Stochastic Processes

Amir Dembo (revised by Kevin Ross)

August 21, 2013

E-mail address: `amir@stat.stanford.edu`

Department of Statistics, Stanford University, Stanford, CA 94305.

Contents

Preface

These are the lecture notes for a one quarter graduate course in Stochastic Processes that I taught at Stanford University in 2002 and 2003. This course is intended for incoming master students in Stanford's Financial Mathematics program, for advanced undergraduates majoring in mathematics and for graduate students from Engineering, Economics, Statistics or the Business school. One purpose of this text is to prepare students to a rigorous study of Stochastic Differential Equations. More broadly, its goal is to help the reader understand the basic concepts of measure theory that are relevant to the mathematical theory of probability and how they apply to the rigorous construction of the most fundamental classes of stochastic processes.

Towards this goal, we introduce in Chapter 1 the relevant elements from measure and integration theory, namely, the probability space and the σ-fields of events in it, random variables viewed as measurable functions, their expectation as the corresponding Lebesgue integral, independence, distribution and various notions of convergence. This is supplemented in Chapter 2 by the study of the conditional expectation, viewed as a random variable defined via the theory of orthogonal projections in Hilbert spaces.

After this exploration of the foundations of Probability Theory, we turn in Chapter 3 to the general theory of Stochastic Processes, with an eye towards processes indexed by continuous time parameter such as the Brownian motion of Chapter 5 and the Markov jump processes of Chapter 6. Having this in mind, Chapter 3 is about the finite dimensional distributions and their relation to sample path continuity. Along the way we also introduce the concepts of stationary and Gaussian stochastic processes.

Chapter 4 deals with filtrations, the mathematical notion of information progression in time, and with the associated collection of stochastic processes called martingales. We treat both discrete and continuous time settings, emphasizing the importance of right-continuity of the sample path and filtration in the latter case. Martingale representations are explored, as well as maximal inequalities, convergence theorems and applications to the study of stopping times and to extinction of branching processes.

Chapter 5 provides an introduction to the beautiful theory of the Brownian motion. It is rigorously constructed here via Hilbert space theory and shown to be a Gaussian martingale process of stationary independent increments, with continuous sample path and possessing the strong Markov property. Few of the many explicit computations known for this process are also demonstrated, mostly in the context of hitting times, running maxima and sample path smoothness and regularity.

Chapter 6 provides a brief introduction to the theory of Markov chains and processes, a vast subject at the core of probability theory, to which many text books are devoted. We illustrate some of the interesting mathematical properties of such processes by examining the special case of the Poisson process, and more generally, that of Markov jump processes.

As clear from the preceding, it normally takes more than a year to cover the scope of this text. Even more so, given that the intended audience for this course has only minimal prior exposure to stochastic processes (beyond the usual elementary probability class covering only discrete settings and variables with probability density function). While students are assumed to have taken a real analysis class dealing with Riemann integration, no prior knowledge of measure theory is assumed here. The unusual solution to this set of constraints is to provide rigorous definitions, examples and theorem statements, while forgoing the proofs of all but the most easy derivations. At this somewhat superficial level, one can cover everything in a one semester course of forty lecture hours (and if one has highly motivated students such as I had in Stanford, even a one quarter course of thirty lecture hours might work).

In preparing this text I was much influenced by Zakai's unpublished lecture notes [**Zak**]. Revised and expanded by Shwartz and Zeitouni it is used to this day for teaching Electrical Engineering Phd students at the Technion, Israel. A second source for this text is Breiman's [**Bre92**], which was the intended text book for my class in 2002, till I realized it would not do given the preceding constraints. The resulting text is thus a mixture of these influencing factors with some digressions and additions of my own.

I thank my students out of whose work this text materialized. Most notably I thank Nageeb Ali, Ajar Ashyrkulova, Alessia Falsarone and Che-Lin Su who wrote the first draft out of notes taken in class, Barney Hartman-Glaser, Michael He, Chin-Lum Kwa and Chee-Hau Tan who used their own class notes a year later in a major revision, reorganization and expansion of this draft, and Gary Huang and Mary Tian who helped me with the intricacies of LaTeX.

I am much indebted to my colleague Kevin Ross for providing many of the exercises and all the figures in this text. Kevin's detailed feedback on an earlier draft of these notes has also been extremely helpful in improving the presentation of many key concepts.

