Advanced Probability Theory

Thierry Lévy *

Office A-314

Tsinghua University, Spring 2018

Contents

*Université Pierre et Marie Curie – LPMA – 4, place Jussieu – F-75005 Paris, thierry.levy@upmc.fr, www.proba.jussieu.fr/pageperso/levy/

References

[1] Patrick Billingsley. *Probability and Measure.* Wiley, 3rd edition, 1995.

[2] Joe Diestel. Uniform integrability: an introduction. *Rend. Istit. Mat. Univ. Trieste*, 23(1):41–80 (1993), 1991. School on Measure Theory and Real Analysis (Grado, 1991).

[3] Rick Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, fourth edition, 2010.

[4] James R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.

[5] David Williams. *Probability with martingales.* Cambridge University Press, 1991.

1 A reminder of elementary probability theory

1.1 Probability spaces

In order to describe an experiment from the point of view of probability theory, one considers a *probability space*, that is, a triple $(\Omega, \mathcal{A}, \mathbb{P})$, where

- Ω is a set: the set of all possible outcomes of the experiment,
- $\mathcal{A}$ is a σ-field on Ω: the set of subsets of Ω that it is possible to consider, to name, to think of,
- $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{A})$, which to each $A \in \mathcal{A}$ associates the probability $\mathbb{P}(A)$ that the outcome of the experiment belongs to A.

By definition, the fact that $\mathcal{A}$ is a σ-field means that $\mathcal{A}$ is a subset of the set $\mathcal{P}(\Omega)$ of all subsets of Ω (in other words, the elements of $\mathcal{A}$ are subsets of Ω) such that

- $\mathcal{A}$ contains the empty set $\varnothing$,
- for all $A \in \mathcal{A}$, the set $\Omega \setminus A$ also belongs to $\mathcal{A}$,
- for all sequence $(A_n)_{n\geq 0}$ of elements of $\mathcal{A}$, the union $\bigcup_{n\geq 0} A_n$ also belongs to $\mathcal{A}$.

The fact that $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{A})$ means that $\mathbb{P}$ is a function $\mathbb{P} : \mathcal{A} \to [0, 1]$ which satisfies

- $\mathbb{P}(\Omega) = 1$,
- for all sequence $(A_n)_{n\geq 0}$ of elements of $\mathcal{A}$ which are pairwise disjoint,

$$\mathbb{P}\left(\bigcup_{n\geq 0} A_n\right) = \sum_{n\geq 0} \mathbb{P}(A_n).$$

Exercise 1.1 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Prove that $\mathbb{P}(\varnothing) = 0$.*

The first fundamental example of probability space is $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, where Ω is a finite set, endowed with the σ-field of all subsets of Ω, and $\mathbb{P}$ is the uniform measure defined, for all $A \subset \Omega$, by $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$. Here, $|A|$ denotes the number of elements of A.

Another fundamental example is $([0, 1], \mathscr{B}_{[0,1]}, \lambda)$, where $\mathscr{B}_{[0,1]}$ is the σ-field of Borel subsets of the real interval $[0, 1]$, that is, the smallest σ-field on $[0, 1]$ which contains all open subsets of $[0, 1]$, and λ is the Lebesgue measure on $([0, 1], \mathscr{B}_{[0,1]})$, the unique measure on this measurable space which for all a, b such that $0 \leq a \leq b \leq 1$ satisfies $\lambda([a, b]) = b - a$.

Let us agree on the following question of terminology: some authors call a set countable if it has the cardinality of $\mathbb{N}$. For us, a countable set is a set which has the cardinality of a subset of $\mathbb{N}$. The difference is that for us, a finite set is countable.

Exercise 1.2 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. An element $\omega \in \Omega$ is called an* atom *of $\mathbb{P}$ if $\{\omega\} \in \mathcal{A}$ and $\mathbb{P}(\{\omega\}) > 0$. Can you give an upper bound to the number of atoms of $\mathbb{P}$ of mass at least $\frac{1}{10}$? Prove that the set of atoms of $\mathbb{P}$ is a countable subset of Ω. Prove that $\{\omega \in \Omega : \omega$ is not an atom of $\mathbb{P}\}$ belongs to $\mathcal{A}$.*

Exercise 1.3 *How many σ-fields does there exist on the set $\{1, 2, 3\}$? Define, for all $n \geq 0$, B_n as the number of σ-fields on a finite set with n elements. Prove that for all $n \geq 0$, the equality $B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$ holds. The number B_n is called the n-th Bell number.*

The most important technical property of a probability measure is the following. Consider a sequence $(A_n)_{n\geq 0}$ of elements of $\mathcal{A}$ which is non-decreasing in the sense that $A_0 \subset A_1 \subset A_2 \subset \ldots$. Then

$$\mathbb{P}\left(\bigcup_{n\geq 0} A_n\right) = \sup\{\mathbb{P}(A_n) : n \geq 0\} = \lim_{n\to\infty} \mathbb{P}(A_n). \tag{1}$$

Exercise 1.4 *Why does the limit on the right-hand side exist ? Prove the two equalities.*

Exercise 1.5 *Prove that for any non-increasing sequence $(A_n)_{n\geq 0}$ of elements of $\mathcal{A}$ one has $\mathbb{P}(\bigcap_{n\geq 0} A_n) = \lim_{n\to\infty} \mathbb{P}(\mathcal{A}_n)$? Is this true on any measure space ? Is (1) true on any measure space ?*

Another important property is this one: for all sequence $(A_n)_{n\geq 0}$ of elements of $\mathcal{A}$, the inequality

$$\mathbb{P}\left(\bigcup_{n\geq 0} A_n\right) \leq \sum_{n\geq 0} \mathbb{P}(A_n)$$

holds.

Yet another important result is formulated in the next exercise.

Exercise 1.6 (Borel-Cantelli lemma) *Let $(A_n)_{n\geq 0}$ be a sequence of elements of $\mathcal{A}$. How would you describe in words the set $S = \bigcap_{p\geq 0} \bigcup_{n\geq p} A_n$? Prove that S belongs to $\mathcal{A}$. Assume now that $\sum_{n\geq 0} \mathbb{P}(A_n) < \infty$ and compute $\mathbb{P}(S)$.*

Exercise 1.7 *Prove that for any family $(A_i)_{i\in I}$ of elements of $\mathcal{A}$ which are pairwise disjoint and such that $\bigcup_{i\in I} A_i$ belongs to $\mathcal{A}$, the inequality*

$$\mathbb{P}\left(\bigcup_{i\in I} A_i\right) \geq \sum_{i\in I} \mathbb{P}(A_i)$$

holds, where the right-hand side is defined by

$$\sum_{i\in I} \mathbb{P}(A_i) = \sup\left\{\sum_{i\in F} \mathbb{P}(A_i) : F \subset I, F \text{ finite}\right\}.$$

Note that the index set I may be uncountable. Can you think of an example where the inequality is strict ?

Exercise 1.8 *Let $(\Omega, \mathcal{A})$ be a measurable space. Prove that a function $\mathbb{P} : \mathcal{A} \to [0,1]$ is a probability measure if and only if the following conditions hold :*

- $\mathbb{P}(\varnothing) = 0$ *and* $\mathbb{P}(\Omega) = 1$,
- *for all* $A, B \in \mathcal{A}$, $\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \cup B) + \mathbb{P}(A \cap B)$,
- *for all sequence* $(A_n)_{n \geq 0}$ *of elements of* $\mathcal{A}$,

$$\mathbb{P}\left(\bigcup_{p \geq 0} \bigcap_{n \geq p} A_n \right) \leq \liminf_{n \to \infty} \mathbb{P}(A_n).$$

Is it possible to remove the assumption that $\mathbb{P}(\varnothing) = 0$ *from this list ?*

1.2 Independence

Let us now turn to the concept of independence. Two events A and B are independent (with respect to $\mathbb{P}$) if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Exercise 1.9 *What can you say about an event which is independent of itself ?*

In order to define the independence of more than two events, we need to be more careful. We say that $A_1, \ldots, A_n$ are independent if for all $k \in \{2, \ldots, n\}$ and all choice of $1 \leq i_1 < \ldots < i_k \leq n$, the equality $\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k})$ holds.

Exercise 1.10 *Find three events A, B, C such that $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ but A and B are not independent. Can you find an example where $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) > 0$?*

It turns out that the best definition is in terms of σ-fields. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that n sub-σ-fields $\mathcal{B}_1, \ldots, \mathcal{B}_n$ of $\mathcal{A}$ are independent if for all choice of events $B_1 \in \mathcal{B}_1, \ldots, B_n \in \mathcal{B}_n$, the equality $\mathbb{P}(B_1 \cap \ldots \cap B_n) = \mathbb{P}(B_1) \ldots \mathbb{P}(B_n)$ holds. We say that an arbitrary family $(\mathcal{B}_i)_{i \in I}$ of sub-σ-fields of $\mathcal{A}$ is independent if any finite sub-family $\mathcal{B}_{i_1}, \ldots, \mathcal{B}_{i_n}$ is independent.

Exercise 1.11 *Let $\mathcal{C}$ be a subset of $\mathcal{P}(\Omega)$. Prove that there exists a unique σ-field $\mathcal{A}$ on Ω such that $\mathcal{C} \subset \mathcal{A}$ and, for all σ-field $\mathcal{B}$ on Ω such that $\mathcal{C} \subset \mathcal{B}$, one has $\mathcal{A} \subset \mathcal{B}$. In words, $\mathcal{A}$ is the smallest σ-field on Ω which contains $\mathcal{C}$. It is called the σ-field generated by $\mathcal{C}$ and it is denoted by $\sigma(\mathcal{C})$.*

For example, the Borel σ-field of $\mathbb{R}$ is the σ-field generated by the class of open subsets of $\mathbb{R}$.

Compute $\sigma(\varnothing)$ and, for all $A \in \mathcal{A}$, $\sigma(\{A\})$. Let $\Omega = A_1 \cup \ldots \cup A_n$ be a partition of Ω. This means that the events $A_1, \ldots, A_n$ are non-empty and pairwise disjoint. Describe $\sigma(\{A_1, \ldots, A_n\})$. What can you say if instead of a finite partition we consider a countable partition $\Omega = \bigcup_{n \geq 0} A_n$? An arbitrary partition $\Omega = \bigcup_{i \in I} A_i$? What is the σ-field on $\mathbb{R}$ generated by $\{\{t\} : t \in \mathbb{R}\}$?

In the following exercise, you will show that our previous two definitions of independence are consistent.

Exercise 1.12 *Let $A_1, \dots, A_n$ be n events. Prove that it is equivalent to say that the events $A_1, \dots, A_n$ are independent or to say that the σ-fields $\sigma(\{A_1\}), \dots, \sigma(\{A_n\})$ are independent.*

The following example warns us against an easily made mistake.

Exercise 1.13 *Let us toss a coin twice. Consider the events "the first coin gives tail", "the second coin gives tail", "the two coins give the same result". Prove that any two of these three events are independent, but the three together are not independent.*

Exercise 1.14 *For $n \geq 2$, set $\Omega = \{\omega = (\omega_1, \dots, \omega_n) \in \{-1, 1\}^n : \omega_1 \dots \omega_n = 1\}$ and consider the uniform probability $\mathbb{P}$ on $(\Omega, \mathcal{P}(\Omega))$. For each $i \in \{1, \dots, n\}$, define $A_i = \{\omega \in \Omega : \omega_i = 1\}$. Compute, for all $k \in \{1, \dots, n\}$ and all $1 \leq i_1 < \dots < i_k \leq n$, the probability $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})$. What is this an example of ? ("Of a silly exercise" is not the expected answer.)*

1.3 Random variables

A random variable is the mathematical notion that represents a quantity whose value depends on the outcome of the experiment which is performed. Formally, a (real-valued) random variable on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is a measurable function $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$. Measurable means by definition that for all $B \in \mathscr{B}_{\mathbb{R}}$, the set $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ belongs to $\mathcal{A}$.

For example, the function $X : [0, 1] \to \mathbb{R}$ defined by $X(t) = t^2 - 1$ is a random variable on the probability space $([0, 1], \mathscr{B}_{[0,1]}, \lambda)$. Random variables are in fact rarely defined by formulas like this one, and this example is rather atypical.

Exercise 1.15 *Consider a random variable $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$. Prove that $\{X^{-1}(B) : B \in \mathscr{B}_{\mathbb{R}}\}$ is a sub-σ-field of $\mathcal{A}$. It is called the σ-field generated by X and it is denoted by $\sigma(X)$.*

Exercise 1.16 *Let $X : \Omega \to \mathbb{R}$ be a function. For all $x \in \mathbb{R}$, consider the subset*

$$\{X \leq x\} = X^{-1}((-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\}$$

of Ω. Prove that X is a random variable if and only if for all $a \in \mathbb{R}$, one has $\{X \leq a\} \in \mathcal{A}$.

It is sometimes convenient to allow infinite values for random variables. To do so, one considers the extended real line $\overline{\mathbb{R}} = [-\infty, +\infty]$. This set is endowed with the σ-field $\mathscr{B}_{\overline{\mathbb{R}}} = \sigma(\mathscr{B}_{\mathbb{R}} \cup \{\{-\infty\}, \{+\infty\}\})$.

Exercise 1.17 *Prove that a subset A of $\overline{\mathbb{R}}$ belongs to $\mathscr{B}_{\overline{\mathbb{R}}}$ if and only if $A \cap \mathbb{R}$ belongs to $\mathscr{B}_{\mathbb{R}}$.*

If you are familiar with the abstract notion of a topology on a set: is $\mathscr{B}_{\overline{\mathbb{R}}}$ the Borel σ-field of a topology on $\overline{\mathbb{R}}$, that is, the σ-field generated by the class of all open subsets of $\overline{\mathbb{R}}$?

Just as for events, we say that n random variables $X_1, \ldots, X_n$ defined on the same probability space are independent if the σ-fields $\sigma(X_1), \ldots, \sigma(X_n)$ are independent. More generally, we say that a family $(X_i)_{i \in I}$ of random variables is independent if every finite sub-family $X_{i_1}, \ldots, X_{i_n}$ is independent.

Exercise 1.18 *Let A be an event. Prove that the indicator function $\mathbb{1}_A$ defined by $\mathbb{1}_A(\omega) = 1$ if $\omega \in A$ and $\mathbb{1}_A(\omega) = 0$ otherwise, is a random variable. Prove that it is equivalent to say that the events $A_1, \ldots, A_n$ are independent or to say that the random variables $\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_n}$ are independent.*

Exercise 1.19 *What can you say about a random variable which is independent of itself?*

Exercise 1.20 *Is it equivalent to say that a finite family of random variables are pairwise independent and to say that they are independent ?*

Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ be a random variable. The function $\mathbb{P} \circ X^{-1} : \mathscr{B}_{\mathbb{R}} \to [0, 1]$ defined by $\mathbb{P} \circ X^{-1}(B) = \mathbb{P}(X^{-1}(B))$ is a probability measure on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$.

Exercise 1.21 *Prove this assertion.*

The probability measure $\mathbb{P} \circ X^{-1}$ is called the distribution, or the law, of X. It is also often denoted by $\mathbb{P}_X$. The line

$$\mathbb{P} \circ X^{-1}(B) = \mathbb{P}_X(B) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}) = \mathbb{P}(\{X \in B\}) = \mathbb{P}(X \in B)$$

gives five common notations for the same quantity. Let us emphasize that $\{X \in B\}$ is a notation for $\{\omega \in \Omega : X(\omega) \in B\}$, that is, for $X^{-1}(B)$.

The distribution of a real-valued random variable is thus a probability measure on the real line. The following rough classification of these probability measures is often used. We say that the distribution of X is discrete, or atomic, if there exists a sequence $x_1, x_2, \ldots$ of atoms of $\mathbb{P}_X$, finite or infinite (but in any case countable, see Exercise 1.2), such that $\mathbb{P}_X(\{x_1\}) + \mathbb{P}_X(\{x_2\}) + \ldots = 1$. In other words, there exists a countable subset $C \subset \mathbb{R}$ such that $\mathbb{P}(X \in C) = 1$. The typical cases are $C = \mathbb{N}$ and $C = \mathbb{Z}$, and we then speak of integer-valued random variables, but there are also other interesting cases of discrete random variables. In this course, we will define and use many random variables with values in $\mathbb{N} \cup \{\infty\}$.

The distribution of a discrete random variable is completely described by a countable set $C = \{x_1, x_2, \ldots\}$ such that $\mathbb{P}(X \in C) = 1$ and the data, for each $i \geq 1$, of the probability $p_i = \mathbb{P}(X = x_i)$. It is the framework of elementary probability theory, especially when C is finite.

A random variable X such that $\mathbb{P}_X$ has no atom, that is, such that $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$, is said to be diffuse.

Exercise 1.22 *Find a diffuse real-valued random variable. Find a real-valued random variable which is neither diffuse nor discrete.*

Among diffuse random variables, there is the practically very important class of random variables for which $\mathbb{P}_X$ is absolutely continuous with respect to the Lebesgue measure. This means that for all Borel subset N of $\mathbb{R}$ such that $\lambda(N) = 0$, one has $\mathbb{P}(X \in N) = 0$. In this case, since we are working with probability measures, which are σ-finite, the Radon-Nikodym theorem can be applied to produce a non-negative measurable function $f : (\mathbb{R}, \mathscr{B}_{\mathbb{R}}) \to (\mathbb{R}_+, \mathscr{B}_{\mathbb{R}_+})$ which has the property that for all Borel subset B of $\mathbb{R}$, the equality

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \int_B f \, d\lambda = \int_{\mathbb{R}} f \mathbb{1}_B \, d\lambda$$

holds. This situation is extremely convenient, because the computation of probabilities is reduced to the computation of ordinary integrals with respect to the Lebesgue measure.

Our quick zoology of probability measures on $\mathbb{R}$ is summarised by Figure 1 below.

Exercise 1.23 *Let $f : \mathbb{R} \to \mathbb{R}_+$ be a measurable function. Prove that the function $\mu : \mathcal{A} \to [0, +\infty]$ defined by $\mu(A) = \int_A f \, d\lambda$ is a measure on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$. We shall write $\mu = f\lambda$. Under which condition on f is μ a probability measure ?*

On $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$, consider the Lebesgue measure λ and the counting measure κ, which by definition is such that $\kappa(A)$ is the number of elements of A if A is finite, and $\kappa(A) = +\infty$ if A is infinite. Prove that λ is absolutely continuous with respect to κ. Prove that there does not exist a measurable function $f : \mathbb{R} \to \mathbb{R}_+$ such that $\lambda = f\kappa$. Is this not in contradiction with the Radon-Nikodym theorem ?

Exercise 1.24 *Prove that every probability measure on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ can be written in a unique way as the sum of a discrete measure and a diffuse measure.*

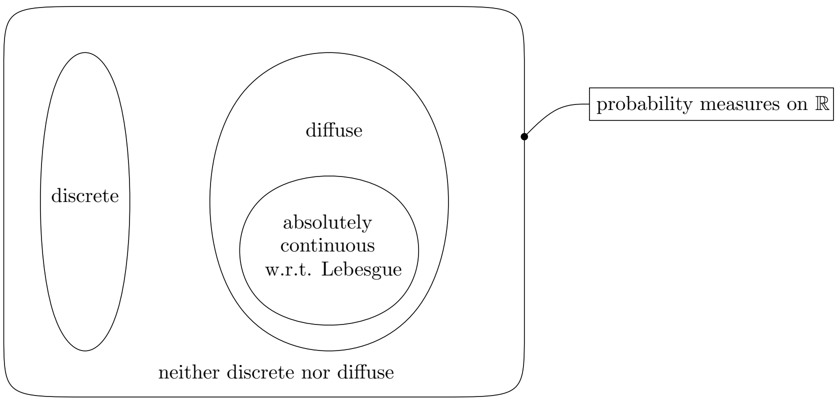


Figure 1: Do you agree with this partition ? Can you find a probability measure in each class ?

1.4 Integration of random variables

Consider a real-valued random variable X. If X admits an integral on $(\Omega, \mathcal{A}, \mathbb{P})$ in the sense of Lebesgue's integration theory, we say that X admits an expectation, which is defined by

$$\mathbb{E}[X] = \int_\Omega X \, d\mathbb{P}.$$

The following exercises help you to remember the key points of the theory of integration.

Exercise 1.25 (Definition of the integral) *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Recall the definition of the integral with respect to $\mathbb{P}$ of a non-negative measurable function $X : (\Omega, \mathcal{A}) \to (\mathbb{R}_+, \mathscr{B}_{\mathbb{R}_+})$, for example as a supremum of elementary integrals of simple functions. Consider an arbitrary measurable real-valued function $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$. Define $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$, so that $X = X^+ - X^-$. Explain how to define the integral of X if one at least of the two functions X^+ and X^- have a finite integral. What is the integral of X if both X^+ and X^- have infinite integrals ? When is it the case that $|X|$ has a finite integral ? (In this case, we say that X is integrable).*

Let us emphasize that a non-negative random variable always admits an expectation, possibly equal to $+\infty$.

Exercise 1.26 (Convergence theorems) *Recall the statement of the monotone convergence theorem. Does the theorem also hold for a decreasing sequence of functions ? Deduce Fatou's lemma from the monotone convergence theorem and then the dominated convergence theorem from Fatou's lemma. (You may find that this is not as difficult as you expect or remember. The monotone convergence theorem is the one most important convergence theorem of the theory of integration, and the other convergence theorems follow relatively easily from it.)*

A few inequalities are of crucial importance. The simplest one, but very useful, is the Markov inequality. It says that for a non-negative random variable X and a non-negative real a, one has

$$a\mathbb{P}(X \geq a) \leq \mathbb{E}[X].$$

Exercise 1.27 *Prove the Markov inequality.*

The Hölder inequality states that if X and Y are non-negative random variables and if $p, q > 1$ are two real numbers such that $\frac{1}{p} + \frac{1}{q} = 1$, then

$$\mathbb{E}[XY] \leq \mathbb{E}[X^p]^{\frac{1}{p}} \mathbb{E}[Y^q]^{\frac{1}{q}}.$$

The case where $p = q = 2$ is an instance of the Cauchy-Schwarz inequality.

One says that a random variable X admits a moment of order $p \geq 1$ if the random variable $|X|^p$ is integrable.

Exercise 1.28 *Prove that if* $1 \leq p < p'$ *and if a random variable* X *admits a moment of order* p'*, then it also admits a moment of order* p*. Find an explicit bound for* $\mathbb{E}[|X|^p]$*.*

Exercise 1.29 *Let* X *be a non-negative random variable. Prove that the function from* $[1,\infty)$ *to* $[0,\infty]$ *defined by* $p \mapsto \log \mathbb{E}[X^p]$ *is convex on* $[1,\infty)$ *(one says that the function* $p \mapsto \mathbb{E}[X^p]$ *is log-convex). Check that the statement and your proof of it make sense and are correct even if some or all moments of* X *are infinite.*

Exercise 1.30 *Let* X *be a non-negative random variable. With the usual agreement that* $\inf \varnothing = +\infty$*, define* $r = \inf\{p \in (0,\infty) : \mathbb{E}[X^p] = \infty\}$*. Prove by giving examples that* r *can be any element of* $[0,\infty]$*. Is it true that* $r = \infty$ *implies that* X *is bounded ? When* $0 < r < \infty$*, is it always true, sometimes true and sometimes false, or never true that* $\mathbb{E}[X^r] < \infty$ *?*

Another important inequality is Minkowski's inequality. It states that for all $p \geq 1$ and for any two non-negative random variables X and Y, the inequality

$$\mathbb{E}[(X+Y)^p]^{\frac{1}{p}} \leq \mathbb{E}[X^p]^{\frac{1}{p}} + \mathbb{E}[Y^p]^{\frac{1}{p}}$$

holds. In particular, the sum of two random variables which admit a moment of order p also admits a moment of order p.

This suggests to define, for all real $p \geq 1$, the set $L^p(\Omega, \mathcal{A}, \mathbb{P})$ as the set of real-valued random variables on Ω which admit a moment of order p, quotiented by the subspace formed by the random variables X such that $\mathbb{E}[|X|^p] = 0$.

Exercise 1.31 *Check that* $L^p(\Omega, \mathcal{A}, \mathbb{P})$ *is a vector space and that the function* $X \mapsto \|X\|_p = \mathbb{E}[|X|^p]^{\frac{1}{p}}$ *is a norm on* $L^p(\Omega, \mathcal{A}, \mathbb{P})$*.*

Exercise 1.32 *Let* X *be a random variable. With the agreement that* $\log 0 = -\infty$ *and* $\log(+\infty) = +\infty$*, prove that*

$$\lim_{p\to\infty} \frac{1}{p} \log \mathbb{E}[|X|^p] = \log \|X\|_\infty,$$

where the norm on the right-hand side is defined by

$$\|X\|_\infty = \inf\{M \in [0,+\infty] : \mathbb{P}(|X| \leq M) = 1\}.$$

Is it true that $\lim_{p\to\infty} \|X\|_p = \|X\|_\infty$*?*

It is a very important fact that the normed vector space $(L^p(\Omega, \mathcal{A}, \mathbb{P}), \|\cdot\|_p)$ is complete. In particular, $(L^2(\Omega, \mathcal{A}, \mathbb{P}), \|\cdot\|_2)$ is a Hilbert space, since the norm $\|\cdot\|_2$ is induced by the scalar product $\langle X, Y\rangle = \mathbb{E}[XY]$.

1.5 Convergence of random variables

Let us conclude this reminder of classical probability theory by defining three notions of convergence of random variables. Let $(X_n)_{n\geq 0}$ be a sequence of real-valued random variables defined on $(\Omega, \mathcal{A}, \mathbb{P})$. Let X be a random variable on the same probability space.

We say that the sequence $(X_n)_{n\geq 0}$ converges almost surely towards X if there exists an event $N \subset \Omega$ such that $\mathbb{P}(N) = 0$ and for all $\omega \notin N$, the sequence $(X_n(\omega))_{n\geq 0}$ converges to $X(\omega)$. This is the pointwise convergence outside a negligible event.

Let $p \geq 1$ be fixed. We say that the sequence $(X_n)_{n\geq 0}$ converges towards X in L^p if X admits a moment of order p and $\|X_n - X\|_p$ converges to 0. This is the norm convergence in the Banach space $L^p(\Omega, \mathcal{A}, \mathbb{P})$.

We say that the sequence $(X_n)_{n\geq 0}$ converges in probability towards X if for all $\varepsilon > 0$, the sequence of probabilities $\mathbb{P}(|X_n - X| > \varepsilon)$ converges to 0 as n tends to infinity.

The next exercises are perhaps more substantial than some of the earlier ones.

Exercise 1.33 (Modes of convergence) *Prove that a sequence which converges almost surely or in L^p for some $p \geq 1$ also converges in probability towards the same limit. Prove that a sequence of random variables has at most one limit in any of the three modes of convergence which we have defined.*

Prove that from a sequence which converges in probability one can extract a sub-sequence which converges almost surely.

Exercise 1.34 (L^p is complete) *Let $(X_n)_{n\geq 1}$ be a Cauchy sequence in the normed vector space $(L^p(\Omega, \mathcal{A}, \mathbb{P}), \|\cdot\|_p)$.*

Prove that there exists a sub-sequence $(X_{n_k})_{k\geq 1}$ of $(X_n)_{n\geq 1}$ such that for all $k \geq 1$ and for all $l \geq k$, one has $\|X_{n_k} - X_{n_l}\|_p^p \leq 2^{-k}$. Prove that for almost all $\omega \in \Omega$, the sequence $(X_{n_k}(\omega))_{k\geq 1}$ is a Cauchy sequence. Denote by $X(\omega)$ its limit.

Use Fatou's lemma to prove that X belongs to L^p and that the sequence $(X_{n_k})_{k\geq 1}$ converges in L^p to X. Prove that the sequence $(X_n)_{n\geq 1}$ converges in L^p to X.

The fact that L^p is complete allows one to use the machinery of functional analysis. In particular, L^2 is a Hilbert space as we already mentioned. This has at least two important consequences.

Firstly, given any closed linear subspace F of $L^2(\Omega, \mathcal{A}, \mathbb{P})$, the orthogonal subspace $F^\perp$ is a closed subspace which satisfies $F \oplus F^\perp = L^2(\Omega, \mathcal{A}, \mathbb{P})$. In particular, there exists an orthogonal projection p_F whose image is exactly F.

Secondly, any continuous linear form on $L^2(\Omega, \mathcal{A}, \mathbb{P})$ is of the form $Y \mapsto \mathbb{E}[XY]$ for some (uniquely defined) $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. This is called the Riesz representation theorem.

Exercise 1.35 (Radon-Nikodym theorem) *Let $(\Omega, \mathcal{A})$ be a measurable space. Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, \mathcal{A})$. Assume that $\mathbb{Q} \ll \mathbb{P}$, that is, $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$, which means that for all $A \in \mathcal{A}$, $\mathbb{P}(A) = 0$ implies $\mathbb{Q}(A) = 0$.*

Define the measure $\mu = \mathbb{P} + \mathbb{Q}$ on $(\Omega, \mathcal{A})$. Note that this is not a probability measure. Prove that $L^2(\Omega, \mathcal{A}, \mu)$ is a subspace of $L^1(\Omega, \mathcal{A}, \mathbb{Q})$ and that the mapping $\mathbb{E}_{\mathbb{Q}}$:

$L^2(\Omega, \mathcal{A}, \mu) \to \mathbb{R}$ defined by $\mathbb{E}_{\mathbb{Q}}[X] = \int_\Omega X \, d\mathbb{Q}$ is a continuous linear form. Deduce that there exists $\delta \in L^2(\Omega, \mathcal{A}, \mu)$ such that for all $X \in L^2(\Omega, \mathcal{A}, \mu)$, one has $\mathbb{E}_{\mathbb{Q}}[X] = \int_\Omega X\delta \, d\mu$.

Prove that $\delta \in [0, 1)$ $\mathbb{P}$-almost surely. Set $D = \frac{\delta}{1-\delta}$. Prove that for all non-negative $X : \Omega \to \mathbb{R}_+$, one has $\int_\Omega X \, dQ = \int_\Omega XD \, dP$. Prove that the same equality holds for any $X \in L^1(\Omega, \mathcal{A}, \mathbb{Q})$.

2 Conditional expectation

There are three main parts in this chapter. The first is the definition of conditional expectation. We will make a lot of comments about it, in order to help you to relate this rather abstract definition to what you already know. The second part is the proof that conditional expectation exists and is unique. Uniqueness is easy, but in my opinion more instructive than many books indicate. Existence is not easy and is commonly proved in one of two different ways. We explain one, the other can for example be found in Durrett's book. The third part is a list of properties which will allow you to manipulate in practical computations the concept of conditional expectation. My own experience of mathematics is that learning a new notion involves repeatedly going from applying rules without questioning them too much and watching them at work on one hand, and thinking about the structure of the theory and going deeper into the understanding of the rules on the other hand.

2.1 Definition and fundamental examples

Definition 2.1 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$, that is, a σ-field on Ω such that $\mathcal{B} \subset \mathcal{A}$. Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ be an integrable random variable.*

A conditional expectation of X given $\mathcal{B}$ *is an integrable real-valued random variable $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ such that the following two conditions hold.*

1. Y is measurable with respect to $\mathcal{B}$.

2. For all $B \in \mathcal{B}$, one has $\int_B X \, d\mathbb{P} = \int_B Y \, d\mathbb{P}$.

We shall prove that a conditional expectation of X given $\mathcal{B}$ always exists and is unique almost surely, and we shall denote it by $\mathbb{E}[X|\mathcal{B}]$.

Exercise 2.1 *Assume that $\mathcal{B} = \{\varnothing, \Omega\}$. Prove that a conditional expectation of X given $\mathcal{B}$ must be a constant random variable and find the unique possible value of this constant.*

Let us start by an example. Suppose that $\mathcal{B}$ is generated by an event $C \in \mathcal{A}$, so that $\mathcal{B} = \{\varnothing, C, C^c, \Omega\}$. Let us assume that $\mathbb{P}(C)$ is neither 0 nor 1. If a conditional expectation of X given $\mathcal{B}$ exists, it must be of the form

$$Y = \alpha \mathbb{1}_C + \beta \mathbb{1}_{C^c},$$

because this is (up to modification on a negligible set) the form of the most general $\mathcal{B}$-measurable random variable. The second condition which Y must satisfy applied with $B = C$ yields

$$\alpha\mathbb{P}(C) = \int_C Y\, d\mathbb{P} = \int_C X\, d\mathbb{P},$$

so that

$$\alpha = \frac{1}{\mathbb{P}(C)} \int_C X\, d\mathbb{P}.$$

Similarly,

$$\beta = \frac{1}{\mathbb{P}(C^c)} \int_{C^c} X\, d\mathbb{P}.$$

The second condition is satisfied for $B = \varnothing$ and a short computation shows that, with the choices of α and β above, it is also satisfied for $B = \Omega$. We have thus proved that a conditional expectation of X given $\mathcal{B}$ exists and is unique. It is equal to

$$\mathbb{E}[X|\mathcal{B}] = \left(\frac{1}{\mathbb{P}(C)} \int_C X\, d\mathbb{P}\right) \mathbb{1}_C + \left(\frac{1}{\mathbb{P}(C^c)} \int_{C^c} X\, d\mathbb{P}\right) \mathbb{1}_{C^c}.$$

We see the connection with elementary conditional probabilities: the value of $\mathbb{E}[X|\mathcal{B}]$ on C is nothing but the expectation of X under the conditional probability $\mathbb{P}^C = \mathbb{P}(\cdot|C)$ defined on $\mathcal{A}$ by

$$\mathbb{P}^C(A) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)}.$$

Let us try to formulate this with words. We know what the expectation of X with respect to the conditional probability $\mathbb{P}^C$ is, as well as its expectation with respect to the conditional probability $\mathbb{P}^{C^c}$. These are two real numbers. Now the conditional expectation $\mathbb{E}[X|\mathcal{B}]$ is a random variable on Ω, a real-valued function on Ω. Consider a point ω in Ω. Think of the σ-field $\mathcal{B}$ as encoding the information to which we have access. What do we know about ω ? We know whether it belongs to C or to C^c, and no more. If it belongs to C, then the conditional expectation of X given $\mathcal{B}$ evaluated at ω is equal to the expectation of X with respect to the conditional probability $\mathbb{P}^C$. If it belongs to C^c, the same conclusion holds with C replaced by C^c.

What we have done with two sets C and C^c can be generalised to an arbitrary finite or countably infinite number of sets. This is the object of the next exercise.

Exercise 2.2 *Let $\{C_1, C_2, \ldots\}$ be a partition of Ω into at least two (and possibly countably infinitely many) disjoint events with positive probability. Let $\mathcal{B} = \sigma(\{C_1, C_2, \ldots\})$ be the σ-algebra generated by this partition. Describe $\mathcal{B}$. Prove that $\mathbb{E}[X|\mathcal{B}]$ exists and is unique, and is given by the formula*

$$\mathbb{E}[X|\mathcal{B}] = \sum_{i \geq 1} \left(\frac{1}{\mathbb{P}(C_i)} \int_{C_i} X\, d\mathbb{P}\right) \mathbb{1}_{C_i}.$$

This first discussion revealed the relation between the definition of $\mathbb{E}[X|\mathcal{B}]$ and the elementary notion of conditional probability. However, it is not general enough for many purposes, and cannot be generalised to the case of an arbitrary sub-σ-field $\mathcal{B}$. Indeed, many interesting σ-fields are not generated by a partition.

Exercise 2.3 *Consider on $\mathbb{R}$ the Borel σ-field $\mathscr{B}_{\mathbb{R}}$ and define*

$$\mathcal{C} = \{C \in \mathscr{B}_{\mathbb{R}} : C + 2\pi = C\},$$

where for any subset C of $\mathbb{R}$, we denote by $C+2\pi$ the result of the translation of C by 2π, that is, $C + 2\pi = \{x + 2\pi : x \in C\}$. Prove that $\mathcal{C}$ is a sub-σ-field of $\mathscr{B}_{\mathbb{R}}$. Is $\mathcal{C}$ generated by a partition of $\mathbb{R}$?

Exercise 2.4 *Prove that every σ-field on a countable set is generated by a partition. Is it true that if a set Ω has the property that every σ-field on Ω is generated by a partition, then Ω is countable ?*

There is another situation where we can study by hand the existence of a conditional expectation. It is a case where, among other things that we will explain in a minute, we assume that $\mathcal{B}$ is generated by a real random variable, say $Z : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$. Recall from Exercise 1.15 the definition of $\sigma(Z)$, the σ-field generated by Z.

Let us make the assumption that the random vector (X, Z) has a distribution which admits a density $f_{(X,Z)}$ with respect to the Lebesgue measure on $\mathbb{R}^2$. This means that for all Borel subset D of $\mathbb{R}^2$, we have

$$\mathbb{P}((X, Z) \in D) = \int_D f_{(X,Z)}(u, v)\, dudv.$$

In this situation, we know that the distribution of Z admits a density with respect to the Lebesgue measure on $\mathbb{R}$ which is the function f_Z given (for almost every $v \in \mathbb{R}$) by

$$f_Z(v) = \int_R f_{(X,Z)}(u, v)\, du.$$

Now let us look for $\mathbb{E}[X|\sigma(Z)]$, which is also denoted by $\mathbb{E}[X|Z]$. It is useful to understand what it means for a random variable to be measurable with respect to $\sigma(Z)$.

Exercise 2.5 *In this exercise, we work with sets and maps between sets, without σ-fields and without any notion of measurability. Let Ω be a set. Let $Y, Z : \Omega \to \mathbb{R}$ be two maps. Prove that Y is constant on every non-empty set of Ω of the form $Z^{-1}(\{t\})$, $t \in \mathbb{R}$, if and only if there exists a map $h : \mathbb{R} \to \mathbb{R}$ such that $Y = h \circ Z$.*

Proposition 2.2 *Let Y, Z be real-valued random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The random variable Y is measurable with respect to $\sigma(Z)$ if and only if there exists a Borel measurable function $h : \mathbb{R} \to \mathbb{R}$ such that $Y = h(Z) = h \circ Z$.*

The proof of this proposition is not essential for our purposes. I give it here for the sake of completeness and pleasure.

Proof. If $Y = h(Z)$ for some measurable function h, then for all $B \in \mathscr{B}_{\mathbb{R}}$, we have $Y^{-1}(B) = (h \circ Z)^{-1}(B) = Z^{-1}(h^{-1}(B)) \in \sigma(Z)$, so that Y is measurable with respect to $\sigma(Z)$.

The most interesting part is the converse. Assume that Y is measurable with respect to $\sigma(Z)$. Let us also assume, for a start, that Y is non-negative.

We shall use the binary expansion of real numbers:

$$14.5 \text{ (decimal)} = 1110.1 \text{ (binary)}.$$

Let us observe two things. Firstly, in the binary expansion of a real x, the digit just to the left of the dot, the 0 in the example above, is the parity of the integer part of x. Let us write it $\lfloor x \rfloor \bmod 2$. Secondly, for all $n \in \mathbb{Z}$, the n-th digit of the binary expansion of x, that is, the digit which is the coefficient of 2^n, is the digit located just to the left of the dot in the expansion of $2^{-n}x$. Putting these remarks together, we find that

$$x = \sum_{n \in \mathbb{Z}} 2^{-n}((2^n x) \bmod 2).$$

We can write the same formula for a non-negative random variable

$$Y = \sum_{n \in Z} 2^{-n}((2^n Y) \bmod 2) = \lim_{N \to +\infty} \sum_{n=-N}^{N} 2^{-n}((2^n Y) \bmod 2).$$

We use now the assumption. Since Y is $\sigma(Z)$ measurable, the set $C_n = \{(2^n Y) \bmod 2 = 1\}$ belongs to $\sigma(Z)$ for all $n \in \mathbb{Z}$. There exists thus a Borel subset B_n of $\mathbb{R}$ such that $C_n = Z^{-1}(B_n)$. Define now a function h on $\mathbb{R}$ by setting

$$h = \sum_{n \in \mathbb{Z}} 2^{-n} \mathbb{1}_{B_n}.$$

Then for all $\omega \in \Omega$,

$$h(Z(\omega)) = \sum_{n \in \mathbb{Z}} 2^{-n} \mathbb{1}_{B_n}(Z(\omega)) = \sum_{n \in \mathbb{Z}} 2^{-n} \mathbb{1}_{C_n}(\omega) = \sum_{n \in \mathbb{Z}} 2^{-n}((2^n Y(\omega)) \bmod 2) = Y(\omega).$$

Observe that h may take the value $+\infty$, but on the set $\{h = +\infty\}$ we can give it the value 0 without changing the last computation, because Y takes real values.

Finally, we have expressed Y explicitely as a function of Z. The case where Y can take negative values is treated as usual by decomposing Y into its non-negative and non-positive parts. □

Exercise 2.6 *Is the σ-field $\mathcal{C}$ defined in Exercise 2.3 of the form $\sigma(Z)$ for some measurable function $Z : \mathbb{R} \to \mathbb{R}$?*

Let us come back to our original problem of computing $\mathbb{E}[X|Z]$. Thanks to the proposition which we have just proved, we know that we are looking for a random variable of the form $h(Z)$ for some function h. This function must satisfy the second defining property of conditional expectation for all $B \in \sigma(Z)$. By definition of $\sigma(Z)$, each $B \in \sigma(Z)$ is of the form $B = Z^{-1}(E)$ for some Borel subset E of $\mathbb{R}$. The condition to be satisfied is

$$\int_B h(Z)\, dP = \int_B X\, dP,$$

which, since $B = Z^{-1}(E)$, can be written as

$$\int_E h(v) f_Z(v)\, dv = \int_{\mathbb{R}\times E} u f_{(X,Z)}(u,v)\, dudv.$$

The last line was deduced from the previous one by a short computation that you might want to carefully check. Now the right-hand side can be written as

$$\int_E \left(\int_{\mathbb{R}} u f_{(X,Z)}(u,v)\, du \right)\, dv.$$

Provided the denominator is not zero, we see that the function h given by

$$h(v) = \frac{\int_{\mathbb{R}} u f_{(X,Z)}(u,v)\, du}{f_Z(v)} = \frac{\int_{\mathbb{R}} u f_{(X,Z)}(u,v)\, du}{\int_{\mathbb{R}} f_{(X,Z)}(u,v)\, du}$$

satisfies the desired relation. Hence, $h(Z)$ is a conditional expectation of X given $\sigma(Z)$.

To what extent is h unique ? The integral of hf_Z with respect to the Lebesgue measure over each Borel subset of $\mathbb{R}$ is prescribed. This means that the integral of h with respect to $\mathbb{P}_Z$ over each Borel subset of $\mathbb{R}$ is prescribed. If h' has the same integral over each Borel subset, then h and h' coincide $\mathbb{P}_Z$-almost surely. Let us state this as an exercise.

Exercise 2.7 *On a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*, let* X *and* X' *be two random variables such that for all* $A \in \mathcal{A}$ *one has* $\int_A X\, d\mathbb{P} = \int_A X'\, d\mathbb{P}$*. Then* $\mathbb{P}(X = X') = 1$*.*

2.2 Uniqueness and existence

The result of Exercise 2.7 is elementary and we may not pay great attention to it, but in the context of conditional expectation, it is very important. Let us see why. Random variables are equivalence classes of measurable functions and therefore cannot be evaluated at a point of Ω. Indeed, two measurable functions which are equal almost everywhere may differ at any point which is not an atom of the underlying probability measure. In contrast, they can be integrated on any event, and this result tells us that this is all there is: a random variable is not meant to be evaluated, but it is certainly made to be integrated over events, and it is characterised by the collection of its integrals over all events.

This is important precisely because the conditional expectation $\mathbb{E}[X|\mathcal{B}]$ is defined not by its values at the points of Ω, but by the values of its integrals over all events of $\mathcal{B}$. It is an instance where a random variable is truly seen as an equivalence class of measurable functions, with no preferred choice of a specific measurable function in this class. This is in a sense the most intrinsic possible definition of a random variable.

These remarks hopefully make the structure of the definition of the conditional expectation $\mathbb{E}[X|\mathcal{B}]$ more natural. The first condition says that it is a $\mathcal{B}$-measurable random variable, and the second condition specifies its integral over every single event of $\mathcal{B}$.

By now, it should have become almost obvious that the conditional expectation is unique.

Lemma 2.3 *Recall the notation of Definition 2.1. If Y and Y' are conditional expectations of X given $\mathcal{B}$, then $Y = Y'$ almost surely.*

Proof. Indeed, Y and Y' are both measurable with respect to $\mathcal{B}$ and they have the same integral over every event belonging to $\mathcal{B}$. □

We have settled the question of the uniqueness, but not yet that of the existence of the conditional expectation. In order to do this, we will introduce a new and quite different point of view on the conditional expectation. This point of view is usually explained in terms of "the best prediction about X which can be made using the information available in $\mathcal{B}$". This is certainly correct, but there is something about this phrasing that I have always found strange, and I will try to explain it in a convincing way.

Let us start by a comment on the usual expectation. Apart from its definition, we can understand the expectation of a random variable through the strong law of large numbers. This law tells us that the arithmetic mean of a sufficiently large sample of independent copies of our random variable will be arbitrarily close to its expectation. However, this can hardly be called a *prediction*, let alone a *best prediction*. In which sense does the expectation of a random variable constitute a prediction about this variable ? If we think about a dice that we roll, there is little hope that it will actually give the value 3.5. A similar comment could be made about a uniform random variable on the interval $[0, 1]$, and about many other random variables. The words *best prediction* should not be heard in the naive sense of *likeliest outcome*.

There is however a precise and simple sense in which the expectation of a random variable constitutes a best prediction of its outcome.

Exercise 2.8 *Let X be a square-integrable random variable. Prove that for all real number c, the following equality holds:*

$$\mathbb{E}[(X - c)^2] = \mathrm{Var}(X) + (c - \mathbb{E}[X])^2.$$

Deduce that there is, among all constant random variables, one which is closer than any other to X in the normed vector space $(L^2(\Omega, \mathcal{A}, \mathbb{P}), \|\cdot\|_2)$.

If we think of $\mathbb{E}[X]$ not as a number but as a constant random variable (and this is very much in the spirit of our study of $\mathbb{E}[X|\mathcal{B}]$ in the case where $\mathcal{B} = \{\varnothing, C, C^c, \Omega\}$), then it is, among all the constant random variables, the one which is closest to X in the L^2 norm. Recalling Exercise 2.1, and introducing $\mathcal{B}_0 = \{\varnothing, \Omega\}$, we can reformulate this as follows: $\mathbb{E}[X|\mathcal{B}_0]$ is, among all random variables which are measurable with respect to $\mathcal{B}_0$, the one which is closest to X in the L^2 distance.

The key to the existence of the conditional expectation is that this is true not just for the σ-field $\mathcal{B}_0$, but for any other sub-σ-field of $\mathcal{A}$.

Theorem 2.4 *Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ be a square-integrable random variable. Let $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$.*

There exists, among all square-integrable random variables which are measurable with respect to $\mathcal{B}$, one which is closer to X in the L^2 distance than any other. This random variable is moreover a conditional expectation of X given $\mathcal{B}$.

Proof. The space of square-integrable random variables which are measurable with respect to $\mathcal{B}$ is the subspace $L^2(\Omega, \mathcal{B}, \mathbb{P})$ of $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Since the L^2 distance on $L^2(\Omega, \mathcal{B}, \mathbb{P})$ is the distance induced by the ambient space $L^2(\Omega, \mathcal{A}, \mathbb{P})$ and since $L^2(\Omega, \mathcal{B}, \mathbb{P})$ endowed with this distance is complete, it is a closed subspace of $L^2(\Omega, \mathcal{A}, \mathbb{P})$.

Let $p : L^2(\Omega, \mathcal{A}, \mathbb{P}) \to L^2(\Omega, \mathcal{B}, \mathbb{P})$ be the orthogonal projection, which exists precisely because $L^2(\Omega, \mathcal{B}, \mathbb{P})$ is a closed linear subspace. Then from the general theory of the geometry of Hilbert spaces, we know that $p(X)$ is the element of $L^2(\Omega, \mathcal{B}, \mathbb{P})$ which is closest to X in the L^2 distance.

Let us now prove that $p(X)$ is a conditional expectation of X given $\mathcal{B}$. To start with, $p(X)$ is measurable with respect to $\mathcal{B}$. Then, let us consider an event $B \in \mathcal{B}$. We have

$$\int_B p(X)\, d\mathbb{P} - \int_B X\, d\mathbb{P} = \int_\Omega (p(X) - X)\mathbb{1}_B\, d\mathbb{P} = \langle p(X) - X, \mathbb{1}_B \rangle_{L^2(\Omega, \mathcal{A}, P)}.$$

But on one hand, $\mathbb{1}_B$ is an element of $L^2(\Omega, \mathcal{B}, \mathbb{P})$. On the other hand, since p is an orthogonal projection, $p(X) - X$ is orthogonal to $L^2(\Omega, \mathcal{B}, \mathbb{P})$. Hence, the scalar product of the right-hand side is equal to 0, so that

$$\int_B p(X)\, d\mathbb{P} = \int_B X\, d\mathbb{P},$$

proving that $p(X)$ is the conditional expectation of X given $\mathcal{B}$. □

We have proved the existence for all square-integrable random variables. It remains to go from square-integrable to integrable random variables (recall that $L^2(\Omega, \mathcal{A}, \mathbb{P}) \subset L^1(\Omega, \mathcal{A}, \mathbb{P})$).

Theorem 2.5 *Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ be an integrable random variable. Let $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$. There exists a conditional expectation of X given $\mathcal{B}$.*

We first need to prove a fundamental property of the conditional expectation, which is its positivity. It is more easily done after proving its linearity. We do both in the following lemma.

Lemma 2.6 *Let $X_1, X_2 : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ be two random variables. Let $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$. Assume that X_1 and X_2 admit conditional expectations with respect to $\mathcal{B}$.*

1. For all reals a, b, the random variable $aX_1 + bX_2$ admits a conditional expectation given $\mathcal{B}$ and $\mathbb{E}[aX_1 + bX_2|\mathcal{B}] = a\mathbb{E}[X_1|\mathcal{B}] + b\mathbb{E}[X_2|\mathcal{B}]$.

2. If $0 \leq X_1 \leq X_2$ almost surely, then $0 \leq \mathbb{E}[X_1|\mathcal{B}] \leq \mathbb{E}[X_2|\mathcal{B}]$ almost surely.

Proof. 1. The random variable $a\mathbb{E}[X_1|\mathcal{B}] + b\mathbb{E}[X_2|\mathcal{B}]$ is integrable and $\mathcal{B}$-measurable. The linearity of the integral implies immediately that it is a conditional expectation of $aX_1 + bX_2$ given $\mathcal{B}$.

2. Thanks to the linearity, it suffices to prove that $0 \leq X$ almost surely implies $0 \leq \mathbb{E}[X|\mathcal{B}]$ almost surely. Consider an integer $k \geq 1$ and the event $B_k = \left\{\mathbb{E}[X|\mathcal{B}] \leq -\frac{1}{k}\right\}$. It belongs to $\mathcal{B}$. Hence, we have the inequalities

$$0 \leq \int_{B_k} X \, d\mathbb{P} = \int_{B_k} \mathbb{E}[X|\mathcal{B}] \, d\mathbb{P} \leq -\frac{1}{k}\mathbb{P}(B_k).$$

This implies $\mathbb{P}(B_k) = 0$. Hence, the event $B = \{\mathbb{E}[X|\mathcal{B}] < 0\}$ satisfies

$$\mathbb{P}(B) = \mathbb{P}\left(\bigcup_{k\geq 1} B_k\right) \leq \sum_{k\geq 1} \mathbb{P}(B_k) = 0.$$

Finally, $\mathbb{E}[X|\mathcal{B}]$ is non-negative almost surely. □

Let us now turn to the proof of the theorem.

Proof. Let us write $X = X^+ - X^-$, where $X^+ = \max(X, 0)$ and $X^- = (-X)^+$. Let us treat the case of X^+ first.

For each $n \geq 1$, let us consider the random variable $\min(X^+, n)$. It is non-negative and bounded by n. It is in particular square-integrable, so that it admits a conditional expectation given $\mathcal{B}$.

For each $m \leq n$, we have $0 \leq \min(X^+, m) \leq \min(X^+, n)$, so that, by the lemma, $0 \leq \mathbb{E}[\min(X^+, m)|\mathcal{B}] \leq \mathbb{E}[\min(X^+, n)|\mathcal{B}]$. Hence, the sequence of random variables $(\mathbb{E}[\min(X^+, n)|\mathcal{B}]))_{n\geq 1}$ is non-decreasing. In particular, it has a limit Y_+ towards which it converges almost surely.

On the other hand, the sequence $(\min(X^+, n))_{n\geq 1}$ is also non-decreasing and converges to X. Hence, for all $B \in \mathcal{B}$, the monotone convergence theorem applied to the equality

$$\int_B \mathbb{E}[\min(X^+, n)|\mathcal{B}] \, d\mathbb{P} = \int_B \min(X^+, n) \, d\mathbb{P}$$

yields

$$\int_B Y_+ \, d\mathbb{P} = \int_B X^+ \, dP.$$

In particular, since Y_+ is non-negative, $\int_\Omega Y_+ \, d\mathbb{P} = \int_\Omega X^+ \, d\mathbb{P} < +\infty$ and Y_+ is integrable. Since Y_+ is the almost sure limit of a sequence of $\mathcal{B}$-measurable random variables, it is also

$\mathcal{B}$-measurable. Finally, the equality above proves that it is the conditional expectation of X^+ given $\mathcal{B}$.

The same argument proves that X^- admits a conditional expectation given $\mathcal{B}$, which we denote by Y_-.

Finally, we claim that $Y = Y_+ - Y_-$ is the conditional expectation of X given $\mathcal{B}$. Indeed, Y is integrable, $\mathcal{B}$-measurable, and for all $B \in \mathcal{B}$, we have

$$\int_B Y \, d\mathbb{P} = \int_B Y_+ - Y_- \, d\mathbb{P} = \int_B Y_+ \, d\mathbb{P} - \int_B Y_- \, d\mathbb{P} = \int_B X^+ \, d\mathbb{P} - \int_B X^- \, d\mathbb{P} = \int_B X \, d\mathbb{P}.$$

This concludes the proof. □

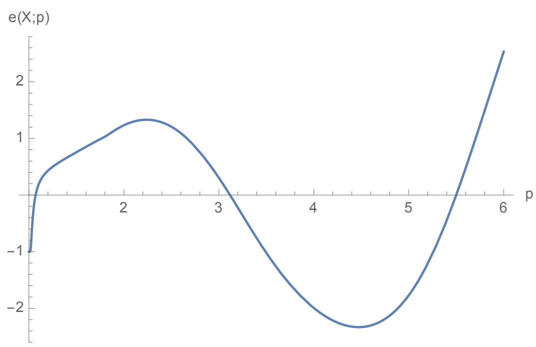
Let us summarise our approach. We analysed the definition of the conditional expectation in a few simple cases and this led us to an understanding of the uniqueness of the conditional expectation. We then proved its existence by using a strategy of proof which is not uncommon when one is dealing with L^1 spaces: we first proved the existence for L^2 random variables, using the rich geometry of Hilbert spaces, and then extended our result to L^1 by an argument of approximation.

Exercise 2.9 *Is it true that among all constant random variables, $\mathbb{E}[X]$ is closer to X than any other in the L^1 distance? Could we have characterised $\mathbb{E}[X|\mathcal{B}]$ for an integrable random variable by a property similar to the one we used for square-integrable random variables?*

Exercise 2.10 *Let X be a random variable that is uniformly distributed on the finite set $\{-2, -1, 1, 10\}$. Compute the expectation of X and draw the graph of the function $c \mapsto \mathbb{E}[|X - c|]$.*

For a general integrable random variable, describe the set of reals c that minimise $\mathbb{E}[|X - c|]$.

Exercise 2.11 *Let X be a bounded random variable. Prove that for all $p \in (1, \infty)$, the function $c \mapsto \|X - c\|_{L^p}$ attains its minimum for a unique value of c, which we denote by $e(X; p)$. What can you say about the function $p \mapsto e(X; p)$? Is it continuous? Does it have a limit as p tends to infinity? As p tends to 1? Here is a part of the graph of this function for a random variable X whose distribution is a slightly perturbed version of the distribution of the previous exercise:*



2.3 Main properties

We are now ready for the third part of this section, where we gather the useful properties of the conditional expectation.

Theorem 2.7 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $X, Y, X_1, X_2, \dots$ be integrable random variables on this space. Let a, b be real numbers. Let $\mathcal{B}, \mathcal{C}$ be sub-σ-fields of $\mathcal{A}$.*

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

2. If X is $\mathcal{B}$-measurable and XY is integrable, then $\mathbb{E}[XY|\mathcal{B}] = X\mathbb{E}[Y|\mathcal{B}]$.

3. If X is $\mathcal{B}$-measurable, then $\mathbb{E}[X|\mathcal{B}] = X$.

4. If X is independent of $\mathcal{B}$, then $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$ almost surely.

5. If $\mathcal{C} \subset \mathcal{B}$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{B}]|\mathcal{C}] = \mathbb{E}[X|\mathcal{C}]$.

6. $\mathbb{E}[\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[X]$.

7. If $X \geq 0$ almost surely, then $\mathbb{E}[X|\mathcal{B}] \geq 0$ almost surely.

8. If $(X_n)_{n\geq 1}$ is a non-decreasing sequence of non-negative random variables, converging to an integrable random variable X, then $\mathbb{E}[X_n|\mathcal{B}] \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}[X|\mathcal{B}]$.

9. If $(X_n)_{n\geq 1}$ is a sequence of non-negative random variables, then $\mathbb{E}[\liminf X_n|\mathcal{B}] \leq \liminf \mathbb{E}[X_n|\mathcal{B}]$.

10. If $(X_n)_{n\geq 1}$ is a sequence of non-negative random variables converging almost surely to a random variable X, and if there exists an integrable random variable Y such that for all $n \geq 1$ one has $|X_n| \leq Y$, then $\mathbb{E}[X_n|\mathcal{B}] \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}[X|\mathcal{B}]$.

11. If $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function and $\phi(X)$ is integrable, then $\phi(\mathbb{E}[X|\mathcal{B}]) \leq \mathbb{E}[\phi(X)|\mathcal{B}]$.

To prove this theorem yourself is one of the best exercises that you can do to at this stage. None of the ten first properties are difficult to prove. Among these, the least simple is the second, in that it is the least directly deduced from the definition. You should prove it after you have proved the eighth point, starting by the case where Y is an indicator function, then a simple function, and finally using an approximation argument.

For the eleventh property, use the fact that a convex function is the supremum of all affine functions which are inferior to it:

$$\phi(x) = \sup_{\substack{a,b\in\mathbb{R} \\ \forall y\in\mathbb{R}, ay+b\leq\phi(y)}} ax + b.$$

Here are some more exercises about conditional expectation. When nothing is specified, $X, Y, \dots$ are integrable random variables on a probability space.

Exercise 2.12 *Let X be a standard Gaussian random variable. Compute $\mathbb{E}[X|X^2]$.*

Exercise 2.13 *Assume that $\mathbb{E}[X|\mathcal{B}]$ is a constant random variable. Prove that this constant is $\mathbb{E}[X]$. Is X necessarily independent of $\mathcal{B}$?*

Exercise 2.14 *Let a be a real number. Let X be a random variable whose distribution admits the density*

$$f_X(x) = \frac{(x+a)^2}{1+a^2} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

with respect to the Lebesgue measure (check that f_X is indeed a density function). Try to understand before computing it that the sign of $\mathbb{E}[X|X^2]$ is almost surely constant, and to guess how it depends on a. Then compute $\mathbb{E}[X|X^2]$.

For all real number t, compute

$$\frac{tf_X(t) - tf_X(-t)}{f_X(t) + f_X(-t)}.$$

Do you see a relation between the two computations ?

Exercise 2.15 *Which relations can you find between $\mathbb{E}[X|Y^2]$ and $\mathbb{E}\left[X\big||Y|\right]$?*

Exercise 2.16 *Last year a student told me that $\mathbb{E}[X|X^2] = 0$ if and only if X and $-X$ have the same distribution. Do you think he was right ?*

Exercise 2.17 *Choose $p \in (0,1)$. Let X and Y be integer-valued random variables such that for all $k, l \in \mathbb{N}$, one has*

$$\mathbb{P}(X = k, Y = l) = (1-p)\left(\frac{p}{e}\right)^k \frac{k^l}{l!}.$$

Compute $\mathbb{E}[Y|X]$.

Exercise 2.18 *Let (X,Y) be a two-dimensional random vector whose distribution admits the density*

$$f_{(X,Y)}(s,t) = e^{-s}\mathbb{1}_{[0,s]}(t)$$

with respect to the Lebesgue measure on $\mathbb{R}^2$. Compute $\mathbb{E}[X|Y]$ and $\mathbb{E}[Y|X]$.

Exercise 2.19 *(Do this exercise only if you know the definition and main properties of Gaussian random vectors.) Let (X,Y) be a two-dimensional Gaussian random vector. Prove that there exists a real number a such that $X - aY$ is independent of Y. Prove that there exists a real number b such that $\mathbb{E}[X|Y] = aY + b$.*

Exercise 2.20 *Let N be a bounded integer-valued random variable. Let $(X_n)_{n\geq 1}$ be a sequence of identically distributed integrable random variables. Assume that $N, X_1, X_2, \ldots$ are independent. Set*

$$S = \sum_{n=1}^{N} X_n.$$

Prove that S is integrable and compute its expectation.

Exercise 2.21 *Consider the probability space* $([0,1), \mathscr{B}_{[0,1)}, \lambda)$*, where* λ *is the Lebesgue measure. Choose* $n \geq 0$ *and consider the* σ*-field*

$$\mathcal{F}_n = \sigma\left(\left\{\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) : k \in \{0 \dots 2^n - 1\}\right\}\right).$$

Define the random variable $X : [0,1) \to \mathbb{R}$ *by setting* $X(t) = t$ *for all* $t \in [0,1)$*. Compute* $\mathbb{E}[X|\mathcal{F}_n]$*.*

Exercise 2.22 *Consider a random vector* (X, Y, Z) *whose distribution admits the density*

$$f_{(X,Y,Z)}(x, y, z) = ce^{-z-2x}\mathbb{1}_{0 \leq x \leq y \leq z}$$

with respect to the Lebesgue measure on $\mathbb{R}^3$*, where* c *is a real constant. Determine the value of* c*, then compute* $\mathbb{E}[X|Y,Z]$ *and* $\mathbb{E}[X|Y]$*.*

Exercise 2.23 *Let* X *and* Y *be two square-integrable random variables such that* $\mathbb{E}[X|Y] = Y$ *and* $\mathbb{E}[Y|X] = X$*. Prove that* $X = Y$*.*

Exercise 2.24 *Consider a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*. Let* $\mathcal{B}$ *be a sub-*σ*-field of* $\mathcal{A}$*. Propose a definition of the conditional expectation* $\mathbb{E}[X|\mathcal{B}]$ *for every random variable* $X : (\Omega, \mathcal{A}) \to [0, +\infty]$*, without any assumption of integrability. Investigate the existence and uniqueness of this conditional expectation, as well as its properties, in the spirit of Theorem 2.7.*

Exercise 2.25 *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space. Let* $\mathcal{B}$ *be a sub-*σ*-field of* $\mathcal{A}$*. Check that* $L^\infty(\Omega, \mathcal{B}, \mathbb{P})$ *acts by multiplication on* $L^1(\Omega, \mathcal{A}, \mathbb{P})$*, in the sense that for all* $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ *and all* $Z \in L^\infty(\Omega, \mathcal{B}, \mathbb{P})$*, the product* ZX *belongs to* $L^1(\Omega, \mathcal{A}, \mathbb{P})$*. Check that* $L^\infty(\Omega, \mathcal{B}, \mathbb{P})$ *acts also by multiplication on* $L^1(\Omega, \mathcal{B}, \mathbb{P})$*.*

Find all linear maps

$$\mathcal{E} : L^1(\Omega, \mathcal{A}, \mathbb{P}) \to L^1(\Omega, \mathcal{B}, \mathbb{P})$$

which preserve the expectation and respect the action by multiplication of $L^\infty(\Omega, \mathcal{B}, \mathbb{P})$ *in the sense that for all* $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ *and all* $Z \in L^\infty(\Omega, \mathcal{B}, \mathbb{P})$*,*

$$\mathcal{E}(ZX) = Z\mathcal{E}(X).$$

In particular, check that $\mathcal{E}$ *is automatically continuous.*

Find all continuous linear maps $\tilde{\mathcal{E}} : L^1(\Omega, \mathcal{A}, \mathbb{P}) \to L^1(\Omega, \mathcal{B}, \mathbb{P})$ *which respect the action by multiplication of* $L^\infty(\Omega, \mathcal{B}, \mathbb{P})$*.*

3 Martingales

3.1 Martingales and conserved quantities

In the study of a deterministic dynamical system, a mechanical system for example, it is very important to identify quantities which stay constant through the evolution of the system. For example, let us think of the motion of a pendulum.

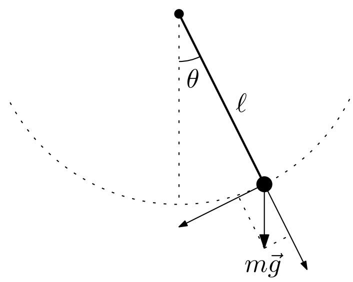


Figure 2: Newton's law writes $m\ell\ddot{\theta} = mg\sin\theta$, that is, $\ddot{\theta} = \frac{g}{\ell}\sin\theta$, where $g = 9.81\ ms^{-2}$ is the gravitation field on the surface of the Earth.

