kernel must be $\mathbb{1} + L$.

It remains to show that X is indeed a Markov process. For this we want to show that We want to show that

$$\mathbb{E}(f(X_{t+s})|\mathfrak{F}_t) = (\mathbb{1} + L)^s f(X_t) \equiv P^s f(X_t), \qquad (5.3.12)$$

from the martingale problem formulation. To see this, we just use the above calculation to see that

$$\begin{aligned}
\mathbb{E}(f(X_{t+r})|\mathfrak{F}_t) &= \mathbb{E}(M_{t+r}|\mathfrak{F}_t) + f(X_0) \\
&\quad + \sum_{s=0}^{t-1}(Lf)(X_s) + \sum_{s=t}^{t+r-1}\mathbb{E}((Lf)(X_s)|\mathfrak{F}_t) \\
&= M_t + f(X_0) + \sum_{s=0}^{t-1}(Lf)(X_s) + \sum_{s=t}^{t+r-1}\mathbb{E}((Lf)(X_s)|\mathfrak{F}_t) \\
&= f(X_t) + \sum_{s=0}^{r-1}\mathbb{E}((Lf)(X_{t+s})|\mathfrak{F}_t)
\end{aligned} \tag{5.3.13}$$

Now let again $r = 1$. Then

$$\mathbb{E}(f(X_{t+1})|\mathfrak{F}_t) = f(X_t) + (Lf)(X_t) = ((\mathbb{1} + L)f)(X_t) \equiv Pf(X_t), \tag{5.3.14}$$

which is (5.3.12) for $r = 1$. Now proceed by induction: assume that (5.3.12) holds for it holds for all bounded measurable functions for $s \leq r-1$. We must show that it then also holds for $s = r$. To do this, we use (5.3.13) for the last sum in (5.3.13),

$$\sum_{s=0}^{r-1}\mathbb{E}((Lf)(X_{t+s})|\mathfrak{F}_t) = \sum_{s=0}^{r-1}(P_s(Lf))(X_t) = (P^r f)(X_t) - f(X_t), \tag{5.3.15}$$

where we undid the telescopic sum. Inserting this into (5.3.13) yields (5.3.12) for $s = r$. Hence (5.3.12) holds for all r, by induction. □

Remark. The full strength of this theorem will come out in the continuous time case, where it remains valid. A crucial point is that it will not be necessary to even consider all bounded functions, but just sufficiently rich classes. This allows to formulate martingale problems even then one cannot write down the generator in a explicit form. The idea of characterising Markov processes by the associated martingale problem goes back to Stroock and Varadhan, see [14].

5.4 Harmonic functions and martingales

We have seen that measures that satisfy $\mu L = 0$ are of special importance in the theory of Markov processes (they are the invariant measures). Also of central importance are functions that satisfy $Lf = 0$. In this section we will assume that the transition kernels of our Markov chains have bounded support, so that for some $K < \infty$, $|X_{t+1} - X_t| \leq K < \infty$ for all t.

Definition 5.8. Let L be the generator of a Markov process. A measurable function that satisfies

$$Lf(x) = 0, \forall x \in S, \tag{5.4.1}$$

is called a *harmonic function*. A function is called *subharmonic* (resp. *superharmonic*, if $Lf \geq 0$, resp. $Lf \leq 0$.

Theorem 5.9. *Let X_t be a Markov process with generator L. Then, a non-negative function f is*

(i) harmonic, if and only if $f(X_t)$ is a martingale;
(ii) subharmonic, if and only if $f(X_t)$ is a submartingale;
(iii) superharmonic, if and only if $f(X_t)$ is a supermartingale;

Proof. Simply use Lemma 5.6. □

Remark. Theorem 5.9 establishes a profound relationship between potential theory and martingales. It also explains, the strange choice of super and sub in martingale theory.

A nice application of the preceding result is the maximum principle.

Theorem 5.10. *Let X be a Markov process and let D be a bounded open domain such that $\mathbb{E}(\tau_{D^c}) < \infty$. Assume that f is a non-negative subharmonic function on D. Then*

$$\sup_{x\in D} f(x) \leq \sup_{x\in D^c} f(x). \tag{5.4.2}$$

Proof. Let us define $T \equiv \tau_{D^c}$. Then $f(X^T)$ is a submartingale, and thus

$$\mathbb{E}\left(f(X_T)|\mathfrak{F}_0\right)(x) \geq f(x). \tag{5.4.3}$$

Since $X_T \in D^c$, it must be true that

$$\sup_{y\in D^c} f(y) \geq \mathbb{E}\left(f(X_T)|\mathfrak{F}_0\right)(x) \geq f(x), \tag{5.4.4}$$

for all $x \in D$, hence the claim of the theorem. Of course we used again the Doob's optional stopping theorem in case (i,c). □

The theorem can be phrased as saying that (sub) harmonic functions take on their maximum on the boundary, since of course the set D^c in (5.4.2) can be replaced by a subset, $\partial D \subset D^c$ such that $\mathbb{P}_x(X_T \in \partial D) = 1$.

The above proof is an example of how intrinsically analytic results can be proven with probabilistic means. The next section will further develop this theme.

5.5 Dirichlet problems

Let us now consider a connected bounded open subset D of S. We define the stopping time $T = \tau_{D^c} \equiv \inf\{t > 0 : X_t \in D^c\}$.

If g is a measurable function on D, we consider the Dirichlet problem associated to a generator, L, of a Markov process, X:

$$\begin{aligned} -(Lf)(x) &= g(x), \quad x \in D, \\ f(x) &= 0, \quad x \in D^c. \end{aligned} \tag{5.5.1}$$

Theorem 5.11. *Assume the $\mathbb{E}(T) < \infty$. Then (5.5.1) has a unique solution given by*

$$f(x) = \mathbb{E}\left(\sum_{t=0}^{T-1} g(X_t)\Big|\mathfrak{F}_0\right)(x) \equiv \mathbb{E}_x\left(\sum_{t=0}^{T-1} g(X_t)\right) \tag{5.5.2}$$

for $x \in D$, and $f(x) = 0$, for $x \in D^c$.

Proof. Consider the martingale M_t from Lemma 5.6. We know from Theorem 4.32 that M^T is also a martingale. Moreover,

$$M_T = f(X_T) - f(X_0) - \sum_{t=0}^{T-1}(Lf)(X_t) = 0 - f(X_0) - \sum_{t=0}^{T-1}(Lf)(X_t). \tag{5.5.3}$$

But we want f such that $-Lf = g$ on D. Thus, (5.5.3) seen as a problem for f, reads

$$M_T = -f(X_0) + \sum_{t=0}^{T-1} g(X_t). \tag{5.5.4}$$

Taking expectations conditioned on $\mathfrak{F}_0$, yields

$$0 = -f(X_0) + \mathbb{E}\left(\sum_{t=0}^{T-1} g(X_t)\Big|\mathfrak{F}_0\right), \tag{5.5.5}$$

or

$$f(x) = \mathbb{E}_x\left(\sum_{t=0}^{T-1} g(X_t)\right) \tag{5.5.6}$$

Here we relied of course on Doob's optimal stopping theorem for $\mathbb{E}(M_T) = 0$.

Thus, any solution of the Dirichlet problem is given by (5.5.6). To verify existence, we just need to check that (5.5.6) solves $-Lf = g$ on D. To do this we use the Markov property "backwards", to see that

$$\begin{aligned} Pf(x) &= P\mathbb{E}_x\left(\sum_{t=0}^{T-1} g(X_t)\right) = \int_D P(x,dy)\mathbb{E}_y\left(\sum_{t=0}^{T-1} g(X_t)\right) + \int_{D^c} P(x,dy0 \\ &= \mathbb{E}_x\left[\sum_{t=1}^{T-1} g(X_t)\right] = \mathbb{E}_x\left[\sum_{t=0}^{T-1} g(X_t)\right] - g(x) = f(x) - g(x). \end{aligned} \tag{5.5.7}$$

This concludes the proof. $\square$

We see that the Markov process produces a solution of the Dirichlet problem. We can express the solution in terms of an integral kernel, called the Green's kernel, $G_D(x,dy)$, as

$$f(x) = \int G_D(x,dy)g(y) \equiv \mathbb{E}_x\left(\sum_{t=0}^{T-1} g(X_t)\right), \tag{5.5.8}$$

or, in more explicit terms,

$$G_D(x,dy) = \sum_{t=0}^{\infty} P_D^t(x,dy), \tag{5.5.9}$$

where

$$P_D^t(x,dy) = \int_D P(x,dz_1) \int_D P(z_1,dz_2) \dots \int_D P(z_{t-1},dy). \tag{5.5.10}$$

The preceding theorem has an obvious extension to more complicated boundary value problems:

Theorem 5.12. *Let D be as above, and let h be a bounded function on D^c. Assume the $\mathbb{E}(T) < \infty$. Then*

$$f(x) \equiv \begin{cases} \mathbb{E}_x \left(\sum_{t=0}^{T-1} g(X_t) \right) + \mathbb{E}_x \left(h(X_T) \right), & x \in D, \\ h(x), & x \in D^c, \end{cases} \tag{5.5.11}$$

is the unique solution of the Dirichlet problem

$$\begin{aligned} -(Lf)(x) &= g(x), \quad x \in D, \\ f(x) &= h(x), \quad x \in D^c. \end{aligned} \tag{5.5.12}$$

Proof. Identical to the previous one. □

Theorem 5.12 is a two way game: it allows to produce solutions of analytic problems in terms of stochastic processes, and it allows to compute interesting probabilistic problems analytically. As an example, assume that $D^c = A \cup B$ with $A \cap B = \emptyset$. Set $h = \mathbb{1}_A$. Then, clearly, for $x \in D$,

$$\mathbb{E}_x(h(X_T)) = \mathbb{P}_x(X_T \in A) \equiv \mathbb{P}_x(\tau_A < \tau_B), \tag{5.5.13}$$

and so $\mathbb{P}_x(X_T \in A)$ can be represented as the solution of the boundary value problem

$$\begin{aligned} (Lf)(x) &= 0, \quad x \in D, \\ f(x) &= 1, \quad x \in A, \\ f(x) &= 0, \quad x \in B. \end{aligned} \tag{5.5.14}$$

The is a generalisation of the *ruin* problem for the random walk that we discussed in Probability 1.

Exercise. Derive the formula for $\mathbb{P}_x(\tau_A < \tau_B)$ directly from the Markov property without using Lemma 5.6.

5.5.1 *Green function, equilibrium potential, and equilibrium measure*

Let us consider the case where the solution of the Dirichlet problem is unique. Then the solution can be written in the form

$$f(x) = \int_D G_{D^c}(x,dz)g(z) + \int_{D^c} H_{D^c}(x,dz)\bar{g}(z), \tag{5.5.15}$$

where

$$G^\lambda_{D^c}(x,A) = \mathbb{E}_x\left[\sum_{t=0}^{\tau_{D^c}-1} \mathbb{1}_{X(t)\in A}\right]$$

is called the *Green kernel*, and

$$\begin{aligned} H^\lambda_{D^c}(x,A) &= \mathbb{E}_x\left[\mathbb{1}_{X(\tau_{D^c})\in A}\right] \\ &= \sum_{t=0}^{\infty} \mathbb{P}_x\left(\tau_{D^c} = t \wedge X(t) \in A\right) \end{aligned} \tag{5.5.16}$$

is called the *Poisson kernel*. The Green kernel can also be characterised as the weak solution of the problem

$$\begin{aligned} -(LG_{D^c}(x,dz) &= \delta_z(dx), \quad \forall x \in D, \\ G_{D^c}(x,dz) &= 0, \qquad\quad \forall x \in D^c. \end{aligned} \tag{5.5.17}$$

Let $A, B \subset S$ be two disjoint subsets. Consider the Dirichlet problem

$$\begin{aligned} (-Lh)(x) &= 0, \quad \forall x \in S \setminus (A \cup B), \\ h(x) &= 1, \quad \forall x \in A, \\ h(x) &= 0, \quad \forall x \in B. \end{aligned} \tag{5.5.18}$$

Suppose that (5.5.18) has a unique solution, e.g. because $\mathbb{E}_x[\tau_{A\cup B}] < \infty$ for all $x \in S$. The harmonic function that solves (5.5.18) will be denoted by $h_{A,B}(x)$ and is called the *equilibrium potential*. We have already seen that

$$h_{A,B}(x) = \mathbb{E}_x\left[\mathbb{1}_A(X(\tau_{A\cup B}))\right] = \mathbb{P}_x\left(\tau_A < \tau_B\right), \qquad x \in S \setminus (A \cup B). \tag{5.5.19}$$

We would like to view this equation as an analytic expression for the probability in the right-hand side. Naturally we would like to obtain such a formula also when $x \in A$ or $x \in B$. However, using the Markov property, we see that

$$\begin{aligned} \mathbb{P}_x\left(\tau_A < \tau_B\right) &= \int_{(A\cup B)^c} P(x,dy)\mathbb{P}_y\left(\tau_A < \tau_B\right) + \int_A P(x,dy) \\ &= \int_S P(x,dy)h_{A,B}(y) = Ph_{A,B}(x) \\ &= (Lh_{A,B})(x) + h_{A,B}(x). \end{aligned} \tag{5.5.20}$$

For $x \in B$, the latter can be also written as (since $h_{A,B}(x) = 0$)

$$\mathbb{P}_x(\tau_A < \tau_B) = (Lh_{A,B})(x), \tag{5.5.21}$$

and for $x \in A$ as

$$-(Lh_{A,B})(x) = 1 - \mathbb{P}_x(\tau_A < \tau_B) = \mathbb{P}_x(\tau_B < \tau_A). \tag{5.5.22}$$

The quantity $e_{A,B}(x) \equiv -Lh_{A,B}(x)$, $x \in A$, is called the *equilibrium measure* on A, and will be the second fundamental object in our study of metastability.

