Advanced Probability

Alan Sola
Department of Pure Mathematics and Mathematical Statistics
University of Cambridge
a.sola@statslab.cam.ac.uk

Michaelmas 2014

Contents

Introduction

These are notes for the second half of the Part III course *Advanced Probability* given at the University of Cambridge in Michaelmas 2014. The content of these notes is to be viewed as examinable, with the exception of parts that are explicitly stated to be non-examinable.

The results (and most of the proofs) presented here are by now classical and can be found in many standard textbooks on the subject; a short list of references is provided at the end. No claim of originality is made.

These notes are based to a large extent on earlier lecture notes by P. Sousi, G. Miermont, and J.R. Norris. Thanks are due to Adam Jones and Perla Sousi for useful comments and suggestions.

1 Conditional expectation

1.1 Basic objects: probability measures, σ-algebras, and random variables

We begin by recalling some fundamental concepts in probability, and setting down notation. Central to everything we do is the notion of a *probability space*: a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set, $\mathcal{F}$ is a σ-algebra, and $\mathbb{P}$ is a probability measure. In the probability context, the subsets of Ω are called *events*.

Definition 1.1. A collection $\mathcal{F}$ of subsets of Ω is said to be a *σ-algebra* on Ω if the following conditions hold:

- $\Omega \in \mathcal{F}$,
- If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$,
- If $\{A_j\}_{j=0}^{\infty}$ is a collection of sets in $\mathcal{F}$, then $\bigcup_{j=0}^{\infty} A_j \in \mathcal{F}$.

Informally speaking, a σ-algebra is a collection of sets that is closed under countable unions, and the operation of taking complements. The pair $(\Omega, \mathcal{F})$ is usually called a *measurable space*.

Definition 1.2. A set function $\mu\colon \mathcal{F} \to \mathbb{R}$ is said to be a *measure* if

- $\mu(A) \geq 0$ for all $A \in \mathcal{F}$,
- $\mu(\varnothing) = 0$,
- $\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j)$ for any countable collection of pairwise disjoint sets in $\mathcal{F}$.

More generally, if μ is a measure and $A \subset \bigcup_{j=1}^{\infty} A_j$, then $\mu(A) \leq \sum_{j=1}^{\infty} \mu(A_j)$; this property is known as *subadditivity*.

We say that $\mathbb{P}$ is a *probability measure* if, in addition to the above requirements, $\mathbb{P}$ satisfies $\mathbb{P}(\Omega) = 1$. The number $\mathbb{P}(A)$, $A \in \mathcal{F}$, is called the *probability* of the event A; we say that A occurs *almost surely*, abbreviated *a.s.*, if $\mathbb{P}(A) = 1$. Let $A, B \in \mathcal{F}$ be two events, and suppose $\mathbb{P}(B) > 0$. Then the *conditional probability of A given B* is defined by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

We often need to build σ-algebras starting with some smaller collection of sets.

Definition 1.3. Let $\mathcal{A}$ be a collection of subsets of Ω, not necessarily forming a σ-algebra. The *σ-algebra generated by $\mathcal{A}$* is defined as

$$\sigma(\mathcal{A}) = \bigcap \{\mathcal{E}\colon \mathcal{E} \text{ is a } \sigma - \text{algebra containing } \mathcal{A}\}.$$

By definition, $\sigma(\mathcal{A})$ is the smallest σ-algebra containing the collection $\mathcal{A}$; such a σ-algebra always exists since the collection of *all* subsets of Ω forms a σ-algebra.

When the base space Ω is endowed with a topology, and hence a notion of open sets, it is often natural to work with the σ-algebra generated by the open sets.

Definition 1.4. Let (Ω, τ) be a topological space. The *Borel σ-algebra* of Ω is the σ-algebra generated by the open sets $\mathcal{O}$ of Ω:

$$\mathcal{B}(\Omega) = \bigcap\{\mathcal{E} \colon \mathcal{E} \text{ is a } \sigma - \text{algebra containing } \mathcal{O}\}.$$

We shall usually write $\mathcal{B}(\mathbb{R})$ for the Borel σ-algebra on the real line, endowed with the usual Euclidean topology.

We are primarily interested in functions defined on probability spaces.

Definition 1.5. A *random variable* X on $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X \colon \Omega \to \mathbb{R}$ that is *measurable* with respect to $\mathcal{F}$; that is, for any open set $V \subset \mathbb{R}$, the pre-image $X^{-1}(V) \in \mathcal{F}$.

To be precise, we should add "real-valued" to the precedining definition, as the concept of random variable can be generalized to include functions taking values in $\mathbb{R}^d$, or any measurable space $(E, \mathcal{E})$. In the first chapters we shall mostly deal with the real-valued case.

The smallest σ-algebra on Ω that makes $X \colon \Omega \to \mathbb{R}$ a measurable map is called the *σ-algebra generated by* X, and is denoted by $\sigma(X)$. The property of being a random variable is preserved under a number of operations. For instance, if $(X_n)_{n=1}^\infty$ is a sequence of random variables, then $\limsup_n X_n$ and $\liminf_n X_n$ are random variables, and if $f \colon \mathbb{R} \to \mathbb{R}$ is a measurable map, and X is a random variable, then $f(X)$ is a random variable.

An important example of a random variable is the *indicator function* $\mathbf{1}(A)$ of an event $A \in \mathcal{F}$, defined via

$$\mathbf{1}(A)(\omega) = \mathbf{1}(\omega \in A) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A \end{cases}.$$

We say that X is a *simple random variable* if X is a finite linear combination of indicator functions, that is,

$$X = \sum_{j=1}^n c_j \mathbf{1}(A_j), \quad A_j \in \mathcal{F}, \; c_j \in \mathbb{R}.$$

A simple random variable is *positive* if all the coefficients satisfy $c_j \geq 0$.

The *expected value* or *expectation* of a postive simple random variable is easy to define: we simply set

$$\mathbb{E}\left[\sum_{j=1}^n c_j \mathbf{1}(A_j)\right] = \sum_{j=1}^n c_j \mathbb{P}(A_j).$$

If X is a non-negative random variable, then X can be obtained as a pointwise limit of positive simple random variables. A concrete way of approximating X is to set

$$X_n(\omega) = 2^{-n}\lfloor 2^n X(\omega)\rfloor \wedge n, \quad n = 1, 2, \ldots;$$

then $X = \lim_{n\to\infty} X_n$ pointwise. (We shall use similar constructions many times; it is probably worth spending a moment thinking about what the X_n look like.) Using such an approximation, we set

$$\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n],$$

and we verify that the limit is independent of the choice of (X_n). Finally, if X is a general random variable, we write

$$X = X^+ - X^-,$$

where $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$, and define

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-],$$

provided at least one of the numbers $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ are finite. We say that a random variable is *integrable* if $\mathbb{E}[|X|] < \infty$; if this is the case, we write $X \in L^1$.

More generally, for $p \geq 1$ and a measurable function $X\colon \Omega \to \mathbb{R}$ on $(\Omega, \mathcal{F}, \mathbb{P})$, we define the *$L^p$ norms*

$$\|X\|_p = (\mathbb{E}[|X|^p])^{1/p} = \left(\int_\Omega |X|^p d\mathbb{P}\right)^{1/p}.$$

We denote the set of measurable functions having finite L^p norm by $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$. When $p = \infty$, we set

$$\|X\|_\infty = \inf\{\lambda \geq 0\colon |X| \leq \lambda \text{ a.s.}\}.$$

A basic theorem in functional analysis states that L^p, $1 \leq p \leq \infty$, is a Banach space (over the reals), that is, a real vector space that is complete with respect to the norm $\|\cdot\|_p$.

The case $p = 2$ is of special importance.

Theorem 1.6. *The space $(L^2, \|\cdot\|_2)$ is a Hilbert space, with inner product $\langle X, Y\rangle = \mathbb{E}[XY]$.*

If $\mathcal{V}$ is a closed subspace, then for all $X \in L^2$, there exists a $Y \in \mathcal{V}$ such that

$$\|X - Y\|_2 = \inf_{Z\in\mathcal{V}} \|X - Z\|_2,$$

and $\langle Y, X - Y\rangle = 0$.

Up to a set of measure 0, this Y is unique.

The random variable Y in Theorem 1.6 is called the *orthogonal projection of X onto $\mathcal{V}$*.

Here are the three main convergence theorems of integration theory.

Theorem 1.7. [Monotone convergence theorem] *Let $(X_n)_{n=1}^\infty$ be a sequence of non-negative random variables, with $X_n \uparrow X$ almost surely as $n \to \infty$. Then*

$$\mathbb{E}[X_n] \uparrow \mathbb{E}[X] \quad \text{as} \quad n \to \infty.$$

Theorem 1.8. [Fatou's lemma] *Let $(X_n)_{n=1}^\infty$ be a sequence of non-negative random variables. Then*

$$\mathbb{E}[\liminf_n X_n] \leq \liminf_n \mathbb{E}[X_n].$$

To remember which way the inequality goes, consider the sequence $X_n = n\mathbf{1}((0,1/n))$ on the unit interval equipped with Lebesgue measure.

Theorem 1.9. [Dominated convergence theorem] *Let $(X_n)_{n=1}^\infty$ be a sequence of random variables. If $X_n \to X$ almost surely, and if, for all n, $|X_n| \leq Y$ almost surely for some integrable random variable Y, then*

$$\mathbb{E}[X_n] \to \mathbb{E}[X] \quad \text{as} \quad n \to \infty.$$

Here is another theorem that is often very useful.

Theorem 1.10. [Jensen's inequality] *Let X be an integrable random variable, and let $\varphi\colon \mathbb{R} \to \mathbb{R}$ be a convex function. Then*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

The notion of *independence* is arguably the key concept that distinguishes probability theory from general measure theory. In elementary probability, the notion of *independence of events* is defined by requiring

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \quad A, B \in \mathcal{F}.$$

This generalizes to σ-algebras as follows.

Definition 1.11. A collection $\{\mathcal{G}_j\}_{j=1}^\infty$ of σ-algebras contained in $\mathcal{F}$ is said to be *independent* if the following holds: for all $G_j \in \mathcal{G}_j$ and all distinct indices $j_1, \ldots, j_n$,

$$\mathbb{P}(G_{j_1} \cap \cdots \cap G_{j_n}) = \prod_{k=1}^n \mathbb{P}(G_{j_k}).$$

We should point out that it is not enough that $\mathbb{P}(G_i \cap G_j) = \mathbb{P}(G_i)\mathbb{P}(G_j)$ for all pairs $i \neq j$; this latter property is known as *pairwise independence*.

Two random variables X and Y are said to be independent if the σ-algebras $\sigma(X)$ and $\sigma(Y)$ are independent in the sense of the preceding definition, and a random variable is independent of a σ-algebra $\mathcal{F}$ if $\sigma(X)$ and $\mathcal{F}$ are independent.

Let $\{A_k\}_{k=1}^\infty$ be a sequence of events. We define

$$\limsup_k A_k = \bigcap_k \bigcup_{l \geq k} A_l \quad \text{and} \quad \liminf_k A_k = \bigcup_k \bigcap_{l \geq k} A_l.$$

It is customary in probability to write $\{A_k \text{ infinitely often}\}$ and $\{A_n \text{ eventually}\}$ for $\limsup A_n$ and $\liminf A_n$, respectively. (Think about what $\omega \in \limsup_n A_n$ entails.)

The *Borel-Cantelli lemmas* will prove very useful.

Lemma 1.12. [Borel-Cantelli lemmas] *Let $\{A_k\}_{k=1}^\infty$ be a sequence of events.*

1. *If $\sum_k \mathbb{P}(A_k) < \infty$, then $\mathbb{P}(A_k \text{ i.o.}) = 0$.*
2. *If the events are independent, and $\sum_k \mathbb{P}(A_k) = \infty$, then $\mathbb{P}(A_k \text{ i.o}) = 1$.*

Note the independence assumption in the second Borel-Cantelli lemma.

1.2 Conditional expectation

We begin by reviewing the definition of *conditional expectation of a random variable with respect to an event.* If X is an integrable random variable, and $A \in \mathcal{F}$ is an event with $\mathbb{P}(A) > 0$, we can define

$$\mathbb{E}[X|A] = \frac{\mathbb{E}[X\mathbf{1}(A)]}{\mathbb{P}(A)}.$$

This produces a number that can be viewed as the average of X over the event A. Our goal now is to generalize the definition of conditional expectation to σ-algebras. In that case, the conditional expectation will be a new random variable that is measurable with respect to the σ-algebra that we are conditioning on. (In the single-event case, this random variable is degenerate.)

Suppose first that $X \in L^1$, and that $\mathcal{G} = \sigma(B_j : j \in \mathbb{N})$ is generated by a countable family of disjoint events, that is, $\Omega = \bigcup_{j=1}^{\infty} B_j$. Let us agree to set $\mathbb{E}[X|B_j] = 0$ if $\mathbb{P}(B_j) = 0$. Guided by the formula defining conditional expectation for a single event, we introduce the random variable

$$Y = \sum_{j=1}^{\infty} \mathbb{E}[X|B_j]\mathbf{1}(B_j). \tag{1.1}$$

We recall that the $\mathbb{E}[X|B_j]$ are simply numbers, and so for $\omega \in \Omega$,

$$Y(\omega) = \sum_{j=1}^{\infty} \mathbb{E}[X|B_j]\mathbf{1}(\omega \in \mathcal{B}_j).$$

Explicitly, if $\omega \in B_1$, say, then $\mathbb{E}[X|\mathcal{G}](\omega) = \mathbb{E}[X|B_1] = \mathbb{E}[X\mathbf{1}(B_1)]/\mathbb{P}(B_1)$.

The random variable Y is measurable with respect to the σ-algebra $\mathcal{G}$ since each $\mathbf{1}(B_j)$ is, and moreoever,

$$\mathbb{E}[|Y|] \leq \sum_{j=1}^{\infty} \mathbb{E}[|X|\mathbf{1}(B_j)] = \mathbb{E}[|X|]$$

since $\{B_j\}_{j=1}^{\infty}$ are disjoint. We conclude that (1.1) defines an integrable, $\mathcal{G}$-measurable random variable.

Now let $G \in \mathcal{G}$, and note that

$$\mathbb{E}[X\mathbf{1}(G)] = \mathbb{E}[Y\mathbf{1}(G)]. \tag{1.2}$$

In the general case, we want $\mathcal{G}$-measurability and (1.2) to be the defining properties of conditional expectation. Our next objective is to prove the following basic result.

Theorem 1.13. *Let X be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. Then there exists a random variable Y with the following properties:*

- *Y is $\mathcal{G}$-measurable,*
- *Y is integrable, and*

$$\mathbb{E}[X\mathbf{1}(A)] = \mathbb{E}[Y\mathbf{1}(A)] \quad \textit{for all} \quad A \in \mathcal{G}.$$

If Z is another random variable satisfying these two requirements, then $Y = Z$ almost surely.

Any random variable satisfying the conditions of Theorem 1.13 is said to be (a version of) the *conditional expectation of X given $\mathcal{G}$*; we then write $Y = \mathbb{E}[X|\mathcal{G}]$ a.s. We can also take the conditional expectation of a random variable with respect to another random variable; this is simply

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)].$$

By approximation, the equality for indicator functions in the second part of Theorem 1.13 can be replaced by the requirement that for all bounded $\mathcal{G}$-measurable random variables Z,

$$\mathbb{E}[XZ] = \mathbb{E}[YZ].$$

The proof we give features some techniques, such as truncation, that we will employ in many places.

Proof of Theorem 1.13. We first demonstrate existence of conditional expectation. We begin by imposing the more restrictive assumption $X \in L^2$; this enables us to appeal to Hilbert space techniques such as orthogonal projection. We verify that for any sub-σ-algebra $\mathcal{G}$, $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. This subspace, together with its orthogonal complement, spans all of $L^2(\Omega, \mathcal{F}, \mathbb{P})$, that is,

$$L^2(\mathcal{F}) = L^2(\mathcal{G}) \oplus L^2(\mathcal{G})^\perp$$

meaning that any $X \in L^2(\mathcal{F})$ can be written in a unique way as

$$X = Y + Z, \quad Y \in L^2(\mathcal{G}), Z \in L^2(\mathcal{G})^\perp.$$

We now set $\mathbb{E}[X|\mathcal{G}] = Y$; this immediately makes $\mathbb{E}[X|\mathcal{G}]$ measurable with respect to $\mathcal{G}$. Now let $A \in \mathcal{G}$; then $\mathbf{1}(A) \in L^2(\mathcal{G})$, and as $Z \in L^2(\mathcal{G})^\perp$, we obtain

$$\mathbb{E}[X\mathbf{1}(A)] = \mathbb{E}[Y\mathbf{1}(A)] + \mathbb{E}[Z\mathbf{1}(A)] = \mathbb{E}[Y\mathbf{1}(A)],$$

the desired second property. As $A \in \mathcal{G}$ was arbitrary, this shows existence of conditional expectation for random variables $X \in L^2$.

We pause to record one important property of conditional expectation: if $X \geq 0$ then $\mathbb{E}[X|\mathcal{G}] \geq 0$ almost surely. To see this, consider the event $\{\mathbb{E}[X|\mathcal{G}] < 0\} \in \mathcal{G}$, use that

$$\mathbb{E}[X\mathbf{1}(\mathbb{E}[X|\mathcal{G}] < 0)] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}(\mathbb{E}[X|\mathcal{G}] < 0)].$$

The left-hand side is non-negative by assumption, the right-hand side is non-positive by construction, and this forces $\mathbb{P}(\mathbb{E}[X|\mathcal{G}] < 0) = 0$.

We now use an approximation scheme to extend the construction to general integrable random variables. Now suppose X is non-negative, and introduce the trunctions $X_n = X \wedge n$, $n = 1, 2, \ldots$. Each X_n is bounded and non-negative, and in particular $X_n \in L^2$ for each n. Thus there exists a sequence of $\mathcal{G}$-measurable random variables Y_n satisfying

$$\mathbb{E}[X_n\mathbf{1}(A)] = \mathbb{E}[Y_n\mathbf{1}(A)]$$

for any $A \in \mathcal{G}$. We now argue that (Y_n) forms an increasing sequence; this follows from the fact that

$$X_{n+1} - X_n \geq 0 \quad \text{implies} \quad \mathbb{E}[X_{n+1} - X_n | \mathcal{G}] \geq 0,$$

plus linearity. Hence $\uparrow Y_n = Y$ exists, and is $\mathcal{G}$-measurable as a pointwise limit of $\mathcal{G}$-measurable functions. We now invoke the monotone convergence theorem, and obtain

$$\mathbb{E}[X\mathbf{1}(A)] = \lim_{n\to\infty} \mathbb{E}[X_n\mathbf{1}(A)] = \lim_{n\to\infty} \mathbb{E}[Y_n\mathbf{1}(A)] = \mathbb{E}[Y\mathbf{1}(A)]$$

for any $A \in \mathcal{G}$. In particular, the choice $A = \Omega$ shows that Y is integrable whenever X is. Hence we may set $\mathbb{E}[X|\mathcal{G}] = Y$, and all the requirements are met.

Finally, we decompose an arbitrary $X \in L^1$ as $X = X^+ - X^-$, apply the previous argument to the positive and negative parts, and set $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}]$. It is then straightforward to check the the random variable so constructed satisfies the assertions of the theorem.

We now establish uniqueness (up to sets of measure 0) of conditional expectation. Let us assume that Z is another random variable satisfying both conclusions of the Theorem. We consider the event $A = \{Y > Z\}$; this is event is in $\mathcal{G}$ as both random variables were assumed $\mathcal{G}$-measurable. Appealing to linearity, we find that

$$\mathbb{E}[(Y - Z)\mathbf{1}(A)] = \mathbb{E}[Y\mathbf{1}(A)] - \mathbb{E}[Z\mathbf{1}(A)] = \mathbb{E}[X\mathbf{1}(A)] - \mathbb{E}[X\mathbf{1}(A)] = 0.$$

This means that $\mathbb{P}(A^c) = 1$ and $Z \geq Y$ almost surely. By reversing the roles of Y and Z we conclude that also $Y \geq Z$ almost surely, and uniqueness follows. □

Here are some immediate consequences of Theorem 1.13 and its proof.

Theorem 1.14. *Let $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. Then the following hold:*

- *If $X \geq 0$, then $\mathbb{E}[X|\mathcal{G}] \geq 0$ almost surely,*
- *Conditional expectation is linear, that is, for any $a, b \in \mathbb{R}$,*

  $$\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$$

  almost surely,
- $|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}]$ *almost surely,*
- $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$*,*
- *If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$ almost surely, and*
- *If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ almost surely.*

The basic convergence theorems for expectation we stated earlier have counterparts for conditional expectation.

Theorem 1.15. *Let $(X_n)_{n=1}^{\infty}$ be a sequence of integrable random variables, let X be a random variable, and let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. Then the following hold:*

- **(Conditional monotone convergence theorem)**
  If $X_n \geq 0$ and $X_n \uparrow X$ almost surely, then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$,
- **(Conditional Fatou's lemma)**
  If $X_n \geq 0$, then $\mathbb{E}[\liminf_n X_n|\mathcal{G}] \leq \liminf_n \mathbb{E}[X_n|\mathcal{G}]$,
- **(Conditional dominated convergence theorem)**
  If $|X_n| \leq Y$ for some random variable Y having $\mathbb{E}[Y] < \infty$, and $X_n \to X$ almost surely, then $\lim_n \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]$.

We leave the proofs to the reader.

Theorem 1.16. [Conditional version of Jensen's inequality] *If $\varphi \colon \mathbb{R} \to \mathbb{R}$ is convex, and $\mathbb{E}[|\varphi(X)|] < \infty$, or else if φ is non-negative, then*

$$\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}] \quad \textit{almost surely.}$$

In particular, the operation of taking conditional expectation is a contraction on L^p; for all $1 \leq p < \infty$,

$$\|\mathbb{E}[X|\mathcal{G}]\|_p \leq \|X\|_p.$$

Proof. It is shown in [7, §6.6] that any convex function $\varphi \colon \mathbb{R} \to \mathbb{R}$ can be obtained as a supremum over affine functions. This means that there exist numbers $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ such that

$$\varphi(x) = \sup_{k \geq 1} (a_k x + b_k).$$

In particular, $\varphi(x) \geq a_k x + b_k$ for all k. By Theorem 1.14, we have, for each k, $\mathbb{E}[\varphi(X)|\mathcal{G}] \geq a_k \mathbb{E}[X|\mathcal{G}] + b_k$ on an event with probability 1; the desired inequality follows after taking the countable union of these events.

To see that conditioning is a contraction, we apply the previous result with $\varphi(x) = |x|^p$ to obtain

$$\mathbb{E}[|X|^p|\mathcal{G}] \geq |\mathbb{E}[X|\mathcal{G}]|^p,$$

and then take expectations on both sides. The desired conclusion now follows from the fact that

$$\mathbb{E}[\mathbb{E}[|X|^p|\mathcal{G}]] = \mathbb{E}[|X|^p]$$

by Theorem 1.14. □

Conditional expectation enjoys the so-called *tower property*; that is, the "smaller σ-algebra wins."

Proposition 1.17. *Let $\mathcal{H} \subset \mathcal{G}$ and $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Then*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}] \ a.s.$$

Proof. By definition, $\mathbb{E}[X|\mathcal{H}]$ is $\mathcal{H}$-measurable. If $A \in \mathcal{H}$, then $A \in \mathcal{G}$ by assumption, and hence

$$\mathbb{E}[\mathbb{E}[X|\mathcal{H}]\mathbf{1}(A)] = \mathbb{E}[X\mathbf{1}(A)] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}(A)].$$

This is valid for any event $A \in \mathcal{H}$, and the conclusion follows. □

The content of the next Proposition is often summarized by saying that we can *take out what is known.*

Proposition 1.18. *Let $X \in L^1$, and let $\mathcal{G}$ be a σ-algebra. If Y is bounded and $\mathcal{G}$-measurable, then*

$$\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}] \ a.s.$$

Proof. Suppose first that Y is the indicator function of some set in $\mathcal{G}$, that is $Y = \mathbf{1}(B)$, $B \in \mathcal{G}$. Then, for any $A \in \mathcal{G}$, we have

$$\mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]\mathbf{1}(A)] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}(A \cap B)] = \mathbb{E}[X\mathbf{1}(A \cap B)] = \mathbb{E}[YX\mathbf{1}(A)],$$

since $\mathbf{1}(A \cap B) = \mathbf{1}(A) \cdot \mathbf{1}(B)$. This implies that $Y\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[YX|\mathcal{G}]$.

By linearity of conditional expectation, the above argument extends to the case where Y is a simple $\mathcal{G}$-measurable random variable. By the monotone convergence theorem, the conclusion of the proposition obtains for random variables $X \geq 0$ and non-negative $\mathcal{G}$-measurable Y.

Finally, we arrive at the full statement by arguing on the positive and negative parts of X and Y. □

The hypotheses of the proposition can be relaxed: the conclusion holds if X and Y are both almost surely non-negative and Y is $\mathcal{G}$-measurable, or else if $X \in L^p$ and $Y \in L^q$, where $1/p + 1/q = 1$, and Y is $\mathcal{G}$ measurable.

This last proposition can be used (provided the appropriate hypotheses are satisfied) in conjunction with Proposition to write

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[XY|\mathcal{H}]] = \mathbb{E}[Y\mathbb{E}[X|\mathcal{H}]];$$

this representation can be useful in cases where $\mathbb{E}[X|\mathcal{H}]$ is easy to compute or estimate.

1.3 Integration with respect to product measures and Fubini's theorem

When dealing with product spaces and functions of several variables, we will often want to form products of probability measures. A natural way to define a product measure would be to prescribe its values on sets that are themselves products: we would like the measure of such a set to be equal to the product of the measures of the factors.

In order to implement this strategy, we first need to discuss the auxiliary notions of π-systems and state a uniqueness of extension theorem.

Definition 1.19. A non-empty collection $\mathcal{P}$ of subsets of Ω is said to be a *π-system* if $\varnothing \in \mathcal{P}$ and if $A \cap B \in \mathcal{P}$ for all $A, B \in \mathcal{P}$.

It is usually easier to work with π-systems than with full σ-algebras, and this makes the following theorem very useful.

Theorem 1.20. [**Uniqueness of extensions**] *Let μ_1 and μ_2 be two measures on a measurable space $(E, \mathcal{E})$, having $\mu_1(E) = \mu_2(E) < \infty$. Suppose $\mathcal{P}$ is a π-system that generates $\mathcal{E}$, and suppose $\mu_1 = \mu_2$ on $\mathcal{P}$.*

Then $\mu_1 = \mu_2$ on $\mathcal{E}$.

Some naturally occurring measure spaces do not have finite measure. However, they can sometimes be divided into countably many pieces, each having finite measure.

Definition 1.21. A measure space $(E, \mathcal{E}, \mu)$ is said to be $\sigma-finite$ if there exists a collection $\{E_n\}_{n=0}^{\infty}$ of measureable sets such that $E = \bigcup_{n=0}^{\infty} E_n$, and $\mu(E_n) < \infty$ for each n.

Example: Lebesgue measure on the real line. (What are some possible choices of E_n's?)

Let us consider two σ-finite spaces $(E_k, \mathcal{E}_k, \mu_k)$, $k = 1, 2$, and form the Cartesian product $E = E_1 \times E_2$ of the underlying spaces. We next consider the π-system given by

$$\mathcal{P} = \{A_1 \times A_2 \in E \colon A_k \in \mathcal{E}_k,\ k = 1, 2\},$$

and the σ-algebra it generates,

$$\mathcal{E}_1 \otimes \mathcal{E}_2 = \sigma(\mathcal{P}).$$

Theorem 1.22. [**Product measures**] *Suppose $(E_k, \mathcal{E}_k, \mu_k)$, $k = 1, 2$, are σ-finite measure spaces. There exists a unique measure $\mu = \mu_1 \otimes \mu_2$ defined on $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2$ such that*

$$\mu(A_1 \times A_2) = \mu(A_1)\mu(A_2),$$

for all $A_k \in \mathcal{E}_k$, $k = 1, 2$.

Fubini's theorem is very useful for computing integrals of functions of several variables. Roughly speaking, it says that the order of integration does not matter as long as there are "no infinities involved."

Theorem 1.23. [**Fubini's theorem**] *Let $(E, \mathcal{E}, \mu)$ be given as the product of two σ-finite measure spaces, as defined above, and let f be a $\mathcal{E}$-measurable function.*

If f is non-negative, or if $\int_E |f| d\mu < \infty$, then

$$\int_E f(x_1, x_2) d\mu(x_1, x_2) = \int_{E_1} \left(\int_{E_2} f(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1) \tag{1.3}$$

$$= \int_{E_1} \left(\int_{E_2} f(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2). \tag{1.4}$$

We caution that the conclusion of the theorem is false for measures that are not σ-finite. (Example: consider a product spaces where the first factor is the unit interval with Lebesgue measure, and the second factor is the unit interval endowed with the discrete σ-algebra and the counting measure. Then compute the relevant integrals of $f = \mathbf{1}(x = y)$.)