The differential equation $\ddot{\theta} = \frac{g}{\ell}\sin\theta$ is not easy to solve explicitly. Let us introduce the function $E = \frac{1}{2}m\ell\dot{\theta}^2 - mg\cos\theta$. It is immediately checked that $\dot{E} = 0$, and a very detailed study of the motion of the pendulum can be done from this single observation.

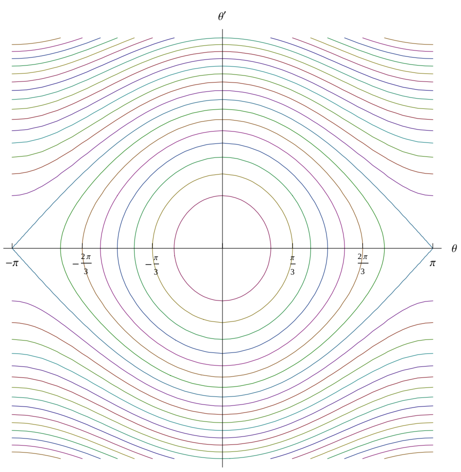


Figure 3: The curve $(\theta, \dot{\theta})$ corresponding to any possible motion of the pendulum is one of these curves (the right and left vertical sides of the picture, corresponding to $\theta = \pi$ and $\theta = -\pi$, should be identified). Which are the four radically different possible kinds of motion ?

In the evolution of a random dynamical system, it is unlikely that any quantity remains constant. However, certain quantities remain constant on average. For example, consider the simple random walk on $\mathbb{Z}$. That is, let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables such that $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$. Set $S_0 = 0$ and, for all $n \geq 1$, $S_n = X_1 + \ldots + X_n$.

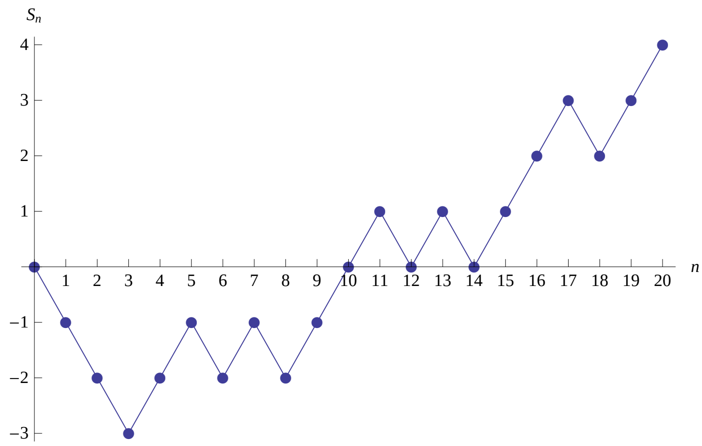


Figure 4: A sample path of the random dynamical system $(S_n)_{n\geq 0}$, also known as the simple random walk.

It is intuitively clear, and not difficult to prove, that the only functions $f : \mathbb{Z} \to \mathbb{R}$ such that $f(S_n)$ does not depend on n are the constant functions. There is nothing which, in this situation, plays exactly the same role as the mechanical energy in the case of the pendulum.

Exercise 3.1 *Find a non-constant function $f : \mathbb{N}\times\mathbb{Z} \to \mathbb{R}$ such that $f(n, S_n)$ is constant.*

However, you know that S_n itself is constant *on average*. More precisely, $\mathbb{E}[S_n] = 0$ does not depend on n. This observation is interesting but it is quite crude and, with some thought, it can be turned into a subtle and very fruitful one.

Indeed, look at the picture above. We know the value of S_n for $n \in \{0, \ldots, 20\}$. What can we say about S_{21} ? Since $S_{20} = 4$, it has probability $\frac{1}{2}$ to be equal to 3, or to 5. In particular, its expected value is $\frac{1}{2}(3 + 5) = 4$. Here, the words "expected value" do not of course refer to the expectation, for we know that $\mathbb{E}[S_{21}] = 0$. They refer to the *conditional expectation* of S_{21} given the information to which we have access. This information consists in the values of $S_0(\omega), \ldots, S_{20}(\omega)$ for the particular ω which corresponds to the particular realisation of the infinite experiment of which we see the first twenty steps.

Hence, the sentence "the expected value of S_{21} is 4" refers to the value at ω of the conditional expectation of S_{21} given the σ-field generated by $S_0, \ldots, S_{20}$. It means that

$$\mathbb{E}[S_{21}|\sigma(S_0, \ldots, S_{20})](\omega) = 4 = S_{20}(\omega).$$

How do we evaluate a random variable, which is only defined almost everywhere, at the point ω? Well, the information to which we have access, contained in the picture above, describes exactly one event of the σ-field $\sigma(S_0, \ldots, S_{20})$, namely the event

$$\{S_0 = 0, S_1 = -1, S_2 = -2, S_3 = -3, S_4 = -2, S_5 = -1, \ldots, S_{19} = 3, S_{20} = 4\}.$$

We know that ω belongs to this event. We also know that $\mathbb{E}[S_{21}|\sigma(S_0, \ldots, S_{20})]$ is constant on this event, equal to 4, which happens to be also the value of S_{20}.

Finally, the random dynamical system $(S_n)_{n\geq 0}$ has the following property: if for some $n \geq 0$ you observe $S_0, \ldots, S_n$, then the values which you observe define an event, on which the (conditional) average of S_{n+1} equals the value of S_n. This is beautifully expressed by the formula

$$\forall n \geq 0,\ \mathbb{E}[S_{n+1}|S_0, \ldots, S_n] = S_n.$$

This is the main defining property of a *martingale*.

Martingales play for random dynamical systems the role played by conserved quantities for deterministic dynamical systems. For example, knowing that $(S_n)_{n\geq 0}$ is a martingale will allow us to answer the question : what is the probability that the first time at which S_n hits the value a is smaller than the first time at which S_n hits the value b ? You can try to find the answer for $a = 2$ and $b = -1$ and see that it is in general not easy to determine this probability.

Exercise 3.2 *Check that $\mathbb{E}[S_n^2] = n$. What do you think about the process $(S_n^2 - n)_{n\geq 0}$?*

In the classical case, it is sometimes not possible, or difficult, to identify a conserved quantity. This is for example the case for the pendulum if we take friction into account. But then, one can prove that $\dot{E} \leq 0$ and this is already a very interesting information. Similarly, martingales have variants called sub- and super-martingales, which correspond to Lyapunov functions of classical dynamical systems. Let us now turn to the systematic study of this class of processes.

3.2 Definition

Definition 3.1 *Let $(\Omega, \mathcal{A})$ be a measurable space. A* filtration *on $(\Omega, \mathcal{A})$ is a non-decreasing sequence of sub-σ-fields of $\mathcal{A}$, that is, a sequence of sub-σ-fields of $\mathcal{A}$ such that the inclusions*

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \ldots \subset \mathcal{A}$$

hold. A probability space $(\Omega, \mathcal{A}, \mathbb{P})$ endowed with a filtration $(\mathcal{F}_n)_{n\geq 0}$ is called a filtered probability space.

Exercise 3.3 *Recall the notation of Exercise 2.21. Prove that $(\mathcal{F}_n)_{n\geq 0}$ is a filtration. Is $\bigcup_{n\geq 0} \mathcal{F}_n$ a σ-field on $[0, 1)$?*

The ambient σ-field $\mathcal{A}$ can always be replaced by $\sigma\left(\bigcup_{n\geq 0}\mathcal{F}_n\right)$, and this is why we will sometimes not mention it in our definitions.

An important example of filtration is that generated by a sequence of random variables. Given a sequence $X = (X_n)_{n\geq 0}$ of random variables (and we will also call such a sequence a stochastic process), we can form a filtration $(\mathcal{F}_n^X)_{n\geq 0}$ by setting, for all $n \geq 0$,

$$\mathcal{F}_n^X = \sigma(X_0, \dots, X_n).$$

Definition 3.2 *A stochastic process $X = (X_n)_{n\geq 0}$ defined on a filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ is said to be* adapted *if for all $n \geq 0$, the random variable X_n is measurable with respect to $\mathcal{F}_n$.*

With the current notation, you can check that X is adapted if and only if $\mathcal{F}_n^X \subset \mathcal{F}_n$ for all $n \geq 0$.

Definition 3.3 (Martingales) *Let $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. Let $X = (X_n)_{n\geq 0}$ be a stochastic process defined on this probability space. One calls X a* martingale *if the following conditions are satisfied.*

1. *X is adapted to $(\mathcal{F}_n)_{n\geq 0}$.*
2. *For all $n \geq 0$, X_n is integrable.*
3. *For all $n \geq 0$,*

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n. \tag{MG}$$

One calls X a supermartingale *if it satisfies the same conditions, but with the equality* (MG) *replaced by the inequality*

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n. \tag{$\overline{\text{MG}}$}$$

One calls X a submartingale *if it satisfies the same conditions with* (MG) *replaced by the other inequality*

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n. \tag{$\underline{\text{MG}}$}$$

The simple random walk is a fundamental example of martingale.

Exercise 3.4 *Consider the simple random walk $S = (S_n)_{n\geq 0}$ as defined in the previous section. Recall in particular the i.i.d. sequence $(X_n)_{n\geq 1}$. Prove that the filtrations $(\mathcal{F}_n^X)_{n\geq 0}$ and $(\mathcal{F}_n^S)_{n\geq 0}$ are equal. Prove that S is a martingale with respect to $(\mathcal{F}_n^S)_{n\geq 0}$. What do you think about $(S_n^2)_{n\geq 1}$?*

Exercise 3.5 *Let $X = (X_n)_{n\geq 0}$ be a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$. Compute, for all $n, m \geq 0$, the conditional expectation $\mathbb{E}[X_n|\mathcal{F}_m]$. Prove that $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ for all $n \geq 0$.*

Exercise 3.6 *Prove that if X is a submartingale (resp. a supermartingale), then the sequence $(\mathbb{E}[X_n])_{n\geq 0}$ is non-decreasing (resp. non-increasing). Mind the misleading vocabulary !*

Exercise 3.7 *Let $X = (X_n)_{n \geq 0}$ be a stochastic process. Prove that if X is a martingale with respect to a filtration $(\mathcal{F}_n)_{n \geq 0}$, then X is a martingale with respect to $(\mathcal{F}_n^X)_{n \geq 0}$. Is the converse true ? What about supermartingales and submartingales ?*

Another important example is the following. Consider an integrable random variable Z and a filtration $(\mathcal{F}_n)_{n \geq 0}$ and define, for all $n \geq 0$, $X_n = \mathbb{E}[Z|\mathcal{F}_n]$. Then, for all $n \geq 0$, X_n is $\mathcal{F}_n$-measurable by definition of the conditional expectation, it is also integrable by definition of the conditional expectation. Thanks to the property labelled 5 in Theorem 2.7, we have

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[Z|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[Z|\mathcal{F}_n] = X_n.$$

Hence, $(X_n)_{n \geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_n)_{n \geq 0}$.

A martingale of the form $(\mathbb{E}[Z|\mathcal{F}_n])_{n \geq 0}$ for some integrable random variable Z is sometimes called a closed martingale.

Exercise 3.8 *Is the simple random walk S a closed martingale ?*

Exercise 3.9 *Let again S be the simple random walk. Define $T = (T_n)_{n \geq 0}$ by setting $T_n = S_n$ if $n \leq 8888$ and $T_n = S_{8888}$ if $n > 8888$. Is T a martingale with respect to the filtration $(\mathcal{F}_n^S)_{n \geq 0}$? Is it a closed martingale ?*

Proposition 3.4 *Let $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ be supermartingales. Let a be a real number.*

1. *If a is positive, then $(aX_n)_{n \geq 0}$ is a supermartingale.*
2. *If a is negative, then $(aX_n)_{n \geq 0}$ is a submartingale.*
3. *$(X_n + Y_n)_{n \geq 0}$ is a supermartingale.*

The proposition obtained by interchanging the words supermartingale and submartingale is also true. In particular, any linear combination of martingales is a martingale.

The proof of this proposition is a verification. For the last sentence, observe that a stochastic process is a martingale if and only if it is both a supermartingale and a submartingale.

As a consequence of the conditional Jensen inequality (property 11 in Proposition 2.7), we have the following result.

Proposition 3.5 *Let $(X_n)_{n \geq 0}$ be a martingale. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function such that for all $n \geq 0$, $\phi(X_n)$ is integrable. Then $(\phi(X_n))_{n \geq 0}$ is a submartingale.*

The proposition also holds if X is a submartingale and ϕ is non-decreasing.

Proof. Let us treat the case where X is a submartingale. The process $(\phi(X_n))_{n \geq 0}$ is adapted and integrable. For all $n \geq 0$, we have

$$\mathbb{E}[\phi(X_{n+1})|\mathcal{F}_n] \geq \phi(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) \geq \phi(X_n),$$

where the first inequality is Jensen inequality and the second a consequence of the fact that ϕ is non-decreasing. This proves that $(\phi(X_n))_{n \geq 0}$ is a submartingale.

The case where X is a martingale is similar and easier. □

3.3 The case where the filtration is generated by partitions

Let us consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. On this probability space, let us consider a filtration $(\mathcal{F}_n)_{n\geq 0}$ and let us make the assumption that each σ-field $\mathcal{F}_n$ is generated by a finite partition of Ω. In this case, the partition which generates $\mathcal{F}_n$ is uniquely defined (as the set of elements of $\mathcal{F}_n \setminus \{\varnothing\}$ which are minimal for inclusion) and we will denote it by Π_n. The inclusion $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ is equivalent to the fact that each element of Π_n is the union of some elements of Π_{n+1} (one says that Π_{n+1} is a finer partition of Ω than Π_n). For the sake of simplicity, let us also assume that $\mathcal{F}_0 = \{\varnothing, \Omega\}$, that is, $\Pi_0 = \{\Omega\}$.

Exercise 3.10 *Check rigourously the two assertions made in the last paragraph.*

There is a natural genealogical structure on $\bigcup_{n\geq 0} \Pi_n$. Indeed, any set $B \in \Pi_{n+1}$ is included in exactly one set $A \in \Pi_n$, and we could say that A is the father of B. This relation of fatherhood gives rise to a genealogical tree whose nodes are the elements of $\bigcup_{n\geq 0} \Pi_n$. In order to follow more easily this discussion, it may be useful to have a look at the picture on the next page.

Let us introduce a notation for the elements of $\bigcup_{n\geq 0} \Pi_n$ which takes its genealogical structure into account. The idea is that each element of Π_n will be labelled by a word of integers of length n which contains all its genealogy.

To start with, $\mathcal{F}_0$ is generated by $\Pi_0 = \{\Omega\}$. There is a unique word of length 0, it is the empty word, and we will accordingly set $A_\varnothing = \Omega$.

Now, $\mathcal{F}_1$ is generated by a partition $\Pi_1 = \{A_1, \ldots, A_{n_\varnothing}\}$ of Ω. The number $n_\varnothing$ happens to be the total number of blocks of Π_1, but it is also more importantly the number of blocks of Π_1 *contained in* $A_\varnothing$. It is the number of children of $A_\varnothing$.

Then, the inductive rule is that the label of an element of Π_{n+1} is a word of $n+1$ integers whose n first letters are the label of the father of that element (the father belongs to Π_n) and whose last letter is the rank of this child in the progeny of the father. As you can guess and see on the picture, there is some freedom in the attribution of the last element, which amounts to a choice of a total order on the set of elements of Π_{n+1} which are contained in a particular element of Π_n.

Let us now consider a process $X = (X_n)_{n\geq 0}$ which is a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$. For each $n \geq 0$, X_n is measurable with respect to $\mathcal{F}_n$ and this is equivalent to the fact that X_n is constant on each block $A_{i_1\ldots i_n}$ of Π_n. Let us denote by $x_{i_1\ldots i_n}$ the value which X_n takes on $A_{i_1\ldots i_n}$.

The martingale property can be easily formulated in terms of the numbers $x_{i_1\ldots x_n}$. Indeed, for each $n \geq 0$, the conditional expectation $\mathbb{E}[X_{n+1}|\mathcal{F}_n]$ is constant on each block $A_{i_1\ldots i_n}$ of Π_n, taking the value $\sum_{j=1}^k x_{i_1\ldots i_n j} \frac{\mathbb{P}(A_{i_1\ldots i_n j})}{\mathbb{P}(A_{i_1\ldots i_n})}$, where k is the number of children of $A_{i_1\ldots i_n}$.

The martingale property is thus equivalent to the equality

$$x_{i_1\ldots i_n} = \sum_{j=1}^k x_{i_1\ldots i_n j} \frac{\mathbb{P}(A_{i_1\ldots i_n j})}{\mathbb{P}(A_{i_1\ldots i_n})} = \frac{\sum_{j=1}^k x_{i_1\ldots i_n j} \mathbb{P}(A_{i_1\ldots i_n j})}{\sum_{j=1}^k \mathbb{P}(A_{i_1\ldots i_n j})}.$$

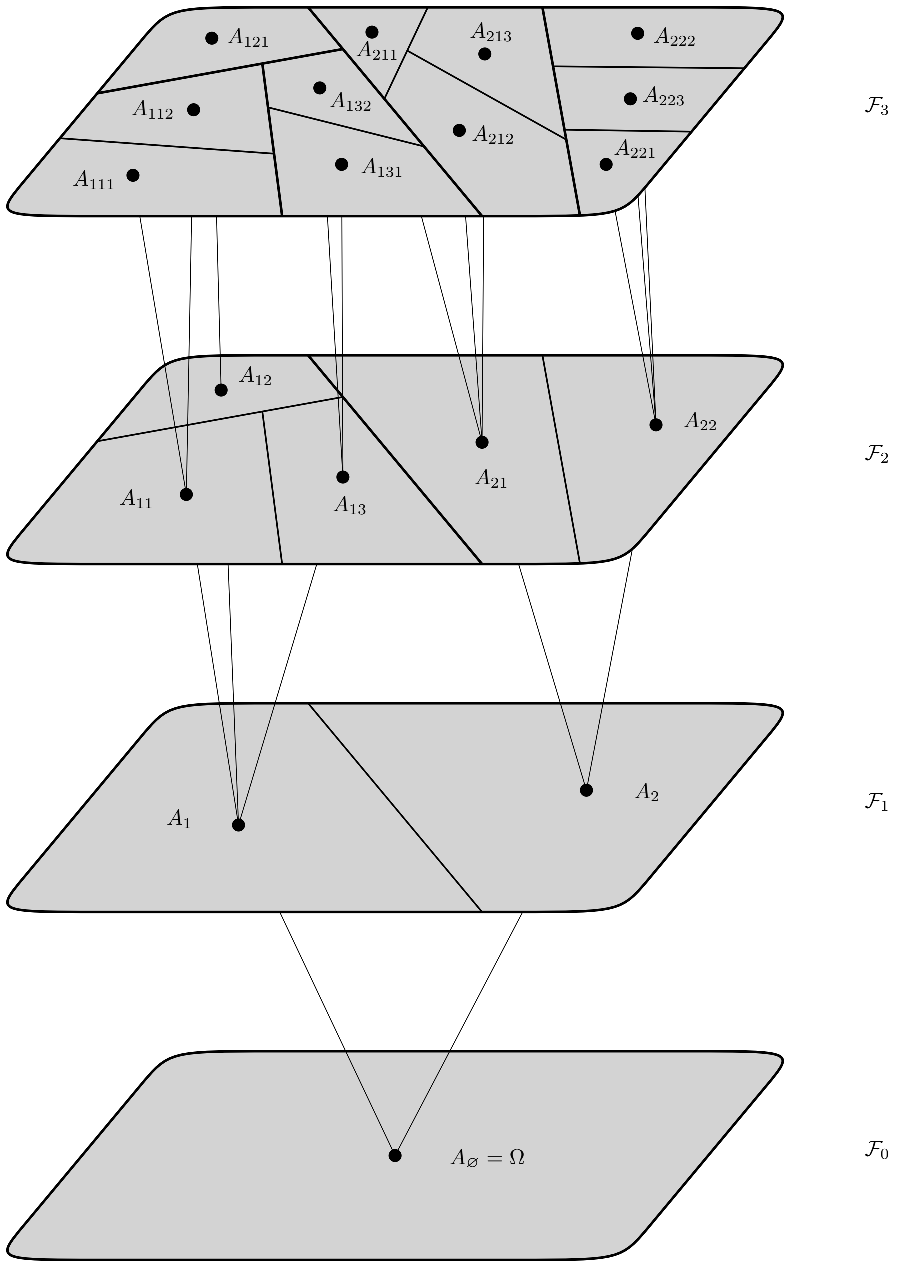


Figure 5: This illustrates the tree structure which underlies a filtration in which every σ-field is generated by a finite partition of Ω. A martingale is a process $X = (X_n)_{n \geq 0}$ such that on each block of the partition which generates $\mathcal{F}_n$, X_n is constant, and the average of X_{n+1} is equal to the value of X_n.

To summarise this discussion:

• we are considering the case of a filtration in which each σ-field is generated by a finite partition of Ω,

• there is a genealogical structure on the set of all blocks of the partitions which generate the σ-fields of our filtration,

• a stochastic process adapted to this filtration is specified by its value on each of these blocks, that is, by a real number attached to each node of the genealogical tree,

• each node of the tree has a natural weight, which is the proportion of the probability of its father that it represents,

• a stochastic process adapted to this filtration is a martingale if and only if the value attached to each node is equal to the weighted average of the values attached to each of its children.

Exercise 3.11 *Recall the notation of Exercise 2.21. Draw the first levels of the genealogical tree of the filtration* $(\mathcal{F}_n)_{n\geq 0}$. *Instead of labelling the sets by the integers* $1, 2, 3, \ldots$ *use the integers starting from* 0, *so that* $\Pi_1 = \{A_0, A_1, \ldots\}$, $\Pi_2 = \{A_{00}, A_{01}, \ldots, A_{10}, A_{11}, \ldots\}$. *Can you characterise the real numbers which belong to the set* $A_{0100110}$? *To the set* $A_{i_1 \ldots i_n}$?

Can you explain how martingales with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$ *can be put in one-to-one correspondence with certain ways of writing real numbers in the circles of the following picture? What about submartingales and supermartingales?*

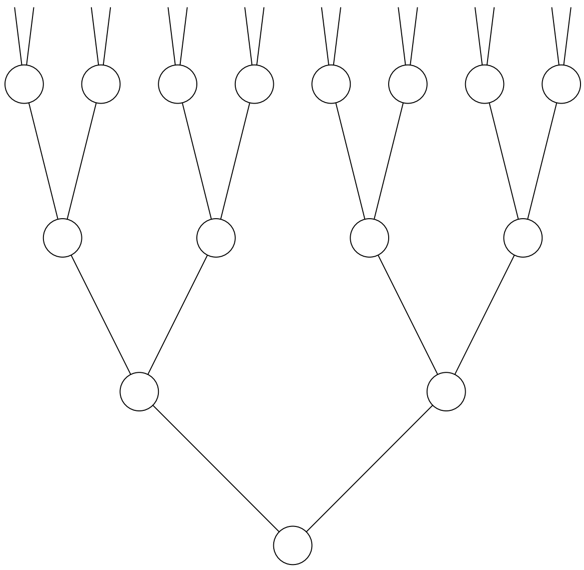


Figure 6: An infinite binary tree.

Exercise 3.12 *Let* X *be a non-negative martingale. By this we mean that for all* $n \geq 0$, $X_n \geq 0$ *almost surely. Prove that for all* n *and* m *such that* $0 \leq n \leq m$, *one has* $X_m = 0$ *on the event* $\{X_n = 0\}$, *that is,* $X_m \mathbb{1}_{\{X_n=0\}} = 0$.

One usually phrases this property by saying that when a non-negative martingale hits zero, it stays equal to zero forever. This formulation implicitly considers the index n *as a time variable, as is customary for stochastic processes.*

Is this property also true of submartingales? Of supermartingales?

3.4 Stochastic integration

Recall the definition of the simple random walk $(S_n)_{n\geq 0}$. You can think of it as representing the (random) unfolding of a game. At each step of the game, a coin is tossed, and according to the result, the player wins or loses 1. There is a variant of the game where the player is allowed to bet any real number x before the coin is tossed. Then his or her gain or loss will be x, depending on the coin. Of course, if the game is to be fair, the player is only allowed to bet *before* the coin is tossed. This means that the amount which he or she bets for the n-th run of the game can be decided only on the basis of the results of the $n-1$ first runs. This notion of fair betting is encoded by the following definition.

Definition 3.6 *Let $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. A stochastic process $H = (H_n)_{n\geq 1}$ is said to be* previsible *if for each $n \geq 1$, the random variable H_n is measurable with respect to $\mathcal{F}_{n-1}$.*

Exercise 3.13 *Check that a previsible process is adapted. What can you say about a previsible martingale ?*

Let $(X_n)_{n\geq 0}$ be a martingale, which we think as representing the successive states of fortune of a player which bets 1 at each turn of the game. Let us consider another player which bets in a fair way, according to the values of a previsible process $(H_n)_{n\geq 1}$. What is his fortune at time n?

Let us be more precise: X_0 is the initial fortune of the player and X_n his fortune after the n-th turn of the game. The gain of this first player during the n-th turn is thus $X_n - X_{n-1}$. The second player bets H_n just before this n-th turn, and gets $H_n(X_n - X_{n-1})$. Let us suppose that the second player starts with nothing. His fortune after the n-th turn is $H_1(X_1 - X_0) + \ldots + H_n(X_n - X_{n-1})$.

Definition 3.7 (Discrete stochastic integral) *Let $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. Let $X = (X_n)_{n\geq 0}$ be a martingale (or a sub- or super-martingale) and $H = (H_n)_{n\geq 1}$ a previsible process, both with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$. The* stochastic integral of H with respect to X *is the process $H \bullet X = ((H \bullet X)_n)_{n\geq 0}$ defined by $(H \bullet X)_0 = 0$ and, for all $n \geq 1$,*

$$(H \bullet X)_n = \sum_{k=1}^{n} H_k(X_k - X_{k-1}).$$

The process $H \bullet X$ is also sometimes denoted by $\int H \, dX$. The strength of this construction lies in the following result, which says that if the correct assumptions of integrability are satisfied, then $H \bullet X$ is still a martingale when X is a martingale.

Theorem 3.8 *Let $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. Let $X = (X_n)_{n\geq 0}$ be a martingale or a sub- or super-martingale and $H = (H_n)_{n\geq 1}$ be a previsible process.*

1. If X is a martingale and each random variable H_n is bounded, then $H \bullet X$ is a *martingale.*

2. If X is a supermartingale and each random variable H_n is bounded and non-negative, then $H \bullet X$ is a supermartingale.

3. In the two previous assertions, the assumption "each random variable H_n is bounded" can be replaced by "all random variables X_n and H_n are square-integrable".

Proof. For all $n \geq 1$, the random variable $(H \bullet X)_n$ is a function of the random variables $X_0, \ldots, X_n, H_1, \ldots, H_n$ which all are $\mathcal{F}_n$-measurable. Thus, $(H \bullet X)_n$ is $\mathcal{F}_n$-measurable and $H \bullet X$ is adapted.

The product of a bounded random variable with an integrable random variable is an integrable random variable. Also, by Hölder inequality, the product of two square-integrable random variables is an integrable random variable. Thus, in all cases considered in the statement, the random variable $(H \bullet X)_n$ is integrable for all $n \geq 0$.

Let us now check the main relation. Let us assume that X is a martingale. Then, for all $n \geq 0$, we must compute the conditional expectation

$$\begin{aligned}
\mathbb{E}[(H \bullet X)_{n+1} | \mathcal{F}_n] &= \mathbb{E}\big[\underbrace{H_1(X_1 - X_0) + \ldots + H_n(X_n - X_{n-1})}_{\mathcal{F}_n\text{-measurable}} + H_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n \big] \\
&= (H \bullet X)_n + \mathbb{E}\big[\underbrace{H_{n+1}}_{\mathcal{F}_n\text{-meas.}} (X_{n+1} - X_n) | \mathcal{F}_n \big] \\
&= (H \bullet X)_n + H_{n+1} \underbrace{\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n]}_{=0 \text{ because } X \text{ mart.}} \qquad \text{(Thm. 2.7, 2)} \\
&= (H \bullet X)_n.
\end{aligned}$$

If X is a supermartingale and H is non-negative, then only the last line changes. Indeed, in this case, $\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] \geq 0$, hence $H_{n+1}\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] \geq 0$ and finally $\mathbb{E}[(H \bullet X)_{n+1} | \mathcal{F}_n] \geq (H \bullet X)_n$. □

Exercise 3.14 *Choose an integer $N \geq 1$ and define H_n to be the constant random variable equal to 1 if $n \leq N$, and to 0 if $n > N$. Describe $H \bullet X$ in terms of X.*

Exercise 3.15 (The clever gambler) *A clever gambler plays the fair coin tossing game described at the beginning of this section. He thinks: "The coin is random, hence it doesn't like to repeat itself. Let me bet 1 on heads for the next turn whenever the coin gives tails, and conversely". Define rigourously the previsible process H which corresponds to his strategy (you will have to make a choice for H_1). Describe as precisely as you can the distribution of the stochastic process $H \bullet S$. Is our player quite as clever as he thinks?*

Exercise 3.16 *Recall once again the notation of Exercise 2.21. For all $n \geq 0$, write $X_n = \mathbb{E}[X | \mathcal{F}_n]$. Prove that for every martingale M with respect to the filtration $(\mathcal{F}_n)_{n \geq 0}$ such that $M_0 = 0$, there exists a previsible process H such that $M = H \bullet X$. It may be useful to have in mind the point of view of Exercise 3.11.*

3.5 Almost sure convergence

As a beautiful application of the construction which we made in the previous section, let us prove of one of the fundamental theorems on martingales. In the next statement, we use the notation $X_n^+ = \max(X_n, 0)$ for the positive part of a random variable X_n.

Theorem 3.9 (Almost sure convergence) *Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. Let $X = (X_n)_{n\geq 0}$ be a submartingale such that $\sup\{\mathbb{E}[X_n^+] : n \geq 0\} < \infty$. Then there exists an integrable random variable X_∞ such that the sequence $(X_n)_{n\geq 0}$ converges almost surely to X_∞.*

This theorem is stated for submartingales. There is also a version for supermartingales, which one reads by replacing X by $-X$: the assumption must be replaced by $\sup\{\mathbb{E}[X_n^-] : n \geq 0\} < \infty$. For martingales, which are both submartingales and supermartingales, any of the two assumptions implies the conclusion.

The following corollary may seem slightly less general than Theorem 3.9, but it is in fact equivalent to it, and perhaps more easily remembered.

Corollary 3.10 *A supermartingale (resp. submartingale, resp. martingale) which is bounded in L^1 converges almost surely.*

The following exercise explains why the assumptions of Theorem 3.9 and Corollary 3.10 are equivalent. For the sake of practising the definitions, it is formulated in terms of supermartingales.

Exercise 3.17 *Let X be a supermartingale. Prove that the sequence $(\mathbb{E}[X_n^-])_{n\geq 0}$ is non-decreasing. Prove that $\sup\{\mathbb{E}[X_n^-] : n \geq 0\}$ is finite if and only if X is bounded in L^1. Can we replace X_n^- by X_n^+ in this statement ?*

Exercise 3.18 *Let X be a (sub-, super-) martingale bounded in L^1. Prove that the almost sure limit of X is an integrable random variable.*

A useful special case of Theorem 3.9 is the following. It is easily remembered by analogy with the fact that a non-increasing sequence of non-negative real numbers is convergent.

Corollary 3.11 *A non-negative supermartingale converges almost surely towards a limit which is an integrable random variable.*

Proof. For all $n \geq 0$, we have

$$\mathbb{E}[|X_n|] = \mathbb{E}[X_n] \leq \mathbb{E}[X_0],$$

so that a non-negative supermartingale is bounded in L^1. The result now follows from Theorem 3.9. □

The main tool for proving the almost sure convergence is the notion of upcrossing of an interval. Let us first define it for a deterministic sequence. Let $x = (x_n)_{n\geq 0}$ be a sequence of real numbers. Let $a < b$ be two real numbers. We call upcrossing of $[a, b]$ by the sequence x any couple (s, t) of integers such that

$$s < t \text{ and } x_s < a < b < x_t.$$

We say that two upcrossings (s_1, t_1) and (s_2, t_2) occur successively if $t_1 < s_2$ or $t_2 < s_1$.

Exercise 3.19 *Prove that the sequence x does not converge to an element of $[-\infty, +\infty]$ if and only if there exists two rational numbers a and b such that $a < b$ and such that there are infinitely many successive upcrossings of $[a, b]$ by x.*

For all $N \geq 1$, let us denote by $u_N(x; a, b)$ the number of successive upcrossings of $[a, b]$ by x before time N. More rigourously, $u_N(x; a, b)$ is the largest integer k such that there exists $2k$ integers $s_1, t_1, \ldots, s_k, t_k$ such that

$$0 \leq s_1 < t_1 < \ldots < s_k < t_k \leq N \text{ and for all } i \in \{1, \ldots, k\},\ x_{s_i} < a < b < x_{t_i}.$$

The sequence $(u_N(x; a, b))_{N\geq 1}$ is non-decreasing and we call $u_\infty(x; a, b)$ its limit, which belongs to $\mathbb{N} \cup \{+\infty\}$.

The statement of the last exercise is that x converges to an element of $[-\infty, +\infty]$ if and only if $u_\infty(x; a, b)$ is finite for all rational a and b.

Let us now turn to the probabilistic case. Let X be a stochastic process. We define for all $N \geq 1$ and for all real numbers $a < b$ the number of upcrossings $U_N(X; a, b)$ of $[a, b]$ by X before time N, which is now an integer-valued random variable. We define $U_\infty(X; a, b)$ as the almost sure limit of the non-decreasing sequence of random variables $(U_N(X; a, b))_{N\geq 1}$.

In order to prove Theorem 3.9, we are going to prove that its assumptions imply that $U_\infty(X; a, b) < +\infty$ for all rational $a < b$ almost surely. One needs to pay attention to the order of the quantifiers "for all rational $a < b$" and "almost surely". Fortunately, there are countably many intervals with rational endpoints, and the two statements

$$\forall a < b \in \mathbb{Q},\ \mathbb{P}(U_\infty(X; a, b) < \infty) = 1 \text{ and } \mathbb{P}(\forall a < b \in \mathbb{Q}, U_\infty(X; a, b) < \infty) = 1$$

are equivalent.

Exercise 3.20 *Check this equivalence.*

The main ingredient of the proof is an estimation of this number of upcrossings, which we will deduce from Theorem 3.8 by an appropriate choice of the previsible process H.

Proposition 3.12 (Doob's upcrossing lemma) *Let X be a supermartingale. Let a, b be two real numbers such that $a < b$. For all $n \geq 1$, one has*

$$\mathbb{E}[U_n(X; a, b)] \leq \frac{1}{b - a}\mathbb{E}[(X_n - a)^-].$$

Proof. Let us think in the same terms as before Definition 3.7. Here is the line of reasoning of a very clever gambler. "I know that the game is fair, so that whenever X_n has reached a value which is too low, it will have to compensate and increase until it takes again higher values. Let me choose two levels a and b, which I call respectively the low level and the high level. I will wait until the first time X reaches a level below a. Then I will start betting 1 at each turn, and stop as soon as X reaches a level above b. And then I will repeat this strategy. I am quite confident that I will make a lot of money that way."

Well, we know by Theorem 3.8 that our gambler is, on average, not going to make money, but rather to lose money. However he is going to help us proving Doob's upcrossing lemma, which is certainly more important.

Let us define the previsible process H which corresponds to his strategy. We define it inductively. First set

$$H_1 = \mathbb{1}_{\{X_0 < a\}},$$

because we start betting at the first turn if and only if X_0 is below the low level a. Then suppose $H_1, \ldots, H_{n-1}$ have been defined, for some $n \geq 2$. If $H_{n-1} = 1$, then X has recently reached a level below a and we are currently betting until it exceeds b. Thus, $H_n = 1$ unless X has just reached such a level, that is, unless $X_{n-1} > b$. If on the contrary $H_{n-1} = 0$, then we are waiting until X passes below a. We thus have $H_n = 0$, unless $X_{n-1} < a$. Altogether, we set

$$H_n = \mathbb{1}_{\{H_{n-1}=1\}} \mathbb{1}_{\{X_{n-1} \leq b\}} + \mathbb{1}_{\{H_{n-1}=0\}} \mathbb{1}_{\{X_{n-1} < a\}}.$$

From this definition it follows that H_1 is $\sigma(X_0)$-measurable, hence $\mathcal{F}_0$-measurable, and for all $n \geq 2$, H_n is $\sigma(H_{n-1}, X_{n-1})$-measurable. By induction, it follows that H_n is $\mathcal{F}_{n-1}$-measurable.

Finally, $H = (H_n)_{n \geq 0}$ is previsible, and since it is bounded and non-negative by definition, the second assertion of Theorem 3.8 applies, allowing us to claim that $H \bullet X$ is a supermartingale.

Incidentally, this is why our very clever gambler is losing money on average : his winnings are given by $H \bullet X$, but $\mathbb{E}[(H \bullet X)_n]$ is a non-increasing sequence.

But the main point is the following inequality : for all $n \geq 1$, we have

$$(H \bullet X)_n \geq (b - a) U_n(X; a, b) - (X_n - a)^-. \tag{2}$$

The first term is what motivates our gambler: each upcrossing of $[a, b]$ by X returns him at least $b - a$. The second term is what he forgot to take into account, and what restores the equity. Sometimes, X will reach a level below a, and it will reach much lower values before it rises again to a level higher than b. In the mean time, our gambler will have reached the limit of his own solvability.

For those of you who want to read a more rigourous proof of (2), let us introduce the sequence of times at which the gambler changes his bet. More precisely, let us set

$$S_1 = \inf\{n \geq 0 : H_{n+1} = 1\} = \inf\{n \geq 0 : X_n < a\}.$$

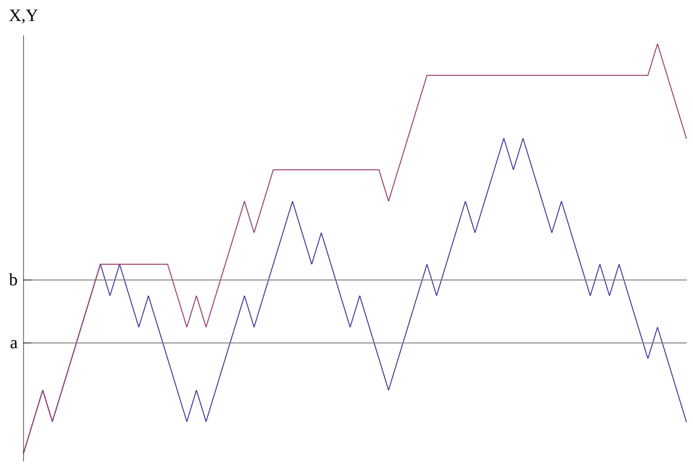


Figure 7: A sample path of X and the corresponding sample path of $H \bullet X$.

Then, set

$$T_1 = \inf\{n \geq S_1 : H_{n+1} = 0\} = \inf\{n \geq S_1 : X_n > b\}.$$

The first upcrossing of $[a, b]$ by X is thus the random interval $[S_1, T_1]$. During this interval, the gambler gains $X_{T_1} - X_{S_1}$. Then, suppose $S_1, T_1, \ldots, S_k, T_k$ defined. We set

$$\begin{cases} S_{k+1} = \inf\{n \geq T_k \quad : H_{n+1} = 1\} = \inf\{n \geq T_k \quad : X_n < a\}, \\ T_{k+1} = \inf\{n \geq S_{k+1} : H_{n+1} = 0\} = \inf\{n \geq S_{k+1} : X_n > b\}. \end{cases}$$

Any of these random times can be infinite. If this occurs, the next times are not defined.

Now let us consider an integer n. There is, among the random times which we have defined, one which is largest among those which are smaller than n. There are two cases, depending on whether this last time is an S or a T.

1. The largest time is S_{k+1} for some k. In this case, $U_n(X; a, b) = k$. Moreover,

$$(H \bullet X)_n = (X_{T_1} - X_{S_1}) + \ldots + (X_{T_k} - X_{S_k}) + (X_n - X_{S_{k+1}}).$$

The first k terms are larger than $b - a$ because $X_{T_l} > b$ and $X_{S_l} < a$ for all $l \in \{1, \ldots, k\}$. The last term is larger than $X_n - a$, hence than $-(X_n - a)^-$. Indeed, recall that for all real number x, we use the notation $x^- = \max(-x, 0)$, so that the inequality $x \geq -x^-$ holds.

Thus, we have

$$(H \bullet X)_n \geq k(b - a) - (X_n - a)^+ = (b - a)U_n(X; a, b) - (X_n - a)^+$$

and (2) is proved.

2. The largest time is T_k for some k. In this case,

$$(H \bullet X)_n = (X_{T_1} - X_{S_1}) + \ldots + (X_{T_k} - X_{S_k}) \geq (b - a)U_n(X; a, b)$$

and (2) is also proved.

3. There is actually a third case, where $S_1 > n$. In this case, both sides of (2) are equal to zero.

Now let us take the expectation on both sides of (2). We find

$$(b - a)\mathbb{E}[U_n(X; a, b)] - \mathbb{E}[(X_n - a)^-] \leq \mathbb{E}[(H \bullet X)_n] \leq \mathbb{E}[(H \bullet X)_0] = 0,$$

the last inequality being due to the fact that $H \bullet X$ is a supermartingale. This concludes the proof of the proposition. □

Proof. (Theorem 3.9) Let X be a supermartingale such that $\sup\{\mathbb{E}[X_n^-] : n \geq 0\} < \infty$. Let us prove that it converges almost surely.

Consider $a, b \in \mathbb{Q}$ such that $a < b$. For all $n \geq 1$, Doob's lemma reads

$$\mathbb{E}[U_n(X; a, b)] \leq \frac{1}{b-a}\mathbb{E}[(X_n - a)^-].$$

On one hand, for all real number x, we have $(x-a)^- \leq x^- + |a|$. On the other hand, the non-decreasing sequence $(U_n(X; a, b))_{n\geq 1}$ converges to $U_\infty(X; a, b)$, so that the monotone convergence theorem entails

$$\mathbb{E}[U_\infty(X; a, b)] \leq \frac{1}{b-a}\left(\sup\{\mathbb{E}[X_n^-] : n \geq 0\} + |a|\right) < \infty$$

Hence,

$$\mathbb{P}(U_\infty(X; a, b) = \infty) = 0.$$

We have proved that for all rationals $a < b$, there are almost surely only finitely many upcrossings of $[a, b]$ by X. By the equivalence checked in Exercise 3.20, it follows that

$$\mathbb{P}(\forall a < b \in \mathbb{Q}, U_\infty(X; a, b) < \infty) = 1.$$

This in turn, by Exercise 3.19, implies that $(X_n)_{n\geq 0}$ converges almost surely. Let us denote by X_∞ its limit. There remains to prove that X_∞ is integrable.

Observe that for all $n \geq 0$, we have $\mathbb{E}[X_0] \geq \mathbb{E}[X_n] = \mathbb{E}[X_n^+] - \mathbb{E}[X_n^-]$, so that

$$\mathbb{E}[X_n^+] \leq \mathbb{E}[X_0] + \mathbb{E}[X_n^-] \leq \mathbb{E}[X_0] + \sup\{\mathbb{E}[X_n^-] : n \geq 0\}.$$

Hence, $\mathbb{E}[X_n^+]$ is also bounded, and we deduce that $\mathbb{E}[|X_n|]$ is bounded, by $\mathbb{E}[X_0] + 2\sup\{\mathbb{E}[X_n^-] : n \geq 0\}$.

The sequence $(X_n)_{n\geq 0}$ is bounded in L^1 and converges almost surely to X_∞. Hence, by Fatou's lemma,

$$\mathbb{E}[|X_\infty|] = \mathbb{E}[\liminf |X_n|] \leq \liminf \mathbb{E}[|X_n|] \leq \sup\{\mathbb{E}[|X_n|] : n \geq 0\} < \infty.$$

In other words, X_∞ is integrable and the proof of the theorem is finished. $\square$

Let us summarise the strategy of the proof.

1. We expressed the convergence of a sequence of real numbers in terms of upcrossings of intervals.

2. By integrating a well-chosen previsible process against our supermartingale, we derived a bound on the number of its upcrossings of a fixed interval.

3. Thanks to the countability of $\mathbb{Q}$, we deduced that a supermartingale which is bounded in L^1 has almost surely only finitely many upcrossings of every interval with rational endpoints.

4. We concluded that a supermartingale bounded in L^1 converges almost surely, and Fatou's lemma allowed us to prove that the limit is integrable.

Exercise 3.21 *Let X be a supermartingale bounded in L^1. For all real numbers a, b such that $a < b$, define*

$$N_{a,b} = \{U_\infty(X; a, b) = \infty\}.$$

Compute $\mathbb{P}(\bigcup_{a<b} N_{a,b})$, where the union is taken over all real *numbers $a < b$.*

Exercise 3.22 *Consider a sequence $(Y_n)_{n\geq 1}$ of independent random variables such that for all $n \geq 1$, one has*

$$\mathbb{P}(Y_n = 0) = 1 - \frac{1}{n^2}, \ \mathbb{P}(Y_n = e^n) = \frac{1}{2n^2} \text{ and } \mathbb{P}(Y_n = -e^n) = \frac{1}{2n^2}.$$

Set $X_0 = 0$ and, for all $n \geq 1$, $X_n = Y_1 + \ldots + Y_n$.

Prove that $(X_n)_{n\geq 0}$ is a martingale in its natural filtration. Prove that $(X_n)_{n\geq 0}$ converges almost surely to a random variable X_∞. Prove that X_∞ is not integrable (Hint: consider the events $J_n = \{Y_1 = e, Y_n = e^n$ and $Y_k = 0$ for all $k \notin \{1, n\}\}$).

Exercise 3.23 *Let $S = (S_n)_{n\geq 0}$ be the simple random walk. Define $H = (H_n)_{n\geq 1}$ inductively by setting $H_1 = 1$ and for all $n \geq 2$,*

$$H_n = \mathbb{1}_{\{H_{n-1}=1\}} \mathbb{1}_{\{S_{n-1} \neq -1\}}.$$

Explain with words what the process $H \bullet S$ is. Prove that $H \bullet S$ is a martingale. Using the following fact (which you can admit if you don't know it):

$$\mathbb{P}(\inf\{k > 0 : S_k = -1\} < \infty) = 1,$$

prove that $H \bullet S$ converges almost surely to a random variable C. What is this random variable ? Is there convergence in L^1 of the sequence $(S_n)_{n\geq 1}$ towards C ?

3.6 Branching processes

Branching processes are a very important class of stochastic processeses, and a rich source of examples of martingales. They constitute a simple model for the evolution of the size of a population, according to the following simple scheme.

The population starts with a certain number $\ell \geq 1$ of indiviuals, who constitute the 0-th generation. Then, at each step of the process, each individual of the existing population gives birth to a certain number of children, and disappears. The number of children of each individual is random, it has the same distribution for each individual, and the numbers of children of the individuals living at a certain generation are independent. Let us give a more formal definition of the process $(X_n)_{n\geq 0}$, where X_n is the size of the population at time n.

Let $(Z_{n,k})_{n,k\geq 0}$ be a family of independent identically distributed integer-valued random variables. Let us exclude the cases where these random variables are 0 almost surely, and 1 almost surely. Let us define inductively $X_0 = \ell$ and, for each $n \geq 0$,

$$X_{n+1} = \sum_{k=1}^{X_n} Z_{n,k}.$$

Let us define a filtration by setting $\mathcal{F}_0 = \{\varnothing, \Omega\}$ and, for all $n \geq 1$,

$$\mathcal{F}_n = \sigma(Z_{m,k} : m < n, k \geq 1).$$

Let m denote the common expectation of the random variables $Z_{n,k}$, that is, the average number of children of an individual of our population. It is a posivite real number or $+\infty$, for the case $\mathbb{P}(Z_{n,k} = 0) = 1$ is excluded. Let us assume that $m < +\infty$.

The following result is extremely helpful in the study of branching processes, and it is an instance of the general idea that it is useful in the study of a random dynamical system to identify that certain quantities are martingales.

Proposition 3.13 *The process $(m^{-n} X_n)_{n\geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$.*

Proof. By definition, X_0 is $\mathcal{F}_0$-measurable. Then, for all $n \geq 1$, X_n is a function of X_{n-1} and $\{Z_{n-1,k} : k \geq 1\}$. By induction, and by definition of $\mathcal{F}_n$, it follows that X_n is $\mathcal{F}_n$-measurable. Thus, $(X_n)_{n\geq 0}$ is adapted, and so is $(m^{-n} X_n)_{n\geq 0}$.

Let us now prove by induction on n that X_n is integrable. For $n = 0$, this is true. Let us assume that X_n is integrable. Then, since X_{n+1} is non-negative, we may consider its expectation, and we must prove that it is not equal to ∞. We have

$$\mathbb{E}[X_{n+1}] = \sum_{x=0}^{\infty} \mathbb{E}\left[\sum_{k=1}^{x} Z_{n,k} \mathbb{1}_{X_n = x}\right].$$

Since X_n is $\mathcal{F}_n$-measurable and the random variables $\{Z_{n,k} : k \geq 1\}$ are independent of $\mathcal{F}_n$, we find

$$\begin{aligned}
\mathbb{E}[X_{n+1}] &= \sum_{x=0}^{\infty} \sum_{k=1}^{x} \mathbb{E}[Z_{n,k}] \mathbb{P}(X_n = x) \\
&= \sum_{x=0}^{\infty} x m \mathbb{P}(X_n = x) \\
&= m \mathbb{E}[X_n].
\end{aligned}$$

Thus, $\mathbb{E}[X_n] = m^n \ell$ is finite.

Let us finally compute $\mathbb{E}[X_{n+1} | \mathcal{F}_n]$. We have

$$\begin{aligned}
\mathbb{E}[X_{n+1} | \mathcal{F}_n] &= \mathbb{E}\left[\sum_{k=1}^{X_n} Z_{n,k} \middle| \mathcal{F}_n\right] \\
&= \sum_{k=1}^{X_n} \mathbb{E}[Z_{n,k}] \\
&= m X_n.
\end{aligned}$$

It follows from this result that $(m^{-n} X_n)_{n\geq 0}$ is a martingale. □

Exercise 3.24 *Check the passage from the first to the second line in the last computation of this proof.*

We may apply Corollary 3.11 to the non-negative martingale $(m^{-n}X_n)_{n\geq 0}$, which thus converges almost surely towards an integrable random variable Y. There are three fairly different cases, depending on the value of m.

- The case $m < 1$. In this case, the almost sure convergence of $(m^{-n}X_n)_{n\geq 0}$ to Y implies that almost surely, $X_n = m^n(m^{-n}X_n)$ converges to $0 \times Y = 0$. Since $(X_n)_{n\geq 0}$ is an integer-valued process, this forces it to be almost surely stationary with limiting value 0. Indeed, a sequence of integers is convergent if and only if it is stationary[1]. Let us write our conclusion in symbols:

  $$\mathbb{P}(\exists N \geq 1, \forall n \geq 1, X_n = 0) = 1.$$

  In words, our conclusion is that when $m < 1$, the population gets extinct with probability 1. This case is called the *subcritical* case.

- The case $m = 1$. In this case, $(X_n)_{n\geq 1}$ itself converges almost surely to Y. Because X is integer-valued, this convergence implies that it is almost surely stationary. In symbols:

  $$\mathbb{P}(\exists p \geq 0, \exists N \geq 1, \forall n \geq 1, X_n = p) = 1.$$

  For each integer $p \geq 0$, we can consider the event

  $$S_p = \{\exists N \geq 1, \forall n \geq 1, X_n = p\}$$

  on which X is stationary at p. We claim that for all $p \geq 1$, the event S_p has probability 0. Indeed, choose $p \geq 1$ and rewrite the event S_p as

  $$S_p = \{\exists N \geq 1, \forall n \geq N, Z_{n,1} + \ldots + Z_{n,p} = p\} = \bigcup_{N\geq 1} \bigcap_{n\geq N} \{Z_{n,1} + \ldots + Z_{n,p} = p\}.$$

  The events $(\{Z_{n,1} + \ldots + Z_{n,p} = p\})_{n\geq 1}$ are independent and they all have the same probability. Since the case $\mathbb{P}(Z_{n,k} = 1) = 1$ is excluded, and since $\mathbb{E}[Z_{n,k}] = 1$, we must have $\mathbb{P}(Z_{n,k} = 0) > 0$ (check this assertion !). So, in particular,

  $$\mathbb{P}(Z_{n,1} + \ldots + Z_{n,p} = p) \leq 1 - \mathbb{P}(Z_{n,1} + \ldots + Z_{n,p} = 0) = 1 - \mathbb{P}(Z_{n,k} = 0)^p < 1.$$

  The version of Borel-Cantelli's lemma for independent events now implies that $\mathbb{P}(S_p) = 0$.

  Since for all $p \geq 1$, the event on which X is stationary with limiting value p has probability 0, it must be that X converges to 0 almost surely. In this case too, the population gets extinct with probability 1. This case is called the *critical* case.

[1] Let me spell this out, since it is so important in the present problem. Let $a = (a_n)_{n\geq 0}$ be a sequence of elements of $\mathbb{Z}$. If a is stationary, it is of course convergent. Let us prove the converse. Assume that a is convergent. Then it is in particular Cauchy. Applying the definition of a Cauchy sequence with $\varepsilon = \frac{1}{2}$, we find that there exists an integer N such that for all $n \geq N$, we have $|a_n - a_N| < \frac{1}{2}$. Since a_n and a_N are both integers, they must be equal. Hence, a is stationary after rank N.

- The case $m > 1$. In this *supercritical* case, which we will not treat in detail, one can show that there is a positive probability that the population survives forever.

Exercise 3.25 *Prove that in the case where $m = 1$, the martingale $(X_n)_{n\geq 0}$ does not converge in L^1 to its almost sure limit. How do you intuitively understand this fact ?*

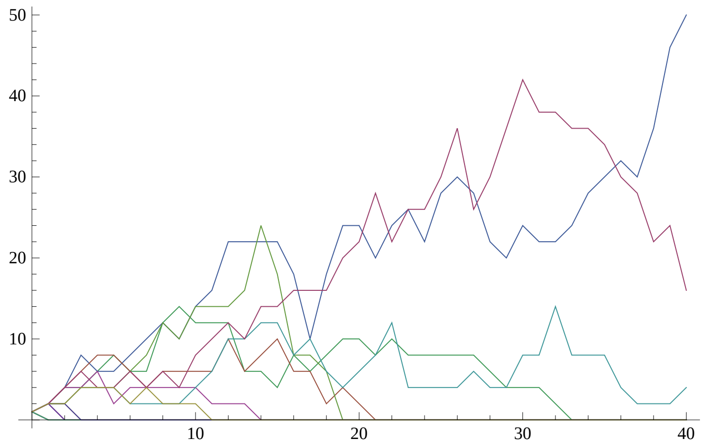


Figure 8: Twenty sample paths of X when $\ell = 1$ and $\mathbb{P}(Z_{n,k} = 0) = \mathbb{P}(Z_{n,k} = 2) = \frac{1}{2}$.

Exercise 3.26 *Prove that in a subcritical branching process, the total population $\sum_{n\geq 0} X_n$ is not only finite almost surely, but has a finite expectation. Compute this expectation. What about the critical case ?*

Exercise 3.27 *This exercise is a preparation for the next. Prove that every random variable Y with values in $\mathbb{N}$ satisfies the inequality*

$$\mathbb{P}(Y \geq 1) \geq \frac{\mathbb{E}[Y]^2}{E[Y^2]}.$$

Exercise 3.28 *Prove that for a subcritical branching process, there exists a positive real $\alpha > 0$ such that for all $n \geq 0$,*

$$\mathbb{P}(X_n \geq 1) \leq e^{-\alpha n}.$$

In words, the probability of survival of the population up to time n decays exponentially fast.

On the other hand, consider a critical branching process such that the random variables $Z_{n,k}$ are square-integrable. Set $\sigma^2 = \mathrm{Var}(Z_{n,k})$. Prove that

$$E[X_n^2] = n\ell\sigma^2 + \ell(\ell - \sigma^2)$$

and deduce that there exists a positive real $\beta > 0$ such that

$$\mathbb{P}(X_n \geq 1) \geq \frac{\beta}{n}.$$

3.7 Convergence in L^1

Critical branching processes are examples of non-negative martingales which converge almost surely (as they must according to Corollary 3.11), but not in L^1 (see Exercise 3.25). In this section, we clarify what it means for a martingale to be convergent in L^1.

Theorem 3.14 *Let $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. Let $X = (X_n)_{n\geq 0}$ be a martingale. The following two conditions are equivalent:*

1. The martingale X converges to a random variable X_∞ almost surely and in L^1.

2. There exists an integrable random variable Z such that $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ for all $n \geq 0$.

Moreover, if these conditions are satisfied, then one can take $Z = X_\infty$ in 2.

Proof. $1 \Rightarrow 2$. Choose $n \geq 0$. For all $m \geq n$, we have $X_n = \mathbb{E}[X_m|\mathcal{F}_n]$. We have

$$\begin{aligned}\mathbb{E}\left[|\mathbb{E}[X_m|\mathcal{F}_n] - \mathbb{E}[X_\infty|\mathcal{F}_n]|\right] &= \mathbb{E}\left[|\mathbb{E}[X_m - X_\infty|\mathcal{F}_n]|\right]\\ &\leq \mathbb{E}\left[\mathbb{E}[|X_m - X_\infty||\mathcal{F}_n]\right]\\ &= \mathbb{E}[|X_m - X_\infty|],\end{aligned}$$

so that $\mathbb{E}[X_m|\mathcal{F}_n]$ converges to $\mathbb{E}[X_\infty|\mathcal{F}_n]$ in L^1 as n tends to infinity. Since $\mathbb{E}[X_m|\mathcal{F}_n]$ does in fact not depend on $m \geq n$, we even have $\mathbb{E}[X_m|\mathcal{F}_n] = \mathbb{E}[X_\infty|\mathcal{F}_n]$. Finally,

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

$2 \Rightarrow 1$. For all $n \geq 0$, one has

$$\mathbb{E}[|X_n|] = \mathbb{E}[|\mathbb{E}[Z|\mathcal{F}_n]|] \leq \mathbb{E}[\mathbb{E}[|Z||\mathcal{F}_n]] = \mathbb{E}[|Z|].$$

Hence, $(X_n)_{n\geq 0}$ is bounded in L^1 and, thanks to Theorem 3.9, converges almost surely to an integrable random variable X_∞. We need to prove that $(X_n)_{n\geq 0}$ converges in L^1 to X_∞.

Let us do this first in the case where Z is bounded by a constant M, that is, under the assumption that $\mathbb{P}(|Z| \leq M) = 1$. Then, for all $n \geq 1$, the positivity of the conditional expectation implies $\mathbb{P}(|X_n| \leq M) = 1$. We can conclude, by the dominated convergence theorem, that $(X_n)_{n\geq 1}$ converges in L^1 to X_∞.

Let us now treat the general case. Let us fix $\varepsilon > 0$. There exists a constant $M > 0$ such that

$$\mathbb{E}\left[|Z - Z\mathbb{1}_{|Z|\leq M}|\right] < \varepsilon.$$

Let us choose such an M and set $\tilde{Z} = Z\mathbb{1}_{|Z|\leq M}$. For all $n \geq 0$, let us define $\tilde{X}_n = \mathbb{E}[\tilde{Z}|\mathcal{F}_n]$. Then for all $n \geq 0$,

$$\mathbb{E}[|X_n - \tilde{X}_n|] = \mathbb{E}[|\mathbb{E}[Z - \tilde{Z}|\mathcal{F}_n]|] \leq \mathbb{E}[|Z - \tilde{Z}|] < \varepsilon.$$

Since $\tilde{Z}$ is a bounded random variable, we know from our study of the bounded case that the martingale $(\tilde{X}_n)_{n\geq 0}$ converges in L^1. Thus, it is a Cauchy sequence in L^1, which is to say that there exists $n_0 \geq 1$ such that for all $n, m \geq n_0$, one has

$$\mathbb{E}[|\tilde{X}_n - \tilde{X}_m|] < \varepsilon.$$

Then, for all $n, m \geq n_0$, we have

$$\mathbb{E}[|X_n - X_m|] \leq \mathbb{E}[|X_n - \tilde{X}_n|] + \mathbb{E}[|\tilde{X}_n - \tilde{X}_m|] + \mathbb{E}[|\tilde{X}_m - X_m|] < 3\varepsilon.$$

Hence, the sequence $(X_n)_{n\geq 0}$ is Cauchy in L^1. Thus, it converges in L^1, and it must be towards X_∞. □

In this proof, we used several facts which it may be useful to review.

Exercise 3.29 *Prove that* $\|\mathbb{E}[X|\mathcal{F}] - \mathbb{E}[Y|\mathcal{F}]\|_{L^1} \leq \|X - Y\|_{L^1}$. *In other words, as a mapping of* L^1 *into itself, conditional expectation is* 1*-Lipschitz. Prove in fact that, as a linear mapping from* L^1 *into itself, it has norm exactly* 1. *What about the norm of* $\mathbb{E}[\cdot|\mathcal{F}]$ *as a linear operator on* L^p *?*

Exercise 3.30 *Let* Z *be an integrable random variable. Let* $\varepsilon > 0$ *be fixed. Prove that there exists* $M > 0$ *such that* $\|Z - Z\mathbb{1}_{|Z|\leq M}\|_{L^1} < \varepsilon$.

Exercise 3.31 *Prove that if a sequence of random variables converges almost surely and in* L^1*, it must be to the same limit.*

The next exercise provides us with a refinement of the preceding theorem.

Exercise 3.32 *Let* $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ *be a filtered probability space. Let* Z *be an integrable random variable. Let* X_∞ *the almost sure and* L^1 *limit of the martingale* $(\mathbb{E}[Z|\mathcal{F}_n])_{n\geq 0}$. *Does the equality* $X_\infty = Z$ *necessarily hold ?*

Define the σ*-field* $\mathcal{F}_\infty = \sigma\left(\bigcup_{n\geq 0} \mathcal{F}_n\right)$. *Prove that* X_∞ *is* $\mathcal{F}_\infty$*-measurable. Prove that for all* $A \in \bigcup_{n\geq 0} \mathcal{F}_n$*, one has*

$$\int_A Z \, dP = \int_A X_\infty \, dP.$$

Prove, using the monotone class theorem, that the same equality holds for all $A \in \mathcal{F}_\infty$. *What is the conclusion of this argument ?*

So far, we have studied the almost sure convergence of martingales, observed that it is not always a convergence in L^1, and understood that it is equivalent for a martingale to be a closed martingale or to be convergent in L^1. If time allows, we shall come back to the question of convergence in L^1. Before that, we want to study convergence in L^p for $p > 1$. As often, the case $p = 2$ is particularly pleasant, but there is a very beautiful theory for general p which relies on Doob's maximal inequality. In order to study it, we will need a fundamental tool in the study of martingales, which is the theory of stopping. This is the subject of the next section.

3.8 Stopping times

Let us remember our understanding of a martingale in the picture involving a gambler. A gambler does not usually play forever, he stops playing at a certain point. The reason why he stops may be a mixture of various ingredients : lack of time, lack of money, bad luck, sense that he has won all that he could, time for dinner. In any case, the time at which the gambler stops playing can depend on how the game went : it is normally a random variable. However, if things are to be fair, the decision of stopping cannot depend on any information about the future of the game. The random time at which the player stops must thus have the following property : the decision of stopping or not at time n must be taken in a deterministic way from the information available at time n. The following definition encodes this property.

Definition 3.15 *Let $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. A random variable $T : \Omega \to \mathbb{N} \cup \{\infty\}$ is a* stopping time *if for all $n \geq 0$, the event $\{T = n\}$ belongs to $\mathcal{F}_n$.*

For example, a random time which is almost surely constant is a stopping time.

Exercise 3.33 *Prove that if T is a stopping time, then $\{T = \infty\}$ belongs to $\mathcal{A}$, and in fact to $\sigma\left(\bigcup_{n\geq 0} \mathcal{F}_n\right)$*

Exercise 3.34 *Prove that T is a stopping time if and only if for all $n \geq 0$, the event $\{T \leq n\}$ belongs to $\mathcal{F}_n$.*

Observe that the definition of a stopping time does not involve the probability $\mathbb{P}$. It only depends on the filtration. Here are a few simple but useful properties of stopping times.

Lemma 3.16 *Let $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. Let S, T be stopping times. Then $S \wedge T = \min(S, T)$, $S \vee T = \max(S, T)$, $S + T$ are stopping times.*

The proof is left as an exercise. A fundamental example of stopping time is the hitting time of a Borel subset by an adapted process.

Exercise 3.35 *Let $X = (X_n)_{n\geq 0}$ be an adapted process. Let B be a Borel subset of $\mathbb{R}$. Prove that $H = \inf\{n \geq 0 : X_n \in B\}$, with the convention $\inf \varnothing = \infty$, is a stopping time.*

As their name indicates, stopping times are meant to stop processes, in particular martingales. In fact, any integer-valued random variable can be used to stop any stochastic process, according to the following definition.

Definition 3.17 *Let $X = (X_n)_{n\geq 0}$ be a stochastic process. Let T be a random variable with values in $\mathbb{N} \cup \{\infty\}$. The stopped process $X^T = (X_{T\wedge n})_{n\geq 0}$ is defined by setting, for all $n \geq 0$ and all $\omega \in \Omega$,*

$$X_{T\wedge n}(\omega) = X_{T(\omega)\wedge n}(\omega) = \begin{cases} X_n(\omega) & \text{if } T(\omega) > n, \\ X_m(\omega) & \text{if } T(\omega) = m \leq n. \end{cases}$$

If T is finite almost surely, then the random variable X_T is well defined by the formula

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

Observe that in general, for X_T to be defined, one need to make sure that T is finite almost surely. If one knows that the sequence $(X_n)_{n\geq 0}$ converges almost surely to X_∞, one might however remove this assumption and set $X_T = X_\infty$ on the event $\{T = \infty\}$.