The following simple observation provides the fundamental connection between the objects we have introduced so far, and leads to a different representation of the equilibrium potential. Pretend that the equilibrium measure $e_{A,B}$ is already known. Then the equilibrium potential satisfies the *inhomogeneous* Dirichlet problem

$$\begin{aligned} -(Lh)(x) &= e_{A,B}(x), \quad && \forall x \in S \setminus B, \\ h(x) &= 0, && \forall x \in B. \end{aligned} \tag{5.5.23}$$

The solution of (5.5.23) can be written in terms of the Green function.

Lemma 5.13. *With the notation introduced above,*

$$h_{A,B}(x) = \int_A G_B(x, dy) e_{A,B}(y). \tag{5.5.24}$$

Note that $e_{a,B}(a) = \mathbb{P}_a(\tau_B < \tau_a)$ has the meaning of an *escape probability*.

5.5.2 Reversibility

Considerable simplifications occur when we assume a certain symmetry property of the transition kernels known as *reversibility* or, in physics terminology, *detailed balance*.

Definition 5.14. A Markov chain with state space S and one-step transition kernel P is called *reversible* if there exists a measure μ on S, such that

$$\int \mu(dx) f(x)(Pg)(x) = \int \mu(dx)(Pf)(x)g(x) \qquad \forall f, g \in L^2(S, \mu). \tag{5.5.25}$$

The measure μ is called the reversible measure of the Markov chain.

The function space $L^2(S, \mu)$ is a natural space to work on when the Markov chain is reversible with respect to μ.

Lemma 5.15. *Let $f \in L^2(S, \mu)$, where μ is invariant with respect to P. Then $Pf \in L^2(S, \mu)$.*

Proof. The claim follows from the fact that P is a contraction in the L^2-norm:

$$\begin{aligned}\int_S \mu(dx)\,[(Pf)(x)]^2 &= \int_S \mu(dx)\left[\int_S P(x,dy)f(y)\right]^2 \\ &\le \int_S \mu(dx)\int_S P(x,dy)f(y)^2 \int_S P(x,dy) \\ &\le \int_S \mu(dx)\int_S P(x,dy)f(y)^2 = \int_S \mu(dx)f(x)^2,\end{aligned} \tag{5.5.26}$$

where we use the Cauchy-Schwartz inequality and the invariance of μ, i.e., $\mu P = \mu$. □

Reversibility can be expressed by saying that the transition kernel P acts as a self-adjoint operator on the Hilbert space $L^2(S,\mu)$.

Lemma 5.16. *If μ is a reversible probability measure for P, then μ is an invariant probability measure for P.*

Proof. Clearly, $f \equiv 1$ is in $L^2(S,\mu)$. Hence, for all bounded measurable functions g,

$$\int_S\int_S \mu(dx)P(x,dy)g(y) = \int_S\int_S P(y,dx)g(y)\mu(dy) = \int_S g(y)\mu(dy). \tag{5.5.27}$$

Hence μ is invariant. □

Note that the converse is not true in general, i.e., an invariant measure is not necessarily reversible.

We next come to the definition of the Dirichlet form.

Lemma 5.17. *6 Let μ be a reversible measure for a Markov process with generator L. Then L defines a non-negative definite quadratic form*

$$\mathscr{E}(f,g) \equiv -\int_S \mu(dx)g(x)(Lf)(x), \tag{5.5.28}$$

called the Dirichlet form.

Proof. It suffices to write out $\mathscr{E}(f,g)$ explicitly. Namely, by reversibility,

$$\begin{aligned}\mathscr{E}(f,g) &= \int\int \mu(dx)g(x)P(x,dy)[f(x)-f(y)] \\ &= \int\int \mu(dx)f(x)P(x,dy)[g(x)-g(y)]\end{aligned} \tag{5.5.29}$$

Symmetrising between the first and the last expression, we get

$$\begin{aligned}
\mathscr{E}(f,g) &= \frac{1}{2}\int\int \mu(dx)P(x,dy)\,[g(x)(f(x)-f(y))+(g(x)-g(y))f(x)] \qquad (5.5.30)\\
&= \frac{1}{2}\int\int \mu(dx)P(x,dy)[f(x)-f(y)][g(x)-g(y)]\\
&\quad+\frac{1}{2}\int\int \mu(dx)\,[g(x)f(x)-P(x,dy)g(y)f(y)]\\
&= \frac{1}{2}\int\int \mu(dx)P(x,dy)[f(x)-f(y)][g(x)-g(y)].
\end{aligned}$$

In the last equality we used of course the invariance of the measure μ. The final expression is manifestly is a non-negative definite quadratic form. □

An important rôle will be played by the analog of the two Green identities for sums.

Lemma 5.18. *Let $f,g\in L^2(S,\mu)$ and let $D\subset S$. Assume that P is reversible with e respect to μ. Then (first Green identity)*

$$\begin{aligned}
&\int_D\int_D \mu(dx)P(x,dy)[f(x)-f(y)][g(x)-g(y)] \qquad (5.5.31)\\
&= -2\int_D \mu(dx)f(x)(Lg)(x)+2\int_D\int_{D^c}\mu(dx)f(x)P(x,dy)[g(x)-g(y)]
\end{aligned}$$

and (second Green identity)

$$\begin{aligned}
&\int_D \mu(dx)\,[f(x)(Lg)(x)-g(x)(Lf)(x)] \qquad (5.5.32)\\
&= \int_D\int_{D^c}\mu(dx)P(x,dy)\,(g(x)[f(x)-f(y)]-f(x)[g(x)-g(y)])\\
&= \int_{D^c}\mu(dy)\,(g(y)(Lf)(y)-f(y)(Lg)(y))\,.
\end{aligned}$$

Proof. To proof the first Green identity, we proceed as in the proof of Lemma 5.17. If $D=S$, then this gives (5.5.31) without the last term. If $D\subsetneq S$, then to produce the full action of L in the second term, we must add terms that are not present and involve points in D^c. These are exactly compensated by the last term.

The first equality in the second Green identity is a trivial consequence of the first Green identity. To get the second line, use reversibility and fill up terms to recover the full action of L by adding a zero consisting of double sums over D^c. Note that the equality between the first and last line is just the statement that L is symmetric. □

Definition 5.19. Let $A,B\subset S$ be two disjoint sets. Then the capacity of the capacitor A,B is defined as

$$\mathrm{cap}(A,B)\equiv\sum_{x\in A}\mu(x)e_{A,B}(x). \qquad (5.5.33)$$

The first Green identity provides an important alternative representation.

Lemma 5.20. *Let $A,B \subset S$ be disjoint. Then* $\mathrm{cap}(A,B)$ *defined in* (5.5.33) *can be expressed as*

$$\mathrm{cap}(A,B) = \mathscr{E}(h_{A,B}, h_{A,B}). \tag{5.5.34}$$

Proof. Just use Lemma 5.18 with $f = g = h_{A,B}$ in combination with the defining properties of the equilibrium potential $h_{A,B}$. $\square$

We have seen that the Dirichlet form computed on the equilibrium potential gives the capacity. We will now show that the equilibrium potential is the solution of a variational problem.

Theorem 5.21. *Let A,B,D be as in the definition of the Dirichlet problem. Let $\mathscr{H}_{A,B}$ be the space of continuous functions f on S such that*

(i) $\mathscr{E}(f,f) < \infty$.
(ii) $f \geq 1$ *on A and* $f \leq 0$ *on B.*

Assume moreover that the corresponding Dirichlet problem has a unique solution, the equilibrium potential $h_{A,B}$. Then

$$\mathrm{cap}(A,B) = \inf_{f \in \mathscr{H}_{A,B}} \mathscr{E}(f,f). \tag{5.5.35}$$

Moreover, if $\mathscr{H}_{A,B} \neq \emptyset$, then the infimum in (5.5.35) *is achieved uniquely on the equilibrium potential, i.e.,* $\mathrm{cap}(A,B) = \mathscr{E}(h_{A,B}, h_{A,B})$.

Proof. Suppose that $\mathscr{H}_{A,B} \neq \emptyset$. Let g be a function with $\mathscr{E}(g,g) < \infty$ such that $g \leq 0$ on A and $g \geq 0$ on both B. Then, for $h \in \mathscr{H}_{A,B}$,

$$\begin{aligned} &\mathscr{E}(h+\varepsilon g, h+\varepsilon g) - \mathscr{E}(h,h) = \varepsilon[\mathscr{E}(h,g) + \mathscr{E}(g,h)] + \varepsilon^2 \mathscr{E}(g,g) \\ &= 2\varepsilon \int_A \mu(dx) g(x)(Lh)(x+) + 2\varepsilon \int_B \mu(dx) g(x)(Lh)(x) \\ &+ 2\varepsilon \int_D \mu(dx) g(x)(Lh)(x) + \varepsilon^2 \mathscr{E}(g,g). \end{aligned} \tag{5.5.36}$$

If h is the equilibrium potential, then the integrals over A and B are positive, since for $x \in A$

$$(Lh)(x) = \int_S P(x,dy)(h(y) - h(x)) = \int_D P(x,dy)(h(y) - 1) \leq 0,$$

and for $x \in B$

$$(Lh)(x) = \int_S P(x,dy)(h(y) - h(x)) = \int_D P(x,dy)(h(y)) \geq 0.$$

The integral over D vanishes since h is harmonic in D and the remaining term is manifestly non-negative. Thus, h is a global minimum of $\mathscr{E}$ in $\mathscr{H}_{A,B}$. Finally, suppose that there is another function f such that $\mathscr{E}(f,f) = \mathscr{E}(h,h)$. Then the identity

$$\mathscr{E}\left(\tfrac{f+h}{2}, \tfrac{f+h}{2}\right) + \mathscr{E}\left(\tfrac{f-h}{2}, \tfrac{f-h}{2}\right) = \tfrac{1}{2}\mathscr{E}(f,f) + \tfrac{1}{2}\mathscr{E}(h,h) \tag{5.5.37}$$

implies that

$$\mathscr{E}\left(\frac{f+h}{2},\frac{f+h}{2}\right) \leq \mathscr{E}(h,h) - \mathscr{E}\left(\frac{f-h}{2},\frac{f-h}{2}\right). \tag{5.5.38}$$

Since h is an absolute minimum, this inequality can only hold if

$$\mathscr{E}(f-h,f-h) = 0. \tag{5.5.39}$$

But the latter means that μ-a.s., $(L(f-g))(x) = 0$ and since outside of D, it holds that $f = g$, uniqueness of the solution of the Dirichlet problem implies $f = g$. $\square$

5.6 Doob's h-transform

Let us consider a discrete time Markov process, X, with generator $L = P - \mathbb{1}$ given. We may want to consider modification of the process. One important type of conditioning is that to reach some set in particular places (e.g. consider a random walk in a finite interval; we may be interested to consider this walk conditioned on the fact that it exits on a specific side of the interval; this may correspond to consider a sequence of games conditioned on the player to win).

How and when can we do this, and what is the nature of the resulting process? In particular, is the resulting process again a Markov process, and if so, what is its generator?

As an example, let us try to condition a Markov process to hit a domain B for the first time in a subset $A \subset B$. We may assume that $\mathbb{E}\tau_B < \infty$. Define $h(x) \equiv \mathbb{P}_x[\tau_A = \tau_B]$, if $x \notin B$. Let $\mathbb{P}$ be the law of X. Let us define a new measure, $\mathbb{P}^h$, on the space of paths as follows: If Y is a $\mathfrak{F}_t$-measurable random variable, then

$$\mathbb{E}^h[Y|\mathfrak{F}_0] = \frac{1}{h(X_0)}\mathbb{E}[h(X_t)Y|\mathfrak{F}_0]. \tag{5.6.1}$$

Lemma 5.22. *With the notation above, if Y is a $\mathfrak{F}_{\tau_B-1}$-measurable function,*

$$\mathbb{E}^h_x[Y] = \mathbb{E}_x[Y|\tau_A = \tau_B]. \tag{5.6.2}$$

Proof. This is an application of the strong Markov property. We have by definition

$$
\begin{aligned}
\mathbb{E}_x^h[Y] &= \frac{1}{h(x)}\mathbb{E}_x[Yh(X_{\tau_B-1})] && (5.6.3)\\
&= \frac{1}{h(x)}\mathbb{E}_x\left[Y\mathbb{E}'_x[\mathbb{1}_{\tau_A=\tau_B}|\mathfrak{F}'_0](X_{\tau_B-1})\right]\\
&= \frac{1}{h(x)}\mathbb{E}_x\left[Y\mathbb{1}_{\tau_A=\tau_B}\circ\theta_{\tau_B-1}\right]\\
&= \frac{1}{\mathbb{P}_x[\tau_A=\tau_B]}\mathbb{E}_x\left[Y\mathbb{1}_{\tau_A=\tau_B}\right]\\
&= \mathbb{E}_x[Y|\tau_A=\tau_B].
\end{aligned}
$$

Here the first equality is just the definition of h and reproduces the form of the right-hand side of the strong Markov property; the second equality is the strong Markov property; the last equality uses that fact that the event $\{\tau_A=\tau_B\}$ depends only on what happens after τ_B-1, and so $\mathbb{1}_{\tau_A=\tau_B}\circ\theta_{\tau_B-1}=\mathbb{1}_{\tau_A=\tau_B}$. □

Let us now look at the transformed measure $\mathbb{P}^h$ in the general case. The first thing to check is of course whether this defines in a consistent way a probability measure. Some thought shows that all that is to show for this is the following lemma.