We end this section by working out an example that connects what we have done for general σ-algebras with more elementary approaches to conditional expectation.

Example 1.24. Suppose X and Y are random variables having a joint density function $f_{X,Y}(x, y)$ on $\mathbb{R}^2$. Then

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)dx$$

is a density function for the random variable Y.

Now let $h: \mathbb{R} \to \mathbb{R}$ be a Borel function such that

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f_X(x)dx = \int_{\mathbb{R}} h(x) \left(\int_{\mathbb{R}} f_{X,Y}(x, y)dy \right) dx < \infty.$$

Our goal is to find a fairly concrete way of representing $\mathbb{E}[h(X)|Y] = \mathbb{E}[h(X)|\sigma(Y)]$. To this end, we suppose g is a bounded Borel function on $\mathbb{R}$, and note that, by Fubini's theorem,

$$\mathbb{E}[h(X)g(Y)] = \int_{\mathbb{R}^2} h(x)g(y)f_{X,Y}(x, y)dxdy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x)f_{X,Y}(x, y)dx \right) g(y)dy. \tag{1.5}$$

Our aim is to rewrite this in such a way that we can extract the expression $g(y)f_Y(y)dy$ on the right-hand side; the obvious way to accomplish this is to factor out f_Y from the inner integral.

For all $y \in \mathbb{R}$ such that $f_Y(y) > 0$, we set

$$d(y) = \int_{\mathbb{R}} h(x)\frac{f_{X,Y}(x, y)}{f_Y(y)}dx,$$

at points where f_Y vanishes we take $d(y) = 0$. We can now rewrite (1.5) as

$$\mathbb{E}[h(X)g(Y)] = \mathbb{E}[d(Y)g(Y)];$$

this holds for all bounded measurable functions g, and hence $d(Y)$ is a version of $\mathbb{E}[h(X)|Y]$. In conclusion, introducing the *conditional density function*

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)}, & f_Y(y) \neq 0 \\ 0, & f_Y(y) = 0 \end{cases},$$

we have shown that

$$\mathbb{E}[h(X)|Y] = \int_{\mathbb{R}} h(x)f_{X|Y}(x|Y)dx.$$

2 Martingales in discrete time

2.1 Basic definitions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A sequence $X = (X_n : n \geq 0)$ of random variables (taking values in $\mathbb{R}$) is called a *stochastic process*. We say that the process X is *integrable* if the random variables X_n are integrable for all n.

Stochastic processes typically arise in applications where random phenomena are observed over a period of time. The first step in constructing a mathematical framework for the study of such dynamical phenomena is formalize the fact that we can potentially extract useful information about the evolution of such a system by continually recording our observations.

Definition 2.1. A *filtration* $(\mathcal{F}_n)_n$ is an increasing family of sub-σ-algebras of $\mathcal{F}$, satisfying $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, for all n. Every process induces a natural filtration $(\mathcal{F}_n^X)_n$, given by

$$\mathcal{F}_n^X = \sigma(X_k : 0 \leq k \leq n).$$

A process X is said to be *adapted* to the filtration $(\mathcal{F}_n)_n$, if X_n is $\mathcal{F}_n$-measurable for all n.

We say that $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ is a *filtered probability space*. It is clear that every process is adapted to its natural filtration. We think of the σ-algebra $\mathcal{F}_n$ as keeping track of the information we have collected up to time-step n. Informally speaking, one way in which processes in discrete time are easier to handle is that this information is only updated at certain separated times; in the continuous-time setting we will need to worry about possible lags between "instantaneous" changes, and corresponding updates. We shall address this problem later through the concept of right-continuous filtrations.

Definition 2.2. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a probability space, endowed with the filtration $(\mathcal{F}_n)_n$ and let $X = (X_n : n \geq 0)$ be an adapted integrable process. We say that

- X is a *martingale* if $\mathbb{E}[X_n|\mathcal{F}_m] = X_m$ a.s., for all $n \geq m$.
- X is a *supermartingale* if $\mathbb{E}[X_n|\mathcal{F}_m] \leq X_m$ a.s., for all $n \geq m$.
- X is a *submartingale* if $\mathbb{E}[X_n|\mathcal{F}_m] \geq X_m$ a.s., for all $n \geq m$.

If a process is a martingale (or super/submartingale) with respect to a given filtration, then it is also a martingale (or super/submartingale) with respect to its own natural filtration. This follows from the minimality of the natural filtration, and the tower property of conditional expectation.

At first, it is easy to forget in which directions the inequalities go in this definition; in fact, as we shall see, supermartingales correspond to *unfavorably* biased games of chance [1].

We ask the reader to verify that the following processes are examples of martingales.

[1] As Durrett puts it, there is nothing "super" about a supermartingale.

Example 2.3. [**Random walk**] Let $(X_k)_{k\geq 1}$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_1] = 0$. Then $S_n = \sum_{k=1}^n X_i$ is a martingale with respect to its natural filtration.

Example 2.4. [**Closed martingale**] Let $(\mathcal{F}_n)_n$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ and let X be an integrable random variable. Then the process

$$Y = (\mathbb{E}[X|\mathcal{F}_n])_{n\geq 0}$$

is a martingale with respect to the filtration $(\mathcal{F}_n)_n$.

The first example contains the origin of martingale theory. Namely, suppose X_1 is a random variable with $\mathbb{P}(X_1 = 1) = 1/2$ and $\mathbb{P}(X_1 = -1) = 1/2$, and $X_2, X_3, \ldots$ are independent copies of X_1. We could use this setup to model a game involving coin tosses, where $X_k = 1$ if the coin comes up heads on the kth toss, and $X_k = -1$ if it is tails. Then S_n represents our net gain or losses after n rounds, if we make a 1 currency unit bet every time. It is clearly in the interest of a gambler to be able to make statements concerning S_n (and perhaps even devise a system for how to bet with a high success rate[2]).

We shall see later that closed martingales are, in a certain sense, canonical examples.

2.2 Stopping times and the optional stopping theorem

Definition 2.5. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ be a filtered probability space. A random variable $T\colon \Omega \to \mathbb{N}$ is said to be a *stopping time* if $\{T \leq n\} \in \mathcal{F}_n$ for all n.

In the definition of stopping time, we could equally well require that $\{T = n\} \in \mathcal{F}_n$ for all n. Equivalence follows from the fact that

$$\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}$$

and conversely, that

$$\{T \leq n\} = \bigcup_{k=0}^{n} \{T = k\}.$$

Example 2.6. Any constant time is a stopping time.

Example 2.7. If S and T are both stopping times, then $S \wedge T = \min\{S, T\}$ is a stopping time.

Example 2.8. [**First hitting times**] Here is a class of concrete example that we shall encounter frequently later on.

Suppose $A \in \mathcal{B}(\mathbb{R})$, and set

$$T_A = \inf\{n \geq 0\colon X_n \in A\};$$

let us agree that $\inf(\varnothing) = \infty$, so that $T_A = \infty$ if X never enters A. Now

$$\{T_A \leq n\} = \bigcup_{k=0}^{n} \{X_k \in A\} \in \mathcal{F}_n$$

since each $\{X_k \in A\}$ is in $\mathcal{F}_k$, and $\mathcal{F}_k \subset \mathcal{F}_n$.

[2] Disclaimer: The lecturer does not, in any way, endorse or recommend real-life gambling.

Here is an example of a random variable that is not a stopping time in general. Consider the last exit time

$$E_A = \sup\{n \geq 0 \colon X_n \in A\}$$

for some $A \in \mathcal{B}(\mathbb{R})$; then $\mathcal{F}_n$ does not tell us whether X comes back to A in the future, and hence E_A is not a stopping time.

Definition 2.9. **[Stopped σ-algebras and stopped processes]** Let T be a stopping time on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$.

The *stopped σ-algebra* is the σ-algebra defined by

$$\mathcal{F}_T = \{A \in \mathcal{F} \colon A \cap \{T \leq n\} \in \mathcal{F}_n \quad \forall\, n\}.$$

If X is an adapted process on the same space, we define the random variable $X_T(\omega)$ by setting, for $T(\omega) < \infty$,

$$X_T(\omega) = X_{T(\omega)}(\omega),$$

and define the *stopped process* $X^T = (X_n^T \colon n \geq 0)$ by letting $X_n^T = X_{T \wedge n}$.

Stopping times are a very powerful tool in the analysis of martingales: if we start out with a martingale and stop it, we still end up with a martingale. As we have seen, martingales have expected values that are constant in time, that is, for any fixed n, we have $\mathbb{E}[X_n] = \mathbb{E}[X_0]$. We will see that stopping times, even though they are random, preserve this property to a large extent.

In order to make these statements precise, we need a couple of lemmas.

Lemma 2.10. *Suppose $S \leq T$. Then $\mathcal{F}_S \subset \mathcal{F}_T$.*

Proof. By assumption, $\{T \leq n\} \subset \{S \leq n\}$ and so $\{T \leq n\} = \{S \leq n\} \cap \{T \leq n\}$. Now let $A \in \mathcal{F}_S$, and write

$$A \cap \{T \leq n\} = A \cap \{S \leq n\} \cap \{T \leq n\}.$$

The latter set is in $\mathcal{F}_n$ for all n, and the lemma follows. □

Lemma 2.11. *Let X be an adapted process, and let T be a stopping time. Then the random variable $X_T \mathbf{1}(T < \infty)$ is $\mathcal{F}_T$-measurable.*

Proof. Since X is an adapted process, we have $\{X_k \in A\} \in \mathcal{F}_n$ for any Borel set A, and any $k \leq n$. Our task is to show that the requirement $\{X_T \mathbf{1}(T < \infty) \in A\} \cap \{T \leq n\} \in \mathcal{F}_n$ holds for all n. To this end, we write

$$\{X_T \mathbf{1}(T < \infty) \in A\} \cap \{T \leq n\} = \bigcup_{k=1}^{n} (\{X_k \in A\} \cap \{T = k\}).$$

To see that each set in the union is in $\mathcal{F}_n$, we now recall that $\{T = k\} \in \mathcal{F}_n$ for $k \leq n$ whenever T is a stopping time; cf. the remark after defintion of stopping time. □

Proposition 2.12. *Let X be an adapted process on a filtered probability space, and let T be a stopping time on the same space. The the process X^T is adapted, and if X is an integrable process then so is X^T.*

Proof. We have already observed that the minimum of two stopping times is a stopping time. This means that, for any $n \in \mathbb{N}$, the random variable $T \wedge n$ is a stopping time. By the preceding lemma, $X_{T\wedge n}$ is $\mathcal{F}_{T\wedge n}$-measurable, hence $\mathcal{F}_n$ measurable since $\mathcal{F}_n \supset \mathcal{F}_{T\wedge n}$. This shows that X^T is adapted.

To establish integrability, we note that

$$\mathbb{E}[|X_{T\wedge n}|] \leq \max_{1\leq k\leq n} \mathbb{E}[|X_k|] \leq \sum_{k=1}^{n} \mathbb{E}[|X_k|].$$

The sum in the right-hand side contains finitely many terms, and hence is finite since the process X was assumed to be integrable. □

The next theorem is central in martingale theory and its applications.

Theorem 2.13. [Optional stopping theorem] *Let $X = (X_n : n \geq 0)$ be a martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$, and T a $(\mathcal{F}_n)$-stopping time.*

1. *The process $X^T = (X_{T\wedge n} : n \geq 0)$ is a martingale.*
2. *If $S \leq T$ are bounded stopping times, then $\mathbb{E}[X_T|\mathcal{F}_S] = X_S$ almost surely. In particular, $\mathbb{E}[X_T] = \mathbb{E}[X_S]$.*
3. *If*
   - *$|X_n| \leq Y$ for some integrable random variable, and T is almost surely finite,*
   - *or X has bounded incremenents, that is $|X_n - X_{n-1}| \leq M$ almost surely for some $M < \infty$, and T satisfies $\mathbb{E}[T] < \infty$,*

   then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Theorem 2.13 holds true also for sub/supermartingales, provided equalities are replaced by appropriate inequalities.

Proof. We already know that X^T is adapted and integrable. It remains to check that X^T has the martingale property, that is to say, that $\mathbb{E}[X_{T\wedge n}|\mathcal{F}_{n-1}] = X_{T\wedge n-1}$ a.s.

We first decompose $X_{T\wedge n}$ as a sum,

$$X_{T\wedge n} = \sum_{k=1}^{n-1} X_k \mathbf{1}(T = k) + X_n \mathbf{1}(T > n - 1);$$

and by linearity of conditional expectation we have

$$\mathbb{E}[X_{T\wedge n}|\mathcal{F}_{n-1}] = \sum_{k=1}^{n-1} \mathbb{E}[X_k \mathbf{1}(T = k)|\mathcal{F}_{n-1}] + \mathbb{E}[X_n \mathbf{1}(T > n - 1)|\mathcal{F}_{n-1}].$$

The first sum is equal to $\sum_{k=1}^{n-1} X_k \mathbf{1}(T = k) = X_T \mathbf{1}(T \leq n-1)$ since the random variables $X_k \mathbf{1}(T = k)$ are all $\mathcal{F}_{n-1}$-measurable. Moreover, since $\{T > n-1\} = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}$, and X is a martingale,

$$\mathbb{E}[X_n \mathbf{1}(T > n-1) | \mathcal{F}_n] = \mathbf{1}(T > n-1)\mathbb{E}[X_n | \mathcal{F}_{n-1}] = \mathbf{1}(T > n-1) X_{n-1}.$$

Hence

$$\mathbb{E}[X_{T \wedge n} | \mathcal{F}_{n-1}] = X_T \mathbf{1}(T \leq n-1) + X_{n-1} \mathbf{1}(T > n-1) = X_{T \wedge n-1}$$

almost surely, and thus X^T is a martingale.

We turn to the second item in the theorem. Since T is assumed bounded, there exists an integer n such that $T \leq n$. Since $S \leq T$, we can write X_T using a telescoping sum:

$$X_T = \sum_{k=0}^{n-1} (X_{k+1} - X_k) \mathbf{1}(S \leq k < T) + X_S;$$

to see this, we first write $X_T = X_T + X_S - X_S = X_T - X_{S+1} + (X_{S+1} - X_S) + X_S$. We then proceed in the same way until we reach $X_{T-1} - X_{T-1}$.

Now let $A \in \mathcal{F}_S$. By the above decomposition,

$$\mathbb{E}[X_T \mathbf{1}(A)] = \mathbb{E}[X_S \mathbf{1}(A)] + \sum_{k=1}^{n-1} \mathbb{E}[(X_{k+1} - X_k) \mathbf{1}(S \leq k < T) \mathbf{1}(A)].$$

In order to conclude that $\mathbb{E}[X_T | \mathcal{F}_S] = X_S$ a.s., we need to show that the sum on the far right is zero. Now since A is an event in the stopped σ-algebra S, we have $\{S \leq k < T\} \cap A \in \mathcal{F}_k$ for $k \leq n$, and hence

$$\mathbb{E}[(X_{k+1} - X_k) \mathbf{1}(S \leq k < T) \mathbf{1}(A)] = 0,$$

by the martingale property. The almost sure equality $\mathbb{E}[X_T | \mathcal{F}_S] = X_S$ follows, and to get $\mathbb{E}[X_T] = \mathbb{E}[X_S]$ we take expectations on both sides.

The last item in the theorem is left to Example Sheet 1. □

In particular, if T is a bounded stopping time, then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$. It is worth emphasizing the following: while $\mathbb{E}[X_{T \wedge n}] = \mathbb{E}[X_0]$ for all n without any assumption whatsoever on the stopping time T, it is in general *not true* that the expected value *at* the stopping time is equal to $\mathbb{E}[X_0]$ unless we impose a boundedness condition on T. We shall give an example later.

As usual, non-negativity leads to stronger results.

Corollary 2.14. *Suppose $X = (X_n : n \geq 0)$ is a non-negative supermartingale, and T is an almost surely finite stopping time. Then $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$.*

2.3 The martingale convergence theorem

We are ready to state and prove our first general theorem concerning martingales. The Martingale Convergence Theorem, due to J.L. Doob, is arguably the central result in martingale theory.

Theorem 2.15. [Martingale convergence theorem] *Let $X = (X_n : n \geq 0)$ be a supermartingale that forms a bounded sequence in L^1, that is,*

$$\sup_{n\geq 0} \mathbb{E}[|X_n|] < \infty.$$

Then $X_\infty = \lim_{n\to\infty} X_n$ exists almost surely, and $\mathbb{E}[|X_\infty|] < \infty$.

A few remarks are in order. It is important to note that the theorem only asserts the existence of a pointwise limit; convergence in L^1 does not hold in general without additional hypotheses. The proof is not constructive, and does not provide us with a means to obtain X_∞ directly.

The basic ideas of the proof is to count the number of *upcrossings* completed by the process X. The intuition is that if the process does not traverse any interval infinitely often, then it has to converge since the L^1-boundedness condition prevents it from escaping to infinity.

Definition 2.16. Let $(x_k)_{k=0}^{\infty}$ be a real sequence, and $a, b \in \mathbb{R}$ with $a < b$ be given. Set $N_0 = -1$, and define, for $k = 1, 2, \ldots$,

$$N_{2k-1} = \inf\{k > N_{2k-2} : x_k \leq a\} \quad N_{2k} = \inf\{k > N_{2k-1} : x_k \geq b\}.$$

The *number of upcrossings of* $[a, b]$ *completed by the sequence by* $n \geq 1$ is defined by

$$U_n[a, b] = U_n(\{x_k\}, [a, b]) = \sup\{k \geq 0 : N_{2k} \leq n\}. \tag{2.1}$$

We shall mostly be interested in the case where the sequence is given by a martingale $X = (X_n : n \geq 0)$; in that case, we usually write $U_n[a, b]$ or $U_n(X, [a, b])$ to denote the number of upcrossings completed by the process by time n. It is worth thinking about upcrossings from the point of view of gambling strategies. We would then try to start betting when a random walk is about to enter an upcrossing of $[a, b]$, and abstain otherwise; at the end of the upcrossing we would then have made a profit of at least $b - a$ units. It turns out (see Example sheet) that this strategy produces another martingale.

We need two lemmas, one from real analysis (which we state without proof), and one that addresses upcrossings associated with supermartingales, and is of some independent interest.

Lemma 2.17. *Let $(x_k)_{k=1}^{\infty}$ be a real sequence. Then $\lim_{k\to\infty} x_k$ exists in $\mathbb{R} \cup \{\pm\infty\}$ if and only if $U_n(\{x_k\}, [a, b]) < \infty$ for all $a, b \in \mathbb{Q}$ with $a < b$.*

Lemma 2.18. [Doob's upcrossings lemma] *Let $X = (X_n : n \geq 0)$ be a supermartingale, and let $a, b \in \mathbb{R}$, with $a < b$. Then*

$$(b - a)\mathbb{E}[U_n[a, b]] \leq \mathbb{E}[(X_n - a)^-]$$

for all $n \geq 0$.

Proof. We begin by recording the basic observation that

$$X_{2k} - X_{N_{2k-1}} \geq b - a, \quad k \geq 1; \tag{2.2}$$

this follows directly from the definition of the times N_{2k-1} and N_{2k}. Next, we check that the N_{2k-1} and N_{2k} are stopping times for the natural filtration of X.

Now consider the sum

$$\sum_{k=1}^{n} (X_{N_{2k}\wedge n} - X_{N_{2k-1}\wedge n}).$$

By definition, we have $N_{2k} \wedge n \geq N_{2k-1} \wedge n$, and both $N_{2k} \wedge n$ and $N_{2k-1} \wedge n$ are bounded stopping times. We now apply the Optional Stopping Theorem 2.13 to the supermartingale X, and obtain

$$\mathbb{E}[X_{N_{2k}\wedge n} - X_{N_{2k-1}\wedge n}] = \mathbb{E}[X_{N_{2k}\wedge n}] - \mathbb{E}[X_{N_{2k-1}\wedge m}] \leq 0,$$

and thus

$$\mathbb{E}\left[\sum_{k=1}^{n} (X_{N_{2k}\wedge n} - X_{N_{2k-1}\wedge n})\right] \leq 0.$$

On the other hand, we have

$$\begin{aligned}\sum_{k=1}^{n} (X_{N_{2k}\wedge n} - X_{N_{2k-1}\wedge n}) &= \sum_{k=1}^{U_n[a,b]} (X_{N_{2k}} - X_{N_{2k-1}}) + (X_n - X_{N_{2U_n[a,b]}+1})\mathbf{1}(N_{2U_n+1} \leq n) \\ &\geq (b-a) \cdot U_n[a,b] + (X_n - X_{N_{2U_n[a,b]}+1})\mathbf{1}(N_{2U_n+1} \leq n),\end{aligned}$$

by the estimate (2.2). The second term on the far right can be bounded by below by

$$(X_n - X_{N_{2U_n[a,b]}-1})\mathbf{1}(N_{2n-1} \leq n) \geq -(X_n - a)^-;$$

we are simply maximizing our possible gambling losses during the last interval (we could be ahead).

Finally, we take expectations,

$$0 \geq \mathbb{E}\left[\sum_{k=1}^{n} (X_{N_{2k}\wedge n} - X_{N_{2k-1}\wedge n})\right] \geq (b-a)\mathbb{E}[U_n[a,b]] - \mathbb{E}[(X_n - a)^-],$$

and the theorem follows after rearranging. □

We now prove the martingale convergence theorem.

Proof of Theorem 2.15. We continue to use the notation adopted in the proof of the upcrossing inequality.

Let us define, for $a, b \in \mathbb{Q}$ with $a < b$,

$$U_\infty[a,b] = \lim_{n\to\infty} U_n[a,b].$$

We first argue that

$$\mathbb{P}(U_\infty[a,b] < \infty) = 1.$$

It follows directly from 2.18 that

$$(b-a)\mathbb{E}[U_n[a,b]] \leq |a| + \mathbb{E}[|X_n|],$$

and by the L^1-boundedness assumption,

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

Now $U_n[a,b] \uparrow U_\infty[a,b]$, and hence, by the monotone convergence theorem, $\mathbb{E}[U_\infty[a,b]] < \infty$, and so the limit is finite almost surely.

The rest of the proof is in essence a consequence of our deterministic lemma concerning upcrossings. Namely, we write

$$\left\{\omega \in \Omega \colon \lim_{n\to\infty} X_n(\omega) \text{ does not exist}\right\} = \bigcup_{a,b\in\mathbb{Q},\, a<b} \left\{\liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)\right\}.$$

Each set in the union is contained in the event $\{U_\infty[a,b] = \infty\}$, which we know occurrs with probability 0. A countable union of sets of sets with probability 0 has probability 0, and so

$$\mathbb{P}\left(\lim_{n\to\infty} X_n(\omega) \text{ does not exist}\right) = 0.$$

Thus $X_n \to X_\infty$ almost surely. An application of Fatou's lemma yields

$$\mathbb{E}[|X_\infty|] \leq \liminf_n \mathbb{E}[|X_n|] \leq \sup_n \mathbb{E}[|X_n|],$$

and so $X_\infty \in L^1$, as claimed. □

2.4 Doob's inequalities

In applications where martingales feature, it is often of interest to control the maximum of a martingale over a time interval. This maximum will be denoted by

$$X_n^* = \sup_{0\leq k\leq n} |X_k|.$$

For instance, if we wish to establish convergence results for a family of stochastic processes indexed by a parameter α, one strategy is to extract a martingale from the process, and then use bounds of the type presented here to reduce the problem to controlling the asymptotics of the single random variable $|X_n|$ in L^p, in terms of α.

Theorem 2.19. **[Doob's maximal inequality]** *Let $X = (X_n \colon n \geq 0)$ be a non-negative submartingale. Then, for $\lambda \in \mathbb{R}$,*

$$\lambda\mathbb{P}(X_n^* \geq \lambda) \leq \mathbb{E}[X_n \mathbf{1}(X_n^* \geq \lambda)] \leq \mathbb{E}[X_n].$$

Proof. Let $\lambda \in \mathbb{R}$ be given, and consider the (not necessarily bounded) stopping time $T = \inf\{k \geq 0 \colon X_k \geq \lambda\}$. Then

$$\{T \leq n\} = \{X_n^* \geq \lambda\}. \tag{2.3}$$

Now with n given as in the theorem, we consider the bounded stopping time $T \wedge n$, and apply the Optional Stopping Theorem 2.13 to get

$$\mathbb{E}[X_n] \geq \mathbb{E}[X_{T\wedge n}].$$

We next write

$$\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_T \mathbf{1}(T \leq n)] + \mathbb{E}[X_n \mathbf{1}(T > n)],$$

and note that on $\{T \leq n\}$, we have $X_T \geq \lambda$, so that

$$\mathbb{E}[X_{T\wedge n}] \geq \lambda \mathbb{P}(T \leq n) + \mathbb{E}[X_n \mathbf{1}(T > n)]. \tag{2.4}$$

Rearranging (2.4), we find that

$$\lambda \mathbb{P}(T \leq n) \leq \mathbb{E}[X_n] - \mathbb{E}[X_n \mathbf{1}(T > n)] = \mathbb{E}[X_n \mathbf{1}(T \leq n)].$$

Reinterpreting everything using (2.3), we arrive at the first inequality of the theorem.

The second inequality follows directly. □

Theorem 2.20. [Doob's L^p-inequality] *Let X be a martingale, or a non-negative submartingale. Then, for all $p > 1$,*

$$\|X_n^*\|_p \leq \frac{p}{p-1} \|X_n\|_p.$$

Proof. It is enough to prove the statement for a non-negative submartingale. For if X is a martingale, then $|X|$ is a non-negative submartingale by Jensen's inequality.

Let $M < \infty$ be fixed, and consider the random variable $X_n^* \wedge M$. By Fubini's theorem, we have

$$\mathbb{E}[(X_n^* \wedge M)^p] = \mathbb{E}\left[\int_0^M p x^{p-1} \mathbf{1}(X_n^* \geq x) dx\right] = \int_0^M p x^{p-1} \mathbb{P}(X_n^* \geq x) dx.$$

By Doob's maximal inequality,

$$\int_0^M p x^{p-1} \mathbb{P}(X_n^* \geq x) dx \leq \int_0^M p x^{p-2} \mathbb{E}[X_n \mathbf{1}(X_n^* \geq x)] dx,$$

and using Fubini's theorem again, we obtain that

$$\int_0^M p x^{p-2} \mathbb{E}[X_n \mathbf{1}(X_n^* \geq x)] dx = \frac{p}{p-1} \mathbb{E}[X_n \cdot (X_n^* \wedge M)^{p-1}].$$

An application of Hölder's inequality (note that $1/q = (p-1)/p$) yields

$$\frac{p}{p-1} \mathbb{E}[X_n \cdot (X_n^* \wedge M)^{p-1}] \leq \frac{p}{p-1} \|X_n\|_p \cdot (\|(X_n^* \wedge M)\|_p)^{p-1},$$

and we conclude that

$$\mathbb{E}[(X_n^* \wedge M)^p] \leq \frac{p}{p-1} \|X_n\|_p \cdot \|(X_n^* \wedge M)\|_p^{p-1}.$$

Dividing by the second factor on the right, we obtain

$$\|X_n^* \wedge M\|_p \leq \frac{p}{p-1} \|X_n\|_p$$

Since $X_n^* \wedge M \uparrow X_n^*$ as $M \to \infty$, and all the random variables are non-negative, an appeal to the monotone convergence theorem completes the proof. □

Exponential bounds are often useful. Here is a very simple example: applying Doob's maximal inequality to the submartingale $Y_n = \exp(\theta X_n)$, we obtain

$$\mathbb{P}(X_n^* \geq \lambda) \leq e^{-\theta\lambda} \mathbb{E}[e^{\theta X_n}],$$

and we can now optimize over $\theta \in \mathbb{R}$ to make the bound as effective as possible. Another result in this vein is the *Azuma-Hoeffding inequality*, which states that if $|X_{n+1} - X_n| \leq c_n$ for all $n \geq 1$, then

$$\mathbb{P}\left(X_n^* > \lambda\right) \leq 2 \exp\left(-\frac{\lambda^2}{2\sum_{k=1}^n c_k^2}\right).$$

See Example Sheet 2 for a proof.

2.5 L^p-convergence for martingales

We return to the question of convergence of martingales. We already know that an L^1-bounded martingale converges almost surely. However, as is often the case of L^1, we need to impose additional assumptions in order to deduce convergence in the space.