The first important result about stopping of martingales by stopping times is the following and we will deduce it from Theorem 3.8.

Proposition 3.18 *Let X be a supermartingale. Let T be a stopping time. The process X^T is a supermartingale.*

Proof. In the picture of the gambler, stopping a process at time T amounts to playing 1 until time T and then playing 0 forever. Let us define accordingly, for all $n \geq 1$,

$$H_n = \mathbb{1}_{\{T\geq n\}} = \mathbb{1}_{\{T\leq n-1\}^c}.$$

We play 1 during the n-th turn if and only if we did not decide to stop just after the end of the $n-1$-th turn. By construction, H is previsible. It is bounded and non-negative. Thus, $H \bullet X$ is a supermartingale. Moreover, for almost all ω, we have

$$\begin{aligned}
(H \bullet X)_n(\omega) &= \sum_{k=1}^{n} H_k(\omega)(X_k(\omega) - X_{k-1}(\omega)) \\
&= \sum_{k=1}^{n} \mathbb{1}_{T(\omega)\geq k}(X_k(\omega) - X_{k-1}(\omega)) \\
&= \sum_{k=1}^{T(\omega)\wedge n} X_k(\omega) - X_{k-1}(\omega) \\
&= X_{T(\omega)\wedge n}(\omega) - X_0(\omega).
\end{aligned}$$

Thus, X^T itself is a supermartingale. □

In particular, $\mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_0]$ and if X is a martingale, the equality holds : one has $\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_0]$. It is tempting to let n tend to infinity in this equation, but one must be very cautious in doing so. In order to be able to say that $\mathbb{E}[X_T] = \mathbb{E}[X_0]$, one must usually combine properties of X and T, in one of many possible ways. Let us see what can go wrong.

Firstly, as we already mentioned, the sequence $(X_{T\wedge n})_{n\geq 0}$ may not converge almost surely. Indeed, the event $\{T = +\infty\}$ may have positive probability, and X may not converge on it. However, if $T < \infty$ almost surely, it is true that $X_{T\wedge n}$ converges almost surely to X_T. Nevertheless, this convergence may not happen in L^1. The classical counter-example is that of the simple random walk (see Exercise 3.23). Indeed, let T be the hitting time of -1 for S, that is, $T = \inf\{n : S_n = -1\}$. Then T is finite almost surely, and S^T converges almost surely to -1, although $0 = \mathbb{E}[S_{T\wedge n}] \neq \mathbb{E}[S_T] = -1$.

Let us give three simple sets of conditions under which the expected result holds.

Theorem 3.19 *Let X be a supermartingale. Let T be a stopping time. If one of the following conditions is satisfied :*

1. T is bounded,

2. T is integrable and there exists $M > 0$ such that for all $n \geq 0$, $|X_{n+1} - X_n| \leq M$,

3. T is almost surely finite and X^T is bounded,

then X_T is integrable and the inequality $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$ holds.

Observe that the three successive sets of assumptions are in decreasing order of strength on T, and of increasing order of strength on X.

Proof. 1. If T is bounded by an integer N, then $\mathbb{E}[X_0] \geq \mathbb{E}[X_{T\wedge N}] = \mathbb{E}[X_T]$.

2. Since T is integrable, it is almost surely finite, so that $X_{T\wedge n}$ converges almost surely to X_T. Moreover, for all $n \geq 1$,

$$|X_{T\wedge n} - X_0| \leq \sum_{k=0}^{(T\wedge n)-1} |X_{k+1} - X_k| \leq MT.$$

Hence, the sequence $(X_{T\wedge n})_{n\geq 0}$ is dominated by the integrable random variable $MT + |X_0|$ and the dominated convergence theorem allows us to conclude that $X_{T\wedge n}$ converges in L^1 to X_T. In particular, X_T is integrable and $\mathbb{E}[X_T] = \lim_{n\to\infty} \mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_0]$.

3. If T is finite almost surely, then $X_{T\wedge n}$ converges almost surely to X_T. If moreover X^T is bounded, then the dominated convergence theorem ensures that the converges holds in L^1, and we conclude as in the previous case. □

Exercise 3.36 *Let S be the simple random walk. Consider two integers a and b such that $a < 0 < b$. Prove that S will almost surely visit the set $\{a, b\}$ (you can prove this without using any sophisticated results on the simple random walk). Compute the probability that S hits a before hitting b. Application : a gambler starts with a fortune of 1. He has decided to stop playing as soon as his fortune reaches 100, or when he has no money anymore. What is the probability that he returns home with empty pockets ?*

Exercise 3.37 *Prove that a random variable T with values in $\mathbb{N} \cup \{+\infty\}$ is a stopping time if and only if the process $H = (H_n)_{n\geq 1}$ defined by $H_n = \mathbb{1}_{\{T\geq n\}}$ is previsible.*

Exercise 3.38 *Let $H = (H_n)_{n\geq 1}$ be a stochastic process. Prove that the following two assertions are equivalent.*

1. H is adapted.

2. For every adapted process X, the process $H \bullet X$ is adapted.

Exercise 3.39 *Let T be a random variable with values in $\mathbb{N} \cup \{+\infty\}$. Prove that the following assertions are equivalent.*

1. For all $n \geq 0$, the event $\{T = n\}$ belongs to $\mathcal{F}_{n+1}$.

2. For every adapted process X, the stopped process X^T is adapted.

3. For every martingale X, the stopped process X^T is adapted.

3.9 Convergence in L^p for $p > 1$

In this section, we shall prove that a supermartingale which is bounded in L^p for some real $p > 1$ converges almost surely and in L^p. We already know that it converges almost surely, for boundedness in L^p implies boundedness in L^1. The point is to establish the convergence in L^p. For this, we shall use Doob's maximal inequality, for which we need the following lemma.

Lemma 3.20 *Let X be a submartingale. Let S and T be two bounded stopping times such that $S \leq T$ almost surely. Then $\mathbb{E}[X_S] \leq \mathbb{E}[X_T]$.*

Proof. Let our gambler start playing at time S and stop at time T. This amounts to defining, for all $n \geq 1$,

$$H_n = \mathbb{1}_{\{S<n\leq T\}} = \mathbb{1}_{\{S\leq n-1\}} \mathbb{1}_{\{T\leq n-1\}^c}.$$

The way we wrote it makes it clear that $H = (H_n)_{n\geq 0}$ is previsible. The same computation as in the proof of Proposition 3.18 yields

$$(H \bullet X)_n = X_{T\wedge n} - X_{S\wedge n}.$$

Let N be an integer such that $S \leq T \leq N$ almost surely. Then $(H \bullet X)_N = X_T - X_S$. Since H is bounded and non-negative, $H \bullet X$ is a submartingale, so that

$$\mathbb{E}[X_T - X_S] = \mathbb{E}[(H \bullet X)_N] \geq \mathbb{E}[(H \bullet X)_0] = 0,$$

as expected. □

Let us use this lemma to prove Doob's maximal inequality.

Proposition 3.21 (Doob's maximal inequality) *Let X be a submartingale. For all integer $n \geq 0$ and all real number a, the following inequalities hold :*

$$a\mathbb{P}\left(\sup_{0\leq k\leq n} X_k \geq a\right) \leq \mathbb{E}[X_n \mathbb{1}_{\{\sup_{0\leq k\leq n} X_k \geq a\}}] \leq \mathbb{E}[X_n^+].$$

Proof. Choose a real number a. Define T as the first hitting time of $[a, +\infty)$ for X:

$$T = \inf\{n \geq 0 : X_n \geq a\}.$$

Consider now an integer $n \geq 0$. The random variable $X_{T\wedge n}$ is equal to X_T, hence greater or equal to a, on the event $\{T \leq n\}$. On the complement of this event, $X_{T\wedge n}$ is equal to X_n. Observing that the events $\{T \leq n\}$ and $\{\sup_{0\leq k\leq n} X_k \geq a\}$ are the same, we thus find

$$X_{T\wedge n} \geq a\mathbb{1}_{\{\sup_{0\leq k\leq n} X_k\geq a\}} + X_n \mathbb{1}_{\{\sup_{0\leq k\leq n} X_k<a\}}.$$

Taking the expectation of both sides of this inequality, we find

$$a\mathbb{P}(\sup_{0\le k\le n} X_k \ge a) \le \mathbb{E}[X_{T\wedge n}] - \mathbb{E}[X_n \mathbb{1}_{\{\sup_{0\le k\le n} X_k < a\}}].$$

The previous lemma applied to the bounded stopping times $T \wedge n$ and n yields the inequality $\mathbb{E}[X_{T\wedge n}] \le \mathbb{E}[X_n]$, which combined with the previous one implies that

$$a\mathbb{P}(\sup_{0\le k\le n} X_k \ge a) \le \mathbb{E}[X_n(1 - \mathbb{1}_{\{\sup_{0\le k\le n} X_k < a\}})] = \mathbb{E}[X_n \mathbb{1}_{\{\sup_{0\le k\le n} X_k \ge a\}}].$$

This proves the first inequality. The second follows from an instance of the inequality $Z\mathbb{1}_A \le Z^+$ which is true for any random variable Z and any event A. □

Although one would quite naturally apply this result with a positive value for a, it seems that the result and the proof do not depend on any assumption on the sign of a. It is interesting to take a moment to think about what the theorem says when $a = 0$ (and even in this case the theorem is not trivial), and when $a < 0$.

Let us now take one further step towards the L^p convergence. For this, let us introduce a notation. If $X = (X_n)_{n\ge 0}$ is a stochastic process, let us define its maximal process $X^* = (X_n^*)_{n\ge 0}$ by setting, for all $n \ge 0$,

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

Proposition 3.22 *Let X be a martingale. Let $p > 1$ be a real number. For all integer $n \ge 1$, one has*

$$\mathbb{E}\left[(X_n^*)^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p].$$

Observe that we are only considering non-negative quantities in this statement. Hence, we do not need to assume that X belongs to L^p.

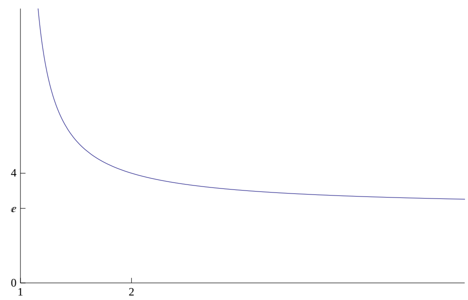


Figure 9: The graph of the function $p \mapsto \left(\dfrac{p}{p-1}\right)^p$.

Proof. By the previous proposition applied to the submartingale $(|X_n|)_{n\geq 0}$, we have, for all $a > 0$,

$$a\mathbb{P}(X_n^* \geq a) \leq \mathbb{E}[|X_n|\mathbb{1}_{\{X_n^*\geq a\}}].$$

Let us multiply both sides of this inequality by a^{p-2} and integrate with respect to a from 0 to $+\infty$. On the left, we find

$$\int_0^{+\infty} a^{p-1}\mathbb{P}(X_n^* \geq a)\, da = \mathbb{E}\left[\int_0^{X_n^*} a^{p-1}\, da\right] = \frac{1}{p}\mathbb{E}[(X_n^*)^p].$$

On the right, we find

$$\begin{aligned}
\int_0^{+\infty} a^{p-2}\mathbb{E}[|X_n|\mathbb{1}_{\{X_n^*>a\}}]\, da &= \mathbb{E}\left[|X_n|\int_0^{X_n^*} a^{p-2}\, da\right] \\
&= \frac{1}{p-1}\mathbb{E}[|X_n|(X_n^*)^{p-1}] \\
&\leq \frac{1}{p-1}\mathbb{E}[|X_n|^p]^{\frac{1}{p}}\mathbb{E}[(X_n^*)^p]^{\frac{p-1}{p}},
\end{aligned}$$

where the last inequality is Hölder's inequality. Combining the two expressions, we find the expected inequality. □

As n tends to infinity, X_n^* increases and converges almost surely to

$$X_\infty^* = \sup\{|X_n| : n \geq 0\}.$$

Let us now prove the convergence result in L^p.

Theorem 3.23 *Let X be a martingale. Let $p > 1$ be a real number. Assume that X is bounded in L^p. Then X converges almost surely and in L^p towards a random variable X_∞ which satisfies*

$$E[|X_\infty|^p] = \sup\{\mathbb{E}[|X_n|^p] : n \geq 0\}.$$

Moreover, one has

$$\mathbb{E}[(X_\infty^*)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_\infty|^p].$$

Proof. Let us assume that X is bounded in L^p. In particular, X is bounded in L^1, hence it converges almost surely to a random variable X_∞.

The previous proposition, the fact that $(X_n^*)_{n\geq 0}$ is a non-decreasing sequence which converges towards X_∞^* and the monotone convergence theorem imply that

$$\mathbb{E}[(X_\infty^*)^p] \leq \left(\frac{p}{p-1}\right)^p \sup\{\mathbb{E}[|X_n|^p] : n \geq 0\} < \infty.$$

Hence, the almost sure converence of X_n to X_∞ is dominated in L^p by X_∞^*. By the dominated convergence theorem, the convergence holds in L^p. In particular, $\mathbb{E}[|X_\infty|^p]$ is the limit of the sequence $(\mathbb{E}[|X_n|^p])_{n\geq 0}$ which, by Jensen's inequality, is non-decreasing. Hence,

$$E[|X_\infty|^p] = \sup\{\mathbb{E}[|X_n|^p : n \geq 0\}.$$

The last inequality follows immediately. $\square$

The results of this section, as well as those of Section 3.5, fall in a category that one could call "rigidity results" for martingales. Doob's upcrossing lemma, which was the main result needed to prove the almost sure convergence of martingales bounded in L^1, can be rephrased by saying that if a martingale oscillates a lot, then its L^1 norm becomes large. Doob's maximal inequality says that if the largest value taken by a martingale is large in L^p, then the martingale itself is large in L^p. Neither of these results hold, even in very weak forms, for arbitrary stochastic processes.

Exercise 3.40 *The two equalities*

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{(-1)^n}{n} = -\log 2$$

are well known. Consider now a sequence $(\varepsilon_n)_{n\geq 1}$ *of i.i.d.random variables such that* $\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = \frac{1}{2}$. *Study the random series*

$$\sum_{n=1}^{\infty} \frac{\varepsilon_n}{n} \quad \text{and more generally} \quad \sum_{n=1}^{\infty} \frac{\varepsilon_n}{n^s},$$

where s *is an arbitrary complex number.*

3.10 Square-integrable martingales

As often in analysis when something can be done for all L^p spaces, the case $p = 2$ enjoys special properties. Martingales are no exception, the more so since the conditional expectation, which lies at the principle of the notion of martingale, is closely related to the geometry of L^2. Let us illustrate this by the following elementary but useful result.

Proposition 3.24 *The increments of a square-integrable martingale are orthogonal in* L^2. *More precisely, if* $X = (X_n)_{n\geq 0}$ *is a martingale on* $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ *such that* $\mathbb{E}[X_n^2] < \infty$ *for all* $n \geq 0$, *then for all integers* m, n, p *such that* $m \leq n \leq p$, *one has*

$$\mathbb{E}[(X_n - X_m)(X_p - X_n)] = 0.$$

Proof. Consider three integers $m \leq n \leq p$. We have

$$\begin{aligned} \mathbb{E}[(X_n - X_m)(X_p - X_n)] &= \mathbb{E}[\mathbb{E}[(X_n - X_m)(X_p - X_n)|\mathcal{F}_n]] \\ &= \mathbb{E}[(X_n - X_m)\mathbb{E}[X_p - X_n|\mathcal{F}_n]] \\ &= \mathbb{E}[(X_n - X_m).0] \\ &= 0, \end{aligned}$$

as expected. □

Exercise 3.41 *Is the converse true ? Is any square-integrable process $X = (X_n)_{n\geq 0}$ such that for all $m \leq n \leq p$ one has $\mathbb{E}[(X_n - X_m)(X_p - X_n)] = 0$ necessarily a martingale ?*

Exercise 3.42 *Let H be a Hilbert space. A curve $x : \mathbb{R} \to H, t \mapsto x_t$ is called a* helix *if for all reals $s \leq t \leq u$, the vectors $x_t - x_s$ and $x_u - x_t$ are perpendicular. Let $x = (x_t)_{t\geq 0}$ be a helix in a Hilbert space H. Prove that for all reals $r \leq s \leq t \leq u$, the vectors $x_s - x_r$ and $x_u - x_t$ are perpendicular. Prove that H is infinite dimensional. Find an example of a helix in your favorite separable infinite-dimensional Hilbert space, and prove that there exists a helix in any infinite-dimensional Hilbert space.*

A collection of orthogonal vectors in a Hilbert space begs us for an application of the Pythagorean theorem.

Proposition 3.25 *Let X be a square-integrable martingale. For all $n \geq 0$, one has*

$$\mathbb{E}[X_n^2] = \mathbb{E}[X_0^2] + \sum_{k=0}^{n-1} \mathbb{E}[(X_{k+1} - X_k)^2].$$

In partiular, X is bounded in L^2 if and only if

$$\sum_{k=0}^{\infty} \mathbb{E}[(X_{k+1} - X_k)^2] < +\infty.$$

Proof. This follows immediately from the previous proposition and the Pythagorean theorem. □

Exercise 3.43 *Prove that if $(Z_n)_{n\geq 0}$ is a sequence of independent random variables such that $\mathbb{E}[Z_n] = 0$ for all $n \geq 0$ and the series $\sum_{n\geq 0} \mathrm{Var}(Z_n)$ converges, then the series $\sum_{n\geq 0} Z_n$ converges almost surely and in L^2.*

The following result is fundamental. It is not specific of the square-integrable case, but we will see that it has very important consequences in this case.

Proposition 3.26 (Doob's decomposition) *Let $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space. Let $X = (X_n)_{n\geq 0}$ be an adapted integrable stochastic process.*

1. There exists a martingale $M = (M_n)_{n\geq 0}$ with $M_0 = 0$ and a previsible process $A = (A_n)_{n\geq 0}$ such that for all $n \geq 0$, one has

$$X_n = X_0 + M_n + A_n.$$

Moreover, this decomposition is unique.

2. The process X is a submartingale if and only if the process A is non-decreasing, that is, for all $n \geq 0$, $\mathbb{P}(A_n \leq A_{n+1}) = 1$.

Proof. If such a decomposition exists, then we must have $A_0 = 0$ and, for all $n \geq 0$,

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] = A_{n+1} - A_n.$$

This formula defines inductively a previsible process $(A_n)_{n\geq 0}$. Moreover, we must have $M_n = X_n - X_0 - A_n$ and the equality

$$\mathbb{E}[(X_{n+1} - A_{n+1}) - (X_n - A_n) | \mathcal{F}_n] = 0$$

shows that $(M_n)_{n\geq 0}$ is a martingale.

The second assertion follows immediately from the first line of computation above. □

For square-integrable martingales, this decomposition leads to the definition of a very important object.

Definition 3.27 *Let X be a square-integrable martingale. Let $X_n^2 = X_0^2 + M_n + A_n$ be the Doob decomposition of the integrable submartingale $(X_n^2)_{n\geq 0}$. The process $(A_n)_{n\geq 0}$ is called the* increasing process *associated to X and is often denoted by $\langle X \rangle = (\langle X \rangle_n)_{n\geq 0}$.*

The process $\langle X \rangle$ plays a crucial role in the study of continuous-time martingales and in the construction of the stochastic integral, one of the main results being the continuous-time analogue of the following property.

Exercise 3.44 *Let X be a martingale. Let H be a bounded previsible process. Prove that for all $n \geq 0$,*

$$\langle H \bullet X \rangle_n = \sum_{k=1}^{n} H_k^2 (\langle X \rangle_k - \langle X \rangle_{k-1}).$$

This can be written informally as

$$\left\langle \int_0^{\cdot} H \, dX \right\rangle_n = \int_0^n H^2 \, d\langle X \rangle.$$

Using the increasing process associated with a square-integrable martingale, we can refine the theorem of convergence. For this, let us introduce

$$\langle X \rangle_\infty = \lim_{n\to\infty} \langle X \rangle_n,$$

which exists almost surely as the limit of a non-decreasing sequence.

Exercise 3.45 *Check that $\mathbb{E}[X_n^2] = \mathbb{E}[\langle X \rangle_n]$ for all $n \geq 0$. Prove that X is bounded in L^2 if and only if $\langle X \rangle_\infty$ is integrable.*

Theorem 3.28 *Let X be a square-integrable martingale. The sequence $(X_n)_{n\geq 0}$ converges almost surely on the event $\{\langle X \rangle_\infty < \infty\}$.*

Proof. Choose an integer $k \geq 0$ and define

$$T_k = \inf\{n \geq 0 : \langle X\rangle_{n+1} > k\}.$$

It is a stopping time, because it is the hitting time of the Borel subset $(k, +\infty)$ of $\mathbb{R}$ by the adapted process $(\langle X\rangle_{n+1})_{n\geq 0}$. We claim that the process $\langle X\rangle^{T_k}$ is the increasing process associated to X^{T_k}.

Firstly, since $X^2 - \langle X\rangle$ is a martingale and T_k a stopping time, the process

$$(X^2 - \langle X\rangle)^{T_k} = (X^{T_k})^2 - \langle X\rangle^{T_k}$$

is a martingale. It remains to prove that $\langle X\rangle^{T_k}$ is previsible.

Choose $n \geq 1$ and B a Borel subset of $\mathbb{R}$. Then

$$\begin{aligned}
\{\langle X\rangle^{T_k}_n \in B\} &= \{\langle X\rangle_{T_k \wedge n} \in B\} \\
&= (\{T_k \geq n\} \cap \{\langle X\rangle_n \in B\}) \cup (\{T_k < n\} \cap \{\langle X\rangle_{T_k} \in B\}) \\
&= (\{T_k \leq n-1\}^c \cap \{\langle X\rangle_n \in B\}) \cup \bigcup_{m=0}^{n-1} (\{T_k = m\} \cap \{\langle X\rangle_m \in B\})
\end{aligned}$$

and this way of writing this event makes it apparent that it belongs to $\mathcal{F}_{n-1}$.

The increasing process of the martingale X^{T_k} is bounded by k by construction, so that this martingale is bounded in L^2, from which it follows that it converges almost surely.

On the event $\langle X\rangle_\infty \leq k$, the stopping time T_k is equal to ∞, and the processes X and X^{T_k} are equal. Hence, X itself converges almost surely on $\{\langle X\rangle_\infty \leq k\}$.

Finally, the equality $\{\langle X\rangle_\infty < \infty\} = \bigcup_{k\geq 1}\{\langle X\rangle_\infty \leq k\}$ implies that X converges almost surely on the whole event $\{\langle X\rangle_\infty < \infty\}$. □

Exercise 3.46 *Give an example of a sequence $(z_n)_{n\geq 0}$ of real numbers such that $\sum_{n\geq 0} z_n^2$ converges, but $\sum_{n\geq 0} z_n$ does not.*

Let $(Z_n)_{n\geq 0}$ be a sequence of independent random variables such that $\mathbb{E}[Z_n] = 0$ for all $n \geq 0$. Set $X_0 = 0$ and, for all $n \geq 1$, $X_n = Z_1 + \ldots + Z_n$. Compute $\langle X\rangle$. What does the last theorem say in this situation ?

3.11 Uniform integrability

In this section, we will give a much more detailed answer to the question of knowing when a martingale which is bounded in L^1 converges in L^1. The crucial tool is the notion of uniform integrability.

Definition 3.29 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A family $(X_i)_{i\in I}$ of integrable random variables is said to be* uniformly integrable *if for all $\varepsilon > 0$ there exists $M > 0$ such that*

$$\forall i \in I,\ \mathbb{E}\left[|X_i|\mathbb{1}_{\{|X_i|>M\}}\right] < \varepsilon.$$

An equivalent formulation of this definition is

$$\lim_{M\to\infty}\left(\sup_{i\in I}\mathbb{E}\left[|X_i|\mathbb{1}_{\{|X_i|>M\}}\right]\right)=0.$$

Exercise 3.47 *Prove that a uniformly integrable family is bounded in L^1.*

Exercise 3.48 *Let $(X_i)_{i\in I}$ be a family of integrable random variables. Prove that it is uniformly integrable as soon as one of the following assumptions is satisfied.*

- *The index set I is finite.*
- *There exists an integrable random variable Z such that $|X_i|\le Z$ for all $i\in I$.*
- *The random variable $\sup_{i\in I}|X_i|$ is integrable.*
- *There exists $p>1$ such that the family $(X_i)_{i\in I}$ is bounded in L^p.*

Consider the probability space $([0,1],\mathscr{B}_{[0,1]},\lambda)$. For each $n\ge 1$, set $m=\lfloor\log_2 n\rfloor$, so that $n=2^m+k$ with $k\in\{0,\dots,2^m-1\}$, and define the random variable

$$X_n=\frac{2^m}{\log(m+1)}\mathbb{1}_{\left[\frac{k}{2^m},\frac{k+1}{2^m}\right]}.$$

Prove that the family $(X_n)_{n\ge1}$ is uniformly integrable but does not satisfy any of the properties above. Incidentally, does the sequence $(X_n)_{n\ge1}$ converge, and in which sense ?

The name *uniform integrability* is justified by the following proposition.

Proposition 3.30 *Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space. Let $(X_i)_{i\in I}$ be a family of random variables which is bounded in L^1. The following two assertions are equivalent.*

1. *The family $(X_i)_{i\in I}$ is uniformly integrable.*
2. *For all $\varepsilon>0$ there exists $\delta>0$ such that for all $A\in\mathcal{A}$ such that $\mathbb{P}(A)<\delta$, one has*

$$\forall i\in I,\ \int_A|X_i|\,d\mathbb{P}<\varepsilon.$$

Proof. $1\Rightarrow 2$. Choose $\varepsilon>0$. Since the family $(X_i)_{i\in I}$ is uniformly integrable, there exists a real $M>0$ such that for all $i\in I$, $\mathbb{E}[|X_i|\mathbb{1}_{|X_i|>M}]<\frac{\varepsilon}{2}$. Let A be any event in $\mathcal{A}$ such that $\mathbb{P}(A)<\frac{\varepsilon}{2M}$. Then

$$\begin{aligned}\int_A|X_i|\,d\mathbb{P}&=\int_{A\cap\{|X_i|\le M\}}|X_i|\,d\mathbb{P}+\int_{A\cap\{|X_i|>M\}}|X_i|\,d\mathbb{P}\\&\le\int_A M\,d\mathbb{P}+\int_{\{|X_i|>M\}}|X_i|\,d\mathbb{P}\\&<M\mathbb{P}(A)+\frac{\varepsilon}{2}\\&<\varepsilon.\end{aligned}$$

$2 \Rightarrow 1$. Let $K > 0$ be such that $\mathbb{E}[|X_i|] \leq K$ for all $i \in I$. Choose $\varepsilon > 0$. Let $\delta > 0$ be such that for all $A \in \mathcal{A}$, $\mathbb{P}(A) < \delta$ implies $\int_A |X_i| \, d\mathbb{P} < \varepsilon$ for all $i \in I$. Set $M = \frac{2K}{\delta}$. Then for all $i \in I$,

$$\mathbb{P}(|X_i| > M) \leq \frac{1}{M} \int_{\{|X_i|>M\}} |X_i| \, dP \leq \frac{K}{M} \leq \frac{\delta}{2} < \delta,$$

so that

$$\int_{\{|X_i|>M\}} |X_i| \, d\mathbb{P} < \varepsilon,$$

and the proof is finished. □

Exercise 3.49 *When did we use the assumption that the family $(X_i)_{i\in I}$ is bounded in L^1? Prove that if the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is such that Ω is a finite set, then the second assertion is true for any family of random variables. Under which assumption on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ can one deduce from the second assertion that the family $(X_i)_{i\in I}$ is bounded in L^1?*

Corollary 3.31 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let Z be an integrable variable. The family $\{\mathbb{E}[Z|\mathcal{B}] : \mathcal{B}$ sub-σ-field of $\mathcal{A}\}$ is uniformly integrable.*

Proof. For each sub-σ-field $\mathcal{B}$ of $\mathcal{A}$, the random variable $\mathbb{E}[Z|\mathcal{B}]$ is integrable. Let us now choose ε. Thanks to the uniform integrability of the family which consists in the single random variable Z, let us consider $\delta > 0$ such that $\mathbb{P}(A) < \delta$ implies $\int_A |Z| \, dP < \varepsilon$. Set $M = \frac{2}{\delta}\mathbb{E}[|Z|]$. Let $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$. We claim that

$$\mathbb{E}\left[|\mathbb{E}[Z|\mathcal{B}]|\mathbb{1}_{\{|\mathbb{E}[Z|\mathcal{B}]|>M\}}\right] < \varepsilon.$$

Indeed, we have

$$\mathbb{P}(|\mathbb{E}[Z|\mathcal{B}]| > M) \leq \frac{1}{M}\mathbb{E}[|\mathbb{E}[Z|\mathcal{B}]|] \leq \frac{1}{M}\mathbb{E}[\mathbb{E}[|Z||\mathcal{B}]] = \frac{1}{M}\mathbb{E}[|Z|] < \delta.$$

Hence,

$$\int_{\{|\mathbb{E}[Z|\mathcal{B}]|>M\}} |\mathbb{E}[Z|\mathcal{B}]| \, d\mathbb{P} \leq \int_{\{|\mathbb{E}[Z|\mathcal{B}]|>M\}} |Z| \, d\mathbb{P} < \varepsilon$$

and the result is proved. □

The reason why uniform integrability is so useful in the context of convergence of martingales is the following result, which is a stronger form of the dominated convergence theorem (but it is only valid on finite measure spaces).

Theorem 3.32 *Let $(X_n)_{n\geq 0}$ be a sequence of integrable random variables. Let X_∞ be a random variable. The following assertions are equivalent.*

1. *The sequence $(X_n)_{n\geq 0}$ converges in L^1 to X_∞.*
2. *The sequence $(X_n)_{n\geq 0}$ is uniformly integrable and converges in probability to X_∞.*

Proof. $1 \Rightarrow 2$. We know that convergence in L^1 implies convergence in probability. Let us prove that $(X_n)_{n\geq 0}$ is uniformly integrable.

Choose $\varepsilon > 0$. Let $\delta_1 > 0$ be such that $\mathbb{P}(A) < \delta_1$ implies $\int_A |X_\infty| \, d\mathbb{P} < \frac{\varepsilon}{2}$. Let n_0 be such that for all $n \geq n_0$, $\mathbb{E}[|X_n - X_\infty|] < \frac{\varepsilon}{2}$. Let $\delta_2 > 0$ be such that for all $n \leq n_0$ and all A with $\mathbb{P}(A) < \delta_2$, we have $\int_A |X_n| \, d\mathbb{P} < \varepsilon$.

Now set $\delta = \min(\delta_1, \delta_2)$. Choose $A \in \mathcal{A}$ such that $\mathbb{P}(A) < \delta$. Choose $n \geq 0$. If $n \leq n_0$, then since $\mathbb{P}(A) < \delta_2$, we have $\int_A |X_n| \, d\mathbb{P}$. If $n \geq n_0$, then

$$\int_A |X_n| \, d\mathbb{P} \leq \int_A |X_n - X_\infty| \, d\mathbb{P} + \int_A |X_\infty| \, d\mathbb{P} < \mathbb{E}[|X_n - X_\infty|] + \frac{\varepsilon}{2} < \varepsilon.$$

$2 \Rightarrow 1$. From Proposition 3.30 it follows that the family $(X_n - X_m)_{n,m\geq 0}$ is uniformly integrable. Choose $\varepsilon > 0$. Let $M > 0$ be such that $\mathbb{E}[|X_n - X_m| \mathbb{1}_{\{|X_n - X_m| > M\}}] < \varepsilon$ for all $n, m \geq 0$. Thus,

$$\begin{aligned}
\mathbb{E}[|X_n - X_m|] &= \mathbb{E}[|X_n - X_m| \mathbb{1}_{\{|X_n - X_m| < \varepsilon\}}] + \mathbb{E}[|X_n - X_m| \mathbb{1}_{\{\varepsilon \leq |X_n - X_m| \leq M\}}] \\
&\qquad + \mathbb{E}[|X_n - X_m| \mathbb{1}_{\{|X_n - X_m| > M\}}] \\
&\leq 2\varepsilon + \mathbb{E}[|X_n - X_m| \mathbb{1}_{\{\varepsilon \leq |X_n - X_m| \leq M\}}] \\
&\leq 2\varepsilon + M\mathbb{P}(|X_n - X_m| > \varepsilon).
\end{aligned}$$

Let n_0 be such that for all $n \geq n_0$, $\mathbb{P}(|X_n - X_\infty| > \frac{\varepsilon}{2}) < \frac{\varepsilon}{2M}$. Then for all $n, m \geq n_0$, we have

$$\mathbb{P}(|X_n - X_m| > \varepsilon) \leq \mathbb{P}\left(|X_n - X_\infty| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_m - X_\infty| > \frac{\varepsilon}{2}\right) < \frac{\varepsilon}{M}.$$

Thus, for all $n, m \geq n_0$ we have

$$\mathbb{E}[|X_n - X_m|] \leq 3\varepsilon.$$

The sequence $(X_n)_{n\geq 0}$ is a Cauchy sequence in L^1. Hence, it converges in L^1, and its limit must be X_∞. □

The proof of the following result is left as an exercise.

Theorem 3.33 *Let $X = (X_n)_{n\geq 0}$ be a martingale. The following assertions are equivalent.*

1. *The martingale X converges in L^1.*
2. *The family $(X_n)_{n\geq 0}$ is uniformly integrable.*
3. *There exists an integrable random variable such that $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ for all $n \geq 0$.*

Moreover, if any of these assertions holds, then the martingale converges almost surely.

We would like to complete this picture by extending the stopping theorem to the case of uniformly integrable martingales. For this, we need to introduce the notion of σ-field associated with a stopping time.

Definition 3.34 *Let* $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ *be a filtered probability space. Let* T *be a stopping time. The collection of events*

$$\mathcal{F}_T = \{A \in \mathcal{A} : \forall n \geq 0, A \cap \{T = n\} \in \mathcal{F}_n\}$$

is a sub-σ*-field of* $\mathcal{A}$ *called the the* σ*-field of the past up to time* T*.*

Exercise 3.50 *Prove that* $\mathcal{F}_T$ *is indeed a* σ*-field. Compute* $\mathcal{F}_T$ *when* $T = n$ *almost surely. Compute* $\mathcal{F}_T$ *when* $T = \infty$ *almost surely.*

Exercise 3.51 *Prove that a subset* A *of* Ω *belongs to* $\mathcal{F}_T$ *if and only if there exists a sequence* $(A_n)_{n\geq 0}$ *of events, and an event* A_∞*, such that* $A_\infty \in \mathcal{A}$ *and* $A_n \in \mathcal{F}_n$ *for every* $n \geq 0$*, such that*

$$A = (A_\infty \cap \{T = \infty\}) \cup \bigcup_{n=0}^{\infty} (A_n \cap \{T = n\}).$$

Lemma 3.35 *Let* S *and* T *two stopping times. If* $S \leq T$*, then* $\mathcal{F}_S \subset \mathcal{F}_T$*.*

Proof. Consider $A \in \mathcal{F}_S$. Choose $n \geq 0$. Then the event

$$A \cap \{T = n\} = \bigcup_{k=0}^{\infty} (A \cap \{T = n\} \cap \{S = k\}) = \bigcup_{k=0}^{n} (A \cap \{S = k\} \cap \{T = n\})$$

belongs to $\mathcal{F}_n$. Hence, A belongs to $\mathcal{F}_T$. □

Lemma 3.36 *Let* $X = (X_n)_{n\geq 0}$ *be an adapted stochastic process. Let* T *be a stopping time. If one of the following assumptions is satisfied :*

1. T *is finite almost surely,*
2. X_n *converges almost surely to* X_∞*,*

then X_T *is well defined and* $\mathcal{F}_T$*-measurable.*

Proof. We know already that X_T is well defined under any of the two assumptions. Let us prove that X_T is $\mathcal{F}_T$ measurable. For this, let us consider a Borel subset B of $\mathbb{R}$ and an integer $n \geq 0$. We have

$$\{X_T \in B\} \cap \{T = n\} = \{X_n \in B\} \cap \{T = n\} \in \mathcal{F}_n.$$

Since this holds for all $n \geq 0$, $\{X_T \in B\}$ belongs to $\mathcal{F}_T$. □

Let us now state and prove the stopping theorem for uniformly integrable martingales.

Theorem 3.37 *Let* X *be a uniformly integrable martingale. Let* T *be a stopping time. Then*

$$X_T = \mathbb{E}[X_\infty | \mathcal{F}_T].$$

In particular, $\mathbb{E}[X_T] = \mathbb{E}[X_\infty] = \mathbb{E}[X_n]$ *for all* $n \geq 0$*. Moreover, if* S *and* T *are two stopping times such that* $S \leq T$*, then*

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

Proof. Let us check that X_T is integrable. We have

$$\begin{aligned}
\mathbb{E}[|X_T|] &= \sum_{n=0}^{\infty} \mathbb{E}[|X_n| \mathbb{1}_{\{T=n\}}] + \mathbb{E}[|X_\infty| \mathbb{1}_{\{T=\infty\}}] \\
&\leq \sum_{n=0}^{\infty} \mathbb{E}\left[\mathbb{E}\left[|X_\infty| \middle| \mathcal{F}_n\right] \mathbb{1}_{\{T=n\}}\right] + \mathbb{E}[|X_\infty| \mathbb{1}_{\{T=\infty\}}] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[|X_\infty| \mathbb{1}_{\{T=n\}}\right] + \mathbb{E}[|X_\infty| \mathbb{1}_{\{T=\infty\}}] \\
&= \mathbb{E}[|X_\infty|] < \infty.
\end{aligned}$$

Let us now choose $A \in \mathcal{F}_T$. We have

$$\begin{aligned}
\mathbb{E}[X_\infty \mathbb{1}_A] &= \sum_{n=0}^{\infty} \mathbb{E}[X_\infty \mathbb{1}_{A \cap \{T=n\}}] + \mathbb{E}[X_\infty \mathbb{1}_{A \cap \{T=\infty\}}] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[\mathbb{E}[X_\infty | \mathcal{F}_n] \mathbb{1}_{A \cap \{T=n\}}] + \mathbb{E}[X_\infty \mathbb{1}_{A \cap \{T=\infty\}}] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[X_n \mathbb{1}_{A \cap \{T=n\}}] + \mathbb{E}[X_\infty \mathbb{1}_{A \cap \{T=\infty\}}] \\
&= \mathbb{E}[X_T \mathbb{1}_A].
\end{aligned}$$

This proves that $X_T = \mathbb{E}[X_\infty | \mathcal{F}_T]$.

It follows that $\mathbb{E}[X_T] = \mathbb{E}[X_\infty]$. We already know that $\mathbb{E}[X_n] = \mathbb{E}[\mathbb{E}[X_\infty | \mathcal{F}_n]] = \mathbb{E}[X_\infty]$. Finally, if $S \leq T$, then since $\mathcal{F}_S \subset \mathcal{F}_T$, we have

$$X_S = \mathbb{E}[X_\infty | \mathcal{F}_S] = \mathbb{E}[\mathbb{E}[X_\infty | \mathcal{F}_T] | \mathcal{F}_S] = \mathbb{E}[X_T | \mathcal{F}_S]$$

and the proof is finished. □

Exercise 3.52 *Let X be a uniformly integrable martingale. Let T be a stopping time. Prove that $X_{T \wedge n}$ converges almost surely and in L^1 to X_T.*

3.12 Backward martingales

The name *backward martingales* may be misleading. We are not going to run time backwards but rather allow time to be unbounded below.

Definition 3.38 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A* backward filtration *is a non-decreasing sequence $(\mathcal{F}_n)_{n \leq 0}$ of sub-σ-fields of $\mathcal{A}$:*

$$\ldots \subset \mathcal{F}_{-n-1} \subset \mathcal{F}_{-n} \subset \ldots \subset \mathcal{F}_{-1} \subset \mathcal{F}_0.$$

A backward martingale on the backward filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\leq 0}, \mathbb{P})$ is a sequence $X = (X_n)_{n\leq 0}$ such that for all $n \leq 0$, X_n is integrable and $\mathcal{F}_n$-measurable, and

$$\mathbb{E}[X_n|\mathcal{F}_{n-1}] = X_{n-1}.$$

Having time unbounded below instead of unbounded above makes a huge difference.

Lemma 3.39 *A backward martingale is uniformly integrable.*

Proof. Indeed, we have for all $n \leq 0$ the equality $X_n = \mathbb{E}[X_0|\mathcal{F}_n]$, and the result is a consequence of Corollary 3.31. □

It is thus not too surprising that the problem of convergence of backward martingales is simpler than that of martingales.

Theorem 3.40 *Let X be a backward martingale. Set $\mathcal{F}_{-\infty} = \bigcap_{n\leq 0} \mathcal{F}_n$. Then, as n tends to $-\infty$, X_n converges almost surely and in L^1 to $\mathbb{E}[X_0|\mathcal{F}_{-\infty}]$.*

Proof. The proof that X converges almost surely is the same as that for usual martingales. Indeed, for all $N \leq 0$, the sequence $X_N, X_{N+1}, \ldots, X_0$ is a usual martingale and for all $a < b$, the number of upcrossings of this martingale can be estimated by Doob's upcrossing lemma. As N tends to $-\infty$, this number of upcrossings converges to the number of upcrossings of the full backward martingale X, and turns out to be finite almost surely for all $a < b$. Thus, X_n converges almost surely as n tends to $-\infty$, towards a random variable which we denote by $X_{-\infty}$.

Since, by the preceding lemma, X is uniformly integrable, it also converges to $X_{-\infty}$ in L^1.

The random variable $X_{-\infty}$ is $\mathcal{F}_n$-measurable for each $n \leq 0$, because it is the limit of the sequence $X_n, X_{n-1}, X_{n-2}, \ldots$. Thus, $X_{-\infty}$ is $\mathcal{F}_{-\infty}$ measurable. Now if A belongs to $\mathcal{F}_{-\infty}$, then the L^1 convergence of X_n to X_∞ implies

$$\int_A X_{-\infty}\, d\mathbb{P} = \lim_{n\to-\infty} \int_A X_n\, d\mathbb{P} = \lim_{n\to-\infty} \int_A \mathbb{E}[X_0|\mathcal{F}_n]\, d\mathbb{P} = \lim_{n\to-\infty} \int_A X_0\, d\mathbb{P} = \int_A X_0\, d\mathbb{P},$$

because A belongs to $\mathcal{F}_n$ for all $n \leq 0$. Hence, $X_{-\infty} = \mathbb{E}[X_0|\mathcal{F}_{-\infty}]$ and the proof is finished. □

This result of convergence has several spectacular consequences, including a proof of the strong law of large numbers. This proof however requires some preparatory results.

Lemma 3.41 *Let $X, Y_1, \ldots, Y_n$ and $X', Y'_1, \ldots, Y'_n$ be random variables. Assume that the random vectors $(X, Y_1, \ldots, Y_n)$ and $(X', Y'_1, \ldots, Y'_n)$ have the same distribution. Assume that X and X' are integrable. Assume finally that $h : \mathbb{R}^n \to \mathbb{R}$ is a measurable function such that*

$$\mathbb{E}[X|\sigma(Y_1, \ldots, Y_n)] = h(Y_1, \ldots, Y_n).$$

Then

$$\mathbb{E}[X'|\sigma(Y'_1, \ldots, Y'_n)] = h(Y'_1, \ldots, Y'_n),$$

with the same function h.

Proof. Let B be a Borel subset of $\mathbb{R}^n$. The assumptions imply that

$$\int_\Omega X \mathbb{1}_B(Y_1, \ldots, Y_n) \, dP = \int_\Omega h(Y_1, \ldots, Y_n) \mathbb{1}_B(Y_1, \ldots, Y_n) \, dP.$$

Denoting by μ the distribution of $(X, Y_1, \ldots, Y_n)$ (which is a probability measure on $\mathbb{R}^{n+1}$) and by ν the distribution of $(Y_1, \ldots, Y_n)$ (which is a probability measure on $\mathbb{R}^n$), this rewrites as

$$\int_{\mathbb{R}^{n+1}} x \mathbb{1}_B(y_1, \ldots, y_n) \mu(dx, dy_1, \ldots, dy_n) = \int_{\mathbb{R}^n} h(y_1, \ldots, y_n) \mathbb{1}_B(y_1, \ldots, y_n) \nu(dy_1, \ldots, dy_n).$$

Since $(X, Y_1, \ldots, Y_n)$ and $(X', Y'_1, \ldots, Y'_n)$ have the same distribution, the last equality also says that

$$\int_\Omega X' \mathbb{1}_B(Y'_1, \ldots, Y'_n) \, dP = \int_\Omega h(Y'_1, \ldots, Y'_n) \mathbb{1}_B(Y'_1, \ldots, Y'_n) \, dP,$$

that is,

$$\mathbb{E}[X'|\sigma(Y'_1, \ldots, Y'_n)] = h(Y'_1, \ldots, Y'_n)$$

as expected. □

This lemma has the following consequence.

Lemma 3.42 *Let $(X_n)_{n \geq 1}$ be an i.i.d. sequence of integrable random variables. For all $n \geq 1$, set $S_n = X_1 + \ldots + X_n$. Then for all $n \geq 1$, and all $k \in \{1, \ldots, n\}$,*

$$\mathbb{E}[X_k|S_n] = \frac{S_n}{n} = \mathbb{E}[X_k|\sigma(S_n, S_{n+1}, S_{n+2}, \ldots)].$$

Proof. For all $k, l \in \{1, \ldots, n\}$, the vectors (X_k, S_n) and (X_l, S_n) have the same distribution. Indeed, assuming $k < l$, the vectors $(X_1, \ldots, X_n)$ and $(X_1, \ldots, \underset{k}{X_l}, \ldots, \underset{l}{X_k}, \ldots, X_n)$ with X_k and X_l exchanged have the same distribution, so that (X_k, S_n) and (X_l, S_n), which are respectively obtained from these two vectors by applying the function

$$(x_1, \ldots, x_n) \mapsto (x_k, x_1 + \ldots + x_n),$$

also have the same distribution.

Let $h : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $\mathbb{E}[X_1|S_n] = h(S_n)$. By the previous lemma, we have $\mathbb{E}[X_k|S_n] = h(S_n)$ for all $k \in \{1, \ldots, n\}$. Hence,

$$S_n = \mathbb{E}[S_n|S_n] = \sum_{k=1}^n \mathbb{E}[X_k|S_n] = nh(S_n),$$

so that $h(S_n) = \frac{S_n}{n}$. This proves the first equality.

In order to prove the second, observe that $\sigma(S_n, S_{n+1}, \ldots) = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots)$. We need to prove that for all event $A \in \sigma(S_n, X_{n+1}, X_{n+2}, \ldots)$, we have

$$\mathbb{E}[X_k \mathbb{1}_A] = \frac{1}{n}\mathbb{E}[S_n \mathbb{1}_A].$$

Let us denote by $\mathcal{C}$ the class of events $A \in \mathcal{A}$ for which this equality holds. It is a λ-system. Let us consider an event $B \in \sigma(S_n)$ and an event $C \in \sigma(X_{n+1}, X_{n+2}, \ldots)$. Since the σ-fields $\sigma(X_k, S_n)$ and $\sigma(X_{n+1}, X_{n+2}, \ldots)$ are independent, we have

$$\mathbb{E}[X_k \mathbb{1}_B \mathbb{1}_C] = \mathbb{E}[X_k \mathbb{1}_B]\mathbb{E}[\mathbb{1}_C] = \frac{1}{n}\mathbb{E}[S_n \mathbb{1}_B]\mathbb{E}[\mathbb{1}_C] = \frac{1}{n}\mathbb{E}[S_n \mathbb{1}_B \mathbb{1}_C].$$

Hence, the λ-system $\mathcal{C}$ contains the π-system of all events of the form $B \cap C$ with $B \in \sigma(S_n)$ and $C \in \sigma(X_{n+1}, X_{n+2}, \ldots)$. Thus, by the monotone class theorem, it contains $\sigma(S_n, X_{n+1}, X_{n+2}, \ldots)$ and the second equality is proved. □

Theorem 3.43 (Strong law of large numbers) *Let $(X_n)_{n\geq 1}$ be an i.i.d. sequence of integrable random variables. Then as n tends to infinity,*

$$\frac{X_1 + \ldots + X_n}{n} \longrightarrow \mathbb{E}[X_1]$$

almost surely and in L^1.

Proof. Set $S_0 = 0$ and, for all $n \geq 1$, $S_n = X_1 + \ldots, X_n$. For all $n \geq 0$, set $\mathcal{F}_{-n} = \sigma(S_n, S_{n+1}, \ldots) = \sigma(S_k : k \geq n)$. Then $(\mathcal{F}_n)_{n\leq 0}$ is a backward filtration. For each $n \geq 1$, we have

$$\mathbb{E}\left[\frac{S_n}{n}\middle|\mathcal{F}_{-n-1}\right] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k|\mathcal{F}_{-n-1}] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k|S_{n+1}, S_{n+2}, \ldots] = \frac{S_{n+1}}{n+1}.$$

In other words, $(\frac{S_{-n}}{-n})_{n\leq 1}$ is a backward martingale with respect to its natural filtration. This implies that the sequence $\left(\frac{S_n}{n}\right)_{n\geq 0}$ converges almost surely and in L^1. Its limit is

$$Y = \lim_{n\to\infty}\frac{X_1 + \ldots + X_n}{n},$$

but for each $n_0 \geq 0$, it is also

$$Y = \lim_{n\to\infty}\frac{X_{n_0} + \ldots + X_n}{n}$$

almost surely, so that Y is measurable with respect to $\sigma(X_n : n \geq n_0)$. Since this holds for all $n_0 \geq 1$, the random variable Y is measurable with respect to the tail σ-field

$$\bigcap_{n\geq 1}\sigma(X_k : k \geq n)$$

and Kolmogorov's 0-1 law asserts that this σ-field is trivial. Hence, the limit Y is a constant. Since $\mathbb{E}[Y] = \mathbb{E}[X_1]$, this constant must be $\mathbb{E}[X_1]$. □

4 Markov chains

Il faut ajouter dans ce chapitre : une présentation du lemme de classe monotone ; une discussion de l'indépendance conditionnelle ; une définition d'une chaîne de Markov avec une filtration qui ne soit pas forcément sa filtration naturelle ; adapter la démonstration de la propriété faible de Markov à cette situation.

4.1 Introduction

Just as martingales, Markov chains are a class of stochastic processes whose definition involves conditional expectation in a crucial way. However, in contrast with martingales, which are a technical tool for the study of dynamical systems, Markov chains are a model for actual random phenomena. The class of Markov chains, or more generally Markov processes, has the main features of a good and successful model, namely: simplicity and versatility. It is mathematically simple enough to be studied in great detail, and allows one to give interesting models of a whole range of real situations.

Both points should however be nuanced. On one hand, the general theory of Markov processes is a very sophisticated theory, of which we are going to study a few fundamental ideas in the simplest interesting situation. On the other hand, Markovian models have their limit when applied to reality, and must often be enhanced to fit any particular actual random phenomenon.

The general idea of Markov processes is that of a random motion in a so-called *state space*, that we will denote by E. This random motion can either be though of as the random motion of a particle in the space E, or as the random evolution in time of the state of a system, each point of E describing such a state. The second point of view is more general than the first, since the position of a particle is a special case of the state of a system. In any case, the main defining property of the random motion, or random evolution, is that it is *memoryless*, in the sense that the way it moves, or evolves, immediately after a given instant of time depends on the history of the movement, or evolution, only through its present location, or state. Another way of stating this property is to say that at every instant of time, the future evolution is independent of the past evolution conditional on the present state.

Markov processes come in several variants, depending on the time being discrete or continuous, and on the space E being discrete or continuous. We are going to study the simplest case, where E is discrete and time is discrete. Although technically simple, this case will allow us to meet most of the fundamental ideas of the theory of Markov processes.

4.2 First definition and first properties

In this chapter, we will call *state space* and denote by E an arbitrary non-empty finite or countable set. Whenever it is necessary, this set will be endowed with the σ-field $\mathcal{P}(E)$ of all subsets of E.

A Markov chain is a sequence of random variables with values in E which satisfies a certain property which we described as absence of memory. In dealing with this property, we will make use of conditional expectations in a way that differs slightly from the way we discussed it in the first chapter of these notes.

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a sub-$\sigma$-field $\mathcal{B}$ of $\mathcal{A}$ and an event $A \in \mathcal{A}$, we will use the notation

$$\mathbb{P}(A|\mathcal{B}) = \mathbb{E}[\mathbb{1}_A|\mathcal{B}]$$

for the conditional probability of A given $\mathcal{B}$. Let us emphasize that this is a random variable, and not a number. In the simple case where $\mathcal{B}$ is the σ-field $\{\varnothing, B, B^c, \Omega\}$ generated by a single event B such that $0 < \mathbb{P}(B) < 1$, we have (see Section 2.1)

$$\mathbb{P}(A|\mathcal{B}) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \mathbb{1}_B + \frac{\mathbb{P}(A \cap B^c)}{\mathbb{P}(B^c)} \mathbb{1}_{B^c}$$

almost surely. The number $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ is traditionally denoted by $\mathbb{P}(A|B)$, and called the conditional probability of A given B, with the important precaution that it is defined only when $\mathbb{P}(B) > 0$.

A situation that we will meet often is summarised in the following exercise.

Exercise 4.1 *Let A be an event and Y a random variable with values in a finite or countable set D endowed with the σ-field $\mathcal{P}(D)$. Suppose that we want to understand the conditional probability $\mathbb{P}(A|\sigma(Y))$, which incidentally is often denoted by $\mathbb{P}(A|Y)$. What is usually easy to compute is, for an element $y \in D$, the conditional probability $\mathbb{P}(A|\{Y = y\})$, which is usually denoted by $\mathbb{P}(A|Y = y)$. However, this conditional probability is defined only if $\mathbb{P}(Y = y) > 0$. Prove that*

$$\mathbb{P}(A|Y) = \sum_{\substack{y \in D \\ \mathbb{P}(Y=y)>0}} \mathbb{P}(A|Y = y) \mathbb{1}_{\{Y=y\}}.$$

The point here is that we consider in the sum only those values y for which $\mathbb{P}(Y = y) > 0$.

Apart from the conditional expectation, the main ingredient in the definition of a Markov chain is the *transition kernel.*

Definition 4.1 *A* transition kernel *on E is a function*

$$\begin{aligned} P : E \times E &\longrightarrow [0, 1] \\ (x, y) &\longmapsto p(x, y) \end{aligned}$$

such that

$$\forall x \in \mathbb{E}, \sum_{y \in E} p(x, y) = 1.$$

The value of P at a couple (x, y) is usually denoted by $p(x, y)$, but also sometimes by $P(x, y)$.

Informally, conditional on the system currently being in the state x, the probably of this state becoming y at the next instant of time is $p(x, y)$.

Exercise 4.2 *Check that it is equivalent to consider a transition kernel on E or to consider a map $E \to \mathrm{Prob}(E)$ from E into the set of all probability measures on $(E, \mathcal{P}(E))$.*

A transition kernel can be though of as, and indeed, when E is finite, identified with a matrix whose rows and columns are indexed by elements of E. The conditions on this matrix are that it should have non-negative entries and that the sum of the entries in each row should be 1.

Just as matrices, transition kernels can be multiplied: if P and P' are transition kernels, then $PP' : E \to [0, +\infty]$ defined by

$$\forall x, y \in E,\ (PP')(x, y) = \sum_{z \in E} P(x, z)P'(z, y)$$

is a transition kernel.

Exercise 4.3 *Check this statement, and check that the product thus defined is associative on the set of transition kernels, in the sense that if P, P', P'' are transition kernels, then $P(P'P'') = (PP')P''$.*

The transition kernel $I : E \times E \to [0, 1]$ defined by

$$I(x, y) = \delta_{x,y} = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

is the unit element of the associative product on the set of all transition kernels. It is of course the identity matrix if E is finite, and the natural analogue of it when E is infinite.

Given a transition kernel P, we define for every non-negative integer n a new transition kernel P^n by setting

$$P^0 = I \text{ and for all } n \geq 1,\ P^n = \underbrace{P \ldots P}_{n \text{ times}}.$$

We can now give a first definition of a Markov chain.

Definition 4.2 (Markov chains, first definition) *Let E be a non-empty, finite or countable set endowed with the σ-field of all its subsets. Let P be a transition kernel on E. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $X = (X_n)_{n \geq 0}$ be a sequence of random variables defined on this probability space, with values in E.*

The sequence X is a Markov chain on E with transition kernel P *if the following condition holds:*

$$\forall n \geq 0, \forall y \in E,\ \mathbb{P}(X_{n+1} = y | X_0, \ldots, X_n) = p(X_n, y).$$

According to the result of Exercise 4.1, this condition can be written in a slightly longer but more elementary way as follows : for all $n \geq 0$, all $y \in E$ and all $x_0, \dots, x_n \in E$ such that $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$,

$$\mathbb{P}(X_{n+1} = y | X_0 = x_0, \dots, X_n = x_n) = p(x_n, y).$$

Let us immediately give an equivalent characterisation of Markov chains.

Proposition 4.3 *With the notation of Definition 4.2, X is a Markov chain on E with transition kernel P if and only if the following condition holds:*

$$\forall n \geq 0, \forall x_0, \dots, x_n \in E,\ \mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_0 = x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n). \tag{3}$$

Proof. Let us start by the 'only if' part, that is, the implication $\Rightarrow$. We prove (3) by induction on n. For $n = 0$, it reduces to $\mathbb{P}(X_0 = x_0) = \mathbb{P}(X_0 = x_0)$, which is true. Let us now assume that (3) has been proved up to rank $n-1$ for some $n \geq 1$ and let us consider $x_0, \dots, x_n$ in E. If $\mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) = 0$, then both sides of (3) vanish, the left-hand side because it is the probability of an event included in a negligible event, and the right-hand side because the product of all factors except the last, being equal by induction to $\mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1})$, is equal to 0.

If, on the other hand, $\mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) > 0$, then

$$\begin{aligned}\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) &= \mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \\ &\qquad \mathbb{P}(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \\ &= \mathbb{P}(X_0 = x_0) p(x_0, x_1) \dots p(x_{n-2}, x_{n-1}) p(x_{n-1}, x_n),\end{aligned}$$

as expected.

Let us now prove the 'if' part, that is, the implication $\Leftarrow$. According to the remark made after the definition of Markov chains, it suffices to consider $x_0, \dots, x_n, y$ in E such that $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$ and compute $\mathbb{P}(X_{n+1} = y | X_0 = x_0, \dots, X_n = x_n)$. Thanks to (3), this conditional expectation is equal to

$$\frac{\mathbb{P}(X_0 = x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n) p(x_n, y)}{\mathbb{P}(X_0 = x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n)} = p(x_n, y),$$

and this finishes the proof. □

It follows from this proposition that if X is a Markov chain on E with transition kernel P, then

$$\forall n \geq 0, \forall y \in E,\ \mathbb{P}(X_{n+1} = y | X_n) = p(X_n, y). \tag{4}$$

However, it is important to realise that this property is much weaker than the property which defines Markov chains. Exercise 4.5 below gives an example of a process that is not a Markov chain but satisfies (4).

Exercise 4.4 *Check the fact that* (4) *holds for a Markov chain.*

Exercise 4.5 *Consider the state space $E = \{a, b, c\}$. Define a random variable Z with values in E such that $\mathbb{P}(Z = b) = \mathbb{P}(Z = c) = \frac{1}{2}$. Define a sequence $X = (X_n)_{n\geq 0}$ of random variables with values in E such that*

$$\forall n \geq 0,\ X_n = \begin{cases} a & \text{if } n \text{ is even,} \\ Z & \text{if } n \text{ is odd.} \end{cases}$$

Prove that there exists a unique transition kernel P such that X satisfies the property (4)*, and prove that X is not a Markov chain with transition kernel P.*

Let us give two simple consequences of this characterisation of Markov chains.

Proposition 4.4 *Let X be a Markov chain on E with transition kernel P. For all $n \geq 0$ and all $y \in E$, one has*

$$\mathbb{P}(X_n = y|X_0) = P^n(X_0, y).$$

Proof. We need to prove that for every element x_0 of E such that $\mathbb{P}(X_0 = x_0) > 0$, the equality $\mathbb{P}(X_n = y|X_0 = x_0) = P^n(x_0, y)$ holds. But for such an x_0, we have

$$\begin{aligned}
\mathbb{P}(X_0 = x_0)\mathbb{P}(X_n = y|X_0 = x_0) &= \mathbb{P}(X_0 = x_0, X_n = y) \\
&= \sum_{x_1,\ldots,x_{n-1}\in E} \mathbb{P}(X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = y) \\
&= \mathbb{P}(X_0 = x_0) \sum_{x_1,\ldots,x_{n-1}\in E} p(x_0, x_1)\ldots p(x_{n-1}, y) \\
&= \mathbb{P}(X_0 = x_0)P^n(x_0, y),
\end{aligned}$$

as expected. □

Exercise 4.6 *Let X be a Markov chain on E with transition kernel P. Check that for every integer $\ell \geq 0$, the sequence $(X_{n\ell})_{n\geq 0}$ is a Markov chain on E with transition kernel P^ℓ.*

Proposition 4.5 *Let X be a Markov chain on E with transition kernel P. Let $N \geq 0$ be an integer. For every $n \geq 0$, set $Y_n = X_{N+n}$. Then $Y = (Y_n)_{n\geq 0}$ is a Markov chain on E with transition kernel P.*

Proof. Let us consider $n \geq 0$ and $y_0, \ldots, y_n \in E$. Firstly, we know from the previous proposition that

$$\begin{aligned}
\mathbb{P}(Y_0 = y_0) &= \sum_{x_0\in\mathbb{E}} \mathbb{P}(X_0 = x_0, Y_0 = y_0) \\
&= \sum_{x_0\in\mathbb{E}} \mathbb{P}(X_0 = x_0, X_N = y_0) \\
&= \sum_{x_0\in\mathbb{E}} \mathbb{P}(X_0 = x_0)P^N(x_0, y_0).
\end{aligned}$$

Secondly,

$$\begin{aligned}
\mathbb{P}(Y_0 = y_0, \ldots, Y_n = y_n) &= \mathbb{P}(X_N = y_0, \ldots, X_{N+n} = y_n) \\
&= \sum_{x_0, \ldots, x_{N-1} \in E} \mathbb{P}(X_0 = x_0, \ldots, X_{N-1} = x_{N-1}, X_N = y_0, \ldots, X_{N+n} = y_n) \\
&= \sum_{x_0, \ldots, x_{N-1} \in E} \mathbb{P}(X_0 = x_0) p(x_0, x_1) \ldots p(x_{N-1}, y_0) p(y_0, y_1) \ldots p(y_{n-1}, y_n) \\
&= \sum_{x_0 \in E} \mathbb{P}(X_0 = x_0) P^N(x_0, y_0) p(y_0, y_1) \ldots p(y_{n-1}, y_n) \\
&= \mathbb{P}(Y_0 = y_0) p(y_0, y_1) \ldots p(y_{n-1}, y_n),
\end{aligned}$$

from which we see that Y is a Markov chain with transition kernel P. □

Exercise 4.7 *Check that for a Markov chain X with transition kernel P, one has, for all integers $0 \leq n \leq m$ and all $y \in E$*

$$\mathbb{P}(X_m = y | X_n) = P^{m-n}(X_n, y).$$

Let us conclude this first section by giving a few examples of Markov chains.

- **Independent random variables.** Let $X = (X_n)_{n \geq 0}$ be a sequence of i.i.d. random variables with values in E, with common distribution μ. Then X is a Markov chain with transition kernel P given by

  $$\forall x, y \in E,\ P(x, y) = \mu(y),$$

  where $\mu(y)$ is a notation for $\mu(\{y\})$.

- **Random walks on $\mathbb{Z}^d$.** Let $d \geq 1$ be an integer. Let μ be a probability measure on $\mathbb{Z}^d$. Let $(\xi_i)_{i \geq 1}$ be an i.i.d. sequence of random variables with values in $\mathbb{Z}^d$ and common distribution μ. Let X_0 be a random variable with values in $\mathbb{Z}^d$ independent of $(\xi_i : i \geq 1)$. For every $n \geq 1$, set

  $$X_n = X_0 + \xi_1 + \ldots + \xi_n.$$

  Then $X = (X_n)_{n \geq 0}$ is a Markov chain on $\mathbb{Z}^d$ with transition kernel P given by

  $$\forall x, y \in E,\ P(x, y) = \mu(y - x).$$

  This Markov chain is called the *random walk on* $\mathbb{Z}^d$ with jump distribution μ.

  Let $(e_1, \ldots, e_d)$ denote the canonical basis of $\mathbb{Z}^d$. In the special case where

  $$\mu = \frac{1}{2d} \sum_{i=1}^{d} (\delta_{e_i} + \delta_{-e_i}),$$

  the random walk is called the *simple random walk* on $\mathbb{Z}^d$. The simple random walk is one of the most classical objects of the theory of probability.

- **Random walk on a graph.** Let A be a subset of the set $\mathcal{P}_2(E)$ of pairs of elements of E. We think of E as the set of vertices of a graph, and of each element $\{x, y\} \in A$ as an unoriented edge joining the vertices x and y.