Lemma 5.23. *Let Y be $\mathfrak{F}_s$-measurable. Then, for any $t\geq s$,*

$$
\mathbb{E}^h[Y|\mathfrak{F}_0]\equiv\frac{1}{h(X_0)}\mathbb{E}[h(X_s)Y|\mathfrak{F}_0]=\frac{1}{h(X_0)}\mathbb{E}[h(X_t)Y|\mathfrak{F}_0]. \qquad (5.6.4)
$$

In particular, $\mathbb{P}^h[\Omega|\mathfrak{F}_0]=1$.

Proof. Just introduce a conditional expectation:

$$
\mathbb{E}[h(X_t)Y|\mathfrak{F}_0]=\mathbb{E}[\mathbb{E}[h(X_t)Y|\mathfrak{F}_s]|\mathfrak{F}_0]=\mathbb{E}[Y\mathbb{E}[h(X_t)|\mathfrak{F}_s]|\mathfrak{F}_0], \qquad (5.6.5)
$$

and use that $h(X_t)$ is a martingale

$$
=\mathbb{E}[Yh(X_s)|\mathfrak{F}_0],
$$

from which the result follows. □

This lemma shows why it is important that h be a harmonic function.
Now we turn to the question of whether the law $\mathbb{P}^h$ is a Markov chain.

Theorem 5.24. *Let X be a Markov chain with generator L and law $\mathbb{P}$. Let h be a harmonic function. Then the h-transformed measure, $\mathbb{P}^h$, is the law of a Markov process with generator L^h, where for any bounded measurable function f,*

$$
L^hf(x)\equiv\frac{1}{h(x)}\int_S P(x,dy)h(y)f(y)-f(x). \qquad (5.6.6)
$$

Proof. To proof this theorem we turn to the martingale problem. We will show that for L^h defined by (5.6.6),

$$M_t^h \equiv f(X_t) - f(X_0) - \sum_{s=0}^{t-1} (L^h f)(X_s) \tag{5.6.7}$$

is a martingale under the law $\mathbb{E}^h$, i.e. that, for $t > t'$,

$$\mathbb{E}^h[M_t^h | \mathfrak{F}_{t'}] = M_{t'}^h. \tag{5.6.8}$$

Note first that, by definition

$$\begin{aligned}\mathbb{E}^h[M_t^h | \mathfrak{F}_{t'}] = \; & \frac{1}{h(X_{t'})} \mathbb{E}[h(X_t) f(X_t) | \mathfrak{F}_{t'}] - f(X_0) - \sum_{s=0}^{t'-1} (L^h f)(X_s) \\ & - \sum_{s=t'}^{t-1} \frac{1}{h(X_{t'})} \mathbb{E}[h(X_s) L^h f(X_s) | \mathfrak{F}_{t'}].\end{aligned} \tag{5.6.9}$$

The middle terms are part of $M_{t'}^h$ and we must consider $\mathbb{E}[f(X_t)h(X_t)|\mathfrak{F}_{t'}]$. This is done by applying the martingale problem for $\mathbb{P}$ and the function fh. This yields

$$\mathbb{E}[f(X_t)h(X_t)|\mathfrak{F}_{t'}] = f(X_{t'})h(X_{t'}) + \sum_{s=t'}^{t-1} \mathbb{E}[(L(fh))(X_s)|\mathfrak{F}_{t'}]$$

Inserting this in (5.6.9) gives

$$\begin{aligned}\mathbb{E}^h[M_t^h | \mathfrak{F}_{t'}] = \; & f(X_{t'}) - f(X_0) - \sum_{s=0}^{t'-1} (L^h f)(X_s) \\ & + \frac{1}{h(X_{t'})} \sum_{s=t'}^{t-1} \Big[\mathbb{E}[(L(fh))(X_s)|\mathfrak{F}_{t'}] - \mathbb{E}[h(X_s) L^h f(X_s) | \mathfrak{F}_{t'}] \Big] \\ = \; & M_{t'}^h \\ & + \frac{1}{h(X_{t'})} \sum_{s=t'}^{t-1} \Big[\mathbb{E}[(L(fh))(X_s)|\mathfrak{F}_{t'}] - \mathbb{E}[h(X_s) L^h f(X_s) | \mathfrak{F}_{t'}] \Big].\end{aligned}$$

The second term will vanish if we choose $L^h f(x) = h(x)^{-1}(L(hf))(x)$, i.e. as defined in (5.24).

Hence we see that under $\mathbb{P}^h$, X solves the martingale problem corresponding to the generator L^h, and so is a Markov chain with transition kernel $P^h = L^h + \mathbb{1}$. The process X under $\mathbb{P}^h$ is called the (Doob) h-transform of the original Markov process. □

Exercise. As a simple example, consider a simple random walk on $\{-N, -N+1, \ldots, N\}$. Assume we want to condition this process on hitting $+N$ before $-N$. Then let

$$h(x) = \mathbb{P}_x[\tau_N = \tau_{\{N\}\cup\{-N\}}] = \mathbb{P}_x[\tau_N < \tau_{-N}].$$

Compute $h(x)$ and use this to compute the transition rates of the h-transformed walk? Plot the probabilities to jump down in the new chain!

5.7 Markov chains with countable state space

The setting of discrete time Markov chains does in some sense not go too well with general state spaces. In fact, in these cases, it is usually more appropriate to consider continuous time. Here we provide some results on Markov chains with countable state space, in particular introduce the notions of recurrence and transience and discuss the existence and uniqueness of invariant distributions.

Much of the theory of Markov chains with countable state space is similar to the case of finite state space. In particular, the notions of communicating classes, irreducibility, and periodicity carry over. There are, however, important new concepts in the case when the state space is infinite. These are the notions of *recurrence* and *transience*. It will be useful to use a notation close to the matrix notation of finite chains. Thus we set

$$P(i,\{j\}) = p(i,j) \tag{5.7.1}$$

We will place us in the setting of an irreducible Markov chain, i.e. the all states in S communicate (i.e. for any $i,j \in S$, $\mathbb{P}_j(\tau_j < \infty) > 0$). We may also for simplicity assume that our chain is aperiodic. In the case of finite state space, we have seen that such chains are ergodic in the sense that there exists a unique invariant probability distribution, and the marginal distributions at time t, converge to this distribution independently of the starting measure. Essentially this is true because the chain is trapped on the finite set. If S is infinite, a new phenomenon is possible: the chain may run "to infinity".

Definition 5.25. Let X be an irreducible aperiodic Markov chain with countable state space S. Then:

(i) X is called *transient*, if for any $i \in S$,

$$\mathbb{P}_i(\tau_i < \infty) < 1; \tag{5.7.2}$$

(ii) X is called *recurrent*, if it is not transient.
(iii) X is called *positive recurrent* or *ergodic*, if, for all $i \in S$,

$$\mathbb{E}_i(\tau_i) < \infty. \tag{5.7.3}$$

Remark. The notion of recurrence and transience can be defined for states rather than for the entire chain. In the case of irreducible and aperiodic chains, all states have the same characteristics.

Some simple consequences of the definition are the following.

Lemma 5.26. *Let X be a Markov chain with countable state space be irreducible. Then X is transient, iff*

$$\mathbb{P}_\ell\left(X_t = \ell, \text{i.o.}\right) = 0. \tag{5.7.4}$$

Proof. Assume that X is transient. Then $\mathbb{P}_\ell(\tau_\ell < \infty) = c < 1$. By the first Borel-Cantelli lemma, (5.7.4) holds if

$$\sum_{t=0}^{\infty} \mathbb{P}_\ell(X_t = \ell) < \infty. \tag{5.7.5}$$

But

$$\sum_{t=0}^{\infty} \mathbb{P}_\ell(X_t = \ell) = \mathbb{E}_\ell\left(\sum_{t=0}^{\infty} \mathbb{1}_{X_t=\ell}\right) = \sum_{n=1}^{\infty} n\mathbb{P}_\ell\left(X_t = \ell, n\text{-times}\right). \tag{5.7.6}$$

Using the strong Markov property,

$$\mathbb{P}_\ell\left(X_t = \ell, n\text{-times}\right) = \mathbb{P}_\ell(\tau_\ell < \infty)^n \mathbb{P}_\ell(\tau_\ell = \infty) = c^n(1-c). \tag{5.7.7}$$

Inserting this equality into (5.7.12) yields that (5.7.11) holds and thus that (5.7.4) is true.

To show the converse, assume that (5.7.4) holds. Then

$$\begin{aligned} 1 = 1 - \mathbb{P}\left(X_t = \ell, \text{i.o.}\right) &= \mathbb{P}\left(X_t = \ell, \text{finitely many times}\right) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_\ell\left(X_t = \ell, n\text{-times}\right) = \sum_{n=0}^{\infty} c^n(1-c). \end{aligned} \tag{5.7.8}$$

The latter sum equals 1 if and only if $c < 1$. Thus X is transitive. $\square$

Positive recurrent chains are called ergodic, because they are ergodic in the same sense as finite Markov chains.

Lemma 5.27. *Let X positive recurrent Markov chain with countable state space, S. Then, for any $j, \ell \in S$,*

$$\mu(j) \equiv \frac{\mathbb{E}_\ell\left(\sum_{t=1}^{\tau_\ell} \mathbb{1}_{X_t=j}\right)}{\mathbb{E}_\ell \tau_\ell}. \tag{5.7.9}$$

is the unique invariant probability distribution of X.

Proof. Define $\nu_\ell(j) = \mathbb{E}_\ell\left[\sum_{t=1}^{\tau_\ell} \mathbb{1}_{X_t=j}\right]$. We show first that ν is an invariant measure. Obviously, $1 = \sum_{m\in S} \mathbb{1}_{X_{\ell-1}=m}$, and hence, using the strong Markov property,

$$
\begin{aligned}
\nu_\ell(j) \equiv \mathbb{E}_\ell\left[\sum_{t=1}^{\tau_\ell} \mathbb{1}_{X_t=j}\right] &= \mathbb{E}_\ell\left[\sum_{m\in S}\sum_{t=1}^{\tau_\ell} \mathbb{1}_{X_t=j}\mathbb{1}_{X_{t-1}=m}\right]\\
&= \sum_{m\in S}\mathbb{E}_\ell\left[\sum_{t=1}^{\tau_\ell} \mathbb{1}_{X_{t-1}=m}\mathbb{P}[X_t=j|\mathfrak{F}_{t-1}](m)\right]\\
&= \sum_{m\in S}\mathbb{E}_\ell\left[\sum_{t=1}^{\tau_\ell} \mathbb{1}_{X_{t-1}=m}\right]p(m,j)\\
&= \sum_{m\in S}\mathbb{E}_\ell\left[\sum_{t=1}^{\tau_\ell} \mathbb{1}_{X_t=m}\right]P(m,j)\\
&= \sum_{m\in S}\nu_\ell(m)P(m,j)
\end{aligned}
$$

Thus μ_ℓ solves the invariance equation and thus is an invariant measure. It remains to show that ν_ℓ is normalisable. But

$$\sum_{j\in\Sigma}\nu_\ell(j) = \mathbb{E}_\ell(t_\ell) < \infty,$$

by assumption. Thus $\nu_\ell(j)/\sum_{i\in S}\nu_\ell(i) = \mu(j)$ is an invariant probability distribution.

Next we want to show uniqueness. Note first that for any irreducible Markov chain (with discrete state space) it holds that, if μ is an invariant measure and $\mu(i) = 0$, for some $i \in S$, then $\mu \equiv 0$. Namely, if for some j,$\mu(j) > 0$, then there exists t finite such that $P^t_{ji} > 0$, and $\mu(i) \geq \mu(j)P^t_{ji} > 0$, in contradiction to the hypothesis.

We will now actually show that ν_ℓ is the only invariant measure such that $\nu_\ell(\ell) = 1$ (which implies the desired uniqueness result immediately). To do so, we will show that for any other invariant measure, ν, such that $\nu(\ell) = 1$, we have that $\nu(j) \geq \nu_\ell(j)$ for all j. For then, $\nu - \nu_\ell$ is a positive invariant measure as well, and being zero in ℓ, must vanish identically. Hence $\nu = \nu_\ell$.