Pleasantly, the situation is better when $p > 1$; in that case boundedness in L^p, that is

$$\sup_{n \geq 0} \|X_n\|_p < \infty,$$

is sufficient for convergence in L^p, essentially because of Doob's inequality. Moreover, L^p-bounded martingales can be described in terms of closed martingales.

Theorem 2.21. *Let $p > 1$, and let $X = (X_n \colon n \geq 0)$ be a martingale. Then the following are equivalent:*

1. *X is bounded in L^p.*
2. *X converges almost surely and in L^p.*
3. *There exists a random variable $\mathfrak{X}$ such that $X_n = \mathbb{E}[\mathfrak{X}|F_n]$.*

Proof. We begin by showing that (*1*) implies (*2*).

On a probability, L^p-boundedness implies boundedness in L^1. Hence $X_\infty = \lim_n n \to \infty X_n$ exists almost surely, and is an integrable random variable. In fact, $X_\infty \in L^p$ since, by Fatou's lemma,

$$\|X_\infty\|_p \leq (\liminf_n \|X_n\|_p)^{1/p},$$

and $\liminf_n \|X_n\|_p \leq \sup_n \|X_n\|_p$.

Now consider the random variables $Y_n = X_n - X_\infty$; we shall show that there exists a random variable in L^p that dominates Y_n, $n \geq 1$. Doob's L^p-inequality tells us that

$$\|X_n^*\|_p \leq \frac{p}{p-1}\|X_n\|_p \leq \frac{p}{p-1}\sup_n \|X_n\|_p < \infty,$$

and then, since X_n^* is a non-decreasing sequence, we can apply the monotone convergence theorem to $X_\infty^* = \lim_n X_n^*$ and obtain

$$\|X_\infty^*\|_p < \infty.$$

Now $|X_n - X_\infty| \leq 2|X_\infty^*|$, and thus $Y_n \to 0$ in L^p.

Next, we show that (*2*) implies (*3*). By assumption, $X_\infty \in L^p$; what we shall show is that $\mathbb{E}[X_\infty|\mathcal{F}_n] = X_n$ almost surely, for each n.

Fixing an $n \geq 1$, we have $\mathbb{E}[X_k|\mathcal{F}_n] = X_n$ for $k \geq n$, by the martingale property. Thus

$$\mathbb{E}[|X_n - \mathbb{E}[X_\infty|\mathcal{F}_n]|^p] = \mathbb{E}[|\mathbb{E}[X_k - X_\infty|\mathcal{F}_n]|^p].$$

By the contractivity of conditional expectation then,

$$\|X_n - \mathbb{E}[X_\infty|\mathcal{F}_n]\|_p \leq \|X_k - X_\infty\|_p;$$

but the right-hand side can be made arbitrarily small by picking $k \geq n$ large enough, since $X_n \to X_\infty$ in L^p. Hence $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$ almost surely.

That (*3*) implies (*1*) follows from the contractivity of conditional expectation on L^p. □

We finally address the case when $p = 1$. To do this, we need to introduce another concept.

Definition 2.22. A collection $(X_j)_{j \in J}$ of random variables is *uniformly integrable* (in brief, UI) if

$$\sup_{j \in J} \mathbb{E}[|X_j|\mathbf{1}(|X_j| > M) \to 0 \quad \text{as} \quad M \to \infty.$$

A martingale $X = (X_n : n \geq 0)$ is said to be a *uniformly integrable (UI) martingale* if the (X_n) form a uniformly integrable collection of random variables.

A uniformly integrable family is bounded in L^1 but the converse is false. However, if a collection of random variables is bounded in L^p, for some $p > 1$, then it is uniformly integrable.

We are not imposing any assumptions about countability in the definition of uniform integrability.

Theorem 2.23. *Let X be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the family of random variables given by*

$$\{\mathbb{E}[X|\mathcal{G}] \colon \mathcal{G} \text{ is a sub } \sigma\text{-algebra of } \mathcal{F}\}$$

is uniformly integrable.

The relevance of uniform integrability in the context of convergence of martingales follows from the next lemma. We omit the proof.

Lemma 2.24. *Let $(X_n)_{n=1}^{\infty}$ and X be integrable random variables, and suppose $X_n \to X$ almost surely as $n \to \infty$. Then $X_n \xrightarrow{L^1} X$ if and only if $(X_n)_{n=1}^{\infty}$ is UI.*

Armed with this lemma, we can adapt the proof in the L^p setting to obtain the following definitive theorem.

Theorem 2.25. *Let $X = (X_n \colon n \geq 0)$ be a submartingale. Then the following are equivalent:*

1. *X is uniformly integrable.*
2. *X converges almost surely and in L^1.*
3. *There exists an integrable random variable $\mathfrak{X}$ such that $X_n = \mathbb{E}[\mathfrak{X}|\mathcal{F}_n]$.*

UI martingales have nice optional stopping properties. Suppose X is a UI martingale, and T is a stopping time, not necessarily bounded. We can then unambiguously define the stopped random variable

$$X_T = \sum_{k=1}^{\infty} X_k \mathbf{1}(T = k),$$

and the corresponding stopped process X^T as before.

Theorem 2.26. *Let $X = (X_n \colon n \geq 0)$ be a UI martingale, and let $S \leq T$ be stopping times.*

$$\mathbb{E}[X_T|\mathcal{F}_S] = \mathcal{F}_S$$

holds almost surely.

2.6 Some applications of martingale techniques

Random walks arise frequently in applications, and they are in many ways the perfect examples of martingales in discrete time: simple enough that many associated quantities can be computed exactly, while at the same time exhibiting interesting and non-trivial behavior.

We define $S = (S_n \colon n \geq 0)$ by setting $S_0 = 0$ and taking $S_n = \sum_{k=1}^{n}$ for $n = 1, 2, \ldots$. The process S is known as the *simple symmetric random walk* on the integers, and we have already seen that it is a martingale with respect to the natural filtration $\mathcal{F}_n = \sigma(X_k \colon k \geq 1)$.

For $c \in \mathbb{Z}$, we define the hitting time

$$T_c = \inf\{n \geq 0\colon S_n = c\};$$

as we have seen previously, this is stopping time. Now suppose $a, b \in \mathbb{Z}_+$. We wish to compute the probability that simple symmetric random walk hits $-a$ before b; this is the quantity $\mathbb{P}(T_{-a} < T_b)$. (Imagine two gamblers, A and B, with initial capital a and b, respectively, playing a fair coin-tossing game with unit stakes. What is the probability that gambler A loses all his money first?)

Let $T = T_{-a} \wedge T_b$. We will argue that $\mathbb{E}[T] < \infty$. If we observe $a+b$ occurrences of $+1$'s, then T has definitely been triggered, and such a string of $+1$'s will arise with probability $2^{-(a+b)}$. If one of the first $X_1, \ldots, X_{a+b}$ was -1 we discard this block, and consider the next block. The occurrence of $a+b$ consecutive $+1$'s independent of the first block, and we see that T is bounded by a geometric random variable of finite mean times $a+b$. Thus $\mathbb{E}[T] < \infty$, as claimed.

We are therefore in the setting of a martingale S with bounded increments, and a stopping time with finite expectation. By the optional stopping theorem then,

$$\mathbb{E}[S_T] = \mathbb{E}[S_0] = 0.$$

By definition, S_T can take one of two values, and so

$$\mathbb{E}[S_T] = -a \cdot \mathbb{P}(T_{-a} < T_b) + b \cdot \mathbb{P}(T_b < T_{-a}) = 0. \tag{2.5}$$

Since $-a \neq b$, we have

$$\mathbb{P}(\Omega) = \mathbb{P}(T_{-a} < \mathcal{T}_b) + P(T_{-a} > T_b) = 1. \tag{2.6}$$

Using (2.5) and (2.6), we can solve for $\mathbb{P}(T_{-a} < T_b)$, and we obtain the so-called *gambler's ruin estimate*

$$\mathbb{P}(T_{-a} < T_b) = \frac{b}{a+b}. \tag{2.7}$$

(If $a = b$ we get $\mathbb{P}(T_{-a} < T_a) = 1/2$, which agrees with our intuition.)

It is important to be careful when applying the optional stopping theorem; when deriving (2.7) we put some effort into showing that $\mathbb{E}[T_{-a} \wedge T_b] < \infty$ for all $a, b \neq 0$. Here is an example of how things can go wrong. Since T_1 is a stopping time, $\mathbb{E}[S_{T_1 \wedge n}] = E[S_0]$ for all n. However, $S_{T_1} = 1$, and

$$1 = \mathbb{E}[S_{T_1}] \neq \mathbb{E}[S_0] = 0.$$

(The walk can spend an arbitrarily long period in $\mathbb{Z}_-$ before coming up to $x = 1$, and we are preventing this when considering $T_1 \wedge n$.)

Here is another example of when stopping times with finite expectation behave much like constant times. Consider an i.i.d. sequence of random variables $(X_k)_{k=1}^{\infty}$ having $\mathbb{E}[X_1] = \mu < \infty$. (The statement that will follow is valid for simple symmetric random walk, but it is not as striking then.) We again form $S_n = \sum_{k=1}^{n} X_k$. By linearity, we clearly have, for any fixed n,

$$\mathbb{E}[S_n] = n \cdot \mathbb{E}[X_1].$$

It can be shown that if T is a stopping time with $\mathbb{E}[T] < \infty$, then in fact

$$\mathbb{E}[S_T] = \mathbb{E}[T]\mathbb{E}[X_1].$$

This last identity is called *Wald's equation*; you will be asked to provide a proof in Example Sheet 2.

Definition 2.27. Let $(\mathcal{G}_n)_{n\leq 0}$ be a sequence of σ-algebras with

$$\cdots \subset \mathcal{G}_{-2} \subset \mathcal{G}_{-1} \subset \mathcal{G}_0 \subset \mathcal{F}.$$

A stochastic process $X = (X_n : n \leq 0)$ is said to be a *backwards martingale* if X is adapted to $(\mathcal{G}_n)$, the random variable X_0 is integrable, and

$$\mathbb{E}[X_{n+1}|\mathcal{G}_n] = X_n \quad \text{almost surely.}$$

We note that the tower property of conditional expectation implies that

$$\mathbb{E}[X_0|\mathcal{G}_n] = X_n \quad \text{a.s.}$$

for all $n < 0$. It is a pleasant feature of backwards martingales that they are automatically uniformly integrable: this follows from the above equality and Theorem 2.23, and ultimately from the assumption that $X_0 \in L^1$. This means that, unlike in the case of martingales, we do not need to distinguish between $p > 1$ and $p = 1$ in the statement of the martingale convergence theorem.

Theorem 2.28. [Martingale convergence theorem for backwards martingales] *Let $X = (X_n : n \leq 0)$ be a backwards martingale with respect to the filtration $(\mathcal{G}_n)_{n\leq 0}$. Suppose $X_0 \in L^p$, for some $1 \leq p < \infty$, and set $\mathcal{G}_{-\infty} = \bigcap_{n\leq 0} \mathcal{G}_n$.*

Then X_n converge almost surely, and in L^p, to the random variable $X_{-\infty} = \mathbb{E}[X_0|\mathcal{G}_{-\infty}]$ as $n \to -\infty$.

The proof is similar to the proofs in the preceding sections (it uses upcrossings, Jensen's inequality in its conditional form, etc.).

In several of our applications, we will need the notion of tail σ-algebra.

Definition 2.29. Let $(X_n)_{n\geq 1}$ be a sequence of random variables. Set $\mathcal{F}_n = \sigma(X_n, X_{n+1}, \ldots)$. The *tail σ-algebra* or *remote future* is defined as

$$\mathcal{F}_\infty = \bigcap_{n=1}^{\infty} \mathcal{F}_n.$$

In words, the tail σ-algebra contains those events whose occurrence is unaffected by changing a finite number of X_k's. Tail σ-algebras have a very surprising feature.

Theorem 2.30. [Kolmogorov's 01-law] *Let $(X_k)_{k\geq 1}$ be a sequence of i.i.d. random variables, and set $\mathcal{F}_n = \sigma(X_k, k \geq n)$. Then $\mathcal{F}_\infty$ is trivial:*

$$\mathbb{P}(A) \in \{0, 1\} \quad \textit{for every} \quad A \in \mathcal{F}_\infty.$$

Proof. Let $\mathcal{G}_n = \sigma(X_k, k \le n)$ and $A \in \mathcal{F}_\infty$. Since $\mathcal{G}_n$ is independent of $\mathcal{F}_{n+1}$, we have that

$$\mathbb{E}[\mathbf{1}(A)|\mathcal{G}_n] = \mathbb{P}(A) \text{ a.s.}$$

Theorem 2.25 gives that $\mathbb{E}[\mathbf{1}(A)|\mathcal{G}_n]$ converges to $\mathbb{E}[\mathbf{1}(A)|\mathcal{G}_\infty]$ a.s. as $n \to \infty$, where $\mathcal{G}_\infty = \sigma(\mathcal{G}_n, n \ge 0)$. Hence we deduce that

$$\mathbb{E}[\mathbf{1}(A)|\mathcal{G}_\infty] = \mathbf{1}(A) = \mathbb{P}(A) \text{ a.s.,}$$

since $\mathcal{F}_\infty \subseteq \mathcal{G}_\infty$, and the theorem follows. □

Kolmogorov's 01-law has ramifications for random infinite series.

Example 2.31. *Let $X_1, X_2, \dots$ be a sequence of independent random variables, and let $S_n = X_1 + \cdots + X_n$. Then the event $\{\lim_{n\to\infty} S_n \text{ exists}\}$ belongs to $\mathcal{F}_\infty$, but $\{\lim_{n\to\infty} S_n \ge 0\} \notin \mathcal{F}_\infty$.*

Thus, Kolmogorov's 0-1 law implies that a random series either converges almost surely, or with probability zero. It turns out that the former occurs provided a simple variance condition is satisfied.

Theorem 2.32. [Kolmogorov's inequality] Suppose $(X_k)_{k\ge1}$ are independent random variables with $\mathbb{E}[X_k] = 0$ and $\mathrm{var}(X_n) = \sigma_k^2 < \infty$, and set $S_n = X_1 + \cdots + X_n$. Then

$$\mathbb{P}\left(\sup_{k\le n} |S_k| > \lambda\right) \le \frac{\mathrm{var}(S_n)}{\lambda^2} = \frac{1}{\lambda^2}\sum_{k=1}^n \sigma_k^2.$$

Proof. See Example Sheet 2. □

Theorem 2.33. *Suppose $X_1, X_2, \dots$ are independent with $\mathbb{E}[X_n] = 0$ and $\mathrm{var}(X_k) = \sigma_k^2 < \infty$. If $\sum_{n=1}^\infty \sigma_k^2 < \infty$ then*

$$\sum_{k=1}^\infty X_k(\omega) = \lim_{N\to\infty} \sum_{k=1}^N X_k(\omega)$$

exists for almost every ω.

Proof. Set $S_N = \sum_{k=1}^N X_k$. Then, by the previous theorem,

$$\mathbb{P}\left(\sup_{N\le k\le M} |S_k - S_N| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\sum_{k=N+1}^M \sigma_k^2.$$

We now let $M \to \infty$ in this inequality, and so obtain

$$\mathbb{P}\left(\sup_{N\le k} |S_k - S_N| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\sum_{k=N+1}^\infty \sigma_k^2 \to 0, \quad N \to \infty,$$

since the full series $\sum_k \sigma_k^2$ converges. But then

$$\mathbb{P}\left(\sup_{k,l\ge N} |S_k - S_l| > 2\varepsilon\right) \le \mathbb{P}\left(\sup_{k\ge N} |S_k - S_N| > \varepsilon\right) \to 0, \quad N \to \infty.$$

This means $S_n(\omega)$ is a Cauchy sequence for almost all $\omega \in \Omega$, and hence the series converges almost surely. □

Example 2.34. **[Random harmonic series]** *Let* $X_1, X_2, \dots$ *be independent random variables with* $\mathbb{P}(X_k = 1) = 1/2$ *and* $\mathbb{P}(X_k = -1) = 1/2$*. Then the series* $\sum_{k=1}^{\infty} X_k/k$ *converges almost surely.*

3 Stochastic processes in continuous time

3.1 Basic definitions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. In the first half of the course, we restricted ourselves to the study of real-valued (or integer-valued) stochastic processes in discrete time, and our main focus was on *martingales*. These are sequences of random variables indexed by $\mathbb{N}$ (or $\mathbb{Z}$ in the case of backwards martingales), satisfying the key requirement that they be measurable with respect to a filtration of the underlying space, an increasing sequence of σ-algebras contained in $\mathcal{F}$, and possess a certain conditional expectation property.

We now wish to extend the theory to stochastic processes indexed by a continuous parameter $t \in I$, where $I \subseteq \mathbb{R}_+$ is a possibly infinite interval. It seems clear how we should attempt to proceed. We define a *filtration* $(\mathcal{F}_t)_t$ to be an increasing collection of sub σ-algebras of $\mathcal{F}$:

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}, \quad s \le t.$$

Definition 3.1. A *stochastic process in continuous time* is a collection $(X_t \colon t \in I)$ of random variables on Ω. A process X is *adapted* to a filtration $(\mathcal{F}_t)$ if X_t is $\mathcal{F}_t$-measurable for all $t \in I$.

Definition 3.2. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space. An adapted process $X = (X_t \colon t \in I)$ is a *martingale* with respect to $(\mathcal{F}_t)_t$ if $\mathbb{E}[|X_t|] < \infty$ and, for $s \le t$,

$$\mathbb{E}[X_t | \mathcal{F}_s] = X_s \text{ almost surely.}$$

Submartingales and supermartingales are defined in the same fashion, with the appropriate choice of inequalities.

Definition 3.3. A *stopping time* T on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ is a random variable $T \colon \Omega \to [0, \infty]$ such that $\{T \le t\} \in \mathcal{F}_t$, for all $t \in I$.

These definitions all feel natural, but as soon as we start trying to prove theorems in this framework, we realize that we will run into difficulties. The main problems we encounter are related to measurability. Recall that the definition of a σ-algebra only requires *countable* unions of sets in the algebra to belong to the σ-algebra. When we work with processes indexed by a continuous parameter, we need to form *uncountable* unions, intersections, and so on, and it turns out that such sets often *fail* to be measurable.

Let us illustrate these problems by considering an archetypal stopping time: the *hitting time* of a Borel set $A \subset \mathbb{R}$:

$$T_A = \inf\{t : X_t \in A\}.$$

In discrete time, we proved this was indeed a stopping time by decomposing the event $\{T_A \le n\}$ as

$$\{T_A \le n\} = \bigcup_{0 \le k \le n} \{X_k \in A\},$$

and using that this was then a countable union of measurable sets. In this setting, the analogous approach leads to an uncountable union, and there is nothing to guarantee that this union remains measurable–indeed, in general, it does not.

Example 3.4. Let X' be a random variable with $\mathbb{P}(X' = 1) = \mathbb{P}(X' = -1) = 1/2$. We introduce the process

$$X_t = \begin{cases} t, & \text{if } t \in [0,1] \\ 1 + (t-1)X', & \text{if } t > 1 \end{cases}.$$

Let $(\mathcal{F}^X)$, $\mathcal{F}^X_t = \sigma(X_s, s \le t)$, denote the natural filtration associated with the process X. The set $A = (1,2)$ is open and hence a Borel set, and the associated hitting time is $T_A = \inf\{t \ge 0 : X_t \in A\}$. In view of how the process is defined, $\{T_A \le 1\} \notin \mathcal{F}_1$, and so T_A is *not* a stopping time.

In this example, the information contained in $\mathcal{F}_1$ is simply not sufficient for us to be able to tell what X will do "immediately after" $t = 1$, and whether X will go on to enter A or not. This suggests that imposing some kind of continuity requirement on $(\mathcal{F}_t)_t$ might be helpful.

Before we address this issue we should think about different ways to view stochastic processes. In discrete time, we can equip the index set $\mathbb{N}$ with the σ-algebra $\mathcal{P}(\mathbb{N})$ that contains *all* subsets of $\mathbb{N}$. Then the mapping

$$(n, \omega) \mapsto X_n(\omega)$$

is measurable with respect to the product σ-algebra $\mathcal{F} \otimes \mathcal{P}(\mathbb{N})$ due to countability. By contrast, if we fix $t \in I$, and turn to a process in continuous time, then $\omega \mapsto X_t(\omega)$ is again a random variable, but we cannot argue in the same way and assert that

$$(t, \omega) \mapsto X_t(\omega)$$

is automatically measurable with respect to $\mathcal{F} \otimes \mathcal{B}(\mathbb{R})$. In fact, we would like the following, stronger, type of measurability property to hold.

Definition 3.5. *A stochastic process* $(X_t \colon t \in I)$ *is said to be progressively measurable if, for any* $A \in \mathcal{B}(\mathbb{R})$ *and* $t \ge 0$*,*

$$\{(s, \omega) \in [0,t] \times \Omega : X_s(\omega) \in A\} \in \mathcal{B}([0,t]) \otimes \mathcal{F}_t.$$

In order to proceed, let us require more of our processes. We say that a process $X = (X_t \colon t \in I)$ is *continuous* if

$$t \mapsto X_t(\omega), \quad \omega \in \Omega \quad \text{fixed},$$

is a continuous function of the real variable t. As $\omega \in \Omega$ varies, these functions of t are called the *sample paths* of the process. Intuitively, a sample path is what we see when we track one particular outcome or realization of a random phenomenon, such as physical Brownian motion, over a period of time, and we often think of processes we observe in nature as being (at least) continuous. The key point is that a continuous function on $I \subseteq \mathbb{R}_+$ is determined by its values on a countable dense subset, for instance $\mathbb{Q}$ or the dyadic rationals, and this allows us to bypass the problems associated with uncountable unions and intersections by restricting ourselves to countable operations, and then extending the results we get by continuity. This is a very appealing strategy, as it suggests that our results on martingales in discrete time should be readily extendable, provided we work in the continuous setting.

There are, however, many natural processes, such as the Poisson process we shall encounter at the end of this chapter, that exhibit a weaker type of continuity. We will work in a slightly more general framework in order to be able to accommodate such jump processes. We will be concerned with functions $t \mapsto f(t)$ that are *right-continuous and admit left limits*, and processes whose sample paths possess these properties. Such processes are frequently called *cadlag* from the French "continu à droite limité à gauche". The abbrevation *RCLL*, or *right-continuous with left limits*, also appears in the literature. It should be pointed out that cadlag functions can only have jump discontinuities; such discontinuities occur when

$$\Delta f(t) = f(t) - \lim_{s \uparrow t} f(s) \neq 0 \quad \text{for some} \quad t \in \mathbb{R}_+.$$

We now set down notation for spaces of continuous functions. We let $C(\mathbb{R}_+, E)$ be the space of continuous functions $f \colon \mathbb{R}_+ \to E$; here, it is understood that E is a topological space. The space $D(\mathbb{R}_+, E)$ consists of all cadlag functions. For the time being, we take $E = \mathbb{R}$, but we shall also consider $E = \mathbb{R}^d$ in this course. The spaces $C(\mathbb{R}_+, E)$ and $D(\mathbb{R}_+, E)$ can be endowed with topologies that turn them into Polish spaces. (We shall return to this shortly.)

The following lemma shows that real-valued cadlag functions are reasonably well-behaved, and that we can argue "countably" when dealing with cadlag processes.

Lemma 3.6. *For each cadlag function $f \colon [a, b] \to \mathbb{R}$, and each $\varepsilon > 0$, there exists a finite partition $P = \{a = t_0 \leq t_1 \leq \cdots \leq t_n = b\}$ such that*

$$\sup\{|f(x) - f(y)| : x, y \in [t_{j-1}, t_j),\ j = 1, \ldots, n\} < \varepsilon.$$

In particular, the set $\Delta_f = \{t \in [a, b] : \Delta f(t) \neq 0\}$ is at most countable.

Proof. Let $\varepsilon > 0$ be given. Since f is right-continuous at a and has limits from the left at any $t > a$, there exists an interval $[a, T_0)$, with $T_0 \leq b$, that admits such a partition. Let T_0^* denote the maximal such T_0, and suppose $T_0^* < b$. Now f is again right-continuous at the point T_0^*, and so there must exist some $T_0^* < T_1 \leq b$ such that $\sup\{|f(x) - f(y)| : x, y \in [T_0^*, T_1)\} < \varepsilon$, and so the previous partition can be extended. This contradicts the maximality of T_0^*, and thus we must have $T_0 = b$, and the first statement follows.

To establish the second assertion, note that $\Delta_f = \bigcup_{k=1}^{\infty} \{t : \Delta f(t) > 1/k\}$. □

Let us return to processes to see how our cadlag assumptions pay off. Suppose $X = (X_t \colon t \in I)$ is cadlag and adapted; we return to the mapping

$$(\omega, t) \mapsto X_t(\omega)$$

and the question of progressive measurability. By right-continuity of X at $s \in [0, t]$, we can write

$$X_s(\omega) = \lim_{n \to \infty} X_s^{(n)}(\omega).$$

using the dyadic construction

$$X_s^{(n)}(\omega) = X_{(k+1)t/2^n}(\omega), \quad \text{for} \quad \frac{kt}{2^n} < s \leq \frac{(k+1)t}{2^n}.$$

Note the use of the right-hand endpoint on each dyadic interval. Each $X^{(n)}$ is $\mathcal{F} \otimes \mathcal{B}((0, t])$-measurable, and hence $X_t(\omega)$ is $\mathcal{F} \otimes \mathcal{B}((0, t])$-measurable as a limit of measurable functions.

Lemma 3.7. *Suppose X is a cadlag process adapted to $(\mathcal{F}_t)_t$. Then X is progressively measurable.*

We return to stopping times. For a cadlag process X and a finite stopping time, we set $X_T(\omega) = X_{T(\omega)}(\omega)$. We obtain a *stopped* process $X^T = (X_t^T : t \in I)$ by setting $X_t^T = X_{T\wedge t}$. The stopped σ-algebra is defined as before,

$$\mathcal{F}_T = \{A \in \mathcal{F} \colon A \cap \{T \le t\} \in \mathcal{F}_t \ \ \forall t \in I\}$$

We now show that a continuous-time analog of Proposition 2.8 holds, provided we agree to work with cadlag processes.

Proposition 3.8. *Let $X = (X_t \colon t \in I)$ be a cadlag adapted process, and let S and T be stopping times with respect to $(\mathcal{F}_t)$. Then the following holds.*

1. *$S \wedge T$ is a stopping time,*
2. *if $S \le T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$,*
3. *$X_T \mathbf{1}(T < \infty)$ is an $\mathcal{F}_T$-measurable random variable,*
4. *X^T is adapted.*

Proof. The first two items are straight-forward.

We prove that the third assertion holds. We first note that a random variable Y is $\mathcal{F}_T$-measurable precisely when $Y\mathbf{1}(T \le t)$ is $\mathcal{F}_t$-measurable for every $t \ge 0$. Namely, if A is a Borel set, then

$$\{Y \in A\} \cap \{T \le t\} = \{Y\mathbf{1}(T \le t) \in A\} \cap \{T \le t\},$$

and since T is a stopping time, the latter intersection is in $\mathcal{F}_t$ if $Y\mathbf{1}(T \le t)$ is $\mathcal{F}_t$-measurable.

Next, we approximate the stopping time T by the dyadic random variable $T_n = 2^{-n}\lceil 2^n T\rceil$. As $n \to \infty$, we note that $T_n \downarrow T$ and $T_n = \infty$ whenever $T = \infty$. Since $\{T_n \le t\} = \{T \le 2^{-n}\lfloor 2^n t\rfloor\}$, we have $\{T_n \le t\} \in \mathcal{F}_t$, and so (T_n) forms a sequence of stopping times.

We denote by $D_n = \{k/2^n : k \in \mathbb{N}\}$ the set of dyadic rationals of level n. From the representation

$$X_{T_n \wedge t} = X_t\mathbf{1}(T_n > t) + \sum_{\{d \in D_n \colon d \le t\}} X_d \mathbf{1}(T_n = d),$$

it follows that $X_{T_n\wedge t}$ is $\mathcal{F}_t$-measurable for each n. Since X is a cadlag process, and T_n converges from above,

$$X_T\mathbf{1}(T < t) = \lim_{n\to\infty} X_{T_n\wedge t}\mathbf{1}(T < t),$$

and it follows that $X_T\mathbf{1}(T < t)$ is $\mathcal{F}_t$-measurable. Finally, we note that

$$X_T\mathbf{1}(T \le t)\mathbf{1}(T < \infty) = X_t\mathbf{1}(T = t) + X_T\mathbf{1}(T < t),$$

which is a sum of $\mathcal{F}_t$ measurable random variables, and so *(3.)* follows.