  For each $x \in E$, we define the set N_x of neighbours of x in the graph (E, A) by

  $$N_x = \{y \in E : \{x, y\} \in A\}.$$

  We make the assumption that for all $x \in E$, the set N_x is non-empty and finite:

  $$\forall x \in E,\ 0 < |N_x| < \infty.$$

  The *random walk on the graph* (E, A) is the Markov chain with transition kernel P given by

  $$\forall x, y \in E,\ P(x, y) = \begin{cases} \frac{1}{|A_x|} & \text{if } y \in A_x \\ 0 & \text{otherwise.} \end{cases}$$

  We did not prove yet that such a Markov chain exists, and we will do it soon.

- **Branching processes.** Recall from Section 3.6 the definition of a branching process. With the notation used there, the sequence $(X_n)_{n\geq 0}$ is a Markov chain on $\mathbb{N}$ with transition kernel

  $$\forall n, m \in \mathbb{N},\ P(x, y) = \mu^{*n}(m),$$

  where μ is the reproduction law of the branching process, that is, the common distribution of all the variables $(Z_{n,k})_{n,k\geq 0}$, and for every integer $n \geq 0$, μ^{*n} is the n-th convolution power of μ, that is, the distribution of the sum of n independent random variables with distribution μ, for example $Z_{0,1} + \ldots + Z_{0,n}$. In particular, $\mu^{*0} = \delta_0$.

Exercise 4.8 *Check that a sequence of i.i.d. random variables, a random walk on $\mathbb{Z}^d$, a branching process, are indeed Markov chains with the claimed kernels.*

4.3 Construction of Markov chains

In this section, we prove that on any finite or countable state space, any transition kernel is the transition kernel of a Markov chain.

Proposition 4.6 *Let E be a finite or countable set. Let P be a transition kernel on E. Let x_0 be an element of E. There exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a sequence $X = (X_n)_{n\geq 0}$ of E-valued random variables defined on this probability space such that $X_0 = x_0$ a.s. and X is a Markov chain with transition kernel P.*

In order to construct such a Markov chain, we will need a source of randomness, which will be provided by the next lemma.

Proposition 4.7 *There exists a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *and a sequence* $U = (U_n)_{n\geq 0}$ *of i.i.d. random variables with common distribution equal to the uniform distribution on the interval* $[0,1]$*.*

It may be that you have always taken this existence result for granted: in this case, you should take a moment to think about the fact that it actually needs a proof.

Proof. Take $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda)$, where λ is the Lebesgue measure. For all $n \geq 1$, define a random variable B_n on this probability space by setting

$$\forall t \in [0,1],\ B_n(t) = \lfloor 2^n t \rfloor - 2\lfloor 2^{n-1} t \rfloor = \sum_{k=0}^{2^{n-1}-1} \mathbb{1}_{\left[\frac{2k+1}{2^n}, \frac{2k+2}{2^n}\right)}.$$

Then $(B_n)_{n\geq 1}$ is an i.i.d. sequence with Bernoulli distribution of parameter $\frac{1}{2}$:

$$\forall n \geq 1,\ \mathbb{P}(B_n = 0) = \mathbb{P}(B_n = 1) = \frac{1}{2}.$$

Now, for all $n \geq 1$, let p_n denote the n-th prime number, so that $p_1 = 2, p_2 = 3, p_3 = 5$ and so on. For all $n \geq 0$, define

$$U_n = \sum_{k=1}^{\infty} 2^{-k} B_{p_{n+1}^k}.$$

Then, by computing for instance its characteristic function, one checks that U_n is uniformly distributed on $[0,1]$ for all $n \geq 0$. Moreover, each variable U_n is built from a subset of the variables $(B_r)_{r\geq 0}$ that is disjoint from the subset used to build all the other U_m's. Thus, $(U_n)_{n\geq 0}$ is an independent sequence of random variables. □

We now turn to the proof of the proposition.

Proof of Proposition 4.6. Take $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda)$ as before and consider a sequence $(U_n)_{n\geq 0}$ of i.i.d. random variables with uniform distribution on $[0,1]$.

We claim that there exists a function $F : E \times [0,1] \to E$ with the property that

$$\forall x, y \in E,\ \lambda(\{t \in [0,1] : F(x,t) = y\}) = p(x,y).$$

In order to build such a function, we order E and write its elements as $E = \{y_1, y_2, \ldots\}$ in an arbitrary fashion. Now choose $x \in E$ and $t \in [0,1)$. There exists a unique integer $k \geq 1$ such that

$$\sum_{l=1}^{k-1} p(x, y_l) \leq t < \sum_{l=1}^{k} p(x, y_l)$$

and we define

$$F(x,t) = y_k.$$

With this definition,

$$\{t \in [0,1] : F(x,t) = y\} = \left(\sum_{l=1}^{k-1} p(x, y_l), \sum_{l=1}^{k} p(x, y_l) \right]$$

is indeed a subset of $[0,1]$ with Lebesgue measure $p(x, y_k)$.

Let us now define $X_0 = x_0$ and inductively, for all $n \geq 0$,

$$X_{n+1} = F(X_n, U_n).$$

Let k_0 be the integer such that $x_0 = y_{k_0}$. Then, for all $n \geq 1$ and all $k_1, \ldots, k_n \geq 1$,

$$\begin{aligned} \mathbb{P}(X_0 = y_{k_0}, X_1 = y_{k_1}, \ldots, X_{k_n} = y_n) &= \prod_{m=1}^{n} \lambda \left((\sum_{l=1}^{k_m - 1} p(y_{k_m - 1}, y_l), \sum_{l=1}^{k_m} p(y_{k_{m-1}}, y_l)] \right) \\ &= p(y_{k_0}, y_{k_1}) \ldots p(y_{k_{n-1}}, y_{k_n}) \end{aligned}$$

and X is a Markov chain issued from $y_{k_0} = x_0$ with transition kernel P. □

We would like now to take a slightly different point of view on Markov chains. So far, we thought of a Markov chain as a sequence of random variables with values in E, and we would like to become familiar with the idea that it is, or at least can be though of as a single random variable with values in the space $E^{\mathbb{N}}$ of sequences of elements of E. Informally, this amouts to realising that $X = (X_n)_{n \geq 0}$ is really a function of two variables, namely $(n, \omega) \mapsto X_n(\omega)$, and changing the priority between the two variables n and ω.

In order to make this idea more precise, we need to define a measurable space with underlying set $E^{\mathbb{N}}$. Concretely, we need to endow the set

$$E^{\mathbb{N}} = \{\omega = (\omega_i)_{i \geq 0} : \forall i \geq 0, \omega_n \in E\}$$

with a σ-field. For this, we start by listing natural functions on this set that we would want to be measurable.

Definition 4.8 *Let E be a finite or countable set. On the set $E^{\mathbb{N}}$, which in the present context is called the* canonical space, *we define for each $n \geq 0$ the function*

$$\begin{array}{rccc} \widehat{X}_n : & E^{\mathbb{N}} & \longrightarrow & E \\ & \omega = (\omega_i)_{i \geq 0} & \longmapsto & \widehat{X}_n(\omega) = \omega_n. \end{array}$$

The collection $\widehat{X} = (\widehat{X}_n)_{n \geq 0}$ of E-valued functions on $E^{\mathbb{N}}$ is called the canonical process.

We think of an element ω of $E^{\mathbb{N}}$ as the complete record of all the successive positions of our particle (or states of our system), from the origin to the end of times. Then, $\widehat{X}_n(\omega)$ is the position at time n of our particle.

In the next definition, recall that the set E is endowed with the σ-field $\mathcal{P}(E)$.

Definition 4.9 *For every $n \geq 0$, we define on $E^{\mathbb{N}}$ the σ-field*

$$\mathcal{C}_n = \sigma(\widehat{X}_0, \ldots, \widehat{X}_n).$$

The elements of the π-system

$$\bigcup_{n=0}^{\infty} \mathcal{C}_n$$

are called cylinders *on $E^{\mathbb{N}}$, and the σ-field*

$$\mathcal{C} = \sigma\left(\bigcup_{n=0}^{\infty} \mathcal{C}_n\right) = \sigma(\widehat{X}_n : n \geq 0)$$

is called the cylinder σ-field *on $E^{\mathbb{N}}$.*

In the present context, $\mathcal{C}$ turns out to be the natural σ-field on $E^{\mathbb{N}}$. This means, in English, and without being too precise, that the functions of a trajectory that we want to consider (and to call measurable) are those which are expressible in terms of finitely many successive positions of the trajectory, or which can be approximated in an appropriate sense by such functions.

It is a useful observation that for all $n \geq 0$, the σ-field $\mathcal{C}_n$ is generated by the partition

$$\begin{aligned}
\mathbb{E}^{\mathbb{N}} &= \bigsqcup_{x_0,\ldots,x_n \in E} \{\widehat{X}_0 = x_0, \ldots, \widehat{X}_n = x_n\} \\
&= \bigsqcup_{x_0,\ldots,x_n \in E} \{\omega \in E^{\mathbb{N}} : \omega_0 = x_0, \ldots, \omega_n = x_n\} \\
&= \bigsqcup_{x_0,\ldots,x_n \in E} \{x_0\} \times \ldots \times \{x_n\} \times E^{\mathbb{N} \setminus \{0,\ldots,n\}}.
\end{aligned}$$

Lemma 4.10 *A map f from a measurable space $(\Omega, \mathcal{A})$ into $(E^{\mathbb{N}}, \mathcal{C})$ is measurable if and only if for all $n \geq 0$, the map $\widehat{X}_n \circ f$ is measurable from $(\Omega, \mathcal{A})$ to $(E, \mathcal{P}(E))$.*

Proof. The 'only if' part of the statement is a consequence of the fact that a composition of measurable maps is measurable. Let us prove the 'if' part: let us assume that for all $n \geq 0$, the map $\widehat{X}_n \circ f$ is measurable. Let us define the class

$$\mathcal{I} = \{C \in \mathcal{C} : f^{-1}(X) \in \mathcal{A}\}$$

of subsets of $E^{\mathbb{N}}$. This is sometimes called the *image σ-field*[2] of $\mathcal{A}$ by f. It is indeed a σ-field (check it!) and it is the largest sub-σ-field of $\mathcal{C}$ that makes f measurable. Of course, we want to show that $\mathcal{I} = \mathcal{C}$. For this, it suffices to to show that $\mathcal{C}_n \subset \mathcal{I}$ for all $n \geq 0$. According to the observation made just before stating the present lemma, it is

[2]More precisely, $\mathcal{I}$ is the intersection of $\mathcal{C}$ and the image σ-field of $\mathcal{A}$ by f.

enough to prove that for all $n \geq 0$ and all $x_0, \ldots, x_n \in E$, the set $\{\widehat{X}_0 = x_0, \ldots, \widehat{X}_n = x_n\}$ belongs to $\mathcal{I}$. This is indeed the case, because

$$\begin{aligned} f^{-1}(\{\widehat{X}_0 = x_0, \ldots, \widehat{X}_n = x_n\}) &= f^{-1}\left(\bigcap_{i=0}^{n}\{\widehat{X}_i = x_i\}\right) \\ &= \bigcap_{i=0}^{n} f^{-1}(\{\widehat{X}_i = x_i\}) \\ &= \bigcap_{i=0}^{n}(\widehat{X}_i \circ f)^{-1}(\{x_i\}) \end{aligned}$$

belongs to $\mathcal{A}$ by assumption. □

We are now in possession of a filtered measurable space with an adapted process on it, namely $(E^{\mathbb{N}}, \mathcal{C}, (\mathcal{C}_n)_{n\geq 0}, \widehat{X} = (\widehat{X}_n)_{n\geq 0})$. The last lemma will allow us to connect this nice space with our more familiar notion of sequence of E-valued random variables.

Definition 4.11 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space on which we are given a sequence $X = (X_n)_{n\geq 0}$ of E-valued random variables. The* trajectory map *of this stochastic process is the map*

$$\begin{aligned} X : \Omega &\longrightarrow E^{\mathbb{N}} \\ \omega &\longmapsto (X_n(\omega))_{n\geq 0}. \end{aligned}$$

A technical remark is in order at this point. Each random variable X_n is really an equivalence class of functions, any two of which coincide outside a $\mathbb{P}$-negligible subset of Ω. Because we are considering a countable collection of random variables, the map X is indeed well defined $\mathbb{P}$-almost surely, in the sense that two distinct choices of a sequence of representatives of the sequence of random variables $(X_n)_{n\geq 0}$ yields two functions X that may be distinct, but coincide outside a $\mathbb{P}$-negligible subset of Ω. This is one respect in which the theory of continous time random processes is technically more demanding than its discrete time analogue.

It follows immediately from Lemma 4.10 and from the identity

$$\forall n \geq 0, X_n = \widehat{X}_n \circ X$$

that the trajectory map is measurable with respect to the σ-fields $\mathcal{A}$ and $\mathcal{C}$.

Exercise 4.9 *What is the trajectory map of the canonical process $(\widehat{X}_n)_{n\geq 0}$ defined on the canonical space $(E^{\mathbb{N}}, \mathcal{C})$?*

We can now give an upgraded and more canonical version of Proposition 4.6.

Theorem 4.12 *Let E be a finite or countable set. Let P be a transition kernel on E. Let x be an element of E. There exists on the measurable space $(E^{\mathbb{N}}, \mathcal{C})$ a unique probability measure $\widehat{\mathbb{P}}_x$ such that $\widehat{X}_0 = x$ $\widehat{\mathbb{P}}_x$-a.s. and under $\widehat{\mathbb{P}}_x$, the canonical process $\widehat{X} = (\widehat{X}_n)_{n\geq 0}$ is a Markov chain on E with transition kernel P.*

Proof. To prove the existence of $\mathbb{P}_x$, let us apply Proposition 4.6 to find a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a Markov chain $X = (X_n)_{n\geq 0}$ defined on this space, issued from x and with transition kernel P. Let us consider the trajectory map X of this process and define

$$\widehat{\mathbb{P}}_x = \mathbb{P} \circ X^{-1}.$$

In English, $\widehat{\mathbb{P}}_x$ is the law of the trajectory of the process $X = (X_n)_{n\geq 0}$. Let us check that $(\widehat{X}_n)_{n\geq 0}$ is under $\widehat{\mathbb{P}}_x$ a Markov chain on E issued from x with transition kernel P. For this, let us choose $n \geq 0$ and $x_0, \ldots, x_n$ in E. According to Proposition 4.3, we must prove that

$$\widehat{\mathbb{P}}_x(\widehat{X}_0 = x_0, \ldots, \widehat{X}_n = x_n) = \delta_{x,x_0} p(x_0, x_1) \ldots p(x_{n-1}, x_n).$$

The following computation is elementary in that it consists exclusively in unfolding definitions:

$$\begin{aligned}
\widehat{\mathbb{P}}_x(\widehat{X}_0 = x_0, \ldots, \widehat{X}_n = x_n) &= \mathbb{P}(X^{-1}(\{\widehat{X}_0 = x_0, \ldots, \widehat{X}_n = x_n\})) \\
&= \mathbb{P}\big(X^{-1}(\{\widehat{X}_0 = x_0\}) \cap \ldots \cap X^{-1}(\{\widehat{X}_n = x_n\})\big) \\
&= \mathbb{P}((\widehat{X}_0 \circ X)^{-1}(\{x_0\}) \cap \ldots \cap (\widehat{X}_n \circ X)^{-1}(\{x_n\})) \\
&= \mathbb{P}(X_0^{-1}(\{x_0\}) \cap \ldots \cap X_n^{-1}(\{x_n\})) \\
&= \mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) \\
&= \delta_{x,x_0} p(x_0, x_1) \ldots p(x_{n-1}, x_n),
\end{aligned}$$

as expected[3].

Let us turn to the uniqueness of $\widehat{\mathbb{P}}_x$. Suppose that $\widehat{\mathbb{Q}}_x$ is a probability measure on $(E^N, \mathcal{C})$ under which the canonical process is a Markov chain on E issued from x and with transition kernel P. The for all $n \geq 0$ and all $x_0, \ldots, x_n \in E$, we have

$$\widehat{\mathbb{P}}_x(\widehat{X}_0 = x_0, \ldots, \widehat{X}_n = x_n) = \widehat{\mathbb{Q}}_x(\widehat{X}_0 = x_0, \ldots, \widehat{X}_n = x_n).$$

According to the remark made just before the statement of Lemma 4.10, this implies that for all $n \geq 0$, the probability measures $\widehat{\mathbb{P}}_x$ and $\widehat{\mathbb{Q}}_x$ agree on $\mathcal{C}_n$. Hence, they agree on the class $\bigcup_{n\geq 0} \mathcal{C}_n$ of all cylinder sets. We conclude the proof by applying the monotone class theorem, as follows.

On any measurable space, the class of measurable sets on which two probability measures agree is a monotone class (also called a λ-system). Applying this general fact in our

[3]The last computation is a proof of the fact, which we could have used directly, that, because $\widehat{\mathbb{P}}_x = \mathbb{P} \circ X^{-1}$ and $(X_0, \ldots, X_n) = (\widehat{X}_0, \ldots, \widehat{X}_n) \circ X$, the random variable $(\widehat{X}_0, \ldots, \widehat{X}_n)$ has under $\widehat{\mathbb{P}}_x$ the same distribution as the random variable $(X_0, \ldots, X_n)$ under $\mathbb{P}$.

particular situation, we find that the class of all elements of $\mathcal{C}$ on which $\widehat{\mathbb{P}}_x$ and $\widehat{\mathbb{Q}}_x$ agree is a λ-system. We just proved that this λ-system contains $\bigcup_{n\geq 0}\mathcal{C}_n$, which is a π-system. The monotone class theorem states that a λ-system which contains a π-system also contains the σ-field generated by this π-system. Hence, the class of sets on which $\widehat{\mathbb{P}}_x$ and $\widehat{\mathbb{Q}}_x$ agree contains the σ-field generated by all cylinder sets, which, by definition, is $\mathcal{C}$. Hence, $\widehat{\mathbb{P}}_x$ and $\widehat{\mathbb{Q}}_x$ agree on $\mathcal{C}$: they are equal. □

We are now in position of giving a second, more sophisticated definition of a Markov chain.

Definition 4.13 (Markov chains, second definition) *Let E be a non-empty, finite or countable. Let P be a transition kernel on E. A Markov chain with transition kernel P on E is a quintuple $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, (\mathbb{P}_x)_{x\in E}, X = (X_n)_{n\geq 0})$ in which $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0})$ is a filtered measurable space on which $X = (X_n)_{n\geq 0}$ is an adapted process which, for all x, is under $\widehat{\mathbb{P}}_x$, and in the sense of Definition 4.2, a Markov chain on E issued from x and with transition kernel P.*

Our discussion of the canonical space shows that every transition kernel on E is the transition kernel of a Markov chain on E in this more sophisticated sense.

Given a Markov chain in the sense of the definition above, we define, for every probability measure μ on E, the measure $\mathbb{P}_\mu$ by

$$\mathbb{P}_\mu = \int_E \mathbb{P}_x \, d\mu(x) = \sum_{x\in E} \mu(\{x\})\mathbb{P}_x.$$

Under $\mathbb{P}_\mu$, the process X is a Markov chain with initial distribution μ and transition kernel P.

4.4 The Markov property

The essence of a Markov chain is the fact that it is memoryless, a property that can be rephrased by saying that it starts afresh at every instant. The Markov property expresses this absence of memory in a very effective way, and extends it to random times: we will prove that, in a certain sense, a Markov chain starts afresh at every stopping time.

In order to articulate the Markov property, we need to introduce one last piece of structure on the canonical space.

Definition 4.14 *The* shift operator *on the canonical space is the map*

$$\begin{array}{cccc} \theta : & E^{\mathbb{N}} & \longrightarrow & E^{\mathbb{N}} \\ & \omega = (\omega_i)_{i\geq 0} & \longmapsto & \theta(\omega) = (\omega_{i+1})_{i\geq 0}. \end{array}$$

We also define $\theta_0 = \mathrm{id}_{E^{\mathbb{N}}}$ and, for all $n \geq 2$,

$$\theta_n = \theta^n = \underbrace{\theta \circ \ldots \circ \theta}_{n \text{ times}}.$$

For all $n \geq 0$, we have simply $\theta_n(\omega) = (\omega_{n+i})_{i\geq 0}$.

Exercise 4.10 *Check that for all $n \geq 0$, the map θ_n is measurable with respect to the σ-field $\mathcal{C}$.*

Finally, let us introduce a classical piece of notation. Let us consider a Markov chain in the sophisticated sense of Definition 4.13. For all $x \in E$, we naturally denote by $\mathbb{E}_x$ the expectation with respect to $\mathbb{P}_x$. Let us extend this notation as follows. Assume that Z is an E-valued random variable on $(\Omega, \mathcal{A})$. Then, for all non-negative random variable Y on $(\Omega, \mathcal{A})$, we define $\mathbb{E}_Z[Y]$ by the formula

$$\mathbb{E}_Z[Y] = \sum_{x\in E} \mathbb{E}_x[Y]\mathbb{1}_{\{Z=x\}}.$$

This can also be written

$$\mathbb{E}_Z[Y] = h(Z), \text{ where } h(x) = \mathbb{E}_x[Y].$$

Let us emphasise that $\mathbb{E}_Z[Y]$ is not a number in general, but a random variable.

We can now state the Markov property. The following statement is called the *weak* Markov property because it involves 'only' deterministic times. Immediately after proving it, we shall strenghen it into the so-called strong Markov property, which covers the case of random times.

Theorem 4.15 (Weak Markov property) *Let P be a transition kernel on the state space E. Let $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, (\mathbb{P}_x)_{x\in E}, X = (X_n)_{n\geq 0})$ be a Markov chain on E with transition kernel P.*

Let $F : (E^{\mathbb{N}}, \mathcal{C}) \to (\mathbb{R}_+, \mathscr{B}_{\mathbb{R}_+})$ be a measurable non-negative function. For all integer $n \geq 0$ and all $x \in E$, we have the equality

$$\mathbb{E}_x[F(\theta_n(X))|\mathcal{F}_n] = \mathbb{E}_{X_n}[F(X)] \quad \mathbb{P}_x\text{-}a.s. \tag{5}$$

of non-negative random variables on $(\Omega, \mathcal{A}, \mathbb{P}_x)$.

Proof. The right-hand side of (5) is a function of X_n, so that it is $\mathcal{F}_n$-measurable. What remains to prove is that for every $A \in \mathcal{F}_n$, both sides of (5) have the same $\mathbb{P}_x$-integral over A. Since both sides of (5) depend linearly on F, and since any measurable positive function can be written as the pointwise limit of an increasing sequence of simple functions, it is enough to prove (5) when F is an indicator function, that is, $F = \mathbb{1}_C$ for some $C \in \mathcal{C}$. Thus, we need to prove that

$$\forall B \in \mathcal{F}_n, \forall C \in \mathcal{C}, \quad \mathbb{E}_x[\mathbb{1}_C(\theta_n(X))\mathbb{1}_A] = \mathbb{E}_x[\mathbb{E}_{X_n}[\mathbb{1}_C(X)]\mathbb{1}_A]. \tag{6}$$

Here we use a monotone class argument to reduce the problem further. The class of all sets $C \in \mathcal{C}$ such that the last equality holds is a λ-system. In order to prove that it contains $\mathcal{C}$, it suffices to prove that it contains the π-system $\bigcup_{m\geq 0} \mathcal{C}_m$, which generates the σ-field $\mathcal{C}$. Thus, it is enough to prove that the equality holds when $C \in \mathcal{C}_m$ for an arbitrary m. Finally, according to the remark made before Lemma 4.10, it suffices to prove that for all choices of $m \geq 0$, $x_0, \dots, x_m \in E$, $y_0, \dots, y_n \in E$, the equality holds for

$$A = \{X_0 = x_0, \dots, X_n = x_n\} \text{ and } C = \{\widehat{X}_0 = y_0, \dots, \widehat{X}_m = y_m\}.$$

In this case, we can compute both sides of (6). The left-hand side is equal to

$$\begin{aligned}
\mathbb{E}_x[\mathbb{1}_C(\theta_n(X))\mathbb{1}_A] &= \mathbb{P}_x(X_0 = x_0, \dots, X_n = x_n, X_n = y_0, \dots, X_{n+m} = y_m) \\
&= \delta_{x,x_0} p(x_0, x_1) \dots p(x_{n-1}, x_n) \delta_{x_n,y_0} p(y_0, y_1) \dots p(y_{m-1}, y_m).
\end{aligned}$$

The right-hand side is

$$\begin{aligned}
\mathbb{E}_x[\mathbb{E}_{X_n}[\mathbb{1}_C(X)]\mathbb{1}_A] &= \mathbb{E}_x[\mathbb{E}_{x_n}[\mathbb{1}_C(X)]\mathbb{1}_A] \\
&= \mathbb{E}_x[\mathbb{1}_A]\mathbb{E}_{x_n}[\mathbb{1}_C(X)] \\
&= \mathbb{P}_x(X_0 = x_0, \dots, X_n = x_n)\mathbb{P}_{x_n}(X_0 = y_0, \dots, X_m = y_m) \\
&= \delta_{x,x_0} p(x_0, x_1) \dots p(x_{n-1}, x_n) \delta_{x_n,y_0} p(y_0, y_1) \dots p(y_{m-1}, y_m),
\end{aligned}$$

and the equality is proved. $\square$

Let us extend the Markov property to random times. For this, we need to remember the definition of a stopping time (see Definition 3.15) and the definition of the σ-field of events prior to a stopping time (see Definition 3.34). We need also to define the shift by a stopping time.

For this, let us consider again a Markov chain in the sense of Definition 4.13. Let T be a stopping time on the measurable space $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0})$. We denote by $\theta_T(X)$ the map from $(\Omega, \mathcal{A})$ to $(E^{\mathbb{N}}, \mathcal{C})$ defined by

$$(\theta_T(X))(\omega) = \theta_{T(\omega)}(X(\omega)) = (X_{T+i}(\omega))_{i\geq 0}.$$

Finally, let us agree on the following convention: if $h : \mathbb{N} \to \mathbb{R}$ is a function and T is a stopping time, we use the notation

$$h(T)\mathbb{1}_{\{T<\infty\}},$$

although T might take the value ∞ and $h(\infty)$ is not defined, to indicate

$$h(T)\mathbb{1}_{\{T<\infty\}} = \sum_{n\geq 0} h(n)\mathbb{1}_{\{T=n\}}.$$

To be clear, this random variable vanishes on the event $\{T = \infty\}$. With all this preparation, we can state the strong Markov property.

Theorem 4.16 (Strong Markov property) *Let P be a transition kernel on the state space E. Let $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, (\mathbb{P}_x)_{x\in E}, X = (X_n)_{n\geq 0})$ be a Markov chain on E with transition kernel P.*

Let $F : (E^{\mathbb{N}}, \mathcal{C}) \to (\mathbb{R}_+, \mathscr{B}_{\mathbb{R}_+})$ be a measurable non-negative function. Let T be a stopping time on $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0})$. For all $x \in E$, we have the equality

$$\mathbb{E}_x[F(\theta_T(X))\mathbb{1}_{\{T<\infty\}}|\mathcal{F}_T] = \mathbb{E}_{X_T}[F(X)]\mathbb{1}_{\{T<\infty\}} \quad \mathbb{P}_x\text{-a.s.} \tag{7}$$

of non-negative random variables on $(\Omega, \mathcal{A}, \mathbb{P}_x)$.

Proof. Just as in the proof of the weak Markov property, the right-hand side of (7), which is a function of X_T, is $\mathcal{F}_T$-measurable, and it suffices to prove that for every $A \in \mathcal{F}_T$, both sides of (7) have the same $\mathbb{P}_x$-integral over A.

Let us choose $A \in \mathcal{F}_T$. For all $n \geq 0$, we have

$$\mathbb{E}_x[F(\theta_T(X))\mathbb{1}_{\{T=n\}}\mathbb{1}_A] = \mathbb{E}_x[F(\theta_T(X))\mathbb{1}_{\{T=n\}\cap A}].$$

Using the weak Markov property and the fact that $\{T = n\} \cap A$ belongs to $\mathcal{F}_n$, we find

$$\mathbb{E}_x[F(\theta_T(X))\mathbb{1}_{\{T=n\}}\mathbb{1}_A] = \mathbb{E}_x[\mathbb{E}_{X_n}[F(X)]\mathbb{1}_{\{T=n\}\cap A}] = \mathbb{E}_x[\mathbb{E}_{X_T}[F(X)]\mathbb{1}_{\{T=n\}}\mathbb{1}_A].$$

Summing this equality over n, we find

$$\mathbb{E}_x[F(\theta_T(X))\mathbb{1}_{\{T<\infty\}}\mathbb{1}_A] = \mathbb{E}_x[\mathbb{E}_{X_T}[F(X)]\mathbb{1}_{\{T<\infty\}}\mathbb{1}_A],$$

and the proof is finished. □

We will see over time that the theorem that we just proved expresses the Markov property in a very convenient and powerful way. For the moment, let us give a simple consequence.

Corollary 4.17 *We use the notation of Theorem 4.16. Let x, y be elements of E and let T be a stopping time such that $T < \infty$ and $X_T = y$ $\mathbb{P}_x$-a.s. Then under $\mathbb{P}_x$, $\theta_T(X)$ is independent of $\mathcal{F}_T$ and has the same distribution as X under $\mathbb{P}_y$.*

Proof. The statement is equivalent to saying that for all non-negative measurable function F on $E^{\mathbb{N}}$ and all $B \in \mathcal{F}_T$,

$$\mathbb{E}_x[F(\theta_T(X))\mathbb{1}_B] = \mathbb{P}_x(B)\mathbb{E}_y[F(X)].$$

Let us compute the expectation on the left-hand side by conditioning with respect to $\mathcal{F}_T$ and using the strong Markov property. We find

$$\mathbb{E}_x[F(\theta_T(X))\mathbb{1}_B] = \mathbb{E}_x[\mathbb{E}_{X_T}[F(X)]\mathbb{1}_B] = \mathbb{E}_x[\mathbb{E}_y[F(X)]\mathbb{1}_B] = \mathbb{P}_x(B)\mathbb{E}_y[F(X)],$$

as expected. □

4.5 Recurrent and transient states

Equipped with the powerful tools developed in the last section, we can embark on the study of the recurrence properties of the states of a Markov chain. From this point on, and until further notice, we fix a Markov chain $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, (\mathbb{P}_x)_{x\in E}, X = (X_n)_{n\geq 0})$ on the state space E with transition kernel P.

For every state $x \in E$, let un introduce a random variable N_x with values in $\mathbb{N} \cup \{\infty\}$ defined by

$$N_x = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = x\}}$$

and a stopping time

$$T_x = \inf\{n \geq 1 : X_n = x\}.$$

The random variable N_x is the total number of visits of the chain to the state x and the stopping time T_x is the first return time of the chain to x. Note that $T_x \geq 1$ by construction, and the usual convention $\inf \varnothing = \infty$ applies.

It is a simple but important observation that the following equality of events holds:

$$\{N_x \geq \mathbb{1}_{\{X_0 = x\}} + 1\} = \{T_x < \infty\}.$$

These two events can be described as 'the chain visits the state x at least once strictly after time to 0'.

We will prove that, starting from x, either the Markov chain visits x infinitely often almost surely, or it visits x a finite number of times which moreover has finite expectation.

Before that, let us introduce the canonical versions of N_x and T_x: let us define on the canonical space

$$\widehat{N}_x = \sum_{n=0}^{\infty} \mathbb{1}_{\{\widehat{X}_n = x\}} \text{ and } \widehat{T}_x = \inf\{n \geq 1 : \widehat{X}_n = x\}.$$

We have $N_x = \widehat{N}_x(X)$ and $T_x = \widehat{T}_x(X)$.

Proposition 4.18 *Let x be an element of E. Exactly one of the following two situations occurs.*

1. $\mathbb{P}_x(T_x < \infty) = 1$. In this case, $N_x = \infty$ $\mathbb{P}_x$-a.s. and one says that x is recurrent.

2. $\mathbb{P}_x(T_x < \infty) < 1$. In this case, $N_x < \infty$ $\mathbb{P}_x$-a.s. and one says that x is transient. *Moreover, in this case,*

$$\mathbb{E}_x[N_x] = \frac{1}{\mathbb{P}_x(T_x = \infty)}.$$

Proof. Let $k \geq 1$ be an integer and let us compute $\mathbb{P}_x(N_x \geq k+1)$.

$$\begin{aligned}
\mathbb{P}_x(N_x \geq k+1) &= \mathbb{P}_x(\widehat{N}_x(X) \geq k+1) \\
&= \mathbb{P}_x(T_x < \infty, \widehat{N}_x(\theta_{T_x}(X)) \geq k) \\
&= \mathbb{E}_x[\mathbb{1}_{\{\widehat{N}_x \geq k\}}(\theta_{T_x}(X))\mathbb{1}_{\{T_x<\infty\}}] \\
&= \mathbb{E}_x[\mathbb{E}_{T_x}[\mathbb{1}_{\{\widehat{N}_x \geq k\}}(X)]\mathbb{1}_{\{T_x<\infty\}}] \\
&= \mathbb{P}_x(N_x \geq k)\mathbb{P}_x(T_x < \infty).
\end{aligned}$$

Since $\mathbb{P}_x(N_x \geq 1) = 1$, we get by induction

$$\forall k \geq 1,\ \mathbb{P}_x(N_x \geq k) = \mathbb{P}_x(T_x < \infty)^{k-1}.$$

If $\mathbb{P}_x(T_x < \infty) = 1$, we deduce that $\mathbb{P}_x(N_x = \infty) = 1$. This is the recurrent case.

On the other hand, if $\mathbb{P}_x(T_x < \infty) < 1$, then

$$\begin{aligned}
\mathbb{E}_x[N_x] &= \sum_{k\geq 1} \mathbb{P}_x(N_x \geq k) \\
&= \sum_{k\geq 0} \mathbb{P}_x(T_x < \infty)^k \\
&= \frac{1}{1 - \mathbb{P}_x(T_x < \infty)},
\end{aligned}$$

so that $\mathbb{E}_x[N_x] < \infty$ and N_x is finite $\mathbb{P}_x$-almost surely. □

The dichotomy between recurrent and transient states suggests the definition of the following function.

Definition 4.19 *The* Green function *of the chain is the function $G : E \times E \to [0, \infty]$ defined by*

$$\forall x, y \in E,\ \ G(x, y) = \mathbb{E}_x[N_y].$$

Proposition 4.20 *1. For all $x, y \in E$, we have*

$$G(x, y) = \sum_{n=0}^{\infty} P^n(x, y).$$

2. The state $x \in E$ is recurrent if and only if $G(x, x) = \infty$.
3. If $x \neq y$, then

$$G(x, y) = \mathbb{P}_x(T_y < \infty)G(y, y).$$

In particular, $G(x, y) \leq G(y, y)$.
4. For all $x \in E$,

$$G(x, x) = 1 + \mathbb{P}_x(T_x < \infty)G(x, x).$$

Proof. 1. By definition of G and the monotonce convergence theorem,

$$G(x,y) = \sum_{n=0}^{\infty} \mathbb{E}_x[\mathbb{1}_{\{X_n=y\}}] = \sum_{n=0}^{\infty} \mathbb{P}_x[X_n = y] = \sum_{n=0}^{\infty} P^n(x,y).$$

2. By the previous proposition, x is recurrent if and only if $\mathbb{E}_x[N_x] < \infty$.
3. Under $\mathbb{P}_x$, we have almost surely $N_y = \widehat{N}_y(\theta_{T_y}(X))\mathbb{1}_{\{T_y<\infty\}}$, so that

$$G(x,y) = \mathbb{E}_x[N_y] = \mathbb{E}_x[\mathbb{E}_y[\widehat{N}_y(X)]\mathbb{1}_{\{T_y<\infty\}}] = \mathbb{P}_x(T_y < \infty)\mathbb{E}_y[N_y] = \mathbb{P}_x(T_y < \infty)G(y,y).$$

4. Under $\mathbb{P}_x$, we have almost surely $N_x = 1 + \widehat{N}_x(\theta_{T_x}(X))\mathbb{1}_{\{T_x<\infty\}}$, and the relation on the Green functions follows. □

The Green function allows us to define a binary relation on E: given two sates x and y, we say that x *leads to* y, and we write $x \to y$, if $G(x,y) > 0$. Thus, x leads to y if and only if starting from x, there is a positive probability[4] of visiting y. If $x \neq y$, then the following four assertions are equivalent:

$$x \to y \quad \Leftrightarrow \quad G(x,y) > 0 \quad \Leftrightarrow \quad \mathbb{P}_x(N_y \geq 1) > 0 \quad \Leftrightarrow \quad \mathbb{P}_x(T_y < \infty) > 0.$$

Lemma 4.21 *On E, the binary relation $\to$ is reflexive and transitive.*

Proof. For every $x \in E$, one has $G(x,x) \geq 1$, so that $x \to x$. This means that $\to$ is a reflexive relation.

To prove that it is transitive, consider three states x, y and z such that $x \to y$ and $y \to z$. If any two of these states are equal, then it is true that $x \to z$. On the other hand, if they are pairwise distinct, then

$$\begin{aligned} G(x,z) &\geq \mathbb{P}_x(N_z \geq 1) \\ &= \mathbb{P}_x(T_z < \infty) \\ &\geq \mathbb{P}_x(T_y < \infty, \widehat{T}_z(\theta_{T_y}(X)) < \infty) \\ &= \mathbb{P}_x(T_y < \infty)\mathbb{P}_y(T_z < \infty) \\ &= \mathbb{P}_x(N_y \geq 1)\mathbb{P}_y(N_z \geq 1) \\ &> 0 \end{aligned}$$

and $x \to z$. □

It is not true in general that the relation $\to$ is symmetric. We will now prove that its restriction to the subset of all recurrent states is symmetric.

[4]The notation may be misleading: $x \to y$ means that starting from x, a visit to y occurs with positive probability. It does not mean that starting from x, a visit to y is certain. This is the difference between $\mathbb{P}_x(N_y \geq 1) > 0$ and $\mathbb{P}_x(N_y \geq 1) = 1$.

Proposition 4.22 *Let x and y be two states. Assume that x is recurrent and x leads to y. Then y is recurrent, y leads to x, and*

$$\mathbb{P}_x(T_y < \infty) = \mathbb{P}_y(T_x < \infty) = 1.$$

Proof. If $y = x$, then there is nothing to prove. Let us assume that $y \neq x$. Then

$$\begin{aligned} 0 = \mathbb{P}_x(T_x = \infty) &\geq \mathbb{P}_x(T_y < \infty, \widehat{T}_x(\theta_{T_y}(X)) = \infty) \\ &= \mathbb{P}_x(T_y < \infty)\mathbb{P}_y(T_x = \infty). \end{aligned}$$

Since, by assumption, $\mathbb{P}_x(T_y < \infty) > 0$, we find $\mathbb{P}_y(T_x = \infty) = 0$, that is, $\mathbb{P}_y(T_x < \infty) = 1$. In particular, $y \to x$.

Since $G(x, y) > 0$ and $G(y, x) > 0$, there exists two integers $a, b \geq 1$ such that

$$P^a(x, y) > 0 \text{ and } P^b(y, x) > 0.$$

Hence,

$$\begin{aligned} G(y, y) &= \sum_{n \geq 0} P^n(y, y) \\ &\geq \sum_{n \geq a+b} P^n(y, y) \\ &\geq \sum_{c \geq 0} P^a(y, x)P^c(x, x)P^b(x, y) \\ &= P^a(x, y)G(x, x)P^b(y, x) \\ &= \infty \end{aligned}$$

and y is recurrent.

Since y is recurrent and $y \to x$, we proved already that $\mathbb{P}_x(T_y < \infty) = 1$. □

Let us write $x \sim y$ if $x \to y$ and $y \to x$. The relation $\sim$ is an equivalence relation on E. It follows from the last proposition that the relations $\to$ and $\sim$ coincide on the set of all recurrent states. Note also that this proposition implies that if x is recurrent and y is transient, then $G(x, y) = 0$.

Proposition 4.23 *Under the assumptions of the previous proposition, we have*

$$\mathbb{P}_x(N_y = \infty) = \mathbb{P}_y(N_x = \infty) = 1.$$

In particular,

$$G(x, y) = G(y, x) = \infty.$$

Proof. Indeed, y is recurrent and for all $k \geq 1$,

$$\mathbb{P}_x(N_y \geq k) = \mathbb{P}_x(T_y < \infty)\mathbb{P}_y(N_y \geq k) = 1.$$

Hence, $\mathbb{P}_x(N_y = \infty) = 1$. □

We can now state the following result which summarises our study.

Theorem 4.24 (Classification of states) *Let R be the subset of E consisting of all recurrent states. Let*

$$R = \bigsqcup_{i \in I} R_i$$

be the partition of R in equivalence classes for the relation $\sim$. Consider a state $x \in E$.

1. If x is recurrent, let $i \in I$ be such that $x \in R_i$. Then $\mathbb{P}_x$-a.s.,

$$N_y = \infty \text{ for all } y \in R_i \text{ and } N_z = 0 \text{ for all } z \in E \setminus R_i.$$

In English, starting from a recurrent state, the chain stays forever in the class of its initial state and visits infinitely often every state of this class.

2. If x is transient, define $T = \inf\{n \geq 0 : X_n \in R\}$. Then

$$\mathbb{P}_x\big(\{T = \infty \text{ and } N_y < \infty \text{ for all } y \in E\} \cup \{T < \infty \text{ and } \exists j \in I, \forall n \geq T, X_n \in R_j\}\big) = 1.$$

In English, starting from a transient state, either the chain never visits a recurrent state, in which case it visits every state a finite number of times, or it eventually visits a recurrent state, in which case it gets stuck forever in the class of the first recurrent state which it visits.

The equivalence classes of R under the relation $\sim$ are called the *recurrence classes* of the chain.

Proof. 1. Consider $y \in R_i$. Then, according to Proposition 4.23, we have $\mathbb{P}_x(N_y = \infty) = 1$. Consider now $z \in E \setminus R_i$. If z is transient, then $G(x, z) = 0$ by Proposition 4.22. If z is recurrent, then $G(x, z) = 0$ by definition of the equivalence class $\sim$.

2. Consider a transient state y. The third assertion of Proposition 4.20 asserts that $G(x, y) \leq G(y, y)$, so that $G(x, y) < \infty$. In particular, N_y is finite $\mathbb{P}_x$-almost surely.

On the event $\{T < \infty\}$, let J be the random element of I such that $X_T \in R_j$. Then

$$\mathbb{P}_x(T < \infty \text{ and } \forall n \geq T, X_n \in R_J) = \mathbb{E}_x[\mathbb{1}_{\{T<\infty\}} \mathbb{P}_{X_T}(\forall n \geq T, X_n \in R_J)].$$

By the first part of the theorem, and since $X_T \in R_J$ by definition of J, we have $\mathbb{P}_{X_T}(\forall n \geq T, X_n \in R_J) = 1$ on the event $T < \infty$. Thus,

$$\mathbb{P}_x(T < \infty \text{ and } \forall n \geq T, X_n \in R_J) = \mathbb{P}(T < \infty),$$

which is what we wanted to prove. □

Definition 4.25 *The Markov chain is said to be* irreducible *if $x \to y$ for all $x, y \in E$.*

Corollary 4.26 *Let us assume that the Markov chain is irreducible. Then we are in exactly one of the following two situations.*

- *All states are recurrent, there is only one recurrence class and*

$$\mathbb{P}_x(\forall y \in E, N_y = \infty) = 1.$$

- *All states are transient and*

$$\mathbb{P}_x(\forall y \in E, N_y < \infty) = 1.$$

If E is a finite set, the we are in the first situation.

Proof. If there is one recurrent state, then by Proposition 4.22, all states are recurrent and there is only one class. Moreover, Proposition 4.23 implies that every state is visited infinitely often $\mathbb{P}_x$-almost surely for every $x \in E$.

If there is no recurrent state, then all states are transient and, for all states x, y, we have $G(x, y) \leq G(y, y) < \infty$, so that N_y is finite $\mathbb{P}_x$-almost surely.

Finally, if E is finite, there exists at least one state that is visited infinitely often, and we must be in the first situation. □

In the first situation, one says that the Markov chain is *irreducible and recurrent*. Let us study an important example.

Theorem 4.27 (Random walks on $\mathbb{Z}$) *Let μ be a probability measure on $\mathbb{Z}$. Let P be the transition kernel of the random walk on $\mathbb{Z}$ with jump distribution μ:*

$$\forall x, y \in \mathbb{Z}, \quad p(x, y) = \mu(y - x).$$

Let ξ be a random variable with distribution μ. Let us assume that $\mathbb{E}[|\xi|] < \infty$.

1. If $\mathbb{E}[\xi] \neq 0$, then every state is transient.

2. If $\mathbb{E}[\xi] = 0$, then every state is recurrent. Moreover, the chain is irreducible if and only if the subgroup of $\mathbb{Z}$ generated by $\{x \in \mathbb{Z} : \mu(x) > 0\}$ is $\mathbb{Z}$ itself.

Proof. 1. For all $x \in \mathbb{Z}$, the strong law of large numbers implies that $|X_n|$ tends to $+\infty$ as n tends to infinity, $\mathbb{P}_x$-almost surely. Hence, N_x is finite $\mathbb{P}_x$-almost surely and x is transient.

2. There is an invariance by translation of the problem which implies that all states have the same nature. It is thus sufficient to prove that 0 is recurrent.

Let us consider an integer $p \geq 0$ and a positive real $\varepsilon > 0$, both to be specified later, and let us estimate

$$\sum_{|x| \leq \varepsilon p} G(0, x)$$

in two ways. Firstly, we want to say that this sum is not too big. For this, we argue that for every $x \in \mathbb{Z}$, we have $G(x, x) = G(0, 0)$ by invariance by translation, and

$$G(0, x) \leq G(x, x) = G(0, 0).$$

Thus,
$$\sum_{|x|\le\varepsilon p} G(0,x) \le (2\varepsilon p+1)G(0,0).$$
Secondly, we want to say that the same sum is not too small. For this, we use the weak law of large numbers, according to which $\frac{X_n}{n}$ converges to 0 in probability. This implies that there exists n_0, which depends on ε, such that for all $n \ge n_0$,
$$\mathbb{P}_0(|X_n| \le \varepsilon n) \ge \frac{1}{2}.$$
For all $p \ge n_0$, we have then
$$\begin{aligned}\sum_{|x|\le\varepsilon p} G(0,x) &= \sum_{|x|\le\varepsilon p}\sum_{n=0}^{\infty} P^n(0,x)\\ &\ge \sum_{|x|\le\varepsilon p}\sum_{n=n_0}^{p} P^n(0,x)\\ &\ge \sum_{n=n_0}^{p} \mathbb{P}_0(|X_n|\le\varepsilon p)\\ &\ge \frac{p-n_0+1}{2}.\end{aligned}$$
Thus, we proved that for all $\varepsilon > 0$, there exists an n_0 such that for all $p \ge n_0$,
$$\frac{p-n_0+1}{2} \le \sum_{|x|\le\varepsilon p} G(0,x) \le (2\varepsilon p+1)G(0,0).$$
Thus, for all $\varepsilon > 0$, we have
$$G(0,0) \ge \lim_{p\to\infty}\frac{p-n_0+1}{4\varepsilon p+2} = \frac{1}{4\varepsilon}.$$
This implies that $G(0,0) = \infty$, and 0 is recurrent.

There remains to study the irreducibility of the chain. Let us denote by S the set $\{x \in \mathbb{Z} : \mu(x) > 0\}$.

Let us first assume that the chain is irreducible. Then in particular 0 leads to 1, which means that there exists $n \ge 0$ such that $P^n(0,1) > 0$. This in turns means that there exists integers $x_1,\dots,x_{n-1}$ such that $p(0,x_1)p(x_1,x_2)\dots p(x_{n-1},1) > 0$. Thus, S contains the integers $x_1, x_2-x_1,\dots,x_{n-1}-x_{n-2}, 1-x_{n-1}$, which add up to 1. The subgroup of $\mathbb{Z}$ generated by S is thus $\mathbb{Z}$.

Conversely, let us assume that S generates $\mathbb{Z}$. Then, let us choose $z \in \mathbb{Z}$ and let us prove that $0 \to z$. Then, there exists $x_1,\dots,x_k$ and $y_1,\dots,y_l$ in S such that
$$z = x_1+\ldots+x_k-y_1-\ldots-y_l.$$

Let us write $x = x_1 + \ldots + x_k$ and $y = y_1 + \ldots + y_l$. Now, we have on one hand

$$P^k(0, x) \geq \mu(x_1) \ldots \mu(x_k) > 0,$$

so that $0 \to x$, and on the other hand

$$P^l(z, x) \geq p(z, z + y_1) \ldots p(z + y_1 + \ldots + y_{l-1}, x) = \mu(y_1) \ldots \mu(y_l) > 0,$$

so that $z \to x$. Since z is recurrent, it follows that $x \to z$, and finally $0 \to z$. □

Exercise 4.11 *With the notation of the theorem, give an example of a probability measure μ such that the random walk is transient, but such that the support of μ generates $\mathbb{Z}$ as an additive group.*

4.6 Invariant measures

As in the previous section, we fix once and for all a Markov chain

$$(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, (\mathbb{P}_x)_{x\in E}, X = (X_n)_{n\geq 0})$$

on the state space E with transition kernel P.

Definition 4.28 *A measure μ on E is an* invariant measure *of the transition kernel P if μ is not the zero measure, μ gives finite mass to every singleton, and*

$$\forall y \in E, \quad \sum_{x\in E} \mu(x)p(x, y) = \mu(y). \tag{8}$$

The conditions that μ is not zero and gives finite mass to every singleton can be written symbolically as

$$\exists x \in E, \mu(x) > 0 \text{ and } \forall x \in E, \mu(x) < \infty.$$

The relation (8) can be written matricially, as follows. Let us remember that P can be thought of as a possibly infinite square matrix with as many rows and columns as E has elements. Let us think of a measure on E as a row, the width (or length) of which is $|E|$, the number of elements of E. In a dual fashion, it would be appropriate to think of a function on E as a column with height EI. Then, if μ is a measure and f a function on E, the matrix product μf is a 1×1 matrix, that is, a number, which is none but the integral of f with respect to μ:

$$\mu f = \int_E f \, d\mu.$$

Exercise 4.12 *Let f be a function on E, seen as a column. The matrix product Pf is well defined and it is a column. Thus, it represents a function on E. What is this function?*

Equation (8) can be written

$$\mu P = \mu.$$

This shows that there is a purely linear algebraic approach to the study of invariant measures. Indeed, if E is finite, μ is an invariant measures if and only if μ^*, the column obtained by transposing μ, is an eigenvector with non-negative coefficients of the transposed kernel P^*, associated with the eigenvalue 1. The piece of linear algebra which deals with stochastic matrices is called the Perron-Frobenius theory, and we shall discuss it later.

Let us give an example of an invariant measure. Let us consider a random walk on $\mathbb{Z}^d$, with arbitrary jump distribution. Thus, η is a probability measure on $\mathbb{Z}^d$ and for all $x, y \in \mathbb{Z}^d$, we have $p(x, y) = \eta(y - x)$. Then the equality

$$\sum_{x \in \mathbb{Z}^d} p(x, y) = \sum_{x \in \mathbb{Z}^d} \eta(y - x) = \sum_{x \in \mathbb{Z}^d} \eta(x) = 1,$$

valid for all $y \in \mathbb{Z}^d$, shows that the counting measure μ, given by

$$\forall x \in \mathbb{Z}^d,\ \mu(x) = 1$$

is an invariant measure of this Markov chain.

It is an elementary observation that any positive multiple of an invariant measure is still an invariant measure. If there exists an invariant measure μ with the property that $\mu(E) < \infty$, it is natural to normalise it and to define

$$\pi = \frac{1}{\mu(E)},$$

which is an invariant probability measure on E. Then, for every bounded function f on E, the matricial relation

$$\mu P f = \mu f$$

can be written in terms of the Markov chain as

$$\mathbb{E}_\pi[f(X_1)] = \int_E f \, d\pi,$$

where $\mathbb{E}_\pi$ denotes the expectation with respect to the probability measure

$$\mathbb{P}_\pi = \sum_{x \in E} \pi(x) \mathbb{P}_x.$$

In other words, if the initial distribution of the Markov chain is an invariant probability measure, then the Markov chain is stationary, in the sense that the distribution of X_n does not depend on n.

There is another instructive way of thinking of (8). Suppose that we consider the evolution of a very large assembly of particles (or sand grains, or people). At the initial time, the quantity of particles (or sand, or people) that is present at each state x is measured by the number $\mu(x)$. For example, $\mu(x)$ is the number of litres of water present at the state x at time 0. Then, between time 0 and time 1, from every state x, and for every state y, a proportion $p(x, y)$ of the water present at x moves to y. Thus, on one hand, the water received by the state y in this process is

$$\sum_{x \in E} \mu(x)p(x, y).$$

On the other hand, the water lost by the same state y is

$$\sum_{x \in E} \mu(y)p(y, x),$$

which incidentally is equal to $\mu(y)$. To say that μ is invariant is to say that the water received by y exactly compensates the water lost by y: this is the equality expressed by (8).

One could demand that more is true, and demand that in the process just described, for all states x and y, the water received by y *from* x exactly compensates the water sent from y *to* x.

Definition 4.29 *The measure μ on E is said to be* reversible *with respect to P if μ is not the zero measure, μ gives finite mass to every singleton, and*

$$\forall x, y \in E,\ \mu(x)p(x, y) = \mu(y)p(y, x). \tag{9}$$

Our discussion should make it clear that the following statement holds.

Proposition 4.30 *A reversible measure is an invariant measure.*

Exercise 4.13 *Prove directly this proposition.*

Let us give an example of a reversible measure. Let us consider on $\mathbb{Z}$ the random walk with jump distribution η given by

$$\eta(1) = p \text{ and } \eta(-1) = q = 1 - p$$

for some $p \in (0, 1)$. This is called the p-biased random walk on $\mathbb{Z}$.

Then the measure μ given by

$$\mu(i) = \left(\frac{p}{q}\right)^i \tag{10}$$

is reversible. If $p \neq q$, that is, if $p \neq \frac{1}{2}$, this measure is not proportional to the counting measure on $\mathbb{Z}$, of which we know already that it is invariant.

Exercise 4.14 *Describe, for every $p \in (0, 1)$, the set of all invariant measures of the p-biased random walk on $\mathbb{Z}$.*

Exercise 4.15 *Recall the random walk on a graph described page 69. Prove that the formula*

$$\mu(x) = |A_x|$$

defines an invariant measure on E.

When a reversible measure exists, it is in general *much* easier to find by solving (9) than by solving (8). In other words, when looking for an invariant measure, one should always start by looking for a reversible measure.

Exercise 4.16 *Draw a small connected graph. Find an invariant probability measure for the random walk on this graph by solving (8) (for this to be tractable by hand, your graph should not have more than a few vertices, a dozen would typically be too much, unless your graph has a lot of symmetry and you find a way to use it). Compare with the effort needed to find a reversible measure for the same random walk.*

Another instance of the fact that reversible measures are nice to work with is given by the following exercise.

Exercise 4.17 *Under the assumption that the Markov chain is irreducible, prove that any two reversible measures are proportional.*

Is it true, under the same assumption of irreducibility, that any two invariant measures are proportional ?

We will soon prove that if the chain is irreducible *and recurrent*, then any two invariant measures are proportional, but this will require more than a two-line proof.

Let us leave reversible measures aside and consider general invariant measures again. The fundamental result is the following.

Theorem 4.31 *Let $x \in E$ be a recurrent state. The formula*

$$\mu(y) = \mathbb{E}_x \left[\sum_{i=0}^{T_x - 1} \mathbb{1}_{\{X_i = y\}} \right]$$

defines an invariant measure on E. Moreover, the support of this measure is

$$\{y \in E : \mu(y) > 0\} = \{y \in E : x \to y\} = \{y \in E : x \sim y\},$$

the communication class of x.

Proof. The first observation is that $\mu(x) = 1$ by definition of T_x. Hence, μ is not identically zero. Let us prove that μ satisfies (8). To start with, whether $y = x$ or $y \neq x$, we have

$$\mu(y) = \mathbb{E}_x \left[\sum_{i=1}^{T_x} \mathbb{1}_{\{X_i = y\}} \right].$$

Now, let us compute by considering not only the i-th state of the walk, but also the state immediately before.

$$\begin{aligned}
\mu(y) &= \mathbb{E}_x\left[\sum_{i=1}^{T_x}\sum_{z\in E}\mathbb{1}_{\{X_{i-1}=z\}}\mathbb{1}_{\{X_i=y\}}\right] \\
&= \sum_{i=1}^{\infty}\sum_{z\in E}\mathbb{E}_x\left[\mathbb{1}_{\{X_{i-1}=z\}}\mathbb{1}_{\{X_i=y\}}\mathbb{1}_{\{i\le T_x\}}\right].
\end{aligned}$$

Now, for all $i \ge 1$, $y, z \in E$, we have

$$\mathbb{E}_x\left[\mathbb{1}_{\{X_{i-1}=z\}}\mathbb{1}_{\{X_i=y\}}\mathbb{1}_{\{i\le T_x\}}\right] = \mathbb{E}_x\Big[\mathbb{E}_x\left[\mathbb{1}_{\{X_{i-1}=z\}}\mathbb{1}_{\{X_i=y\}}\mathbb{1}_{\{i\le T_x\}}|\mathcal{F}_{i-1}\right]\Big].$$

Using the fact that $\{i \le T_x\} = \{T_x \le i-1\}^c$ belongs to $\mathcal{F}_{i-1}$, we find

$$\begin{aligned}
\mathbb{E}_x\left[\mathbb{1}_{\{X_{i-1}=z\}}\mathbb{1}_{\{X_i=y\}}\mathbb{1}_{\{i\le T_x\}}\right] &= \mathbb{E}_x\Big[\mathbb{1}_{\{X_{i-1}=z\}}\mathbb{E}_x\left[\mathbb{1}_{\{X_i=y\}}|\mathcal{F}_{i-1}\right]\mathbb{1}_{\{i\le T_x\}}\Big] \\
&= \mathbb{E}_x\left[\mathbb{1}_{\{X_{i-1}=z\}}p(X_{i-1},y)\mathbb{1}_{\{i\le T_x\}}\right] \\
&= p(z,y)\mathbb{E}_x\left[\mathbb{1}_{\{X_{i-1}=z\}}\mathbb{1}_{\{i\le T_x\}}\right].
\end{aligned}$$

Thus,

$$\begin{aligned}
\mu(y) &= \sum_{z\in E}p(z,y)\sum_{i=1}^{\infty}\mathbb{E}_x\left[\mathbb{1}_{\{X_{i-1}=z\}}\mathbb{1}_{\{i\le T_x\}}\right] \\
&= \sum_{z\in E}p(z,y)\mathbb{E}_x\left[\sum_{i=1}^{T_x}\mathbb{1}_{\{X_{i-1}=z\}}\right] \\
&= \sum_{z\in E}p(z,y)\mathbb{E}_x\left[\sum_{i=0}^{T_x-1}\mathbb{1}_{\{X_{i-1}=z\}}\right] \\
&= \sum_{z\in E}\mu(z)p(z,y).
\end{aligned}$$

Note that, by construction, for all $y \in E$,

$$\mu(y) \le \mathbb{E}_x\left[\sum_{i=0}^{\infty}\mathbb{1}_{\{X_i=y\}}\right] = \mathbb{E}_x[N_y] = G(x,y).$$

In particular, if $x \not\to y$, then $\mu(y) = 0$.

Let now y be a state such that $x \to y$ and $y \to x$. Thus, there exist integers n and m such that $P^n(x,y) > 0$ and $P^m(y,x) > 0$. Since $\mu P^n = \mu P^m = \mu$, we have in particular

$$\mu(y) = \sum_{z\in E}\mu(z)P^n(z,y) \ge \mu(x)P^n(x,y) > 0,$$

so that $\mu(y) > 0$, and

$$1 = \mu(x) = \sum_{z \in E} \mu(z) P^m(z, x) \geq \mu(y) P^m(y, x),$$

so that $\mu(y) < \infty$.

This concludes the proof that μ is an invariant measure with support equal to the communication class of x. □

Exercise 4.18 *Does every Markov chain admit an invariant measure ?*

Exercise 4.19 *Prove that any invariant measure of an irreducible Markov chain gives a positive mass to every state. In other words, such a measure μ satisfies $\forall x \in E, \mu(x) > 0$.*

The second fundamental result is the following.

Theorem 4.32 *Assume that the Markov chain is irreducible and recurrent. Then any two invariant measures are proportional.*

Proof. Let μ be an invariant measure. Let $x \in E$ be such that $\mu(x) > 0$. Dividing μ by $\mu(x)$, we may and will assume that $\mu(x) = 1$. This normalisation being made, we want to prove that μ is equal to the measure ν defined by

$$\forall y \in E, \ \nu(y) = \mathbb{E}_x \left[\sum_{i=0}^{T_x - 1} \mathbb{1}_{\{X_i = y\}} \right].$$

For this, we will prove that $\mu \geq \nu$, in the sense that

$$\forall y \in E, \ \mu(y) \geq \nu(y). \tag{11}$$

Let us take this inequality for granted for one moment and explain how it implies $\mu = \nu$. The point is that, up to the fact that it may be identically zero, $\mu - \nu$ is an invariant measure of our Markov chain, which gives mass 0 to the state x, and which therefore must vanish identically[5]. More precisely, consider $y \in E$ and an integer m such that $P^m(y, x) > 0$. We have

$$0 = (\mu - \nu)(x) = \sum_{z \in E} (\mu - \nu)(z) P^m(z, x) \geq (\mu - \nu)(y) P^m(y, x),$$

from which it follows that $\mu(y) = \nu(y)$.

Let us now prove (11). For this, we prove that for all $n \geq 0$ and all $y \in E$,

$$\mu(y) \geq \mathbb{E}_x \left[\sum_{i=0}^{(T_x - 1) \wedge n} \mathbb{1}_{\{X_i = y\}} \right]. \tag{12}$$

[5] This may be a good time to search Exercise 4.19 if you did not yet do so.

If $y = x$, then both sides are equal to 1, for all $n \geq 0$, and the inequality holds.

Let us assume that $y \neq x$ and prove the result by induction on n. If $n = 0$, both sides of (12) are equal to 0. Let us assume that the result is proved at rank n. Then, proceding in a way that is very similar to the proof of Theorem 4.31, we have

$$\begin{aligned}
\mathbb{E}_x\left[\sum_{i=0}^{(T_x-1)\wedge(n+1)} \mathbb{1}_{\{X_i=y\}}\right] &= \mathbb{E}_x\left[\sum_{i=1}^{T_x\wedge(n+1)} \mathbb{1}_{\{X_i=y\}}\right] \\
&= \sum_{z\in E}\sum_{i=1}^{n+1} \mathbb{E}_x[\mathbb{1}_{\{X_{i-1}=z\}}\mathbb{1}_{\{X_i=y\}}\mathbb{1}_{\{i\leq T_x\}}] \\
&= \sum_{z\in E} \mathbb{E}_x\left[\sum_{i=1}^{T_x\wedge(n+1)} \mathbb{1}_{\{X_{i-1}=z\}}\right] p(z,y) \\
&= \sum_{z\in E} \mathbb{E}_x\left[\sum_{i=0}^{(T_x-1)\wedge n} \mathbb{1}_{\{X_{i-1}=z\}}\right] p(z,y) \\
&\leq \sum_{z\in E} \mu(z)p(z,y) \\
&= \mu(y).
\end{aligned}$$

Using the monotone convergence theorem to let n tend to infinity in (12), we obtain (11), and the proof is finished. □

Exercise 4.20 *Compare your understanding of the invariant measures of the p-biased random walk on $\mathbb{Z}$ with the previous theorem.*

We proved that an irreducible recurrent Markov chain admits, up to a multiplicative constant, a unique invariant measure. We will distinguish between the cases where these invariant measures are finite, and the case where they are infinite.

Proposition 4.33 *Assume that the chain is irreducible and recurrent. Then exactly one of the following two situations occurs.*
1. All invariant measures have infinite total mass and for all $x \in E$, we have $\mathbb{E}_x[T_x] = \infty$. In this case, the Markov chain is called null recurrent.
2. All invariant measures have finite total mass. There exists a unique invariant probability measure π. For all $x \in E$, we have

$$\pi(x) > 0,\ \mathbb{E}_x[T_x] < \infty \text{ and } \pi(x) = \frac{1}{\mathbb{E}_x[T_x]}.$$

In this case, the chain is called positive recurrent.

Proof. Let x be a state and let μ be the unique invariant measure on E such that $\mu(x) = 1$, which is given by Theorem 4.31. We have

$$\mu(E) = \sum_{y \in E} \mathbb{E}_x \left[\sum_{i=0}^{T_x - 1} \mathbb{1}_{\{X_i = y\}} \right] = \mathbb{E}_x \left[\sum_{i=0}^{T_x - 1} \sum_{y \in E} \mathbb{1}_{\{X_i = y\}} \right] = \mathbb{E}_x[T_x].$$

Since all invariant measures are proportional, they are either all finite, or all infinite. If they are all infinite, the previous computation shows that $\mathbb{E}_x[T_x] = \infty$ for all $x \in E$.