Now we clearly have that

$$\nu(i) = \sum_{j\neq\ell} p(j,i)\nu(j) + p(\ell,i), \tag{5.7.10}$$

since $\nu(\ell) = 1$, by hypothesis. We want to think of $p(\ell,i)$ as

$$p(\ell,i) = \mathbb{E}_\ell\left(\mathbb{1}_{\tau_\ell\geq 1}\mathbb{1}_{X_s=i}\right).$$

Now iterate the same relation in the first term in (5.7.10). Thus

$$
\begin{aligned}
\nu(i) &= \sum_{j_1,j_2\neq\ell} p(j_2,j_1)p(j_1,i)\nu(j_2) + \sum_{j_1\neq\ell} p(\ell,j_1)p(j_1,i) + \mathbb{E}_\ell\left(\mathbb{1}_{\tau_\ell\geq 1}\mathbb{1}_{X_1=i}\right)\\
&= \sum_{j_1,j_2\neq\ell} p(j_2,j_1)p(j_1,i)\nu(j_2) + \mathbb{E}_\ell\left(\sum_{s=1}^{2\wedge\tau_\ell}\mathbb{1}_{X_s=i}\right).
\end{aligned}
\tag{5.7.11}
$$

Further iteration yields for any $n \in \mathbb{N}$

$$\begin{aligned}\nu(i) &= \sum_{j_1, j_2, \dots j_n \neq \ell} p(j_n, j_{n-1}) \dots p(j_2, j_1) p(j_1, i) \nu(j_n) + \mathbb{E}_\ell \left(\sum_{s=1}^{n \wedge \tau_\ell} \mathbb{1}_{X_s = i} \right) \\ &\geq \mathbb{E}_\ell \left(\sum_{s=1}^{n \wedge \tau_\ell} \mathbb{1}_{X_s = i} \right). \end{aligned} \tag{5.7.12}$$

This implies $\nu(i) \geq \nu_\ell(i)$, as desired, and the proof is complete. $\square$

Corollary 5.28. *An ergodic Markov chain satisfies*

$$\mu(j) = \frac{1}{\mathbb{E}_j(\tau_j)}. \tag{5.7.13}$$

Proof. Just set $\ell = j$ in the definition of $\mu(j)$, and note that $\nu_j(j) = \mathbb{E}_j(\sum_{t=1}^{\tau_j} \mathbb{1}_{X_t = x}) = 1$. $\square$

We have seen that positive recurrence is needed to ensure the existence of an invariant probability measures. Next we show that if the chain is in addition aperiodic, we get convergence towards this invariant measure.

Let us show first that the existence of strictly positive invariant probability measure ensures positive recurrence.

Lemma 5.29. *Let X be an irreducible Markov chain with countable state space. If there exists an invariant probability measure μ, then $\mu(i) = 1/\mathbb{E}_i \tau_i$, and X is positive recurrent.*

Proof. Since μ is a probability measure, due to irreduciblity for any ℓ there exists n such that ℓ für geeignetes n gelten, dass $\mu(\ell) = \sum_{i \in S} \mu(i)(p^n)_{i\ell} > 0$. Then $\lambda(j) \equiv \mu(j)/\mu(\ell)$ is an invariant measure satisfying $\lambda(i) = 1$. We have seen above that $\lambda(k) \geq \nu_\ell(k)$. Hence

$$\mathbb{E}_\ell \tau_\ell = \sum_{i \in S} \nu_\ell(i) \leq \sum_{i \in S} \frac{\mu(i)}{\mu(\ell)} = \frac{1}{\mu(\ell)} < \infty. \tag{5.7.14}$$

Therefore X is positive recurrent. $\square$

We can now state our first ergodic theorem.

Theorem 5.30. *Let X be an irreducible, aperiodic, and positive recurrent Markov chain with countable state space. Let P denote its transition kernel and μ its unique invariant probability measure. Then, for any initial distribution π_0, we have that for all $i \in S$,*

$$\lim_{n \uparrow \infty} (\pi_0 P^n)_i = \mu(i). \tag{5.7.15}$$

Proof. The proof uses the method of "coupling". Let π_0 be our initial distribution.die Anfangsverteilung unserer Kette X. We construct a second Markov chain, indpendent of X with the same transition kernel but initial distribution μ. Then we define the stopping time T with respect to the filtration $\mathfrak{F}_n \equiv \sigma(X_0, Y_0, X_1, Y_1, \dots, X_n, Y_n)$ as

$$T \equiv \inf\{n : X_n = Y_n = i\}, \tag{5.7.16}$$

where $i \in S$ is an arbitrary state in S.

We show first that T is almost surely finite. To do this, we consider the pair $W = (X,Y)$ as a Markov chain with state space $S \times S$. Its transition kernel $\widetilde{P}$ has elements

$$\tilde{p}_{(ik)(jm)} \equiv p_{ij} p_{km}. \tag{5.7.17}$$

The initial distribution of this chain is $\tilde{\pi}_0((jk)) = \pi_0(j)\mu(k)$. Since P is irreducible and aperiodic, for any i, j, k, ℓ there exists n, such that

$$\tilde{p}^n_{(ik)(jm)} = p^n_{ij} p^n_{km} > 0. \tag{5.7.18}$$

Hence W is irreducible. Furthermore, it is evident that the invariant distribution of W is given by $\tilde{\mu}$

$$\tilde{\mu}((jk)) = \mu(j)\mu(k) > 0. \tag{5.7.19}$$

Hence W is positive recurrent. Since $T = \inf\{n \geq 0 : W_n = (ii)\}$, we have $\mathbb{E}T < \infty$ and hence $\mathbb{P}(T < \infty) = 1$.

Next we construct a new Markov chain with state space S as

$$Z_n = \begin{cases} X_n, & \text{wenn}\, n < T \\ Y_n, & \text{wenn}\, n \geq T. \end{cases} \tag{5.7.20}$$

This chain has the same law as X. It follows that

$$\begin{aligned} \mathbb{P}(X_n = i) &= \mathbb{P}(Z_n = i) \\ &= \mathbb{P}(Z_n = i \wedge \{n < T\}) + \mathbb{P}(Z_n = i \wedge \{n \geq T\}) \\ &= \mathbb{P}(X_n = i \wedge \{n < T\}) + \mathbb{P}(Y_n = i \wedge \{n \geq T\}) \\ &= \mathbb{P}(Y_n = i) + -\mathbb{P}(Y_n = i \wedge \{n < T\}) + \mathbb{P}(X_n = i|\{n < T\}) \\ &= \mu(i) + (\mathbb{P}(Y_n = i|n < T) - \mathbb{P}(X_n = i|n < T))\,\mathbb{P}(n < T). \end{aligned} \tag{5.7.21}$$

The expression in the brackets is smaller than one while the coefficient $\mathbb{P}(n < T)$ tends to zero, as $n \uparrow \infty$. This proves the theorem. □

Remark. Note that both irreducibility and aperiodicity were used in the proof. It is clear from elementary examples that the conclusion cannot hold in for periodic Markov chains.

Let us remark that for any transient states i of a Markov chainetransiente Zustände, i, it holds thst for any $j \in S$ and any invariant measure, μ,

$$\lim_{n\uparrow\infty}(p^n)_{ji} = 0 = \mu(i).$$

Namely, by Lemma 5.7.15

$$\sum_{n=0}^{\infty}(p^n)_{ji} \leq \mathbb{E}_i\left(\sum_{n=0}^{\infty}\mathbb{1}_{X_n=i}\right) < \infty.$$

This implies the claim.

Finally we note that the stong ergodic theorem that we know for irreducible Markov chains with finite state space holds also for positive recurrent chains with countable state space. The proof is identical to that in the finite state space, given that we already know existence und uniqueness of an invariant probability measure.

Chapter 6
Random walks and Brownian motion

The goal of this chapter is to introduce Brownian motion as a continuous time stochastic process with continuous paths and to explain its connection to random walks through Donsker's invariance principle. A very detailed source on Brownian motion is the classical book by Itô and McKean [7].

6.1 Random walks

The innocent looking stochastic processes

$$S_n \equiv \sum_{i=1}^{n} X_i, \tag{6.1.1}$$

with $X_i, i \in \mathbb{N}$ iid random variables are generally called *random walks* and receive a considerable attention in probability theory. A special case is the so-called *simple random walk on* $\mathbb{Z}^d$, characterized by the fact that the random variables X_i take values in the set of $\pm$ unit vectors in the lattice $\mathbb{Z}^d$. Consequently, $S_n \in \mathbb{Z}^d$, is a stochastic process with discrete state space. Obviously, S_n is a Markov chain, and, moreover, the coordinate processes, S_n^μ, $\mu = 1, \ldots d$, are sub-, super-, or martingales, depending on whether $\mathbb{E}(X_0^\mu)$ is positive, negative, or zero.

Let us focus on the centred case, $\mathbb{E}(X_1) = 0$. In this case we have seen that $Z_n \equiv n^{-1/2} S_n$ converges in distribution to a Gaussian random variable. By considering the process coordinate wise, it will also be enough to think about $d = 1$. We now want to extend this result to a convergence result on the level of stochastic process. That is, rather than saying something about the position of the random walk at a time n, we want to trace the entire trajectories of the process and try give a description of their statistical properties in terms of some limiting stochastic process.

It is rather clear from the central limit theorem that we must consider a rescaling like

$$Z_n(t) \equiv n^{-1/2} \sum_{k=1}^{[tn]} X_k. \tag{6.1.2}$$

In that case we have from the central limit theorem, that for any $t \in (0,1]$,

$$Z_n(t) \xrightarrow{\mathscr{D}} B_t,$$

($[x]$ denotes the lower integer part of x) where B_t is a centred Gaussian random variable with variance t. Moreover, for any finite collection of indices $t_1, \dots, t_\ell$, define $Y_n(i) \equiv Z_n(t_i) - Z_n(t_{i-1})$. Then the random variables $Y_n(i)$ are independent and it is easy to see that they converge, as $n \to \infty$, jointly to a family of independent centered Gaussian variables with variances $t_i - t_{i-1}$. This implies that the finite dimensional distributions of the processes $Z_n(t), t \in (0,1]$, converge to the finite dimensional distributions of the Gaussian process with covariance $C(s,t) = s \wedge t$, that we introduced in Section 3.3.2 and that we have preliminarily called Brownian motion.

We now want to go a step further and discuss the properties of the paths of our processes.

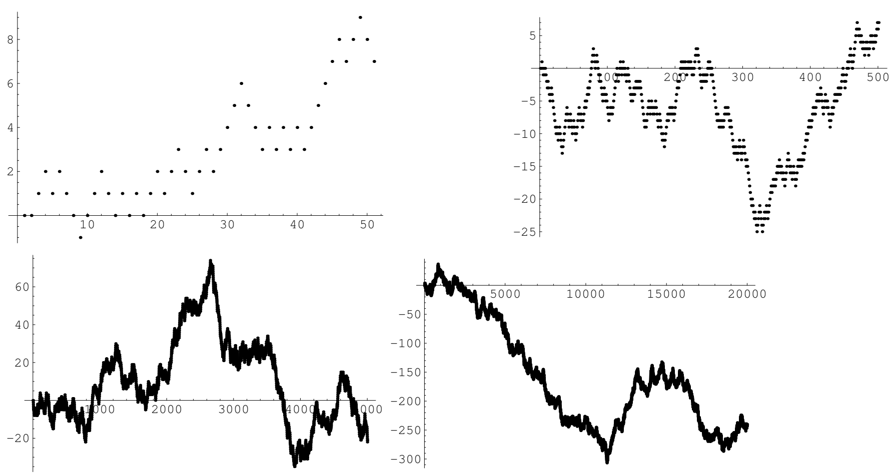


Fig. 6.1 Paths of S_n for various values of n.

From looking at pictures, it is clear that the limiting process B_t should have rather continuous looking sample paths.

6.2 Construction of Brownian motion

Before stating the desired convergence result, we have to define and construct the limiting object, the Brownian motion.

Definition 6.1. A stochastic process $\{B_t \in \mathbb{R}^d, t \in \mathbb{R}_+\}$, defined on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, is called a d-dimensional Brownian motion starting in 0, iff

(o) $B_0 = 0$, a.s..
(i) For any $p \in \mathbb{N}$, and any $0 = t_0 < t_1 < \cdots < t_p$, the random variables $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_p} - B_{t_{p-1}}$, are independent and each $B_{t_i} - B_{t_{i-1}}$ is a centered Gaussian r.v. with variance $t_i - t_{i-1}$.
(ii) For any $\omega \in \Omega$, the map $t \mapsto B_t(\omega)$ is continuous.

The question is whether such a process exists. The first property can, as we have seen, be established with the help of Kolmogorov's theorem. The problem with this is that it constructs the process on the space $((\mathbb{R}^d)^{\mathbb{R}_+}, \mathfrak{B}^{\mathbb{R}_+}(\mathbb{R}^d))$; but the second requirement, the continuity of the sample paths, is not a measurable property with respect to the product σ-algebra. Therefore, we have to proceed differently. In fact, we want to construct Brownian motion as a random variable with values in the space $C(\mathbb{R}_+, \mathbb{R}^d)$.

Theorem 6.2. *Brownian motion exists.*

Proof. We consider the case $d = 1$, the extension to higher dimensions is straightforward. We consider a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ on which an infinite family of independent standard Gaussian random variables is defined. We define the so-called *Haar-functions*, h_n^k on $[0,1]$ via

$$
\begin{aligned}
h_0^0(t) &\equiv 1, \\
h_n^k(t) &\equiv 2^{(n-1)/2} \left[\mathbb{1}_{[(2k)2^{-n},(2k+1)2^{-n})}(t) - \mathbb{1}_{[(2k+1)2^{-n},(2k+2)2^{-n})}(t) \right]
\end{aligned}
\tag{6.2.1}
$$

for $k \in \{0, \dots, 2^{n-1} - 1\}$ and $n \geq 1$. We set $I(n) \equiv \{0, \dots, 2^{n-1} - 1\}$ for $n \geq 1$ and $I(0) = \{0\}$. The functions h_n^k, $n \in \mathbb{N}$, $k \in I(n)$ form a complete orthonormal system of functions in $L^2([0,1])$, as one may easily check. Now set

$$
f_n^k(t) = \int_0^t h_n^k(u) du, \tag{6.2.2}
$$

and set

$$
B_t^{(n)} \equiv \sum_{m=0}^{n} \sum_{k \in I(m)} f_m^k(t) X_{m,k} \tag{6.2.3}
$$

for $t \in [0,1]$, where $X_{m,k}$ are our independent standard normal random variables. We will show that (i) the continuous functions $B^{(n)}(\omega)$ converge uniformly, almost surely, and hence to continuous functions, and (ii) that the covariances of $B^{(n)}$ converge to the correct limit. The limit, modified to be $B_t(\omega) \equiv 0$ when $B_t^{(n)}(\omega)$ does not converge to a continuous function, will then be Brownian Motion on $[0,1]$.