To deduce the last item from the third, we consider the stopping time $T \wedge t$. Then $X_{T\wedge t}$ is $\mathcal{F}_{T\wedge t}$-measurable, and by *(2.)*, $\mathcal{F}_{T\wedge t} \subset \mathcal{F}_t$. It follows that X_T is adapted with respect to $(\mathcal{F}_t)$. □

We are now in a position to deal with hitting times. We restrict ourselves to continuous processes, and begin with the case of closed sets.

Proposition 3.9. *Suppose A is a closed set and $X = (X_t : t \in I)$ is a continuous process, adapted to the filtration $(\mathcal{F}_t)_t$. Then the hitting time of A,*

$$T_A = \inf\{t \in I : X_t \in A\}$$

is an $(\mathcal{F}_t)_t$-stopping time.

Proof. Let $d(x, A) = \inf_{y \in A} |x - y|$ denote the distance from a point $x \in \mathbb{R}$ to the closed set A.

Suppose that, for some sequence $(q_n) \subset \mathbb{Q} \cap [0, t] \cap I$, the distance $d(X_{q_n}(\omega), A)$ tends to 0 as $n \to \infty$. Then by boundedness there exists a subsequence (q_{n_k}) such that $q_{n_k} \to s \leq t$, and since X is continuous, $d(X_{q_{n_k}}, A) \to 0$. Since A is a closed set, we deduce that $X_s \in A$.

Conversely, if $X_s \in A$ for some $s \in [0, t] \cap I$, then $X_{q_n} \to X_s$ for any sequence of rational numbers in $[0, t] \cap I$ that converges to s, and hence also $d(X_{q_n}, A) \to 0$ for such a sequence.

Thus, the event $\{T_A \leq t\}$ can be represented using an enumeration of the rationals in $[0, t] \cap I$,

$$\{T_A \leq t\} = \bigcap_{k=1}^{\infty} \bigcup_{s \in \mathbb{Q} \cap [0,t] \cap I} \{d(X_s, A) \leq 1/k\} \in \mathcal{F}_t,$$

and this shows that T_A is a stopping time. □

Before we proceed to deal with hitting times for open sets, we make the notion of "continuity" of a filtration precise.

Definition 3.10. Let $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ be a filtration. For each t we define

$$\mathcal{F}_{t+} = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}.$$

If $\mathcal{F}_{t+} = \mathcal{F}_t$ for all t, the filtration $(\mathcal{F}_t)$ is said to be *right-continuous.*[3]

In other words, if a filtration is right-continuous, then "infinitesimal sneak previews" of future events make no difference.

One proves that for open sets $U \subset \mathbb{R}$, hitting times are stopping times with respect to the (potentially bigger) σ-algebras $(\mathcal{F}_{t+})$

Proposition 3.11. *Suppose U is an open set in $\mathbb{R}$ and $(X_t : t \in I)$ is a continuous process, adapted to the filtration $(\mathcal{F}_t)$. Then*

$$T_U = \inf\{t \in I : X_t \in U\}$$

is an $(\mathcal{F}_{t+})$-stopping time.

[3]Thus, right-continuity is used in two contexts: to describe a property of sample paths, and to describe a property of σ-algebras.

Proof. Since X is adapted and U is open, $\{X_s \in U\} \in \mathcal{F}_s$. By continuity,

$$\{T_U < t\} = \bigcup_{s \in \mathbb{Q} \cap [0,t] \cap I} \{X_s \in U\},$$

and since this is a countable union for each t, we have $\{T_U < t\} \in \mathcal{F}_t$. Next, we may write $\{T_U \leq t\} = \bigcap_k \{T_U < t + 1/k\}$, and this intersection belongs to $\mathcal{F}_{t+}$ by definition. □

In general, it is not true that T_U is a stopping time with respect to $(\mathcal{F}_t)$.

3.2 Canonical processes and sample paths

We return to our previous discussion concerning the definition of a stochastic processes in continuous time. We have realized that it is desirable to ensure our stochastic processes, or more precisely, their sample paths $\omega \mapsto X_t(\omega)$, enjoy certain regularity properties. The question still remains whether any reasonable such processes exist, and if so, how to describe them. So far, we have primarily been viewing a stochastic process as a collection of random variables indexed by $t \in I \subseteq \mathbb{R}$, and it is not clear that prescribing the distribution of each random variable is consistent with requiring continuity properties of sample paths.

It is useful to change perspective at this point. Consider a space of functions, contained in the collection of all mappings

$$\mathbb{R}^I = \{f : I \to \mathbb{R}\},$$

distinguished by some desirable property. Two natural such spaces are $C(I, \mathbb{R})$, the space of continuous functions, or $D(I, \mathbb{R})$, the space of cadlag functions. As was mentioned earlier, these spaces can be endowed with reasonable topologies. In the case of $C(I, \mathbb{R})$, the topology induced by

$$\|f\|_\infty = \sup\{|f(x)| \colon x \in I\}$$

will do; it can be shown that it turns $C(I, \mathbb{R})$ into a complete and separable space with metric induced by the norm. (If I is unbounded, some modifications involving cutoffs need to be made, but we will not dwell on that here.) For the sake of definiteness, let us focus on this case for now. Once we have a topology, it makes sense to speak of open sets and Borel sets in $C(I, \mathbb{R})$, and the Borel σ-algebra generated by the open sets. We can then introduce *random variables taking values in a function space*. By this we simply mean a *measurable mapping* $X \colon \Omega \to C(I, \mathbb{R})$ from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into $C(I, \mathbb{R})$, and a stochastic process can be viewed as a single random variable. The *law of the process* X is the measure μ_X on $C(I, \mathbb{R})$ that is obtained as the push-forward of $\mathbb{P}$,

$$\mu_X(A) = \mathbb{P}(X \in A), \quad A \in \mathcal{B}(C(I, \mathbb{R})).$$

To connect the abstract viewpoint with our previous definition of a stochastic process, we introduce the *coordinate processes* $(X_t, t \in I)$, where $X_t \colon C(I, \mathbb{R}) \to \mathbb{R}$, by setting

$$\omega \in C(I, \mathbb{R}) \mapsto X_t(\omega) = \omega(t).$$

The coordinate mapping is measurable with respect to the Borel σ-algebra of $C(I, \mathbb{R})$ (see Example Sheet 3 for a proof), and so induces a real-valued stochastic process on I. While the abstract setup is elegant, it is not so easy to construct measures on spaces of functions, or measurable mappings into them.

Since the full law μ_X of a stochastic process is a rather unwieldy object to work with in practice, it is convenient to find simpler objects that still carry sufficient information about X.

Definition 3.12. Let $J = (t_1, \dots, t_n) \subset I$ be a finite set. The *finite-dimensional distribution* of a stochastic process $(X_t \colon t \in I)$ with index set $(t_1, \dots, t_n)$ is defined as the law μ_J of the $\mathbb{R}^J$-valued random variable $(X_{t_1}, \dots, X_{t_n})$.

This suggests a strategy for constructing stochastic processes. First, we specify what properties we would like the finite-dimensional distributions of the process to have. This is relatively easy. Then, provided the finite-dimensional distributions satisfy certain consistency criteria, a result of Kolmogorov asserts the existence of a measure on the space R^I having precisely these finite-dimensional distributions, and so the existence of our stochastic process. It is not clear, however, that this measure is supported on the spaces $C(I, \mathbb{R})$ or $D(I, \mathbb{R})$, that is, that the sample paths of a process constructed in this way are continuous or cadlag–this requires additional work.

Definition 3.13. Let X and X' be two processes defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X' is *a version* of X if

$$\mathbb{P}(X_t = X'_t) = 1 \quad \text{for every} \;\; t.$$

We say that X and X' are *indistinguishable* if

$$\mathbb{P}(X_t = X'_t \quad \forall t \in I) = 1.$$

Note carefully where the quantifiers are placed in these definitions.

The sample paths of indistinguishable processes have the same properties almost surely. Two versions of the same process have the same finite-dimensional distributions,

$$\mathbb{P}((X_{t_1}, \dots, X_{t_n}) \in A) = \mathbb{P}((X'_{t_1}, \dots, X'_{t_n}) \in A), \quad A \subset \mathcal{B}(R^J),$$

but they do not necessarily enjoy the same sample path properties. Here is an example of two different processes whose finite-dimensional distributions are the same, but whose path properties are different.

Example 3.14. Define $X = (X_t \colon t \in [0, 1])$ to be identically 0. Next, let T be any random variable on $[0, 1]$ with continuous distribution. (For instance, the uniform distribution will do.) Now define a second process by setting $X'_t = \mathbf{1}(T = t)$. Then X' is a version of X, since for every t, $\mathbb{P}(X_t = X'_t) = \mathbb{P}(T \neq t) = 1$. Moreover, X and X' have the same finite-dimensional distributions: they are point masses at the origin.

However,

$$\mathbb{P}(X'_t = X_t \;\; \forall t \in [0, 1]) = \mathbb{P}(X'_t = 0 \;\; \forall t \in [0, 1]) = 0.$$

and so X' is not indisinguishable from X, and moreover X' is not continuous.

We end this section by introducing another technical concept. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space, and let $\mathcal{N}$ denote the collection of sets in Ω that are contained in some set of measure 0 with respect to $\mathbb{P}$.

Definition 3.15. We say that a filtration $(\mathcal{F}_t)$ satisfies the *usual hypotheses*[4] if it is right-continuous, that is $\mathcal{F}_t = \mathcal{F}_{t+}$ for all t, and $\mathcal{F}_t$ contains $\mathcal{N}$.

The requirement of right-continuity is more substantial than the other condition. Namely, we can always define an *augmented filtration* $(\bar{\mathcal{F}}_t)_t$ by setting

$$\bar{\mathcal{F}}_t = \sigma(\mathcal{F}_t, \mathcal{N}),$$

and one proves that a $\mathcal{F}_t$-adapted process remains $\bar{\mathcal{F}}_t$-adapted, and satisfies $\mathbb{E}[X_t|\mathcal{F}_s] = \mathbb{E}[X_t|\bar{\mathcal{F}}_s]$ a.s.

3.3 Martingale regularization theorem

We have discussed cadlag processes at some length, but has not clear to what extent this hypothesis imposes restrictions on the main class of stochastic processes we are interested in at the moment, namely martingales in continuous time. Fortunately, the answer turns out to be rather encouraging.

Theorem 3.16. [Martingale regularization theorem] *Suppose $(X_t \colon t \in I)$ is a martingale with respect to a given filtration $(\mathcal{F}_t)_t$. Then there exists a process X' with the following properties:*

- *the sample paths of X' are cadlag almost surely*
- *X' is a martingale with respect to the filtration $(\mathcal{F}_t^*)$ given by $\mathcal{F}_t^* = \sigma(\mathcal{F}_{t+}, \mathcal{N})$, and satisfies*
$$X_t = \mathbb{E}[X_t'|\mathcal{F}_t] \quad \textit{almost surely},$$
for all $t \in I$.

If the filtration $(\mathcal{F}_t)$ satisfies the usual hypotheses, then X' is a cadlag version of X.

Proof. As a first step, we show that we have good control over certain discrete-time martingales that we can extract from X.

Let $P \subset \mathbb{Q} \cap I$ be bounded, and consider a finite subset $Q = \{q_1, \dots, q_n\} \subset P$, ordered in increasing order. The process $X^Q = (X_k^Q \colon 1 \leq k \leq n)$ defined by setting $X_k^Q = X_{q_k}$ is a martingale with respect to $(\mathcal{F}_{q_k})$. Since $|X^Q|$ is a submartingale, Doob's maximal inequality applies and yields

$$\lambda \mathbb{P}\left(\sup_{1 \leq k \leq n} |X_k^Q| > \lambda \right) \leq \mathbb{E}[|X_{q_n}|].$$

[4] This terminology is not very exciting, but it is by now standard.

Since P was assumed bounded , we have, with a suitable choice of T, a uniform bound

$$\lambda \mathbb{P}\left(\sup_{q_k \in Q} |X_{q_k}| > \lambda\right) \leq \mathbb{E}[|X_T|] \tag{3.1}$$

for any finite subset $Q \subset P$. By monotonicity then, after exhausting P by finite subsets Q, dividing by λ, and letting $\lambda \to \infty$, we deduce that $\mathbb{P}(\sup_{t\in P} |X_t| < \infty) = 1$ and so X is almost surely bounded when restricted to bounded subsets of the rationals.

Now that we have shown that X does not get too big on bounded subsets of rationals, we count upcrossings to rule out that X fluctuates to much to have a limit. Let $J \subset I \cap \mathbb{Q}$ be a bounded time-interval, and pick $a, b \in \mathbb{Q}$ with $a < b$. We compute the upcrossings of X over J using finite $Q \subset J$,

$$N([a, b], J) = \sup_{Q\subset J} N([a, b], Q),$$

and associate, as before, a martingale X^Q with each such subset. Applying Doob's upcrossings theorem, Theorem 2.16, we find that, for some suitable T,

$$(b - a)\mathbb{E}[N_{q_n}([a, b], Q)] \leq \mathbb{E}[(X_{q_n} - a)^-] \leq \mathbb{E}[(X_T - a)^-];$$

the last inequality holds due to the fact that $(X - a)^-$ is a submartingale. The bound on the right-hand side is independent of Q, meaning that $\mathbb{E}[N([a, b], J)] < C$ follows upon taking the supremum over all finite $Q \subset J$. Hence $N([a, b], J)$ is almost surely finite for any pair of rationals $a < b$, whenever J is bounded.

Still keeping the bounded set $J \subset I \cap \mathbb{Q}$ fixed, we observe that

$$\mathbb{P}\left(\bigcap_{a,b\in\mathbb{Q},a<b} \{N([a, b], J) < \infty\}\right) = 1, \tag{3.2}$$

since we are taking a countable intersection of events of probability 1.

We now exhaust $I \cap \mathbb{Q}$ (which may be unbounded) by a sequence of increasing bounded sets (J_n) consisting of rational numbers. Running the above argument for each one of these rational intervals, we obtain bounds in terms of $\mathbb{E}[|X_{T_n}|]$, as in (3.1). Using (3.2) on each J_n and finally intersecting over n, we obtain a set $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$, and such that X has the finite-upcrossing property on each J_n for any rationals $a < b$, and has $\sup_{J_n} |X_t| < \infty$, on Ω_0.

A lemma from real analysis states that a real-valued function that is bounded on bounded subsets of the rationals, and has a finite number of (a, b)-upcrossings on every such set, for any $a, b \in \mathbb{Q}$, admits left and right limits at every point. Hence, by our previous reasoning, the right-hand limit

$$X'_{t+}(\omega) = \lim_{s\downarrow t,\ s\in\mathbb{Q}} X_s(\omega), \quad \omega \in \Omega_0, \tag{3.3}$$

exists for $t \in I$, as does the corresponding left-hand limit. By setting $X'_t(\omega) = 0$ for every $t \in I$ whenever $\omega \in \Omega_0^c$, and using (3.3) on Ω_0, we define a process $(X'_t \colon t \in I)$ that is adapted to $\mathcal{F}^*_t$ (since the augmented filtration contains the null sets). The prescription for X' in 3.3 also ensures that X' is cadlag since any $X_t(\omega)$, $\omega \in \Omega_0$ is right-continuous and has left limits. If this were not the case, so that $|X'_{s_n} - X'_t| > \varepsilon$ for some sequence $s_n \downarrow t$, then by considering a rational sequence $q_n \downarrow t$ having $q_n > s_n$ and $|X'_{s_n} - X_{q_n}| < \varepsilon/2$, we would obtain $|X_{q_n} - X'_t| \geq \varepsilon/2$.

We prove that X' is a martingale with respect to $(\mathcal{F}^*_t)$, and satisfies $\mathbb{E}[X'_t|\mathcal{F}_t] = X_t$ for each $t \in I$. To achieve this, we use results on backwards martingales. Let $(q_n) \in \mathbb{Q}$ be a decreasing sequence, with $q_n \downarrow t$. By definition then, and since X is a martingale, for $r > q_0$,

$$X'_t = \lim_{n\to\infty} X_{q_n} = \lim_{n\to\infty} \mathbb{E}[X_r|\mathcal{F}_{q_n}] \text{ almost surely.}$$

The process $(X^B_n : n \leq 0)$ defined by $X^B_n = \mathbb{E}[X_r|\mathcal{F}_{q_{-n}}]$ is a backwards martingale with respect to the filtration $(\mathcal{F}_{q_{-k}})$ and so converges, by Theorem 2.31, almost surely and in L^1,

$$\lim_{n\to\infty} \mathbb{E}[X_r|\mathcal{F}_{q_{-n}}] = \mathbb{E}[X_r|\mathcal{F}_{t+}].$$

Therefore, the equality

$$X'_t = \mathbb{E}[X_r|\mathcal{F}_{t+}]$$

follows. We note that $\mathcal{F}_t \subset \mathcal{F}_{t+}$, as well as $t < r$, and so, using the tower property of conditional expectation and the martingale property of X, we finally arrive at $\mathbb{E}[X'_t|\mathcal{F}_t] = X_t$, almost surely. A similar argument involving backwards martingales establishes that $\mathbb{E}[X'_t|\mathcal{F}^*_s] = X'_s$, almost surely. □

When working with martingales in continuous time, we shall typically assume that filtrations do satisfy the usual hypotheses, and consider cadlag versions of the processes. The previous theorem shows that this assumption is justifiable.

3.4 Convergence and Doob's inequalities in continuous time

We are now ready to state the continuous-time analogues of Doob's inequalities, and the convergence theorems for martingales. For $t > 0$, we define, as in the discrete-time setting,

$$X^*_t = \sup_{s\leq t} |X_s|.$$

The arguments that establish the validity of these results are similar to the ones used in the proof of the martingale regularization theorem: having the cadlag property at hand allows us to restrict to countable times, where Doob's upcrossings theorem, and the other results in Chapter 2 are available.

Theorem 3.17. [Martingale convergence theorem] *Suppose $(X_t : t \in I)$ is a cadlag martingale, bounded in L^1. Then there exists a random variable $X \in L^1$ such that $X_t \to X_\infty$ almost surely, as $t \to \infty$.*

Theorem 3.18. [Doob's maximal inequality] *Let $(X_t : t \geq 0)$ be a cadlag martingale. Then, for all $\lambda \geq 0$,*

$$\lambda\mathbb{P}(X^*_t \geq \lambda) \leq \mathbb{E}[|X_t|], \quad t > 0.$$

Proof. Let us consider the usual dyadic discretization, with $t > 0$ given: that is, for $n = 1, 2, \ldots$, we set

$$D_n = \{kt2^{-n} : k = 0, 1, \ldots, 2^n\}.$$

We then define martingales $X^{(n)}$ and filtrations $\mathcal{F}^{(n)}$ by setting

$$X_k^{(n)} = X_{\frac{kt}{2^n}}, \quad k = 0, \ldots, 2^n,$$

and, similarly,

$$\mathcal{F}_k^{(n)} = \mathcal{F}_{\frac{kt}{2^n}}.$$

Applying Doob's maximal inequality in discrete time to these martingales, we find that, for $\lambda > 0$,

$$\lambda \mathbb{P}\left(\max_{k \le 2^n} |X_k^{(n)}| > \lambda\right) \le \mathbb{E}[|X_{2^n}^{(n)}|] = \mathbb{E}[|X_t|].$$

We note that the expectation in the right-hand side is independent of n. We next consider the events

$$A_n = \left\{\sup_{s \in D_n} |X_s| > \lambda\right\}, \quad n = 1, 2 \ldots.$$

These events are increasing in n, that is $A_n \subset A_{n+1}$, and since X is cadlag by assumption,

$$\left\{\sup_{s \le t} |X_s| > \lambda\right\} = \bigcup_{n=1}^{\infty} A_n.$$

Hence,

$$\mathbb{P}\left(\sup_{s \le t} |X_s| > \lambda\right) = \lim_{n \to \infty} \mathbb{P}(A_n) \le \frac{1}{\lambda} \mathbb{E}[|X_t|].$$

To get from strict inequality to inequality, we repeat the argument using a sequence with $\lambda_n \uparrow \lambda$. □

Theorem 3.19. [**Doob's L^p-inequality**] *Let $(X_t : t \ge 0)$ be a cadlag martingale. For all $p > 1$,*

$$\|X_t^*\|_p \le \frac{p}{p-1} \|X_t\|_p, \quad t > 0.$$

Theorem 3.20. [**L^p martingale convergence theorem,** $p > 1$] *Let X be a cadlag martingale and $p > 1$, then the following statements are equivalent:*

1. *X is bounded in $L^p(\Omega, \mathcal{F}, \mathbb{P}) : \sup_{t \ge 0} \|X_t\|_p < \infty$*
2. *X converges a.s. and in L^p to a random variable X_∞*
3. *There exists a random variable $Z \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ such that*

$$X_t = \mathbb{E}[Z | \mathcal{F}_t] \text{ a.s.}$$

In the case $p = 1$, we again need to impose the condition of uniform integrability to have convergence in L^1.

The following continuous-time version of the optional stopping theorem will frequently be of use to us.

Theorem 3.21. *Let $X = (X_t \colon t \geq 0)$ be a cadlag uniformly integrable martingale, and let $S \leq T$ be stopping times. Then*

$$\mathbb{E}[X_T|\mathcal{F}_S] = X_S \quad \textit{almost surely.}$$

3.5 Kolmogorov's continuity criterion

Suppose a stochastic process has been constructed, for instance via finite-dimensional distributions and extension. Kolomogorov's continuity criterion provides us with a sufficient condition for a stochastic process in continuous time to admit a version whose sample paths have additional regularity properties.

Kolmogorov's criterion complements the martingale regularization theorem since we are not assuming that the process X be a martingale.

Theorem 3.22. [Kolmogorov's continuity criterion] *Let $X = (X_t \colon t \in [0,T])$ be a stochastic process. Suppose there exist $\alpha > 0, \beta > 0$ such that the condition*

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq C|t-s|^{1+\beta}, \quad s,t \in [0,T],$$

is satisfied with some constant $C < \infty$.

Then, for every $\gamma \in (0, \beta/\alpha)$, there exists a verion of X that is (locally) Hölder continuous with exponent γ. That is, $\mathbb{P}(X_t = \tilde{X}_t) = 1$ for all $t \in [0,T]$, and, for some almost surely positive random variable h,

$$\mathbb{P}\left(\omega \in \Omega \colon \sup_{s,t \in [0,T],\ 0<t-s<h(\omega)} \frac{|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)|}{|t-s|^\gamma} \leq K\right) = 1.$$

Our proof follows [5].

Proof. We take $T = 1$ for notational simplicity; our arguments will rely on dyadic partitions of the unit interval.

Let $\varepsilon > 0$ be given. We begin by invoking Chebyshev's inequality to obtain

$$\mathbb{P}(|X_t - X_s| > \varepsilon) \leq \frac{1}{\varepsilon^\alpha}\mathbb{E}[|X_t - X_s|^\alpha] \leq C\frac{|t-s|^{1+\beta}}{\varepsilon^\alpha}. \tag{3.4}$$

This preliminary estimate shows that $X_s \to X_t$ in probability as $s \to t$.

As a first step towards a stronger statement, we consider dyadic partitions of $[0,1]$:

$$D_n = \{k2^{-n} \colon k = 0, \ldots, 2^n\}, \quad \text{for } n = 1, 2, \ldots\ .$$

The union $D = \bigcup_{n=1}^\infty D_n$ forms a countable dense subset in the unit interval. We pick $\varepsilon = 2^{-\gamma n}$, with $\gamma < \beta/\alpha$. Then, for two adjacent dyadic rationals, we deduce from (3.4) that

$$\mathbb{P}(|X_{k2^{-n}} - X_{(k-1)2^{-n}}| > \varepsilon) \leq C2^{-n(1+\beta-\alpha\gamma)} \tag{3.5}$$

Summing up over 2^n such pairs, we obtain the bound

$$\mathbb{P}\left(\max_{1\le k\le 2^n} |X_{k2^{-n}} - X_{(k-1)2^{-n}}| > \varepsilon\right) \le C2^{-n(\beta-\alpha\gamma)}.$$

Since $\gamma < \beta/\alpha$, the infinite series $\sum_n 2^{-n(\beta-\alpha\gamma)}$ converges. By the Borel-Cantelli lemma then, there exists an event Ω_0 of probability 1, such that for $\omega \in \Omega_0$,

$$\sup_{1\le k\le 2^n} |X_{k2^{-n}}(\omega) - X_{(k-1)2^{-n}}(\omega)| \le 2^{-n\gamma}, \quad n > N(\omega), \tag{3.6}$$

for some integer $N(\omega)$, and thus Hölder continuity holds on level n.

The estimate (3.6) is key; what remains is to match things across dyadic partitions of different levels, and then to define a version of X using the density of the dyadic rationals in the unit interval.

Set $m = n+1$ for $n > N(\omega)$. If $s, t \in D_m$ and $t - s < 1/2^n$, then s, t must be neighbors in D_m. Since each interval determined by D_n contains two intervals determined by D_{n+1}, we have, by (3.6)

$$|X_t(\omega) - X_s(\omega)| < 2 \cdot 2^{-\gamma(n+1)}.$$

Using an inductive argument then, we deduce that for any $m > n$,

$$|X_t(\omega) - X_s(\omega)| \le 2\sum_{k=n+1}^{m} 2^{-\gamma k}, \quad \text{for } s, t \in D_m \text{ with } 0 < t - s < 2^{-n}. \tag{3.7}$$

Namely, suppose (3.7) holds for $m = n+1, \ldots, M-1$ and $0 < t - s < 2^{-n}$. Let $s, t \in D_M$ and consider

$$t_* = \inf\{q \in D_{M-1} \colon q \ge s\} \quad \text{and} \quad t^* = \sup\{q \in D_{M-1} \colon q \le t\}.$$

Then $s \le t_* \le t^* \le t$, and moreover, $t_* - s \le 2^{-M}$, and $t - t^* \le 2^{-M}$. Applying (3.6) to the pairs s, t_* and t^*, t, exploiting the inductive hypothesis for the pair t_*, t^*, and adding up, we arrive at the desired conclusion.

These estimates in fact show that $(X_t(\omega) \colon t \in D)$ is uniformly continuous on the event Ω_0. Namely, let $s, t \in D$ satisfy $0 < t - s < 2^{-N(\omega)}$. We then find $n > N(\omega)$ with $1/2^{n+1} \le t - s \le 1/2^n$, and obtain

$$|X_t(\omega) - X_s(\omega)| \le 2\sum_{k=n+1}^{\infty} 2^{-\gamma k} = 2^{-\gamma(n+1)}\sum_{k=1}^{\infty}(2^{-\gamma})^k \le K2^{-\gamma(n+1)},$$

for $K > 2/(1-2^{-\gamma})$. It now follows that $|X_t(\omega) - X_s(\omega)| \le K|t-s|^\gamma$ for such $s, t \in D$.

We are now in a position to define the desired process $\tilde{X}$. For $\omega \in \Omega_0^c$, we set $\tilde{X}_t(\omega) = 0$ for $0 < t < 1$, and we let

$$\tilde{X}_t(\omega) = X_t(\omega), \quad \text{for } \omega \in \Omega_0,\ t \in D.$$

For non-dyadic times, we pick a sequence $(d_n)_{n=1}^\infty$, with $d_n \in D$, such that $d_n \to t$. Then, by uniform continuity on D, the sequence of numbers $(X_{d_n}(\omega))_{n=1}^\infty$ converges to a limit, and we may define

$$\tilde{X}_t(\omega) = \lim_{d_n \to t} X_{d_n}(\omega) \quad \text{for } \omega \in \Omega_0,\ t \notin D.$$

The process $\tilde{X}$ then inherits the Hölder continuity of X on D.

For $t \in D$, we have $\tilde{X}_t = X_t$ almost surely. Suppose then that $t \notin D$, $0 < t < 1$. By (3.4), we know that $X_{d_n} \to X_t$ in probability when $d_n \to t$ and (d_n) is a dyadic sequence. Since also $\tilde{X}_t = \lim_n X_{d_n}$ almost surely for such a sequence, it then follows that

$$\mathbb{P}(\tilde{X}_t = X_t) = 1$$

for any $t \in [0, 1]$. Hence $\tilde{X}$ is a version of X, and the proof is complete. □

3.6 The Poisson process

In this course, we shall focus on two stochastic processes in continuous time that are in some sense canonical: Brownian motion, to be constructed later, and the Poisson process, which we describe here. Unlike Brownian motion, the Poisson process can be constructed in a rather straight-forward way.