If they are all finite, the same computation shows that $\mathbb{E}_x[T_x]$ is finite for all $x \in E$. Moreover, the unique invariant probability measure is

$$\pi = \frac{1}{\mu(E)} \mu = \frac{1}{\mathbb{E}_x[T_x]} \mu,$$

from which it follows that

$$\pi(x) = \frac{1}{\mathbb{E}_x[T_x]},$$

and the proof is finished. □

Proposition 4.34 *Assume that the chain is irreducible. If there exists an invariant probability measure, then the chain is recurrent.*

Proof. Let y be a state such that $\pi(y) > 0$. Let us compute $G(y, y)$. Remember that for every state x, we have $G(x, y) \leq G(y, y)$. Thus,

$$G(y, y) = \sum_{x \in E} \pi(x) G(y, y) \geq \sum_{x \in E} \pi(x) G(x, y).$$

Now,

$$\sum_{x \in E} \pi(x) G(x, y) = \sum_{x \in E} \sum_{n=0}^{\infty} \pi(x) P^n(x, y) = \sum_{n=0}^{\infty} (\pi P^n)(y) = \sum_{n=0}^{\infty} \pi(y) = \infty.$$

Thus, $G(y, y) = \infty$ and y is recurrent. Thus, since the chain is irreducible, all states are recurrent, and the chain is positive recurrent. □

4.7 Brief summary

Let us summarise the results of the last two sections. Firstly, about recurrence, transience, and communication between states.

- The state space E of a Markov chain is partitioned into two disjoint subsets: the set of recurrent states and the set of transient states. Recall that x is recurrent, by definition, if and only if

$$\mathbb{P}_x(T_x < \infty) = 1.$$

- On E, there is the relation $\to$ defined by
$$x \to y \Leftrightarrow G(x,y) > 0.$$
It is reflexive and transitive. There is also the relation $\sim$ defined by
$$x \sim y \Leftrightarrow (x \to y \text{ and } y \to x).$$
It is an equivalence relation, the classes of which are called the communication classes. If E is itself a communication class, the chain is said to be irreducible.

- A communication class consists either exclusively of recurrent states, or exclusively of transient states. One speaks of recurrent and transient classes.

- Given two communication classes C and D, the existence of $x \in C$ and $y \in D$ such that $x \to y$ implies that for every $x \in C$ and every $y \in D$, one has $x \to y$. In this case, one writes $C \succ D$. This defines a binary relation on the set $E/\sim$ of communication classes.

- The relation $\succ$ on $E/\sim$ is a (partial) order[6] of which the recurrent classes are minimal elements. Let us emphasize that there can exist transient minimal classes. It is also possible that there be no minimal class at all.

Now about invariant measures.

- The support of any invariant measure is a union of communication classes. Moreover, if the support of an invariant measure contains a class C, then it also contains every class D such that $C \succ D$.

Let us now consider irreducible chains. There are three cases: transient, null recurrent and positive recurrent.

- Transient irreducible chains can have no invariant measure at all, a unique (up to multiplication) invariant measure, or several non-proportional invariant measures. In any case, any invariant measure of a transient irreducible chain has infinite total mass.

- Recurrent irreducible chains admit, up to multiplication, exactly one invariant measure.

- Null recurrent chains are the recurrent irreducible chains for which all invariant measures are infinite. The expected return time at every state is infinite[7].

- Positive recurrent chains are the recurrent irreducible chains for which all invariant measures are finite, or equivalently for which there exists an invariant probability measure. If π is this invariant probability measure, then the relation
$$\pi(x)\mathbb{E}_x[T_x] = 1$$
holds for every state x.

[6] As the notation suggests, we think of C being 'larger' than D if $C \succ D$.

[7] It is however not true that the hitting time of every state starting from every other is infinite.

4.8 The ergodic theorem

In this section and the next, we will study the asymptotic behaviour of our Markov chain when time tends to infinity.

The main question is to determine, for all x and y, the behaviour as n tends to infinity of $P^n(x,y) = \mathbb{P}_x(X_n = y)$.

If y is transient, then $G(x,y) \le G(y,y) < \infty$, so that

$$\lim_{n\to\infty} P^n(x,y) = 0.$$

With recurrent states, the situation is more interesting.

Theorem 4.35 *Consider a Markov chain that is irreducible and recurrent. Let μ be an invariant measure of this chain. Let $f : E \to \mathbb{R}^+$ and $g : E \to \mathbb{R}^+$ be two non-negative real-valued functions on E. Assume that $\int_E g\, d\mu > 0$ and that at least one of the functions f and g has finite integral with respect to μ. The for all $x \in E$,*

$$\frac{\sum_{i=0}^n f(X_i)}{\sum_{i=0}^n g(X_i)} \underset{n\to\infty}{\longrightarrow} \frac{\int_E f\, d\mu}{\int_E g\, d\mu} \quad \mathbb{P}_x - a.s.$$

Proof. Let $x \in E$ be a state. Let us introduce the sequence of stopping times

$$T_x^{(0)} = 0 \text{ and, for all } k \ge 1,\ T_x^{(k)} = \inf\{n > T_x^{(k-1)} : X_n = x\}.$$

Note that for all $k \ge 1$,

$$T_x^{(k)} = \widehat{T}_x(\theta_{T_x^{(k-1)}}(X)).$$

In particular, $T_x^{(1)} = T_x$ and the sequence $(T_x^{(k)})_{k\ge1}$ is a sequence of independent and identically distributed random variables.

Since x is recurrent, we have

$$\mathbb{P}_x(\forall k \ge 0, T_x^{(k)} < \infty) = 1.$$

For all $k \ge 1$, let us define

$$\xi_k = \sum_{n=T_x^{(k-1)}}^{T_x^{(k)}-1} f(X_n),$$

the sum of the values of f on the states visited by the chain during its k-th excursion from x.

The Markov property implies that the sequence $(\xi_k)_{k\ge1}$ of non-negative random variables is i.i.d. The strong law of large numbers asserts that

$$\frac{\xi_1 + \ldots + \xi_k}{k} \overset{\mathbb{P}_x-a.s.}{\underset{k\to\infty}{\longrightarrow}} \mathbb{E}[\xi_1].$$

On the other hand,

$$\begin{aligned}\mathbb{E}[\xi_1] &= \mathbb{E}\left[\sum_{n=0}^{T_x-1}\sum_{y\in E}\mathbb{1}_{\{X_n=y\}}f(y)\right]\\ &= \sum_{y\in E}\mathbb{E}\left[\sum_{n=0}^{T_x-1}\mathbb{1}_{\{X_n=y\}}\right]f(y)\\ &= \int_E f\,d\nu,\end{aligned}$$

where ν is the unique invariant measure on E such that $\nu(x)=1$. This measure is none other than $\frac{1}{\mu(x)}\mu$, so that

$$\mathbb{E}[\xi_1] = \frac{1}{\mu(x)}\int_E f\,d\nu.$$

For all $n\geq 0$, let us define

$$N_{x,n} = \sum_{i=0}^{n}\mathbb{1}_{\{X_i=y\}},$$

the number of visits to x before n, of which we think as the number of the excursion at x which is going on at time n. Since x is recurrent, we have

$$N_{x,n} \overset{\mathbb{P}_x-a.s.}{\underset{n\to\infty}{\longrightarrow}} +\infty.$$

By construction, we have, for all $n\geq 0$,

$$T_x^{(N_{x,n}-1)} \leq n < T_x^{(N_{x,n})} \quad \mathbb{P}_x - a.s.$$

Thus, for all $n\geq 0$, we have

$$\xi_1+\ldots+\xi_{N_{x,n}-1} \leq \sum_{i=0}^{n} f(X_i) \leq \xi_1+\ldots+\xi_{N_{x,n}}.$$

Dividing by $N_{x,n}$ and letting n tend to infinity, we find

$$\frac{1}{N_{x,n}}\sum_{i=0}^{n} f(X_i) \underset{n\to\infty}{\longrightarrow} \frac{1}{\mu(x)}\int_E f\,d\mu \quad \mathbb{P}_x - a.s.$$

The same argument can be applied to g, and by dividing the two a.s. convergences, one obtains the expected result. □

Corollary 4.36 *Suppose that the Markov chain is irreducible and recurrent.*
1. Assume that the chain is positive recurrent. Let π denote the unique invariant probability measure. Then for all probability measure ν on E and all $y\in E$, we have

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_{\{X_i=y\}} \underset{n\to\infty}{\longrightarrow} \pi(y) \quad \mathbb{P}_\nu - a.s.$$

In particular, for every non-negative function f on E,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \underset{n\to\infty}{\longrightarrow} \int_E f\, d\pi \quad \mathbb{P}_\nu - a.s.$$

2. Assume that the chain is null recurrent. Then for all probability measure ν on E and all $y \in E$, we have

$$\frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{\{X_i=y\}} \underset{n\to\infty}{\longrightarrow} 0. \quad \mathbb{P}_\nu - a.s.$$

Proof. It suffices to apply the ergodic theorem to the function $g = 1$. □

This corollary explains the names *positive recurrent* and *null recurrent*: in the positive recurrent case, the chain spends a positive proporition of the time in each state, whereas in the null recurrent case, the asymptotic proportion of time spent in any given state is 0.

This corollary shows also, in the positive recurrent case, that for all $x, y \in E$, the sequence $(P^n(x,y))_{n\geq 0}$ converges in the sense of Cesàro to $\pi(y)$. We will now study when this convergence holds in the usual sense.

Consider, as an example, the transition kernel

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

on the set $E = \{1, 2\}$. The invariant probability measure of this Markov chain is the uniform measure $\frac{1}{2}\delta_1 + \frac{1}{2}\delta_2$, but under $\mathbb{P}_1$ for instance, the distribution of the chain does not converge to the invariant distribution. Instead, it is alternatively equal to δ_1 and δ_2. This example shows that a phenomenon of cyclicity, or periodicity, can prevent the convergence of a Markov chain to its invariant measure. We will study this phenomenon and prove that it is the only possible obstruction to the convergence in distribution of a Markov chain.

We will need a small amount of arithmetic. Firstly, we say that a subset I of $\mathbb{N}$ is a *semigroup* if it contains 0 and is stable under addition:

$$\forall x, y \in I,\ x + y \in I.$$

We will need two properties of semigroups.

Lemma 4.37 *Let $I \subset \mathbb{N}$ be a semigroup. Let d be the g.c.d. of the elements of I.*
1. The subgroup of $\mathbb{Z}$ generated by I is the set

$$I - I = \{x - y : x, y \in \mathbb{Z}\}.$$

2. The equality $I - I = d\mathbb{Z}$ holds.
3. If $d = 1$, then there exists an integer n_0 such that I contains every integer larger than n_0.

Proof. 1. The subgroup of $\mathbb{Z}$ generated by I certainly contains $I - I$. Our claim is thus equivalent to the fact that $I - I$ is a subgroup of $\mathbb{Z}$. Since $I - I$ contains 0, it suffices to check that it is stable by subtraction. But for all $n, m, n', m' \in I$, we have

$$(n - m) - (n' - m') = \underbrace{(n + m')}_{\in I} - \underbrace{(n' + m)}_{\in I} \in I - I.$$

2. There exists an integer e such that $I - I = e\mathbb{Z}$. On one hand, since 0 belongs to I, we have $I \subset I - I \subset e\mathbb{Z}$, so that e is a common divisor of I. On the other hand, any common divisor of I is also a common divisor of $I - I$, hence of e. Thus, $e = d$.

3. Let us assume that $d = 1$. Then by the first assertion, I contains two consecutive integers, say i and $i + 1$. Hence, I, being stable under addition, contains the set

$$\{ai + b(i + 1) : a, b \geq 0\}.$$

We claim that this set contains all integers larger than i^2. Indeed, by reducing n modulo i^2 and then the remainder modulo i, write $n = qi^2 + ri + s$ with $q \geq 1$ and $r, s \in \{0, \dots, i-1\}$. Then $r - s > -i$ and the equality

$$n = (qi + (r - s))i + s(i + 1)$$

shows that n belongs to I. □

Definition 4.38 *Let $x \in E$ be a state. Define the set*

$$I_x = \{n \geq 0 : P^n(x, x) > 0\}.$$

The g.c.d. of this set is called the period *of x and it is denoted by d_x.*

Let us observe that I_x contains 0 and is stable by addition. Indeed, for all $n, m \geq 0$, we have

$$P^{n,m}(x, x) \geq P^n(x, x)P^m(x, x),$$

so that $n + m$ belongs to I_x as soon as n and m do. In particular, according to Lemma 4.37,

$$I_x - I_x = d_x\mathbb{Z}.$$

Proposition 4.39 *If the chain is irreducible, then all states have the same period.*

Proof. Consider two states x and y. By irreducibility, there exists integers k and l such that $P^k(x, y) > 0$ and $P^l(y, x) > 0$. It follows that

$$k + I_y + l \subset I_x.$$

Thus, for all $n, m \in I_y$, we have

$$n - m = (k + n + l) - (k + m + l) \in I_x - I_x = d_x\mathbb{Z}.$$

This shows that d_x is a common divisor of all the elements of I_y, so that d_x divides d_y. By symmetry, d_y divides d_x, and $d_x = d_y$. □

Definition 4.40 *An irreducible chain is said to be* aperiodic *if the common period of all states is* 1*.*

Proposition 4.41 *Assume that the Markov chain is irreducible and aperiodic. For all $x, y \in E$, there exists an integer n_0 such that for all $n \geq n_0$, $P^n(x,y) > 0$.*

Proof. Consider $x, y \in E$. Let $n_2 \geq 0$ be such that $P^{n_2}(x,y) > 0$. Such an integer n_2 exists because the chain is irreducible. Then, since $d_x = 1$, there exists n_1 such that I_x contains every integer larger than n_1. Set $n_0 = n_1 + n_2$. For all $n \geq n_0$, write $n = m + n_2$ with $m \geq n_1$. Then

$$P^n(x,y) \geq P^m(x,x)P^{n_2}(x,y) > 0,$$

and the result is proved. □

Theorem 4.42 *Assume that the chain is irreducible, positive recurrent, and aperiodic. Then, for all $x \in E$, we have*

$$\lim_{n\to\infty} \sum_{y\in E} |P^n(x,y) - \pi(y)| = 0,$$

where π is the unique invariant probability.

Proof. Let us define the Markovian kernel $\overline{P}$ on $E \times E$ by

$$\overline{p}((x_1,x_2),(y_1,y_2)) = p(x_1,y_1)p(x_2,y_2).$$

Let $(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_n)_{n\geq 0}, (\overline{\mathbb{P}}_{(x_1,x_2)})_{(x_1,x_2)\in E^2}, \overline{X} = (X_n^1, X_n^2)_{n\geq 0})$ be a Markov chain with transition kernel $\overline{P}$.

This chain $\overline{X}$ is irreducible. Indeed, consider (x_1,x_2) and (y_1,y_2) in $E\times E$. By aperiodicity, there exists n_1 and n_2 such that for all $n \geq n_1$ (resp. $n \geq n_2$), we have $P^n(x_1,y_1) > 0$ (resp. $P^n(x_2,y_2) > 0$). For $n = \max(n_1,n_2)$, we have $\overline{P}^n((x_1,x_2),(y_1,y_2)) > 0$.

Moreover, the probability measure $\pi \otimes \pi$ is invariant for this chain. Indeed, for all $(y_1,y_2) \in E^2$,

$$\begin{aligned}\sum_{(x_1,x_2)\in E\times E} (\pi\otimes\pi)(x_1,x_2)\overline{P}((x_1,x_2),(y_1,y_2)) &= \sum_{(x_1,x_2)\in E\times E} \pi(x_1)p(x_1,y_1)\pi(x_2)p(x_2,y_2) \\ &= \pi(y_1)\pi(y_2) \\ &= (\pi\otimes\pi)(y_1,y_2).\end{aligned}$$

It follows from the fact that it admits an invariant probability measure that the chain is positive recurrent.

For all $x, y \in E$,

$$P^n(x,y) - \pi(y) = \overline{\mathbb{P}}_{\pi\otimes\delta_x}(X_n^2 = y) - \overline{\mathbb{P}}_{\pi\otimes\delta_x}(X_n^1 = y) = \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbb{1}_{\{X_n^2=y\}} - \mathbb{1}_{\{X_n^1=y\}}].$$

Let us consider the stopping time

$$T = \inf\{n \geq 0 : X_n^1 = X_n^2\}.$$

We have

$$\begin{aligned} P^n(x,y) - \pi(y) &= \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbb{1}_{\{T>n\}}(\mathbb{1}_{\{X_n^2=y\}} - \mathbb{1}_{\{X_n^1=y\}})] \\ &\quad + \sum_{k=0}^{n}\sum_{z\in E}\overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbb{1}_{\{T=k,X_k^1=X_k^2=z\}}(\mathbb{1}_{\{X_n^2=y\}} - \mathbb{1}_{\{X_n^1=y\}})]. \end{aligned}$$

For $k \in \{0,\dots,r-1\}$ and $z \in E$, we have

$$\begin{aligned} \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbb{1}_{\{T=k,X_k^1=X_k^2=z\}}\mathbb{1}_{\{X_n^2=y\}}] &= \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbb{1}_{\{T=k,X_k^1=X_k^2=z\}}]P^{n-k}(z,y) \\ &= \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbb{1}_{\{T=k,X_k^1=X_k^2=z\}}\mathbb{1}_{\{X_n^1=y\}}], \end{aligned}$$

so that the double sum of the last expression vanishes. Thus,

$$\sum_{y\in E}|P^n(x,y) - \pi(y)| \leq 2\overline{\mathbb{P}}_{\pi\otimes\delta_x}(T > n).$$

Since the chain $\overline{X}$ is recurrent, T is finite $\overline{\mathbb{P}}_{\pi\otimes\delta_x}$-almost surely, and the last quantity tends to 0 as n tends to infinity. □

Tsinghua University 2017–2018

Advanced Probability

Bbrownian Motion

Shi Zhan

Contents

Chapter 4

Construction of Brownian motion

Brownian motion lies in the intersection of several important families of random processes (martingales, Markov processes, Gaussian processes), and is the fundamental example in each theory. This chapter gives a brief introduction to Brownian motion, providing an account of basic properties.

1. Some historical dates

The expression "Brownian motion" originates from the highly irregular movement of pollen grains on the surface of water, observed by the Scottish botanist Robert Brown in 1828. Bachelier (1900) and Einstein (1905) studied quantitatively this irregular movement in finance and in physics, respectively. Wiener, in 1923, established the mathematical model of Brownian motion, which we are going to study in this chapter, whereas many deep properties were discovered by Paul Lévy (1939, 1948). In 1973, Black and Scholes used Brownian motion to model options prices in financial mathematics in terms of the formula bearing now their name, which led to the Nobel Prize of Economics awarded to Scholes in 1997.

2. Before starting: Gaussian random vectors

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let ξ be a Gaussian $\mathscr{N}(0, 1)$ random variable. Its complex moment generating function is given by

$$\mathbb{E}[\mathrm{e}^{z\xi}] = \mathrm{e}^{z^2/2}, \qquad z \in \mathbb{C}.$$

In particular, the characteristic function of ξ is

$$\mathbb{E}[\mathrm{e}^{it\xi}] = \mathrm{e}^{-t^2/2}, \qquad t \in \mathbb{R}.$$

Theorem 2.1. (Gaussian tail distribution). *If ξ is a Gaussian $\mathscr{N}(0,\,1)$ random variable, then for any $x>0$,*

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\Big(\frac{1}{x}-\frac{1}{x^3}\Big)\mathrm{e}^{-x^2/2} &\le \mathbb{P}(\xi>x)\le \frac{1}{\sqrt{2\pi}}\frac{1}{x}\mathrm{e}^{-x^2/2},\\ \mathbb{P}(\xi>x) &\le \mathrm{e}^{-x^2/2}.\end{aligned}$$

Remark 2.2. (i) We have $\mathbb{P}(\xi>x)\sim\frac{1}{\sqrt{2\pi}}\frac{1}{x}\mathrm{e}^{-x^2/2}$, $x\to\infty$.

(ii) The upper bound $\mathrm{e}^{-x^2/2}$ is less precise but simpler than $\frac{1}{(2\pi)^{1/2}}\frac{1}{x}\mathrm{e}^{-x^2/2}$. It is useful in situations where we do not need much precision. □

Proof of Theorem 2.1. Is an exercise. □

Let $\mu\in\mathbb{R}$ and $\sigma>0$. We say that η is a Gaussian $\mathscr{N}(\mu,\,\sigma^2)$ if it has density

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\Big(-\frac{(y-\mu)^2}{2\sigma^2}\Big),\qquad y\in\mathbb{R}.$$

Clearly, η is Gaussian $\mathscr{N}(\mu,\,\sigma^2)$ if and only if $\eta=\sigma\xi+\mu$, where ξ is $\mathscr{N}(0,\,1)$.

Remark 2.3. It is convenient from now on to consider the Dirac measure δ_μ at $\mu\in\mathbb{R}$ as a (degenerate) Gaussian distribution. □

Proposition 2.4. (Convergence of sequences of Gaussian random variables). *Let (ξ_n) be a sequence of random variables such that for any n, ξ_n is $\mathscr{N}(\mu_n,\sigma_n^2)$.*

(i) If $\xi_n\to\xi$ in distribution, then ξ is $\mathscr{N}(\mu,\sigma^2)$, with $\mu:=\lim_{n\to\infty}\mu_n$ and $\sigma:=\lim_{n\to\infty}\sigma_n$.

(ii) If $\xi_n\to\xi$ in probability, then it also converges in L^p, for any $p\in[1,\infty)$.

Proof. Is an exercise. □

Definition 2.5. (Gaussian random vectors). *We say that $(\xi_1,\cdots,\xi_n)$ is a Gaussian random vector if any linear combination of its components is Gaussian.*

Remark 2.6. If $(\xi_1,\cdots,\xi_n)$ is a Gaussian random vector, then each of its component is a Gaussian random variable. Attention: the converse is wrong! □

Theorem 2.7. *(i) Let $\xi:=(\xi_1,\cdots,\xi_n)$ be a Gaussian random vector. Then $\xi_1,\cdots,\xi_n$ are independent if and only if the covariance matrix of ξ diagonal.*

(ii) Let $(\xi_1,\cdots,\xi_n,\eta_1,\cdots,\eta_m,\theta_1,\cdots,\theta_\ell)$ be a Gaussian vector. Then $(\xi_1,\cdots,\xi_n)$ and $(\eta_1,\cdots,\eta_m)$ are independent if and only if $\mathrm{Cov}(\xi_i,\,\eta_j)=0$, *$\forall i\le n$, $j\le m$.*

3. Construction of Brownian motion

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

Definition 3.1. *A family of real-valued random variables $X := (X_t,\, t \in \mathbf{T})$ is a* **Gaussian random process** *(or: Gaussian process, or: Gaussian stochastic process) if for any $n \geq 1$ and any $(t_1, \cdots, t_n) \in \mathbf{T}^n$, $(X_{t_1}, \cdots, X_{t_n})$ is a Gaussian random vector. We say that X is centered if $\mathbb{E}(X_t) = 0$, $t \in \mathbf{T}$.*

Definition 3.2. *A real-valued process $B = (B_t,\, t \geq 0)$ is said to be* **Brownian motion** *if it satisfies the following conditions:*

(i) For any n and any $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$, $B_{t_n} - B_{t_{n-1}}$, $\cdots$, $B_{t_2} - B_{t_1}$, B_{t_1} are independent.

(ii) For any $t \geq s \geq 0$, $B_t - B_s$ is Gaussian $\mathscr{N}(0, t-s)$.

Remark 3.3. One says that Brownian motion is of independent and stationary increments, or simply a Lévy process. □

Proposition 3.4. *A process $X = (X_t, t \geq 0)$ is Brownian motion with $X_0 = 0$* a.s. *if and only if it is centered Gaussian with covariance*

$$\mathbb{E}(X_s X_t) = \min\{s, t\} =: s \wedge t, \qquad s \geq 0, \ \ t \geq 0.$$

Proof. "$\Rightarrow$" Let $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$. By assumption, $X_{t_n} - X_{t_{n-1}}$, $\cdots$, $X_{t_2} - X_{t_1}$, X_{t_1} are independent Gaussian; hence $(X_{t_1}, X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}})$ is a Gaussian random vector, and so is $(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$. This implies that X is a Gaussian process, which is obviously centered.

Let us check the covariance of X. Let $t \geq s \geq 0$. We have $\mathbb{E}(X_s X_t) = \mathbb{E}(X_s(X_t - X_s)) + \mathbb{E}(X_s^2)$. Since X_s and $X_t - X_s$ are independent, we have $\mathbb{E}(X_s(X_t - X_s)) = 0$, whereas by (ii), $\mathbb{E}(X_s^2) = s$. Hence $\mathbb{E}(X_s X_t) = s$.

"$\Leftarrow$" Let $t \geq s \geq 0$. By assumption, $X_t - X_s$ is centered Gaussian, with variance $\mathbb{E}(X_t - X_s)^2 = \mathbb{E}(X_t^2) + \mathbb{E}(X_s^2) - 2\mathbb{E}(X_s X_t) = t + s - 2s = t - s$: we have property (ii).

Let $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$. We know that $(X_{t_n} - X_{t_{n-1}}, \cdots, X_{t_2} - X_{t_1}, X_{t_1})$ is a Gaussian random vector, whose covariance matrix is diagonal: indeed, for $j > i$, $\mathbb{E}[(X_{t_j} - X_{t_{j-1}})(X_{t_i} - X_{t_{i-1}})] = \mathbb{E}(X_{t_j} X_{t_i}) - \mathbb{E}(X_{t_{j-1}} X_{t_i}) - \mathbb{E}(X_{t_j} X_{t_{i-1}}) + \mathbb{E}(X_{t_{j-1}} X_{t_{i-1}}) = t_i - t_i - t_{i-1} + t_{i-1} = 0$. By Theorem 2.7(i), the components of this Gaussian random vector are independent. □

Remark 3.5. (i) By definition, it is easily using the π-λ theorem, that if B is Brownian motion, then $(B_t - B_0,\, t \geq 0)$ is Brownian motion starting at the origin, independent of B_0.

Unless stated otherwise, we will always assume $B_0 = 0$ a.s. We say that B is standard Brownian motion because it is sometimes of interest to study the centered Gaussian process with covariance $\sigma^2(s \wedge t)$, with $\sigma^2 > 0$.

(ii) For any real number $T > 0$, we call Brownian motion on $[0,\, T]$, any centered Gaussian process with covariance $s \wedge t$, for $(s,\, t) \in [0,\, T]^2$. $\square$

Theorem 3.6. (Wiener, 1923). *Brownian motion exists.*

Proof. (Lévy 1948). We start by constructing a Brownian motion defined on $[0, 1]$.

Let $(\xi_{k,n},\, 0 \leq k \leq 2^n,\, n \geq 0)$ be a family of i.i.d. Gaussian $\mathscr{N}(0,\, 1)$ random variables. Let $(X_n(t),\, t \in [0, 1],\, n \geq 0)$ be a family of random variables such that: (it is convenient de make a graph here)

- for any $n \geq 0$, $t \mapsto X_n(t)$ is affine on each interval of type $[\frac{k}{2^n},\, \frac{k+1}{2^n}]$;
- $X_0(0) = 0$, $X_0(1) = \xi_{0,0}$;
- $X_n(\frac{2j}{2^n}) = X_{n-1}(\frac{2j}{2^n})$, $X_n(\frac{2j+1}{2^n}) = X_{n-1}(\frac{2j+1}{2^n}) + \frac{\xi_{2j+1,n}}{2^{(n+1)/2}}$.

It is easy to check (left as an exercise) that for any $n \geq 0$, $(X_n(\frac{k}{2^n}),\, 0 \leq k \leq 2^n)$ is a centered Gaussian random vector with $\mathbb{E}[X_n(\frac{k}{2^n})X_n(\frac{\ell}{2^n})] = \frac{k}{2^n} \wedge \frac{\ell}{2^n}$.

Let $n \geq 0$. We see that $(X_n(t),\, t \in [0, 1])$ is a centered Gaussian process because $\sum_{i=1}^m a_i X_n(t_i)$ is a linear combination of $(X_n(\frac{k}{2^n}),\, k = 0, 1, \cdots, 2^n)$.

Consider the event $A_n := \{\sup_{t\in[0,1]} |X_n(t) - X_{n-1}(t)| > 2^{-n/4}\}$. We have

$$\begin{aligned}
\mathbb{P}(A_n) &= \mathbb{P}\Big(\bigcup_{j=0}^{2^{n-1}-1} \Big\{ \sup_{t\in[\frac{2j}{2^n}, \frac{2j+2}{2^n}]} |X_n(t) - X_{n-1}(t)| > 2^{-n/4} \Big\}\Big) \\
&= \mathbb{P}\Big(\bigcup_{j=0}^{2^{n-1}-1} \Big\{ \frac{|\xi_{2j+1,n}|}{2^{(n+1)/2}} > 2^{-n/4} \Big\}\Big) \leq \sum_{j=0}^{2^{n-1}-1} \mathbb{P}\Big(|\xi_{2j+1,n}| > 2^{(n+2)/4} \Big).
\end{aligned}$$

By symmetry and Theorem 2.1, $\mathbb{P}(|\xi_{2j+1,n}| > 2^{(n+2)/4}) \leq 2\exp(-2^{n/2})$; so we have $\mathbb{P}(A_n) \leq 2^n \exp(-2^{n/2})$, which implies that $\sum_n \mathbb{P}(A_n) < \infty$. By the Borel–Cantelli lemma, there exists an $E \in \mathscr{F}$ with $\mathbb{P}(E) = 1$, such that for all $\omega \in E$, $\sup_{t\in[0,1]} |X_n(t,\omega) - X_{n-1}(t,\omega)| \leq 2^{-n/4}$, $\forall\, n \geq n_0(\omega)$. [For $\omega \notin E$, we can take, for example, $X(t,\, \omega) := 0$, $\forall t \in [0,\, 1]$.] In particular, $X_n(\bullet, \omega)$ converges uniformly on $[0,\, 1]$, to a continuous limit denoted by $X(\bullet, \omega)$. By Proposition 2.4, $X = (X(t),\, t \in [0, 1])$ is a centered Gaussian process.

Let us check the covariance matrix of X. Let $0 \le s \le t \le 1$ and $n \ge 0$. There exists a pair $(k,\, \ell)$ with $k \le \ell$ such that $s \in [\frac{k}{2^n},\, \frac{k+1}{2^n}]$ and $t \in [\frac{\ell}{2^n},\, \frac{\ell+1}{2^n}]$. Since X_n is affine on $[\frac{k}{2^n},\, \frac{k+1}{2^n}]$, we have $X_n(s) = \alpha X_n(\frac{k}{2^n}) + (1-\alpha) X_n(\frac{k+1}{2^n})$, where $\alpha := k+1-2^n s \in [0,1]$. Similarly, $X_n(t) = \beta X_n(\frac{\ell}{2^n}) + (1-\beta) X_n(\frac{\ell+1}{2^n})$, with $\beta := \ell + 1 - 2^n t \in [0,1]$. It follows that

$$\mathbb{E}[X_n(s)X_n(t)] = \frac{\alpha\beta k}{2^n} + \frac{(1-\alpha)\beta((k+1)\wedge \ell)}{2^n} + \frac{\alpha(1-\beta)k}{2^n} + \frac{(1-\alpha)(1-\beta)(k+1)}{2^n}.$$

The expression on the right-hand side is, for $n \to \infty$,

$$\frac{\alpha\beta k}{2^n} + \frac{(1-\alpha)\beta k}{2^n} + \frac{\alpha(1-\beta)k}{2^n} + \frac{(1-\alpha)(1-\beta)k}{2^n} + \mathcal{O}(\frac{1}{2^n}),$$

which is $\frac{k}{2^n} + \mathcal{O}(\frac{1}{2^n}) = s + \mathcal{O}(\frac{1}{2^n})$. Letting $n \to \infty$, and by Proposition 2.4 again, we obtain $\mathbb{E}[X(s)X(t)] = s = s \wedge t$. Consequently, $(X(t),\, t \in [0,1])$ is a Brownian motion defined on $[0,1]$.

To conclude, if $(B^m_t,\, t \in [0,1])$, $m \ge 0$, is a sequence of independent Brownian motions on $[0,1]$, then

$$B_t := B^{\lfloor t \rfloor}_{t - \lfloor t \rfloor} + \sum_{0 \le m < \lfloor t \rfloor} B^m_1,$$

is Brownian motion on $[0, \infty)$. □

4. Regularisation of sample paths

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a complete probability space, i.e., $\mathscr{F}$ contains all $\mathbb{P}$-negligible sets.

Let $B = (B_t,\, t \ge 0)$ be Brownian motion. The functions $t \mapsto B_t(\omega)$, for $\omega \in \Omega$, are called sample paths (or: trajectories) of B. At this stage, we know nothing about them: it is even not clear (nor true, in general) that these functions are measurable. The aim of this section is to prove that, after a "slight modification" of B, we will have continuous sample paths.

Definition 4.1. *Let $(X_t,\, t \in \mathbf{T})$ and $(\widetilde{X}_t,\, t \in \mathbf{T})$ be processes indexed by the same set $\mathbf{T}$. We say that $\widetilde{X}$ is a modification (or : version) of X if*

$$\forall\, t \in \mathbf{T}, \qquad \mathbb{P}[X_t = \widetilde{X}_t] = 1.$$

So for any $t_1,\, t_2,\, \cdots,\, t_n$, the random vectors $(\widetilde{X}_{t_1}, \cdots, \widetilde{X}_{t_n})$ and $(X_{t_1}, \cdots, X_{t_n})$ have the same distribution. In particular, if X is Brownian motion, then so is $\widetilde{X}$. On the other hand, the trajectories of $\widetilde{X}$ may have a totally different behaviour from those of X. It can happen that the trajectories of $\widetilde{X}$ are all continues while those of X are all discontinuous.

Definition 4.2. *Two processes X and $\widetilde{X}$ are undistinguishable if*

$$\mathbb{P}[\,\forall\, t \in \mathbf{T}, \quad X_t = \widetilde{X}_t] = 1.$$

This means the set $\{\,\forall\, t \in \mathbf{T}, \quad X_t = \widetilde{X}_t\}$ contains a measurable event of probability 1. Notice that, a priori, we know nothing about the measurability of this set.

If X and $\widetilde{X}$ are undistinguishable, then $\widetilde{X}$ is obviously a modification of X. The notion of undistinguishability, however, is stronger: two undistinguishable processes almost surely have the same trajectories.

Assume that $\mathbf{T} = I$ is an interval of $\mathbb{R}$, and that the trajectories of X and $\widetilde{X}$ are a.s. continuous, then $\widetilde{X}$ is a modification of X if and only if X and $\widetilde{X}$ are undistinguishable: indeed, if $\widetilde{X}$ is a modification of X, then a.s. for all $t \in I \cap \mathbb{Q}$, $X_t = \widetilde{X}_t$. By continuity, a.s. for all $t \in I$, $X_t = \widetilde{X}_t$, which means that X and $\widetilde{X}$ are undistinguishable.

Theorem 4.3. (Kolmogorov's criterion). *Let $X = (X_t,\, t \in I)$ be a process indexed by an interval $I \subset \mathbb{R}$, taking values in a complete metric space (E, d). Suppose there exist $p > 0$, $\varepsilon > 0$ and $C > 0$ such that*

$$\mathbb{E}[\, d(X_s,\, X_t)^p] \le C\, |t-s|^{1+\varepsilon}, \qquad \forall\, s, t \in I.$$

Then there exists a modification $\widetilde{X}$ of X whose trajectories are locally Hölder continuous for exponent α, for any $\alpha \in (0, \frac{\varepsilon}{p})$, i.e., for all $T > 0$ and $\alpha \in (0, \frac{\varepsilon}{p})$, there exists $C_\alpha(T, \omega) > 0$ such that

$$d(\widetilde{X}_s(\omega),\, \widetilde{X}_t(\omega)) \le C_\alpha(T, \omega)\, |t-s|^\alpha, \qquad \forall\, s, t \in I,\ s, t \le T.$$

In particular, there exists a continuous modification of X, which is unique in the sense of undistinguishability.

Proof. The uniqueness is clear from the discussions in the previous paragraph.

We need to prove the existence. For notational simplification, we assume $I = [0,\, 1]$. By assumption, for $a > 0$ and $s,\, t \in [0,\, 1]$,

$$\mathbb{P}\{d(X_s,\, X_t) \ge a\} \le \frac{\mathbb{E}[d(X_s,\, X_t)^p]}{a^p} \le \frac{C\, |t-s|^{1+\varepsilon}}{a^p}.$$

Applying this inequality to $s = \frac{i-1}{2^n}$ and $t = \frac{i}{2^n}$ and $a = 2^{-n\alpha}$ gives

$$\mathbb{P}\{d(X_{(i-1)/2^n},\, X_{i/2^n}) \ge 2^{-n\alpha}\} \le \frac{C}{2^{(1+\varepsilon-p\alpha)n}}, \qquad i = 1,\, 2,\, \cdots,\, 2^n.$$

As such,

$$\mathbb{P}\{\exists i \leq 2^n : \; d(X_{(i-1)/2^n},\, X_{i/2^n}) \geq 2^{-n\alpha}\} \leq \frac{C}{2^{(\varepsilon - p\alpha)n}},$$

which is summable in n since $p\alpha < \varepsilon$. By the Borel–Cantelli lemma, there exists $A \in \mathscr{F}$ with $\mathbb{P}(A) = 1$ such that for all $\omega \in A$, there exists $n_0 = n_0(\omega) < \infty$ satisfying

$$d(X_{(i-1)/2^n},\, X_{i/2^n}) < 2^{-n\alpha}, \qquad \forall n \geq n_0, \;\; \forall 1 \leq i \leq 2^n. \tag{4.1}$$

Let D denote the set of all dyadic numbers in $[0,\, 1)$, i.e., the set of all $t \in [0,\, 1)$ that can be written as

$$t = \sum_{k=1}^{\ell} \frac{\varepsilon_k}{2^k},$$

with $\varepsilon_k = 0$ or 1. Consider $s,\, t \in D$ with $s < t$. Let $q \geq 0$ be the largest integer satisfying $t - s \leq 2^{-q}$. Let $k := \lfloor 2^q s \rfloor$, and $k \leq \lfloor 2^q t \rfloor \leq k+1$. We can find integers $\ell \geq 0$ and $m \geq 0$, such that

$$\begin{aligned} s &= \frac{k}{2^q} + \frac{\varepsilon_{q+1}}{2^{q+1}} + \cdots + \frac{\varepsilon_{q+\ell}}{2^{q+\ell}}, \\ t &= \frac{k}{2^q} + \frac{\widetilde{\varepsilon}_q}{2^q} + \frac{\widetilde{\varepsilon}_{q+1}}{2^{q+1}} + \cdots + \frac{\widetilde{\varepsilon}_{q+m}}{2^{q+m}}, \end{aligned}$$

where $\varepsilon_j,\, \widetilde{\varepsilon}_j = 0$ or 1 (if $q = 0$, then $k = 0$). If we write

$$\begin{aligned} s_i &= \frac{k}{2^q} + \frac{\varepsilon_{q+1}}{2^{q+1}} + \cdots + \frac{\varepsilon_{q+i}}{2^{q+i}}, \qquad 0 \leq i \leq \ell, \\ t_j &= \frac{k}{2^q} + \frac{\widetilde{\varepsilon}_q}{2^q} + \frac{\widetilde{\varepsilon}_{q+1}}{2^{q+1}} + \cdots + \frac{\widetilde{\varepsilon}_{q+j}}{2^{q+j}}, \qquad 0 \leq j \leq m, \end{aligned}$$

then for $\omega \in A$,

$$\begin{aligned} d(X_s,\, X_t) &= d(X_{s_\ell},\, X_{t_m}) \\ &\leq d(X_{s_0},\, X_{t_0}) + \sum_{i=1}^{\ell} d(X_{s_{i-1}},\, X_{s_i}) + \sum_{j=1}^{m} d(X_{t_{j-1}},\, X_{t_j}) \\ &\leq K_\alpha(\omega)\, 2^{-q\alpha} + \sum_{i=1}^{\ell} K_\alpha(\omega)\, 2^{-(q+i)\alpha} + \sum_{j=1}^{m} K_\alpha(\omega)\, 2^{-(q+j)\alpha}, \end{aligned}$$

où

$$K_\alpha(\omega) := \sup_{n\geq 1} \max_{1 \leq i \leq 2^n} \frac{d(X_{(i-1)/2^n}, X_{i/2^n})}{2^{-n\alpha}}$$

which is finite according to (4.1). So, for $\omega \in A$,

$$d(X_s,\, X_t) \leq 2K_\alpha(\omega) \sum_{i=0}^{\infty} 2^{-(q+i)\alpha} = \frac{2K_\alpha(\omega) 2^{-q\alpha}}{1 - 2^{-\alpha}} \leq \frac{2^{1+\alpha} K_\alpha(\omega)}{1 - 2^{-\alpha}}\, (t-s)^\alpha,$$

because $2^{-(q+1)} < t - s$. Thus, a.s. the function $t \mapsto X_t(\omega)$ is Hölder continuous on D and a fortiori uniformly continuous on D. Since (E, d) is complete, this function a.s. has a unique continuous extension to $I = [0, 1]$, and the extension is also Hölder continuous for exponent α. More precisely, define, for all $t \in [0, 1]$,

$$\widetilde{X}_t(\omega) := \lim_{s \to t,\, s \in D} X_s(\omega)$$

if $\omega \in A$, and $\widetilde{X}_t(\omega) := x_0$ (any element of E) if $\omega \notin A$. Then the trajectories of $\widetilde{X}$ are Hölder continuous for exponent α.

It remains to check that $\widetilde{X}$ is a modification of X. Let $t \in [0, 1]$. By assumption,

$$\lim_{s \to t} X_s = X_t, \qquad \text{in probability.}$$

By definition, $\widetilde{X}_t$ is the a.s. limit of X_s when $s \to t$ and $s \in D$. Hence $\widetilde{X}_t = X_t$ a.s. $\square$

Corollary 4.4. *Let $B = (B_t,\, t \geq 0)$ be Brownian motion. The process B admits a modification whose trajectories are locally Hölder continuous for exponent $\frac{1}{2} - \varepsilon$, for all $\varepsilon \in (0, \frac{1}{2})$.*

In particular, B admits a continuous modification.

Proof. Fix $\varepsilon \in (0, \frac{1}{2})$. Let $t,\, s \geq 0$. Since $B_t - B_s$ is Gaussian $\mathscr{N}(0, |t - s|)$, we have, for all $p > 0$, $\mathbb{E}[\,|B_t - B_s|^p] = C_p\,(t - s)^{p/2}$, where $C_p := \mathbb{E}[\,|\mathscr{N}(0, 1)|^p] < \infty$. It suffices to take p sufficiently large such that $\frac{1}{2} - \varepsilon < \frac{(p/2)-1}{p}$ to see that B admits a modification whose trajectories are locally Hölder continuous for exponent $\frac{1}{2} - \varepsilon$. $\square$

So if B is Brownian motion, one can always replace it with a process $\widetilde{B}$ such that $\forall\, t \geq 0$, $\mathbb{P}(\widetilde{B}_t = B_t) = 1$ and that the trajectories of $\widetilde{B}$ are continuous (and even locally Hölder continuous for exponent $\frac{1}{2} - \varepsilon$). From now on, we make this replacement systematically : this means that **in the definition of Brownian motion, one adds the condition that the trajectories are a.s. continuous**.

It is natural to ask whether the trajectories of Brownian motion can be locally Hölder continuous for exponent $\frac{1}{2}$. The answer is negative, which we will see later on. For the moment, we prove the following result which will give us an interesting corollary.

Proposition 4.5. *For any $\gamma > 1/2$, we have*

$$\mathbb{P}\Big[\forall\, t \geq 0 :\ \limsup_{h \to 0+} \frac{|B_{t+h} - B_t|}{h^\gamma} = \infty\Big] = 1\,.$$

Remark 4.6. It is possible to strengthen Proposition 4.5. As a matter of fact, Dvoretzky (1963) proves the existence of a $c > 0$ such that a.s.,

$$\forall\, t \geq 0,\ \limsup_{h \to 0+} \frac{|B_{t+h} - B_t|}{\sqrt{h}} \geq c\,. \qquad \square$$

Proof of Proposition 4.5. [The proof of the proposition is not part of the examination program.] Let $\gamma > 1/2$. Since

$$\Big\{ \exists\, t \geq 0 : \limsup_{h \to 0+} \frac{|B_{t+h} - B_t|}{h^\gamma} < \infty \Big\} \subset \bigcup_{m=1}^\infty \bigcup_{\ell=1}^\infty \bigcup_{k=1}^\infty \Big\{ \exists\, t \in [0, m] : |B_{t+h} - B_t| \leq \ell\, h^\gamma,\ \forall\, h \in (0, \frac{1}{k}] \Big\},$$

it suffices to prove that for $m \geq 1$, $\ell \geq 1$ and $\delta > 0$,

$$\mathbb{P}\Big\{ \exists\, t \in [0, m] : |B_{t+h} - B_t| \leq \ell\, h^\gamma,\ \forall\, h \in (0, \delta] \Big\} = 0\,.$$

Consider $A_{i,n} := \{\exists\, t \in [\frac{i}{n}, \frac{i+1}{n}] : |B_{t+h} - B_t| \leq \ell\, h^\gamma, \forall\, h \in (0,\, \delta]\ \}$. It suffices to check that for all $m \geq 1$, $\ell \geq 1$ and $\delta > 0$, we have $\sum_{i=0}^{nm-1} \mathbb{P}(A_{i,n}) \to 0$, $n \to \infty$.

Let $K > 2$ be an integer with $(K-2)(\gamma - \frac{1}{2}) > 1$. Let $n > n_0 := \lfloor K/\delta \rfloor$. If $\omega \in A_{i,n}$, and let t be as in the definition of $A_{i,n}$ (attention: t depends on ω), then $|B_{\frac{i+j}{n}} - B_t| \leq \ell(\frac{i+j}{n} - t)^\gamma \leq \ell\,(\frac{j}{n})^\gamma$ as long as $0 < \frac{i+j}{n} - t \leq \delta$ (a fortiori, if $2 \leq j \leq K$) ; this implies $|B_{\frac{i+j}{n}} - B_{\frac{i+j-1}{n}}| \leq 2\ell\,(\frac{K}{n})^\gamma$ for $3 \leq j \leq K$. Accordingly,

$$A_{i,n} \subset \bigcap_{j=3}^K \Big\{ |B_{\frac{i+j}{n}} - B_{\frac{i+j-1}{n}}| \leq 2\ell\,(\frac{K}{n})^\gamma \Big\}.$$

The events on the right-hand side being independent, we obtain:

$$\mathbb{P}(A_{i,n}) \leq \prod_{j=3}^K \mathbb{P}\Big\{ |B_{\frac{i+j}{n}} - B_{\frac{i+j-1}{n}}| \leq 2\ell\,(\frac{K}{n})^\gamma \Big\}.$$

Since $B_{\frac{i+j}{n}} - B_{\frac{i+j-1}{n}}$ is Gaussian $\mathscr{N}(0,\, \frac{1}{n})$, and since[1] $\mathbb{P}(|\mathscr{N}(0,\, 1)| < x) \leq (\frac{2}{\pi})^{1/2}\, x, \forall x > 0$, we have $\mathbb{P}(A_{i,n}) \leq \prod_{j=3}^K (\frac{2}{\pi})^{1/2}\, \frac{2\ell\, K^\gamma}{n^{\gamma - 1/2}} = \frac{c}{n^{(\gamma-1/2)(K-2)}}$, where $c := [(\frac{2}{\pi})^{1/2}\, 2\ell\, K^\gamma]^{K-2}$. Consequently, $\sum_{i=0}^{nm-1} \mathbb{P}(A_{i,n}) \leq m\, \frac{c}{n^{(\gamma-1/2)(K-2)-1}} \to 0$, $n \to \infty$, as $(K-2)(\gamma - \frac{1}{2}) > 1$. $\square$

Corollary 4.7. (Paley, Wiener and Zygmund 1933). *Almost surely, $t \mapsto B_t$ is nowhere differentiable.*

[1] The density of $\mathscr{N}(0,\, 1)$ is bounded by $\frac{1}{(2\pi)^{1/2}}$.

Since a function of finite variation is almost everywhere differentiable, this yields:

Corollary 4.8. *Almost surely, $t \mapsto B_t$ is not of finite variation in any interval (a, b) with $a < b$.*

We will give a refinement of this result in next chapter.

5. The canonical process and the Wiener measure

Consider $C(\mathbb{R}_+, \mathbb{R})$, the space of all real-valued functions on $\mathbb{R}_+$, endowed with the topology of uniform convergence on compacts:

$$d(\mathrm{w}, \widetilde{\mathrm{w}}) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{d_n(\mathrm{w}, \widetilde{\mathrm{w}})}{1 + d_n(\mathrm{w}, \widetilde{\mathrm{w}})},$$

where $d_n(\mathrm{w}, \widetilde{\mathrm{w}}) := \sup_{t \in [0,n]} |\mathrm{w}(t) - \widetilde{\mathrm{w}}(t)|$.

Let $(X_t, t \geq 0)$ be the process of projections:

$$X_t(\mathrm{w}) := \mathrm{w}(t), \qquad \mathrm{w} \in C(\mathbb{R}_+, \mathbb{R}).$$

The next result identifies the σ-field $\sigma(X_t, t \geq 0)$ generated by these projections (i.e., the smallest σ-field making all X_t measurable) with the Borel σ-field $\mathscr{C}(\mathbb{R}_+, \mathbb{R})$.

Lemma 5.1. *We have $\sigma(X_t, t \geq 0) = \mathscr{C}(\mathbb{R}_+, \mathbb{R})$.*

Proof. Is an exercise. $\square$

Let $Z = (Z_t, t \in \mathbb{R}_+)$ be a continuous process defined on $(\Omega, \mathscr{F}, \mathbb{P})$. Consider the mapping φ

$$\begin{array}{ccc} \Omega & \longrightarrow & C(\mathbb{R}_+, \mathbb{R}) \\ \omega & \longmapsto & \varphi(\omega) = (t \mapsto Z_t(\omega)), \end{array}$$

which is measurable according to Lemma 5.1. We call the **law** (or: distribution) of Z, the image-measure of $\mathbb{P}$ by φ. By the π-λ theorem, the law of Z is determined by the finite-dimensional distributions $(Z_{t_1}, \cdots, Z_{t_n})$: indeed, two measures on $C(\mathbb{R}_+, \mathbb{R})$ are identical if they attribute same value to sets of type $(X_{t_1}(\mathrm{w}), \cdots, X_{t_n}(\mathrm{w})) \in A$, $A \in \mathscr{B}(\mathbb{R}^n)$ (Borel σ-field of $\mathbb{R}^n$).

In the special case where Z is Brownian motion, this particular image-measure of Z is denoted by $\mathbb{W}$. It is a probability measure on $C(\mathbb{R}_+, \mathbb{R})$ such that $\mathbb{W}\{\mathrm{w} : \mathrm{w}(0) = 0\} = 1$, and that for all $n \geq 1$, $0 = t_0 < t_1 < t_2 < \cdots < t_n$ and $A \in \mathscr{B}(\mathbb{R}^n)$,

$$\mathbb{W}\Big\{\mathrm{w} : (X_{t_1}(\mathrm{w}), \cdots, X_{t_n}(\mathrm{w})) \in A\Big\}$$
$$= \int_A \frac{1}{(2\pi)^{n/2}\sqrt{t_1(t_2 - t_1)\cdots(t_n - t_{n-1})}} \exp\Big(-\frac{1}{2}\sum_{k=1}^{n} \frac{(x_k - x_{k-1})^2}{t_k - t_{k-1}}\Big)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n \,,$$

with $x_0 := 0$, because the integrand is the density function of the Gaussian random vector $(Z_{t_1}, Z_{t_2}, \cdots, Z_{t_n})$. This formula characterises the probability measure $\mathbb{W}$, and does not depend on the choice of Brownian motion in the construction. We call $\mathbb{W}$ the **Wiener measure**, and the process of projections $(X_t,\, t \geq 0)$ is the **canonical process** (of Brownian motion). Summarizing, we have

Theorem 5.2. *There exists a unique probability measure (which is the Wiener measure) on $C(\mathbb{R}_+, \mathbb{R})$ under which the process of projections $(X_t,\, t \geq 0)$ is a Brownian motion.*

The canonical process of Brownian motion often immediately answers the measurability question. For example, for all $t > 0$, $\sup_{0\leq s\leq t} X_s$ and $\int_0^t X_s^2\, \mathrm{d}s$ are random variables. In terms of generic Brownian motion, we know that for all $t > 0$, $\sup_{0\leq s\leq t} B_s$ and $\int_0^t B_s^2\, \mathrm{d}s$ are random variables if B is Brownian motion.

Let $x \in \mathbb{R}$. Let $\mathbb{W}_x$ be the image-measure of $\mathbb{W}$ by the mapping $\mathrm{w} \mapsto \mathrm{w} + x$. Clearly, $\mathbb{W}_x\{\mathrm{w} : \mathrm{w}(0) = x\} = 1$. The process of projections $(X_t,\, t \geq 0)$ under $\mathbb{W}_x$ is called Brownian motion starting at $X_0 = x$. It is a Lévy process whose trajectories are a.s. continuous, $X_0 = x$, a.s., such that $\forall\, t \geq s$, $X_t - X_s$ is Gaussian $\mathscr{N}(0,\, t-s)$. It coincides with the notion of Brownian motion in Definition 3.2.

Chapter 5

Brownian motion and the Markov property

In the previous chapter, we mainly studied Brownian motion as a Gaussian process. We now study the Markov property of Brownian motion, and leave martingale properties to the next chapter.

1. Elementary properties

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a complete probability space, and let $B = (B_t,\, t \geq 0)$ be Brownian motion.

Proposition 1.1. *The following processes are Brownian motions:*

(i) $X_t = -B_t$. *(symmetry)*

(ii) $X_t = tB_{1/t},\, X_0 = 0$. *(time inversion)*

(iii) $a > 0$ *fixed,* $X_t = \frac{1}{a^{1/2}} B_{at}$. *(scaling)*

(iv) $T > 0$ *fixed,* $X_t = B_T - B_{T-t},\, t \in [0,\, T]$. *(time reversal)*

Proof. Is trivial. It suffices to check, for each of the processes, that X is a centered Gaussian process with covariance $s \wedge t$. Only Part (ii) needs some special care because the trajectories are not necessarily continuous at 0: this however, does not cause any trouble because X is, according to Kolmogorov's criterion, undistinguishable to Brownian motion. $\square$

Example 1.2. (Brownian bridge). Let $b_t = B_t - tB_1,\, t \in [0,\, 1]$. It is a centered Gaussian process with a.s. continuous trajectories and with covariance $(s \wedge t) - st$. We call b a Brownian bridge.

The process $(b_t,\, t \in [0,\, 1])$ is independent of the random variable B_1.

If b is a Brownian bridge, so is $(b_{1-t},\, t \in [0,\, 1])$.

If b is a Brownian bridge, then $B_t = (1+t)b_{t/(1+t)}$, $t \geq 0$, is Brownian motion. Note that $b_t = (1-t)B_{t/(1-t)}$. □

Example 1.3. By continuity, $\lim_{t\to 0+} B_t = 0$, a.s., which, by time inversion, leads to:

$$\lim_{t\to\infty} \frac{B_t}{t} = 0, \qquad \text{a.s.}$$

It is possible to prove this directly, first using law of the large numbers (which says that $\frac{B_n}{n} \to 0$, a.s.), and then proves that B "does not move very much" during $[n,\, n+1]$. Left as an exercise. □

2. The simple Markov property

In this section, we denote by $\mathscr{F}_t$ (the completion of) the σ-field generated by $(B_s,\, 0 \leq s \leq t)$.

Theorem 2.1. (Simple Markov property). *Let $s \geq 0$. The process $(\widetilde{B}_t := B_{t+s} - B_s,\, t \geq 0)$ is Brownian motion, independent of $\mathscr{F}_s$.*

Proof. It is immediately checked that $\widetilde{B}$ is a centered Gaussian process with a.s. continuous trajectories and with covariance $\mathbb{E}(\widetilde{B}_t\widetilde{B}_{t'}) = t \wedge t'$: it is Brownian motion.

To prove independence, it suffices to show that for $0 \leq t_1 < \cdots < t_n$ and $0 < s_1 < \cdots < s_m \leq s$, the random vectors $(\widetilde{B}_{t_1},\, \cdots,\, \widetilde{B}_{t_n})$ and $(B_{s_1},\, \cdots,\, B_{s_m})$ are independent. Since $\mathrm{Cov}(\widetilde{B}_{t_i},\, B_{s_j}) = \mathbb{E}[(B_{s+t_i} - B_s)B_{s_j}] = 0$ (because $s \geq s_j$), and since $(\widetilde{B}_{t_1},\, \cdots,\, \widetilde{B}_{t_n},\, B_{s_1},\, \cdots,\, B_{s_m})$ is a Gaussian random vector, we see that $(\widetilde{B}_{t_1},\, \cdots,\, \widetilde{B}_{t_n})$ and $(B_{s_1},\, \cdots,\, B_{s_m})$ are independent. □

Proposition 2.2. *Let $s \geq 0$, and define*

$$\mathscr{F}_{s+} := \bigcap_{u>s} \mathscr{F}_u.$$

The process $(\widetilde{B}_t := B_{t+s} - B_s,\, t \geq 0)$ is independent of $\mathscr{F}_{s+}$.

Proof. It suffices to check that for $A \in \mathscr{F}_{s+}$, $0 \le t_1 < t_2 < \cdots < t_n$ and continuous and bounded $F : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}\Big[\mathbf{1}_A F(\widetilde{B}_{t_1}, \cdots, \widetilde{B}_{t_n})\Big] = \mathbb{P}(A)\,\mathbb{E}\Big[F(B_{t_1}, \cdots, B_{t_n})\Big]. \tag{2.1}$$

Let $\varepsilon > 0$. By the Markov property, $t \mapsto B_{t+s+\varepsilon} - B_{s+\varepsilon}$ is independent of $\mathscr{F}_{s+\varepsilon}$, and is, a fortiori, independent of $\mathscr{F}_{s+}$. Hence

$$\mathbb{E}\Big[\mathbf{1}_A F(B_{t_1+s+\varepsilon} - B_{s+\varepsilon}, \cdots, B_{t_n+s+\varepsilon} - B_{s+\varepsilon})\Big] = \mathbb{P}(A)\,\mathbb{E}\Big[F(B_{t_1}, \cdots, B_{t_n})\Big].$$

Letting $\varepsilon \to 0$, and by continuity of trajectories and the dominated convergence theorem, we obtain (2.1). $\square$

Theorem 2.3. (Blumenthal 0–1 law). *The σ-field $\mathscr{F}_{0+}$ is trivial, in the sense that $\forall\, A \in \mathscr{F}_{0+}$, $\mathbb{P}(A) = 0$ or 1.*

Proof. By Proposition 2.2, $\mathscr{F}_{0+}$ is independent of $\sigma(B_t,\, t \ge 0)$, and thus of the completion of $\sigma(B_t,\, t \ge 0)$. Let $A \in \mathscr{F}_{0+}$. Since $A \in \mathscr{F}_{0+} = \bigcap_{u>0} \mathscr{F}_u$, which is contained in the completion of $\sigma(B_t,\, t \ge 0)$, it follows that A is independent of itself. $\square$

Example 2.4. Let $\tau := \inf\{t > 0 :\, B_t > 0\}$. Then $\tau = 0$, a.s.

To prove this, we note that

$$\{\tau = 0\} = \bigcap_{s>0,\, s\in\mathbb{Q}} \{\sup_{0\le u\le s} B_u > 0\} \in \mathscr{F}_{0+}.$$

For $t > 0$, $\mathbb{P}(\tau \le t) \ge \mathbb{P}(B_t > 0) = \frac{1}{2}$. So $\mathbb{P}(\tau = 0) = \lim_{t\to 0+} \mathbb{P}(\tau \le t) \ge \frac{1}{2}$. By Blumenthal's 0–1 law, $\mathbb{P}(\tau = 0) = 1$.

Since $-B$ is Brownian motion, we see that $\inf\{t \ge 0 :\, B_t < 0\} = 0$, a.s. So B visits both $\mathbb{R}_+^*$ and $\mathbb{R}_-^*$ in any neighbourhood of 0. Consequently, there exists a.s. a sequence $(t_n = t_n(\omega))_{n\ge 0}$, decreasing to 0, such that $B_{t_n}(\omega) = 0$, $\forall\, n$. In particular, $\inf\{t > 0 :\, B_t = 0\} = 0$, a.s.

By inversion of time, $\{t > 0 :\, B_t = 0\}$ is a.s. unbounded.

On the other hand, we have seen that $\mathbb{P}(\forall\, t > 0,\, \sup_{0\le s\le t} B_s > 0) = 1$. Let $x > 0$. We have

$$\mathbb{P}\Big(\sup_{s\in[0,\,t]} B_s > x\Big) = \mathbb{P}\Big(\sup_{s\in[0,\,1]} B_s > \frac{x}{t^{1/2}}\Big).$$

We let $t \to +\infty$. The probability expression on the right-hand side tends to 1. So $\mathbb{P}(\sup_{s\ge 0} B_s > x) = 1$, $\forall\, x > 0$. In other words, $\sup_{s\ge 0} B_s = +\infty$, a.s. By symmetry, $\inf_{s\ge 0} B_s = -\infty$, a.s. (In particular, this confirms that $\{t > 0 :\, B_t = 0\}$ is a.s. unbounded.) If $T_a := \inf\{t > 0 :\, B_t = a\}$, $a \in \mathbb{R}$ (notation: $\inf \varnothing := \infty$), then $\mathbb{P}(T_a < \infty,\, \forall\, a \in \mathbb{R}) = 1$.$\square$

Example 2.5. Let $(t_n)_{n\ge 1}$ be a sequence decreasing to 0. Then a.s. $B_{t_n} > 0$ for infinitely many n, and $B_{t_n} < 0$ for infinitely many n.

Indeed, let $A_n := \{B_{t_n} > 0\}$, then $\mathbb{P}(\limsup A_n) = \lim_{N\to\infty} \mathbb{P}(\cup_{n=N}^{\infty} A_n)$, which is $\ge \limsup_{N\to\infty} \mathbb{P}(A_N) = \frac{1}{2}$. Since $\limsup A_n \in \mathscr{F}_{0+}$, it follows that $\mathbb{P}(\limsup A_n) = 1$. □

Example 2.6. We have

$$\limsup_{t\to\infty} \frac{B_t}{t^{1/2}} = \infty, \qquad \liminf_{t\to\infty} \frac{B_t}{t^{1/2}} = -\infty, \qquad \text{a.s.}$$

In fact, fix a constant $K > 0$, and let $A_n := \{\sqrt{n}\, B_{1/n} > K\}$. We have $\mathbb{P}(\limsup A_n) \ge \limsup_{N\to\infty} \mathbb{P}(A_N) = \mathbb{P}(B_1 > K) > 0$, so by Blumenthal's 0–1 law, $\mathbb{P}(\limsup A_n) = 1$, and a fortiori, $\limsup_{t\to 0} \frac{B_t}{t^{1/2}} = \infty$, a.s. (In particular, a.s. the trajectories of B are not $\frac{1}{2}$-Hölder.) We obtain the desired result by means of inversion of time and of symmetry. □

3. Semi-group of Brownian motion

Let $\mathscr{F}_t$ be as before (the completion of) $\sigma(B_s,\, 0 \le s \le t)$. For $x \in \mathbb{R}$, let $\mathbb{P}_x$ denote the probability such that B is Brownian motion with $\mathbb{P}_x(B_0 = x) = 1$.[1] So $\mathbb{P} = \mathbb{P}_0$.

The simple Markov property says that under $\mathbb{P}_x$, $(\widetilde{B}_t := B_{t+s} - B_s,\, t \ge 0)$ is Brownian motion independent of $\mathscr{F}_s$.

We can state the Markov property of Brownian motion in the following, more familiar way: conditionally on $\mathscr{F}_s$, $(\widehat{B}_t := B_{t+s},\, t \ge 0)$ is Brownian motion starting at $y = B_s$. In fact, $\widehat{B}_t = \widetilde{B}_t + B_s$, and for $0 \le t_1 < t_2 < \cdots < t_n$ and continuous and bounded $F : \mathbb{R}^n \to \mathbb{R}$,

$$\begin{aligned} \mathbb{E}_x\Big[F(\widehat{B}_{t_1}, \cdots, \widehat{B}_{t_n}) \,|\, \mathscr{F}_s \Big] &= \mathbb{E}_x\Big[F(\widetilde{B}_{t_1} + y, \cdots, \widetilde{B}_{t_n} + y) \,|\, \mathscr{F}_s \Big] \\ &= \mathbb{E}_0\Big[F(B_{t_1} + y, \cdots, B_{t_n} + y) \Big] \\ &= \mathbb{E}_y\Big[F(B_{t_1}, \cdots, B_{t_n}) \Big], \end{aligned}$$

with $y := B_s$.

By the Markov property, for $s > 0$ and Borel function $f : \mathbb{R} \to \mathbb{R}_+$,

$$\mathbb{E}_x[\, f(B_{t+s}) \,|\, \mathscr{F}_s \,] = P_t f(B_s) = \mathbb{E}_x[\, f(B_{t+s}) \,|\, B_s \,],$$

[1] Alternatively, we can work on the canonical space of Brownian motion, can define $\mathbb{W}_x$ to be the image-measure of $\mathbb{W}$ by the mapping $\mathrm{w} \mapsto \mathrm{w} + x$. As such, the process of projections $(X_t,\, t \ge 0)$ under $\mathbb{W}_x$ is Brownian motion with $\mathbb{W}_x(X_0 = x) = 1$.

where
$$P_t f(x) := \mathbb{E}_x[f(B_t)] = \int_{\mathbb{R}} \frac{1}{(2\pi t)^{1/2}} \exp\Big(-\frac{(x-y)^2}{2t}\Big) f(y)\,\mathrm{d}y.$$
Hence $P_{t+s}f(x) = \mathbb{E}_x(f(B_{t+s})) = \mathbb{E}_x[\,\mathbb{E}_x(f(B_{t+s})\,|\,\mathscr{F}_t)] = \mathbb{E}_x[P_s f(B_t)] = P_t(P_s f)(x)$, i.e., $P_t(P_s f) = P_{t+s} f$.