Amir Dembo

Stanford, California
January 2008

CHAPTER 1

Probability, measure and integration

This chapter is devoted to the mathematical foundations of probability theory. Section 1.1 introduces the basic measure theory framework, namely, the probability space and the σ-fields of events in it. The next building block are random variables, introduced in Section 1.2 as measurable functions $\omega \mapsto X(\omega)$. This allows us to define the important concept of expectation as the corresponding Lebesgue integral, extending the horizon of our discussion beyond the special functions and variables with density, to which elementary probability theory is limited. As much of probability theory is about asymptotics, Section 1.3 deals with various notions of convergence of random variables and the relations between them. Section 1.4 concludes the chapter by considering independence and distribution, the two fundamental aspects that differentiate probability from (general) measure theory, as well as the related and highly useful technical tools of weak convergence and uniform integrability.

1.1. Probability spaces and σ-fields

We shall define here the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ using the terminology of measure theory. The *sample space* Ω is a set of all possible outcomes $\omega \in \Omega$ of some random experiment or phenomenon. Probabilities are assigned by a set function $A \mapsto \mathbf{P}(A)$ to A in a subset $\mathcal{F}$ of all possible sets of outcomes. The *event space* $\mathcal{F}$ represents both the amount of information available as a result of the experiment conducted and the collection of all events of possible interest to us. A pleasant mathematical framework results by imposing on $\mathcal{F}$ the structural conditions of a σ-field, as done in Subsection 1.1.1. The most common and useful choices for this σ-field are then explored in Subsection 1.1.2.

1.1.1. The probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

We use 2^Ω to denote the set of all possible subsets of Ω. The event space is thus a subset $\mathcal{F}$ of 2^Ω, consisting of all allowed events, that is, those events to which we shall assign probabilities. We next define the structural conditions imposed on $\mathcal{F}$.

DEFINITION 1.1.1. *We say that $\mathcal{F} \subseteq 2^\Omega$ is a σ-*field *(or a σ-algebra), if*
(a) $\Omega \in \mathcal{F}$,
(b) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$ as well (where $A^c = \Omega \setminus A$).
(c) If $A_i \in \mathcal{F}$ for $i = 1, 2 \ldots$ then also $\bigcup_{i=1}^\infty A_i \in \mathcal{F}$.

REMARK. Using DeMorgan's law you can easily check that if $A_i \in \mathcal{F}$ for $i = 1, 2 \ldots$ and $\mathcal{F}$ is a σ-field, then also $\bigcap_i A_i \in \mathcal{F}$. Similarly, you can show that a σ-field is closed under countably many elementary set operations.

Definition 1.1.2. *A pair* $(\Omega, \mathcal{F})$ *with* $\mathcal{F}$ *a* σ*-field of subsets of* Ω *is called a* measurable space. *Given a measurable space, a* probability measure $\mathbf{P}$ *is a function* $\mathbf{P}: \mathcal{F} \to [0,1]$*, having the following properties:*
(a) $0 \leq \mathbf{P}(A) \leq 1$ *for all* $A \in \mathcal{F}$.
(b) $\mathbf{P}(\Omega) = 1$.
(c) (Countable additivity) $\mathbf{P}(A) = \sum_{n=1}^{\infty} \mathbf{P}(A_n)$ *whenever* $A = \bigcup_{n=1}^{\infty} A_n$ *is a countable union of disjoint sets* $A_n \in \mathcal{F}$ *(that is,* $A_n \bigcap A_m = \emptyset$*, for all* $n \neq m$*).*
A probability space *is a triplet* $(\Omega, \mathcal{F}, \mathbf{P})$*, with* $\mathbf{P}$ *a probability measure on the measurable space* $(\Omega, \mathcal{F})$*.*

The next exercise collects some of the fundamental properties shared by all probability measures.

Exercise 1.1.3. *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a probability space and* A, B, A_i *events in* $\mathcal{F}$*. Prove the following properties of every probability measure.*

(a) Monotonicity. *If* $A \subseteq B$ *then* $\mathbf{P}(A) \leq \mathbf{P}(B)$.
(b) Sub-additivity. *If* $A \subseteq \cup_i A_i$ *then* $\mathbf{P}(A) \leq \sum_i \mathbf{P}(A_i)$.
(c) Continuity from below: *If* $A_i \uparrow A$*, that is,* $A_1 \subseteq A_2 \subseteq \ldots$ *and* $\cup_i A_i = A$*, then* $\mathbf{P}(A_i) \uparrow \mathbf{P}(A)$.
(d) Continuity from above: *If* $A_i \downarrow A$*, that is,* $A_1 \supseteq A_2 \supseteq \ldots$ *and* $\cap_i A_i = A$*, then* $\mathbf{P}(A_i) \downarrow \mathbf{P}(A)$.
(e) Inclusion-exclusion rule:

$$\begin{aligned}\mathbf{P}\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_i \mathbf{P}(A_i) - \sum_{i<j} \mathbf{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbf{P}(A_i \cap A_j \cap A_k) \\ &- \cdots + (-1)^{n+1} \mathbf{P}(A_1 \cap \cdots \cap A_n)\end{aligned}$$