Let $(T_k)_{k=1}^{\infty}$ be a sequence of iid. random variables on some probability space Ω, having exponential distribution with parameter $\lambda > 0$. That is, we let

$$\mathbb{P}(T_1 > t) = e^{-\lambda t}, \quad t \geq 0, \tag{3.8}$$

so that T_1 has a probability density function given by

$$f_\lambda(t) = \lambda e^{-\lambda t} \mathbf{1}(t \geq 0), \quad t \in \mathbb{R}.$$

We often refer to the sequence (T_k) as the *waiting times* of the process we are about to define. Next, we set $S_0 = 0$, and introduce the *jump times*

$$S_n = \sum_{k=1}^{n} T_k, \quad n \geq 1.$$

We define a *Poisson process of intensity* $\lambda > 0$ by taking

$$N_t = \sup\{n \geq 0 \colon S_n \leq t\}, \quad t \geq 0.$$

Clearly, the resulting continuous-time process $N = (N_t \colon t \geq 0)$ is integer-valued. Moreover, its sample paths are cadlag by definition, with jump discontinuities of size 1 occurring at random times S_n.

We endow the space $(\Omega, \mathcal{F}, \mathbb{P})$ upon which N is defined with the natural filtration $(\mathcal{F}_t^N)$ generated by the process. By induction one proves that the jump times S_n are Gamma(n, λ)-distributed:

$$f_{S_n}(s) = \frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s}, \quad s \geq 0.$$

This, together with (3.8) allows us to compute that, for $t > 0$,

$$\begin{aligned}
\mathbb{P}(N_t = n) = \mathbb{P}(S_n \leq t < S_{n+1}) &= \mathbb{P}(\{S_n \leq t\} \cap \{S_n + T_{n+1} > t\}) \\
&= \int_0^t f_{S_n}(s) \mathbb{P}(T_{n+1} > t - s) ds \\
&= \frac{(\lambda t)^n}{n!} e^{-\lambda t}.
\end{aligned}$$

Thus, the random variables N_t have Poisson distributions, and

$$\mathbb{E}[N_t] = \sum_{k=0}^{\infty} k\mathbb{P}(N_t = k) = e^{-\lambda t} \sum_{k=1}^{\infty} \frac{(\lambda t)^k}{(k-1)!} = \lambda t. \tag{3.9}$$

The lack-of-memory property of the exponential distribution used in the definition of the waiting times,

$$\mathbb{P}(T > t + s | T > t) = \mathbb{P}(T > s), \quad s, t > 0,$$

allows us to deduce that the Poisson process enjoys the following property.

Proposition 3.23. *Let $N = (N_t \colon t \geq 0)$ be a Poisson process with intensity $\lambda > 0$. Then for $0 \leq s < t$, $N_t - N_s$ is a random variable, independent of $\mathcal{F}_s^N$, and having Poisson distribution with parameter $\lambda(t - s)$.*

The proposition (or rather, its extension to finite tuples $t_1 < \ldots < t_n$) describes the finite-dimensional distributions of the Poisson process. In particular, the process has *independent and stationary increments.*

It follows from the Proposition and from (3.9) that

$$\mathbb{E}[N_t - N_s] = \lambda(t - s), \quad s < t,$$

and similarily,

$$\mathbb{E}[(N_t - N_s)^2] = \lambda(t - s) + \lambda^2(t - s)^2.$$

Thus $\mathbb{E}[N_t - N_s] \to 0$ as $s \to t$, but Kolmogorov's criterion 3.22 cannot be immediately applied. Fortunately, in this instance, the construction of the process already guarantees that it is cadlag.

Proposition 3.24. [Compensated Poisson process] *Let N be a Poisson process of intensity $\lambda > 0$. Then the process $(M_t \colon t \geq 0)$ defined by*

$$M_t = N_t - \lambda t, \quad t \geq 0,$$

is a martingale with respect to $(\mathcal{F}_t^N)$.

Proof. That N is adapted and integrable follows immediately.

For $s < t$, the independence guaranteed by Proposition 3.23, together with (3.9), yields that

$$\mathbb{E}[N_t - \lambda t | \mathcal{F}_s] = \mathbb{E}[N_t - N_s - \lambda(t - s) + N_s - \lambda s | \mathcal{F}_s] = N_s - \lambda s,$$

and thus M is a martingale. □

A similar computation shows that N is a submartingale. Representing the Poisson process N using the compensated process,

$$N_t = \lambda t + M_t,$$

we recover a special instance of the *Doob-Meyer decomposition* (cf. Example Sheet 2) of a submartingale: $t \mapsto \lambda t$ is an increasing previsible process, and M is a martingale. Compensated Poisson processes will play a prominent role in our analysis of Lévy processes and their jumps.

4 Weak convergence

4.1 Basic definitions

We have dealt with two different modes of convergence for random objects so far:almost sure convergence, and convergence in probability. These concepts are defined for random variables supported on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In some cases, however, it is natural to consider sequences of random variables that map into the same space, for instance $\mathbb{R}$ or $\mathbb{R}^n$, but whose underlying probability spaces need not coincide. In this chapter, we introduce a convergence concept that applies in this latter setting, and is defined in terms of the laws of such random variables. At the same time, we take the opportunity to extend some of our formalism to the framework of general metric spaces.

We shall discuss the concept of *tightness*. This is a notion that aims at preventing the phenomenon of "escape of mass to infinity." Finally, we review the notion of characteristic function (or Fourier transform) of a random variable, or rather, its law, and we show how weak convergence is related to pointwise convergence of characteristic functions.

In what follows, (M, d) denotes a metric space. Given a metric, we can make sense of the notion of open sets in M, and we can endow M with its Borel σ-algebra $\mathcal{B}(M)$, the smallest σ-algebra containing the open sets. Unless otherwise stated, all measures under consideration in this chapter will be defined on this type of measurable space. We begin with some topological considerations. Let A^0 denote the interior of a Borel set $A \subset M$, and $\bar{A}$ its closure with respect to d. The *boundary* of a set is

$$\partial A = \bar{A} \setminus A^0.$$

The distance between a point $p \in M$ and $A \subset M$ is defined as

$$d(p, A) = \inf\{d(p, q) \colon q \in A\}.$$

We will be concerned with continuous functions $f \colon M \to \mathbb{R}$, and the space $C_b(M)$ consisting of *bounded continuous functions* will play a central role. In this context, continuity properties of a function depend on the choice of metric d. There is a natural pairing between a probability measure μ on $(M, d, \mathcal{B})$ and a function $f \in C_b(M)$ that produces a mapping $\mu \colon C_b(M) \to \mathbb{R}$. This pairing is obtained through the operation of integration,

$$\mu(f) = \int_M f d\mu.$$

The concept of weak convergence is defined in terms of this duality.

Definition 4.1. Let $(\mu_n) = (\mu_n \colon n \geq 0)$ be a sequence of probability measures defined on a metric space (M, d) endowed with its Borel σ-algebra $\mathcal{B}$. We say that (μ_n) *converges weakly* to a Borel measure μ as $n \to \infty$, and we write

$$\mu_n \Rightarrow \mu,$$

if

$$\mu_n(f) \to \mu(f) \quad \text{for all } f \in C_b(M).$$

Any weak limit measure μ is a probability measure since $\mu(1) = \lim_{n\to\infty} \mu_n(1) = 1$.

Here are some simple examples and non-examples of weakly convergent sequences of measures.

Example 4.2. Consider a sequence of points $(p_n)_{n=0}^{\infty}$ in M with $d(p_n, p) \to 0$, for some $p \in M$. Letting δ_p denote a point mass at a point $p \in M$, we consider the sequence of probability measures $(\mu_n) = (\delta_{p_n})$.

Then (μ_n) is weakly convergent, with limit δ_p, since $f(p_n) \to f(p)$ as $n \to \infty$ for any continuous function f.

Example 4.3. Let $M = \mathbb{R}$, and consider point masses $(\mu_n\colon \geq 1) = (\delta_n)$ charging the natural numbers in increasing order. The sequence of measures (μ_n) is not weakly convergent.

Remark 4.4. From the point of view of functional analysis, the concept we are discussing here is essentially that of weak-$*$ convergence: by the Riesz representation theorem, the space of (signed) measures with bounded total variation can be identified with the dual space of (compactly supported) continuous functions.

However, the terminology "weak convergence" remains standard in probability theory (despite some attempts to introduce a different name), and we shall stick to that convention in these notes.

In the next definitions, we specialize to $M = \mathbb{R}^d$, equipped with the Euclidean metric induced by the inner product [5]

$$\langle x, y\rangle = x_1 y_1 + \cdots + x_d y_d, \quad x, y \in \mathbb{R}^d.$$

Thus, for $x \in \mathbb{R}^d$, we have $\|x\| = \langle x, x\rangle^{1/2}$.

Definition 4.5. Let X be a random variable taking values in $\mathbb{R}^d$, and let $\mu = \mu_X$ denote the law of X. The *characteristic function* $\varphi = \varphi_X$ of X is defined as

$$\varphi_X(t) = \mathbb{E}[e^{i\langle t, X\rangle}] = \int_{\mathbb{R}^d} e^{i\langle t, x\rangle} d\mu_X(x), \quad t \in \mathbb{R}^d. \tag{4.1}$$

Since $|e^{i\xi}| = 1$, the characteristic function is well defined as a function on $\mathbb{R}^d$. Moreover, φ is continuous, and $\varphi(0) = 1$. We shall make use of the following fact without proof: if two characteristic functions satisfy

$$\varphi_X(t) = \varphi_Y(t), \quad \text{for all } t \in \mathbb{R}^d,$$

then $X = Y$ in law.

Remark 4.6. This is another example of how probability theory and harmonic analysis differ in terminology. An analyst would probably refer to what we have dubbed characteristic function as the *Fourier-Stieltjes transform* of the measure μ_X, and would most likely include a minus sign, and possibly a normalizing constant, in the exponent.[6]

[5] We use d for dimension here, and reserve n for indices of sequences.

[6] This may be mildly annoying when comparing formulas coming from different references.

Example 4.7. The characteristic function of a standard $\mathcal{N}(0,1)$-distributed random variable on $\mathbb{R}$ is

$$\varphi_X(t) = e^{-\frac{t^2}{2}}.$$

The characteristic function of a random variable having uniform distribution in $[-a,a]$ is

$$\varphi_X(t) = \frac{\sin at}{at}.$$

These examples illustrate the fact that if φ_X is real, then X and $-X$ have the same distribution.

We will be able to reformulate weak convergence in terms of distribution functions, a concept that should be familiar from basic probability theory. The *distribution function* of a probability measure on the real line is defined as

$$F_\mu(x) = \mu((-\infty, x]), \quad x \in \mathbb{R}.$$

Recall that if X is a real-valued random variable, then

$$F_X(x) = \mathbb{P}(X \leq x),$$

and that if X has *probability density function* f_X, then

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy,$$

and $F_X'(x) = f_X(x)$ at continuity points of f. Distribution functions are non-decreasing in x and right-continuous, and

$$\lim_{x\to-\infty} F_\mu(x) = 0 \quad \text{and} \quad \lim_{n\to\infty} F_\mu(x) = 1.$$

We hinted at the usefulness of weak convergence when dealing with random variables defined on different probability spaces. Let us make this more precise; in this definition we return to the full generality of metric spaces.

Definition 4.8. Let $(X_n) = (X_n \colon n \geq 0)$ be a sequence of random variables, defined on some collection of probability spaces $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, and taking values in a metric space (M,d). We say that (X_n) converges *in distribution* to a random variable X, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, if the sequence of associated laws (μ_{X_n}) converges weakly to the law of X as $n \to \infty$.

Phrased differently, the random variables X_n converge in distribution to X if, for all $f \in C_b(M)$,

$$\mathbb{E}_{\mathbb{P}_n}[f(X_n)] \to \mathbb{E}_{\mathbb{P}}[f(X)].$$

Here, $\mathbb{E}_{\mathbb{Q}}[\cdot] = \int_\Omega (\cdot) d\mathbb{Q}$ denotes expectation with respect to the probability measure $\mathbb{Q}$.

We shall sometimes use $X_n \Rightarrow X$ as shorthand notation for convergence in distribution of a sequence of random variables; to be precise, what is meant by this is $\mu_{X_n} \Rightarrow \mu_X$.

4.2 Characterizations of weak convergence

We might ask whether weak convergence of (μ_n) to μ is equivalent to having $\mu_n(A) \to \mu(A)$ for all Borel sets A, but in general, this is not the case. Consider the sequence of point masses $(\delta_{1/n})$: we verify that $\delta_{1/n} \Rightarrow \delta_0$, but $\delta_{1/n}((0,1)) = 1$ for all n, and $\delta_0((0,1)) = 0$.

However, the following is true.

Theorem 4.9. *Suppose (μ_n) is a sequence of probability measures on a metric space (M, d) endowed with its Borel σ-algebra. The following statements are equivalent:*

1. $\mu_n \Rightarrow \mu$ *as* $n \to \infty$
2. $\liminf_{n\to\infty} \mu_n(U) \geq \mu(U)$, *for all open sets* U
3. $\limsup_{n\to\infty} \mu_n(V) \leq \mu(V)$, *for all closed sets* V
4. $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for all sets with* $\mu(\partial A) = 0$

Proof. We begin with the implication *(1.)* $\implies$ *(2.)*. For $U = \Omega$, the inequality is trivial. Suppose $U \neq \Omega$, and consider the functions

$$f_m(x) = 1 \wedge m d(x, U^c), \quad x \in M, \ m = 1, 2, \ldots.$$

Note that f_m is continuous and bounded, and that $f_m(x) \leq \mathbf{1}(x \in U)$ for all m. As U is open, the set U^c is closed, and so $f_m(x) \to \mathbf{1}(U)$ as $m \to \infty$. In view of this, and the assumption of weak convergence,

$$\liminf_{n\to\infty} \mu_n(U) \geq \liminf_{n\to\infty} \mu_n(f_m) = \mu(f_m),$$

and an application of the monotone convergence theorem to (f_m) leads

$$\liminf_{n\to\infty} \mu_n(U) \geq \mu(U),$$

the desired inequality.

The equivalence of the second and third items follows from the fact that U is open if and only if $V = U^c$ is closed, and $\mathbb{P}(U) + \mathbb{P}(U^c) = 1$.

Next, we show that *(2.)* and *(3.)* imply *(4.)*. Suppose the Borel set A satisfies $\mu(\partial A) = 0$. Then

$$\mu(\partial A) = \mu(\bar{A} \setminus A^0) = 0,$$

and $\mu(A^0) = \mu(\bar{A})$, both numbers being equal to $\mu(A)$. Finally,

$$\limsup_{n\to\infty} \mu_n(\bar{A}) \leq \mu(\bar{A}) = \mu(A),$$

and

$$\liminf_{n\to\infty} \mu_n(A^0) \geq \mu(A^0) = \mu(A),$$

and the assertion follows since $A^0 \subset \bar{A}$.

We finish by proving that *(4.)* implies *(1.)*. Suppose $f \in C_b(M)$ is non-negative with $\|f\|_\infty < C$. Then, by Fubini's theorem,

$$\int_M f(x)d\mu_n(x) = \int_M \left(\int_0^\infty \mathbf{1}(f(x) \geq \lambda)d\lambda \right) d\mu_n(x) = \int_0^C \mu_n(\{f \geq \lambda\})d\lambda. \tag{4.2}$$

Now, by the continuity of f, the set $\{f \geq \lambda\}$ is closed, and $\{f > \lambda\}$ is open and contained in its interior. Thus $\partial\{f \geq \lambda\}$ is contained in $\{f = \lambda\}$. Since μ is a probability measure, we can only have $\mu(\{f = \lambda\}) > 0$ for a countable set of $\lambda \in \Lambda \subset \mathbb{R}$. In particular, it follows that

$$\int_\Lambda \mu(\{f = \lambda\})d\lambda = 0,$$

and thus, in the integral on the right-hand side in (4.2), it suffices to restrict our attention to $\lambda \in [0, C]$ such that $\mu(\partial\{f \geq \lambda\}) = 0$. Using *(3.)* and applying the dominated convergence theorem to the resulting integrals, we deduce

$$\mu_n(f) \to \mu(f) \quad \text{as } n \to \infty.$$

To obtain the result for general $f \in C_b(M)$, we perform the usual approximation and splitting into positive and negative parts. □

We next connect weak convergence with distribution functions.

Theorem 4.10. *For a sequence (μ_n) of probability measures on $\mathbb{R}$, the following are equivalent.*

As $n \to \infty$,

1. *(μ_n) converges weakly to μ*
2. *the distribution functions $F_{\mu_n}(x)$ converge to $F_\mu(x)$ at every point of continuity of F_μ*

Proof. We begin by proving that *(1.)* implies *(2.)*.

Let $x \in \mathbb{R}$ be a point of continuity of F_μ. By Theorem 4.9, we can deduce

$$F_{\mu_n}(x) = \mu_n((-\infty, x]) \to \mu((-\infty, x]) = F_\mu(x),$$

provided $\mu(\partial(-\infty, x]) = \mu(\{x\}) = 0$. But this follows upon writing

$$\mu(\{x\}) = \mu((-\infty, x]) - \lim_{n\to\infty} \mu\left(\left(-\infty, x - \frac{1}{n}\right]\right),$$

expressing things in terms of the distribution function, and exploiting the continuity of F_μ at x.

Next is the implication *(2.)* $\implies$ *(1.)*. Let U be an arbitrary open set in $\mathbb{R}$. The open intervals form a (countable) base for the topology of $\mathbb{R}$, and so we may write

$$U = \bigcup_{k=1}^\infty (a_k, b_k),$$

using a countable collection of disjoint intervals (a_k, b_k). Since distribution functions are right-continuous and non-decreasing, their points of discontinuity form countable sets. This means that the continuity points of F_{μ_n} are dense in (a_k, b_k), and so there are plenty of such points $a'_k, b'_k \in (a_k, b_k)$ with $a'_k < b'_k$. On each interval then,

$$\mu_n((a_k, b_k)) = F_{\mu_n}(b_k-) - F_{\mu_n}(a_k) \geq F_{\mu_n}(b'_k) - F_{\mu_n}(a'_k) = \mu_n((a'_k, b'_k)),$$

and so

$$\liminf_{n\to\infty} \mu_n((a, b)) \geq \mu((a'_k, b'_k)),$$

and then $\liminf_{n\to\infty} \mu_n((a, b)) \geq \mu((a, b))$ persists in the limit as $a'_k \downarrow a$ $b'_k \uparrow b_k$.

Now $\liminf_n \mu_n(U) = \liminf_n \sum_k \mu_n((a_k, b_k))$, and after applying Fatou's lemma and the previously established inequality,

$$\liminf_{n\to\infty} \sum_{k=1}^{\infty} \mu_n((a_k, b_k)) \geq \sum_{k=1}^{\infty} \liminf_{n\to\infty} \mu_n((a_k, b_k)) \geq \sum_{k=1}^{\infty} \mu((a_k, b_k)).$$

The last expression is equal to $\mu(U)$, and we are done. □

The relation between different modes of convergence is addressed in the following proposition.

Proposition 4.11. *Let (X_n) be a sequence of random variables, and let X be another random variable defined on the same space. If (X_n) converges to X in probability, then (X_n) converges to X in distribution.*

Suppose (X_n) converges in distribution to the constant $c \in \mathbb{R}$. Then (X_n) converges in probability to the degenerate random variable c.

Proof. See Example Sheet 3. □

Let $C_c(M)$ denote the space of continuous functions on M having compact support. It is immediate that $C_c(M) \subset C_b(M)$. For completeness, we mention the following mode of convergence which is sometimes more convenient.

Definition 4.12. Let (μ_n) be a sequence of bounded measures on $(M, d, \mathcal{B})$. We say that (μ_n) converges *vaguely* to a measure μ as $n \to \infty$ if

$$\mu_n(f) \to \mu(f) \quad \text{for all } f \in C_c(M).$$

4.3 Tightness

We return to the problem of sequences of probability measures where mass escapes to infinity. At first glance, the following theorem seems to imply that distributions functions automatically enjoy certain compactness properties. We do not give a proof here.

Theorem 4.13. [**Helly's selection theorem**] *For every sequence (F_n) of distribution functions, there exists a subsequence (F_{n_k}) and a right-continuous and non-decreasing function F such that $\lim_{k\to\infty} F_{n_k}(x) = F(x)$ at every point of continuity of F.*

Sometimes, this result is good enough. The problem in general is that the limit function in the theorem need not be a distribution function.

Example 4.14. Let $a, b, c > 0$ be real constants, with $a + b + c = 1$, and set G be any distribution function. For $n = 1, 2, \ldots$, consider the distribution functions

$$F_n(x) = a\mathbf{1}(x \geq n) + b\mathbf{1}(x \geq -n) + cG(x), \quad x \in \mathbb{R}.$$

As $n \to \infty$, we have $F_n(x) \to b + cG(x)$, and

$$\lim_{x\to-\infty} F(x) = b \quad \text{and} \quad \lim_{x\to\infty} F(x) = 1 - a.$$

Having seen what can go wrong, we define a notion that counteracts the escape of mass to infinity.

Definition 4.15. A sequence (μ_n) of probability measures on a metric space (M, d) is said to be *tight* if, for every $\varepsilon > 0$, there exists a compact set $K \subset M$ such that

$$\sup_n \mu_n(M \setminus K) \leq \varepsilon.$$

If the metric space under consideration is itself compact, then every sequence of probability measures is tight.

The fundamental result concerning tightness is the following.

Theorem 4.16. [**Prokhorov's theorem**] *Let (μ_n) be a tight sequence of probability measures on a metric space (M, d). Then there exists a subsequence (n_k) such that*

$$\mu_{n_k} \Rightarrow \mu,$$

where μ is a probability measure on (M, d).

Proof. We give the proof in the easiest case: $M = \mathbb{R}$, equipped with the Euclidean metric.

We argue on the distribution functions F_{μ_n}, and the values they take on $\mathbb{Q}$. Considering successive subsequences of (F_{μ_n}), and using a diagonal argument on these subsequences, we extract a non-decreasing limit function $F\colon \mathbb{Q} \to [0, 1]$ with $F_{\mu_{n_k}}(r) \to F(r)$ as $k \to \infty$, for every $r \in \mathbb{Q}$. This is achieved by considering an enumeration the rationals $\{q_j\}$, and then noting that for q_j fixed, the sequence $(F_{\mu_{n_k}}(q_j))$ is contained in the unit interval. Since this is a compact set, the Bolzano-Weierstrass theorem guarantees the existence of a convergent subsequence.

We may then extend F to a cadlag function on all of $\mathbb{R}$ in the usual manner, by setting $F(x) =\downarrow_{r\in\mathbb{Q}} F(r)$. Then, by monotonicity, $\lim_{k\to\infty} F_{\mu_{n_k}}(x) = F(x)$ at all points of continuity of F.

We claim that the function F thus constructed is the distribution function of some probability measure μ on $\mathbb{R}$. Since (μ_n) is a tight collection, there exists a $K > 0$ such that $\mu_n([-K, K]) \geq 1 - \varepsilon$ for all n. In terms of distribution functions, we then have, for all n,

$$F_{\mu_n}(-K) \leq \varepsilon \quad \text{and} \quad F_{\mu_n}(K) \geq 1 - \varepsilon.$$

By making sure to choose K so that K and $-K$ are points of continuity of F, we obtain $F(-K) \leq \varepsilon$ and $F(K) \geq 1-\varepsilon$ in the limit as $K \to \infty$. As continuity points are dense in $\mathbb{R}$, we may conclude that

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} F(x) = 1.$$

It is now a straight-forward matter to associate a probability measure with F. We define a set function μ on intervals of the form $(a, b]$, with $a < b$, by assigning

$$\mu((a, b]) = F(b) - F(a).$$

Since the collection of such intervals generate the Borel σ-algebra $\mathcal{B}(\mathbb{R})$, we can extend μ to a Borel measure that inherits $\mu(\mathbb{R}) = 1$ from the corresponding property of F. This completes the proof. □

In practice, when dealing with convergence in distribution of random variables or stochastic processes, tightness is often the first property one checks. Prokhorov's theorem then guarantees the existence of subsequential limits at least. Proving that a common limit actually exists can be more challenging.

4.4 Characteristic functions and Lévy's theorem

Characteristic functions often provide a convenient tool for establishing convergence in distribution of random variables. Before stating and proving a result in this direction, due to P. Lévy, we establish a lemma that relates the decay of measures at infinity to the smoothness at 0 of the associated characteristic functions.

Lemma 4.17. *Let X be a random variable in $\mathbb{R}^d$. Then, there exists a constant $C = C(d) > 0$ such that, for all $K > 0$,*

$$\mathbb{P}(\|X\| > K) \leq CK^d \int_{[-1/K,1/K]^d} (1 - \mathrm{Re}\varphi_X(t))dt.$$

Proof. We give a proof in the case $d = 1$; the general case can be dealt with in an analogous manner, but the notation becomes more cumbersome.

As a first step, one verifies using calculus that $f(x) = \sin x/x$ has

$$\left|\frac{\sin x}{x}\right| \leq \sin 1, \quad x > 1,$$

so that

$$\mathbf{1}(|u| \geq 1) \leq C(1 - f(u))$$

with a suitable constant (say $C = 1/(1-\sin 1)$). Taking expectations on both sides of the previous estimate, we obtain

$$\begin{aligned}\mathbb{P}(|X| \geq K) = \mathbb{P}\left(\left|\frac{X}{K}\right| \geq 1\right) &\leq C\mathbb{E}\left[1 - \frac{\sin(X/K)}{(X/K)}\right] \\ &= C\int_{\mathbb{R}} \left(1 - \frac{\sin(x/K)}{(x/K)}\right) d\mu(x).\end{aligned}$$

The rest of the proof reduces to computations using Fubini's theorem. Namely, we have

$$\begin{aligned}\int_{\mathbb{R}} \frac{\sin(\lambda x)}{\lambda x} d\mu(x) &= \int_{\mathbb{R}} \left(\frac{e^{i\lambda x} - e^{-i\lambda x}}{2i\lambda x} \right) d\mu(x) \\ &= \frac{1}{2\lambda} \mathrm{Re} \int_{\mathbb{R}} \left(\int_{-\lambda}^{\lambda} e^{ixu} du \right) d\mu(x) \\ &= \frac{1}{2\lambda} \int_{-\lambda}^{\lambda} \left(\mathrm{Re} \int_{\mathbb{R}} e^{iux} d\mu(x) \right) du = \frac{1}{2\lambda} \int_{-\lambda}^{\lambda} \mathrm{Re}\varphi_X(u) du.\end{aligned}$$

Setting $\lambda = 1/K$, the desired statement now follows from the previous estimate.

□

Theorem 4.18. [**Lévy's convergence theorem**] *Suppose $\mu_{X_n} \Rightarrow \mu_X$ as $n \to \infty$. Then $\varphi_{X_n}(t) \to \varphi_X(t)$ as $n \to \infty$, for all $t \in \mathbb{R}^d$.*

On the other hand, suppose that for a sequence (X_n) of random variables, it holds that $\varphi_{X_n}(t) \to \psi(t)$ for every $t \in \mathbb{R}^d$ for some function $\psi \colon \mathbb{R}^d \to \mathbb{C}$, continuous at the origin.

Then ψ is the characteristic function of some random variable X, and moreover $\mu_{X_n} \Rightarrow \mu_X$ as $n \to \infty$.

Proof. The first implication is the easier one. If X_n converges in distribution to X as $n \to \infty$, then for every f continuous and bounded,

$$\mu_n(f) \to \mu(f) \quad \text{as } n \to \infty.$$

For $t \in \mathbb{R}^d$ fixed, pick $f_t(x) = e^{i\langle t,x\rangle} = \cos\langle t,x\rangle + i\sin\langle t,x\rangle$. The real and imaginary parts of the complex function f_t are continuous and bounded, and hence, by linearity,

$$\varphi_{X_n}(t) = \mu_n(e^{i\langle t,\cdot\rangle}) \to \mu(e^{i\langle t,\cdot\rangle}) = \varphi_X(t).$$

As $t \in \mathbb{R}^d$ was arbitrary, the result follows.

We turn to the second part of the theorem, and begin by showing that (μ_{X_n}) form a tight sequence. Lemma 4.17 yields that that, for $K > 0$ and all n,

$$\mathbb{P}(\|X_n\| > K) \leq CK^d \int_{[-1/K,1/K]^d} (1 - \mathrm{Re}\varphi_{X_n}(t))dt.$$

Noting that $|1 - \mathrm{Re}\varphi_{X_n}(t)| \leq 2$ for all n, we invoke the dominated convergence theorem to obtain

$$\lim_{n\to\infty} \int_{[-1/K,1/K]^d} (1 - \mathrm{Re}\varphi_{X_n}(t))dt = \int_{[-1/K,1/K]} (1 - \mathrm{Re}\psi(t))dt.$$

We recall that ψ was assumed to be continuous at 0, and hence, given $\varepsilon > 0$, we may pick K so large that

$$K^d \int_{[-1/K,1/K]^d} (1 - \mathrm{Re}\psi(t))dt \leq \frac{\varepsilon}{2C},$$

meaning that there exists an $N > 0$ with

$$\mathbb{P}(\|X_m\| > K) \leq \varepsilon, \quad \text{for } m \geq N.$$

After possibly increasing K to accomodate for a finite number of variables X_m having $m \leq N$, we can make this inequality valid all n, thus establishing the tightness of (μ_{X_n}).