Feller property: if $f \in C_0$ (continuous, with $\lim_{|x|\to\infty} f(x) = 0$), then $P_t f \in C_0$, and $\lim_{t\downarrow 0} P_t f = f$ uniformly on $\mathbb{R}$. Left as an exercise.

Generator: if $f \in C_c^2$ (class C^2 of compact support), then $\lim_{t\downarrow 0} \frac{P_t f(x) - f(x)}{t} = \frac{1}{2} f''(x)$. Left as an exercise.

Relation with the heat equation: let $u(t,x) := P_t f(x)$. We have $u(0,x) = f(x)$. If f is a bounded Borel function, then
$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}.$$
Left as an exercise.

4. Strong Markov property

Let B be Brownian motion defined on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and let $\mathscr{F}_t$ be (the completion of) $\sigma(B_s,\, 0 \le s \le t)$. We write $\mathscr{F}_\infty$ for (the completion of) $\sigma(B_s,\, s \ge 0)$. The Markov property tells us that for any $s \ge 0$, $(B_{t+s} - B_s,\, t \ge 0)$ is Brownian motion, independent of $\mathscr{F}_s$. In this section, we extend this property to random times s.

Definition 4.1. *A mapping $T : \Omega \to \mathbb{R}_+ \cup \{\infty\}$ is called a stopping time if for any $t \ge 0$, $\{T \le t\} \in \mathscr{F}_t$.*

Example 4.2. The constant time $T = t$ is a stopping time. Another example is $T := T_a$, where $T_a := \inf\{t > 0 :\ B_t = a\}$: indeed, for $a \ge 0$, $\{T_a \le t\} = \{\sup_{s\in[0,\,t]} B_s \ge a\} \in \mathscr{F}_t$.

However, $T := \sup\{s \le 1 :\ B_s = 0\}$ is not a stopping time (this will be a consequence of the strong Markov property below and of Example 2.4). $\square$

Definition 4.3. *Let T be a stopping time. The σ-field generated by T is*
$$\mathscr{F}_T := \Big\{ A \in \mathscr{F}_\infty :\ \ \forall\, t \ge 0,\ A \cap \{T \le t\} \in \mathscr{F}_t \Big\}.$$

Example 4.4. We claim that T and $B_T\,\mathbf{1}_{\{T<\infty\}}$ are $\mathscr{F}_T$-measurable. For $B_T\,\mathbf{1}_{\{T<\infty\}}$, it suffices to see that a.s.,

$$B_T\,\mathbf{1}_{\{T<\infty\}} = \lim_{n\to\infty}\sum_{i=0}^{\infty}\mathbf{1}_{\{\frac{i}{2^n}<T\le\frac{i+1}{2^n}\}}B_{\frac{i}{2^n}},$$

and that $B_s\,\mathbf{1}_{\{s<T\}}$ and $\mathbf{1}_{\{T\le t\}}$ are $\mathscr{F}_T$-measurable. □

Theorem 4.5. (Strong Markov property). *Let T be a stopping time. Let $x\in\mathbb{R}$. Under P_x, conditionally on $\{T<\infty\}$, the process $\widetilde{B}:=(B_{T+t}-B_T,\,t\ge 0)$ is Brownian motion starting at 0, independent of $\mathscr{F}_T$.*[2]

Proof. Suppose first $T<\infty$, $\mathbb{P}_x$-a.s. We are going to prove, for $A\in\mathscr{F}_T$, $0\le t_1<\cdots<t_n$ and $F:\mathbb{R}^n\to\mathbb{R}_+$ continuous and bounded,

$$\mathbb{E}_x\Big[\mathbf{1}_A\,F(\widetilde{B}_{t_1},\cdots,\widetilde{B}_{t_n})\Big]=\mathbb{P}_x(A)\,\mathbb{E}\Big[F(B_{t_1},\cdots,B_{t_n})\Big]. \tag{4.1}$$

This will yield that $\widetilde{B}$ is Brownian motion starting at 0 (by taking $A=\Omega$), and that it is independent of $\mathscr{F}_T$.

Observe that

$$\sum_{k=0}^{\infty}\mathbf{1}_{\{\frac{k-1}{2^m}<T\le\frac{k}{2^m}\}}F(B_{\frac{k}{2^m}+t_1}-B_{\frac{k}{2^m}},\cdots,B_{\frac{k}{2^m}+t_n}-B_{\frac{k}{2^m}})$$

converges a.s. (when $m\to\infty$) to $F(\widetilde{B}_{t_1},\cdots,\widetilde{B}_{t_n})$. By dominated convergence,

$$\begin{aligned}&\mathbb{E}_x\Big[\mathbf{1}_A\,F(\widetilde{B}_{t_1},\cdots,\widetilde{B}_{t_n})\Big]\\ =\;&\lim_{m\to\infty}\sum_{k=0}^{\infty}\mathbb{E}_x\Big[\mathbf{1}_A\mathbf{1}_{\{\frac{k-1}{2^m}<T\le\frac{k}{2^m}\}}F(B_{\frac{k}{2^m}+t_1}-B_{\frac{k}{2^m}},\cdots,B_{\frac{k}{2^m}+t_n}-B_{\frac{k}{2^m}})\Big].\end{aligned}$$

For each k, $A\cap\{\frac{k-1}{2^m}<T\le\frac{k}{2^m}\}\in\mathscr{F}_{\frac{k}{2^m}}$. By the Markov property,

$$\begin{aligned}\mathbb{E}_x\Big[\mathbf{1}_A\,F(\widetilde{B}_{t_1},\cdots,\widetilde{B}_{t_n})\Big]&=\lim_{m\to\infty}\sum_{k=0}^{\infty}\mathbb{E}_x\Big[\mathbf{1}_A\mathbf{1}_{\{\frac{k-1}{2^m}<T\le\frac{k}{2^m}\}}\Big]\,\mathbb{E}\Big[F(B_{t_1},\cdots,B_{t_n})\Big].\\ &=\mathbb{P}_x(A)\,\mathbb{E}\Big[F(B_{t_1},\cdots,B_{t_n})\Big],\end{aligned}$$

[2] We have only defined the process $\widetilde{B}$ on $\{T<\infty\}$, though its values have no influence on the statement of the theorem. On $\{T=\infty\}$, we could take $\widetilde{B}_t:=0$ for $t\ge 0$.

from which (4.1) follows.

When $\mathbb{P}_x(T = \infty) > 0$, the same argument gives

$$\mathbb{E}_x\Big[\mathbf{1}_{A\cap\{T<\infty\}}\, F(\widetilde{B}_{t_1}, \cdots, \widetilde{B}_{t_n})\Big] = \mathbb{P}_x(A \cap \{T < \infty\})\, \mathbb{E}\Big[F(B_{t_1}, \cdots, B_{t_n})\Big],$$

and the desired result follows again. □

Example 4.6. Let $T_a := \inf\{t > 0 :\ B_t = a\}$. By the strong Markov property, the process $(T_a,\, a \geq 0)$ is a Lévy process, with non-decreasing trajectories. Furthermore, for all $c > 0$, the processes $(\frac{1}{c^2} T_{ca},\, a \geq 0)$ and $(T_a,\, a \geq 0)$ have the same finite-dimensional distributions. One says that $(T_a,\, a \geq 0)$ is a stable subordinator of index $\frac{1}{2}$. □

Theorem 4.7. (Reflection principle). *Let* $S_t = \sup_{s\in[0,\,t]} B_s$, $t > 0$. *Then*

$$\mathbb{P}(S_t \geq a,\ B_t \leq b) = \mathbb{P}(B_t \geq 2a - b), \qquad a \geq 0,\ \ b \leq a. \tag{4.2}$$

In particular, for any fixed $t > 0$, S_t *has the same distribution as* $|B_t|$.

Remark 4.8. The identity in law between S_t and $|B_t|$ is valid only for each fixed $t > 0$. The processes $(S_t,\, t \geq 0)$ and $(|B_t|,\, t \geq 0)$ obviously have different behaviours (for example, the former is non-decreasing, which is not the case with the latter). □

Proof of Theorem 4.7. Recall that $T_a < \infty$, a.s. We have

$$\mathbb{P}(S_t \geq a,\ B_t \leq b) = \mathbb{P}(T_a \leq t,\ B_t \leq b) = \mathbb{P}(T_a \leq t,\ \widetilde{B}_{t-T_a} \leq b - a),$$

where $\widetilde{B}_s := B_{s+T_a} - B_{T_a} = B_{s+T_a} - a$. So $\mathbb{P}(S_t \geq a,\ B_t \leq b) = \mathbb{P}\{(T_a,\, \widetilde{B}) \in A_t\}$, where $A_t := \{(s,\, \mathrm{w}) \in \mathbb{R}_+ \times C(\mathbb{R}_+,\, \mathbb{R}) :\ s \leq t,\, \mathrm{w}(t-s) \leq b - a\}$ is measurable with respect to $\mathscr{B}(\mathbb{R}_+) \otimes \mathscr{C}(\mathbb{R}_+,\, \mathbb{R})$.

By the strong Markov property, $\widetilde{B}$ is Brownian motion, independent of $\mathscr{F}_{T_a}$, a fortiori of T_a. In particular, $(T_a,\, -\widetilde{B})$ has the same distribution as $(T_a,\, \widetilde{B})$. Therefore,

$$\mathbb{P}(S_t \geq a,\ B_t \leq b) = \mathbb{P}(T_a \leq t,\ -\widetilde{B}_{t-T_a} \leq b - a) = \mathbb{P}(T_a \leq t,\ B_t \geq 2a - b),$$

proving (4.2) because $\{B_t \geq 2a - b\} \subset \{T_a \leq t\}$.

To complete the proof of the theorem, it remains to note that

$$\mathbb{P}(S_t \geq a) = \mathbb{P}(S_t \geq a,\, B_t \geq a) + \mathbb{P}(S_t \geq a,\, B_t \leq a) = 2\mathbb{P}(B_t \geq a) = \mathbb{P}(|B_t| \geq a). \qquad □$$

Corollary 4.9. *Let $t > 0$. The pair (S_t, B_t) has density*

$$f_{(S_t,B_t)}(a,b) = \frac{2(2a-b)}{(2\pi t^3)^{1/2}} \exp\Big(- \frac{(2a-b)^2}{2t}\Big) \mathbf{1}_{\{a>0,\, b<a\}} \, .$$

Example 4.10. By Theorem 4.7, for any $t > 0$,

$$\mathbb{P}(T_a \le t) = \mathbb{P}(S_t \ge a) = \mathbb{P}(|B_t| \ge a) = \mathbb{P}(t^{1/2}\, |B_1| \ge a) = \mathbb{P}\Big(\frac{a^2}{B_1^2} \le t\Big).$$

Hence T_a is distributed as $\frac{a^2}{B_1^2}$, $\forall a \in \mathbb{R}$. As a consequence, for $a \neq 0$,

$$f_{T_a}(t) = \frac{|a|}{(2\pi t^3)^{1/2}} \exp\Big(- \frac{a^2}{2t}\Big) \mathbf{1}_{\{t>0\}} \, .$$

In particular, $\mathbb{E}(T_a) = \infty$ if $a \neq 0$. $\square$

Reformulation of the strong Markov property.

An $\mathbb{R}^d$-valued process is Brownian motion if its components are independent real-valued Brownian motions. Most of the properties we have studied so far are valid for multidimensional Brownian motion; for example, the strong Markov property for multidimensional Brownian motion is proved using exactly the same proof.

We now reformulate the strong Markov property in the canonical space. Let $\Omega := C(\mathbb{R}_+, \mathbb{R}^d)$ endowed with the Wiener measure $\mathbb{P}$; let $(B_t,\, t \ge 0)$ denote the canonical process. For any $t \ge 0$, let $\mathscr{F}_t$ be (the completion of) $\sigma(B_s,\, s \in [0,\, t])$ and let $\mathscr{F}_\infty$ be (the completion of) $\sigma(B_s,\, s \ge 0)$. Let $\mathbb{P}_x$ be the probability under which B is $\mathbb{R}^d$-valued Brownian motion with $\mathbb{P}_x(B_0 = x) = 1$.

Let us introduce the shift operators as follows: for $s \ge 0$, $\theta_s : \Omega \to \Omega$ is defined by

$$(\theta_s \mathrm{w})(t) := \mathrm{w}(t+s), \qquad t \ge 0.$$

In other words, $B_t \circ \theta_s = B_{t+s}$.

Theorem 4.11. (Strong Markov property) *Let T be a stopping time. Let $F : \Omega \to \mathbb{R}_+$ be measurable. For $x \in \mathbb{R}^d$,*

$$\mathbb{E}_x[\mathbf{1}_{\{T<\infty\}}\, (F \circ \theta_T) \,|\, \mathscr{F}_T] = \mathbf{1}_{\{T<\infty\}}\, \mathbb{E}_{B_T}(F), \qquad \mathbb{P}_x\text{-a.s.}$$

[It sometimes helps to write as $\mathbb{E}_x[\mathbf{1}_{\{T<\infty\}}\, F(B_{T+\bullet}) \,|\, \mathscr{F}_T] = \mathbf{1}_{\{T<\infty\}}\, \mathbb{E}_{B_T}[F(B)]$, $\mathbb{P}_x$-a.s.]

Proof. On the set $\{T(\mathrm{w}) < \infty\}$,

$$(\theta_T \mathrm{w})(t) = \mathrm{w}(T+t) = \mathrm{w}(T) + (\mathrm{w}(T+t) - \mathrm{w}(T)) = B_T(\mathrm{w}) + \widetilde{B}_t(\mathrm{w})\,,$$

where $\widetilde{B}_t(\mathrm{w}) := B_{T+t}(\mathrm{w}) - B_T(\mathrm{w})$. So

$$\mathbb{E}_x[\mathbf{1}_{\{T<\infty\}}\,(F \circ \theta_T) \,|\, \mathscr{F}_T] = \mathbb{E}_x[\mathbf{1}_{\{T<\infty\}}\,F(B_T + \widetilde{B}) \,|\, \mathscr{F}_T]\,.$$

Since B_T is $\mathscr{F}_T$-measurable, whereas $\widetilde{B}$ is independent of $\mathscr{F}_T$ (conditionally on $T < \infty$) and is distributed as B under $\mathbb{P}_0$, it follows that

$$\mathbb{E}_x[\mathbf{1}_{\{T<\infty\}}\,(F \circ \theta_T) \,|\, \mathscr{F}_T] = \mathbf{1}_{\{T<\infty\}} h(B_T)$$

where $h(y) := \mathbb{E}_0[F(y+B)] = \mathbb{E}_y[F(B)]$. $\square$

Chapter 6

Brownian motion and martingales

We first prove a few general properties for continuous-time martingales, and then apply them to Brownian motion.

1. Continuous-time martingales

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A **filtration** $(\mathscr{F}_t)_{t\geq 0}$ on this space is a non-decreasing family of sub-σ-fields of $\mathscr{F}$. We say that $(\Omega, \mathscr{F}, (\mathscr{F}_t), \mathbb{P})$ is a filtered probability space. We always assume that it is complete, i.e., for any $t \geq 0$, $\mathscr{F}_t$ is complete.

Example 1.1. Let B be Brownian motion, and let

$$\mathscr{F}_t := \sigma\{B_s,\, 0 \leq s \leq t\}.$$

Then $(\mathscr{F}_t)$ is a filtration. More generally, if $(X_t,\ t \geq 0)$ is a process, then $(\mathscr{F}_t := \sigma\{X_s,\, 0 \leq s \leq t\},\ t \geq 0)$ is a filtration, and is called the "canonical filtration" of $(X_t,\ t \geq 0)$. □

A process $(X_t,\, t \geq 0)$ is right-continuous (resp. left-continuous) if its trajectories are a.s. right-continuous (resp. left-continuous).

Let $(\Omega, \mathscr{F}, (\mathscr{F}_t), \mathbb{P})$ a filtered probability space. We define[1]

$$\mathscr{F}_\infty := \sigma\Big(\bigcup_{t\geq 0} \mathscr{F}_t\Big),$$

the σ-field generated by all the elements of all the σ-fields $\mathscr{F}_t$, $t \geq 0$.

[1] In some books, $\sigma\Big(\bigcup_{t\geq 0} \mathscr{F}_t\Big)$ is denoted by $\bigvee_{t\geq 0} \mathscr{F}_t$.

A mapping $T:\,\Omega\to\mathbb{R}_+\cup\{+\infty\}$ is a **stopping time** if $\forall\,t\ge 0$, $\{T\le t\}\in\mathscr{F}_t$. If T is a stopping time, we define the σ-field generated by T:

$$\mathscr{F}_T := \{A\in\mathscr{F}_\infty:\ \forall\,t\ge 0,\, A\cap\{T\le t\}\in\mathscr{F}_t\}.$$

If $T=t$, then $\mathscr{F}_T=\mathscr{F}_t$. Furthermore, T is $\mathscr{F}_T$-measurable.

Exercise 1.2. Let S and T be stopping times.

(i) If $S\le T$, then $\mathscr{F}_S\subset\mathscr{F}_T$.

(ii) Both $S\wedge T$ and $S\vee T$ are stopping times, and $\mathscr{F}_{S\wedge T}=\mathscr{F}_S\cap\mathscr{F}_T$. Moreover, $\{S\le T\}\in\mathscr{F}_{S\wedge T}$, $\{S=T\}\in\mathscr{F}_{S\wedge T}$, $\{S<T\}\in\mathscr{F}_{S\wedge T}$.

(iii) $S+T$ is a stopping time.

Exercise 1.3. Let T be a stopping time. Then

$$T_n := \sum_{k=0}^{\infty}\frac{k}{2^n}\,\mathbf{1}_{\{\frac{k-1}{2^n}<T\le\frac{k}{2^n}\}} + (+\infty)\,\mathbf{1}_{\{T=\infty\}}$$

is a non-increasing sequence of stopping times such that $T_n(\omega)\downarrow T(\omega)$ for all $\omega\in\Omega$.

A process $(X_t,\,t\ge 0)$ is said to be adapted to the filtration $(\mathscr{F}_t)$ if for any t, X_t is $\mathscr{F}_t$-measurable. For example, $(X_t,\,t\ge 0)$ is adapted to its canonical filtration.

Exercise 1.4. If (X_t) is an $\mathbb{R}^d$-valued adapted and right-continuous (or: left-continuous) process, and if T is a stopping time, then $X_T\,\mathbf{1}_{\{T<\infty\}}$ is $\mathscr{F}_T$-measurable.

Moreover, if $X_t\to X_\infty$ a.s., then X_T is $\mathscr{F}_T$-measurable.

Example 1.5. If (X_t) is continuous and adapted, then for any closed set F,

$$T_F := \inf\{t\ge 0:\ X_t\in F\}$$

is a stopping time.

In fact, $\{T_F\le t\}=\{\inf_{s\in[0,\,t]} d(X_s,F)=0\}$ (by continuity of the trajectories), which is $\{\inf_{s\in[0,\,t]\cap\mathbb{Q}} d(X_s,F)=0\}$. □

Let $(\Omega,\mathscr{F},(\mathscr{F}_t),\mathbb{P})$ be a complete filtered probability space.

Definition 1.6. *We say that $(X_t,\, t \geq 0)$ is a martingale [resp. supermartingale; submartingale] with respect to $(\mathscr{F}_t)_{t\geq 0}$ if*

(i) *(X_t) is adapted with respect to $(\mathscr{F}_t)_{t\geq 0}$;*

(ii) *$\forall\, t \geq 0$, $\mathbb{E}(|X_t|) < \infty$;*

(iii) *$\forall\, s < t$, $\mathbb{E}(X_t \,|\, \mathscr{F}_s) = X_s$,* a.s. *[resp., $\leq X_s$; $\geq X_s$].*

Exercise 1.7. Let $(B_t, t \geq 0)$ be Brownian motion, and let $(\mathscr{F}_t)$ be its canonical filtration. Then the following processes are martingales.

(i) $(B_t, t \geq 0)$.

(ii) $(B_t^2 - t,\, t \geq 0)$.

(iii) For any $\theta \in \mathbb{R}$, $(\mathrm{e}^{\theta B_t - \frac{\theta^2}{2} t},\, t \geq 0)$.

A process $(X_t,\, t \geq 0)$ has independent increments if for $0 \leq t_1 < t_2 < \cdots < t_n$, $X_{t_n} - X_{t_{n-1}}, X_{t_{n-1}} - X_{t_{n-2}}, \cdots, X_{t_1}$ are independent (so by the π-λ theorem, for all $s < t$, $X_t - X_s$ is independent of $\mathscr{F}_s$, if $(\mathscr{F}_t)$ denotes the canonical filtration of (X_t)). Brownian motion has independent increments. Another example is the Poisson process $N_t := \sum_{n=1}^{\infty} \mathbf{1}_{\{T_n \leq t\}}$, where $T_n := W_1 + \cdots W_n$, and (W_i) are i.i.d. exponential random variables. More generally, a Lévy process has independent (and stationary) increments. So, this is the case with $(T_a,\, a \geq 0)$, where $T_a := \inf\{t \geq 0 :\, B_t = a\}$.

Exercise 1.8. Let $(X_t,\, t \geq 0)$ be a process having independent increments, and let $(\mathscr{F}_t)$ be its canonical filtration.

(i) If for all t, $\mathbb{E}(|X_t|) < \infty$, then $\widetilde{X}_t := X_t - \mathbb{E}(X_t)$ is a martingale.

(ii) If for all t, $\mathbb{E}(X_t^2) < \infty$, then $Y_t := \widetilde{X}_t^2 - \mathbb{E}(\widetilde{X}_t^2)$ is a martingale.

(iii) Let $\theta \in \mathbb{R}$. If $\mathbb{E}(\mathrm{e}^{\theta X_t}) < \infty$ for all $t \geq 0$, then $(Z_t := \frac{\mathrm{e}^{\theta X_t}}{\mathbb{E}[\mathrm{e}^{\theta X_t}]},\, t \geq 0)$ is a martingale.

It is easily seen that most properties of discrete-time martingales have an obvious analogue for continuous-time martingales. For example, if (X_t) is a martingale and if f is a convex function such that $\mathbb{E}(|f(X_t)|) < \infty$, $\forall\, t$, then $(f(X_t))$ is a submartingale.

Theorem 1.9. (Doob's L^p inequality). *Let $p > 1$. Let (X_s) be a right-continuous martingale. Then for any $t \geq 0$,*

$$\Big\| \sup_{s\in[0,t]} |X_s| \Big\|_p \leq q\, \|X_t\|_p \,,$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. Let $0 \leq t_1 < t_2 < \cdots < t_k = t$. Then $Y_n := X_{t_{n\wedge k}}$ is a discrete-time martingale. By Doob's L^p inequality for discrete-time martingales,

$$\Big\| \max_{0\leq i\leq k} |X_{t_i}| \Big\|_p \leq q \, \|X_t\|_p \, .$$

Let $D \subset \mathbb{R}_+$ be a countable set containing t. Par the monotone convergence theorem,

$$\Big\| \sup_{s\in D} |X_s| \Big\|_p \leq q \, \|X_t\|_p \, .$$

Since the trajectories of (X_s) is a.s. right-continuous, we get $\sup_{s\in[0,t]\cap D} |X_s| = \sup_{s\in[0,t]} |X_s|$ a.s., which yields the theorem. □

Corollary 1.10. *Let (X_s) be a right-continuous martingale. Then*

$$\Big\| \sup_{s\geq 0} |X_s| \Big\|_p \leq q \sup_{s\geq 0} \|X_s\|_p \, .$$

We now establish some convergence results.

Theorem 1.11. *Let $(X_t, \ t \geq 0)$ be a right-continuous submartingale satisfying*[2]

$$\sup_{t\geq 0} \mathbb{E}(X_t^+) < \infty.$$

Then $X_\infty := \lim_{t\to\infty} X_t$ exists a.s.*, and* $\mathbb{E}(|X_\infty|) < \infty$.

Proof. (i) Let $D \subset \mathbb{R}_+$ be countable and dense. Let $a < b$ and let $X([0, t] \cap D)$ be the number of crossings of $(X_s, \ s \in [0, t] \cap D)$ over $[a, b]$. Then

$$\mathbb{E}\left[X_{ab}([0, t] \cap D)\right] \leq \frac{\mathbb{E}[(X_t - a)^+]}{b-a} \leq \frac{1}{b-a}\Big(\sup_{s\geq 0} \mathbb{E}(X_s^+) + |a| \Big).$$

Let $t \to \infty$. By the monotone convergence theorem, we obtain $\mathbb{E}[X_{ab}(D)] < \infty$, where $X_{ab}(D)$ is the number of crossings of $(X_s, \ s \in D)$ over $[a, b]$. Hence a.s., for all rational numbers $a < b$, $X_{ab}(D) < \infty$. This implies that $X_\infty := \lim_{t\to\infty, \, t\in D} X_t$ exists a.s. By Fatou's lemma, $\mathbb{E}(|X_\infty|) \leq \liminf_{t\to\infty, \, t\in D} \mathbb{E}(X_t) \leq \sup_{s\geq 0} \mathbb{E}(|X_s|) < \infty$.

(ii) Let us prove $X_s \to X_\infty$ a.s.

Let $A \in \mathscr{F}$ be such that $\mathbb{P}(A) = 1$, and that for all $\omega \in A$, $s \mapsto X_s(\omega)$ is right-continuous on $\mathbb{R}_+$, and $X_\infty(\omega) = \lim_{t\to\infty, t\in D} X_t(\omega)$ exists and is finite.

Let $\varepsilon > 0$ and $\omega \in A$. There exists $t_0 = t_0(\omega) < \infty$ such that $|X_t(\omega) - X_\infty(\omega)| \leq \varepsilon$ for $t \geq t_0$ and $t \in D$. Since $s \mapsto X_s(\omega)$ is right-continuous, and since D is dense in $\mathbb{R}_+$, we have $|X_s(\omega) - X_\infty(\omega)| \leq 2\varepsilon$ for all $s \geq t_0$. In other words, $X_s \to X_\infty$ a.s. □

[2] In the exercise class, we are going to see that for submartingales, $\sup_{t\geq 0} \mathbb{E}(X_t^+) < \infty$ is equivalent to $\sup_{t\geq 0} \mathbb{E}(|X_t|) < \infty$.

Corollary 1.12. *Let $(X_t,\ t \ge 0)$ be a nonnegative and right-continuous* **super***martingale. Then $X_\infty := \lim_{t\to\infty} X_t$ exists* a.s., *and $\mathbb{E}(X_\infty) \le \mathbb{E}(X_0)$.*

Theorem 1.13. *Let $p > 1$. If $(X_t,\ t \ge 0)$ is a right-continuous martingale satisfying*

$$\sup_{t\ge 0} \mathbb{E}(|X_t|^p) < \infty,$$

then $X_t \to X_\infty$ a.s. *and in L^p.*

Proof. (i) By Jensen's (or more generally, Hölder's) inequality, $\sup_{t\ge 0} \mathbb{E}(|X_t|) < \infty$. Theorem 1.11 tells us that $X_t \to X_\infty$ a.s.

(ii) By Doob's inequality, $\mathbb{E}(\sup_{t\ge 0} |X_t|^p) < \infty$. Since $X_t \to X_\infty$ a.s., the dominated convergence theorem implies that $X_t \to X_\infty$ in L^p. □

Theorem 1.14. *Let $(X_t,\ t \ge 0)$ be a uniformly integrable and right-continuous martingale. Then*

(i) $X_t \to X_\infty$ in L^1;

(ii) $X_t \to X_\infty$ a.s.*;*

(iii) $X_t = \mathbb{E}(X_\infty \,|\, \mathscr{F}_t)$, a.s.

Proof. (ii) The uniform integrability implies $\sup_{t\ge 0} \mathbb{E}(|X_t|) < \infty$. So by Theorem 1.11, $X_t \to X_\infty$ a.s., and $\mathbb{E}(X_\infty) < \infty$.

(i) The L^1 convergence, being a consequence of convergence in probability and uniform integrability, follows from (ii).

(iii) Let $t \ge s$ and $A \in \mathscr{F}_s$. Then $\mathbb{E}(X_t \mathbf{1}_A) = \mathbb{E}(X_s \mathbf{1}_A)$. We let $t \to \infty$. The L^1 convergence implies $\mathbb{E}(X_t \mathbf{1}_A) \to \mathbb{E}(X_\infty \mathbf{1}_A)$. So $\mathbb{E}(X_s \mathbf{1}_A) = \mathbb{E}(X_\infty \mathbf{1}_A)$ for all $A \in \mathscr{F}_s$. In other words, $X_s = \mathbb{E}(X_\infty \,|\, \mathscr{F}_s)$, a.s. □

Theorem 1.15. (Doob's optional sampling theorem). *Let $(X_t,\ t \ge 0)$ be a right-continuous martingale and let $S \le T$ be stopping times. If $(X_t,\ t \ge 0)$ is uniformly integrable, then*

$$\mathbb{E}(X_T \,|\, \mathscr{F}_S) = X_S \qquad \text{a.s.} \tag{1.1}$$

In particular, $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ for any stopping time T.

Proof. (i) Let us prove first $\mathbb{E}(|X_T|) < \infty$.

Define

$$T_n := \sum_{k=0}^{\infty} \frac{k}{2^n} \mathbf{1}_{\{\frac{k-1}{2^n} < T \le \frac{k}{2^n}\}} + (+\infty)\, \mathbf{1}_{\{T=\infty\}}.$$

We have seen that (T_n) is a non-increasing sequence of stopping times converging pointwise to T. For any n, $(X_{\frac{i}{2^{n+1}}},\ i \ge 0)$ is a discrete-time $(\mathscr{F}_{\frac{i}{2^{n+1}}})$-martingale and is uniformly integrable[3]. The optional sampling theorem (for discrete-time $(\mathscr{F}_{\frac{i}{2^{n+1}}})$-martingales) allows us to see that $\mathbb{E}(X_{T_n} \,|\, \mathscr{F}_{T_{n+1}}) = X_{T_{n+1}}$ and $\mathbb{E}(X_{T_n}) = \mathbb{E}(X_0)$. Let $Y_k := X_{T_{-k}}$, $k \le 0$. Then $(Y_k,\ k \le 0)$ is a discrete-time backward martingale, indexed by $\{0, -1, -2, \ldots\}$, so $Y_k \to Y_{-\infty}$ a.s. and in L^1 when $k \to -\infty$. On the other hand, the right-continuity of the trajectories implies that $Y_k \to X_T$ a.s. Hence $X_T = Y_{-\infty}$ is integrable.

(ii) Let (S_n) be the non-increasing sequence of stopping times associated with S. We also have $X_{S_n} \to X_S$ a.s. and in L^1. Since $S_n \le T_n$, the optional sampling theorem (for discrete-time $(\mathscr{F}_{\frac{i}{2^n}})$-martingales) says $X_{S_n} = \mathbb{E}[\, X_{T_n} \,|\, \mathscr{F}_{S_n}\,]$; thus for all $A \in \mathscr{F}_S \subset \mathscr{F}_{S_n}$, $\mathbb{E}[\, X_{S_n} \mathbf{1}_A\,] = \mathbb{E}[\, X_{T_n} \mathbf{1}_A\,]$. Letting $n \to \infty$ and taking the L^1 convergence into account, we obtain: $\mathbb{E}[\, X_S \mathbf{1}_A\,] = \mathbb{E}[\, X_T \mathbf{1}_A\,]$. Since $A \in \mathscr{F}_S$ is arbitrary, we have $\mathbb{E}[\, X_S \,|\, \mathscr{F}_S\,] = \mathbb{E}[\, X_T \,|\, \mathscr{F}_S\,]$ a.s. Remembering that X_S is $\mathscr{F}_S$-measurable, this yields (1.1). □

Theorem 1.16. (Doob's optional sampling theorem). *Let $(X_t,\ t \ge 0)$ be a right-continuous martingale and let $S \le T$ be stopping times. If T is* **bounded**, *then*

$$\mathbb{E}(X_T \,|\, \mathscr{F}_S) = X_S \qquad \text{a.s.}$$

In particular, $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ for any **bounded** *stopping time T.*

Proof. Let $a > 0$ be such that $S \le T \le a$. The proof of Theorem 1.15 remains valid. The only problem being the uniform integrability of the discrete-time $(\mathscr{F}_{\frac{i}{2^{n+1}}})$-martingale $(X_{\frac{i}{2^{n+1}}},\ i \ge 0)$, it suffices to consider the discrete martingale $(X_{\frac{i}{2^{n+1}} \wedge a},\ i \ge 0)$, which, for any fixed n, is uniformly integrable, being a finite collection of integrable random variables. □

2. Brownian motion as a martingale

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a complete probability space. Let B be one-dimensional Brownian motion with $B_0 = 0$. Let $(\mathscr{F}_t)$ be (the completion of) the canonical filtration of B.

[3] In particular, X_∞ is a.s. well defined.

Exercise 2.1. (Wald identities). Let T be a stopping time such that $\mathbb{E}(T) < \infty$. Then $\mathbb{E}(B_T) = 0$ and $\mathbb{E}(B_T^2) = \mathbb{E}(T)$.

Example 2.2. Let $T_a := \inf\{t \ge 0 : \ B_t = a\}$. Let $\theta > 0$. We know that $(\mathrm{e}^{\theta B_t - \frac{\theta^2}{2} t}, \ t \ge 0)$ is a martingale. If $a > 0$, then $\theta B_{t\wedge T_a} - \frac{\theta^2}{2}(t \wedge T_a) \le \theta a$, so $(M_t := \mathrm{e}^{\theta B_{t\wedge T_a} - \frac{\theta^2}{2}(t\wedge T_a)}, \ t \ge 0)$ is a continuous and bounded martingale, thus uniformly integrable, and $M_\infty = \mathrm{e}^{\theta a - \frac{\theta^2}{2} T_a}$ (recalling that $T_a < \infty$ a.s.). We have, by the optional sampling theorem,

$$1 = \mathbb{E}\Big[\mathrm{e}^{\theta a - \frac{\theta^2}{2} T_a}\Big].$$

In other words, $\mathbb{E}[\mathrm{e}^{-\frac{\theta^2}{2} T_a}] = \mathrm{e}^{-\theta a}$. (In particular, $\mathbb{P}(T_a < \infty) = 1$.)

In general, for all $a \in \mathbb{R}$, the Laplace transform of T_a is given by $\mathbb{E}[\mathrm{e}^{-\frac{\theta^2}{2} T_a}] = \mathrm{e}^{-\theta |a|}$, $\theta \ge 0$, or, equivalently,

$$\mathbb{E}\Big[\mathrm{e}^{-\lambda T_a}\Big] = \mathrm{e}^{-|a|\sqrt{2\lambda}}, \qquad \lambda \ge 0.$$

One easily checks that this is in agreement with the density function of T_a (for $a \ne 0$) given in the previous chapter. □

Example 2.3. Let $((X_t, Y_t), \ t \ge 0)$ be $\mathbb{R}^2$-valued Brownian motion with $(X_0, Y_0) = (0, 1)$. Let $T := \inf\{t \ge 0 : \ Y_t = 0\}$. What is the law of X_T?

By the previous example, we have, for all $\theta \ge 0$, $\mathbb{E}[\mathrm{e}^{-\frac{\theta^2}{2} T}] = \mathrm{e}^{-\theta}$. Since T is independent of $\sigma(X_t, \ t \ge 0)$, we obtain: for all $a \in \mathbb{R}$,

$$\mathbb{E}[\,\mathrm{e}^{iaX_T}] = \mathbb{E}[\,\mathrm{e}^{ia\sqrt{T}\,X_1}] = \mathbb{E}[\,\mathrm{e}^{-\frac{a^2}{2} T}] = \mathrm{e}^{-|a|}.$$

In other words, X_T has the standard Cauchy distribution. □

Example 2.4. Let $a > 0$ and $b > 0$, and let $T_{a,b} := \inf\{t \ge 0 : \ B_t = -a \text{ ou } B_t = b\} = T_{-a} \wedge T_b$, which is a stopping time. We are interested in the law of $T_{a,b}$.

There are different approaches to determine this law. For example, one can use the strong Markov property. We use here the optional sampling theorem.

Let $\theta \in \mathbb{R}$, and consider the following continuous martingale:

$$M_t := \sinh(\theta(B_t + a))\,\mathrm{e}^{-\frac{\theta^2}{2} t}\,.$$

Since $(M_{t\wedge T_{a,b}},\ t\ge 0)$ is a continuous and bounded martingale, with its a.s. limit (when $t\to\infty$) $\sinh(\theta(B_{T_{a,b}}+a))\,\mathrm{e}^{-\frac{\theta^2}{2}T_{a,b}}$, the optional sampling theorem says that

$$\begin{aligned}\sinh(\theta a) &= \mathbb{E}\Big[\sinh(\theta(B_{T_{a,b}}+a))\,\mathrm{e}^{-\frac{\theta^2}{2}\,T_{a,b}}\Big]\\ &= \sinh(\theta(a+b))\mathbb{E}\Big[\mathrm{e}^{-\frac{\theta^2}{2}\,T_b}\,\mathbf{1}_{\{T_b<T_{-a}\}}\Big].\end{aligned}$$

Hence

$$\mathbb{E}\Big[\mathrm{e}^{-\frac{\theta^2}{2}\,T_b}\,\mathbf{1}_{\{T_b<T_{-a}\}}\Big]=\frac{\sinh(\theta a)}{\sinh(\theta(a+b))}. \tag{2.1}$$

Exchanging the roles of a and b (which amounts to replacing B by $-B$), we also obtain:

$$\mathbb{E}\Big[\mathrm{e}^{-\frac{\theta^2}{2}\,T_{-a}}\,\mathbf{1}_{\{T_b>T_{-a}\}}\Big]=\frac{\sinh(\theta b)}{\sinh(\theta(a+b))}. \tag{2.2}$$

So

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{-\frac{\theta^2}{2}\,T_{a,b}}] &= \mathbb{E}\Big[\mathrm{e}^{-\frac{\theta^2}{2}\,T_b}\,\mathbf{1}_{\{T_b<T_{-a}\}}\Big]+\mathbb{E}\Big[\mathrm{e}^{-\frac{\theta^2}{2}\,T_{-a}}\,\mathbf{1}_{\{T_b>T_{-a}\}}\Big]\\ &= \frac{\sinh(\theta a)+\sinh(\theta b)}{\sinh(\theta(a+b))}\\ &= \frac{\cosh(\frac{\theta(a-b)}{2})}{\cosh(\frac{\theta(a+b)}{2})},\end{aligned}$$

i.e.,

$$\mathbb{E}\Big[\mathrm{e}^{-\lambda\,T_{a,b}}\Big]=\frac{\cosh(\frac{a-b}{2}\sqrt{2\lambda}\,)}{\cosh(\frac{a+b}{2}\sqrt{2\lambda}\,)},\qquad \lambda\ge 0.$$

One can also send $\theta\to 0$ in (2.1) and (2.2) to get

$$\mathbb{P}(T_b<T_{-a})=\frac{a}{a+b},\qquad \mathbb{P}(T_b>T_{-a})=\frac{b}{a+b}.$$

This gives us the law of $\sup_{0\le t\le T_{-1}}B_t$: for all $x>0$,

$$\mathbb{P}\Big(\sup_{0\le t\le T_{-1}}B_t\ge x\Big)=\mathbb{P}(T_x<T_{-1})=\frac{1}{1+x}.$$

In words, $\sup_{0\le t\le T_{-1}}B_t$ has the same law as $\frac{1-U}{U}$, where U is uniformly distributed in $(0,\,1)$.

A special case is $a=b$. Let $T_a^*:=\inf\{t\ge 0:\ |B_t|=a\}$. Then

$$\mathbb{E}[\mathrm{e}^{-\frac{\theta^2}{2}\,T_a^*}]=\frac{1}{\cosh(\theta a)}.$$

[In the literature, it is known that $\mathbb{E}[\exp(-\frac{\theta^2}{2}\int_0^a B_s^2\,\mathrm{d}s)]=\frac{1}{\sqrt{\cosh(\theta a)}}$. As a consequence, for any fixed $a\ge 0$, T_a^* has the same distribution as $\int_0^a X_s^2\,\mathrm{d}s+\int_0^a Y_s^2\,\mathrm{d}s$, where X and Y are independent Brownian motion.] □

Chapter 7

Further properties of Brownian motion

We describe a few other properties of Brownian motion. The content of this chapter is **not** part of the program of the examination.

1. Law of the iterated logarithm

It is trivial that, for $t \to \infty$, $\frac{B_t}{t^{1/2}}$ converges in distribution to the Gaussian $\mathscr{N}(0, 1)$ law; indeed, the law of $\frac{B_t}{t^{1/2}}$ even does not depend on t. When $t \to \infty$ along natural numbers, this is also a consequence of the central limit theorem, and it is known that in the central limit theorem, convergence in distribution cannot be replaced by convergence in probability, and a fortiori, not by a.s. convergence. In the exercise class, however, we have proved the following law of the iterated logarithm:

$$\limsup_{t\to\infty} \frac{B_t}{(2t \log\log t)^{1/2}} = 1 \qquad \text{a.s.}$$

A simple argument tells us that in the law of the iterated logarithm, B_t can be replaced by $|B_t|$, $\sup_{s\in[0,\,t]} B_s$, or $\sup_{s\in[0,\,t]} |B_s|$. So for example, we also have

$$\limsup_{t\to\infty} \frac{1}{(2t \log\log t)^{1/2}} \Big(\sup_{s\in[0,\,t]} B_s - \inf_{s\in[0,\,t]} B_s \Big) = 1 \qquad \text{a.s.}$$

The law of the iterated logarithm is not as precise as it looks like. For example, it does not tell us about the probability that the curve $(2t \log\log t)^{1/2}$ is crossed by Brownian motion infinitely often.[1] The question is answered by the following integral criterion of Kolmogorov

[1] We should be careful about what we mean by "infinitely often": there exists an increasing sequence $t_n \uparrow \infty$ such that $B_{t_n} \geq (2t_n \log\log t_n)^{1/2}$ for all n.

(also referred to sometimes as the Erdős–Feller–Kolmogorov–Petrovski or EFKP integral test): if $h : \mathbb{R} \to \mathbb{R}_+$ is non-decreasing, then

$$\mathbb{P}\Big(B_t \geq t^{1/2}h(t), \text{ infinitely often}\Big) = 0 \text{ or } 1,$$

according as whether $\int_1^\infty \frac{h(t)}{t}\,\mathrm{e}^{-\frac{1}{2}h^2(t)}\,\mathrm{d}t$ converges or diverges. In particular, we see that the curve $(2t\log\log t)^{1/2}$ is crossed by Brownian motion infinitely often.

Another question is about the lower limits of Brownian motion: we trivially have

$$\liminf_{t\to\infty} \frac{B_t}{(2t\log\log t)^{1/2}} = -1 \qquad \text{a.s.}$$

Chung (1948) gives a non trivial result for the lower limits of Brownian motion

$$\liminf_{t\to\infty} \frac{(\log\log t)^{1/2}}{t^{1/2}} \sup_{s\in[0,\,t]} |B_s| = \frac{\pi}{8^{1/2}} \qquad \text{a.s.}$$

Feller (1951) studies the lower limits of the range:

$$\liminf_{t\to\infty} \frac{(\log\log t)^{1/2}}{t^{1/2}} \Big(\sup_{s\in[0,\,t]} B_s - \inf_{s\in[0,\,t]} B_s\Big) = \frac{\pi}{2^{1/2}} \qquad \text{a.s.}$$

What happens if we replace $\sup_{s\in[0,\,t]} |B_s|$ by $\sup_{s\in[0,\,t]} B_s$? Intuitively, as far as the lower limits are concerned, it is clear that $\sup_{s\in[0,\,t]} B_s$ can be far smaller than $\sup_{s\in[0,\,t]} |B_s|$: this is confirmed by Hirsch (1965): a.s.,

$$\liminf_{t\to\infty} \frac{(\log t)^a}{t^{1/2}} \sup_{s\in[0,\,t]} B_s$$

is either 0 if $a \leq 1$, or infinite if $a > 1$.

2. Modulus of continuity

Paul Lévy established the following modulus of continuity for Brownian motion:

$$\lim_{h\to 0} \frac{\sup_{t\in[0,\,1-h]} |B_{t+h} - B_t|}{[2h\log(\frac{1}{h})]^{1/2}} = 1 \qquad \text{a.s.}$$

$$\lim_{h\to 0} \frac{\sup_{t\in[0,\,1-h]} \sup_{s\in[0,\,h]} |B_{t+s} - B_t|}{[2h\log(\frac{1}{h})]^{1/2}} = 1 \qquad \text{a.s.}$$

It is a simple consequence of the following probabilistic estimate: for any $\varepsilon > 0$, there exists a constant $C(\varepsilon) > 0$ such that for all $x > 0$ and $0 < h < 1$,

$$\mathbb{P}\Big(\sup_{ts\in[0,\,1-h]} \sup_{s\in[0,\,h]} |B_{t+s} - B_t| > h^{1/2}x\Big) \leq \frac{C(\varepsilon)}{h}\,\mathrm{e}^{-\frac{x^2}{2+\varepsilon}}\,.$$

In the literature, the size of exceptional sets such as $\{t \in [0, 1] : \limsup_{h\to 0} \frac{|B_{t+h}-B_t|}{[2h\log(\frac{1}{h})]^{1/2}} > a\}$ (for some fixed $0 < a < 1$) is studied.

Note that the modulus of continuity has the following lower limits counterpart (Csörgő and Révész 1979), sometimes referred to as the modulus of non differentiability:

$$\lim_{h\to 0} \frac{[\log\log(\frac{1}{h})]^{1/2}}{h^{1/2}} \inf_{t\in[0,\,1-h]} \sup_{s\in[0,\,h]} |B_{t+s} - B_t| = \frac{\pi}{8^{1/2}} \qquad \text{a.s.}$$

3. Donsker's theorem

Let $(\xi_i,\, i \ge 1)$ be a sequence of i.i.d. random variables with $\mathbb{E}(\xi_1) = 0$ and $0 < \sigma^2 := \mathbb{E}(\xi_1^2) < \infty$. Let $S_0 := 0$ and $S_n := \xi_1 + \cdots + \xi_n$, $n \ge 1$. Define

$$X_t^{(n)} := \frac{S_{\lfloor nt\rfloor}}{\sigma\, n^{1/2}} + (nt - \lfloor nt\rfloor)\frac{1}{\sigma n^{1/2}}\xi_{\lfloor nt\rfloor+1}, \qquad t \in [0,\, 1].$$

We view $(X^{(n)},\, n \ge 1)$ as a sequence of random variables taking values in $C([0,\,1],\,\mathbb{R})$. The Donsker invariance principle says that $(X^{(n)},\, n \ge 1)$ converges weakly in $C([0,\,1],\,\mathbb{R})$ to $(B_t,\, n \ge 1)$. By weak convergence $X_n \to X$ in a metric space E, we means that $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all continuous and bounded functions $f : E \to \mathbb{R}$. We have so far been familiar with the notion of weak convergence for random variables taking values in a Euclidian space.

In $C([0,\,1],\,\mathbb{R})$, weak convergence implies (but is not equivalent to) that for all $p \ge 1$ and all $0 \le t_1 < t_2 < \cdots < t_m \le 1$, $(X_{t_1}^{(n)}, \cdots, X_{t_p}^{(n)}) \to (B_{t_1}, \cdots, B_{t_p})$ in law. The latter can be trivially checked by means of the central limit theorem.

It is easily checked that if $X_n \to X$ weakly in E, and if $g : E \to \widetilde{E}$ is a.s. continuous (where $\widetilde{E}$ is a metric space), then $g(X_n) \to g(X)$ weakly in $\widetilde{E}$. As such, the Donsker invariance principle has many simple applications. For example, taking $g : C([0,\,1],\,\mathbb{R}) \to \mathbb{R}$ to be $g(x) := x(1)$ (for $x \in C([0,\,1],\,\mathbb{R})$), we recover the usual central limit theorem. We can also take $g : C([0,\,1],\,\mathbb{R}) \to \mathbb{R}^2$ to be $g(x) := (\sup_{t\in[0,\,1]} x(t),\, x(1))$, we see that $\frac{1}{\sigma\, n^{1/2}}(\max_{0\le i\le n} S_i,\, S_n) \to (\sup_{s\in[0,\,1]} B_s,\, B_1)$ in law. In particular, $\frac{1}{\sigma\, n^{1/2}} \max_{0\le i\le n} S_i \to \mathscr{N}(0,\,1)$ in distribution. Another example is that $\frac{1}{\sigma\, n^{1/2}} \sum_{i=0}^n S_i \to \int_0^1 B_s\,\mathrm{d}s$ in distribution, and we have seen that the limit law $\int_0^1 B_s\,\mathrm{d}s$ is Gaussian $\mathscr{N}(0,\,\frac{1}{3})$.

It is possible to strengthen the Donsker invariance principle into an almost sure invariance principle, if we are allowed to redefine the random variables. Komlós, Major and Tusnády (KMT) in 1975 prove that there exist a probability space $(\widetilde{\Omega},\, \widetilde{\mathscr{F}},\, \widetilde{\mathbb{P}})$, an i.i.d. sequence $(\widetilde{\xi}_i,\, i \ge 1)$ and Brownian motion $(\widetilde{B}_t,\, t \in [0,\,1])$ on this space, with $\widetilde{\xi}_1$ under $\widetilde{\mathbb{P}}$ having the same law of ξ_1 under $\mathbb{P}$, such that when $n \to \infty$,

$$\max_{0\le i\le n} |\widetilde{S}_i - \sigma\, \widetilde{B}_i| = O(\log n), \qquad \widetilde{\mathbb{P}}\text{-a.s.},$$

where $\widetilde{S}_i$ is defined in terms of $(\widetilde{\xi}_i,\, i \ge 1)$ exactly as S_i in terms of $(\xi_i,\, i \ge 1)$.

The KMT strong invariance principle allows us to conclude that

$$\limsup_{n\to\infty} \frac{S_n}{(2n \log\log n)^{1/2}} = 1, \qquad \text{a.s.},$$

the latter known as the Hartman–Wintner law of the iterated logarithm. It also yields the corresponding EFKP integral test for (S_n), originally proved by Erdős for sums of Bernoulli random variables and by Kolmogorov for Brownian motion.

4. Variations of Brownian motion

Let $B = (B_t,\, t \in [0,\, 1])$ be Brownian motion.

Theorem 4.1. (Lévy). *Fix $t > 0$. Let $\Pi := \{0 = t_0 < t_1 < \cdots < t_p = t\}$ be a sequence of subdivisions of $[0,\, t]$. Write $\|\Pi\| := \max_{1\le i\le p}(t_i - t_{i-1})$. Then*

$$\lim_{\|\Pi\|\to 0} \sum_{i=1}^{p} (B_{t_i} - B_{t_{i-1}})^2 = t, \qquad \textit{in } L^2(\mathbb{P}).$$

If, moreover, $\Pi_1 \subset \Pi_2 \subset \cdots$, then we also have almost sure convergence.

Proof. Let us first prove L^2 convergence. Define

$$Y_i := (B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i-1}), \qquad 1 \le i \le p\,.$$

Then $(Y_i,\, 1 \le i \le p)$ are i.i.d. centered, with (writing $a := t_i - t_{i-1}$) $\mathbb{E}(Y_i^2) = a^2\, \mathbb{E}[(B_1^2 - 1)^2] = a^2(\mathbb{E}(B_1^4) - 1) = 2a^2$. Accordingly,

$$\begin{aligned} \mathbb{E}\Big[\Big(\sum_{i=1}^{p} (B_{t_i} - B_{t_{i-1}})^2 - t\Big)^2\Big] &= \mathbb{E}\Big[\Big(\sum_{i=1}^{p} Y_i\Big)^2\Big] = \sum_{i=1}^{p} \mathrm{Var}(Y_i) \\ &= 2\sum_{i=1}^{p} (t_i - t_{i-1})^2 \\ &\le 2t \max_{1\le i\le p} (t_i - t_{i-1}) \to 0. \end{aligned}$$

This yields $L^2(\mathbb{P})$ convergence.

We prove a.s. convergence only[2] for the special case $t_i = t_i^n := \frac{i}{2^n}$, $0 \le i \le 2^n$. We have seen that

$$\mathbb{E}\Big[\Big(\sum_{i=1}^{2^n} Y_i\Big)^2\Big] \le 2\sum_{i=1}^{2^n}(t_i^n - t_{i-1}^n)^2 = \frac{1}{2^{n-1}}.$$

By Tchebychev's inequality,

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{2^n} Y_i\Big| > \frac{1}{n}\Big) \le \frac{n^2}{2^{n-1}},$$

which is summable in $n \ge 1$. By the Borel–Cantelli lemma, there exists $A \in \mathscr{F}$ with $\mathbb{P}(A) = 1$, such that for all $\omega \in A$, we can find $n_0 = n_0(\omega) < \infty$ satisfying

$$\Big|\sum_{i=1}^{2^n} Y_i\Big| \le \frac{1}{n}, \qquad \forall\, n \ge n_0\,,$$

from which convergence a.s. follows. □

Remark 4.2. An immediate consequence of Theorem 4.1 is that a.s., Brownian motion is of infinite variation on any interval, which we already know. □

By an abuse of language, we state Theorem 4.1 by saying that the quadratic variation of Brownian motion on an interval is the length of the interval. However, the genuine definition of the quadratic variation of B is the limit (if exists in some sense) of $\sum_{i=1}^p (B_{t_i} - B_{t_{i-1}})^2$ when $\|\Pi\| \to 0$, without the restriction $\Pi_1 \subset \Pi_2 \subset \cdots$. Unfortunately, Lévy also shows that[3]

$$\limsup_{\|\Pi\|\to 0} \sum_{i=1}^p (B_{t_i} - B_{t_{i-1}})^2 = \infty, \qquad \text{a.s.},$$

without the assumption that $\Pi_1 \subset \Pi_2 \subset \cdots$

A way to avoid an explosion is to replace the function x^2 by a smaller function. For example, Lévy shows that for any $\alpha > 2$,

$$\sup_{\Pi} \sum_{i=1}^p f(B_{t_i} - B_{t_{i-1}}) < \infty, \qquad \text{a.s.},$$

with $f(x) := |x|^\alpha$. If we use Lévy modulus of continuity for Brownian, we immediately see that this holds with $f(x) := \frac{x^2}{\log^* \frac{1}{|x|}}$, with $\log^* y := \max\{1, \log y\}$. Taylor (1972) shows

[2]The proof of a.s. convergence in the general case is more technical. We refer to Proposition II.2.12 in the book of Revuz and Yor, "*Continuous Martingales and Brownian Motion*" (third edition), Springer, 1999.

[3]See p. 48 of the book of D. Freedman, "*Brownian Motion and Diffusion*", Holden-Day, 1971.

that we can do better: we can take $g(x) := \frac{x^2}{\log^* \log^* \frac{1}{|x|}}$; Taylor (1972) proves the following quantitative result:

$$\lim_{\delta \to 0} \sup_{\|\Pi\| \leq \delta} \sum_{i=1}^{p} g(B_{t_i} - B_{t_{i-1}}) = 2, \qquad \text{a.s.},$$

and that $\sum_{i=1}^{p} f(B_{t_i} - B_{t_{i-1}})$ is unbounded a.s. if f is such that $\frac{f(x)}{g(x)} \to \infty$, $x \to 0$.

5. Multidimensional Brownian motion

Let $(B_t,\, t \geq 0)$ be d-dimensional Brownian motion. The law of the iterated logarithm holds as usual:

$$\limsup_{t \to \infty} \frac{|B_t|}{(2t \log \log t)^{1/2}} = 1\,, \qquad \text{a.s.}$$

However, multidimensional Brownian motion has some special properties that are not shared with one-dimensional Brownian motion. For example, it is known that in dimension $d = 2$, points are polar, in the sense that for any $x \in \mathbb{R}^2 \backslash \{0\}$, $\mathbb{P}(B_t = x, \text{ for some } t \geq 0) = 0$, and that $\mathbb{P}(B_t = 0, \text{ for some } t > 0) = 0$. The latter is a consequence of the former and the Markov property: $\mathbb{P}(B_t = 0, \text{ for some } t > 0) = \lim_{n\to\infty} \mathbb{P}(B_t = 0, \text{ for some } t > \frac{1}{n})$, and for each n, $\mathbb{P}(B_t = 0, \text{ for some } t > \frac{1}{n}) = \mathbb{E}[\mathbb{P}_{B_{\frac{1}{n}}}(B_t = 0, \text{ for some } t \geq 0)] = 0$ because $B_{\frac{1}{n}} \neq 0$ a.s., and for any $x \in \mathbb{R}^2 \backslash \{0\}$, $\mathbb{P}_x(B_t = 0, \text{ for some } t \geq 0) = \mathbb{P}(B_t = -x, \text{ for some } t \geq 0) = 0$.

A fortiori, points are polar for Brownian motion in dimension $d \geq 2$: it suffices to consider the first two coordinates.

However, two-dimensional Brownian motion is neighbourhood-recurrent, in the sense that given any ball $B(x,\, r)$ with $r > 0$ in $\mathbb{R}^2$, there exists a.s. a (random) sequence $(t_n) \uparrow \infty$ such that $B_{t_n} \in B(x,\, r)$, $\forall n$.

The neighbourhood-recurrence is not shared by Brownian motion in space. Let us prove that for $d \geq 3$, $\lim_{t\to\infty} |B_t| = \infty$ a.s.

Let $a \in \mathbb{R}$. We have

$$\mathbb{P}(|B_n| < n^a) = \mathbb{P}(|B_1| < n^{a - \frac{1}{2}}) \leq n^{(a - \frac{1}{2})d}\,,$$

the last inequality being a consequence of the fact that the density of $|\mathscr{N}(0,\, \mathrm{Id})|$ is bounded by 1. So if $a < \frac{1}{2} - \frac{1}{d}$, then $\sum_n \mathbb{P}(|B_n| < n^a) < \infty$. By the Borel–Cantelli lemma, $\liminf_{n\to\infty} \frac{|B_n|}{n^a} \geq 1$ a.s.

On the other hand, we have seen in a previous exercise class that for any $\varepsilon > 0$, $\lim_{n\to\infty} \frac{1}{n^\varepsilon} \sup_{t\in[n,\, n+1]} |B_{t+1} - B_t| = 0$ a.s. [This is a consequence of the reflection principle. It was proved in the exercise class only for one-dimensional Brownian motion, but we see immediately that it holds in any dimension, by considering each coordinate of Brownian motion.] We therefore can remove the restriction that $t \to \infty$ along integers, and conclude that for any $0 < a < \frac{1}{2} - \frac{1}{d}$ (which is possible if $d \geq 3$), $\liminf_{t\to\infty} \frac{|B_t|}{t^a} \geq 1$ a.s. Since $a \in (0, \frac{1}{2} - \frac{1}{d})$ is arbitrary, it yields that for all $a \in (0, \frac{1}{2} - \frac{1}{d})$, $\lim_{t\to\infty} \frac{|B_t|}{t^a} = \infty$ a.s. In particular, $|B_t| \to \infty$ a.s.

We can improve our argument and show that for all $a \in (0, \frac{1}{2})$, $\lim_{t\to\infty} \frac{|B_t|}{t^a} = \infty$ a.s. Dvoretzky and Erdős (1951) prove that the following is true: a.s., $\liminf_{t\to\infty} \frac{|B_t|}{t^{1/2}/(\log t)^a}$ is ∞ if $a > \frac{1}{d-2}$, and is 0 otherwise. [They actually have obtained an integral criterion.]

Since $|B_t| \to \infty$ $(t \to \infty)$, we can define the process of future infima: $J_t := \inf_{s\geq t} |B_s|$, $t \geq 0$. The lower limits of J_t are identical to those of $|B_t|$: you can replace $|B_t|$ by J_t in the result of Dvoretzky and Erdős. Concerning the upper limits of J_t, Erdős and Taylor (1962) prove that

$$\limsup_{t\to\infty} \frac{J_t}{(2t \log\log t)^{1/2}} = 1\,, \qquad \text{a.s.}$$

[Are you surprised? What would you do if you wanted to distinguish the upper limits of J_t from those of $|B_t|$?]

6. For further reading

We have only mentioned a few elementary properties of Brownian motion. For further reading: the book of Revuz and Yor: "*Continuous Martingales and Brownian Motion*" (third edition), Springer, 1999. For fractal nature of Brownian motion, see the book of Mörters and Peres: "*Brownian Motion*", Cambridge, 2010.

Tsinghua University *2017–2018*

"Advanced Probability" (Part III: Brownian motion)

Exercise sheet #III.1:

Construction of Brownian motion

Exercice 1. Let ξ be a Gaussian $\mathscr{N}(0, 1)$ random variable. Let $x > 0$.

(i) Prove that $\frac{1}{(2\pi)^{1/2}}\,(\frac{1}{x} - \frac{1}{x^3})\,\mathrm{e}^{-x^2/2} \le \mathbb{P}(\xi > x) \le \frac{1}{(2\pi)^{1/2}}\,\frac{1}{x}\,\mathrm{e}^{-x^2/2}$.

(ii) Prove that[1] $\mathbb{P}(\xi > x) \le \mathrm{e}^{-x^2/2}$.

Exercice 2. Let ξ be a Gaussian $\mathscr{N}(0, 1)$ random variable.

(i) Compute $\mathbb{E}(\xi^4)$ and $\mathbb{E}(|\xi|)$.

(ii) Compute $\mathbb{E}(\mathrm{e}^{a\xi})$, $\mathbb{E}(\xi\mathrm{e}^{a\xi})$ and $\mathbb{E}(\mathrm{e}^{a\xi^2})$, with $a \in \mathbb{R}$.

(iii) Let $b \ge 0$. Let η be a Gaussian $\mathscr{N}(0, 1)$ random variable, independent of ξ. Prove that $\mathbb{E}(\mathrm{e}^{b\xi^2}) = \mathbb{E}(\mathrm{e}^{\lambda\xi\eta})$, where $\lambda := (2b)^{1/2}$.

Exercice 3. Let ξ, ξ_1, ξ_2, $\cdots$ be real-valued random variables. Assume that for each n, ξ_n is Gaussian $\mathscr{N}(\mu_n, \sigma_n^2)$, with $\mu_n \in \mathbb{R}$ and $\sigma_n \ge 0$, and that $\xi_n \to \xi$ in law. Prove that ξ is Gaussian.

Exercice 4. Let ξ, ξ_1, ξ_2, $\cdots$ be random variables. Assume that for any n, ξ_n is Gaussian $\mathscr{N}(\mu_n, \sigma_n^2)$, where $\mu_n \in \mathbb{R}$ and $\sigma_n \ge 0$, and that $\xi_n \to \xi$ in probability. Prove that ξ_n converges in L^p, for all $p \in [1, \infty)$.

Exercice 5. Let (ξ, η, θ) be an $\mathbb{R}^3$-valued Gaussian random vector. Assume $\mathbb{E}(\xi) = \mathbb{E}(\eta) = \mathbb{E}(\xi\eta) = 0$, $\sigma_\xi^2 := \mathbb{E}(\xi^2) > 0$ and $\sigma_\eta^2 := \mathbb{E}(\eta^2) > 0$.

(i) Prove that $\mathbb{E}(\theta \,|\, \xi, \eta) = \mathbb{E}(\theta \,|\, \xi) + \mathbb{E}(\theta \,|\, \eta) - \mathbb{E}(\theta)$.

(ii) Prove that $\mathbb{E}(\xi \,|\, \xi\eta) = 0$.

(iii) Prove that $\mathbb{E}(\theta \,|\, \xi\eta) = \mathbb{E}(\theta)$.

Exercice 6. Let $(\xi_{k,n},\, k \ge 0,\, n \ge 0)$ be a collection of i.i.d. Gaussian $\mathscr{N}(0, 1)$ random variables. For all $n \ge 0$, we define the process $(X_n(t),\, t \in [0, 1])$ with $t \mapsto X_n(t)$ being affine on each of the intervals $[\frac{i}{2^n}, \frac{i+1}{2^n}]$, $0 \le i \le 2^n - 1$, in the following way $X_0(0) := 0$, $X_0(1) := \xi_{0,0}$, and by induction, for $n \ge 1$,

$$\begin{aligned} X_n(\frac{2i}{2^n}) &:= X_{n-1}(\frac{2i}{2^n}), \qquad 0 \le i \le 2^{n-1}, \\ X_n(\frac{2j+1}{2^n}) &:= X_{n-1}(\frac{2j+1}{2^n}) + \frac{\xi_{2j+1,n}}{2^{(n+1)/2}}, \qquad 0 \le j \le 2^{n-1} - 1. \end{aligned}$$

[1] We will see that $\mathbb{P}(\xi > x) \le \frac{1}{2}\mathrm{e}^{-x^2/2}$.

Prove that for all $n \geq 0$, $(X_n(\frac{k}{2^n}),\, 0 \leq k \leq 2^n)$ is a centered Gaussian vector such that $\mathbb{E}[X_n(\frac{k}{2^n})X_n(\frac{\ell}{2^n})] = \frac{k}{2^n} \wedge \frac{\ell}{2^n}$, for $0 \leq k,\, \ell \leq 2^n$.

Exercice 7. Let $(B_t^m,\, t \in [0,\, 1])$, for $m \geq 0$, be a sequence of independent Brownian motions defined on $[0,\, 1]$. Let

$$B_t := B_{t-\lfloor t \rfloor}^{\lfloor t \rfloor} + \sum_{0 \leq m < \lfloor t \rfloor} B_1^m, \qquad t \geq 0.$$

Prove that $(B_t,\, t \geq 0)$ is Brownian motion.