Let us now prove (i). The point here is that, of course, that the functions $f_n^k(t)$ are very small, namely,

$$
|f_n^k(t)| \leq 2^{-(n+1)/2}.
$$

Moreover, for given t, there is only one value of k such that $f_n^k(t) \neq 0$. Therefore,

$$\begin{aligned}\mathbb{P}\left[\sup_{0\le t\le 1}|B_t^{(n)}-B_t^{(n-1)}|>a_n\right] &= \mathbb{P}\left[\sup_{0\le t\le 1}|\sum_{k\in I(n)} f_n^k(t)X_{n,k}|>a_n\right] \qquad (6.2.4)\\ &= \mathbb{P}\left[\sup_{k\in I(n)}|X_{n,k}|>2^{(n+1)/2}a_n\right]\\ &\le 2^n\mathbb{P}\left[|X_{n,1}|>2^{(n+1)/2}a_n\right]\\ &\le 2^n\frac{e^{-a_n^2 2^n}}{\sqrt{\pi/2}a_n 2^{(n+1)/2}}=\frac{2^{n/2}e^{-a_n^2 2^n}}{\sqrt{\pi}a_n},\end{aligned}$$

where we used the very useful bound

$$\mathbb{P}[|X|>u]\le\frac{1}{u\sqrt{\pi/2}}e^{-u^2/2} \qquad (6.2.5)$$

for Gaussian probabilities. Now we are close to being done: Choose a sequence a_n such that $\sum_{n=0}^{\infty}a_n<\infty$ and

$$\sum_{n=1}^{\infty}\mathbb{P}\left[\sup_{0\le t\le 1}|B_t^{(n)}-B_t^{(n-1)}|>a_n\right]<\infty.$$

Clearly, the choice $a_n=2^{-n/4}$ will do. Then, by the Borel-Cantelli lemma,

$$\mathbb{P}\left[\sup_{0\le t\le 1}|B_t^{(n)}-B_t^{(n-1)}|>a_n \text{ i.o.}\right]=0,$$

and hence for all δ,

$$\mathbb{P}\left[\forall_{\delta>0}\,\exists_{n<\infty}\,\forall_{m>n}\sup_{0\le t\le 1}|B_t^{(m)}-B_t^{(n)}|<\delta\right]=1. \qquad (6.2.6)$$

which implies that almost surely, the sequence $B^{(n)}$ converges uniformly in the interval $[0,1]$. Since uniformly convergent sequences of continuous functions converge to continuous functions, $\lim_{n\to\infty}B_t^{(n)}\equiv B_t(\omega)$ in $C([0,1],\mathbb{R})$, for almost all ω.

To check (ii), we compute the covariances:

$$\begin{aligned}\mathbb{E}(B_t^{(n)}B_s^{(n)}) &= \sum_{m=0}^{n}\sum_{k\in I(m)}\sum_{m'=0}^{n}\sum_{k'\in I(m')} f_m^k(t)f_{m'}^{k'}(s)\mathbb{E}(X_{m,k}X_{m',k'})\\ &= \sum_{m=0}^{n}\sum_{k\in I(m)} f_m^k(t)f_m^k(s) \qquad (6.2.7)\\ &= \int_0^1 du\int_0^1 dv\,\mathbb{1}_{[0,t]}(u)\mathbb{1}_{[0,s]}(v)\sum_{m=0}^{n}\sum_{k\in I(m)} h_m^k(u)h_m^k(v).\end{aligned}$$

Taking the $n\to\infty$ we obtain

$$\lim_{n\to\infty} \mathbb{E}(B_t^{(n)} B_s^{(n)}) = \int_0^1 du \int_0^1 dv \mathbb{1}_{[0,t]}(u) \mathbb{1}_{[0,s]}(v) \sum_{m=0}^{\infty} \sum_{k\in I(m)} h_m^k(u) h_m^k(v)$$
$$= \int_0^1 du \mathbb{1}_{[0,t]}(u) \mathbb{1}_{[0,s]}(u) = s \wedge t \tag{6.2.8}$$

due to the fact that the system h_n^k is a complete orthonormal system. Now note that from the definition of Brownian motion, for $s < t$,

$$\mathbb{E} B_t B_s = \mathbb{E}\left[\left((B_t - B_s) + B_s\right) B_s\right] = \mathbb{E} B_s^2 = s = t \wedge s,$$

so the limiting covariance is that of Brownian motion. Finally, since $B_t^{(n)}$ are Gaussian whose covariances converge, the limit is necessarily Gaussian with the limiting covariance (Exercise! Hint: Show that the Fourier transforms converge!).

This provides B_t on $[0,1]$. To construct B_t for $t \in (k, k+1]$, just take $k+1$ independent copies of the B we just constructed, say $B_{t,1}, \ldots, B_{t,k+1}$, via

$$B_t = \sum_{i=1}^{k} B_{1,i} + B_{t-k,k+1}.$$

Finally, to construct d-dimensional Brownian motion, take d independent copies of B_t, say $B_{t,1}, \ldots, B_{t,d}$ and let e^{μ}, $\mu = 1, \ldots, d$, be a orthonormal basis of $\mathbb{R}^d$. Then set

$$\widehat{B}_t \equiv \sum_{\mu=1}^{d} e^{\mu} B_{t,\mu}. \tag{6.2.9}$$

It is easily checked that this process is a Brownian motion in $\mathbb{R}^d$. This concludes the existence proof. $\square$

Having constructed the random variable B_t in $C(\mathbb{R}_+, \mathbb{R}^d)$, we can now define its distribution, the so-called *Wiener measure*.

For this is it useful to observe that

Lemma 6.3. *The smallest σ-algebra, $\mathfrak{C}$, on $C(\mathbb{R}_+, \mathbb{R}^d)$ that makes all coordinate functions, $t \mapsto w(t)$, measurable coincides with the Borel-σ-algebra, $\mathfrak{B} \equiv \mathfrak{B}(C(\mathbb{R}_+, \mathbb{R}^d))$, of the metrisable space $C(\mathbb{R}_+, \mathbb{R}^d)$ equipped with the topology of uniform convergence on compact sets.*

Proof. First, $\mathfrak{C} \subset \mathfrak{B}$ since all functions $t \mapsto w(t)$ are continuous and hence measurable with respect to the Borel-σ-algebra $\mathfrak{B}$. To prove that $\mathfrak{B} \subset \mathfrak{C}$, we note that the topology of uniform convergence is equivalent to the metric topology relative to the metric

$$d(w, w') \equiv \sum_{n\in\mathbb{N}} 2^{-n} \sup_{0\leq t\leq n} \left(|w(t) - w'(t)| \wedge 1\right). \tag{6.2.10}$$

We thus have to show that any ball with respect to this distance is measurable with respect to $\mathfrak{C}$. But since w are continuous functions,

$$\sup_{t\in[0,n]}\left(|w(t)-w'(t)|\wedge 1\right)=\sup_{t\in[0,n]\cap\mathbb{Q}}\left(|w(t)-w'(t)|\wedge 1\right),$$

we see that e.g. the set $\{w : d(w,0) < \rho\}$ is in fact in $\mathfrak{C}$. $\square$

Note that by construction, the map $\omega \mapsto B(\omega)$ is measurable, since the maps $\omega \mapsto B_t(\omega)$ are measurable for all t, and by definition of $\mathfrak{C}$, all coordinate maps $B \mapsto B_t$ are measurable. Thus the following definition makes sense.

Definition 6.4. Let B_t a Brownian motion in $\mathbb{R}^d$ defined on a probability space $(\Omega,\mathfrak{F},\mathbb{P})$. The probability measure on $(C(\mathbb{R}_+,\mathbb{R}^d),\mathfrak{B}(C(\mathbb{R}_+,\mathbb{R}^d)))$ given as the image of $\mathbb{P}$ under the map $\omega \mapsto \{B_t(\omega), t\in\mathbb{R}_+\}$ is called the *d-dimensional Wiener measure*.

Note that uniqueness of the Wiener measure is a consequence of the Kolmogorov-Daniell theorem, since we have already seen that the finite-dimensional distributions are fixed by the prescription of the covariances.

6.3 Donsker's invariance principle

We are now in the position to prove *Donsker's theorem*.

Theorem 6.5. *Let X_i be independent, identically distributed random variables with mean zero and variance one. Let $Z_n(t)$ be as defined in (6.1.2). Then the processes $Z_n(t), t\in[0,1]$, converge in distribution to Brownian motion. More precisely, if B_t is a Brownian motion, then there exists a sequence of processes $\widetilde{Z}_n(t), t\in[0,1]$ such that the process $\widetilde{Z}_n(t), t\in[0,1]$ has the same distribution as $Z_n(t), t\in[0,1]$, and for all $\varepsilon>0$,*

$$\lim_{n\uparrow\infty}\mathbb{P}\left(\sup_{t\in[0,1]}\|\widetilde{Z}_n(t)-B_t\|>\varepsilon\right)=0. \tag{6.3.1}$$

Remark. The assertion of the theorem implies what is called *weak convergence* in the uniform topology on $[0,1]$. This means the following: Take any function $F : B([0,1],\mathbb{R}) \to \mathbb{R}$, that is continuous in the uniform topology, meaning that for any $\varepsilon>0$, one can find $\delta>0$, such that whenever two functions w,w' satisfy $\sup_{t\in[0,1]}|w(t)-w'(t)|<\delta$, then $|F(w)-F(w')|<\varepsilon$. Then

$$\lim_{n\uparrow\infty}\mathbb{E}F(Z_n)=\mathbb{E}F(B). \tag{6.3.2}$$

This is easily proven from the assertion of our theorem: First,

$$\mathbb{E}F(Z_n)=\mathbb{E}F(\widetilde{Z}_n). \tag{6.3.3}$$

Next,

$$\begin{aligned}\left|\mathbb{E}\left(F(\widetilde{Z}_n)-F(B)\right)\right| &\leq \left|\mathbb{E}\left((F(\widetilde{Z}_n)-F(B)\mathbb{1}_{\sup_{t\in[0,1]}|\widetilde{Z}_n(t)-B(t)|\leq\delta}\right)\right| \\ &+\left|\mathbb{E}\left((F(\widetilde{Z}_n)-F(B)\mathbb{1}_{\sup_{t\in[0,1]}|\widetilde{Z}_n(t)-B(t)|>\delta}\right)\right| \\ &\leq \varepsilon + C\mathbb{P}\left(\sup_{t\in[0,1]}\|\widetilde{Z}_n(t)-B_t\|>\varepsilon\right).\end{aligned} \tag{6.3.4}$$

This implies that

$$\lim_{n\uparrow\infty}\left|\mathbb{E}F(\widetilde{Z}_n)-F(B)\right|=0. \tag{6.3.5}$$

Obviously, the interval $[0,1]$ can be replaced with any other finite interval.

Proof. We will give an interesting proof of this theorem which will not use what we already know about finite dimensional distributions. For simplicity we consider the case $d=1$ only. It will be based on the famous *Skorokhod embedding*. What this will do is to construct any desired random walk from a Brownian motion. This goes a follows: we assume that F is the common distribution function of our random variables X_i, assumed to have finite second moments σ^2. We now want to construct stopping times, T, for the Brownian motion, B, such that (i) the law of B_T is F, and (ii) $\mathbb{E}(T)=\sigma^2$. This is a little tricky. First, we construct a probability measure on $(-\mathbb{R}_+)\times\mathbb{R}_+$, from the restrictions, $F_\pm$, of F to the positive and negative axis:

$$\mu(da,db)\equiv\gamma(b-a)dF_-(a)dF_+(b). \tag{6.3.6}$$

where γ provides the normalization, i.e.,

$$\gamma^{-1}=\int_0^\infty bdF_+(b)=-\int_{-\infty}^0 adF_-(a). \tag{6.3.7}$$

We need some elementary facts that follow easily once we know that B_t is a Markov chain with continuous time and generator $\Delta/2$:

Lemma 6.6. *Let $a<0<b$ and $\tau\equiv\inf\{t>0:B_t\notin(a,b)\}$. Then*
(i) $\mathbb{P}(B_\tau=a)=\frac{b}{b-a}$;
(ii)$\mathbb{E}(\tau)=|ab|$.

Proof. As we will discuss shortly, B_t is a martingale and let us anticipate that Doob's optional stopping theorem also holds for Brownian motion. Then $0=\mathbb{E}(B_\tau)=b\mathbb{P}[B_\tau=b]+a\mathbb{P}[B_\tau=a]=b+(a-b)\mathbb{P}[B_\tau=a]$, which gives (i). To prove (ii) consider

$$M_t=(B_t-a)(b-B_t)+t,$$

which is a martingale with $M_0=-ba$. On the other hand (again assuming that we can use the optional stopping theorem,

$$\mathbb{E}(M_0)=\mathbb{E}(M_\tau)=\mathbb{E}(\tau)+\mathbb{E}((B_\tau-a)(b-B_\tau))=\mathbb{E}(\tau)+0.$$

which gives the claimed result. □

The Skorokhod embedding is now constructed by choosing $\alpha < 0 < \beta$ at random from μ, and $T = \inf\{t > 0 : B_t \notin (\alpha, \beta)\}$. Then:

Theorem 6.7. *The law of B_T is F and $\mathbb{E}(T) = \sigma^2$.*

Proof. Let $b > 0$. Then

$$\mathbb{P}(B_T \in db) = \int_{-\infty}^{0} \frac{-a}{b-a} \gamma(b-a) dF_+(b) dF_-(a) = dF_+(b).$$

Analogously, for $a < 0$, $\mathbb{P}(B_T \in da) = dF_-(a)$. This proves the first assertion. Finally, by a simple computation,

$$\mathbb{E}(T) = \int_0^\infty \int_{-\infty}^0 \mu(da, db)|ab| = \int_{-\infty}^{\infty} x^2 F(dx) = \sigma^2.$$

This proves (ii). $\square$

Exercise. Construct the Skorokhod embedding for the simple random walk on $\mathbb{Z}$.