Then, by Prokhorov's theorem, there exists a subsequence (X_{n_k}) that converges in distribution to a probability measure on $\mathbb{R}^d$, which we identify with the law of a random variable. But since (φ_{X_n}) converges pointwise to φ_X, we obtain $\psi = \varphi_X$, showing that ψ is a characteristic function.

The proof will be complete once we have shown that (X_n) converges in distribution to X. Suppose this fails. Then there exists a function $f \in C_b$, and a subsequence (m_k), such that

$$|\mathbb{E}[f(X_{m_k})] - \mathbb{E}[f(X)]| > \varepsilon$$

for some positive ε. Since the laws $(\mu_{X_{m_k}})$ form a tight sequence, we may extract another subsequence that converges in distribution. But since $\psi = \varphi_Y$ for any such limiting random variable Y, this leads to a contradiction. □

In particular, Lévy's theorem implies the following equivalence for a sequence of random variables on $\mathbb{R}^d$:

$$\mu_{X_n} \Rightarrow \mu_X \quad \text{if and only if} \quad \varphi_{X_n}(t) \to \varphi_X(t), \quad \text{for all } t \in \mathbb{R}^d. \tag{4.3}$$

5 Brownian motion

5.1 Basic definitions

In Chapter 3, we discussed stochastic processes in continuous time, and at the end of that chapter, we encountered our first concrete examples of such processes, namely, *Poisson processes of intensity* $\lambda > 0$. These are stochastic processes in continuous time that take values in $\mathbb{N}$, and have cadlag sample paths. Poisson processes are in many ways canonical, and are important in many applications we cannot discuss here. We shall, however, encounter variants of Poisson processes in the last chapter of these notes.

Our construction of Poisson processes was based on independent, exponentially distributed waiting times, and derived jump times, but we were later able to verify that a Poisson process N defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ enjoys the following key properties:

- $N(0) = 0$,
- for any sequence of times $0 < t_1 < t_2 < \cdots < t_n$, the random variables

$$N_{t_k} - N_{t_{k-1}}, \quad 1 \leq k \leq n,$$

  are independent,
- and $N_t - N_s$, for $t > s$, has $\mathrm{Po}(\lambda(t-s))$-distribution.

In terms of formulas, the latter property reads

$$\mathbb{P}(N_t - N_s = k) = \frac{\lambda^k (t-s)^k e^{-\lambda}}{k!}, \quad k = 1, 2, \ldots.$$

The pleasant features of the Poisson process ultimately follow from the properties of the exponential distribution, a distribution distinguished by its memorylessness.

The Central Limit Theorem features another canonical distribution on $\mathbb{R}$: the *normal* or *Gaussian distribution*. In view of this, and the other remarkable features of this distribution, it seems natural from the mathematical perspective to study stochastic processes related to normal random variables, as natural models of "random fluctuations."

Moreover, Gaussian processes can be used to model phenomena in nature that exhibit randomness such as the physical Brownian motion observed by R. Brown and later analyzed by A. Einstein. Finally, in his pioneering work "Théorie de la Spéculation", L. Bachelier indicated how such processes appear in the analysis of financial markets.

We set down some notation. In what follows, we shall frequently work on the Euclidean space $\mathbb{R}^d$; we let I_d denote the d-dimensional identity matrix having 1's on the diagonal, and 0 in all other entries. The set of all *orthogonal matrices* will be denoted by $O(d)$; this consists of the $d \times d$ matrices having transpose equal to the inverse:

$$M^T M = I_d.$$

Multiplication of a vector in $\mathbb{R}^d$ by a $d \times d$ matrix will be indicated by juxtaposition. As before, we let $\mathcal{N}(\mu, \sigma)$ be the *normal distribution* on $\mathbb{R}^d$ with mean $\mu \in \mathbb{R}^d$, and $d \times d$ covariance matrix

$$\sigma = (\sigma_{jk})_{1 \le j,k \le d}.$$

In the case $d = 1$, its probability density function is given by the classical Gauss kernel

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma}}. \tag{5.1}$$

In the general case, we have

$$f_{0,\sigma I_d}(x) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|x\|^2}{2\sigma}}$$

We are now ready to define the process we propose to study in this chapter.

Definition 5.1. Let $B = (B_t \colon t \ge 0)$ be a continuous $\mathbb{R}^d$-valued stochastic process defined on some probability space. We say that B is a *standard Brownian motion* or *Wiener process* on $\mathbb{R}^d$ if

- $B_0 = 0$,
- for any sequence of times $0 < t_1 < t_2 < \cdots < t_n$, the random variables

  $$B_{t_k} - B_{t_{k-1}}, \quad 1 \le k \le n,$$

  are independent,
- and $B_t - B_s$, for $t > s$, is $\mathcal{N}(0, (t-s)I_d)$-distributed.

We shall also work with Brownian motions started at a point $x \in \mathbb{R}^d$, that is, with $B_0 = x$ (sometimes almost surely). One way to obtain such a process is to take $B_t = x + W_t$, where W is a standard Brownian motion. We shall frequently write $\mathbb{P}_x$ for the law of a Brownian motion started at $x \in \mathbb{R}^d$, and we let $\mathbb{E}_x[\cdot]$ denote the corresponding expectated values. The subscript will typically be omitted when B is a standard Brownian motion.

We recognize two properties of the Poisson process in the above definition: Brownian motion has *independent and stationary increments*. Note that we are including continuity of the process in its definition. The second and third items above specify the finite-dimensional distributions of the process, and so the law of a Brownian motion is uniquely determined.

It is not at all clear a priori that an object satisfying these (natural) requirements actually exists. Fortunately, N. Wiener was able to present a rigorous construction of Brownian motion in the 1920's–hence its alternative name. We shall return to the construction of Brownian motion in the next section.

For the time being, we assume existence, and begin by listing some useful properties that follow immediately from its definition in terms of the normal distribution.

Proposition 5.2. **[Rotational invariance, scaling, and time inversion]**
Let $B = (B_t : t \geq 0)$ be a standard Brownian motion.

For any orthogonal matrix $M \in O(d)$, the process $MB = (MB_t : t \geq 0)$ is a standard Brownian motion.

Moreover, for any $c > 0$, the process $W = (c^{-1/2} B_{ct} : t \geq 0)$ is a standard Brownian motion.

Finally, the process $Z = (Z_t : t \geq 0)$ defined by $Z_0 = 0$ and

$$Z_t = tB_{1/t}, \quad t > 0,$$

is a standard Brownian motion.

In particular, the process $-B$ is a Brownian motion. In two dimensions, this result can be generalized to say that Brownian motion is *conformally invariant*, that is, the image of a Brownian motion under a conformal map $f : U \subset \mathbb{C} \to \mathbb{C}$ (which can be viewed infinitesimally as a dilation and rotation) is again a Brownian motion, after a suitable time-change.

Proof. The first two assertions can be established by appealing directly to the multidimensional normal distribution. Computations show that the finite-dimensional distributions of the processes MB and W agree with those of B, and since continuity is satisfied as well, the assertions follow by an appeal to the monotone class theorem.

We now establish the third claim. We know (see Example Sheet 3) that Gaussian processes are determined by their means and their covariances. We begin by verifying that the process Z has the same covariance structure as B; it is immediate from the definition that $\mathbb{E}[Z_t] = 0$ for all t. Now, for $s < t$, we have

$$\mathrm{cov}(Z_s, Z_t) = \mathrm{cov}(sB_{1/s}, tB_{1/t}) = st\ \mathrm{cov}(B_{1/s}, B_{1/t}) = st\frac{1}{t} = s,$$

and so the claim regarding covariances follows. This means that the finite-dimensional distributions of B and Z agree.

Almost sure pathwise continuity of Z follows from that of B for any $t > 0$, since $Z_t(\omega)$ is obtained as the composition of continuous functions. At 0, continuity follows once we exploit the continuity of Z_t for $t > 0$ to represent the event $\{\lim_{t\to 0} Z_t = 0\}$ as

$$\bigcap_n \bigcup_m \bigcap_{q \in \mathbb{Q}\cap(0,1/m]} \left\{ |Z_q| \leq \frac{1}{n} \right\},$$

and use that Z_t and B_t have the same distribution for $t > 0$ to deduce that this countably written event has probability 1. □

An consequence of the last item is a law of large numbers for Brownian motion:

$$\lim_{t\to\infty} \frac{B_t}{t} = \lim_{t\to\infty} X_{1/t} = 0 \quad \text{almost surely.}$$

We shall later obtain much more precise statements concerning the asymptotics of B.

The following result also follows from the definition of Brownian motion, in particular, from independence of increments.

Proposition 5.3. [**Markov property**]
For any $s \geq 0$, the process $(B_{t+s} - B_s \colon t \geq 0)$ is a standard Brownian motion, independent of $\mathcal{F}_s^B$.

Roughly speaking, the Markov property, which is enjoyed by Brownian motion and many other important stochastic processes. Roughly speaking, says that "the future and past of the process are independent, given the present."

Finally, we introduce a number of filtrations of $(\Omega, \mathcal{F}, \mathbb{P})$ we shall need in our study of Brownian motion. As usual, we let $(\mathcal{F}_t^B)$ denote the natural filtration associated with B, that is

$$\mathcal{F}_t^B = \sigma(B_s \colon s \leq t), \quad t > 0.$$

The right-continuous filtration $(\mathcal{F}_t^{B+})$ is defined by

$$\mathcal{F}_t^{B+} = \bigcap_{u>t} \mathcal{F}_u^B.$$

Sometimes it is convenient to work with bigger filtration. We say that B is an $(\mathcal{F}_t)$-Brownian motion provided $\mathcal{F}_t^B \subset \mathcal{F}_t$, so that B is adapted to $(\mathcal{F}_t)$, and $\sigma(X_t - X_s \colon t \geq s)$ and $\mathcal{F}_s$ are independent, for all $s \geq 0$.

Finally, we let $\mathcal{B}((\mathbb{R}^d)^{[0,\infty)})$ denote the smallest σ-algebra on the space of all functions $f \colon [0,\infty) \to \mathbb{R}^d$ that contains cylinder sets. These are sets of the form

$$B = \{\omega \in (\mathbb{R}^d)^{[0,\infty)} \colon (\omega(t_1), \ldots, \omega(t_n)) \in A\},$$

where $\bar{t} = (t_1, \ldots, t_n)$ is an arbitrary finite tuple with $t_k \in [0,\infty)$, and $A \in \mathcal{B}(\mathbb{R}^{nd})$.

5.2 Construction of Brownian motion

We now address the question of existence of Brownian motion on some probability space. This problem was first solved by N. Wiener in 1923.

Several different constructions of Brownian are possible, and one option is to use some of the techniques developed in Chapter 3. To ease notation, we indicate this construction in the case $d = 1$. We begin by stating a result alluded to in that chapter. Suppose that, for any sequence of non-negative numbers

$$\bar{t} = (t_1, t_2, \ldots, t_n) \in \mathbb{R}^n,$$

we are given a Borel probability measure $P_{\bar{t}}$. Then the family of finite-dimensional distributions $\{P_{\bar{t}}\}_{\bar{t} \in T}$, where the index $\bar{t}$ runs over all finite sequences of the above type, is said to be *consistent* provided the following two conditions holds. Firstly,

$$P_{\bar{t}}(A_1 \times A_2 \times \cdots \times A_n) = P_{\sigma(\bar{t})}(A_{\sigma(t_1)} \times A_{\sigma(t_2)} \times \cdots \times A_{\sigma(t_n)})$$

for any Borel sets $A_1, A_2, \ldots, A_n$ in $\mathbb{R}$, and for any permutation σ of $\bar{t} = (t_1, \ldots, t_n)$, and secondly, letting $\bar{t}^* = (t_1, \ldots, t_{n-1})$,

$$P_{\bar{t}}(A \times \mathbb{R}) = P_{\bar{t}^*}(A), \quad A \in \mathcal{B}(\mathbb{R}^{n-1}).$$

It can be shown that if μ is a measure on the space $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R}^{[0,\infty)}))$ of *all* real-valued functions on the right half-line, then its finite dimensional distribution, as defined in Chapter 3, satisfy these consistency criteria. It is the converse that allows us to construct Brownian motion.

Theorem 5.4. [**Daniell-Kolmogorov consistency theorem**] *Suppose $\{P_{\bar{t}}\}$ is a consistent family of finite-dimensional distributions. Then there exists a probability measure $\mathbb{P}$ on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R}^{[0,\infty)})$ such that*

$$P_{\bar{t}}(A) = \mathbb{P}\left(\omega \in \mathbb{R}^{[0,\infty)} \colon (\omega(t_1), \dots, \omega(t_n)) \in A\right), \quad A \in \mathcal{B}(\mathbb{R}^n)$$

holds for any finite set of indices $\bar{t}$.

We omit the proof, which is given in [5].

Recall the definition of the Gauss kernel (5.1). With a view towards invoking the Daniell-Kolmogorov theorem, we consider a collection of distribution functions on $\mathbb{R}^n$

$$F_{\bar{t}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f_{0,t_1}(u_1) f_{u_1, t_2 - t_1}(u_2) \cdots f_{u_{n-1}, t_n - t_{n-1}}(u_n) du_1 \cdots du_n,$$

defined for any tuple $\bar{t}$ with $0 < t_1 < \cdots < t_n$. Computing with the Gauss kernel, we verify that if $(B_{t_1}, \dots, B_{t_n})$ has distribution $F_{\bar{t}}$, then $B_{t_k} - B_{t_{k-1}}$ is itself normally distributed, and moreover these increments are independent.

Next, fairly straight-forward computations confirm that these distribution functions induce a consistent family of finite-dimensional distributions via

$$P_{\bar{t}}(A) = \mathbb{P}_{F_{\bar{t}}}((B_{t_1}, \dots, B_{t_2}) \in A)), \quad A \in \mathcal{B}(\mathbb{R}^n),$$

where $(B_{t_1}, \dots, B_{t_n})$ has distribution function $F_{\bar{t}}$. We deduce the following result, which proves existence of a version (in the sense of Chapter 3) of Brownian motion.

Proposition 5.5. *There exists a probability measure $\mathbb{P}$ on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R}^{[0,\infty)}))$ under which the coordinate process*

$$X_t(\omega) = \omega(t), \quad \omega \in \mathbb{R}^{[0,\infty)},$$

has independent increments, and $X_t - X_s$, $t > s$, is distributed according to $\mathcal{N}(0, t - s)$.

This means that there exists at least a (non-continuous) version of Brownian motion. To finish our construction, we wish to appeal to Kolmogorov's continuity criterion to deduce the existence of the desired continuous process.

Lemma 5.6. *Let B be a standard Brownian motion. Then, for each $n \in \mathbb{N}$, there exists a constant $c_n > 0$ such that*

$$\mathbb{E}[|B_t - B_s|^{2n}] = c_n |t - s|^n.$$

Proof. By stationarity and the scaling relation for Brownian motion, $B_t - B_s$ coincides in law with $(t - s)^{1/2} B_1$. Hence

$$\mathbb{E}[|B_t - B_s|^{2n}] = |t - s|^n \mathbb{E}[|B_1|^{2n}].$$

The random variable B_1 has the normal distribution with variance 1, and in particular has moments of all orders. Hence the desired inequality holds with $c_n = \mathbb{E}[|B_1|^{2n}]$. □

Note that the proof does not use continuity, only stationarity and normality of increments of a process. Hence, for $t > s$, and $n = 2$, we can apply the same reasoning to the coordinate process obtained from the Daniell-Kolmogorov construction to arrive at our desired result.

Theorem 5.7. [**Wiener's theorem**] *There exists a probability measure $\mathbb{P}$ on the space $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R}^{[0,\infty)}))$ and a stochastic process B on this space, such that under $\mathbb{P}$, the process B is a Brownian motion.*

Proof. By Proposition 5.5, the previous lemma, and 3.22, for any fixed $N \in \mathbb{N}$, there exists a continuous process $(B_t^N : t \in [0, N])$ satisfying all the requirements on a Brownian motion up to time N. We consider the subsets

$$\Omega_N = \{\omega \in \mathbb{R}^{[0,\infty)} : B_t(\omega) = B_t^N(\omega), t \in \mathbb{Q}\}$$

and observe that $\mathbb{P}(\cap_{N=1}^{\infty} \Omega_N) = 1$. On this intersection, all the processes B^N are continuous and agree on the rationals, and hence are continuous on all of $[0, \infty)$. On the complement, we set $B_t(\omega) = 0$, and then $t \mapsto B_t(\omega)$ are continuous functions. □

Our construction has produced a probability measure on the (very large) space $\mathbb{R}^{[0,\infty)}$. Later, in connection with Donsker's invariance principle, we shall place such a probability measure on the nicer space $C[0, \infty)$.

Finally, we can construct Brownian motion on $\mathbb{R}^d$ by taking d independent Brownian motions $W_1, \ldots, W_d$ on $\mathbb{R}$, and set $B = (W_1, \ldots, W_d)$. One then verifies that the process B satisfies all the requirements on a higher-dimensional Brownian motion.

5.3 Regularity and roughness of Brownian paths

The arguments in our construction can be used to show that the sample paths of Brownian motion are not only continuous, but actually almost surely Hölder continuous for small enough exponents.

Theorem 5.8. *Brownian motion is almost surely locally γ-Hölder continuous for any exponent $\gamma \in (0, 1/2)$.*

Proof. From the work in Lemma 5.6, and Kolmogorov's theorem 3.22, it follows that since

$$\mathbb{E}[|B_t - B_s|^{2n}] \leq C|t - s|^{1+(n-1)},$$

Brownian motion is $(n-1)/2n$-Hölder for every $n \geq 1$. □

On the other hand, the sample paths of Brownian motion look rather rough, and it is natural to ask whether they are differentiable. The answer turns out to be negative in a rather strong way.

Theorem 5.9. *For any $t \geq 0$, Brownian motion is almost surely not differentiable at t.*

Proof. To prove this, we may take $t = 0$; the general case then follows by considering translations $\tilde{B}_s = B_{t+s} - B_t$, $s \geq 0$.

By scaling invariance, the process

$$W_t^n = \frac{1}{n^2} B_{n^4 t}, \quad t \geq 0,$$

is a Brownian motion for every $n = 1, 2, \ldots$. We introduce the events

$$A_n = \left\{ \frac{|B_t|}{t} > n \ \text{ for some } \ t \in \left[0, \frac{1}{n^4}\right] \right\},$$

and our goal is to show that these events occur with high probability. We have

$$\begin{aligned} \mathbb{P}(A_n) &\geq \mathbb{P}\left(\frac{|B_{1/n^4}|}{1/n^4} > n \right) \\ &= \mathbb{P}\left(n^2 |B_{1/n^4}| > \frac{1}{n} \right) \\ &= \mathbb{P}\left(|W_1^n| > \frac{1}{n} \right), \end{aligned}$$

and since W_1^n has normal distribution with mean 0 and variance 1, the latter probability tends to 1 as $n \to \infty$.

The events $(A_n)_{n=1}^{\infty}$ form a contracting sequence, and hence

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n \right) = \lim_{n \to \infty} \mathbb{P}(A_n) = 1.$$

Finally, the lower bound $|B_t(\omega)|/t > n$, for some $t \in [0, 1/n^4]$ for each integer $n = 1, 2, \ldots$ implies that $B_t(\omega)$ is not differentiable at $t = 0$, and we are done. □

More refined arguments show that Brownian paths are in fact almost surely *nowhere* differentiable. The hard part is to show that the event of probability 1 does not change with different times t–note that we cannot immediately argue by countability here.

Theorem 5.10. *Almost surely, the sample paths of Brownian motion are not Lipschitz continuous at any point. In particular, the sample paths of Brownian motion are almost surely not differentiable at any point.*

For a proof of this result, which was discovered by R.E.A.C. Paley, N. Wiener, and A. Zygmund, see for instance [3, Chapter 7] or [5, Chapter 2].

5.4 Blumenthal's 01-law and martingales for Brownian motion

We have already noted that $B_{t+s} - B_s$ is independent of $\mathcal{F}_s^B$. We now generalize this statement to the larger filtration $(\mathcal{F}_t^{B+})$. We recall that this right-continuous filtration admits more stopping time than the natural filtration; hitting times of open sets furnish examples of this.

Proposition 5.11. *For any $s \geq 0$, the process $(B_{t+s} - B_s \colon t \geq 0)$ is independent of $\mathcal{F}_s^{B+}$.*

Proof. Let $s \geq 0$ be given, and consider a sequence (s_k) with $s_k \downarrow s$. By continuity then,

$$\lim_{k\to\infty} (B_{t+s_k} - B_{s_k}) = B_{t+s} - B_s \quad \text{almost surely.}$$

Now for any tuple $t_1 < \ldots < t_n$, and any k, the collection $(B_{t_1+s_k} - B_{s_k}, \ldots, B_{t_n+s_k} - B_{s_k})$ is independent of $\mathcal{F}_{s_k}^B$ and hence of $\mathcal{F}_s^{B+}$, and then so is the limit process by a monotone class argument. □

Theorem 5.12. [**Blumenthal's 01-law**] *The germ σ-algebra $\mathcal{F}_0^{B+}$ is trivial, that is, any $A \in \mathcal{F}_0^{B+}$ has $\mathbb{P}(A) \in \{0, 1\}$.*

Proof. We apply Proposition 5.11 with $s = 0$: any $A \in \mathcal{F}_t^B$, for $t \geq 0$, is independent of $\mathcal{F}_0^{B+}$. Hence, in particular, $A \in \mathcal{F}_0^{B+} = \bigcap \mathcal{F}_t^B$ is independent of itself. This means that

$$[\mathbb{P}(A)]^2 = \mathbb{P}(A \cap A) = \mathbb{P}(A),$$

and the assertion of the theorem follows. □

Blumenthal's 01-law is very useful in the study of almost sure path properties of Brownian motion.

Theorem 5.13. *Let B be a standard Brownian motion on $\mathbb{R}$. Set*

$$\tau^* = \inf\{t > 0 \colon B_t > 0\} \quad \textit{and} \quad \tau_* = \inf\{t > 0 \colon B_t = 0\}.$$

Then

$$\mathbb{P}(\tau^* = 0) = \mathbb{P}(\tau_* = 0) = 1.$$

Proof. The first step is to observe that $\{\tau^* = 0\} \in \mathcal{F}_{1/n}^B$, for all $n \geq 1$. This follows once we represent this event as

$$\{\tau^* = 0\} = \bigcap_{1/k \leq 1/n} \{B_\varepsilon > 0 \text{ for some } \varepsilon \in (0, 1/k)\},$$

and recall that $\mathcal{F}_0^{B+} = \bigcap_{t>0} \mathcal{F}_t^B$.

By Blumenthal's 01-law then, the event $\{\tau^* = 0\}$ occurs with probability 0 or 1. The assertion of the theorem will follow if we can establish that the probability is in fact positive. Now, for any $t > 0$,

$$\mathbb{P}(\tau^* \leq t) \geq \mathbb{P}(B_t > 0) = \frac{1}{2};$$

the last equality follows from the symmetry of the normal distribution. Thus, letting t tend to zero, we obtain $\mathbb{P}(\tau^* = 0) \geq 1/2$, and $\mathbb{P}(\tau^* = 0) = 1$ follows.

Next, we note that, by symmetry,

$$\mathbb{P}(\inf\{t > 0 \colon B_t < 0\} = 0) = 1.$$

Continuity of the sample paths of B and the intermediate value theorem now yield $\tau_* = 0$ almost surely. □

The set of zeros $\mathcal{Z}$ of a one-dimensional Brownian motion B has expected Lebesgue measure 0; this follows from Fubini's theorem since

$$\mathbb{E}(|\mathcal{Z}(\omega) \cap [0,t]|) = \int_0^t \mathbb{P}(B_s = 0) ds = 0.$$

The truth is that the zeros of Brownian motion is almost surely an uncountable set and in fact, forms a fractal set.

In higher dimensions, the natural analog of the previous result involved hitting times of cones of the form

$$\mathcal{C} = \{\lambda y \colon \lambda > 0,\ y \in U \subset \mathbb{S}^d\}. \tag{5.2}$$

Here, $U \neq \varnothing$ is an open subset of $\mathbb{S}^d = \{x \in \mathbb{R}^d \colon \|x\| = 1\}$, the unit sphere in $\mathbb{R}^d$.

Proposition 5.14. *Let B be a standard Brownian motion on $\mathbb{R}^d$, and let $\mathcal{C}$ be as in* (5.2)*. Then*

$$\tau_{\mathcal{C}} = \inf\{t > 0 \colon B_t \in \mathcal{C}\}$$

has $\mathbb{P}(\tau_{\mathcal{C}} = 0) = 1$.

Proof. By definition, the sets in (5.2) are invariant under multiplication by positive scalars. Hence, for any $t > 0$,

$$\mathbb{P}(B_t \in \mathcal{C}) = \mathbb{P}(t^{1/2} B_1 \in \mathcal{C}) = \mathbb{P}(B_1 \in \mathcal{C}).$$

By assumption, $\mathcal{C}$ has non-empty interior, and since B_1 has d-dimensional normal distribution, $\mathbb{P}(B_1 \in \mathcal{C}) > 0$. Arguing as in the proof of the previous Theorem, we deduce from Blumenthal's 01-law that $\mathbb{P}(\tau_{\mathcal{C}} = 0) = 1$. □

There exist a number of martingales associated with Brownian motion and its filtrations in a natural way. The simplest two are the following two processes.

Proposition 5.15. *Let $B = (B_t \colon t \geq 0)$ be a Brownian motion. Then the processes*

$$B = (B_t \colon t \geq 0) \quad \textit{and} \quad \tilde{B} = (B_t^2 - t \colon t \geq 0),$$

are martingales with respect to the filtration $(\mathcal{F}_t^{B+})$.

Proof. In both cases, adaptedness and integrability are immediate. For Brownian motion itself, the martingale property follows from independence of increments. To establish the martingale property of $\tilde{B}$, we complete the square and write, for $s < t$,

$$\mathbb{E}[B_t^2 - t | \mathcal{F}_s^{B+}] = \mathbb{E}[(B_t - B_s)^2 | \mathcal{F}_s^{B+}] + 2\mathbb{E}[B_t B_s | \mathcal{F}_s^{B+}] - \mathbb{E}[B_s^2 | \mathcal{F}_s^{B+}] - t.$$

The first term on the right-hand side is equal to $t - s$ by independence. In the second term, we take out the $\mathcal{F}_s^{B+}$-measurable factor B_s and use the martingale property of B. Combining this, we obtain

$$\mathbb{E}[B_t^2 - t | \mathcal{F}_s^{B+}] = t - s + 2B_s^2 - B_s^2 - t = B_s^2 - s,$$

which shows that the process $(B_t^2 - t)$ is indeed a martingale. □

Using these martingales, one can show gambler's ruin estimates for one-dimensional Brownian motion. For $x \in \mathbb{R}$, we define the stopping times

$$T_x = \inf\{t > 0 \colon B_t = x\}.$$

and we find that (see Example sheet 3), for $x, y > 0$,

$$\mathbb{P}(T_{-y} < T_x) = \frac{x}{x+y} \quad \text{and} \quad \mathbb{E}[T_x \wedge T_{-y}] = xy.$$

We next turn to a martingale which actually characterizes Brownian motion.

Proposition 5.16. *Let B be a Brownian motion taking values in $\mathbb{R}^d$, and define*

$$Z^u = \exp\left(i\langle u, B_t\rangle + \frac{t\|u\|^2}{2}\right), \quad u \in \mathbb{R}^d. \tag{5.3}$$

Then Z^u is an $(\mathcal{F}_t^{B+})$-martingale for all $u \in \mathbb{R}^d$.

One can show that a continuous process for which Z^u is a martingale is an $(\mathcal{F}_t)$-Brownian motion.

Proof. This amounts to a computation. We first recall that $\varphi_X(u) = \exp(-u^2/2)$ for X distributed according to $\mathcal{N}(0,1)$; in d dimensions and with variance t, we obtain

$$\mathbb{E}[e^{i\langle u, X\rangle}] = e^{-t\frac{\|u\|^2}{2}}, \quad u \in \mathbb{R}^d.$$

Conditioning and using translation invariance plus independence, we find that

$$\begin{aligned}\mathbb{E}[Z_t^u|\mathcal{F}_s] &= \mathbb{E}\left[e^{i\langle u, B_s\rangle + s\frac{\|u\|^2}{2}} e^{i\langle u, B_t - B_s\rangle + (t-s)\frac{\|u\|^2}{2}} \,\Big|\, \mathcal{F}_s^{B+}\right] \\ &= Z_s^u \mathbb{E}\left[e^{i\langle u, B_{t-s}\rangle + (t-s)\frac{\|u\|^2}{2}}\right] = Z_s^u.\end{aligned}$$

□

We close this section by describing a general procedure for generating martingales in continuous time starting from a Brownian motion.