Exercice 8. Prove that $\mathscr{C}(\mathbb{R}_+,\, \mathbb{R})$, the Borel σ-field of $C(\mathbb{R}_+,\, \mathbb{R})$, coincides with $\sigma(X_t,\, t \geq 0)$, the σ-field generated by the process of projections $(X_t,\, t \geq 0)$.

Exercice 9. Let $T := \inf\{t \geq 0 :\, B_t = 1\}$ (with $\inf \varnothing := \infty$). Prove that[2] $\mathbb{P}(T < \infty) \geq \frac{1}{2}$.

Exercice 10. **(i)** Prove that $(-B_t,\, t \geq 0)$ is Brownian motion.

(ii) **(Scaling)** Prove that for any $a > 0$, $(\frac{1}{a^{1/2}}\, B_{at},\, t \geq 0)$ is Brownian motion.

Exercice 11. **(i)** Let $\xi := \int_0^1 B_t\, \mathrm{d}t$. Determine the law of ξ.

(ii) Let $\eta := \int_0^2 B_t\, \mathrm{d}t$. Determine $\mathbb{E}(B_1 \,|\, \eta)$.

(iii) Prove that $B_7 - B_2$ is independent of $\sigma(B_s,\, s \in [0,\, 1])$.

(iv) Let $\mathscr{F}_1 := \sigma(B_s,\, s \in [0,\, 1])$. Determine $\mathbb{E}(B_5 \,|\, \mathscr{F}_1)$ and $\mathbb{E}(B_5^2 \,|\, \mathscr{F}_1)$.

Exercice 12. **(i)** Prove or disprove: for all $t > 0$, $\int_0^t B_s^2\, \mathrm{d}s$ has the same distribution as $t^2 \int_0^1 B_s^2\, \mathrm{d}s$.

(ii) Prove or disprove: the processes $(\int_0^t B_s^2\, \mathrm{d}s,\, t \geq 0)$ and $(t^2 \int_0^1 B_s^2\, \mathrm{d}s,\, t \geq 0)$ have the same distribution.

Exercice 13. Let T be a random variable having the exponential law of parameter 1, independent of B. Determine the law of B_T.

Exercice 14. **(i)** Prove that $\int_0^1 \frac{B_s}{s}\, \mathrm{d}s$ is a.s. well defined.

(ii) Let $\beta_t := B_t - \int_0^t \frac{B_s}{s}\, \mathrm{d}s$. Prove that $(\beta_t,\, t \geq 0)$ is Brownian motion.

Exercice 15. Prove that $\int_0^\infty |B_s|\, \mathrm{d}s = \infty$ a.s.

Exercice 16. Let $T := \inf\{t \geq 0 :\, |B_t| = 1\}$ (with $\inf \varnothing := \infty$).

(i) Prove that $T < \infty$ a.s.

(ii) Prove that T and $\mathbf{1}_{\{B_T = 1\}}$ are independent.

[2] Later on, we will see that $T < \infty$ a.s.

Exercice 17. Let $B := (B_t,\, t \in [0,\, 1])$ be Brownian motion defined on $[0,\, 1]$. For all $t \in [0,\, 1]$, let

$$\begin{aligned}\mathscr{F}_t &:= \sigma(B_s,\ s \in [0,\, t]),\\ \mathscr{G}_t &:= \mathscr{F}_t \vee \sigma(B_1) = \sigma(\{C;\ C \in \mathscr{F}_t \text{ or } C \in \sigma(B_1)\}).\end{aligned}$$

(i) Let $0 \le s < t \le 1$. Prove that

$$\mathbb{E}[(B_t - B_s) \,|\, \mathscr{G}_s] = \frac{t-s}{1-s}\,(B_1 - B_s).$$

(ii) Consider the process $\beta := (\beta_t,\, t \in [0,\, 1])$ defined by

$$\beta_t := B_t - \int_0^t \frac{B_1 - B_s}{1-s}\,\mathrm{d}s, \qquad t \in [0,\, 1].$$

Prove that for $0 \le s < t \le 1$, $\mathbb{E}(\beta_t \,|\, \mathscr{G}_s) = \beta_s$ a.s.

Exercice 18. Let $\mathscr{F}_1 := \sigma(B_s,\ s \in [0,\, 1])$, and let $a \in \mathbb{R}$. Let $\mathbb{Q}$ be the probability measure on $\mathscr{F}_1$ defined by $\mathbb{Q}(A) := \mathbb{E}(\mathrm{e}^{aB_1 - \frac{a^2}{2}}\,\mathbf{1}_A)$, $A \in \mathscr{F}_1$. Define $\gamma_t := B_t - at$, $t \in [0,\, 1]$. Prove that $(\gamma_t,\, t \in [0,\, 1])$ is Brownian motion under $\mathbb{Q}$.

Tsinghua University *2017–2018*

"Advanced Probability" (Part III: Brownian motion)

Exercise sheet #III.2:
Brownian motion and the Markov property

Exercice 1. Let $\mathscr{A}_1 \subset \mathscr{F}, \cdots, \mathscr{A}_n \subset \mathscr{F}$ be π-systems, satisfying $\Omega \in \mathscr{A}_i$, $\forall i$. Assume

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n), \qquad \forall A_i \in \mathscr{A}_i.$$

Then $\sigma(\mathscr{A}_1), \cdots, \sigma(\mathscr{A}_n)$ are independent.

Exercice 2. **(i) (Time reversal)** Fix $a > 0$. Prove that $(B_a - B_{a-t},\, t \in [0,\, a])$ is Brownian motion on $[0,\, a]$.

(ii) (Time inversion) Prove that $X := (X_t,\, t \ge 0)$ defined by $X_t := t\, B_{\frac{1}{t}}$ (for $t > 0$) and $X_0 := 0$ is Brownian motion.

Exercice 3. Prove that there exists a constant $a > 0$ (that does not depend on ω) such that $\inf_{t\in[0,\,2]} B_t$ has the same distribution as $a \inf_{t\in[0,\,1]} B_t$.

Exercice 4. **(Brownian bridge)** Let $b_t = B_t - tB_1$, $t \in [0,\, 1]$. It is a centered Gaussian process with a.s. continuous trajectories and with covariance $(s \wedge t) - st$. We call b a Brownian bridge.

(i) The process $(b_t,\, t \in [0,\, 1])$ is independent of the random variable B_1.

(ii) If b is a Brownian bridge, so is $(b_{1-t},\, t \in [0,\, 1])$.

(iii) If b is a Brownian bridge, then $B_t = (1 + t)b_{t/(1+t)}$, $t \ge 0$, is Brownian motion. Note that $b_t = (1 - t)B_{t/(1-t)}$.

Exercice 5. Prove that

$$\lim_{t\to\infty} \frac{B_t}{t} = 0, \qquad \text{a.s.}$$

Hint: Use time inversion.

Exercice 6. Let $(t_n)_{n\ge1}$ be a sequence of positive real numbers decreasing towards 0. Prove that a.s., $B_{t_n} > 0$ for infinitely many n, and $B_{t_n} < 0$ for infinitely many n.

Exercice 7. Prove that when $t \to \infty$, $(\int_0^t \mathrm{e}^{B_s}\, \mathrm{d}s)^{1/t^{1/2}} \to \mathrm{e}^{|N|}$ in law, where N is a Gaussian $\mathscr{N}(0,\, 1)$ random variable.

Exercice 8. **(i)** Prove that $0 < \sup_{t\geq 0}(|B_t| - t) < \infty$ a.s. and that $0 < \sup_{t\geq 0} \frac{|B_t|}{1+t} < \infty$ a.s.

(ii) Prove that $\sup_{t\geq 0}(|B_t| - t)$ and $(\sup_{t\geq 0} \frac{|B_t|}{1+t})^2$ have the same distribution.

Hint: Use the scaling property.

(iii) Prove that for any $p > 0$, $\mathbb{E}\{[\sup_{t\geq 0}(|B_t| - t)]^p\} < \infty$.

(iv) Prove that there exists a constant $C < \infty$ such that for any non-negative random variable T (not necessarily a stopping time!), $\mathbb{E}(|B_T|) \leq C\,[\mathbb{E}(T)]^{1/2}$.

Hint: Write, for any $a > 0$, $|B_T| = (|B_T| - aT) + aT$, and prove that $\mathbb{E}(|B_T| - aT) \leq \frac{1}{a}\,\mathbb{E}[\sup_{t\geq 0}(|B_t| - t)]$.

Exercice 9. Let $S_t := \sup_{s\in[0,t]} B_s$, $t \geq 0$. Prove that $S_2 - S_1$ is distributed as $\max\{|N| - |\widetilde{N}|, 0\}$, where N and $\widetilde{N}$ are independant Gaussian $\mathscr{N}(0, 1)$ random variables.

Exercice 10. Let $d_1 := \inf\{t \geq 1 : B_t = 0\}$ and $g_1 := \sup\{t \leq 1 : B_t = 0\}$.

(i) Is d_1 a stopping time?

(ii) Determine the law of d_1, and the law of g_1.

Exercice 11. Define $T_1 := \inf\{t > 0 : B_t = 1\}$ and $\tau := \inf\{t \geq T_1 : B_t = 0\}$.

(i) Is τ a stopping time?

(ii) Determine the law of τ.

Exercice 12. **(i)** Study convergence in probability of $\frac{\log(1+B_t^2)}{\log t}$ (quand $t \to \infty$).

(ii) Study a.s. convergence of $\frac{\log(1+B_t^2)}{\log t}$.

Exercice 13. Prove, *without using inversion of time* (but using instead the law of large numbers and the reflection principle), that $\frac{B_t}{t} \to 0$ a.s. when $t \to \infty$.

Exercice 14. The aim of this exercise is to prove $T < \infty$ a.s., where $T := \inf\{t \geq 0 : B_t = (1 + t)^{1/2}\}$ ($\inf \varnothing := \infty$).

Ken says : Since T is $\mathscr{F}_{0+}$-measurable, we know from the Blumenthal 0–1 law that $\mathbb{P}\{T < \infty\}$ is either 0 or 1. But $\mathbb{P}\{T < \infty\} \geq \mathbb{P}\{B_1 \geq 2^{1/2}\} > 0$, so $T < \infty$ a.s.

What do you think of Ken's argument?

Exercice 15. **(i)** Prove that $\int_0^\infty \sin^2(B_t)\,\mathrm{d}t = \infty$ a.s.

(ii) More generally, prove that if $f : \mathbb{R} \to \mathbb{R}$ is continuous which is not identically 0, then $\int_0^\infty f^2(B_t)\,\mathrm{d}t = \infty$ a.s.

Exercice 16. *(This exercise is not part of the examination program.)* Let $\mathscr{Z} := \{t \geq 0 : B_t = 0\}$. Prove that a.s., $\mathscr{Z}$ is closed, unbounded, with no isolated point.

Exercice 17. **(i)** Let $[a, b]$ and $[c, d]$ be disjoint intervals of $\mathbb{R}_+$. Prove that $\sup_{t\in[a,b]} B_s \neq \sup_{t\in[c,d]} B_s$ a.s.

(ii) Prove that a.s., each local maximum of B is a strict local maximum.

(iii) Prove that a.s., the set of times at which B realises local maxima is countable and dense in $\mathbb{R}_+$.

Exercice 18. **(i)** Let $a > 0$ and let $T_a := \inf\{t \geq 0 : B_t = a\}$. Recall that $\mathbb{E}[\mathrm{e}^{-\lambda T_a}] = \mathrm{e}^{-a(2\lambda)^{1/2}}$, $\forall \lambda \geq 0$. Prove that $\mathbb{P}(T_a \leq t) \leq \exp(-\frac{a^2}{2t})$, for all $t > 0$.

(ii) Prove that if ξ is a Gaussian $\mathscr{N}(0, 1)$ random variable, then $\mathbb{P}(\xi \geq x) \leq \frac{1}{2}\mathrm{e}^{-x^2/2}$, $\forall x > 0$.

Exercice 19. **(i)** Prove that for all $t > 0$ and all $\varepsilon > 0$, $\mathbb{P}\{\sup_{s\in[0,\,t]} |B_s| \leq \varepsilon\} > 0$.

(ii) Prove that there exists $c \in (0, \infty)$ such that $\mathbb{P}\{\sup_{s\in[0,\,1]} |B_s| \leq \varepsilon\} \geq \mathrm{e}^{-c/\varepsilon^2}$, $\forall \varepsilon \in (0, 1]$.

(iii) Prove that for all $t > 0$ and all $x > 0$, $\mathbb{P}\{\sup_{s\in[0,\,t]} |B_s| \geq x\} > 0$.

Exercice 20. **(Law of the iterated logarithm)** *(This exercise is not part of the examination program.)* Let $S_t := \sup_{s\in[0,t]} B_s$, and let $h(t) := (2t \log\log t)^{1/2}$.

(i) Let $\varepsilon > 0$. Prove that $\sum_n \mathbb{P}\{S_{t_{n+1}} \geq (1+\varepsilon)h(t_n)\} < \infty$, where $t_n = (1+\varepsilon)^n$. Prove that $\limsup_{t\to\infty} \frac{S_t}{h(t)} \leq 1$, a.s.

(ii) Prove that

$$\limsup_{t\to\infty} \frac{\sup_{s\in[0,t]} |B_s|}{h(t)} \leq 1, \qquad \text{a.s.}$$

(iii) Let $\theta > 1$, and let $s_n = \theta^n$. Prove that for all $\alpha \in (0, (1-\frac{1}{\theta})^{1/2})$, we have $\sum_n \mathbb{P}\{B_{s_n} - B_{s_{n-1}} > \alpha\, h(s_n)\} = \infty$. Prove that $\limsup_{t\to\infty} \frac{B_t}{h(t)} \geq \alpha - \frac{2}{\theta^{1/2}}$, a.s.

(iv) Prove that

$$\limsup_{t\to\infty} \frac{B_t}{h(t)} = 1, \qquad \text{a.s.}$$

(v) Let $X_1(t) := |B_t|$, $X_2(t) := S_t$, and $X_3(t) := \sup_{s\in[0,t]} |B_s|$. What can you say about $\limsup_{t\to\infty} \frac{X_i(t)}{h(t)}$ for $i = 1, 2,$ ou 3 ?

(vi) What can you say about $\liminf_{t\to\infty} \frac{B_t}{h(t)}$? And about $\limsup_{t\to 0} \frac{B_t}{[2t \log\log(1/t)]^{1/2}}$?

Exercice 21. Let $(P_t)_{t\geq 0}$ denote the semi-group of Brownian motion. Prove that if $f \in C_0$ (continuous function satisfying $\lim_{|x|\to\infty} f(x) = 0$), then $P_t f \in C_0$, $\forall t \geq 0$, and $\lim_{t\downarrow 0} P_t f = f$ uniformly on $\mathbb{R}$.

Exercice 22. Prove that if $f \in C_c^2$ (C^2 function with compact support), then

$$\lim_{t\downarrow 0} \frac{(P_t f)(x) - f(x)}{t} = \frac{1}{2} f''(x), \qquad x \in \mathbb{R}.$$

Exercice 23. Let f be a bounded Borel function on $\mathbb{R}$, and let $u(t,x) := (P_t f)(x)$ (for $t \geq 0$ and $x \in \mathbb{R}$). Prove that

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}, \qquad t > 0,\ x \in \mathbb{R}.$$

Tsinghua University *2017–2018*

"Advanced Probability" (Part III: Brownian motion)

Exercise sheet #III.3:
Brownian motion and martingales

Exercice 1. Let $a > 0$, and let $T_a^* := \inf\{t \geq 0 : |B_t| = a\}$. Prove that T_a^* has the same distribution as $\frac{a^2}{\sup_{s\in[0,1]} B_s^2}$.

Exercice 2. Let ξ and η be integrable random variables. Let $\mathscr{G} \subset \mathscr{F}$ be a sigma-algebra.

(i) Prove that $\mathbb{E}(\xi \,|\, \mathscr{G}) \leq \mathbb{E}(\eta \,|\, \mathscr{G})$, a.s., if and only if $\mathbb{E}(\xi\, \mathbf{1}_A) \leq \mathbb{E}(\eta\, \mathbf{1}_A)$ for all $A \in \mathscr{G}$.

(ii) Prove that $\mathbb{E}(\xi \,|\, \mathscr{G}) = \mathbb{E}(\eta \,|\, \mathscr{G})$, a.s., if and only if $\mathbb{E}(\xi\, \mathbf{1}_A) = \mathbb{E}(\eta\, \mathbf{1}_A)$ for all $A \in \mathscr{G}$.

Exercice 3. Let $(X_n,\, n \geq 0)$ be a sequence of real-valued random variables and let X_∞ be a real-valued random variable. Prove that $X_n \to X_\infty$ in L^1 (when $n \to \infty$) if and only if $X_n \to X_\infty$ in probability and $(X_n,\, n \geq 0)$ is uniformly integrable.

Exercice 4. Let $(X_t,\, t \geq 0)$ be a family of of real-valued random variables and let X_∞ be a real-valued random variable. Prove that if $X_t \to X_\infty$ in probability (when $t \to \infty$) and if $(X_t,\, t \geq 0)$ is uniformly integrable, then $X_t \to X_\infty$ in L^1.

Prove that the converse is, in general, not true.

Exercice 5. Let S and T be stopping times.

(i) Prove that $\mathscr{F}_S \subset \mathscr{F}_T$.

(ii) Prove that both $S \wedge T$ and $S \vee T$ are stopping times, and $\mathscr{F}_{S\wedge T} = \mathscr{F}_S \cap \mathscr{F}_T$. Moreover, $\{S \leq T\} \in \mathscr{F}_{S\wedge T}$, $\{S = T\} \in \mathscr{F}_{S\wedge T}$, $\{S < T\} \in \mathscr{F}_{S\wedge T}$.

(iii) Prove that $S + T$ is a stopping time. [Hint: both S and T are $\mathscr{F}_{S\vee T}$-measurable.]

Exercice 6. Let T be a stopping time. Then

$$T_n := \sum_{k=0}^{\infty} \frac{k}{2^n}\, \mathbf{1}_{\{\frac{k-1}{2^n} < T \leq \frac{k}{2^n}\}} + (+\infty)\, \mathbf{1}_{\{T=\infty\}}$$

is a non-increasing sequence of stopping times such that $T_n(\omega) \downarrow T(\omega)$ for all $\omega \in \Omega$.

Exercice 7. Let T Be a stopping time. Let $(X_t,\, t \geq 0)$ is an $\mathbb{R}^d$-valued adapted right-continuous (or left-continuous) process.

(i) Let $Y : \Omega \to \mathbb{R}^d$. Prove that $Y\,\mathbf{1}_{\{T<\infty\}}$ is $\mathscr{F}_T$-measurable if and only if $\forall\, t$, $Y\,\mathbf{1}_{\{T\le t\}}$ is $\mathscr{F}_t$-measurable.

(ii) Prove that for any t, the mapping $[0,\, t] \times \Omega \to \mathbb{R}^d$ defined by $(s,\, \omega) \mapsto X_s(\omega)$ is $\mathscr{B}([0,\, t]) \otimes \mathscr{F}_t$-measurable, where $\mathscr{B}([0,\, t])$ denotes the Borel σ-field of $[0,\, t]$.

(iii) Prove that $X_T\,\mathbf{1}_{\{T<\infty\}}$ is $\mathscr{F}_T$-measurable.

Exercice 8. Let $(X_t,\ t \ge 0)$ be a submartingale. Prove that for all $t \ge 0$, we have $\sup_{s\in[0,t]} \mathbb{E}(|X_s|) < \infty$.

Exercice 9. Let $(B_t, t \ge 0)$ be Brownian motion, and let $(\mathscr{F}_t)$ be its canonical filtration. Then the following processes are martingales:

(i) $(B_t, t \ge 0)$.

(ii) $(B_t^2 - t,\, t \ge 0)$.

(iii) For any $\theta \in \mathbb{R}$, $(\mathrm{e}^{\theta B_t - \frac{\theta^2}{2} t},\, t \ge 0)$.

Exercice 10. Let $(X_t,\, t \ge 0)$ be a process with independent increments, and let $(\mathscr{F}_t)$ be its canonical filtration.

(i) If for all t, $\mathbb{E}(|X_t|) < \infty$, then $\widetilde{X}_t := X_t - \mathbb{E}(X_t)$ is a martingale.

(ii) If for all t, $\mathbb{E}(X_t^2) < \infty$, then $Y_t := \widetilde{X}_t^2 - \mathbb{E}(\widetilde{X}_t^2)$ is a martingale.

(iii) Let $\theta \in \mathbb{R}$. If $\mathbb{E}(\mathrm{e}^{\theta X_t}) < \infty$ for all $t \ge 0$, then $(Z_t := \frac{\mathrm{e}^{\theta X_t}}{\mathbb{E}[\mathrm{e}^{\theta X_t}]},\, t \ge 0)$ is a martingale.

Exercice 11. Let $X := (X_t,\ t \ge 0)$ be a martingale such that $\sup_{t\ge 0} \mathbb{E}(|X_t|) < \infty$.

(i) Prove that for all $t \ge 0$, $\mathbb{E}(X_n^+ \,|\, \mathscr{F}_t)$ converges (when $n \to \infty$) a.s. to a real-valued random variable, denoted by α_t .

(ii) Prove that $(\alpha_t,\, t \ge 0)$ is a martingale.

(iii) Prove that X is the difference of two non-negative martingales.

Exercice 12. Let ξ be a real-valued random variable. Let $X_t := \mathbb{P}(\xi \le t \,|\, \mathscr{F}_t)$. Prove that $(X_t,\, t \ge 0)$ is a submartingale.

Exercice 13. Let $(X_t,\ t \ge 0)$ be a submartingale. Prove that $\sup_{t\ge 0} \mathbb{E}(X_t^+) < \infty$ if and only if $\sup_{t\ge 0} \mathbb{E}(|X_t|) < \infty$.

Exercice 14. Let $(X_t,\ t \ge 0)$ be a martingale. If there exists $\xi \in L^1(\mathbb{P})$ such that for all $t \ge 0$, $\mathbb{E}(\xi \,|\, \mathscr{F}_t) = X_t$ a.s., we say that $(X_t,\ t \ge 0)$ is closed by ξ.

Prove that a right-continuous martingale is closed if and only if it is uniformly integrable.

Exercice 15. (Discrete backward submartingales) Let $(\mathscr{F}_n,\, n \le 0)$ be a sequence of sub-σ-fields of $\mathscr{F}$ satisfying $\mathscr{F}_n \subset \mathscr{F}_{n+1}$ for all $n \le 0$. Let $(X_n,\, n \le 0)$ be such that $\forall\, n$,

X_n is $\mathscr{F}_n$-measurable et integrable, and that $\mathbb{E}(X_{n+1} \,|\, \mathscr{F}_n) \ge X_n$ a.s. We call $(X_n,\, n \le 0)$ a backward submartingale.

(i) Let $a < b$. Let $U_n(X;\, a,\, b)$ be the number of up-crossings along $[a,\, b]$ by $X_n, \cdots, X_{-1}$, X_0. Prove that $\mathbb{E}[U_n(X;\, a,\, b)] \le \frac{\mathbb{E}[(X_0-a)^+]}{b-a}$.

(ii) Prove that $X_n \to X_{-\infty}$ a.s. when $n \to -\infty$.

(iii) Assume from now on that $\inf_{n\le 0} \mathbb{E}(X_n) > -\infty$. Prove that $X_n \to X_{-\infty}$ in L^1.

Hint: Only uniform integrability needs proved. By considering $X_n - \mathbb{E}(X_0 \,|\, \mathscr{F}_n)$, you can argue that X_n may be assumed to take values in $(-\infty,\, 0]$.

(iv) Prove that $X_{-\infty} \le \mathbb{E}(X_0 \,|\, \mathscr{F}_{-\infty})$ a.s., where $\mathscr{F}_{-\infty} := \bigcap_{n\le 0} \mathscr{F}_n$.

(v) (P. Lévy) Let ξ be a real-valued random variable with $\mathbb{E}(|\xi|) < \infty$. Prove that $\mathbb{E}(\xi \,|\, \mathscr{F}_n) \to \mathbb{E}(\xi \,|\, \mathscr{F}_{-\infty})$ a.s. and in L^1, as $n \to -\infty$.

Exercice 16. Let $(X_t,\, t \ge 0)$ be a continuous and non-negative martingale. Let $T := \inf\{t \ge 0 :\, X_t = 0\}$ (with $\inf \varnothing := \infty$). Prove that a.s. on $\{T < \infty\}$, we have $X_t = 0$, $\forall t \ge T$.

Exercice 17. Let $(X_t,\; t \ge 0)$ be a right-continuous submartingale, and let S and T be bounded stopping times. Prove that

$$\mathbb{E}(X_T \,|\, \mathscr{F}_S) \ge X_{T\wedge S}, \qquad \text{a.s.}$$

Exercice 18. Let $(X_t,\; t \ge 0)$ be a right-continuous martingale. Let T be a stopping time.

(i) Prove that $(X_{T\wedge t},\; t \ge 0)$ is a right-continuous martingale.

(ii) Prove that if $(X_t,\; t \ge 0)$ is uniformly integrable, then so is $(X_{T\wedge t},\; t \ge 0)$.

Exercice 19. Let $(X_t,\, t \ge 0)$ be a non-negative and right-continuous *super*martingale. Recall that $X_t \to X_\infty$ a.s. in this case. Prove that if $\mathbb{E}(X_\infty) = \mathbb{E}(X_0)$, then $(X_t,\, t \ge 0)$ is a uniformly integrable martingale.

Exercice 20. Let $X = (X_t,\, t \ge 0)$ be a non-negative continuous submartingale. We write $S_t := \sup_{s\in[0,\, t]} X_s$, $t \ge 0$.

(i) Prove that for all $\lambda > 0$ and all $t \ge 0$, $\lambda\, \mathbb{P}(S_t > 2\lambda) \le \mathbb{E}[\, X_t\, \mathbf{1}_{\{X_t > \lambda\}}]$.

We can use the following inequality: for all $a > 0$, $a\, \mathbb{P}(S_t > a) \le \mathbb{E}[\, X_t\, \mathbf{1}_{\{S_t > a\}}]$ (this follows from the maximal inequality for discrete-time submartingales and the continuity of the trajectories).

(ii) Prove that $\frac{1}{2}\, \mathbb{E}[S_t] \le 1 + \mathbb{E}[\, X_t \log_+ X_t\,]$, wher $\log_+ x := \log \max(x, 1)$.

(iii) Let $(Y_t,\, t \ge 0)$ be a continuous and uniformly integrable martingale. We assume that $\mathbb{E}[\, |Y_\infty|\, \log_+ |Y_\infty|\,] < \infty$. Prove that $\sup_{t\ge 0} |Y_t|$ is integrable.

Exercice 21. For any martingale $X := (X_t,\, t \geq 0)$, we say that it is square-integrable if $\mathbb{E}(X_t^2) < \infty$, $\forall t \geq 0$, and that it is bounded in L^2 if $\sup_{t\geq 0} \mathbb{E}(X_t^2) < \infty$.

(i) Prove that if X is a right-continuous martingale and is bounded in L^2, then it is uniformly integrable, with $\mathbb{E}(\sup_{t\geq 0} X_t^2) < \infty$.

(ii) Let X and Y be right-continuous martingales that are bounded in L^2. Let S and T be stopping times. Prove that $\mathbb{E}(X_S Y_T) = \mathbb{E}(X_{S\wedge T} Y_{S\wedge T})$.

(iii) Let X and Y be right-continuous and square-integrable martingales. Let S and T be bounded stopping times. Prove that $\mathbb{E}(X_S Y_T) = \mathbb{E}(X_{S\wedge T} Y_{S\wedge T})$.

Exercice 22. Let $S \leq T$ be bounded stopping times. Prove that $\mathbb{E}[(B_T - B_S)^2] = \mathbb{E}(B_T^2) - \mathbb{E}(B_S^2) = \mathbb{E}(T - S)$.

Exercice 23. **(i)** Let $(X_t,\, t \geq 0)$ be a non-negative and continuous martingale such that $X_t \to 0$, a.s. $(t \to \infty)$. Prove that for all $x > 0$, $\mathbb{P}(\sup_{t\geq 0} X_t \geq x \,|\, \mathscr{F}_0) = 1 \wedge \frac{X_0}{x}$, a.s.

(ii) Let B be Brownian motion. Determine the law of $\sup_{t\geq 0}(B_t - t)$.

Exercice 24. Let $\gamma \neq 0$, $a > 0$ and $b > 0$. Let $T_x := \inf\{t > 0 :\, B_t + \gamma t = x\}$, $x = -a$ or b. Compute $\mathbb{P}(T_{-a} > T_b)$.

Hint: You can use the martingale $(\mathrm{e}^{-2\gamma\,(B_t+\gamma t)},\, t \geq 0)$.

Exercice 25. **(First Wald identity)** Let T be a stopping time such that $\mathbb{E}(T) < \infty$. Prove that B_T is integrable and that $\mathbb{E}(B_T) = 0$.

Exercice 26. **(Second Wald identity)** Let T be a stopping time such that $\mathbb{E}(T) < \infty$. Prove that B_T has a finite second moment and that $\mathbb{E}(B_T^2) = \mathbb{E}(T)$.

Tsinghua University *2017–2018*

"Advanced Probability" (Part III: Brownian motion)

Exercise sheet #III.1:
Construction of Brownian motion

Exercice 1. **(i)** We have

$$\mathbb{P}(\xi > x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \mathrm{e}^{-u^2/2}\,\mathrm{d}u \le \frac{1}{\sqrt{2\pi}} \frac{1}{x} \int_x^\infty u\mathrm{e}^{-u^2/2}\,\mathrm{d}u = \frac{1}{\sqrt{2\pi}} \frac{1}{x} \mathrm{e}^{-x^2/2},$$

giving the desired upper bound. For the lower bound, we note that by integration by parts,

$$\mathbb{P}(\xi > x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \mathrm{e}^{-u^2/2}\,\mathrm{d}u = \Big[-\frac{1}{\sqrt{2\pi}} \frac{1}{u} \mathrm{e}^{-u^2/2} \Big]_x^\infty - \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{1}{u^2} \mathrm{e}^{-u^2/2}\,\mathrm{d}u.$$

This yields the desired lower bound because $\int_x^\infty \frac{1}{u^2}\mathrm{e}^{-u^2/2}\,\mathrm{d}u \le \frac{1}{x^3}\int_x^\infty u\mathrm{e}^{-u^2/2}\,\mathrm{d}u = \frac{1}{x^3}$.

(ii) By the Markov inequality, for any $\lambda > 0$,

$$\mathbb{P}(\xi > x) \le \mathrm{e}^{-\lambda x}\mathbb{E}[\mathrm{e}^{\lambda \xi}] = \mathrm{e}^{-\lambda x + \lambda^2/2},$$

which yields the desired inequality by taking $\lambda = x$. □

Exercice 2. **(i)** We have $\mathbb{E}(\xi^4) = 3$, $\mathbb{E}(|\xi|) = (\frac{2}{\pi})^{1/2}$.

(ii) We have $\mathbb{E}(\mathrm{e}^{a\xi}) = \mathrm{e}^{a^2/2}$, $\mathbb{E}(\xi \mathrm{e}^{a\xi}) = a\mathrm{e}^{a^2/2}$. As for $\mathbb{E}(\mathrm{e}^{a\xi^2})$, it is seen that $\mathbb{E}(\mathrm{e}^{a\xi^2}) = \infty$ if $a \ge \frac{1}{2}$, whereas $\mathbb{E}(\mathrm{e}^{a\xi^2}) = (1-2a)^{-1/2}$ if $a < \frac{1}{2}$.

(iii) By conditioning on ξ, we habe, by (ii), $\mathbb{E}(\mathrm{e}^{\lambda \xi \eta} \,|\, \xi) = \mathrm{e}^{\lambda^2\xi^2/2}$, which is nothing else but $\mathrm{e}^{b\xi^2}$. Taking expectation on both sides gives the desired conclusion. □

Exercice 3. For any random variable ξ, we denote its characteristic function by φ_ξ. By assumption, $\varphi_{\xi_n}(t) = \exp(i\mu_n t - \frac{\sigma_n^2}{2}t^2)$ converges pointwise to $\varphi_\xi(t)$. So $\exp(-\frac{\sigma_n^2}{2}t^2) \to |\varphi_\xi(t)|$ for any $t \in \mathbb{R}$. As a consequence, $\sigma_n^2 \to \sigma^2 \ge 0$ (the possibility that $\sigma_n^2 \to \infty$ is excluded as $\mathbf{1}_{\{t=0\}}$ is not a characteristic function, being discontinuous at point 0).

Suppose that (μ_n) is unbounded. Then there exists a subsequence (μ_{n_k}) tending to $+\infty$ (or to $-\infty$, but the argument will be identical). Let $a \in \mathbb{R}$. The distribution function F_ξ of ξ being non-decreasing, we can find $b \ge a$ which is a point of continuity of F_ξ. Hence

$$F_\xi(a) \le F_\xi(b) = \lim_{k\to\infty} \mathbb{P}(\xi_{n_k} \le b) \le \frac{1}{2},$$

as for large k, $\mathbb{P}(\xi_{n_k} \leq b) \leq \mathbb{P}(\xi_{n_k} \leq \mu_{n_k}) = \frac{1}{2}$. So $F_\xi(a) \leq \frac{1}{2}$ for all $a \in \mathbb{R}$, which is absurd because F_ξ is a distribution function and its limit at $+\infty$ is 1.

The sequence (μ_n) is thus bounded. Let $\mu \in \mathbb{R}$ and $\nu \in \mathbb{R}$ be limits along subsequences, then $\mathrm{e}^{i\mu t} = \mathrm{e}^{i\nu t}$ for all $t \in \mathbb{R}$, which is possible only if $\mu = \nu$. So the sequence (μ_n) converges, to a limit, denoted by $\mu \in \mathbb{R}$. Since $\sigma_n \to \sigma$, we have $\varphi_\xi(t) = \exp(i\mu t - \frac{\sigma^2}{2} t^2)$. In other words, ξ is Gaussian $\mathscr{N}(\mu, \sigma^2)$. $\square$

Exercice 4. We use what we have proved in the previous exercise. For $a \in \mathbb{R}$, we have

$$\mathbb{E}(\mathrm{e}^{a\xi_n}) = \exp\Big(a\mu_n + \frac{a^2\sigma_n^2}{2}\Big).$$

Since $\mathrm{e}^{|x|} \leq \mathrm{e}^x + \mathrm{e}^{-x}$, we have, for all $a \geq 0$, $\sup_n \mathbb{E}(\mathrm{e}^{a|\xi_n|}) < \infty$. A fortiori, $\sup_n \mathbb{E}(|\xi_n|^{p+1}) < \infty$; hence $\sup_n \mathbb{E}(\,|\xi_n - \xi|^{p+1}) < \infty$. This implies that $(|\xi_n - \xi|^p)$ is uniformly integrable. Since $|\xi_n - \xi|^p \to 0$ in probability, the convergence takes place also in L^1.

Exercice 5. **(i)** Let $a \in \mathbb{R}$ and $b \in \mathbb{R}$. It is clear that $(\xi, \eta, \theta - a\xi - b\eta)$, being a linear transform of the Gaussian random variable (ξ, η, θ), is also a Gaussian random variable. So $\theta - a\xi - b\eta$ and (ξ, η) are independent if and only if $\mathrm{Cov}(\theta - a\xi - b\eta, \xi) = \mathrm{Cov}(\theta - a\xi - b\eta, \eta) = 0$.

We have $\mathrm{Cov}(\theta - a\xi - b\eta, \xi) = \mathrm{Cov}(\xi, \theta) - a\sigma_\xi^2$, and $\mathrm{Cov}(\theta - a\xi - b\eta, \eta) = \mathrm{Cov}(\eta, \theta) - b\sigma_\eta^2$. Choosing from now on $a := \mathrm{Cov}(\xi, \theta)/\sigma_\xi^2$ and $b := \mathrm{Cov}(\eta, \theta)/\sigma_\eta^2$, it is seen that $\theta - a\xi - b\eta$ is independent of (ξ, η). Accordingly,

$$\begin{aligned} \mathbb{E}(\theta \,|\, \xi, \eta) &= \mathbb{E}(\theta - a\xi - b\eta \,|\, \xi, \eta) + a\xi + b\eta \\ &= \mathbb{E}(\theta - a\xi - b\eta) + a\xi + b\eta = \mathbb{E}(\theta) + a\xi + b\eta. \end{aligned}$$

On the other hand, $\theta - a\xi$ is independent of ξ: indeed, $(\xi, \theta - a\xi)$ is a Gaussian random vector, with $\mathrm{Cov}(\xi, \theta - a\xi) = 0$; hence $\mathbb{E}(\theta \,|\, \xi) = \mathbb{E}(\theta - a\xi \,|\, \xi) + a\xi = \mathbb{E}(\theta - a\xi) + a\xi = \mathbb{E}(\theta) + a\xi$. Similarly, $\mathbb{E}(\theta \,|\, \eta) = \mathbb{E}(\theta) + b\eta$. As a consequence,

$$\mathbb{E}(\theta \,|\, \xi, \eta) = \mathbb{E}(\theta) + a\xi + b\eta = \mathbb{E}(\theta \,|\, \xi) + \mathbb{E}(\theta \,|\, \eta) - \mathbb{E}(\theta).$$

(ii) Let $A \in \sigma(\xi\eta)$. By definition, there exists a Borel set $B \subset \mathbb{R}$ such that $A = \{\omega : \xi(\omega)\eta(\omega) \in B\}$. So $\mathbf{1}_A = \mathbf{1}_B(\xi\eta)$.

Since (ξ, η) is a *centered* Gaussian random vector, it is distributed as $(-\xi, -\eta)$. Thus $\mathbb{E}[\xi\, \mathbf{1}_B(\xi\eta)] = \mathbb{E}[(-\xi)\mathbf{1}_B((-\xi)(-\eta))] = -\mathbb{E}[\xi\, \mathbf{1}_B(\xi\eta)]$, i.e., $\mathbb{E}[\xi\, \mathbf{1}_B(\xi\eta)] = 0$. In other words, $\mathbb{E}(\xi\, \mathbf{1}_A) = 0$, $\forall A \in \sigma(\xi\eta)$, which means that $\mathbb{E}(\xi \,|\, \xi\eta) = 0$.

(iii) We have $\mathbb{E}(\theta \,|\, \xi\eta) = \mathbb{E}(\theta - a\xi - b\eta \,|\, \xi\eta) + a\,\mathbb{E}(\xi \,|\, \xi\eta) + b\,\mathbb{E}(\eta \,|\, \xi\eta)$. By (ii), $\mathbb{E}(\xi \,|\, \xi\eta) = 0$; similarly, $\mathbb{E}(\eta \,|\, \xi\eta) = 0$. It follows that $\mathbb{E}(\theta \,|\, \xi\eta) = \mathbb{E}(\theta - a\xi - b\eta \,|\, \xi\eta)$. We have seen that $\theta - a\xi - b\eta$ is independent of (ξ, η); so $\mathbb{E}(\theta - a\xi - b\eta \,|\, \xi\eta) = \mathbb{E}(\theta - a\xi - b\eta) = \mathbb{E}(\theta)$, which yields the desired identity. $\square$

Exercice 6. We prove by induction in n. The case $n=0$ is trivial. Assume that the desired conclusion holds for $n-1$. It is clear that $(X_n(\frac{k}{2^n}),\, 0\le k\le 2^n)$ is a Gaussian random vector (which is obviously centered), being a linear function of independent Gaussian vectors $(X_{n-1}(\frac{k}{2^{n-1}}),\, 0\le k\le 2^{n-1})$ and $(\xi_{k,n},\, 0\le k\le 2^n)$. It remains to check the covariance. We distinguish two possible situations.

First situation: there is at least an even number among k and ℓ, say $k=2k_1$. In this case, $X_n(\frac{k}{2^n}) = X_{n-1}(\frac{k_1}{2^{n-1}})$, and the desired identity $\mathrm{Cov}(X_{n-1}(\frac{k}{2^n}),\, X_{n-1}(\frac{\ell}{2^n})) = \frac{k}{2^n}\wedge\frac{\ell}{2^n}$ is trivial by the induction hypothesis if ℓ is even; if, however, ℓ is odd, say $\ell = 2\ell_1+1$, then $X_n(\frac{\ell}{2^n}) = \frac12 X_{n-1}(\frac{\ell_1}{2^{n-1}}) + \frac12 X_{n-1}(\frac{\ell_1+1}{2^{n-1}}) + \frac{\xi_{\ell,n}}{2^{(n+1)/2}}$; since $\xi_{\ell,n}$ is independent of $X_{n-1}(\frac{k_1}{2^{n-1}})$, we obtain:

$$\begin{aligned}
&\mathrm{Cov}\Big(X_n(\frac{k}{2^n}),\, X_n(\frac{\ell}{2^n})\Big)\\
=\;& \frac12\,\mathrm{Cov}\Big(X_{n-1}(\frac{k_1}{2^{n-1}}),\, X_{n-1}(\frac{\ell_1}{2^{n-1}})\Big) + \frac12\,\mathrm{Cov}\Big(X_{n-1}(\frac{k_1}{2^{n-1}}),\, X_{n-1}(\frac{\ell_1+1}{2^{n-1}})\Big),
\end{aligned}$$

which, by the induction hypothesis, is $\frac12(\frac{k_1}{2^{n-1}}\wedge\frac{\ell_1}{2^{n-1}}) + \frac12(\frac{k_1}{2^{n-1}}\wedge\frac{\ell_1+1}{2^{n-1}}) = \frac{k}{2^n}\wedge\frac{\ell}{2^n}$ as desired.

Second (and last) situation: both k and ℓ odd numbers, say $k=2k_1+1$ and $\ell=2\ell_1+1$. In this case, we have $X_n(\frac{k}{2^n}) = \frac12 X_{n-1}(\frac{k_1}{2^{n-1}}) + \frac12 X_{n-1}(\frac{k_1+1}{2^{n-1}}) + \frac{\xi_{k,n}}{2^{(n+1)/2}}$ and $X_n(\frac{\ell}{2^n}) = \frac12 X_{n-1}(\frac{\ell_1}{2^{n-1}}) + \frac12 X_{n-1}(\frac{\ell_1+1}{2^{n-1}}) + \frac{\xi_{\ell,n}}{2^{(n+1)/2}}$. Since $\xi_{k,n}$ and $\xi_{\ell,n}$ are independent of $(X_{n-1}(t),\, t\in[0,\,1])$, we have, by the induction hypothesis,

$$\begin{aligned}
\mathrm{Cov}\Big(X_n(\frac{k}{2^n}),\, X_n(\frac{\ell}{2^n})\Big) \;=\;& \frac14(\frac{k_1}{2^{n-1}}\wedge\frac{\ell_1}{2^{n-1}}) + \frac14(\frac{k_1}{2^{n-1}}\wedge\frac{\ell_1+1}{2^{n-1}}) +\\
&+\frac14(\frac{k_1+1}{2^{n-1}}\wedge\frac{\ell_1}{2^{n-1}}) + \frac14(\frac{k_1+1}{2^{n-1}}\wedge\frac{\ell_1+1}{2^{n-1}}) + \frac{1}{2^{n+1}}\mathrm{Cov}(\xi_{k,n},\, \xi_{\ell,n}).
\end{aligned}$$

It is then easily checked that the sum of the five terms on the right-hand side is indeed $\frac{k}{2^n}\wedge\frac{\ell}{2^n}$.

By induction, we conclude that $\mathrm{Cov}[X_n(\frac{k}{2^n})X_n(\frac{\ell}{2^n})] = \frac{k}{2^n}\wedge\frac{\ell}{2^n}$. $\square$

Exercice 7. Clearly, the trajectories of B are a.s. continuous. It is easily checked that B is a centered Gaussian process with covariance $\mathrm{Cov}(B_t,\, B_s) = t\wedge s$ for all $s\ge 0$ and $t\ge 0$. $\square$

Exercice 8. For all $t\ge 0$, X_t is continuous, thus measurable with respect to $\mathscr{C}(\mathbb{R}_+,\mathbb{R})$. Consequently, $\sigma(X_t,\, t\ge 0)\subset \mathscr{C}(\mathbb{R}_+,\mathbb{R})$.

Conversely, for all $\mathrm{w}_0\in C(\mathbb{R}_+,\mathbb{R})$, $\delta_n(\mathrm{w},\mathrm{w}_0) = \sup_{t\in[0,n]\cap\mathbb{Q}} |\mathrm{w}(t)-\mathrm{w}_0(t)|$ is $\sigma(X_t,\, t\ge 0)$-measurable, and so is $d(\mathrm{w},\mathrm{w}_0)$. Le F be a closed subset of $C(\mathbb{R}_+,\mathbb{R})$, and let (w_n) be a sequence that is dense in F (because the space is separable), then

$$F = \{\mathrm{w}\in C(\mathbb{R}_+,\mathbb{R}) : d(\mathrm{w},F)=0\} = \{\mathrm{w}\in C(\mathbb{R}_+,\mathbb{R}) : \inf_n d(\mathrm{w},\mathrm{w}_n)=0\},$$

which is an element of $\sigma(X_t,\, t\ge 0)$. Hence, $\mathscr{C}(\mathbb{R}_+,\mathbb{R})\subset\sigma(X_t,\, t\ge 0)$.

It is also possible to directly prove that all the open sets are $\sigma(X_t,\, t \geq 0)$-measurable, by means of the following property[1]: if a metric space is separable, then all opens sets are countable unions of open balls. $\square$

Exercice 9. Let $t > 0$. We have $\mathbb{P}(T < \infty) \geq \mathbb{P}(T \leq t) \geq \mathbb{P}(B_t \geq 1)$. Since $\mathbb{P}(B_t \geq 1) \to \frac{1}{2}$ when $t \to \infty$, we obtain: $\mathbb{P}(T < \infty) \geq \frac{1}{2}$. $\square$

Exercice 10. Both are centered Gaussian processes with covariance $s \wedge t$ and with a.s. continuous trajectories. $\square$

Exercice 11. (i) By definition, ξ is the a.s. limit of $\xi_n := 2^{-n} \sum_{i=1}^{2^n} B_{i/2^n}$, and a fortiori, the weak limit. For each n, ξ_n is Gaussian (because Brownian motion is a Gaussian process). By Exercice 4, ξ is Gaussian, with $\mathbb{E}(\xi) = \lim_{n\to\infty} \mathbb{E}(\xi_n)$ and $\mathrm{Var}(\xi) = \lim_{n\to\infty} \mathrm{Var}(\xi_n)$.

Since $\mathbb{E}(\xi_n) = 0$, $\forall n$, we have $\mathbb{E}(\xi) = 0$.

Since $\mathrm{Var}(\xi_n) = 2^{-2n} \sum_{i=1}^{2^n} \sum_{j=1}^{2^n} (\frac{i}{2^n} \wedge \frac{j}{2^n}) \to \int_0^1 \int_0^1 (s \wedge t)\, \mathrm{d}s\, \mathrm{d}t = \frac{1}{3}$, we have $\mathrm{Var}(\xi) = \frac{1}{3}$.

Conclusion : ξ is Gaussian $\mathscr{N}(0,\, \frac{1}{3})$.

(ii) Let $a \in \mathbb{R}$ and $b \in \mathbb{R}$. Exactly as in (i), we see that $aB_1 + b\eta$ is Gaussian, and centered; in other words, $(B_1,\, \eta)$ is a centered Gaussian random vector. Moreover, $\mathbb{E}(B_1) = 0 = \mathbb{E}(\eta)$, $\mathbb{E}(B_1^2) = 1$, $\mathbb{E}(\eta^2) = \frac{8}{3}$, and $\mathbb{E}(B_1\eta)$ is, by Fubini's theorem (why?), $= \int_0^2 \mathbb{E}(B_1 B_t)\, \mathrm{d}t = \int_0^2 (1 \wedge t)\, \mathrm{d}t = \frac{3}{2}$. Hence $(B_1,\, \eta)$ has the Gaussian law $\mathscr{N}(\begin{pmatrix} 0 \\ 0 \end{pmatrix},\, \begin{pmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & \frac{8}{3} \end{pmatrix})$.

In particular, $\mathbb{E}(B_1 \,|\, \eta) = \frac{\mathbb{E}(B_1\eta)}{\mathbb{E}(\eta^2)}\, \eta = \frac{9}{16}\, \eta$.

(iii) Let $n \geq 1$, and let $(s_1,\, \cdots,\, s_n) \in [0,\, 1]^n$. Then $(B_7 - B_2,\, B_{s_1},\, \cdots,\, B_{s_n})$ is a centered Gaussian random vector. Since $\mathrm{Cov}(B_7 - B_2,\, B_{s_i}) = \mathrm{Cov}(B_7,\, B_{s_i}) - \mathrm{Cov}(B_2,\, B_{s_i}) = s_i - s_i = 0$ for all $i \leq n$, an important property (which one?) of Gaussian random vectors tells us that $B_7 - B_2$ is independent of $(B_{s_1},\, \cdots,\, B_{s_n})$. This implies that $B_7 - B_2$ is independent of $\sigma(B_s,\, s \in [0,\, 1])$.

(iv) Exactly as in the previous question, we see that $B_5 - B_1$ is indepenfent of $\mathscr{F}_1$. In particular, $\mathbb{E}(B_5 \,|\, \mathscr{F}_1) = \mathbb{E}(B_5 - B_1 \,|\, \mathscr{F}_1) + \mathbb{E}(B_1 \,|\, \mathscr{F}_1) = \mathbb{E}(B_5 - B_1) + B_1 = B_1$, et $\mathbb{E}(B_5^2 \,|\, \mathscr{F}_1) = \mathbb{E}((B_5 - B_1)^2 \,|\, \mathscr{F}_1) + 2B_1 \mathbb{E}(B_5 \,|\, \mathscr{F}_1) - B_1^2 = \mathbb{E}((B_5 - B_1)^2) + 2B_1^2 - B_1^2 = 4 + B_1^2$. $\square$

Exercice 12. (i) The answer is yes, by the scaling property.

(ii) The answer is no: the trajectories of the second process are a.s. C^∞, whereas those of the first are a.s. not C^2. $\square$

Exercice 13. The measurability of B_T is clear if we work in the canonical space of Brownian motion. Let us compute its characteristic function.

[1]Let G be an open set, and let D be a countable set that is dense, then for all $x \in G$, there exist $x_D \in D$ and $n_x \geq 1$ sufficiently large such that $x \in B(x_D,\, \frac{1}{n_x}) \subset G$. Thus $G = \cup_{x\in G} B(x_D,\, \frac{1}{n_x})$. The family $\{B(x_D,\, \frac{1}{n_x}),\, x \in G\}$ is countable, being a subset of $\{B(x,\, \frac{1}{n}),\, x \in D,\, n \geq 1\}$.

Let $x \in \mathbb{R}$. We have $\mathbb{E}[e^{ixB_T} \mid T] = e^{-x^2T/2}$, so $\mathbb{E}[e^{ixB_T}] = \mathbb{E}[e^{-x^2T/2}] = \frac{2}{2+x^2}$. In other words, B_T has density $(1/\sqrt{2})e^{-\sqrt{2}|x|}$ ("two-sided exponential law" of parameter $\sqrt{2}$). $\square$

Exercice 14. (i) By Fubini–Tonelli, $\mathbb{E}(\int_0^1 |\frac{B_s}{s}| \, ds) = \int_0^1 \mathbb{E}(|\frac{B_s}{s}|) \, ds = c\int_0^1 s^{-1/2} \, ds < \infty$, where $c := \mathbb{E}(|B_1|) < \infty$. A fortiori, $\int_0^1 |\frac{B_s}{s}| \, ds < \infty$ a.s. Consequently, $\int_0^1 \frac{B_s}{s} \, ds$ is a.s. well defined.

[One can also directly prove that $\int_0^1 \frac{B_s}{s} \, ds$ is a.s. well defined by means of the Hölder continuity of B.]

(ii) Exactly as in (i), we see that for all $t > 0$, $X_t := \int_0^t \frac{B_s}{s} \, ds$ is well defined a.s. So a.s., the process $(X_t, \, t \geq 0)$ is well defined (why?), with continuous trajectories, and so is $(\beta_t := B_t - X_t, \, t \geq 0)$.

As in a previous exercise, we see that for all n and all real numbers $a_1, \cdots, a_n$, $\sum_{i=1}^n a_i \beta_{t_i}$ is centered Gaussian. As a consequence, β is a centered Gaussian process.

It remains to check the covariance. Let $t \geq s > 0$. We have $\mathbb{E}(X_t B_s) = s + s\log(\frac{t}{s})$ (why?), $\mathbb{E}(X_s B_t) = s$ and $\mathbb{E}(X_s X_t) = 2s + s\log(\frac{t}{s})$. Hence $\mathbb{E}(\beta_t \beta_s) = \mathbb{E}(B_t B_s) - \mathbb{E}(X_t B_s) - \mathbb{E}(X_s B_t) + \mathbb{E}(X_t X_s) = s$ as desired. Consequently, β is Brownian motion. $\square$

Exercice 15. Let $X_t := \int_0^t |B_s| \, ds$, $t \geq 0$. By scaling, for all $t > 0$, X_t is distributed as $t^{3/2} X_1$. For all $x > 0$, we have $\mathbb{P}\{X_\infty \geq x\} \geq \mathbb{P}\{X_t \geq x\} = \mathbb{P}\{X_1 \geq \frac{x}{t^{3/2}}\}$ which converges to $\mathbb{P}\{X_1 > 0\} = 1$ when $t \to \infty$. Since this holds for all $x > 0$, we get $X_\infty = \infty$ a.s. $\square$

Exercice 16. **(i)** For all $t > 0$, we have $\mathbb{P}(T < \infty) \geq \mathbb{P}(T \leq t) \geq \mathbb{P}(\{B_t \geq 1\} \cup \{B_t \leq -1\}) = \mathbb{P}(B_t \geq 1) + \mathbb{P}(B_t \leq -1) = 2\,\mathbb{P}(B_t \geq 1)$. Since $\mathbb{P}(B_t \geq 1) \to \frac{1}{2}$ when $t \to \infty$, we get $\mathbb{P}(T < \infty) \geq 1$. In other words, $T < \infty$ a.s.

(ii) For bounded Borel function $f : \mathbb{R}_+ \to \mathbb{R}$ and by symmetry of Brownian motion (replacing B by $-B$), we have $\mathbb{E}[f(T)\,\mathbf{1}_{\{B_T=1\}}] = \mathbb{E}[f(T)\,\mathbf{1}_{\{B_T=-1\}}]$; hence

$$\mathbb{E}[f(T)\,\mathbf{1}_{\{B_T=1\}}] = \frac{1}{2}\mathbb{E}[f(T)] = \mathbb{P}(B_T = 1)\mathbb{E}[f(T)],$$

the last identity following from the fact that $\mathbb{P}(B_T = 1) = \frac{1}{2}$ (taking $f \equiv 1$ in the previous identity). Similarly, $\mathbb{E}[f(T)\,\mathbf{1}_{\{B_T=-1\}}] = \mathbb{P}(T = -1)\mathbb{E}[f(B_T)]$. This yields the desired independence. $\square$

Exercice 17. **(i)** Write

$$B_t - B_s = \frac{t-s}{1-s}\,(B_1 - B_s) + \frac{1-t}{1-s}\,(B_t - B_s) - \frac{t-s}{1-s}\,(B_1 - B_t).$$

Clearly, $\frac{t-s}{1-s}\,(B_1 - B_s)$ is $\mathscr{G}_s$-measurable. We now prove that $X := \frac{1-t}{1-s}\,(B_t - B_s) - \frac{t-s}{1-s}\,(B_1 - B_t)$ is independent of $\mathscr{G}_s$. It suffices to prove that for all n and all $0 \leq s_1 < \cdots < s_n \leq s$, X is independent of $(B_{s_1}, \cdots, B_{s_n}, B_1)$.

Since $(X,\, B_{s_1},\, \cdots,\, B_{s_n},\, B_1)$ is a Gaussian vector, it suffices to check that $\mathrm{Cov}(X,\, B_{s_i}) = \mathrm{Cov}(X,\, B_1) = 0$, $\forall i$. We have $\mathrm{Cov}(X,\, B_{s_i}) = \frac{1-t}{1-s}\,(s_i - s_i) - \frac{t-s}{1-s}\,(s_i - s_i) = 0$ and $\mathrm{Cov}(X,\, B_1) = \frac{1-t}{1-s}\,(t-s) - \frac{t-s}{1-s}\,(1-t) = 0$, as desired.

So X is independent of $\mathscr{G}_s$: we have $\mathbb{E}[X \,|\, \mathscr{G}_s] = \mathbb{E}[X] = 0$. As a consequence, $\mathbb{E}[(B_t - B_s) \,|\, \mathscr{G}_s] = \frac{t-s}{1-s}\,(B_1 - B_s)$.

(ii) [The integral $\int_0^1 \frac{B_1 - B_s}{1-s}\,\mathrm{d}s$ is a.s. well defined by the local Hölder continuity of Brownian sample paths.]

Let $1 \geq t > s \geq 0$. By (i), $\mathbb{E}[B_t \,|\, \mathscr{G}_s] = B_s + \frac{t-s}{1-s}\,(B_1 - B_s)$,and $\mathbb{E}[(B_1 - B_u) \,|\, \mathscr{G}_s] = B_1 - B_s - \frac{u-s}{1-s}\,(B_1 - B_s) = \frac{1-u}{1-s}\,(B_1 - B_s)$ for $u \geq s$. By Fubini's theorem (of which the application is easily justified),

$$\begin{aligned}
\mathbb{E}[\beta_t \,|\, \mathscr{G}_s] &= \mathbb{E}[B_t \,|\, \mathscr{G}_s] - \int_s^t \frac{\mathbb{E}[(B_1 - B_u) \,|\, \mathscr{G}_s]}{1-u}\,\mathrm{d}u - \int_0^s \frac{B_1 - B_u}{1-u}\,\mathrm{d}u \\
&= B_s + \frac{t-s}{1-s}\,(B_1 - B_s) - \int_s^t \frac{1}{1-u}\,\frac{1-u}{1-s}\,(B_1 - B_s)\,\mathrm{d}u - \int_0^s \frac{B_1 - B_u}{1-u}\,\mathrm{d}u,
\end{aligned}$$

which is nothing else but β_s. $\square$

Exercice 18. The trajectories of γ are $\mathbb{P}$-continuous and thus also $\mathbb{Q}$-continuous (the two probabilities being equivalent on $\mathscr{F}_1$). It remains to check that for $0 := t_0 < t_1 < \cdots < t_n \leq 1$, $B_{t_n} - B_{t_{n-1}},\, \cdots,\, B_{t_2} - B_{t_1},\, B_{t_1}$ are independent Gaussian random variables under $\mathbb{Q}$. We consider the characteristic function. Let $(x_1,\, \cdots,\, x_n) \in \mathbb{R}^n$. Then

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[\mathrm{e}^{i\sum_{k=1}^n x_k(\gamma_{t_k} - \gamma_{t_{k-1}})}] &= \mathbb{E}[\mathrm{e}^{aB_1 - \frac{a^2}{2} + i\sum_{k=1}^n x_k(B_{t_k} - B_{t_{k-1}})}] \\
&= \mathrm{e}^{-\frac{a^2}{2} - ia\sum_{k=1}^n x_k(t_k - t_{k-1})}\,\mathbb{E}[\mathrm{e}^{a(B_1 - B_{t_n}) + \sum_{k=1}^n (ix_k + a)(B_{t_k} - B_{t_{k-1}})}]\,,
\end{aligned}$$

which is

$$= \mathrm{e}^{-\frac{a^2}{2} - ia\sum_{k=1}^n x_k(t_k - t_{k-1})}\,\mathrm{e}^{\frac{a^2}{2}(1-t_n) + \sum_{k=1}^n \frac{(ix_k + a)^2}{2}\,(t_k - t_{k-1})} = \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^n x_k^2(t_k - t_{k-1})}\,.$$

This implies (i) the desired independence under $\mathbb{Q}$, and (ii) that the law of $\gamma_{t_k} - \gamma_{t_{k-1}}$ under $\mathbb{Q}$ is Gaussian $\mathscr{N}(0,\, t_k - t_{k-1})$. $\square$

Tsinghua University *2017–2018*
"Advanced Probability" (Part III: Brownian motion)

Exercise sheet #III.2:
Brownian motion and the Markov property

Exercice 1. Fix $A_2 \in \mathscr{A}_2, \cdots, A_n \in \mathscr{A}_n$. Consider

$$\mathscr{M}_1 := \{C_1 \in \sigma(\mathscr{A}_1) : \mathbb{P}(C_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(C_1)\, \mathbb{P}(A_2) \cdots \mathbb{P}(A_n)\}.$$

It is easily checked by definition that $\mathscr{M}_1$ is a λ-system[2], whereas by assumption, $\mathscr{A}_1 \subset \mathscr{M}_1$, et $\mathscr{A}_1$ is a π-system. So by the π-λ theorem, $\mathscr{M}_1 = \sigma(\mathscr{A}_1)$; in other words,

$$\mathbb{P}(C_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(C_1)\, \mathbb{P}(A_2) \cdots \mathbb{P}(A_n), \qquad \forall C_1 \in \sigma(\mathscr{A}_1),\ \forall A_2 \in \mathscr{A}_2, \cdots, \forall A_n \in \mathscr{A}_n.$$

To continue, let us fix $C_1 \in \sigma(\mathscr{A}_1)$, $A_3 \in \mathscr{A}_3, \cdots, A_n \in \mathscr{A}_n$, and consider

$$\mathscr{M}_2 := \{C_2 \in \sigma(\mathscr{A}_2) : \mathbb{P}(C_1 \cap C_2 \cap A_3 \cap \cdots \cap A_n) = \mathbb{P}(C_1)\, \mathbb{P}(C_2)\, \mathbb{P}(A_3) \cdots \mathbb{P}(A_n)\}.$$

Again, $\mathscr{M}_2$ is a λ-system, and we have proved in the previous step that it contains the π-system $\mathscr{A}_2$. Hence $\mathscr{M}_2 = \sigma(\mathscr{A}_2)$. Iterating the procedure, we arrive at:

$$\mathbb{P}(C_1 \cap \cdots \cap C_n) = \mathbb{P}(C_1) \cdots \mathbb{P}(C_n), \qquad \forall C_1 \in \sigma(\mathscr{A}_1), \cdots, \forall C_n \in \sigma(\mathscr{A}_n),$$

which means that $\sigma(\mathscr{A}_1), \cdots, \sigma(\mathscr{A}_n)$ are independent. $\square$

Exercice 2. In both situations, it is easily checked that the process is centered Gaussian with covariance $s \wedge t$. For time reversal, the continuity of trajectories is obvious. For time inversion, one may feel that there could be a continuity problem at 0: this however, does not cause any trouble because X is, according to Kolmogorov's criterion, undistinguishable to Brownian motion. $\square$

Exercice 3. By scaling, $\inf_{t\in[0,\,2]} B_t$ has the same distribution as $2^{1/2} \inf_{t\in[0,\,1]} B_t$. $\square$

Exercice 4. **(i)** Let $0 \le t_1 < t_2 < \cdots < t_n \le 1$. Then $(b_{t_1}, \cdots, b_{t_n}, B_1)$ is a Gaussian random vector, with $\mathrm{Cov}(b_{t_i}, B_1) = \mathrm{Cov}(B_{t_i}, B_1) - \mathrm{Cov}(t_i B_1, B_1) = t_i - t_i = 0$, $\forall i$. So a property of Gaussian vectors tells us that $(b_{t_1}, \cdots, b_{t_n})$ is independent of B_1.

(ii)–(iii) By checking covariance. $\square$

[2]The assumption $\Omega \in \mathscr{A}_1$ is used here to guarantee $\Omega \in \mathscr{M}_1$.

Exercice 5. By continuity, $\lim_{t\to 0+} B_t = 0$, a.s., which yields the desired conclusion by time inversion. $\square$

Exercice 6. Let $A_n := \{B_{t_n} > 0\}$. We have $\mathbb{P}(A_n) = \frac{1}{2}$, $\forall n$, so $\mathbb{P}(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} \downarrow \mathbb{P}(\cup_{k\ge n} A_k) \ge \limsup_{n\to\infty} \mathbb{P}(A_n) = \frac{1}{2}$. On the other hand, by Blumenthal's 0–1 law, we know that $\mathbb{P}(\limsup_{n\to\infty} A_n)$ is either 0 or 1; so $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$. In other words, a.s., $B_{t_n} > 0$ for infinitely many n.

By considering $-B$ which is also Brownian motion, we see that a.s., $B_{t_n} < 0$ for infinitely many n. $\square$

Exercice 7. By scaling, for any fixed $t > 0$, $(\int_0^t \mathrm{e}^{B_s}\,\mathrm{d}s)^{1/t^{1/2}}$ is distributed as

$$\Big(t\int_0^1 \mathrm{e}^{t^{1/2}B_u}\,\mathrm{d}u\Big)^{1/t^{1/2}} = \exp\Big(\frac{\log t}{t^{1/2}} + \frac{1}{t^{1/2}}\log\int_0^1 \mathrm{e}^{t^{1/2}B_u}\,\mathrm{d}u\Big).$$

The continuity of trajectories of B implies that $\frac{1}{t^{1/2}}\log\int_0^1 \mathrm{e}^{t^{1/2}B_u}\,\mathrm{d}u \to \sup_{u\in[0,1]} B_u$ a.s., so $\exp(\frac{\log t}{t^{1/2}} + \frac{1}{t^{1/2}}\log\int_0^1 \mathrm{e}^{t^{1/2}B_u}\,\mathrm{d}u) \to \exp(\sup_{u\in[0,1]} B_u)$ a.s.