The σ-field $\mathcal{F}$ always contains at least the set Ω and its complement, the empty set $\emptyset$. Necessarily, $\mathbf{P}(\Omega) = 1$ and $\mathbf{P}(\emptyset) = 0$. So, if we take $\mathcal{F}_0 = \{\emptyset, \Omega\}$ as our σ-field, then we are left with no degrees of freedom in choice of $\mathbf{P}$. For this reason we call $\mathcal{F}_0$ the *trivial* σ*-field*.

Fixing Ω, we may expect that the larger the σ-field we consider, the more freedom we have in choosing the probability measure. This indeed holds to some extent, that is, as long as we have no problem satisfying the requirements (a)-(c) in the definition of a probability measure. For example, a natural question is when should we expect the maximal possible σ-field $\mathcal{F} = 2^{\Omega}$ to be useful?

Example 1.1.4. *When the sample space* Ω *is finite we can and typically shall take* $\mathcal{F} = 2^{\Omega}$*. Indeed, in such situations we assign a probability* $p_\omega > 0$ *to each* $\omega \in \Omega$ *making sure that* $\sum_{\omega \in \Omega} p_\omega = 1$*. Then, it is easy to see that taking* $\mathbf{P}(A) = \sum_{\omega \in A} p_\omega$ *for any* $A \subseteq \Omega$ *results with a probability measure on* $(\Omega, 2^{\Omega})$*. For instance, when we consider a single coin toss we have* $\Omega_1 = \{\mathrm{H}, \mathrm{T}\}$ *(*$\omega = \mathrm{H}$ *if the coin lands on its head and* $\omega = \mathrm{T}$ *if it lands on its tail), and* $\mathcal{F}_1 = \{\emptyset, \Omega, \{\mathrm{H}\}, \{\mathrm{T}\}\}$*. Similarly, when we consider any finite number of coin tosses, say* n*, we have* $\Omega_n = \{(\omega_1, \ldots, \omega_n) : \omega_i \in \{\mathrm{H}, \mathrm{T}\}, i = 1, \ldots, n\}$*, that is* Ω_n *is the set of all possible* n*-tuples of coin tosses, while* $\mathcal{F}_n = 2^{\Omega_n}$ *is the collection of all possible sets of* n*-tuples of coin tosses. The same construction applies even when* Ω *is infinite, provided it is countable. For instance, when* $\Omega = \{0, 1, 2, \ldots\}$ *is the set of all non-negative integers and* $\mathcal{F} = 2^{\Omega}$*, we get the Poisson probability measure of parameter* $\lambda > 0$ *when starting from* $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$ *for* $k = 0, 1, 2, \ldots$*.*

When Ω is uncountable such a strategy as in Example 1.1.4 will no longer work. The problem is that if we take $p_\omega = \mathbf{P}(\{\omega\}) > 0$ for uncountably many values of ω, we shall end up with $\mathbf{P}(\Omega) = \infty$. Of course we may define everything as before on a countable subset $\widehat{\Omega}$ of Ω and demand that $\mathbf{P}(A) = \mathbf{P}(A \cap \widehat{\Omega})$ for each $A \subseteq \Omega$. Excluding such trivial cases, to genuinely use an uncountable sample space Ω we need to restrict our σ-field $\mathcal{F}$ to a *strict subset* of 2^Ω.

1.1.2. Generated and Borel σ-fields. Enumerating the sets in the σ-field $\mathcal{F}$ it not a realistic option for uncountable Ω. Instead, as we see next, the most common construction of σ-fields is then by implicit means. That is, we demand that certain sets (called the *generators*) be in our σ-field, and take the smallest possible collection for which this holds.

Definition 1.1.5. *Given a collection of subsets $A_\alpha \subseteq \Omega$, where $\alpha \in \Gamma$ a not necessarily countable index set, we denote the smallest σ-field $\mathcal{F}$ such that $A_\alpha \in \mathcal{F}$ for all $\alpha \in \Gamma$ by $\sigma(\{A_\alpha\})$ (or sometimes by $\sigma(A_\alpha, \alpha \in \Gamma)$), and call $\sigma(\{A_\alpha\})$ the* σ-field generated *by the collection $\{A_\alpha\}$. That is,*
$\sigma(\{A_\alpha\}) = \bigcap\{\mathcal{G} :\ \mathcal{G} \subseteq 2^\Omega \text{ is a } \sigma-\text{field}, \quad A_\alpha \in \mathcal{G} \quad \forall \alpha \in \Gamma\}.$

Definition 1.1.5 works because the intersection of (possibly uncountably many) σ-fields is also a σ-field, which you will verify in the following exercise.