In what follows, we work on $\mathbb{R}_+ \times \mathbb{R}^d$, and we write $(t, x) = (t, x_1, \ldots, x_d)$ for a point in the space. We shall think of $t \geq 0$ as being time. The linear partial differential operator

$$\partial_t - \frac{1}{2}\Delta = \partial_t - \frac{1}{2}\sum_{k=1}^{d} \partial_{x_j}^2 \tag{5.4}$$

acts in a well-defined way on any real-valued function $f = f(t, x)$ that is continuously differentiable in the first variable, and is twice continuously differentiable with respect to each x_k, $k = 1, \ldots, n$. Let us write $C_b^{1,2} = C_b^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$ for the set of all functions that satisfy these requirements, are bounded and have bounded partials up to orders $(1,2)$. Associated with this operator is the *heat equation*

$$\partial_t u(t,x) = \frac{1}{2}\Delta u(t,x);$$

this equation describes diffusion of heat, and is the fundamental example of a *parabolic partial differential equation.*

We change notation, and set

$$p_t(x,y) = \frac{1}{(2\pi t)^{d/2}} e^{-\frac{\|x-y\|^2}{2t}}, \quad (t,x) \in (\mathbb{R}_+ \setminus \{0\}) \times \mathbb{R}^d.$$

It amounts to a computation to verify that the transition density p_t solves the heat equation,

$$\left(\partial_t - \frac{1}{2}\Delta\right) p_t(x,y) = 0, \quad \text{for} \quad t > 0, \quad x, y \in \mathbb{R}^d.$$

Since also $\lim_{t\to 0}[p_t * g](x) = g(x)$, in a suitable sense, the kernel p_t is actually a *fundamental solution* to the Cauchy problem for the heat equation (see [4] for a discussion). These facts, in conjunction with the Markov property of Brownian motion, lead to the following result.

Theorem 5.17. *Let $B = (B_t \colon t \geq 0)$ be a d-dimensional Brownian motion, started at $x \in \mathbb{R}^d$, and let $f \in C_b^{1,2}$.*

Then the process $M = (M_t \colon t \geq 0)$ given by

$$M_t = f(t, B_t) - f(0, B_0) - \int_0^t \left(\partial_t + \frac{1}{2}\Delta\right) f(s, B_s) ds$$

is a martingale with respect to the filtration $(\mathcal{F}_t^{B+})$.

Proof. In view of our boundedness and continuity assumptions, integrability and adaptedness follow.

In order to establish that M_t is a martingale, we therefore need to show that

$$\mathbb{E}[M_{t+s} - M_s | F_s^+] = 0, \quad t > s,$$

holds almost surely. To accomplish this, we shall use the fact that, for any $s \geq 0$, the process $W_t = B_{t+s} - B_s$ is a Brownian motion, and is independent of $\mathcal{F}_s^+$.

From the definition of M_t, we obtain

$$\begin{aligned} M_{t+s} - M_s &= f(t+s, B_{t+s}) - f(s, B_s) - \int_s^{t+s} \left(\partial_r + \frac{1}{2}\Delta\right) f(r, B_r) dr \\ &= f(t+s, B_{t+s}) - f(s, B_s) - \int_0^t \left(\partial_r + \frac{1}{2}\Delta\right) f(r+s, B_s + W_t) dr \\ &= I_1 + I_2. \end{aligned}$$

We begin by considering $I_1 = f(t+s, B_{t+s}) - f(s, B_s)$. In view of the Markov property, we find that

$$\begin{aligned}
\mathbb{E}[I_1|\mathcal{F}_s^+] &= \mathbb{E}[f(t+s, B_{t+s} - B_s + B_s)|\mathcal{F}_s^+] - f(s, B_s) \\
&= \mathbb{E}[f(t+s, B_s + W_t)|F_s^+] - f(s, B_s) \\
&= \int_{\mathbb{R}^d} f(t+s, B_s + x) p_t(0, x) dx - f(s, B_s);
\end{aligned}$$

where the integral in the last line represents expectation with respect to a Brownian motion started at the (random) point B_s.

We next turn to the integral constituting I_2. Applying Fubini's theorem, and the Markov property, we obtain

$$\begin{aligned}
-\mathbb{E}[I_2|\mathcal{F}_s^+] &= \int_0^t \mathbb{E}\left[\left(\partial_r + \frac{1}{2}\Delta\right) f(r+s, B_{r+s}) \Big| \mathcal{F}_s^+\right] dr. \\
&= \int_0^t \int_{\mathbb{R}^d} \left(\partial_r + \frac{1}{2}\Delta\right) f(r+s, B_s + x) p_r(0, x) dx\ dr.
\end{aligned}$$

Splitting the latter integral as a sum, and then integrating by parts, we obtain

$$\begin{aligned}
\int_0^t \int_{\mathbb{R}^d} \partial_r f(r+s, B_s + x) p_r(0, x) dx dr = &\int_{\mathbb{R}^d} f(t+s, B_s + x) p_t(0, x) dx - f(s, B_s) \\
&- \int_0^t \int_{\mathbb{R}^d} f(r+s, B_s + x) \partial_r p_r(0, x) dx dr.
\end{aligned}$$

Here we have interpreted evaluation of the integrand at 0 as a limit. Making use of the fact that $\partial_t p_t = (1/2)\Delta p_t$, we then apply integration by parts to the remaining integral. Since the terms involving Laplacian cancel out, we arrive at

$$-\mathbb{E}[I_2|\mathcal{F}_s^+] = \int_{\mathbb{R}^d} f(t+s, B_s + x) p_t(0, x) dx - f(s, B_s).$$

Thus

$$\mathbb{E}[I_1|\mathcal{F}_s^+] + \mathbb{E}[I_2|\mathcal{F}_s^+] = 0,$$

and it follows that M_t is a martingale. □

In particular, if $f(t, x) = f(x)$ and f is harmonic, that is $\Delta f(x) = 0$, then the theorem implies that the process $M = (f(B_t) \colon t \geq 0)$ is a martingale. In the case where harmonicity only holds on a subdomain $D \subset \mathbb{R}^d$, we can frequently use stopping times to obtain processes that are martingales up to the time when B exits D.

5.5 Strong Markov property and reflection principle

We shall now prove that Brownian motion enjoys the *strong Markov property*: the assertion of Proposition 5.3 continues to hold when the sure time s is replaced by a stopping time T.

Theorem 5.18. [**Strong Markov property**] *Let τ be an almost surely finite stopping time. Then the process*

$$W^\tau = (B_{t+\tau} - B_\tau : t \geq 0)$$

is a standard Brownian motion independent of $\mathcal{F}^{B+}_\tau$.

Proof. We begin by once again considering the discrete approximations $\tau_n = 2^{-n}\lceil 2^n\tau \rceil$. As in Section 3, (τ_n) is a sequence of stopping times, taking dyadic values, with $\tau_n \downarrow \tau$ as $n \to \infty$. In addition, we already know that, for any $k \geq 0$, the process

$$B^k_t = B_{t+k2^{-n}} - B_{k2^{-n}}$$

is a Brownian motion, and is independent of $\mathcal{F}^B_{k2^{-n}}$.

We shall now prove that

$$\tilde{B}^n_t = B_{t+\tau_n} - B_{\tau_n}$$

is also a Brownian motion, and that this process is independent of $\mathcal{F}^{B+}_{\tau_n}$. Since the stopping times (τ_n) take values in a countable set, we can decompose them into a countable sum involving the events $\{\tau_n = k2^{-n}\}$, where $\tilde{B}^n$ is expressible in terms of B^k, use independence on each such event, and then reassemble everything at the end.

Let us implement this strategy and show that, for an m-tuple $t_1 < \cdots < t_m$, an event $A \in \mathcal{F}^{B+}_\tau$, and $f \in C_b(\mathbb{R}^m)$, we have

$$\mathbb{E}[f(B_{t_1+\tau} - B_\tau, \ldots, B_{t_m+\tau} - B_\tau)\mathbf{1}(A)] = \mathbb{E}[f(B_{t_1}, \ldots, B_{t_m})]\mathbb{P}(A). \tag{5.5}$$

We begin by establishing this for our discretized stopping times, in which case we may write

$$\begin{aligned}&\mathbb{E}[f(B_{t_1+\tau_n} - B_{\tau_n}, \ldots, B_{t_m+\tau_n} - B_{\tau_n})\mathbf{1}(A)] \\ &\qquad = \sum_{k=0}^{\infty} \mathbb{E}[f(B_{t_1+k2^{-n}} - B_{t_1}, \ldots, B_{t_m+k2^{-n}} - B_{t_m})\mathbf{1}(A \cap \{\tau_n = k2^{-n}\}].\end{aligned}$$

By independence and the (weak) Markov property, and since $A \cap \{\tau_n = k2^{-n}\} \in \mathcal{F}^{B+}_{k2^{-n}}$, each summand can be factorized as $]\mathbb{E}[f(B_{t_1}, \ldots, B_{t_m})]\mathbb{P}(A \cap \{\tau_n = k2^{-n}\})$, and since $\mathbb{E}[f(B_{t_1}, \ldots, B_{t_m})]$ does not depend on n, the right-hand side in the displayed equation is equal to

$$\mathbb{E}[f(B_{t_1}, \ldots, B_{t_m})] \sum_{k=0}^{\infty} \mathbb{P}(A \cap \{\tau_n = k2^{-n}\}) = \mathbb{E}[f(B_{t_1}, \ldots, B_{t_m})]\mathbb{P}(A);$$

this last expression does not depend on n.

Since B is a continuous process and f is bounded, the dominated convergence theorem now yields that

$$\lim_{n\to\infty} \mathbb{E}[f(B_{t_1+\tau_n} - B_{\tau_n}, \ldots, B_{t_m+\tau_n} - B_{\tau_n})\mathbf{1}(A)] = \mathbb{E}[f(B_{t_1+\tau} - B_\tau, \ldots)\mathbf{1}(A),$$

and by our previous computation, (5.5) follows. Taking $A = \Omega$ in (5.5) now shows that W has the finite-dimensional distributions of a Brownian motion. Since finite-dimensional distributions characterize continuous processes, it follows that W has the law of a Brownian motion, as claimed.

It remains to show that W is independent of $\mathcal{F}_\tau^{B+}$. By approximating $\mathbf{1}(B)$ for $B \in \mathcal{B}(\mathbb{R}^m)$ by continuous functions, we deduce from (5.5) that

$$\mathbb{P}((B_{t_1+\tau} - B_\tau, \dots, B_{t_m+\tau} - B\tau) \in B) \cap A) = \mathbb{P}((B_{t_1+\tau} - B_\tau, \dots, B_{t_m+\tau} - B_\tau) \in B)\mathbb{P}(A).$$

This shows that independence holds for cylinder sets. We next check that the sets $B \in \mathcal{B}(C[0,\infty))$ for which

$$\mathbb{P}(\{W \in B\} \cap A) = \mathbb{P}(W \in B)\mathbb{P}(A)$$

holds form a monotone class. As we have seen that this monotone class contains the cylinder sets, independence of W from $\mathcal{F}_\tau^{B+}$ then follows from the monotone class theorem. □

From the strong Markov property, we deduce results concerning reflection of one-dimensional Brownian motion. As usual, we define the hitting time of level $a \in \mathbb{R}$

$$\tau_a = \inf\{t > 0 \colon B_t = a\};$$

since B is a continuous process, this is an $(\mathcal{F}_t^{B+})$-stopping time.

Theorem 5.19. *Let B be a one-dimensional Brownian motion. For $a \in \mathbb{R}$ fixed, the process*

$$\tilde{B}_t = \begin{cases} B_t, & t < \tau_a \\ 2a - B_t, & t \geq \tau_a \end{cases} \tag{5.6}$$

is a Brownian motion

Proof. Invoking the strong Markov property, we find that $W = (B_{\tau_a+t} - a \colon t \geq 0)$ is a standard Brownian motion that is independent of $(B_t \colon \tau \geq t \geq 0)$, and so is the reflected process $-W$.

This means that the pair (B, W) has the same law as $(B, -W)$. From the first pair, we obtain a continuous process via the mapping

$$\Phi \colon (B, W) \mapsto B\mathbf{1}(t \leq \tau_a) + (a + W_{t-\tau_a})\mathbf{1}(t > \tau_a),$$

which has the same law as $\Phi(B, -W)$. By definition, $\Phi(B, W) = B$ while $\Phi(B, -W) = \tilde{B}$, and since the mapping Φ is measurable, equality in law of B and $\tilde{B}$ follow. □

This *reflection principle* allows us to study the distribution of the *running maximum of Brownian up to time* $t > 0$. This is the process

$$M_t = \sup_{0 \leq s \leq t} B_s.$$

Theorem 5.20. *Let B be Brownian motion on $\mathbb{R}$. For $a > 0$, the process $M_t = \sup_{0 \leq s \leq t} B_s$ has*

$$\mathbb{P}(M_t > a) = 2\mathbb{P}(B_t > a) = \mathbb{P}(|B_t| > a).$$

Proof. Let $a > 0$ be given, and consider, as before, the stopping time $\tau_a = \inf\{t > 0\colon B_t = a\}$. Since $\{B_t > a\} \subset \{M_t > a\}$, we obtain

$$\{M_t > a\} = \{B_t > a\} \cup (\{M_t > a\} \cap \{B_t \leq a\}).$$

Note that the event on the right-hand side is given as disjoint union. Since

$$\{M_t > a\} \cap \{B_t \leq a\} = \{\tilde{B}_t \geq a\},$$

where $\tilde{B}$ is defined as in the previous theorem, the statement follows from the reflection principle 5.19 via

$$\mathbb{P}(M_t \geq a) = \mathbb{P}(B_t > a) + \mathbb{P}(\tilde{B}_t \geq a).$$

□

As an immediate corollary, we obtain the density of the *passage time*

$$\tau_a = \inf\{t > 0\colon B_t = a\}.$$

Corollary 5.21. *For $a > 0$, the random variable τ_a has density*

$$f_{\tau_a}(t) = \frac{a}{\sqrt{2\pi t^3}} e^{-\frac{a^2}{2t}}, \quad t > 0.$$

Proof. We note that

$$\mathbb{P}(\tau_a \leq t) = \mathbb{P}(M_t \geq a),$$

and then, using Theorem 5.20 and performing a change of variables, we find that

$$\mathbb{P}(\tau_a \leq t) = \sqrt{\frac{2}{\pi}} \int_{\frac{a}{\sqrt{t}}}^{\infty} e^{-\frac{x^2}{2}}\, dx.$$

The desired result now follows upon differentiation with respect to the variable t. □

5.6 Recurrence, transience, and connections with partial differential equations

We shall now study the long-time behavior of Brownian motion, making heavy use of martingales constructed as in Theorem 5.17 and the optional stopping theorem.

Definition 5.22. A process $X = (X_t\colon t \geq 0)$ on $\mathbb{R}^d$ started at x_0 is said to be *point-recurrent* or simply *recurrent* if, for each $x \in \mathbb{R}^d$, the set

$$\{t \geq 0\colon X_t = x\} \subset [0, \infty)$$

is unbounded $\mathbb{P}_{x_0}$-almost surely.

We say that X is *neighborhood-recurrent* if, for each $x \in \mathbb{R}^d$, the sets

$$\{t \geq 0\colon \|X_t - x\| \leq \varepsilon\}$$

are unbounded $\mathbb{P}_{x_0}$-almost surely for each $\varepsilon > 0$.

The process X is *transient* if $\|X_t\| \to \infty$ holds almost surely as $t \to \infty$.

The recurrence and transience properties of Brownian motion depend on the dimension of the ambient space. As we shall see, the decay of the *Green's function* for the Laplacian plays an important role.

Theorem 5.23. *Brownian motion on $\mathbb{R}$ is point recurrent. Planar Brownian motion is neighborhood-recurrent, but not point-recurrent.*

For $d \geq 3$, d-dimensional Brownian motion is transient.

Proof. The statement concerning one-dimensional Brownian motion follows from the fact that

$$\limsup_{t\to\infty} B_t = \infty \quad \text{and} \quad \liminf_{t\to\infty} B_t = -\infty;$$

cf. Proposition 5.14.

We turn to a proof of neighborhood recurrence in $\mathbb{R}^2$. We begin by considering the function

$$G(z) = \log \|z\|, \quad z \in \mathbb{R}^2 \setminus \{0\}.$$

A computation (using polar coordinates for instance) shows that G is harmonic away from the origin, that is, that

$$\Delta G(z) = 0, \quad z \neq 0.$$

In fact, G is the Green's function for the Dirichlet problem on the unit disk $\mathbb{D} = \{z \in \mathbb{R}^2 \colon \|z\| < 1\}$; that is, formally, G solves the problem[7]

$$\begin{cases} \Delta G(z) = \delta_0, & z \in \mathbb{D}, \\ G(\zeta) = 0, & \zeta \in \partial\mathbb{D}. \end{cases}$$

For $\varepsilon, R \in \mathbb{R}$ with $0 < \varepsilon < R < \infty$, we consider the annular region

$$A(\varepsilon, R) = \{z \in \mathbb{R}^2 \colon \varepsilon < \|z\| < R\},$$

and a function $f \in C_b^2(\mathbb{R}^2)$ that satisfies

$$f(x) = G(x) = \log \|x\|, \quad x \in A(\varepsilon, R).$$

In particular then, $\Delta f(x) = 0$ on $A(\varepsilon, R)$. Moreover, by Theorem 5.17, the process $M = (M_t \colon t \geq 0)$ defined by

$$M_t = f(B_t) - \int_0^t \Delta f(B_s) ds$$

is a martingale with respect to the filtration $(\mathcal{F}_t^+)$ associated with a Brownian motion started at a point $x \in A(\varepsilon, R)$.

We define two stopping times through

$$\tau_{\varepsilon,R} = \inf\{t > 0 \colon \|B_t\| = \varepsilon, R\},$$

[7]We are suggestively using z for a point in $\mathbb{R}^2$, which can be identified with the complex plane. There are many fascinating connections between complex analysis, potential theory, and Brownian motion, but they are sadly outside the scope of this course.

and introduce the almost surely bounded stopping times

$$T = \tau_\varepsilon \wedge \tau_R.$$

Since the stopped process M^T given by $M_t^T = M_{t\wedge T}$ is a bounded martingale, we can apply the optional stopping theorem in continuous time, cf. Theorem 3.21. As the integral term in the initial definition of M_t drops out, this then yields the relation

$$\mathbb{E}_x[M_T] = \mathbb{E}_x[\log|B_t|] = \mathbb{E}_x[\log|B_0|] = \log\|x\|, \tag{5.7}$$

or, formulated in terms of the stopping times,

$$\log\varepsilon\ \mathbb{P}_x(\tau_\varepsilon < \tau_R) + \log R\ \mathbb{P}_x(\tau_R < \tau_\varepsilon).$$

Solving for $\mathbb{P}_x(\tau_\varepsilon < \tau_R)$ using $\mathbb{P}_x(\tau_\varepsilon < \tau_R) + \mathbb{P}_x(\tau_R < \tau_\varepsilon) = 1$, we arrive at

$$\mathbb{P}_x(\tau_\varepsilon < \tau_R) = \frac{\log\left(\frac{R}{\|x\|}\right)}{\log\left(\frac{R}{\varepsilon}\right)}. \tag{5.8}$$

We now let R tend to infinity in (5.8) to obtain

$$\mathbb{P}_x(\tau_\varepsilon < \infty) = 1;$$

that $T_R \to \infty$ as $R \to \infty$ follows by continuity of Brownian motion. This means that for $x \in A(\varepsilon, \infty)$,

$$\mathbb{P}_x(\|B_t\| \leq \varepsilon \text{ for some } t > 0) = 1.$$

Next, the strong Markov property implies that $W = (B_{t+n} - B_n \colon t \geq 0)$ is a Brownian motion independent of $\mathcal{F}_n^+$. Hence

$$\begin{aligned}\mathbb{P}_x(\|B_t\| \leq \varepsilon \text{ for some } t \geq n) &= \mathbb{P}_x(\|B_{t+n} - B_n + B_n\| \leq \varepsilon \text{ for some } t > 0)\\ &= \int_{\mathbb{R}^2} \mathbb{P}_0(\|W_t + y\| \leq \varepsilon \text{ for some } t > 0)\mathbb{P}_x(B_n = y)dy\\ &= \int_{\mathbb{R}^2} \mathbb{P}_y(\|B_t\| \leq \varepsilon \text{ for some } t > 0)p_n(x,y)dy,\end{aligned}$$

but the latter integral is equal to 1 for all n since the integrand is 1. This shows that $\{t > 0 \colon \|B_t\| \leq \varepsilon\}$ is unbounded, as claimed.

It remains to prove that B is not point-recurrent. Returning to the equation (5.8), we first send $\varepsilon \to 0$ to get

$$\mathbb{P}_x(\tau_0 < \tau_R) = 0.$$

This means that, almost surely, Brownian motion hits the outer circle $\{\|x\| = R\}$ before it reaches the origin. Taking the limit $R \to \infty$, we conclude that

$$\mathbb{P}_x(B_t = 0 \text{ for some } t > 0) = 0.$$

Again appealing to the Markov property, we obtain that for $a > 0$,

$$\mathbb{P}_0(B_t = 0 \text{ for some } a > 0) = 0,$$

and taking the limit $a \to 0$, we finally arrive at

$$\mathbb{P}_0(B_t = 0 \text{ for some } t > 0) = 0,$$

thus completing the proof.

In the higher-dimensional case $d \geq 3$, one works with the function

$$G(x) = \frac{1}{\|x\|^{d-2}}, \quad x \in \mathbb{R}^d, \|x\| > 0.$$

Again using optional stopping, one obtains

$$\mathbb{P}_x(\tau_\varepsilon < \tau_R) = \frac{\|x\|^{2-d} - R^{2-d}}{\varepsilon^{2-d} - R^{2-d}},$$

where annular regions are now replaced by spherical shells in the definition of the stopping times. This shows, after an appeal to the Borel-Cantelli lemma, that only finitely many events of the form $\{\|B_t\| < n \text{ for all } t \geq \tau_{n^3}\}$ occur, and transitivity follows.

We leave the details as an exercise (see Example Sheet 4). □

Theorem 5.17 and the proof of Theorem 5.23 clearly indicate a connection between Brownian motion and the Laplacian. The fact that the Dirichlet problem for the Laplacian can be solved using Brownian motion serves as a further illustration.

The classical *Dirichlet problem* runs as follows: Given a bounded connected open set $D \subset \mathbb{R}^d$, and a continuous function $f \in C(\partial D)$, find a function $u \colon \bar{D} \to \mathbb{R}$ such that $u \in C(\bar{D})$ and

$$\begin{cases} \Delta u(x) = 0, & x \in D; \\ u(\xi) = f(\xi), & \xi \in \partial D \end{cases}. \tag{5.9}$$

We shall refer to an open connected set in $\mathbb{R}^d$ as a *domain*. Counterexamples due to S. Zaremba and H. Lebesgue show that the Dirichlet problem does not admit a solution witout some additional regularity assumptions on the domain D. A necessary and sufficient condition was obtained by Wiener; we content ourselves with a stronger sufficient condition, given in terms of exterior cones.

Definition 5.24. Let $D \subset \mathbb{R}^d$ be an open connected set. We say that D satisfies the *Zaremba cone condition* at a boundary point $\xi \in \partial D$ if there exists an open non-empty cone $\mathcal{C}$ with apex at ξ such that, for some $r > 0$,

$$\mathcal{C} \cap B(\xi, r) \subset D^c.$$

We are now ready to present S. Kakutani's solution to the Dirichlet problem.

Theorem 5.25. *Suppose $D \subset \mathbb{R}^d$ is a bounded domain with the property that every boundary point satisfies the Zaremba cone condition. For $B = (B_t : t \geq 0)$ a Brownian motion in $\mathbb{R}^d$, define*

$$\tau = \tau_{\partial D} = \inf\{t \geq 0 \colon B_t \in \partial D\}.$$

Then the unique solution to the Dirichlet problem (5.9) *is given by*

$$u(x) = \mathbb{E}_x[f(B_\tau)], \quad x \in \bar{D}. \tag{5.10}$$

This theorem generalizes to elliptic second-order partial differential operators (and beyond) via the theory of diffusion processes and stochastic differential equations. [8] For the Dirichlet problem, a necessary and sufficient condition has been given by Wiener, but this falls outside the scope of the present course.

We shall need two classical results concerning harmonic functions in the proof of Theorem 5.25. We refer the reader to [4, Chapter 2.2] for background material and demonstrations.

In what follows, $S(B(x,r)) = \int_{\partial B(x,r)} ds$ denotes the surface area of a sphere in $\mathbb{R}^d$ with radius $r > 0$ and center at x, expressed in terms of the area element ds.

Theorem 5.26. [**Mean-value property of harmonic functions**] *Suppose u is measurable and bounded, and has the property that, for every $x \in D$, and every $r > 0$ such that $B(x,r) \subset D$,*

$$u(x) = \frac{1}{S(B(x,r))} \int_{\partial B(x,r)} u(s) ds.$$

Then u is harmonic in D, that is, $\Delta u(x) = 0$ for all $x \in D$.

Theorem 5.27. [**Maximum principle**] *Suppose $u \in C^2(D) \cap C(\bar{D})$ is harmonic in D. Then*

$$\max_{x \in \bar{D}} u(x) = \max_{\xi \in \partial D} u(\xi).$$

The maximum principle implies, in particular, that the Dirichlet problem can have at most one solution. Starting with two solutions u_1, u_2, set $u_\pm = \pm(u_1 - u_2)$: these function u are harmonic and vanish on the boundary of D, and hence $u_1 = u_2$ in D.

Proof. Since D is assumed bounded, the function u is bounded. Moreover, for any $x \in D$ and $r > 0$,

$$\tau_r = \inf\{t > 0 \colon B_t \in \partial B(x,r)\}$$

is a finite stopping time $\mathbb{P}_x$-almost surely.

We now use the strong Markov property to show that

$$u(x) = \mathbb{E}_x[f(B_\tau)]$$

[8] More on this in the course in Stochastic Calculus.

enjoys the mean-value property, and hence is harmonic. Namely, we have

$$\begin{aligned} u(x) &= \mathbb{E}_x[f(B_\tau)] \\ &= \mathbb{E}_x\left[\mathbb{E}_x[f(B_\tau)|\mathcal{F}_{\tau_r}]\right] \\ &= \mathbb{E}_x\left[\mathbb{E}_{B_{\tau_r}}[f(B_\tau)]\right] = \mathbb{E}_x[u(B_{\tau_r}).] \end{aligned}$$

Since Brownian motion is invariant with respect to rotations, and since Lebesgue measure is the unique rotationally invariant measure on the sphere,

$$\mathbb{E}_x[u(B_{\tau_r})] = \frac{1}{S(B(x,r))}\int_{\partial B(x,r)} u(s)ds,$$

and the harmonicity of u follows.

It remains to show that u is continuous in $\bar{D}$ and attains the correct boundary values. This we shall prove using the cone condition.

Since D is a bounded domain, we can find, for any given $\varepsilon > 0$, a number $\delta > 0$ such that

$$|f(\eta) - f(\xi)| < \varepsilon \quad \text{provided} \quad |\xi - \eta| < \delta.$$

For a given $\xi \in \partial D$, we let $(x_k)_{k=1}^\infty$ be a sequence of points in D satisfying

$$|\xi - x_k| < \frac{\delta}{2^k}.$$

A Brownian motion started at $x \in D$ will either hit the boundary ∂D before it exits the ball $B(x,\delta)$, in which case we obtain the desired estimate $|f(\xi) - f(B_\tau)| < \varepsilon$, or else $\tau_\delta = \inf\{t > 0\colon B_t \in \partial B(\xi,\delta)\} < \tau$. With this in mind, and using the triangle inequality, we obtain

$$\begin{aligned} |u(x_k) - u(\xi)| &= |\mathbb{E}_{x_k}[f(B_\tau)] - f(\xi)| \\ &\le \mathbb{E}_{x_k}[|f(B_\tau) - f(\xi)|] \\ &\le \varepsilon\, \mathbb{P}_{x_k}(\tau < \tau_\delta) + 2\sup_{\xi\in\partial D}|f(\xi)|\, \mathbb{P}_{x_k}(\tau_\delta < \tau) \\ &\le \varepsilon + M\mathbb{P}_{x_k}(\tau_\delta < \tau). \end{aligned}$$

The proof will be complete once we show that the probability $\mathbb{P}_{x_k}(\tau_\delta < \tau)$ becomes arbitrarily small when k becomes large.