As a consequence, $(\int_0^t \mathrm{e}^{B_s}\,\mathrm{d}s)^{1/t^{1/2}} \to \exp(\sup_{u\in[0,1]} B_u)$ in law; the limit is distributed as $\mathrm{e}^{|N|}$ (by the reflection principle). $\square$

Exercice 8. **(i)** It suffices to recall that $\frac{B_t}{t} \to 0$ a.s. for $t\to\infty$ and that $\limsup_{t\to 0} \frac{B_t}{t^{1/2}} = \infty$ a.s..

(ii) Let $x > 0$. We have $\mathbb{P}\{\sup_{t\ge 0}(B_t - t) < x\} = \mathbb{P}\{B_t - t < x,\ \forall t \ge 0\}$. By scaling, the probability is

$$\begin{aligned} &= \mathbb{P}\{x^{1/2}B_{t/x} - t < x,\ \forall t\ge 0\} \\ &= \mathbb{P}\{x^{1/2}B_s - sx < x,\ \forall s\ge 0\} \\ &= \mathbb{P}\{\frac{B_s}{1+s} < x^{1/2},\ \forall s\ge 0\}, \end{aligned}$$

from which the desired identity in law follows.

(iii) By (ii), it suffices to check $\mathbb{E}\{[\sup_{t\ge 0}\frac{|B_t|}{1+t}]^{2p}\} < \infty$.

By the reflection principle, $\mathbb{E}\{[\sup_{t\in[0,\,1]} B_t]^{2p}\} < \infty$. By symmetry, $\mathbb{E}\{[\sup_{t\in[0,\,1]}(-B_t)]^{2p}\} < \infty$. So $\mathbb{E}\{[\sup_{t\in[0,\,1]}|B_t|]^{2p}\} < \infty$. A fortiori, $\mathbb{E}\{[\sup_{t\in[0,\,1]}\frac{|B_t|}{1+t}]^{2p}\} < \infty$.

It remains to check $\mathbb{E}\{[\sup_{t\ge 1}\frac{|B_t|}{1+t}]^{2p}\} < \infty$. We have seen that $\mathbb{E}\{[\sup_{t\in[0,\,1]}|B_t|]^{2p}\} < \infty$. By inversion of time, this yields $\mathbb{E}\{[\sup_{t\ge 1}\frac{|B_t|}{t}]^{2p}\} < \infty$. A fortiori, $\mathbb{E}\{[\sup_{t\ge 1}\frac{|B_t|}{1+t}]^{2p}\} < \infty$.

(iv) We assume $0 < \mathbb{E}(T) < \infty$ (because otherwise, there is nothing to prove).

By scaling, $\mathbb{E}(|B_T| - aT) = \mathbb{E}(\frac{1}{a}|B_{a^2T}| - aT) = \frac{1}{a}\,\mathbb{E}(|B_{a^2T}| - a^2T)$, which is obviously bounded by $\frac{1}{a}\,\mathbb{E}[\sup_{t\ge 0}(|B_t| - t)]$.

So $\mathbb{E}(|B_T|) \le \frac{K}{a} + a\,\mathbb{E}(T)$, with $K := \mathbb{E}[\sup_{t\ge 0}(|B_t| - t)] \in (0,\,\infty)$. Since this holds for all $a > 0$, we take $a := [\frac{K}{\mathbb{E}(T)}]^{1/2}$ to see that $\mathbb{E}(|B_T|) \le 2\,[K\,\mathbb{E}(T)]^{1/2}$. $\square$

Exercice 9. Put $\beta_s := B_{s+1} - B_1$, $s \geq 0$. By the Markov property, β is Brownian motion, independent of $\mathscr{F}_1$, a fortiori of (S_1, B_1).

Write $\widetilde{S}_t := \sup_{s\in[0,t]} \beta_s$. Then $\sup_{s\in[1,2]} B_s = \widetilde{S}_1 + B_1$; hence $S_2 = \max\{S_1,\, \widetilde{S}_1 + B_1\}$. In other words, $S_2 - S_1 = \max\{0,\, \widetilde{S}_1 - (S_1 - B_1)\}$. Since $\widetilde{S}_1$ and $S_1 - B_1$ are independent (see the previous paragraph), both having the law of $|B_1|$ (by the reflection principle, the desired identity in law follows. $\square$

Exercice 10. **(i)** Fix $t \geq 0$. Let us check $\{d_1 \leq t\} \in \mathscr{F}_t$.

If $t < 1$, then $\{d_1 \leq t\} = \varnothing \in \mathscr{F}_t$. If $t \geq 1$, we have

$$\{d_1 \leq t\} = \Big\{ \inf_{s\in[1,t]\cap\mathbb{Q}} |B_s| = 0 \Big\} \in \mathscr{F}_t.$$

Conclusion: d_1 is a stopping time.

(ii) Let $t \geq 1$. Applying the Markov property at time 1, we get

$$\mathbb{P}\{d_1 \leq t\} = \int_{-\infty}^{\infty} \mathbb{P}\{B_1 \in \mathrm{d}x\}\, \mathbb{P}\{T_{-x} \leq t-1\}.$$

Let N and $\widetilde{N}$ be independent Gaussian $\mathscr{N}(0,\,1)$ random variables. We know that T_{-x} is distributed as $\frac{x^2}{N^2}$. Hence

$$\mathbb{P}\{d_1 \leq t\} = \mathbb{P}\Big(\frac{\widetilde{N}^2}{N^2} \leq t-1\Big).$$

As consequence, $(d_1 - 1)^{1/2}$ has the standard Cauchy distribution. In other words,

$$\mathbb{P}(d_1 \in \mathrm{d}t) = \frac{1}{\pi}\, \frac{\mathbf{1}_{\{t>1\}}}{t(t-1)^{1/2}}\, \mathrm{d}t.$$

Let us now study the law of g_1. For all $t \in [0,\,1)$,

$$\begin{aligned} \mathbb{P}(g_1 \leq t) &= \int_{-\infty}^{\infty} \mathbb{P}\{B_t \in \mathrm{d}x\}\, \mathbb{P}\{T_{-x} > 1-t\} \\ &= \mathbb{P}\Big(\frac{t\widetilde{N}^2}{N^2} > 1-t\Big) \\ &= \mathbb{P}\Big(\frac{1}{1+(\widetilde{N}/N)^2} < t\Big). \end{aligned}$$

Thus g_1 is distributed as $\frac{1}{1+C^2}$, where C is a standard Cauchy random variable. We have

$$\mathbb{P}(g_1 \in \mathrm{d}t) = \frac{1}{\pi}\, \frac{\mathbf{1}_{\{0<t<1\}}}{t(1-t)^{1/2}}\, \mathrm{d}t.$$

We say that g_1 has the **Arcsine law**, because $\mathbb{P}(g_1 \leq t) = \frac{2}{\pi}\arcsin(t^{1/2})$.

Observe that we could have determined the law of g_1 from the law of d_1 by means of the scaling property: $\{g_1 < t\} = \{d_t > 1\}$, where $d_t := \inf\{s \geq t :\, B_s = 0\}$ has the same law as td_1. $\square$

Exercice 11. **(i)** Let us first prove that for any finite stopping time $T \geq 0$, $\tau = \inf\{t \geq T : B_t = 0\}$ is a stopping time. This was proved in the previous exercise when T is a constant. If T takes countably many values, say (t_n), then

$$\{\tau \leq t\} = \bigcup_{n:\, t_n \leq t} \{T = t_n\} \cap \Big\{ \inf_{s \in [t_n, t] \cap \mathbb{Q}} |B_s| = 0 \Big\} \in \mathscr{F}_t \,,$$

which means τ is a stopping time.

In the general case, for all n, let

$$T_n := \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbf{1}_{\{\frac{k}{2^n} < T \leq \frac{k+1}{2^n}\}},$$

which is a non-increasing stopping times tending to T. By what we have just proved, $\tau_n := \inf\{t \geq T_n : B_t = 0\}$ is a stopping time; hence

$$\{\tau \leq t\} = \Big(\{T \leq t\} \cap \{B_T = 0\} \Big) \cup \Big(\{T \leq t\} \cap \{B_T \neq 0\} \cap \bigcup_{n=1}^{\infty} \{\tau_n \leq t\} \Big),$$

which is an element of $\mathscr{F}_t$. As a conclusion, τ is a stopping time.

(ii) By the strong Markov property, τ is distributed as $T_1 + \widetilde{T}_{-1}$, where $\widetilde{T}_{-1}$ is an independent copy of T_1. So τ is distributed as T_2, thus also as $4T_1$. The density of τ is

$$\mathbb{P}(\tau \in \,\mathrm{d}t) = (\frac{2}{\pi t^3})^{1/2} \exp\Big(-\frac{2}{t} \Big) \,\mathrm{d}t,$$

for $t > 0$. □

Exercice 12. **(i)** By scaling, for all fixed $t \geq 0$, $\log(1 + B_t^2)$ has the same distribution as $\log(1 + tB_1^2)$. Since $B_1 \neq 0$ a.s., we have $\frac{\log(1+tB_1^2)}{\log t} \to 1$ a.s. So $\frac{\log(1+B_t^2)}{\log t} \to 1$ in law. The limit being a constant, the convergence holds also in probability. Conclusion: $\frac{\log(1+B_t^2)}{\log t} \to 1$ in probability.

(ii) If $\frac{\log(1+B_t^2)}{\log t}$ converged a.s., it would converge a.s. to 1. But $\{t : \; B_t = 0\}$ is a.s. unbounded, which makes it impossible to converge a.s. to 1. Conclusion: $\frac{\log(1+B_t^2)}{\log t}$ does not converge a.s. □

Exercice 13. By the strong law of large numbers, $\frac{B_n}{n} \to 0$ a.s. for $n \to \infty$. It remains to check $\frac{1}{n} \sup_{t \in [n,\, n+1]} |B_t - B_n| \to 0$ a.s.

Let $\varepsilon > 0$. Let $A_n := \{\sup_{t \in [n,\, n+1]} |B_t - B_n| > n^{\varepsilon}\}$. We have $\mathbb{P}(A_n) = \mathbb{P}(\sup_{s \in [0,\, 1]} |B_s| > n^{\varepsilon}) \leq 2\,\mathbb{P}(\sup_{s \in [0,\, 1]} B_s > n^{\varepsilon})$. By the reflection principle, $\sup_{s \in [0,\, 1]} B_s$ is distributed as $|B_1|$. So $\mathbb{P}(A_n) \leq 2\,\mathbb{P}(|B_1| > n^{\varepsilon}) = 4\,\mathbb{P}(B_1 > n^{\varepsilon}) \leq 2 \exp(-\frac{n^{2\varepsilon}}{2})$, which yields $\sum_n \mathbb{P}(A_n) < \infty$. By the Borel–Cantelli lemma, $\limsup_{n \to \infty} n^{-\varepsilon} \sup_{t \in [n,\, n+1]} |B_t - B_n| \leq 1$ a.s. A fortiori, $\frac{1}{n} \sup_{t \in [n,\, n+1]} |B_t - B_n| \to 0$ a.s. □

Exercice 14. Ken's argument is wrong, because T is not $\mathscr{F}_{0+}$-measurable. As a matter of fact, whenever $t > 0$, T is not $\mathscr{F}_t$-measurable.

To prove $T < \infty$ a.s., it suffices to recall that $\limsup_{t\to\infty} \frac{B_t}{t^{1/2}} = \infty$ a.s. $\square$

Exercice 15. (i) We define inductively two sequences of stopping times $(\tau_i)_{i\geq 1}$ and $(T_i)_{i\geq 1}$ as follows: $\tau_1 := 0$, $T_i := \inf\{t > \tau_i : |B_t| = 1\}$ and $\tau_{i+1} := \inf\{t > T_i : B_t = 0\}$ for $i \geq 1$. The strong Markov property tells us that $\int_{\tau_i}^{T_i} \sin^2(B_t)\, \mathrm{d}t$, $i \geq 1$, are i.i.d. In particular, $\sum_{i\geq 1} \int_{\tau_i}^{T_i} \sin^2(B_t)\, \mathrm{d}t = \infty$ a.s. A fortiori, $\int_0^\infty B_t^2\, \mathrm{d}t \geq \sum_{i\geq 1} \int_{\tau_i}^{T_i} \sin^2(B_t)\, \mathrm{d}t = \infty$ a.s.

(ii) Same argument as in (i), replacing $\inf\{t > \tau_i : |B_t| = 1\}$ by $\inf\{t > \tau_i : |B_t| = a\}$, where $a > 0$ is such that $f^2(x) \in (0, a)$. $\square$

Exercice 16. That $\mathscr{Z}$ is a closed set comes from the continuity of $t \mapsto B_t$. We have also seen in the class that $\mathscr{Z}$ is a.s. unbounded. It remains to show that $\mathscr{Z}$ has a.s. no isolated point.

For $t \geq 0$, let $\tau_t := \inf\{s \geq t : B_s = 0\}$ which is a stopping time. Clearly, $\tau_t < \infty$ a.s., and $B_{\tau_t} = 0$. The strong Markov property telles us that τ_t is not an isolated zero point of B. So a.s. for all $r \in \mathbb{Q}_+$, τ_r is not an isolated zero point.

Let $t \in \mathscr{Z} \backslash \{\tau_r,\, r \in \mathbb{Q}_+\}$. It suffices to show that t is not an isolated zero point. Consider a rational sequence $(r_n) \uparrow\uparrow t$. Clearly, $r_n \leq \tau_{r_n} < t$. So $\tau_{r_n} \to t$; thus t is not an isolated zero point.[3] $\square$

Exercice 17. **(i)** Let $b < c$. By the Markov property, $\sup_{t\in[c,d]} B_s - B_c$ is independent of $(B_c, \sup_{t\in[a,b]} B_s)$, and is distributed as $(d-c)^{1/2}\,|N|$, with N denoting a standard Gaussian $\mathscr{N}(0,\, 1)$ random variable. Since $\mathbb{P}(N = x) = 0$ for all $x \in \mathbb{R}$, we obtain the desired result.

(ii) By (i), a.s. for all non-negative rationals $a < b < c < d$, $\sup_{t\in[a,b]} B_s \neq \sup_{t\in[c,d]} B_s$. If B had a non strict local maximum, there would be two disjoint closed intervals with rational extremity points, on which B would have the same maximal value, which is impossible.

(iii) Let M denote the set of times at which B realises the local minima. Consider the mapping:

$$[a, b] \mapsto \inf\Big\{t \geq a : B_t = \sup_{s\in[a,b]} B_s\Big\},$$

for all rationals $0 \leq a < b$. According to (i), the image of this mapping contains M a.s., so M is a.s. countable.

Since a.s. there exists no interval on which B is monotone (because B is nowhere differentiable), B admits a local maximum on each interval with rational extremity points: M is a.s. dense. $\square$

[3] It is known in analysis (see page 72 of the book by Hewitt, E. and Stromberg, K.: *Real and Abstract Analysis*. Springer, New York, 1969) that a closed set with no isolated point is uncountable. So $\mathscr{Z}$ is a.s. uncountable.

Exercice 18. **(i)** Let $\lambda > 0$. We have $\mathbb{P}(T_a \le t) = \mathbb{P}(\mathrm{e}^{-\lambda T_a} \ge \mathrm{e}^{-\lambda t}) \le \mathrm{e}^{\lambda t}\,\mathbb{E}(\mathrm{e}^{-\lambda T_a}) = \mathrm{e}^{\lambda t - a(2\lambda)^{1/2}}$.

Choosing $\lambda := \frac{a^2}{2t^2}$ yields the desired inequality.

(ii) Let $S_1 := \sup_{s\in[0,\,1]} B_s$. By (i), we have, for all $a > 0$, $\mathbb{P}(S_1 \ge a) = \mathbb{P}(T_a \le 1) \le \mathrm{e}^{-a^2/2}$. According to the reflection principle, S_1 has the law of the modulus of a standard Gaussian random variable: the desired conclusion follows immediately. $\square$

Exercice 19. (i) By scaling, $\mathbb{P}\{\sup_{s\in[0,\,t]} |B_s| \le \varepsilon\} = \mathbb{P}\{\sup_{s\in[0,\,\frac{4t}{\varepsilon^2}]} |B_s| \le 2\}$. So it suffices to check that for all $a > 0$, $\mathbb{P}\{\sup_{s\in[0,\,a]} |B_s| \le 2\} > 0$.

Let $T^* := \inf\{t \ge 0 : |B_t| = 1\}$. Let $\delta > 0$ be such that $p := \mathbb{P}\{T^* > \delta\} > 0$. By symmetry, $\mathbb{P}\{T^* > \delta,\ B_{T^*} = 1\} = \mathbb{P}\{T^* > \delta,\ B_{T^*} = -1\} = \frac{p}{2} > 0$. It follows from the strong Markov property that $\mathbb{P}\{\sup_{s\in[0,\,a]} |B_s| \le 2\} \ge (\frac{p}{2})^N > 0$, where $N := \lceil \frac{a}{\delta} \rceil$.

(ii) Already proved in (i).

(iii) We have $\mathbb{P}\{\sup_{s\in[0,\,t]} |B_s| \ge x\} \ge \mathbb{P}\{B_t \ge x\} = \mathbb{P}\{B_1 \ge \frac{x}{t^{1/2}}\} > 0$, as B_1 is a standard Gaussian random variable. $\square$

Exercice 20. **(i)** Let $A_n := \{S_{t_{n+1}} \ge (1+\varepsilon)h(t_n)\}$. We have

$$\mathbb{P}(A_n) = \mathbb{P}\Big(|B_1| \ge [2(1+\varepsilon)\log\log t_n]^{1/2}\Big) \le 2\exp\Big(-(1+\varepsilon)\log\log t_n\Big),$$

as $\mathbb{P}(N \ge x) \le \mathrm{e}^{-x^2/2}$ for all $x \ge 0$. Hence $\sum \mathbb{P}(A_n) < \infty$. By the Borel–Cantelli lemma, there exists $A \in \mathscr{F}$ with $\mathbb{P}(A) = 1$ such that for all $\omega \in A$, $\exists\, n_0 = n_0(\omega) < \infty$,

$$n \ge n_0 \implies S_{t_{n+1}} < (1+\varepsilon)(2t_n \log\log t_n)^{1/2}.$$

Therefore, for $t \in [t_n, t_{n+1}]$,

$$S_t \le S_{t_{n+1}} < (1+\varepsilon)(2t_n\log\log t_n)^{1/2} \le (1+\varepsilon)(2t\log\log t)^{1/2},$$

which implies $\limsup_{t\to\infty} \frac{S_t}{h(t)} \le 1+\varepsilon$, a.s. It suffices now to let $\varepsilon \to 0$ along a sequence of rational numbers to reach the desired conclusion.

(ii) Since $-B$ is also Brownian motion, it follows from (i) that $\limsup_{t\to\infty} \frac{\sup_{s\in[0,t]}(-B_s)}{h(t)} \le 1$, a.s. The desired result follows.

(iii) Let $E_n := \{B_{s_n} - B_{s_{n-1}} > \alpha h(s_n)\}$. The events (E_n) are independent. Furthermore,

$$\begin{aligned}\mathbb{P}(E_n) &= \mathbb{P}\Big(B_1 > \alpha(\frac{2\log\log s_n}{1-\theta^{-1}})^{1/2}\Big)\\ &\sim \frac{1}{(2\pi)^{1/2}}\frac{1}{\alpha[2(\log\log s_n)/(1-\theta^{-1})]^{1/2}}\exp\Big(-\frac{\alpha^2\log\log s_n}{1-\theta^{-1}}\Big),\end{aligned}$$

which yields $\sum_n \mathbb{P}(E_n) = \infty$ (because $\alpha < (1-\theta^{-1})^{1/2}$). By the Borel–Cantelli lemma, there exists $E \in \mathscr{F}$ with $\mathbb{P}(E) = 1$ such that for all $\omega \in E$,

$$B_{s_n} - B_{s_{n-1}} > \alpha(2s_n\log\log s_n)^{1/2}, \qquad \text{for infinitely many } n.$$

On the other hand, by (ii), a.s. for all sufficiently large n,

$$|B_{s_{n-1}}| \le 2(2s_{n-1}\log\log s_{n-1})^{1/2} \le \frac{2}{\theta^{1/2}}\,(2s_n\log\log s_n)^{1/2}\,.$$

The desired inequality follows.

(iv) By (iii), $\limsup_{t\to\infty}\frac{B_t}{h(t)} \ge 1$ a.s., which, together with (i), implies the desired result.

(v) The "limsup" expression is 1 a.s. (for all i).

(vi) By symmetry, $\liminf_{t\to\infty}\frac{B_t}{h(t)} = -1$ a.s.

By inversion of time, $\limsup_{t\to 0}\frac{B_t}{(2t\log\log(1/t))^{1/2}} = 1$ a.s. □

Exercice 21. Let $t>0$. We have

$$(P_tf)(x) = \frac{1}{(2\pi)^{1/2}}\int_{\mathbb{R}} f(x+t^{1/2}z)\mathrm{e}^{-z^2/2}\,\mathrm{d}z.$$

By the dominated convergence theorem (because f is bounded and continuous), we have $P_tf\in C_0$.

Let us prove that $\lim_{t\downarrow 0}P_tf = f$ uniformly on $\mathbb{R}$. Write

$$(P_tf)(x)-f(x) = \frac{1}{(2\pi)^{1/2}}\int_{\mathbb{R}}\mathrm{e}^{-z^2/2}[f(x+t^{1/2}z)-f(x)]\,\mathrm{d}z.$$

(The dominated convergence theorem allows us immediately to see that $P_tf\to f$ pointwise.) Let $\varepsilon>0$. Since f is bounded, there exists $M>0$ such that $\int_{|z|>M}\mathrm{e}^{-z^2/2}\|f\|_\infty\,\mathrm{d}z<\varepsilon$. For $|z|\le M$, as f is uniformly continuous on $\mathbb{R}$, there exists $\delta>0$ such that for $t\le\delta$, we have $\sup_{|z|\le M}|f(x+t^{1/2}z)-f(x)|\le\varepsilon$, $\forall x\in\mathbb{R}$. Consequently, for all $t\le\delta$, $|P_tf(x)-f(x)|\le \frac{2\varepsilon}{(2\pi)^{1/2}}+\varepsilon\le 2\varepsilon$, $\forall x\in\mathbb{R}$. □

Exercice 22. Write

$$\frac{(P_tf)(x)-f(x)}{t} = \frac{1}{(2\pi)^{1/2}}\int_0^\infty \frac{f(x+t^{1/2}z)+f(x-t^{1/2}z)-2f(x)}{t}\,\mathrm{e}^{-z^2/2}\,\mathrm{d}z.$$

We let $t\to 0$. Since $f\in C^2$, we have $\frac{f(x+t^{1/2}z)+f(x-t^{1/2}z)-2f(x)}{t}\to z^2f''(x)$, and there exists a constant $K<\infty$ such that for all $t\le 1$, $\frac{f(x+t^{1/2}z)+f(x-t^{1/2}z)-2f(x)}{t}\le Kz^2$ (we use, moreover, the assumption that f is of compact support). Since $z^2\mathrm{e}^{-z^2/2}$ is integrable, it follows from the dominated convergence theorem that $\frac{(P_tf)(x)-f(x)}{t}\to\frac{1}{(2\pi)^{1/2}}\int_0^\infty z^2f''(x)\,\mathrm{e}^{-z^2/2}\,\mathrm{d}z = \frac{1}{2}f''(x)$. □

Exercice 23. Fix $t>0$ and $x\in\mathbb{R}$. We have

$$u(t,x) = \frac{1}{(2\pi)^{1/2}}\int_{\mathbb{R}} f(r)\,\frac{1}{t^{1/2}}\,\exp\Big(-\frac{(r-x)^2}{2t}\Big)\,\mathrm{d}r.$$

Since f is bounded, we can use the dominated convergence theorem to take the partial derivative (with respect to t) under the integral sign:

$$\frac{\partial u(t,\, x)}{\partial t} = \frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} f(r) \Big(-\frac{1}{2t^{3/2}} + \frac{(r-x)^2}{2t^{5/2}} \Big) \exp\Big(-\frac{(r-x)^2}{2t} \Big) \, \mathrm{d}r.$$

Similarly, thanks again to the boundedness of f and to the dominated convergence theorem, we can take the second partial derivative (with respect to x) under the integral sign, to see that

$$\frac{\partial^2 u(t,\, x)}{\partial x^2} = \frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} f(r) \, \frac{1}{t^{1/2}} \Big(-\frac{1}{t} + \frac{(r-x)^2}{t^2} \Big) \exp\Big(-\frac{(r-x)^2}{2t} \Big) \, \mathrm{d}r.$$

It is readily observed that $\frac{\partial u(t,x)}{\partial t} = \frac{1}{2} \frac{\partial^2 u(t,x)}{\partial x^2}$. $\square$

Tsinghua University *2017–2018*

"Advanced Probability" (Part III: Brownian motion)

Exercise sheet #III.3:
Brownian motion and martingales

Exercice 1. Let $t > 0$. Then $\mathbb{P}(T_a \le t) = \mathbb{P}(\sup_{s\in[0,\,t]} |B_s| \ge a)$, which, by scaling, equals to $\mathbb{P}(t^{1/2} \sup_{u\in[0,\,1]} |B_u| \ge a)$. As such, T_a and $\frac{a^2}{\sup_{s\in[0,\,1]} B_s^2}$ have the same distribution function: they have the same law. $\square$

Exercice 2. **(i)** Without loss of generality, we may assume $\xi = 0$ (otherwise, we replace η by $\eta + \xi$). We need to prove that $\mathbb{E}(\eta \,|\, \mathscr{G}) \ge 0$ a.s. $\Leftrightarrow$ $\mathbb{E}(\eta \, \mathbf{1}_A) \ge 0$, $\forall A \in \mathscr{G}$.

"$\Rightarrow$" Assume $\mathbb{E}(\eta \,|\, \mathscr{G}) \ge 0$ a.s. Then for all $A \in \mathscr{G}$, we have, by the definition of conditional expectation, $\mathbb{E}(\eta \, \mathbf{1}_A) = \mathbb{E}[\, \mathbf{1}_A \, \mathbb{E}(\eta \,|\, \mathscr{G})]$, which is non-negative because by assumption, $\mathbb{E}(\eta \,|\, \mathscr{G}) \ge 0$ a.s.

"$\Leftarrow$" Assume $\mathbb{E}(\eta \, \mathbf{1}_A) \ge 0$, $\forall A \in \mathscr{G}$.

Write $\theta := \mathbb{E}(\eta \,|\, \mathscr{G})$ which is $\mathscr{G}$-mesurable. Let $B := \{\omega : \theta(\omega) < 0\} \in \mathscr{G}$. By assumption, $\mathbb{E}(\eta \, \mathbf{1}_B) \ge 0$. We observe that $\mathbb{E}(\eta \, \mathbf{1}_B) = \mathbb{E}[\, \mathbb{E}(\eta \, \mathbf{1}_B \,|\, \mathscr{G})] = \mathbb{E}[\, \mathbf{1}_B \, \mathbb{E}(\eta \,|\, \mathscr{G})] = \mathbb{E}[\, \mathbf{1}_B \, \theta]$; as such, saying that $\mathbb{E}(\eta \, \mathbf{1}_B) \ge 0$ is equivalent to saying that $\mathbb{E}[\, \mathbf{1}_B \, \theta] \ge 0$. Since $\mathbf{1}_B \, \theta \le 0$, this is possible only if $\mathbf{1}_B \, \theta = 0$ a.s., i.e., $\theta \ge 0$ a.s.

(ii) It is a consequence of (i), by considering the pair $(-\xi, -\eta)$ in place of $(-\xi, -\eta)$. $\square$

Exercice 3. "$\Leftarrow$" Without loss of generality, we may assume $X_\infty = 0$ (otherwise, we consider $X_n - X_\infty$ in place of X_t, by observing that $(X_n - X_\infty,\, t \ge 0)$ is also uniformly integrable).

Let $\varepsilon > 0$. We fix $a > 0$ sufficiently large such that $\mathbb{E}(|X_n| \, \mathbf{1}_{\{|X_n|>a\}}) < \varepsilon$, $\forall n \ge 0$. Then $\mathbb{E}(|X_n|) = \mathbb{E}(|X_n| \, \mathbf{1}_{\{\varepsilon \le |X_n| \le a\}}) + \mathbb{E}(|X_n| \, \mathbf{1}_{\{|X_n|>a\}}) + \mathbb{E}(|X_n| \, \mathbf{1}_{\{|X_n|<\varepsilon\}}) \le a\mathbb{P}(|X_n| \ge \varepsilon) + \varepsilon + \varepsilon$. Letting $n \to \infty$, and since $X_n \to 0$ in probability, we get $\limsup_{n\to\infty} \mathbb{E}(|X_n|) \le 2\varepsilon$, which yields $X_t \to 0$ in L^1 because $\varepsilon > 0$ can be as small as possible.

"$\Rightarrow$" Assume that $X_n \to X_\infty$ in L^1.

Convergence in probability follows immediately from convergence in L^1. To prove that $(X_n,\, n \ge 0)$ is uniformly integrable, it suffices to check (a) $\sup_{n\ge1} \mathbb{E}(|X_n|) < \infty$; (b) for all $\varepsilon > 0$, there exists $\delta > 0$ such that $\forall B \in \mathscr{F}$, $\mathbb{P}(B) < \delta \Rightarrow \sup_{n\ge1} \mathbb{E}(|X_n| \, \mathbf{1}_B) < \varepsilon$.

Condition (a) is a straightforward consequence of convergence in L^1. Let us check condition (b). Let $B \in \mathscr{F}$. We have $\mathbb{E}(|X_n| \, \mathbf{1}_B) \le \mathbb{E}(|X_\infty| \, \mathbf{1}_B) + \mathbb{E}(|X_n - X_\infty|)$. Let $\varepsilon > 0$. There exists $n_0 < \infty$ such that $\mathbb{E}(|X_n - X_\infty|) < \frac{\varepsilon}{2}$, $\forall n \ge n_0$. On the other hand, there exists $\delta > 0$

sufficiently small such that if $\mathbb{P}(B) < \delta$, then $\mathbb{E}(|X_\infty|\,\mathbf{1}_B) < \frac{\varepsilon}{2}$, and $\max_{0\le n\le n_0} \mathbb{E}(|X_n|\,\mathbf{1}_B) \le \varepsilon$. Hence $\sup_{n\ge 0} \mathbb{E}(|X_n|\,\mathbf{1}_B) \le \varepsilon$ for all B with $\mathbb{P}(B) < \delta$: condition (b) is satisfied. $\square$

Exercice 4. The first part is proved using exactly the same argument as in the previous, replacing everywhere n by t.

To see the converse is not true in general, it suffices to consider an example of $(X_t,\, t \in [0,\, 1])$ that is not uniformly integrable, and let $X_t := 0$ for $t > 1$. Then $X_t \to 0$ in L^1 but $(X_t,\, t \ge 0)$ is not uniformly integrable. $\square$

Exercice 5. (i) Let $A \in \mathscr{F}_S$. Then $A \cap \{T \le t\} = (A \cap \{S \le t\}) \cap \{T \le t\} \in \mathscr{F}_t$.

(ii) We have $\{S \wedge T \le t\} = \{S \le t\} \cup \{T \le t\} \in \mathscr{F}_t$ and $\{S \vee T \le t\} = \{S \le t\} \cap \{T \le t\} \in \mathscr{F}_t$.

By (i), $\mathscr{F}_{S\wedge T} \subset \mathscr{F}_S \cap \mathscr{F}_T$. Conversely, if $A \in \mathscr{F}_S \cap \mathscr{F}_T$, then

$$A \cap \{S \wedge T \le t\} = (A \cap \{S \le t\}) \cup (A \cap \{T \le t\}) \in \mathscr{F}_t\,;$$

thus $A \in \mathscr{F}_{S\wedge T}$. Consequently, $\mathscr{F}_{S\wedge T} = \mathscr{F}_S \cap \mathscr{F}_T$.

Finally, $\{S \le T\} \cap \{T \le t\} = \{S \le t\} \cap \{T \le t\} \cap \{S \wedge t \le T \wedge t\} \in \mathscr{F}_t$, because $S \wedge t$ and $T \wedge t$ being $\mathscr{F}_{S\wedge t}$-measurable and $\mathscr{F}_{T\wedge t}$-measurable respectively, are $\mathscr{F}_t$-measurable. Hence $\{S \le T\}$ is $\mathscr{F}_T$-measurable. Similarly, $\{S \le T\} \cap \{S \le t\} = \{S \le t\} \cap \{S \wedge t \le T \wedge t\} \in \mathscr{F}_t$, which yields $\{S \le T\} \in \mathscr{F}_S$. Therefore, $\{S \le T\} \in \mathscr{F}_S \cap \mathscr{F}_T = \mathscr{F}_{S\wedge T}$.

By exchanging S and T, we have, $\{T \le S\} \in \mathscr{F}_{S\wedge T}$. Hence $\{S = T\} = \{S \le T\} \cap \{T \le S\} \in \mathscr{F}_{S\wedge T}$, and $\{S < T\} = \{S \le T\} \backslash \{S = T\} \in \mathscr{F}_{S\wedge T}$.

(iii) Since S and T are $\mathscr{F}_{S\vee T}$-measurable, so is $S+T$. We have $\{S+T \le t\} = \{S+T \le t\} \cap \{S \vee T \le t\} \in \mathscr{F}_t$, because $\{S+T \le t\} \in \mathscr{F}_{S\vee T}$. $\square$

Exercice 6. Clearly, (T_n) decreases pointwise to T. It suffices to check that each T_n is a stopping time. Since T_n is $\mathscr{F}_T$-measurable, and since $T_n \ge T$, we have $\{T_n \le t\} = \{T_n \le t\} \cap \{T \le t\} \in \mathscr{F}_t$, because $\{T_n \le t\} \in \mathscr{F}_T$. $\square$

Exercice 7. (i) It suffices to observe that for all $A \in \mathscr{B}(\mathbb{R}^d)$ with $0 \notin A$, $\{Y\,\mathbf{1}_{\{T\le t\}} \in A\} = \{Y \in A\} \cap \{T \le t\}$.

(ii) We first assume that $(X_s,\, s \ge 0)$ is right-continuous. For any $n \ge 1$, let

$$X_s^{(n)} := X_{t\wedge \frac{(\lfloor ns/t\rfloor+1)t}{n}}, \qquad s \in [0,\, t].$$

Then $X_s^{(n)}(\omega) \to X_s(\omega)$ by the right-continuity of the trajectories. For any $A \in \mathscr{B}(\mathbb{R}^d)$,

$$\begin{aligned} &\big\{(s,\omega):\ s \in [0,\, t],\, X_s^{(n)}(\omega) \in A\big\} \\ &= \bigcup_{k=1}^{n} \Big([\frac{(k-1)t}{n}, \frac{kt}{n}) \times \{X_{\frac{kt}{n}} \in A\}\Big) \cup \Big(\{t\} \times \{X_t \in A\}\Big) \\ &\in \mathscr{B}([0,\, t]) \otimes \mathscr{F}_t\,. \end{aligned}$$

Hence $(s,\omega) \mapsto X_s(\omega)$ on $[0,\, t] \times \Omega$ is $\mathscr{B}([0,\, t]) \otimes \mathscr{F}_t$-measurable.

The proof is similar if $(X_s,\, s \geq 0)$ is left-continuous; it suffices to consider instead $X_s^{(n)} := X_{\frac{\lfloor ns/t \rfloor\, t}{n}}$.

(iii) We apply (i) to $Y = X_T\, \mathbf{1}_{\{T<\infty\}}$; so it suffices to check that for all t, $Y\, \mathbf{1}_{\{T\leq t\}} = X_{T\wedge t}\, \mathbf{1}_{\{T\leq t\}}$ is $\mathscr{F}_t$-measurable.

Note that $X_{T\wedge t}$ is the composition of the following two mappings:

$$\begin{array}{rcl} (\Omega,\ \mathscr{F}_t) & \longrightarrow & ([0,\, t] \times \Omega,\ \mathscr{B}([0,\, t]) \otimes \mathscr{F}_t) \\ \omega & \longmapsto & (T(\omega) \wedge t,\ \omega) \end{array}$$

and

$$\begin{array}{rcl} ([0,t] \times \Omega,\ \mathscr{B}([0,\, t]) \otimes \mathscr{F}_t) & \longrightarrow & (\mathbb{R}^d,\ \mathscr{B}(\mathbb{R}^d)) \\ (s,\omega) & \longmapsto & X_s(\omega) \end{array}$$

both of which are measurable. So $X_{T\wedge t}$, as well as $X_{T\wedge t}\, \mathbf{1}_{\{T\leq t\}}$, are $\mathscr{F}_t$-measurable. $\square$

Exercice 8. Since $(X_t^+,\, t \geq 0)$ is a submartingale, we have $\mathbb{E}(X_s^+) \leq \mathbb{E}(X_t^+)$ for $s \leq t$. On the other hand, $\mathbb{E}(X_s) \geq \mathbb{E}(X_0)$, which implies $\sup_{s\in[0,t]} \mathbb{E}(|X_s|) \leq 2\, \mathbb{E}(X_t^+) - \mathbb{E}(X_0) < \infty$. $\square$

Exercice 9. **(i)** For any t, $\mathbb{E}(|B_t|) < \infty$ and B_t is $\mathscr{F}_t$-measurable. Let $t > s \geq 0$. Since $B_t - B_s$ is independent of $\mathscr{F}_s$, we have $\mathbb{E}(B_t - B_s \,|\, \mathscr{F}_s) = \mathbb{E}(B_t - B_s)$, which vanishes because $B_t - B_s$ has the Gaussian $\mathscr{N}(0,\, t-s)$ law. So $\mathbb{E}(B_t \,|\, \mathscr{F}_s) = B_s$ a.s.

(ii) For any t, $\mathbb{E}(B_t^2) < \infty$ and B_t^2 is $\mathscr{F}_t$-measurable. Let $t > s$, $\mathbb{E}(B_t^2 - t \,|\, \mathscr{F}_s) = \mathbb{E}[(B_t - B_s + B_s)^2 \,|\, \mathscr{F}_s)] - t$, and for all $x \in \mathbb{R}$, $\mathbb{E}[(B_t - B_s + x)^2] = \mathrm{Var}(B_t - B_s) + x^2 = t - s + x^2$, so we get $\mathbb{E}(B_t^2 - t \,|\, \mathscr{F}_s) = t - s + B_s^2 - t = B_s^2 - s$ a.s.

(iii) For any t, $\mathbb{E}(\mathrm{e}^{\theta B_t - \frac{\theta^2}{2}t}) < \infty$ and $\mathrm{e}^{\theta B_t - \frac{\theta^2}{2}t}$ is $\mathscr{F}_t$-measurable. Let $t > s$. We have $\mathbb{E}[\mathrm{e}^{\theta B_t - \frac{\theta^2}{2}t} \,|\, \mathscr{F}_s] = \mathrm{e}^{\frac{\theta^2}{2}2(t-s)} \mathrm{e}^{\theta B_s - \frac{\theta^2}{2}t} = \mathrm{e}^{\theta B_s - \frac{\theta^2}{2}s}$. $\square$

Exercice 10. Similar to the solution to the previous exercise. $\square$

Exercice 11. **(i)** Fix $t \geq 0$. Let $\xi_n := \mathbb{E}(X_n^+ \,|\, \mathscr{F}_t)$.

For $m > n \geq t$, $\xi_n = \mathbb{E}\{\, [\mathbb{E}(X_m \,|\, \mathscr{F}_n)]^+ \,|\, \mathscr{F}_t\} \leq \mathbb{E}\{\, \mathbb{E}(X_m^+ \,|\, \mathscr{F}_n) \,|\, \mathscr{F}_t\} = \mathbb{E}\{X_m^+ \,|\, \mathscr{F}_t\} = \xi_m$. So the sequence $(\xi_n)_{n\geq t}$ is a.s. non-decreasing. In particular, it converges a.s., whose limit is denoted by α_t.

By the monotone convergence theorem, $\mathbb{E}(\alpha_t) = \lim_{n\to\infty} \uparrow \mathbb{E}(\xi_n)$. We observe that $\mathbb{E}(\xi_n) = \mathbb{E}(X_n^+) \leq \sup_{t\geq 0} \mathbb{E}(|X_t|)$, which implies $\mathbb{E}(\alpha_t) \leq \sup_{t\geq 0} \mathbb{E}(|X_t|) < \infty$. In particular, $\alpha_t < \infty$ a.s.

(ii) We have seen that for any t, α_t is integrable, and is clearly $\mathscr{F}_t$-measurable (being the pointwise limit of $\mathscr{F}_t$-measurable random variables). Let us check the characteristic identity.

Let $s < t$, and let $A \in \mathscr{F}_s$. Since α_t is the limit of the non-decreasing sequence (ξ_n), it follows from the monotone convergence theorem that $\mathbb{E}(\alpha_t \mathbf{1}_A) = \lim_{n\to\infty} \uparrow \mathbb{E}(\xi_n \mathbf{1}_A)$. For $n \geq t$, we have $\mathbb{E}(\xi_n \mathbf{1}_A) = \mathbb{E}(X_n^+ \mathbf{1}_A)$, thus $\mathbb{E}(\alpha_t \mathbf{1}_A) = \lim_{n\to\infty} \uparrow \mathbb{E}(X_n^+ \mathbf{1}_A)$. Similarly, $\mathbb{E}(\alpha_s \mathbf{1}_A) = \lim_{n\to\infty} \uparrow \mathbb{E}(X_n^+ \mathbf{1}_A)$. It follows that $\mathbb{E}(\alpha_t \mathbf{1}_A) = \mathbb{E}(\alpha_s \mathbf{1}_A)$. Since $A \in \mathscr{F}_s$ is arbitrary, we deduce that $\mathbb{E}(\alpha_t \,|\, \mathscr{F}_s) = \alpha_s$ a.s.

[We note that for question (i) and (ii), it suffices to have a submartingale X satisfying $\sup_{t\geq 0} \mathbb{E}(X_t^+) < \infty$.]

(iii) By considering $-X$ in place of X, we see that $\mathbb{E}(X_n^- \,|\, \mathscr{F}_t)$ converges a.s. (when $n \to \infty$) to a limit, denoted by β_t, and that $(\beta_t,\, t \geq 0)$ is a non-negative martingale. We have $X_t = \alpha_t - \beta_t$, $\forall t \geq 0$. □

Exercice 12. Let $0 \leq s < t$. Let us check that $\mathbb{E}(X_t \,|\, \mathscr{F}_s) \geq X_s$ a.s.

By definition, $X_t \geq \mathbb{P}(\xi \leq s \,|\, \mathscr{F}_t)$; so $\mathbb{E}[X_t \,|\, \mathscr{F}_s] \geq \mathbb{E}[\,\mathbb{P}(\xi \leq s \,|\, \mathscr{F}_t) \,|\, \mathscr{F}_s] = \mathbb{P}(\xi \leq s \,|\, \mathscr{F}_s) = X_s$. □

Exercice 13. "$\Leftarrow$" Obvious.

"$\Rightarrow$" Suppose $\sup_{t\geq 0} \mathbb{E}(X_t^+) < \infty$. Since $|X_t| = 2X_t^+ - X_t$ and $\mathbb{E}(X_t) \geq \mathbb{E}(X_0)$, we have $\sup_{t\geq 0} \mathbb{E}(|X_t|) \leq 2\sup_{t\geq 0} \mathbb{E}(X_t^+) - \mathbb{E}(X_0) < \infty$. □

Exercice 14. If X is closed by ξ, then $X_t = \mathbb{E}(\xi \,|\, \mathscr{F}_t)$ is uniformly integrable.

Conversely, we assume that X is right-continuous and uniformly integrable. Then $X_t \to X_\infty$ a.s. and in L^1, with $X_t = \mathbb{E}(X_\infty \,|\, \mathscr{F}_t)$. By definition, this means X is closed by X_∞. □

Exercice 15. **(i)** It follows from the usual inequality for the number of up-crossings.

(ii) By (i) and the monotone convergence theorem, $\mathbb{E}[U_\infty(X;\, a,\, b)] \leq \frac{\mathbb{E}[(X_0-a)^+]}{b-a}$, where $U_\infty(X;\, a,\, b)$ denotes the number of up-crossings along the interval $[a,\, b]$ by $(X_n,\, n \leq 0)$. A fortiori, $U_\infty(X;\, a,\, b) < \infty$ a.s.; hence $\mathbb{P}(U_\infty(X;\, a,\, b) < \infty,\, \forall a < b \text{ rationals}) = 1$. This yields the a.s. existence of $\lim_{n\to-\infty} X_n$.

(iii) In view of a.s. convergence proved in (ii), it only remains to prove that $(X_n,\, n \leq 0)$ is uniformly integrable. Since $(\mathbb{E}[X_0 \,|\, \mathscr{F}_n],\, n \leq 0)$ is uniformly integrable, it suffices, for the proof of convergence in L^1, to verify that the submartingale $(X_n - \mathbb{E}[X_0 \,|\, \mathscr{F}_n],\, n \leq 0)$ is uniformly integrable. As such, we can assume, without loss of generality, that $X_n \leq 0$ for all $n \leq 0$.

When $n \to -\infty$, $\mathbb{E}(X_n) \to A = \inf_{n\leq 0} \mathbb{E}(X_n) \in]-\infty, 0]$. Let $\varepsilon > 0$. There exists $N < \infty$ such that $\mathbb{E}(X_{-N}) - A \leq \varepsilon$, and a fortiori $\mathbb{E}(X_{-N}) - \mathbb{E}(X_n) \leq \varepsilon$, $\forall\, n \leq 0$. Let $a > 0$. We have, for $n \le -N$,

$$\begin{aligned}
\mathbb{E}[\,|X_n|\,\mathbf{1}_{\{|X_n|>a\}}\,] &= -\mathbb{E}[\,X_n\,\mathbf{1}_{\{X_n<-a\}}\,] \\
&= -\mathbb{E}(X_n) + \mathbb{E}[\,X_n\,\mathbf{1}_{\{X_n\ge -a\}}\,] \\
&\le -\mathbb{E}(X_n) + \mathbb{E}[\,X_{-N}\,\mathbf{1}_{\{X_n\ge -a\}}\,] \\
&= -\mathbb{E}(X_n) + \mathbb{E}(X_{-N}) - \mathbb{E}[\,X_{-N}\,\mathbf{1}_{\{X_n<-a\}}\,] \\
&\le \varepsilon + \mathbb{E}[\,|X_{-N}|\,\mathbf{1}_{\{|X_n|>a\}}\,].
\end{aligned}$$

By the Markov inequality, $\mathbb{P}(|X_n|>a) \le \frac{-\mathbb{E}(X_n)}{a} \le \frac{-A}{a} = \frac{|A|}{a}$. Hence we can choose a so large that $\mathbb{E}[\,|X_{-N}|\,\mathbf{1}_{\{|X_n|>a\}}\,] \le \varepsilon$. Then

$$\sup_{n\le -N} \mathbb{E}[\,|X_n|\,\mathbf{1}_{\{|X_n|>a\}}\,] \le 2\varepsilon.$$

On the other hand, we can choose a sufficiently large such that $\mathbb{E}[\,|X_n|\,\mathbf{1}_{\{|X_n|>a\}}\,] \le \varepsilon$ for $n = 0$, $-1, \cdots, -N$. Consequently, $(X_n,\, n\le 0)$ is uniformly integrable (and $\mathbb{E}(|X_{-\infty}|) < \infty$).

(iv) Since $X_n \le \mathbb{E}(X_0 \,|\, \mathscr{F}_n)$, we have, for all $A \in \mathscr{F}_{-\infty}$ (A is, a fortiori, an element of $\mathscr{F}_n$),

$$\mathbb{E}[X_n\,\mathbf{1}_A] \le \mathbb{E}[X_0\,\mathbf{1}_A].$$

Since $X_n \to X_{-\infty}$ in L^1, by letting $n\to -\infty$, we get $\mathbb{E}[X_{-\infty}\,\mathbf{1}_A] \le \mathbb{E}[X_0\,\mathbf{1}_A]$. Since $X_{-\infty}$ is $\mathscr{F}_n$-measurable (for all $n\le 0$) hence $(\mathscr{F}_{-\infty})$-measurable, this implies that $X_{-\infty} \le \mathbb{E}(X_0\,|\,\mathscr{F}_{-\infty})$, a.s.

(v) Let $X_n := \mathbb{E}(\xi\,|\,\mathscr{F}_n)$, $n\le 0$, which is a backward martingale. By (ii) and (iii), $X_n \to X_{-\infty}$ a.s. and in L^1, where

$$X_{-\infty} = \mathbb{E}[X_0\,|\,\mathscr{F}_{-\infty}] = \mathbb{E}[\mathbb{E}(\xi\,|\,\mathscr{F}_0)\,|\,\mathscr{F}_{-\infty}] = \mathbb{E}[\xi\,|\,\mathscr{F}_{-\infty}], \qquad \text{a.s.,}$$

as desired. □

Exercice 16. Fix $n\ge 1$. We apply the optional sampling theorem to the uniformly integrable martingale $(X_{t\wedge n},\, t\ge 0)$ and to the pair of stopping times T and $T+t$, to see that $\mathbb{E}(X_{(T+t)\wedge n}\,|\,\mathscr{F}_T) = X_{T\wedge n}$. Let $n\to\infty$. By the conditional Fatou's lemma, $\mathbb{E}(X_{T+t}\,|\,\mathscr{F}_T) \le X_T$, hence $\mathbb{E}(X_{T+t}\,\mathbf{1}_{\{T<\infty\}}\,|\,\mathscr{F}_T) \le X_T\,\mathbf{1}_{\{T<\infty\}} = 0$. This is possible only if $X_{T+t}\,\mathbf{1}_{\{T<\infty\}} = 0$ a.s., i.e., $X_{T+t} = 0$ a.s. on $\{T<\infty\}$.

Summarizing: a.s. on $\{T<\infty\}$, we have $X_{T+t} = 0$, $\forall t \in \mathbb{R}_+ \cap \mathbb{Q}$. The continuity of X tells us that we can remove the restriction $t\in\mathbb{Q}$. □

Exercice 17. We have

$$\begin{aligned}
\mathbb{E}[X_T\,|\,\mathscr{F}_S] &= \mathbb{E}[X_{T\wedge S}\,\mathbf{1}_{\{T\le S\}}\,|\,\mathscr{F}_S] + \mathbb{E}[X_{T\vee S}\,\mathbf{1}_{\{T>S\}}\,|\,\mathscr{F}_S] \\
&= X_{T\wedge S}\,\mathbf{1}_{\{T\le S\}} + \mathbf{1}_{\{T>S\}}\mathbb{E}[X_{T\vee S}\,|\,\mathscr{F}_S] \\
&\ge X_{T\wedge S}\,\mathbf{1}_{\{T\le S\}} + \mathbf{1}_{\{T>S\}}X_S = X_{T\wedge S},
\end{aligned}$$

as desired. $\square$

Exercice 18. **(i)** The right-continuity of the trajectories is obvious. Let us prove that $(X_{T\wedge t},\ t \geq 0)$ is a martingale with respect to $(\mathscr{F}_t)$.

For $t \geq 0$, it is clear that $\mathbb{E}(|X_{T\wedge t}|) < \infty$ (a consequence of the optional sampling theorem) and that $X_{T\wedge t}$ is $\mathscr{F}_t$-measurable (being $\mathscr{F}_{T\wedge t}$-measurable). Let $t > s \geq 0$. Applying the previous exercise gives $\mathbb{E}(X_{T\wedge t} \,|\, \mathscr{F}_s) = X_{(T\wedge t)\wedge s}$, which is $X_{T\wedge s}$.

(ii) If $(X_t,\ t \geq 0)$ is uniformly integrable, then the optional sampling theorem says that $X_{T\wedge t} = \mathbb{E}(X_\infty \,|\, \mathscr{F}_{T\wedge t})$, which yields the uniform integrability of $(X_{T\wedge t},\ t \geq 0)$ by recalling that for any integrable random variable ξ, $(\mathbb{E}(\xi \,|\, \mathscr{G}),\ \mathscr{G} \subset \mathscr{F}\ \sigma\text{-field})$ is uniformly integrable. $\square$

Exercice 19. By the conditional Fatou's lemma, $\mathbb{E}(X_\infty \,|\, \mathscr{F}_t) \leq X_t$ a.s. Taking expectation on both sides gives $\mathbb{E}(X_\infty) \leq \mathbb{E}(X_t)$ which is $\leq \mathbb{E}(X_0)$ because X is a *super*martingale. By assumption, $\mathbb{E}(X_\infty) = \mathbb{E}(X_0)$, which is possible only if $\mathbb{E}(X_\infty \,|\, \mathscr{F}_t) = X_t$ a.s., i.e., only if is a uniformly integrable martingale. $\square$

Exercice 20. **(i)** For all $a > 0$, $a\,\mathbb{P}(S_t \geq a) \leq \mathbb{E}[\,X_t\,\mathbf{1}_{\{S_t \geq a\}}]$. So

$$
\begin{aligned}
2\lambda\,\mathbb{P}(S_t \geq 2\lambda) &\leq \mathbb{E}[\,X_t\,\mathbf{1}_{\{S_t \geq 2\lambda\}}] \leq \mathbb{E}[\,X_t\,\mathbf{1}_{\{X_t > \lambda\}}] + \mathbb{E}[\,X_t\,\mathbf{1}_{\{X_t \leq \lambda,\ S_t \geq 2\lambda\}}] \\
&\leq \mathbb{E}[\,X_t\,\mathbf{1}_{\{X_t > \lambda\}}] + \lambda\,\mathbb{P}(S_t \geq 2\lambda),
\end{aligned}
$$

from which the desired inequality follows.

(ii) We have

$$
\begin{aligned}
\frac{1}{2}\,\mathbb{E}[S_t] &= \int_0^\infty \mathbb{P}(S_t \geq 2\lambda)\,\mathrm{d}\lambda \leq 1 + \int_1^\infty \mathbb{P}(S_t \geq 2\lambda)\,\mathrm{d}\lambda \\
&\leq 1 + \int_1^\infty \mathbb{E}[\,\lambda^{-1} X_t\,\mathbf{1}_{\{X_t > \lambda\}}]\,\mathrm{d}\lambda.
\end{aligned}
$$

By Fubini's theorem, the last integral equals $\mathbb{E}[\int_1^{X_t} \lambda^{-1} X_t\,\mathbf{1}_{\{X_t \geq 1\}}\,\mathrm{d}\lambda\,] = \mathbb{E}[\,X_t \log_+ X_t\,]$. We obtain the desired result.

(iii) By assumption, $Y_t = \mathbb{E}(Y_\infty \,|\, \mathscr{F}_t)$. Since $x \mapsto |x| \log_+ |x| =: \varphi(x)$ is convex, Jensen's inequality says that $\varphi(Y_t) \leq \mathbb{E}[\,\varphi(Y_\infty) \,|\, \mathscr{F}_t\,]$; hence $\sup_{t\geq 0} \mathbb{E}[\,\varphi(Y_t)\,] \leq \mathbb{E}[\,\varphi(Y_\infty)\,] < \infty$. By (ii) (applied to $X_t := |Y_t|$, $t \geq 0$, which is a non-negative submartingale), $\frac{1}{2}\,\mathbb{E}(\sup_{s\in[0,\,t]} |Y_s|) \leq 1 + \mathbb{E}[\,\varphi(Y_t)\,] \leq 1 + \mathbb{E}[\,\varphi(Y_\infty)\,]$. It follows from the monotone convergence theorem that $\mathbb{E}(\sup_{t\geq 0} |Y_t|) \leq 2 + 2\,\mathbb{E}[\,\varphi(Y_\infty)\,] < \infty$. $\square$

Exercice 21. **(i)** That $\mathbb{E}(\sup_{t\geq 0} X_t^2) < \infty$ is a consequence of Doob's inequality. In particular, $\mathbb{E}(\sup_{t\geq 0} |X_t|) < \infty$; a fortiori, X is uniformly integrable.

(ii) Since $|X_S| \leq \sup_{t\geq 0} |X_t|$, we have $\mathbb{E}(X_S^2) < \infty$. Similarly, $\mathbb{E}(Y_T^2) < \infty$. Hence by the Cauchy–Schwarz inequality, $\mathbb{E}(|X_S Y_T|) < \infty$.

Applying the optional sampling theorem to the uniformly integral martingale Y gives

$$\begin{aligned}\mathbb{E}(X_S Y_T \mathbf{1}_{\{S\le T\}} \,|\, \mathscr{F}_S) &= X_S \mathbf{1}_{\{S\le T\}} \mathbb{E}(Y_{T\vee S} \,|\, \mathscr{F}_S) \\ &= X_S \mathbf{1}_{\{S\le T\}} Y_S \\ &= X_{S\wedge T} Y_{S\wedge T} \mathbf{1}_{\{S\le T\}},\end{aligned}$$

from which it follows that

$$\mathbb{E}(X_S Y_T \mathbf{1}_{\{S\le T\}}) = \mathbb{E}(X_{S\wedge T} Y_{S\wedge T} \mathbf{1}_{\{S\le T\}}).$$

On the other hand, $X_S Y_T \mathbf{1}_{\{S>T\}} = X_{S\wedge T} Y_{S\wedge T} \mathbf{1}_{\{S>T\}}$. Hence

$$\mathbb{E}(X_S Y_T \mathbf{1}_{\{S>T\}}) = \mathbb{E}(X_{S\wedge T} Y_{S\wedge T} \mathbf{1}_{\{S>T\}}).$$

Consequently, $\mathbb{E}(X_S Y_T) = \mathbb{E}(X_{S\wedge T} Y_{S\wedge T})$.

(iii) The same proof as in (ii), except in two places:

- to justify the integrability of $X_S Y_T$, let $a > 0$ be such that $S \le a$, then $\mathbb{E}(X_S^2) \le \mathbb{E}(\sup_{u\in[0,\,a]} X_u^2) \le 4\mathbb{E}(X_a^2) < \infty$, and similarly, $\mathbb{E}(Y_T^2) < \infty$, so $\mathbb{E}(|X_S Y_T|) < \infty$;
- to justify $\mathbb{E}(Y_{T\vee S} \,|\, \mathscr{F}_S) = Y_S$, we apply the optional sampling theorem to Y and to the pair of *bounded* stopping times $T \vee S$ and S. □

Exercice 22. Since S and T are bounded, Doob's inequality implies that $\mathbb{E}(B_s^2) < \infty$ and that $\mathbb{E}(B_T^2) < \infty$. We have

$$\begin{aligned}\mathbb{E}[(B_T - B_S)^2] &= \mathbb{E}(B_S^2) + \mathbb{E}(B_T^2) - 2\mathbb{E}[\,\mathbb{E}(B_S B_T \,|\, \mathscr{F}_S)\,] \\ &= \mathbb{E}(B_S^2) + \mathbb{E}(B_T^2) - 2\mathbb{E}[\,B_S \mathbb{E}(B_T \,|\, \mathscr{F}_S)\,],\end{aligned}$$

because B_S is $\mathscr{F}_S$)-measurable. Applying the optional sample theorem to B and to the pair of *bounded* stopping times S and T yields $\mathbb{E}(B_T \,|\, \mathscr{F}_S) = B_S$, which, in turn, implies that

$$\mathbb{E}[(B_T - B_S)^2] = \mathbb{E}(B_S^2) + \mathbb{E}(B_T^2) - 2\mathbb{E}[\,B_S^2] = \mathbb{E}(B_T^2) - \mathbb{E}(B_S^2).$$

We now apply the optional sample theorem to $(B_t^2 - t,\ t \ge 0)$ and to the pair of *bounded* stopping times T and 0, to see that $\mathbb{E}(B_T^2 - T) = 0$; thus $\mathbb{E}(B_T^2) = \mathbb{E}(T)$. Similarly, $\mathbb{E}(B_S^2) = \mathbb{E}(S)$. Hence $\mathbb{E}(B_T^2) - \mathbb{E}(B_S^2) = \mathbb{E}(T - S)$. □

Exercice 23. **(i)** Let $T := \inf\{t \ge 0 :\ X_t \ge x\}$ which is a stopping time. Clearly, $(X_{t\wedge T},\ t \ge 0)$ is a continuous martingale, and is uniformly integrable (because $|X_{t\wedge T}| \le x + X_0$), closed by X_T (with the notation $X_\infty := 0$). By the optional sampling theorem, $\mathbb{E}(X_T \,|\, \mathscr{F}_0) = X_0$. We observe that

$$\begin{aligned}\mathbb{E}[X_T \,|\, \mathscr{F}_0] &= \mathbb{E}[X_T \mathbf{1}_{\{T<\infty\}} \,|\, \mathscr{F}_0] + \mathbb{E}[X_\infty \mathbf{1}_{\{T=\infty\}} \,|\, \mathscr{F}_0] \\ &= \mathbb{E}[(x \vee X_0) \mathbf{1}_{\{T<\infty\}} \,|\, \mathscr{F}_0] \\ &= (x \vee X_0)\, \mathbb{P}[T < \infty \,|\, \mathscr{F}_0],\end{aligned}$$

which yields

$$\mathbb{P}[T < \infty \mid \mathscr{F}_0] = \frac{X_0}{x \vee X_0} = 1 \wedge \frac{X_0}{x}.$$

It suffices then to remark that $\{T < \infty\} = \{\sup_{t \geq 0} X_t \geq x\}$.

(ii) Let $X_t := \mathrm{e}^{2(B_t - t)}$ which is a continuous martingale. Since a.s. $\frac{B_t}{t} \to 0$ $(t \to \infty)$, we have $B_t - t = (\frac{B_t}{t} - 1)t \to -\infty$, a.s., and thus $X_t \to 0$ a.s. By (i), $\mathbb{P}\{\sup_{t \geq 0} X_t \geq x\} = 1 \wedge \frac{1}{x}$, $x > 0$, which means $\mathbb{P}\{\sup_{t \geq 0}(B_t - t) \geq a\} = \mathrm{e}^{-2a}$, $a > 0$. In other words, $\sup_{t \geq 0}(B_t - t)$ has the exponential law of parameter 2 (i.e., with mean $\frac{1}{2}$). □

Exercice 24. Consider the martingale $(X_t := \mathrm{e}^{-2\gamma B_t - 2\gamma^2 t},\ t \geq 0)$. Since $\mathrm{e}^{-2\gamma B_{t \wedge T_{a,b}} - 2\gamma^2 (t \wedge T_{a,b})} \leq \mathrm{e}^{2|\gamma|(a+b)}$, we see that $(X_{T_{a,b} \wedge t},\ t \geq 0)$ is a continuous and bounded martingale, closed by $X_{T_{a,b}}$. Applying the optional sample theorem to this uniformly integrable martingale, we obtain:

$$\begin{aligned} 1 &= \mathbb{E}[\mathrm{e}^{-2\gamma B_{T_{a,b}} - 2\gamma^2 T_{a,b}}] \\ &= \mathbb{E}[\mathrm{e}^{2\gamma a}\, \mathbf{1}_{\{T_{-a} < T_b\}}] + \mathbb{E}[\mathrm{e}^{-2\gamma b}\, \mathbf{1}_{\{T_{-a} > T_b\}}] \\ &= \mathrm{e}^{2\gamma a} - \mathrm{e}^{2\gamma a}\, \mathbb{P}(T_{-a} > T_b) + \mathrm{e}^{-2\gamma b}\, \mathbb{P}(T_{-a} > T_b), \end{aligned}$$

which yields[4] $\mathbb{P}(T_{-a} > T_b) = \frac{\mathrm{e}^{2\gamma a} - 1}{\mathrm{e}^{2\gamma a} - \mathrm{e}^{-2\gamma b}}$. □

Exercice 25. Both $(B_{t \wedge T},\ t \geq 0)$ and $(B^2_{t \wedge T} - t \wedge T,\ t \geq 0)$ are continuous martingales, with $\mathbb{E}(B^2_{t \wedge T}) = \mathbb{E}(t \wedge T) \leq \mathbb{E}(T)$; hence $\sup_t \mathbb{E}(B^2_{t \wedge T}) \leq \mathbb{E}(T) < \infty$. Consequently, $(B_{t \wedge T},\ t \geq 0)$ is a uniformly integrable martingale, closed by B_T (in particular, B_T is integrable). Applying the optional sampling theorem to this uniformly integrable martingale yields $\mathbb{E}(B_T) = \mathbb{E}(B_{0 \wedge T}) = 0$. □

Exercice 26. By Doob's inequality,

$$\mathbb{E}\left[\sup_{t \geq 0} B^2_{t \wedge T}\right] \leq 4 \sup_{t \geq 0} \mathbb{E}\left[B^2_{t \wedge T}\right] \leq 4\mathbb{E}(T) < \infty,$$

so $(B^2_{t \wedge T},\ t \geq 0)$ is uniformly integrable. Since $(t \wedge T,\ t \geq 0)$ is also uniformly integrable (being bounded by T), $(B^2_{t \wedge T} - t \wedge T,\ t \geq 0)$ is a continuous and uniformly integrable martingale, closed by $B^2_T - T$ (in particular, B_T has a finite second moment). Applying the optional sampling theorem to this uniformly integrable martingale yields $\mathbb{E}(B^2_T - T) = 0$. In other words, $\mathbb{E}(B^2_T) = \mathbb{E}(T)$. □

[4] Letting $a \to \infty$, we see that $\mathbb{P}(T_b < \infty)$ is 1 if $\gamma > 0$, and is $\mathrm{e}^{2\gamma b}$ if $\gamma < 0$, which is in agreement with the previous exercise, because $\mathbb{P}(T_b < \infty) = \mathbb{P}\{\sup_{t \geq 0}(B_t + \gamma t) \geq b\}$.