Exercise 1.1.6. *Let $\mathcal{A}_\alpha$ be a σ-field for each $\alpha \in \Gamma$, an arbitrary index set. Show that $\bigcap_{\alpha\in\Gamma} \mathcal{A}_\alpha$ is a σ-field. Provide an example of two σ-fields $\mathcal{F}$ and $\mathcal{G}$ such that $\mathcal{F} \cup \mathcal{G}$ is not a σ-field.*

Different sets of generators may result with the same σ-field. For example, taking $\Omega = \{1, 2, 3\}$ it is not hard to check that $\sigma(\{1\}) = \sigma(\{2, 3\}) = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$.

Example 1.1.7. *An example of a generated σ-field is the* Borel σ-field *on $\mathbb{R}$. It may be defined as $\mathcal{B} = \sigma(\{(a, b) : a, b \in \mathbb{R}\})$.*

The following lemma lays out the strategy one employs to show that the σ-fields generated by two different collections of sets are actually identical.

Lemma 1.1.8. *If two different collections of generators $\{A_\alpha\}$ and $\{B_\beta\}$ are such that $A_\alpha \in \sigma(\{B_\beta\})$ for each α and $B_\beta \in \sigma(\{A_\alpha\})$ for each β, then $\sigma(\{A_\alpha\}) = \sigma(\{B_\beta\})$.*

Proof. Recall that if a collection of sets $\mathcal{A}$ is a subset of a σ-field $\mathcal{G}$, then by Definition 1.1.5 also $\sigma(\mathcal{A}) \subseteq \mathcal{G}$. Applying this for $\mathcal{A} = \{A_\alpha\}$ and $\mathcal{G} = \sigma(\{B_\beta\})$ our assumption that $A_\alpha \in \sigma(\{B_\beta\})$ for all α results with $\sigma(\{A_\alpha\}) \subseteq \sigma(\{B_\beta\})$. Similarly, our assumption that $B_\beta \in \sigma(\{A_\alpha\})$ for all β results with $\sigma(\{B_\beta\}) \subseteq \sigma(\{A_\alpha\})$. Taken together, we see that $\sigma(\{A_\alpha\}) = \sigma(\{B_\beta\})$. ∎

For instance, considering $\mathcal{B}_\mathcal{Q} = \sigma(\{(a, b) : a, b \in \mathcal{Q}\})$, we have by the preceding lemma that $\mathcal{B}_\mathcal{Q} = \mathcal{B}$ as soon as we show that any interval (a, b) is in $\mathcal{B}_\mathcal{Q}$. To verify this fact, note that for any real $a < b$ there are rational numbers $q_n < r_n$ such that $q_n \downarrow a$ and $r_n \uparrow b$, hence $(a, b) = \cup_n (q_n, r_n) \in \mathcal{B}_\mathcal{Q}$. Following the same approach, you are to establish next a few alternative definitions for the Borel σ-field $\mathcal{B}$.

Exercise 1.1.9. *Verify the alternative definitions of the* Borel σ-field $\mathcal{B}$*:*

$$\begin{aligned}\sigma(\{(a, b) : a < b \in \mathbb{R}\}) &= \sigma(\{[a, b] : a < b \in \mathbb{R}\}) = \sigma(\{(-\infty, b] : b \in \mathbb{R}\}) \\ &= \sigma(\{(-\infty, b] : b \in \mathcal{Q}\}) = \sigma(\{O \subseteq \mathbb{R} \text{ open }\})\end{aligned}$$

Hint: *Any $O \subseteq \mathbb{R}$ open is a countable union of sets (a, b) for $a, b \in \mathcal{Q}$ (rational).*

If $A \subseteq \mathbb{R}$ is in $\mathcal{B}$ of Example 1.1.7, we say that A is a *Borel set*. In particular, all open or closed subsets of $\mathbb{R}$ are Borel sets, as are many other sets. However,

PROPOSITION 1.1.10. *There exists a subset of $\mathbb{R}$ that is not in $\mathcal{B}$. That is, not all sets are Borel sets.*

Despite the above proposition, all sets encountered in practice are Borel sets.