By assumption, we can choose δ small enough so that there exists a cone with apex at ξ and $\mathcal{C} \cap B(\xi,\delta) \in D^c$. We introduce the stopping time

$$\tau_{\mathcal{C}} = \inf\{t > 0\colon B_t \in \mathcal{C}\},$$

and use the Markov property to write down the estimate

$$\mathbb{P}_x(\tau_\delta < \tau) \le \mathbb{P}_x(\tau_\delta < \tau_C) \le \prod_{j=1}^{k} \sup_{x_j \in B(\xi, 2^{-k+j}\delta)} \mathbb{P}_{x_j}(\tau_{2^{-k+j}\delta} < \tau_{\mathcal{C}}).$$

Finally, since $\mathbb{P}_{x_j}(\tau_{2^{-k}\delta} < \tau_{\mathcal{C}}) = \varepsilon' < 1$ for every k by the invariance of the cone under multiplication, the right-hand side is bounded by $(\varepsilon')^k$, and tends to 0 as $k \to \infty$. □

5.7 Donsker's invariance principle

Our final topic in this chapter on Brownian motion is M. Donsker's remarkable *invariance principle*. This result can be viewed as a generalization of the Central Limit Theorem to processes: Brownian motion is the limit of suitable rescaled random walks, embedded in continuous time through linear interpolation.

While we have already constructed Brownian motion, the invariance principle can also be used to give a construction of B. The details are carried out in [5]; one advantage of this approach is that one argues on weakly convergent measures on the metric space $C[0,\infty)$, endowed with its Borel σ-algebra. In this way one obtains a probability measure $\mathbb{P}$ on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$, the *Wiener measure*, under which the coordinate process

$$B_t(\omega) = \omega(t)$$

is a standard Brownian motion on $\mathbb{R}$.

Throughout this section, we work with stochastic processes on $\mathbb{R}$. We begin by proving a result, interesting in its own right, that provides an embedding of random walks into Brownian motion via a non-decreasing sequence of stopping times. We do this in two steps.

Theorem 5.28. [**Skorokhod representation, part I**]
Suppose X is a random variable with $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = \sigma^2 < \infty$. Then there exists a stopping time for one-dimensional Brownian motion such that B_T and X coincide in law, and

$$\mathbb{E}[T] = \mathbb{E}[X^2].$$

To see how such a statement could possibly be true, we consider the following simple situation. Suppose X takes values in $\{-a, b\}$, where $a, b > 0$, and has mean 0. By necessity, then, the law of X is

$$\mu_X = \frac{b}{b+a}\delta_{-a} + \frac{a}{b+a}\delta_b.$$

We introduce stopping times for standard Brownian motion by taking

$$\tau_{-a} = \inf\{t > 0 \colon B_t = -a\} \quad \text{and} \quad \tau_b = \inf\{t > 0 \colon B_t = b\},$$

and obtain an almost surely finite stopping time by setting

$$T = \tau_{-a} \wedge \tau_b.$$

We know (see Example sheet 3) that

$$\mathbb{P}(\tau_{-a} < \tau_b) = \frac{b}{a+b} \quad \text{and} \quad \mathbb{P}(\tau_b < \tau_{-a}) = \frac{a}{a+b}, \tag{5.11}$$

and so $\mu_{B_T} = \mu_X$ in this simple example. Moreover, since

$$\mathbb{E}[T] = \mathbb{E}[\tau_{-a} \wedge \tau_b] = ab. \tag{5.12}$$

and

$$\mathbb{E}[B_T^2] = -a\frac{b}{a+b} + b\frac{a}{a+b} = ab,$$

the second assertion also holds.

As we shall see in the proof, the basic idea is to reduce the problem to the simple situation we have just analyzed.

Proof. Since X was assumed to have zero mean, that is, $\int_{\mathbb{R}} u d\mu_X(u) = 0$, it follows that

$$\int_{-\infty}^{0} (-u) d\mu_X(u) = \int_{0}^{\infty} v d\mu_X(v) = C$$

for some constant $C > 0$. Now, letting f be a bounded Borel function, we can write

$$\begin{aligned}
C \int f d\mu &= \left(\int_0^\infty f(v) d\mu_X(v) \right) \int_{-\infty}^{0} -u d\mu(u) + \left(\int_{-\infty}^{0} f(u) d\mu_X(u) \right) \int_0^\infty v d\mu(v) \\
&= \int_{v=0}^{\infty} \int_{u=-\infty}^{0} (v f(u) - u(f(v)) d\mu(u) d\mu(v) \\
&= \int \int \left(\frac{v}{v-u} f(u) - \frac{u}{v-u} f(v) \right) (v-u) d\mu(u) d\mu(v) \\
&= \int \int \left(\frac{v}{v-u} f(u) - \frac{u}{v-u} f(v) \right) d\tilde{\nu}(u,v),
\end{aligned}$$

and the expression in parentheses is of the same form as in our example.

Guided by this, we now define, on the same probability space as the Brownian motion and the random variable X, an independent random vector (U, V) having a law with

$$\nu_{(U,V)}(A_1 \times A_2) = \frac{1}{C} \int_{A_1} \int_{A_2} (u+v) d\mu(-u) d\mu(v),$$

for $A_1 \subset (0, \infty)$ and $A_2 \subset (-\infty, 0)$. We let $\mathcal{F}_t = \sigma(\mathcal{F}_0, \mathcal{F}_t^B)$, with $\mathcal{F}_0 = \sigma(U, V)$, and we next define

$$T = \inf\{t \geq 0 \colon B_t = -U \text{ or } V\}. \tag{5.13}$$

Now B is a $\mathcal{F}_t$-Brownian motion, and T is a stopping time. By (5.11) and (5.12), conditional on U and V, it holds that

$$\mathbb{P}(B_T = V | U, V) = \frac{U}{U+V} \quad \text{and} \quad \mathbb{P}(B_T = -U | U, V) = \frac{V}{U+V},$$

as well as $\mathbb{E}[T|U,V] = UV$. Hence, by our previous computations involving the arbitrary Borel function f and culminating in the expression defining ν, we conclude that B_T coincides in law with X.

Finally, we compute, again using the definition of C:

$$\begin{aligned}
\mathbb{E}[T] = \mathbb{E}[\mathbb{E}[T|U,V]] &= \int_0^\infty \int_0^\infty \frac{1}{C} uv(u+v) d\mu(-u) d\mu(v) \\
&= \int_{-\infty}^{0} (-u)^2 d\mu(u) + \int_0^\infty v^2 d\mu(v) = \sigma^2.
\end{aligned}$$

Thus $\mathbb{E}[T] = \mathbb{E}[X^2]$ and the theorem is proved. □

The next step is to embed an entire random walk into a Brownian motion. For notational simplicity, we assume that the variance of the underlying random variables is equal to 1; the general case is treated via rescaling.

Theorem 5.29. [**Skorokhod representation, part II**]
Let $(X_k)_{k=1}^{\infty}$ be a sequence of i.i.d. random variables with zero mean and unit variance, and set $S_n = \sum_{k=1}^{n} X_k$.

Then there exists a sequences of stopping times $(T_n)_{n=1}^{\infty}$ such that

$$0 = T_0 \leq T_1 \leq T_2 \leq \cdots ,$$

the random variable S_n coincides with B_{T_n} in law, and the increments $T_n - T_{n-1}$ are independent and identically distributed, with mean 1.

Proof. We generate a sequence of i.i.d. pairs (U_k, V_k) as in the previous theorem, and let B be an independent Brownian motion.

Letting $T_0 = 0$, we inductively define stopping times

$$T_n = \inf\{t \geq T_{n-1} \colon B_t - B_{T_{n-1}} \notin (U_n, V_n)\}.$$

Invoking the strong Markov property of Brownian motion, we see that $B_{T_n} - B_{T_{n-1}}$ is independent with law μ_X, the increments have mean 1, and both random variables are independent of $\mathcal{F}_{T_{n-1}}$. □

In order to state Donsker's theorem, we introduce the notation $[x]$ for the integer part of a real number x.

Theorem 5.30. [**Donsker's invariance principle**]
Let $(X_k)_{k=1}^{\infty}$ be a sequence of i.i.d. random variables with

$$\mathbb{E}[X_k] = 0 \quad \textit{and} \quad \mathbb{E}[X_k^2] = 1.$$

Set $S_n = \sum_{k=1}^{n} X_k$ and define a process $\mathcal{S} = (\mathcal{S}_t \colon 0 \leq t \leq 1)$ by linear interpolation:

$$\mathcal{S}_t = S_{[t]} + (t - [t])(S_{[t]+1} - S_{[t]}). \tag{5.14}$$

Then the laws of the processes

$$\mathcal{S}^N = \left(\frac{1}{\sqrt{N}} \mathcal{S}_{Nt} \colon 0 \leq t \leq 1 \right)$$

converge weakly to Wiener measure on $(C[0,1], \mathcal{B}(C[0,1]))$.

Recall that weak convergence means that

$$\mathbb{E}[f(\mathcal{S}^N)] \to \mathbb{E}[f(B)] \quad \text{as} \quad N \to \infty$$

for every continuous and bounded function F on the metric space we obtain by endowing $C[0,1]$ with the uniform norm.

The mapping $f_m\colon C[0,1] \to \mathbb{R}$ given by

$$f_m\colon \omega \mapsto \max\{\omega(t)\colon 0 \le t \le 1\}$$

is continuous, and the composition $g(f_m)$ with a bounded continuous $g\colon \mathbb{R} \to \mathbb{R}$, furnishes an example of such a function.

This particular choice also illustrates the usefulness of the invariance principle. Assuming the validity of Donsker's result, it follows that the law of the random variable

$$\max_{0\le l\le n} \frac{S_l}{\sqrt{n}}$$

converges to that of the running maximum

$$M_1 = \sup_{0\le t\le 1} B_t,$$

and for M_t we have already obtained the explicit formula

$$\mathbb{P}(M_1 \ge a) = 2\mathbb{P}(B_1 \ge a).$$

We now turn to the proof of Donsker's invariance principle.

Proof. Let us define, for $N = 1, 2, \ldots$, the processes

$$B_t^N = \sqrt{N} B_{t/N}, \quad 0 \le t \le 1.$$

By Brownian scaling, each B^N is again a standard Brownian motion. We shall use this rescaling to construct certain auxilliary processes.

We invoke the Skorokhod embedding theorem, but with B^N instead of B, to obtain a sequence of stopping times

$$(T_n^N)_{n=1}^\infty \quad \text{for each } N = 1, 2, \ldots$$

We then set $S_n^N = B_{T_n^N}^N$ and form a sequence of continuous-time processes S^N by linear interpolation, as in the statement of the theorem. By construction then, for all $N \ge 1$,

$$((S_t^N), (T_n^N)) \quad \text{coincide in law with} \quad ((\mathcal{S}_t), (T_n)).$$

With this in mind, we rescale,

$$\tilde{T}_n^N = \frac{1}{N} T_n^N, \quad \tilde{S}_t^N = \frac{1}{\sqrt{N}} S_{Nt}^N,$$

so that $\tilde{S}^N$ has the same law as the target process $\mathcal{S}^N$. In view of the definition of B^N as a rescaling of B, we also have

$$\tilde{S}_{n/N}^N = \frac{1}{\sqrt{N}} S_n^N = \frac{1}{\sqrt{N}} B_{T_n^N}^N = B_{\tilde{T}_n^N} \quad \text{for all } n.$$

Now, since $T_n - T_{n-1}$ has mean 1, the strong law of large numbers yields

$$\lim_{n\to\infty} \frac{T_n}{n} = 1,$$

and hence

$$\frac{1}{N} \sup_{1\leq n\leq N} |T_n - n| \to 0 \quad \text{as } N \to \infty.$$

In particular, for any $\delta > 0$,

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{1\leq n\leq N} \left| \tilde{T}_n^N - \frac{n}{N} \right| > \delta \right) = 0.$$

Since $\tilde{S}^N_{n/N} = B_{\tilde{T}_n^N}$, and Brownian motion is continuous, the intermediate value theorem applies to guarantee the existence of an $s \in [\tilde{T}_n^N, \tilde{T}_{n+1}^N]$ such that

$$\tilde{S}_t^N = B_s$$

for every $t \in [n/N, (n+1)/N]$. The idea is that this choice of s allows us to write

$$\tilde{S}_t^N - B_t = \tilde{S}_t^N - B_t + B_s - B_s = -(B_t - B_s),$$

so that showing that $\tilde{S}_t^N - B_t$ tends to zero reduces to controlling the increments of the continuous function B_t.

For $\delta > 0$, we now bound the probability of the unfavorable events $A_N = \{|\tilde{S}_t^N - B_t| > \varepsilon \text{ for some } t \in [0,1]\}$ as follows:

$$\Omega_{B,N} \subset \{|B_t - B_s| > \varepsilon \text{ for some } t \in [0,1], \text{ and } |s - t| < \delta + 1/N\} \cup \{|\tilde{T}_n^N - n/N| > \delta\}.$$

We know that the probability of the second event tends to 0 as $N \to \infty$ for any $\delta > 0$. On the other hand, Brownian motion is uniformly continuous on the unit interval, and so the first probability can be made arbitrarily small, for N large enough, by choosing δ sufficiently small. In view of this

$$\mathbb{P}\left(\sup_{t\in[0,1]} |\tilde{S}_t^N - B_t| > \varepsilon \right) \to 0 \quad \text{as} \quad N \to \infty,$$

that is $\tilde{S}^N$ tends to B in probability in $C[0,1]$ equipped with the uniform metric.

Since $\tilde{S}^N$ has the same law as $\mathcal{S}^N$, and since convergence in probability implies weak convergence, it follows that the law of $\mathcal{S}^N$ to Wiener measure. The proof is complete. □

6 Poisson random measures and Lévy processes

6.1 Basic definitions

In these notes, we have focused on two specific examples of stochastic processes in continuous time: the Poisson process and Brownian motion. Both of these processes have cadlag paths, and independent and stationary increments. It turns out that they are members of a rather general class of processes named Lévy processes.

A real-valued stochastic process $X = (X_t \colon t \geq 0)$, defined on some process space $(\Omega, \mathcal{F}, \mathbb{P})$, is called a *Lévy process* if

- $X_0 = 0$
- X has independent and stationary increments
- X is *stochastically continuous*: for every $\varepsilon > 0$ and $t \geq 0$,

$$\lim_{s \to t} \mathbb{P}(|X_t - X_s| > \varepsilon) = 0.$$

It is clear that standard Brownian motion satisfies these requirements, and so does the Poisson process. One can show (see [1, Chapter 2]) that any Lévy process has a cadlag modification that is itself a Lévy process.

Processes that enjoy the cadlag property are, in a way, rather well-behaved. For instance, we saw in Chapter 3 that such processes can only have a countable number of discontinuities. However, unlike Brownian motion, which has no discontinuities at all, and the Poisson process, which has jump discontinuities at random times, all of size 1, a "typical" Lévy process exhibits jump discontinuities whose sizes can vary considerably–from "infinitesimal" to large. A systematic discussion of Lévy processes is beyond the scope of this course, but this chapter will be devoted to a concept that is useful in the study of the jump processes, namely Poisson random measures.

Recall that each X_t, $t \geq 0$, is a random variable, and so the characteristic function is well-defined of X is defined for each $t > 0$. We end this preliminary discussion by quoting without proof a basic result in the theory of Lévy processes, namely the *Lévy-Khinchin formula*: the characteristic function of a Lévy process is of the form

$$\varphi_{X_t}(u) = e^{t\eta(u)}, \quad u \in \mathbb{R},$$

where

$$\eta(u) = ibu - a\frac{u^2}{2} + \int_{\mathbb{R}\setminus\{0\}} [e^{iux} - 1 - iux\mathbf{1}(|x| < 1)]d\nu(x). \tag{6.1}$$

In this expression, we have $b \in \mathbb{R}$ and $a > 0$, and ν is a so-called Lévy measure: a Borel measure having

$$\int_{\mathbb{R}\setminus\{0\}} (|x|^2 \wedge 1)d\nu(x) < \infty.$$

In general, it is not the case that Lévy measures are finite, or have finite first moment: it may well be that $\int |x| d\nu(x) = \infty$.

Looking closer at the formula (6.1), we recognize within the first two terms several probabilistic objects we have already encountered. The expression $\exp(ibtu)$ is the characteristic function of the deterministic process $X_t = bt$, while $\varphi_{B_t}(u) = \exp(-atu^2/2)$ corresponds to that of a Brownian motion, rescaled by the factor $a > 0$. Finally, assuming for a moment that ν does have finite first moment, we can absorb $\int iux d\nu(x) = iu \int x d\nu(x)$ into the term ibu, leaving the expression

$$\int_{\mathbb{R}\setminus\{0\}} [e^{iux} - 1] d\nu(x)$$

A computation along the lines of (3.9) shows that the Poisson process of intensity $\lambda > 0$ has

$$\varphi_{N_t}(u) = e^{\lambda t \left(e^{iu} - 1\right)}; \tag{6.2}$$

this coincides with the previous integral in the special case $\nu = \lambda \delta_1$. Conversely, the measure $\nu = \sum_{k=1}^{n} \lambda_k \delta_{x_k}$ yields precisely the characteristic function we would observe in a process X given as the sum of n independent Poisson processes of intensities λ_k and different jump sizes x_k.

This suggests the following, admittedly very heuristic, interpretation of the "unknown" term in the Lévy-Khinchin formula: it should arise from a superposition of many Poisson processes, all independent of the Brownian motion that yields $-au^2/2$, and with jump sizes that are real. (A clear weakness here is that we are assuming finite first moments.) In the next section, we shall discuss this kind of process.

The celebrated *Lévy-Ito decomposition* puts these heuristic observation on a rigorous footing, and provides a precise description of an arbitrary Lévy process as a sum of simpler independent processes, featuring several types of objects we have encountered, or are about to discuss: Brownian motion, Poisson processes, and martingales.

6.2 Poisson random measures

In the Poisson process, the increments $N_t - N_s$ are distributed acording to the $\mathrm{Po}(\lambda(t-s))$ distribution for $t > s$. Random variables having the Poisson distribution enjoy the *addition property*: if $(N_k)_{k=1}^{\infty}$ are independent and N_k has $\mathrm{Po}(\lambda_k)$-distribution, then

$$\sum_{k=1}^{n} N_k \sim \mathrm{Po}\left(\sum_{k=1}^{n} \lambda_k\right).$$

Moreover, the *splitting property* holds: if N has $\mathrm{Po}(\lambda)$, and $(X_n)_{n=1}^{\infty}$ are independent random variables with $\mathbb{P}(X_n = m) = p_m$, then

$$N_m = \sum_{k=1}^{N} \mathbf{1}(X_k = m) \sim \mathrm{Po}(\lambda p_m).$$

This suggests the following general definition.

Definition 6.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $(S, \mathcal{S}, \mu)$ be a σ-finite measure space. An *independently scattered random measure* on S is a map

$$M\colon \Omega \times \mathcal{S} \to \mathbb{N} \cup \{\infty\}$$

that satisfies, for any collection $(A_k)_{k=1}^{\infty}$ of disjoint sets in $\mathcal{S}$,

- $M(\varnothing) = 0$
- $M\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} M(A_k)$
- the random variables $(M(A_k))_{k=1}^{n}$ are independent

If the random variables $M(A_k)$ have distribution $\mathrm{Po}(\mu(A_k))$ we say that M is a *Poisson random measure* with intensity μ.

We let S^* denote the space of measures on $(S, \mathcal{S})$ taking values in the non-negative integers, or infinity, and define a mapping $X\colon S^* \to \mathbb{N} \cup \{\infty\}$ indexed by sets $A \in \mathcal{S}$:

$$X_A(m) = m(A), \quad m \in S^*.$$

We equip the space S^* with the σ-algebra $\mathcal{S}^* = \sigma(X_A\colon A \in \mathcal{S})$.

Theorem 6.2. *There exists a unique probability measure on $(S^*, \mathcal{S}^*)$ such that, under μ^*, the map X is a Poisson random measure with intensity measure μ.*

Proof. Suppose first that $\lambda = \mu(S) < \infty$. We consider a random variable N and a sequence of random variables $(X_n)_{n=1}^{\infty}$, defined on some common probability space and all independent, such that N is $\mathrm{Po}(\lambda)$ and X_n has law μ/λ. Then, by the splitting property, the prescription

$$M(A) = \sum_{n=1}^{N} \mathbf{1}(X_n \in A), \quad A \in \mathcal{S}, \tag{6.3}$$

defines a Poisson random measure with the desired intensity μ.

In the general case where μ is merely σ-finite, we let $(S_k)_{k=1}^{\infty}$ be disjoint sets with $\bigcup_k S_k = S$ and $\mu(S_k) < \infty$. On each S_k we can construct independent Poisson random measures M_k with respective intensities $\mu_k = \mu_{|S_k}$. We then define

$$M(A) = \sum_{k=1}^{\infty} M_k(A \cap S_k), \quad A \in \mathcal{S},$$

and verify that this produces a Poisson random measure. The law of M on S^* then yields the desired measure μ^* on S^*.

To establish uniqueness, we consider the π-system $\mathcal{A}^*$ consisting of sets of the form

$$A^* = \{m \in S^*\colon m(A_1) = n_1, \ldots, m(A_k) = n_k\}$$

where $(A_j)_{j=1}^k$ are disjoints sets in $\mathcal{S}^*$, and $n_1, \dots, n_k \in \mathbb{N}$. If ν^* is a measure that turns X into Poisson random measure, then

$$\mu^*(A^*) = \prod_{j=1}^{k} e^{-\mu(A_j)} \frac{\mu(A_j)^{n_j}}{(n_j)!},$$

and it follows that such a measure must coincide with μ^* since $\mathcal{S}^*$ is generated by the π-system $\mathcal{A}^*$. □

Theorem 6.3. *Suppose M is a Poisson random measure on $(S, \mathcal{S})$ with intensity measure μ, and let $f: S \to \mathbb{R}$ be square integrable with respect to μ. Then the random variable*

$$X = M(f) = \int_S f(x) dM(x)$$

is integrable, has characteristic function

$$\mathbb{E}[e^{iuM(f)}] = \exp\left(\int_S [e^{iuf(x)} - 1] d\mu(x) \right),$$

and moreover,

$$\mathbb{E}[M(f)] = \int_S f(x) d\mu(x) \quad \textit{and} \quad \mathrm{var}(M(f)) = \int_S [f(x)]^2 d\mu(x).$$

Proof. We again start with the case $\lambda = \mu(S) < \infty$. In this case $M(S)$ is finite. Starting with the simplest case $f = \mathbf{1}(A)$ for some $A \in \mathcal{S}$, we obtain the random variable

$$X(\omega) = \int_S \mathbf{1}(A) dM(x) = M(A)(\omega).$$

This readily extends to step functions of the form $f \sum_{k=1}^n c_k \mathbf{1}(A_k)$, where the A_k's are disjoint sets in $\mathcal{S}$, and $c_k \geq 0$: we set

$$X = \sum_{k=1}^{n} c_k M(A_k).$$

We then extend the definition to non-negative $f \in L^2(\mu)$ by approximating from below by step functions, and taking limits in L^2.

By uniqueness, we may assume that M is written in the form (6.3). Conditioning and using independence, we find that

$$\mathbb{E}[e^{iuX} | N = n] = (\mathbb{E}[e^{iuf(X_1)}])^n = \left(\int_S e^{iuf(x)} \frac{d\mu(x)}{\lambda} \right)^n.$$

Next, by the law of total expectation,

$$\mathbb{E}[e^{iuX}] = \sum_{n=0}^{\infty} \mathbb{E}[e^{iuX} | N = n] \mathbb{P}(N = n) = \sum_{n=0}^{\infty} \left(\int_S e^{iuf(x)} \frac{d\mu(x)}{\lambda} \right)^n e^{-\lambda} \frac{\lambda^n}{n!},$$

and the desired exponential expression follows upon summation, and using the fact that $\lambda = \int_S d\mu$.

The case where μ is not necessarily finite but f is integrable is dealt with by approximating f from below by functions having $\mu(|f_n| > 0) < \infty$, and using monotone convergence. Finally, we extend to general functions in the usual manner, by considering positive and negative parts.

The formulas for the expectation and variance are obtained by differentiating with respect to u and then setting $u = 0$. (See the Chapter on Weak Convergence.) □

6.3 Jumps of a Lévy process

We now introduce a notion of time into our discussion of Poisson random measures. Letting $(S, \mathcal{S}, \nu)$ be a σ-finite measure space, we consider the product measure μ on $(0, \infty) \times S$ determined by

$$\mu((0,t] \times A) = t\nu(A), \quad t > 0, A \in \mathcal{S}.$$

We now take M to be a Poisson random measure with intensity μ, and consider integrals of the form $Y_t = \int_{(0,t]\times S} f(x)dM(x)$ The process Y_t is cadlag, and a jump process–whereas the Poisson process jumps up by 1 at random times, the distribution of jumps of Y_t is governed by ν, the spatial part of the intensity measure μ.

We also introduce the *compensated Poisson random measure*

$$\tilde{M} = M - \mu.$$

The key point is that compensated Poisson random measures are centered, that is, have zero mean: if $A \subset \mathcal{S}$, then

$$\mathbb{E}[\tilde{M}((0,t] \times A)] = \mathbb{E}[M((0,t] \times A)] - t\nu(A) = 0.$$

Recall that the Poisson process is cadlag, but is not a martingale. However, by subtracting off the expected value of the process, we obtain the compensated Poisson process

$$\tilde{N}_t = N_t - \lambda t, \quad t \geq 0,$$

which does satisfy the martingale property. An analogous statement holds for integrals with respect to compensated Poisson random measures.

For $f \in L^2(\nu)$, we consider the process

$$X_t = \int_{(0,t]\times S} f(x)d\tilde{M}(x),$$

and its natural filtration $(\mathcal{F}_t)_{t\geq 0}$. We show that X enjoys the martingale property. Integrability and adaptedness is straight-forward, and if $s < t$, then since Poisson random measures are independently scattered,

$$\mathbb{E}[X_t|\mathcal{F}_s] = \mathbb{E}\left[\int_{(s,t]\times S} fd\tilde{M} + \int_{(0,s]\times S} fd\tilde{M}\Big|\mathcal{F}_s\right] = \int_{(0,s]\times S} fd\tilde{M} + \mathbb{E}\left[\int_{(s,t]\times S} fd\tilde{M}\right].$$

The second term on the right-hand side satisfies

$$\mathbb{E}\left[\int_{(s,t]\times S} f d\tilde{M}\right] = \mathbb{E}\left[\int_{(s,t]\times S} f dM\right] - (t-s)\int_S f(x)d\nu(x) = 0$$

by Theorem 6.3, and so

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s$$

as desired

We summarize our findings in the following theorem.

Theorem 6.4. *Let f be a ν-square integrable function on S, and define the process*

$$X_t = \int_{(0,t]\times S} f(x)d\tilde{M}(s,x).$$

Then $X = (X_t \colon t \geq 0)$ is a cadlag martingale with independent and stationary increments, and moreover

$$\mathbb{E}[e^{iuX_t}] = \exp\left(t\int_S [e^{iuf(x)} - 1 - iuf(x)]d\nu(x)\right)$$

and

$$\mathbb{E}[X_t^2] = t\int_S [f(x)]^2 d\nu(x).$$

This last theorem identifies the last summand in our heuristic decomposition of a Lévy process in the case $\int_{\mathbb{R}\setminus\{0\}} |x|d\nu(x) < \infty$. We see that it arises as an integral, with $f(x) = x$, against a compensated Poisson random measure, and is a cadlag L^2 martingale. To summarize our intuititive understanding up to this point, we want to think of a Lévy process with Lévy measure with finite first moment as

$$X_t = bt + aB_t + C_t,$$

where bt is a deterministic drift, B_t is a Brownian motion, and C_t is a cadlag martingale arising from a Poisson random measure, and all the processes are independent. For a rigorous discussion, including proofs of the Lévy-Khinchin formula in full generality (which requires a limiting procedure for the jump terms) and of the Lévy-Ito decomposition, we refer the reader to [1].

References

[1] D. Applebaum, Lévy processes and stochastic calculus, Cambridge studies in advanced mathematics, Cambridge University Press, 2009.

[2] R.F. Bass, Stochastic Processes, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2011.

[3] R. Durrett, Probability: Theory and examples, Cambridge University Press, 2010.

[4] L.C. Evans, Partial differential equations, Graduate Studies in Mathematics **19**, American Mathematical Society, 1998.

[5] I. Karatzas and S. Shreve, Brownian motion and stochastic calculus, Graduate texts in mathematics, Springer-Verlag New York, 1998

[6] P. Mörters and Y. Peres, Brownian motion, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2010.

[7] D. Williams, Probability with martingales, Cambridge Mathematical Textbooks, Cambridge University Press, 1991.