We can now define a sequence of stopping times $T_1 = T$, $T_2 = T_1 + T_2', \dots$, where T_i' are independent and constructed in the same way as T on the Brownian motions $B_{T_{i-1}+t} - B_{T_{i-1}}$. Then it follows immediately from the preceding theorem that:

Theorem 6.8. *The process $\widetilde{S}_n, n \in \mathbb{N}$ where $\widetilde{S}_n \equiv B_{T_n}$, for all $n \in \mathbb{N}$, has the same distribution as the process $S_n \equiv \sum_{i=1}^n X_i$, where X_i are iid with distribution F. Similarly, the process there are stopping times T_k^n such that $\widetilde{Z}_n(t) \equiv B_{T^n_{[nt]}}$ has the same distribution as $Z_n(t)$ and T_k^n have the same distribution as T_k/n..*

Proof. Let X_i be iid with distribution functions F. By Theorem 6.7, the random variables $\widetilde{X}_i \equiv B_{T_i} - B_{T_{i-1}}$ are iid with the same distribution as X_i. Therefore, $S_n(t)$ has the same law as B_{T_n}. Then $Z_n(t)$ has the same distribution as $n^{-1/2} B_{T_{[nt]}}$. However, we can also construct the Skorokhod embedding to reproduce the random variables $n^{-1/2} X_i$ as $B_{T_i^n} - B_{T_{i-1}^n}$. Then $Z_n(t)$ also has the same distribution as $\widetilde{Z}_n(t) \equiv B_{T^n[nt]}$.

Now we use an important property of Brownian motion:

Lemma 6.9. *For any $a \in \mathbb{R}_+$, the processes B_t and $B_t^a \equiv a^{-1} B_{ta^2}$ have the same distribution.*

Proof. Obviously, B^a is a Gaussian process. It suffices to show that B and B^a have the same covariance. But trivially

$$\mathbb{E} B_t^a B_s^a = a^{-2} \mathbb{E} B_{a^2 t} B_{a^2 s} = a^{-2} (a^2 t) \wedge (a^2 s) = s \wedge t.$$

which is the covariance of B. $\square$

From the scaling property it follows easily that T_i^n have the same law as T_i/n. This proves the theorem. $\square$

The Skorokhod embedding now provides the means to prove Donsker's theorem. Namely, we will show that the process $\widetilde{Z}_n(t)$ converges *uniformly* to B_t in probability. This is possible, since it is coupled to B_t realisationwise, unlike the original $Z_n(t)$ which would not know which particular $B_t(\omega)$ it should stick with. We will set $\sigma^2 = 1$.

Note first that by the continuity of Brownian motion, for any $\varepsilon > 0$, we can find $\delta > 0$ such that

$$\mathbb{P}(\exists u,t \in [0,1]; |u-t| \leq \delta \text{ s.d. } |B_u - B_t| > \varepsilon) \leq \varepsilon/2.$$

Next, by the independence of the T_i', and the law of large numbers,

$$\lim_{n\to\infty} \frac{T_n}{n} = \mathbb{E}(T) = 1, \text{ a.s.} \tag{6.3.8}$$

Thus

$$\lim_{n\to\infty} n^{-1} \sup_{k\leq n} |T_k - k| = 0, \text{ a.s.} \tag{6.3.9}$$

This holds since otherwise there exits with positive probability a sequence $k_n \uparrow \infty$, where $k_n n$, such that for all n, $|T_{k_n}/k_n - 1| \geq \varepsilon n/k_n \geq \varepsilon$, for some $\varepsilon > 0$. But this contradicts (6.3.8) Therefore, there exists n_1 such that for all $n \geq n_1$,

$$\mathbb{P}\left[n^{-1} \sup_{k\leq n} |T_k - k| \geq \delta/3\right] \leq \varepsilon/2. \tag{6.3.10}$$

Since T_i^n have the same law as T_i/n, this impies that also

$$\mathbb{P}\left[\sup_{k\leq n} |T_k^n - k/n| \geq \delta/3\right] \leq \varepsilon/2. \tag{6.3.11}$$

Finally, the process $\widetilde{Z}_n(t)$ will coincide for any $t = k/n$ with $B_{T_k^n}$. Therefore,

$$\begin{aligned}
&\mathbb{P}\left[\sup_{0\leq t\leq 1} \left|\widetilde{Z}_n(t) - B_t\right| \geq \varepsilon\right] \\
&\leq \mathbb{P}\left[\sup_{0\leq t\leq 1} \left|\widetilde{Z}_n(t) - B_t\right| \geq \varepsilon, |T_k^n - k/n| \leq n\delta/3, \forall_{k\leq n}\right] \\
&\quad + \mathbb{P}\left[\sup_{k\leq n} |T_k^n - k/n| \geq \delta/3\right] \\
&\leq \mathbb{P}\left[\exists k \leq n, t \in [0,1], |k/n - t| \leq \delta : \left|B_{T_k^n} - B_t\right| \geq \varepsilon\right] + \varepsilon/2 \\
&\leq \mathbb{P}\left[\exists u,t \in [0,1], |u-t| \leq \delta : |B_u - B_t| \geq \varepsilon\right] + \varepsilon/2 \leq \varepsilon..
\end{aligned}$$

This implies that the difference between $\widetilde{Z}_n(t)$ and B_t converges uniformly in $t \in [0,1]$ to zero in probability. On the other hand, $\widetilde{Z}_n(t)$ has the same law as $Z_n(t)$. This implies weak convergence as claimed. $\square$

6.4 Martingale and Markov properties

Although we have not studied with full rigor the concepts of martingales and Markov processes in continuous time, Brownian motion is a good example to get provisionally acquainted with them. The nice thing here is that we know already that it has continuous paths, so that we need not worry about discontinuities; moreover, a path is determined by knowing it on a dense set of times, say the rational numbers, so we also need not worry about unaccountability.

Proposition 6.10. *Brownian motion is a continuous time martingale, in the sense that, if $\mathfrak{F}_t$ is a filtration such that B_t is adapted, for any $s < t$,*

$$\mathbb{E}[B_t|\mathfrak{F}_s] = B_s. \tag{6.4.1}$$

Proof. Of course we have not defined what a continuous time filtration is, but we will not worry at this moment, and just take $\mathfrak{F}_t$ as the σ-algebra generated be $\{B_s\}_{s\leq t}$. Now we know that $B_t = B_t - B_s + B_s$, where $B_t - B_s$ and B_s are independent. Thus

$$\mathbb{E}[B_t|\mathfrak{F}_s] = \mathbb{E}[B_t - B_s + B_s|\mathfrak{F}_s] = \mathbb{E}[B_t - B_s|\mathfrak{F}_s] + \mathbb{E}[B_s|\mathfrak{F}_s] = 0 + B_s,$$

as claimed. $\square$

Next we show that Brownian motion is also a Markov process. As a definition of a continuous time Markov process, we adopt the obvious generalisation of (3.3.8).

Definition 6.11. A stochastic process with state space S and index set $\mathbb{R}_+$ is called a *continuous time Markov process*, if there exists a two-parameter family of probability kernels, $P_{s,t}$, satisfying the Chapman-Kolmogorov equations,

$$P_{s,t}(x,A) = \int_S P_{r,t}(y,A)P_{s,r}(x,dy), \quad \forall r \in (s,t), A \in \mathfrak{B}, \tag{6.4.2}$$

such that for all $A \in \mathfrak{B}$, $s < t \in \mathbb{R}_+$,

$$\mathbb{P}[B_t \in A|\mathfrak{F}_s](\omega) = P_{s,t}(B_s(\omega),A), \text{ a.s..} \tag{6.4.3}$$

This definition may not sound abstract enough, because it stipulates that we search for the kernels $P_{s,t}$; one may replace this by saying that

$$\mathbb{P}[B_t \in A|\mathfrak{F}_s] \tag{6.4.4}$$

is independent of the σ-algebras $\mathfrak{F}_r$, for all $r < s$; or in other words, that $\mathbb{P}[B_t \in A|\mathfrak{F}_s](\omega)$ is a function of $B_s(\omega)$, a.s.. You can see that we will have to worry a little bit about these definitions in general, but by the continuity of Brownian motion, we may just look at rational times and then no problem arises. We come to these things in the next course. We see that the two definitions are really the same, using the existence of regular conditional probabilities: namely, $P_{s,t}$ will be just the regular version of $\mathbb{P}[B_t \in A|\mathfrak{F}_s]$.

Proposition 6.12. *Brownian motion in dimension d is a continuous time Markov process with transition kernel*

$$P_{s,t}(x,A) = \frac{1}{(2\pi(t-s))^{d/2}} \int_A \exp\left(-\frac{|y-x|^2}{2(t-s)}\right) dy. \tag{6.4.5}$$

Proof. The proof is next to trivial from the defining property (i) of Brownian motion and left as an exercise. □

We now come, again somewhat informally, to the martingale problem associated with Brownian motion.

Theorem 6.13. *Let f be a two time differentiable function with bounded second derivatives. Let B_t be Brownian motion. Then*

$$M_t = f(B_t) - f(B_0) - \frac{1}{2}\int_0^t \Delta f(B_s) ds \tag{6.4.6}$$

is a martingale.

Proof. We consider for simplicity only the case $d = 1$; the general case works the same way. We proceed as in the discrete time case.

$$\begin{aligned} \mathbb{E}[M_{t+r}|\mathfrak{F}_t] &= f(B_t) - f(B_0) - \frac{1}{2}\int_0^t f''(B_s)ds \\ &\quad + \mathbb{E}[f(B_{t+r}) - f(B_t)|\mathfrak{F}_t] - \frac{1}{2}\int_0^r \mathbb{E}[f''(B_{t+s})|\mathfrak{F}_t]ds \\ &= M_t + \frac{1}{\sqrt{2\pi r}}\int_{\mathbb{R}} f(y)\exp\left(-\frac{(y-B_t)^2}{2r}\right) dy - f(B_t) \\ &\quad - \frac{1}{2}\int_0^r \frac{1}{\sqrt{2\pi s}}\int_{\mathbb{R}} f''(y)\exp\left(-\frac{(y-B_t)^2}{2s}\right) dyds \\ &= M_t \end{aligned} \tag{6.4.7}$$

The last inequality holds since, using integration by parts

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi s}}\int_{\mathbb{R}} f''(y)\exp\left(-\frac{(y-x)^2}{2s}\right)dy \qquad (6.4.8)\\
&= \int_{\mathbb{R}} f(y)\frac{d^2}{dy^2}\frac{1}{\sqrt{2\pi s}}\exp\left(-\frac{(y-x)^2}{2s}\right)dy\\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(y)\left[-s^{-3/2}+(y-x)^2 s^{-5/2}\right]\exp\left(-\frac{(y-x)^2}{2s}\right)dy\\
&= 2\int_{\mathbb{R}} f(y)\frac{d}{ds}\frac{1}{\sqrt{2\pi s}}\exp\left(-\frac{(y-x)^2}{2s}\right)dy
\end{aligned}$$

Integrating the last expression in (6.4.7) over s yields

$$\frac{2}{\sqrt{2\pi r}}\int_{\mathbb{R}} f(y)\exp\left(-\frac{(x-y)^2}{2r}\right)dy - f(x),$$

where we used that

$$\lim_{h\downarrow 0}\frac{2}{\sqrt{2\pi h}}\int_{\mathbb{R}} f(y)\exp\left(-\frac{(x-y)^2}{2h}\right)dy = f(x).$$

Inserting this into (6.4.7) concludes the proof. □

Note that we really used that the function

$$e(t,x) \equiv \frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{\|x\|^2}{2t}\right) \qquad (6.4.9)$$

satisfies the (parabolic) partial differential equation

$$\frac{\partial}{\partial t}e(x,t) = \frac{1}{2}\Delta e(x,t), \qquad (6.4.10)$$

with the (singular) initial condition

$$e(x,t) = \delta(x), \qquad (6.4.11)$$

(where δ here denotes the Dirac-delta function, i.e., for any bounded integrable function $\int_{\mathbb{R}}\delta(x)f(x)dx = f(0)$). $e(t,x)$ is called the *heat kernel* associated to (one-dimensional) Brownian motion.

Remark. Let us note that if we rewrite (6.4.6) in the form

$$f(B_t) = f(B_0) + M_t + \frac{1}{2}\int_0^t \Delta f(B_s)ds, \qquad (6.4.12)$$

it formally resembles the Itô formula (4.5.4) that we derived formally in Section 4. The martingale M_t should then play the rôle of the stochastic integral, i.e. we would like to think of

$$M_t = \int_0^t \nabla f(B_s)\cdot dB_s.$$

It will turn out that this is indeed a correct interpretation if and that (6.4.12) is the Itô formula for Brownian motion.

The preceding theorem justifies to call $L = \frac{\Delta}{2}$ the generator of Brownian motion, and to think of (6.4.7) as the associated martingale problem. The connection between Markov processes and potential theory, established for discrete time Markov processes, also carries over to Brownian motion; in this case, this links to the classical potential theory associated to the Laplace operator Δ.

6.5 Sample path properties

We have constructed Brownian motion on a space of continuous paths. What else can we say about the properties of theses paths? The striking feature is that Brownian paths are almost surely *nowhere* differentiable!

The following theorem shows that it is not even Lipshitz continuous anywhere:

Theorem 6.14. *For almost all ω, $B(\omega)$ is nowhere Lipshitz continuous.*

Proof. Let $K > 0$ and define

$$A_{n,K} \equiv \{\omega \in \Omega : \exists_{s\in[0,1]} \forall_{|t-s|\leq 2/n} |B_t - B_s| \leq K|t-s|\}. \tag{6.5.1}$$

Clearly

$$A_{n,K} \subset \cup_{k=2}^{n} \left\{ |B_{j/n} - B_{(j-1)/n}| \leq 4K/n, \text{for } j \in \{k-1, k, k+1\} \right\}. \tag{6.5.2}$$

Now

$$\begin{aligned} \mathbb{P}[A_{n,K}] &\leq (n-1)\left(\mathbb{P}[|B_{1/n} - B_0| \leq 4K/n]\right)^3 \\ &\leq (n-1)\left(\mathbb{P}[|B_{1/n}| \leq 4K/n]\right)^3 \leq Cn^{-1/2} \end{aligned} \tag{6.5.3}$$

for some finite constant $C = (8K/\sqrt{2\pi})^3$. Now $A_{n,K} \subset A_{n+1,K}$, and so for all n and all K,

$$\mathbb{P}[A_{n,K}] \leq \lim_{\ell\to\infty} \mathbb{P}[A_{\ell,K}] = 0. \tag{6.5.4}$$

Finally, by monotonicity of the Lipshitz property, it follows that

$$\mathbb{P}[\exists_{K<\infty} A_{n,K}] \leq \sum_{K\in\mathbb{N}} \mathbb{P}[A_{n,K}] = 0.$$

□

Remark. The argument used in the proof can be extended to show that Brownian motion is nowhere Hölder continuous with exponent larger than $1/2$. Namely, for $\alpha > 1/2$, let k be chosen such that $k(\alpha - 1/2) > 1$. Then define

$$A_{n,K} \equiv \{\omega \in \Omega : \exists_{s\in[0,1]} \forall_{|t-s|\leq k/n} |B_t - B_s| \leq K|t-s|^{\alpha}\}. \tag{6.5.5}$$

We then obtain that

$$\begin{aligned} \mathbb{P}[A_{n,K}] &\leq (n-1)\left(\mathbb{P}[|B_{1/n} - B_0| \leq 2kK/n^{\alpha}]\right)^k \\ &\leq (n-1)\left(\mathbb{P}[|B_{1/n}| \leq 2kK/n^{\alpha}]\right)^k \leq Cn^{-k(\alpha-1/2)+1} \end{aligned} \tag{6.5.6}$$

which yields the conclusion as in the case $\alpha = 1$.

An important notion is that of the *quadratic variation*. Let $t_k^n \equiv (k2^{-n}) \wedge t$ and set

$$[B]_t^n \equiv \sum_{k=1}^{\infty} [B_{t_k^n} - B_{t_{k-1}^n}]^2. \tag{6.5.7}$$

Lemma 6.15. *With probability one, as $n \to \infty$, $[B]_t^n \to t$, uniformly on compact intervals.*

Proof. Note that all sums over k contain only finitely many non-zero terms, and that all the summands in (6.5.12) are independent random variables, satisfying (for $t_k^n \leq t$)

$$\mathbb{E}\left(B_{t_k^n} - B_{t_{k-1}^n}\right)^2 = 2^{-n}, \tag{6.5.8}$$

$$\operatorname{var}\left(\left(B_{t_k^n} - B_{t_{k-1}^n}\right)^2\right) = 2^{-2n}. \tag{6.5.9}$$

Thus

$$\mathbb{E}[B]_t^n = t, \quad \operatorname{var}([B]_t^n) = 2^{-n}t, \tag{6.5.10}$$

and thus

$$\lim_{n\to\infty} [B]_t^n = t, \text{ a.s.}. \tag{6.5.11}$$

By telescopic expansion,

$$\begin{aligned} B_t^2 - B_0^2 &= \sum_{k=1}^{\infty} \left(B_{t_n^k}^2 - B_{t_{k-1}^n}^2\right) \\ &= \sum_{k=1}^{\infty} \left(B_{t_n^k} - B_{t_{k-1}^n}\right)\left(B_{t_n^k} + B_{t_{k-1}^n}\right) \\ &= \sum_{k=1}^{\infty} 2B_{t_{k-1}^n}\left(B_{t_k^n} - B_{t_{k-1}^n}\right) + [B]_t^n. \end{aligned} \tag{6.5.12}$$

Now set

$$V_t^n \equiv B_t^2 - [B]_t^n = \sum_{k=1}^{\infty} 2B_{t_{k-1}^n}\left(B_{t_k^n} - B_{t_{k-1}^n}\right). \tag{6.5.13}$$

One can check easily that for any n, V^n is a martingale. Then also

$$V_t^n - V_t^{n+1} = [B]_t^{n+1} - [B]_t^n \tag{6.5.14}$$

is a martingale. If we accept that Doob's $\mathscr{L}^2$-inequality (Theorem 4.21) applies in the continuous martingale case as well, we get that, for any $T < \infty$,

$$\left\| \sup_{0\leq t\leq T} \left([B]_t^{n+1} - [B]_t^n\right) \right\|_2 \leq 2 \sup_{0\leq t\leq T} \left\| [B]_t^{n+1} - [B]_t^n \right\|_2 = 2\sqrt{T2^{-n-1}}, \qquad (6.5.15)$$

where the last inequality is obtained by explicit computation. This implies that $[B]_t^n$ converges uniformly on compact intervals. $\square$

Remark. Lemma 6.15 plays a crucial rôle in stochastic calculus. It justifies the claim that $d[B]_t = dt$. If we go with this into our "discrete Itô formula (Section 4.6), this means this justifies in a more precise way the step from Eq. (4.5.2) to Eq. (4.5.4).

Remark. The definition of the quadratic variation we adopt here via di-adic partitions is different from the "true" quadratic variation that would be

$$\sup\left\{ \sum_{k=1}^n [B_{t_k} - B_{t_{k-1}}]^2,\ n\in\mathbb{N}, 0 = t_0 < t_1 < \cdots < t_n = 1 \right\},$$

which can be shown to be infinite almost surely (note that the choices of the t_i can be adapted to the specific realization of the BM). The diadic version above is, however, important in the construction of stochastic integrals.

Remark. The fact that the quadratic variation of BM converges to t implies that the linear variation,

$$\sum_{k=1}^{\infty} \left| B_{t_k^n} - B_{t_{k-1}^n} \right|$$

is infinite on every interval. This means in particular that the length of a Brownian path between any times t, t' is infinite.

6.6 The law of the iterated logarithm

How precisely random phenomena can be controlled is witnessed by the so-called *law of the iterated logarithm (LIL)*. It states (not in its most general form) that

Theorem 6.16. *Let $S_n = \sum_{i=1}^n X_i$, where X_i are independent identically distributed random variables with mean zero and variance σ^2. Then*

$$\mathbb{P}\left[\limsup_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\ln\ln n}} = 1\right] = 1. \qquad (6.6.1)$$

Remark. Just as the CLT, the LIL has extensions to the case of non-identically distributed random variables. For a host of results, see [4], Chapter 10. Furthermore, there are extensions to the case of martingales, under similar conditions as for the CLT.

The nicest proof of this result passes though the analogous result for Brownian motion and then uses the Skorokhod embedding theorem. The proof below follows [12].

Thus we want to first prove:

Theorem 6.17. *Let B_t be a one-dimensional Brownian motion. Then*

$$\mathbb{P}\left[\limsup_{t\to\infty}\frac{B_t}{\sqrt{2t\ln\ln t}}=1\right]=1, \tag{6.6.2}$$

and

$$\mathbb{P}\left[\limsup_{t\downarrow 0}\frac{B_t}{\sqrt{2t\ln\ln(1/t)}}=1\right]=1. \tag{6.6.3}$$

Proof. Note first that the two statements are equivalent since the two processes B_t and $tB_{1/t}$ have the same law (Exercise!).

We concentrate on (6.6.3). Set $h(t)=\sqrt{2t\ln\ln(1/t)}$. Basically, the idea is to use exponentially shrinking subsequences $t_n\equiv\theta^n$ in such a way that the variables B_{t_n} are essentially independent. Then, for the lower bound, it is enough to show that along such a subsequence, the $h(t_n)$ is reached infinitely often: this will prove that the lim sup is as large as claimed. For the upper bound, one shows that along such subsequences, the threshold $h(t_n)$ is not exceeded, and then uses a maximum inequality for martingales to control the intermediate values of t.

We first show that $\limsup_{t\downarrow 0}(\cdots)\leq 1$. For this we will assume that we can use Doob's submartingale inequality, Theorem 4.18 also in the continuous time case. Define

$$Z_t\equiv\exp\left(\alpha B_t-\frac{1}{2}\alpha^2 t\right). \tag{6.6.4}$$

A simple calculation shows that Z_t is a martingale (with $\mathbb{E}(Z_t)=1$), and so

$$\mathbb{P}\left[\sup_{s\leq t}(B_s-\alpha s/2)>\beta\right]=\mathbb{P}\left[\sup_{s\leq t}e^{\alpha B_s-\alpha^2 s/2}>e^{\alpha\beta}\right]\leq e^{-\alpha\beta}\mathbb{E}(Z_t)=e^{-\alpha\beta}.$$

Let $\theta,\delta\in(0,1)$, and chose $t_n=\theta^n$, $\alpha_n=\theta^{-n}(1+\delta)h(\theta^n)$, and $\beta_n=\frac{1}{2}h(\theta^n)$. Then

$$\mathbb{P}\left[\sup_{s\leq\theta^n}(B_s-\alpha_n s/2)>\beta_n\right]\leq n^{-(1+\delta)}(\ln 1/\theta)^{-(1+\delta)},$$

since $\alpha_n\beta_n=(1+\delta)\ln\ln\theta^{-n}=(1+\delta)(\ln n+\ln\ln\theta^{-1})$. Therefore, the Borel-Cantelli lemma implies that, almost surely, for all but finitely many values of n,

$$\sup_{s\leq\theta^n}\left(B_s-\frac{s}{2}(1+\delta)\theta^{-n}h(\theta^n)\right)\leq\frac{1}{2}h(\theta^n).$$

It follows that

$$\sup_{s\leq\theta^n} B_s \leq \frac{\theta^n}{2}(1+\delta)\theta^{-n}h(\theta^n) + \frac{1}{2}h(\theta^n) = \frac{1}{2}(2+\delta)h(\theta^n) \tag{6.6.5}$$

and so for any $\theta^{n+1} \leq t \leq \theta^n$,

$$B_t \leq \sup_{s\leq\theta^n} B_s \leq \frac{1}{2}(2+\delta)\theta^{-1/2}h(t), \tag{6.6.6}$$

hence, almost surely,

$$\limsup_{t\downarrow 0} B_t/h(t) \leq \frac{1}{2}\theta^{-1/2}(2+\delta). \tag{6.6.7}$$

Since this holds for any $\delta > 0$ and $\theta < 1$ almost surely, it holds along any countable subsequence $\delta_k \downarrow 0$, $\theta_k \uparrow 1$, almost surely, and

$$\limsup_{t\downarrow 0} B_t/h(t) \leq 1, \text{ a.s..} \tag{6.6.8}$$

To prove the converse inequality, consider the event

$$A_n \equiv \{B_{\theta^n} - B_{\theta^{n+1}} > (1-\theta)^{1/2}h(\theta^n)\}.$$

The events are independent, and their probability can be bounded easily using that for any $u > 0$,

$$\frac{1}{2\pi}\int_u^\infty e^{-x^2/2}dx \geq \frac{1}{u\sqrt{2\pi}}e^{-u^2/2}\left(1-2u^{-2}\right). \tag{6.6.9}$$

This implies that

$$\begin{aligned}\mathbb{P}[A_n] &= \frac{1}{\sqrt{2\pi(\theta^n(1-\theta))}}\int_{(1-\theta)^{1/2}h(\theta^n)}^\infty \exp\left(-\frac{x^2}{2\theta^n(1-\theta)}\right)dx \qquad (6.6.10)\\ &= \frac{1}{\sqrt{2\pi}}\int_{\theta^{-n/2}h(\theta^n)}^\infty \exp\left(-\frac{x^2}{2}\right)dx \\ &\geq \frac{\exp\left(-\theta^{-n}h(\theta^n)^2/2\right)}{\sqrt{2\pi}\theta^{-n/2}h(\theta^n)}\left(1-2\theta^n h(\theta^n)^{-2}\right) \equiv \gamma_n.\end{aligned}$$

Now $\theta^{-n}h(\theta^n)^2 = 2\ln n + 2\ln\ln(1/\theta)$, and so

$$\gamma_n \geq C\frac{1}{n\sqrt{\ln n}},$$

so that $\sum_n \gamma_n = +\infty$; hence, the second Borel-Cantelli lemma implies that, with probability one, A_n happens infinitely often, i.e. for infinitely many n,

$$B_{\theta^n} \geq (1-\theta)^{1/2}h(\theta^n) + B_{\theta^{n+1}}.$$

Now, the upper bound (6.6.8) also holds for $-B_t$, so that, almost surely, for all but finitely many n,

$$B_{\theta^{n+1}} \geq -h(\theta^{n+1}).$$

But by some simple estimates,

$$h(\theta^{n+1}) = \theta^{1/2} h(\theta^n) \sqrt{\frac{\ln\ln(\theta^{-n}\theta^{-1})}{\ln\ln(\theta^{-n})}} \leq \theta^{-1/2} h(\theta^n) \left(1 + O(\ln\theta^{-1}/n)\right),$$

so that, for infinitely many n,

$$B_{\theta^n} \geq \left((1-\theta)^{1/2} - 2\theta^{1/2}\right) h(\theta^n).$$

This implies that

$$\limsup_{n\to\infty} B_{\theta^n}/h(\theta^n) \geq \left((1-\theta)^{1/2} - 2\theta^{1/2}\right), \tag{6.6.11}$$

for all $\theta > 0$; hence,

$$\limsup_{t\to\infty} B_t/h(t) \geq 1, \tag{6.6.12}$$

which completes the proof. $\square$

From the LIL for Brownian motion one can prove the LIL for random walk using the Skorokhod embedding.

Proof. (of Theorem 6.16). From the construction of the Skorokhod embedding, we know that we may choose $S_n(\omega) = B_{T_n}(\omega)$. The strong law of large numbers implies that $T_n/n \to 1,$ a.s., and so also $h(T_n)/h(n) \to 1,$ a.s.. Thus the upper bound follows trivially:

$$\limsup_{n\to\infty} \frac{S_n}{h(n)} = \limsup_{n\to\infty} \frac{B_{T_n}}{h(T_n)} \leq \limsup_{t\to\infty} \frac{B_t}{h(t)} = 1. \tag{6.6.13}$$

To prove the complementing lower bound, note that by Kolmogorov's $0-1$-law, $\rho \equiv \limsup_{n\to\infty} \frac{S_n}{h(n)}$ is almost surely a constant (since the limsup is measurable with respect to the tail-σ-algebra. Assume $\rho < 1$; then, there exists $n_0 < \infty$, such that for all $n \geq n_0$, $\frac{B_{T_n}}{h(T_n)} < \rho$. We will show that this leads to a contradiction with (6.6.2) of Theorem 6.17. To show this, we must show that the Brownian motion cannot rise too far in the intervals $[T_n, T_{n+1}]$. But recall that T_{n+1} is defined as the stopping time at the random interval $[\alpha, \beta]$ of the Brownian motion B_t. We will want to show that in no such interval can the BM climb by more than $\varepsilon\sqrt{2n\ln\ln n}$. An explicit computation shows that

$$\phi(x) \equiv \mathbb{P}\left[\sup_{t\leq T_1} B_t > x\right] = \gamma \int_{-\infty}^{0} dF_-(a) \int_x^{\infty} dF_+(b)(b-a)\frac{-a}{x-a}, \tag{6.6.14}$$

where the ratio $\frac{-a}{x-a}$ is the probability that the BM reaches x before a (i.e. before T_1) (the logic of the formula is that for B_t to exceed x before T_1, the random variable β

must be larger than x, and then B_t may not hit the lower boundary before reaching x). Now we will be done by Borel-Cantelli, if

$$\sum_n \phi(\varepsilon\sqrt{2n\ln\ln n}) < \infty,$$

or in fact the stronger but simpler condition

$$\sum_n \phi(\varepsilon\sqrt{n}) < \infty \tag{6.6.15}$$

holds for all $\varepsilon > 0$. For than, except finitely often,

$$\sup_{T_n<t<T_{n+1}} B_t \leq h(n)(\rho+\varepsilon),$$

which implies

$$\limsup_{t\to\infty} \frac{B_t}{h(t)} < \rho+\varepsilon,$$

which can be made smaller than 1, thus contradicting the result for BM.

We are left we checking (6.6.15). We may decompose ϕ as

$$\begin{aligned}\Phi(x) = \gamma\int_{-\infty}^{0} dF_-(a)\int_x^\infty dF_+(b)(b-x)\frac{-a}{x-a} \qquad (6.6.16)\\ +\gamma\int_{-\infty}^{0} dF_-(a)\int_x^\infty dF_+(b)|a| \equiv \phi_1(x)+\phi_2(x).\end{aligned}$$

Now $\sum_n \phi_2(\varepsilon\sqrt{n}) < \infty$ if $\int_0^\infty \phi_2(\varepsilon\sqrt{x}) < \infty$. Recalling the formula for γ, (6.3.7), we see that

$$\int_0^\infty \phi_2(\varepsilon\sqrt{x})dx = \int_0^\infty (1-F(\varepsilon\sqrt{x}))dx = \varepsilon^{-2}\int_0^\infty (1-F(t))tdt < \mathbb{E}(X^2) < \infty.$$

To deal with ϕ_1, use that $x-a > x$, and then as before

$$\phi_1(x) \leq x^{-1}\int_x^\infty (b-x)dF_+(x)$$

Comparing the sum to an integral, we must check the finiteness of

$$\int dx\frac{1}{\varepsilon\sqrt{x}}\int_{\varepsilon\sqrt{x}}^\infty dF_+(b)(b-\varepsilon\sqrt{x}) = 2\varepsilon^{-2}\int dt\int_t^\infty dF_+(b)(b-t),$$

which again hold since F has finite second moment. This concludes the proof. $\square$

Remark. On can show more than what we did. For one thing, not only is $\limsup_t B_t/h(t) = +1$ (and hence by symmetry $\liminf_t B_t/h(t) = -1$), a.s., it is also true that the set of limit points of the process $B_t/h(t)$ is the entire interval $[-1,1]$; i.e., for any $a \in [-1,1]$, there exist subsequences t_n, such that $\lim_n B_{t_n}/h(t_n) = a$.

The following theorem, called *Lévy's theorem* , is closely related to the LIL.

Theorem 6.18. *Let B be Brownian motion. Then*

$$\mathbb{P}\left(\limsup_{\delta\downarrow 0}\sup_{t\in[0,1]}\frac{B_{t+\delta}-B_t}{\sqrt{2\delta|\ln\delta|}}=1\right)=1. \tag{6.6.17}$$

Remark. This theorem implies in particular that Brownian motion is almost surely Hölder continuous with exponent α, i.e.

$$\mathbb{P}\left(\limsup_{\delta\downarrow 0}\sup_{t\in[0,1]}\delta^{-\alpha}|B_{t+\delta}-B_t|=0\right)=1, \tag{6.6.18}$$

for any $\alpha<1/2$. This is a basic property of Brownian motion that one should memorise. But Theorem 6.18 is sharper than that. It states that almost surely, on any compact interval, there will be points where *BM* increases like $\sqrt{\delta|\ln\delta|}$, *faster* than what one would guess from the LIL, which states that at any given point, it increases like $\sqrt{\delta\ln|\ln\delta|}$!

Proof. The proof we give her is due to Lévy and differs from that of the LIL in that it does not use maximum inequalities for the upper bound, but a new technique, called *chaining*. We first prove the lower bound. For $\varepsilon\in(0,1)$, Here it is enough to exhibit candidates for the highly singular behaviour:

$$\begin{aligned}
&\mathbb{P}\left(\max_{k\leq 2^n}\left(B_{k2^n}-B_{(k-1)2^n}\right)\leq(1-\varepsilon)\sqrt{2^{1-n}\ln 2^n}\right) \qquad (6.6.19)\\
&=\left[1-\mathbb{P}\left(B_{2^{-n}}>(1-\varepsilon)\sqrt{2^{1-n}\ln 2^n}\right)\right]^{2^n}\\
&=\left[1-\mathbb{P}\left(B_1>(1-\varepsilon)\sqrt{2\ln 2^n}\right)\right]^{2^n}\\
&\leq\left[1-\frac{1}{\sqrt{2\pi 2n\ln 2}}\exp\left(-(1-\varepsilon)^2\ln 2^n\right)\right]^{2^n}\\
&\leq\exp\left(-\frac{2^{-n(1-\varepsilon)^2+n}}{\sqrt{2\pi 2n\ln 2}}\right)\sim\exp\left(-2^{2\varepsilon n}\right),
\end{aligned}$$

which tends to zero and is summable over n, for any $\varepsilon>0$. By the first Borel-Cantelli lemma, this implies that the event considered can happen only for finitely many values of n, almost surely. Thus

$$\mathbb{P}\left(\limsup_{\delta\downarrow 0}\sup_{t\in[0,1]}\frac{B_{t+\delta}-B_t}{\sqrt{2\delta|\ln\delta|}}\geq 1\right)=1.$$

The lower bound is more tricky and uses an interesting technique of chaining. We first establish that the required conditions hold on a 2^{-n} grid. By convention we set $h(\varepsilon)\equiv\sqrt{2\varepsilon|\ln\varepsilon|}$. Then we estimate

$$\begin{aligned}
&\mathbb{P}\left(\max_{j+j_i-j_2\le 2^{\varepsilon n}, j_i\le 2^n} h(j2^{-n})\left|B_{j_2 2^{-n}}-B_{j_1 2^{-n}}\right|>1+2\varepsilon\right) \qquad (6.6.20)\\
&\le 2^{(1+\varepsilon)n}\mathbb{P}\left(\left|B_{j2^{-n}}\right|>(1+2\varepsilon)h(j2^{-n})\right)\\
&= 2^{(1+\varepsilon)n}\mathbb{P}\left(|B_1|>(1+\varepsilon)\sqrt{2\ln 2^{n(1-\varepsilon)}}\right)\\
&\le 2^{(1+\varepsilon)n}\frac{2}{\sqrt{2\pi 2n\ln 2}}\exp(-(1+2\varepsilon)^2\ln 2^{n(1-\varepsilon)})\le 2^{-2\varepsilon n}.
\end{aligned}$$

This bound is summable over n, so that, by the first Borel-Cantelli lemma, almost surely, there exists an $n(\omega)<\infty$, such that for all $n\ge n(\omega)$,

$$\max_{j+j_i-j_2\le 2^{\varepsilon n}, j_i\le 2^n} h(j2^{-n})\left|B_{j_2 2^{-n}}-B_{j_1 2^{-n}}\right|\le 1+\varepsilon.$$

We may chose $n(\omega)$ in such a way that $2^{(n+1)\varepsilon-1}>2$ and $2^{-n(1-\varepsilon)}<1/e$, and

$$\sum_{m=n+1}^{\infty} h(2^-m)\le \varepsilon h(2^{-(1-\varepsilon)(n+1)}), \qquad (6.6.21)$$

for all $n>n(\omega)$.

Now let t_2-t_1 in $[0,1]$ be such that $\delta=t_2-t_1<2^{-n(\omega)(1-\varepsilon)}$ and chose $n\ge n(\omega)$ such that $2^{-(n+1)(1-\varepsilon)}\le\delta\le 2^{-n(1-\varepsilon)}$. Obviously, we can represent the numbers t_i in a binary representation as

$$\begin{aligned}
t_1 &= j_1 2^{-n}-2^{-n_1}-2^{-n_2}-\dots,\\
t_2 &= j_2 2^{-n}+2^{-m_1}+2^{-m_2}=\dots.
\end{aligned}$$

By our estimates, we have then the bound

$$\begin{aligned}
|B_{t_2}-B_{t_1}| &\le |B_{j_1 2^{-n}}-B_{t_1}|+|B_{j_2 2^{-n}}-B_{t_2}|+|B_{j_1 2^{-n}}-B_{j_2 2^{-n}}| \qquad (6.6.22)\\
&\le 2\sum_{m>n}(1+\varepsilon)h(2^{-m})+(1+\varepsilon)h(j2^{-n})\\
&\le 2(1+\varepsilon)\varepsilon h(2^{-(n+1)(1-\varepsilon)})+(1+\varepsilon)h(j2^{-n})\\
&\le (1+4\varepsilon)h(\delta).
\end{aligned}$$

This provides the upper bound and concludes the proof. $\square$

References

1. H. Bauer and R.B. Burckel. *Probability theory and elements of measure theory*. Academic Press London, 1981.
2. P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1995.
3. Anton Bovier. *Statistical mechanics of disordered systems*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2006.
4. Y.S. Chow and H. Teicher. *Probability theory*. Springer Texts in Statistics. Springer-Verlag, New York, 1997.
5. J.L. Doob. *Measure theory*, volume 143. Springer, 1994.
6. H.-O. Georgii. *Gibbs measures and phase transitions*, volume 9 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1988.
7. Kiyoshi Itô and Henry P. McKean, Jr. *Diffusion processes and their sample paths*. Die Grundlehren der Mathematischen Wissenschaften, Band 125. Academic Press Inc., Publishers, New York, 1965.
8. O. Kallenberg. *Random measures*. Akademie-Verlag, Berlin, 1983.
9. M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991.
10. M. Ledoux and M. Talagrand. *Probability in Banach Spaces: isoperimetry and processes*, volume 23. Springer, 1991.
11. M.M. Rao. *Measure theory and integration*, volume 265. CRC, 2004.
12. L.C.G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales*, volume 2. Cambridge Univ Pr, 2000.
13. B. Simon. *The statistical mechanics of lattice gases. Vol. I*. Princeton Series in Physics. Princeton University Press, Princeton, NJ, 1993.
14. D.W. Stroock and S.R.S. Varadhan. *Multidimensional diffusion processes*, volume 233. Springer, 1979.
15. M.E. Taylor. *Measure theory and integration*, volume 76. Amer Mathematical Society, 2006.

Index