Often there is no explicit enumerative description of the σ-field generated by an infinite collection of subsets. A notable exception is $\mathcal{G} = \sigma(\{[a, b] : a, b \in \mathbf{Z}\})$, where one may check that the sets in $\mathcal{G}$ are all possible unions of elements from the countable collection $\{\{b\}, (b, b+1), b \in \mathbf{Z}\}$. In particular, $\mathcal{B} \neq \mathcal{G}$ since for example $(0, 1/2) \notin \mathcal{G}$.

EXAMPLE 1.1.11. *One example of a probability measure defined on $(\mathbb{R}, \mathcal{B})$ is the* Uniform probability measure *on $(0, 1)$, denoted U and defined as following. For each interval $(a, b) \subseteq (0, 1)$, $a < b$, we set $U((a, b)) = b - a$ (the length of the interval), and for any other open interval I we set $U(I) = U(I \cap (0, 1))$.*

Note that we did not specify $U(A)$ for each Borel set A, but rather only for the generators of the Borel σ-field $\mathcal{B}$. This is a common strategy, as under mild conditions on the collection $\{A_\alpha\}$ of generators each probability measure $\mathbf{Q}$ specified only for the sets A_α can be uniquely extended to a probability measure $\mathbf{P}$ on $\sigma(\{A_\alpha\})$ that coincides with $\mathbf{Q}$ on all the sets A_α (and these conditions hold for example when the generators are all open intervals in $\mathbb{R}$).

EXERCISE 1.1.12. *Check that the following are Borel sets and find the probability assigned to each by the uniform measure of the preceding example: $(0, 1/2) \cup (1/2, 3/2)$, $\{1/2\}$, a countable subset A of $\mathbb{R}$, the set T of all irrational numbers within $(0, 1)$, the interval $[0, 1]$ and the set $\mathbb{R}$ of all real numbers.*

EXAMPLE 1.1.13. *Another classical example of an uncountable Ω is relevant for studying the experiment with an infinite number of coin tosses, that is, $\Omega_\infty = \Omega_1^{\mathbf{N}}$ for $\Omega_1 = \{\mathrm{H}, \mathrm{T}\}$ (recall that setting $\mathrm{H} = 1$ and $\mathrm{T} = 0$, each infinite sequence $\omega \in \Omega_\infty$ is in correspondence with a unique real number $x \in [0, 1]$ with ω being the binary expansion of x). The σ-field should at least allow us to consider any possible outcome of a finite number of coin tosses. The natural σ-field in this case is the minimal σ-field having this property, that is, $\mathcal{F}_c = \sigma(A_{n,\theta}, \theta \in \{\mathrm{H}, \mathrm{T}\}^n, n < \infty)$, for the subsets $A_{n,\theta} = \{\omega : \omega_i = \theta_i, i = 1 \ldots, n\}$ of Ω_∞ (e.g. $A_{1,\mathrm{H}}$ is the set of all sequences starting with H and $A_{2,\mathrm{TT}}$ are all sequences starting with a pair of T symbols). This is also our first example of a stochastic process, to which we return in the next section.*

Note that any countable union of sets of probability zero has probability zero, but this is not the case for an uncountable union. For example, $U(\{x\}) = 0$ for every $x \in \mathbb{R}$, but $U(\mathbb{R}) = 1$. When we later deal with continuous time stochastic processes we should pay attention to such difficulties!

1.2. Random variables and their expectation

Random variables are numerical functions $\omega \mapsto X(\omega)$ of the outcome of our random experiment. However, in order to have a successful mathematical theory, we limit our interest to the subset of *measurable* functions, as defined in Subsection

1.2.1 and study some of their properties in Subsection 1.2.2. Taking advantage of these we define the mathematical expectation in Subsection 1.2.3 as the corresponding Lebesgue integral and relate it to the more elementary definitions that apply for simple functions and for random variables having a probability density function.

1.2.1. Indicators, simple functions and random variables. We start with the definition of a random variable and two important examples of such objects.

Definition 1.2.1. *A* Random Variable *(R.V.) is a function* $X : \Omega \to \mathbb{R}$ *such that* $\forall \alpha \in \mathbb{R}$ *the set* $\{\omega : X(\omega) \leq \alpha\}$ *is in* $\mathcal{F}$ *(such a function is also called a* $\mathcal{F}$-measurable or, simply, measurable function*).*

Example 1.2.2. *For any* $A \in \mathcal{F}$ *the function* $I_A(\omega) = \begin{cases} 1, \omega \in A \\ 0, \omega \notin A \end{cases}$ *is a R.V.*

since $\{\omega : I_A(\omega) \leq \alpha\} = \begin{cases} \Omega, \alpha \geq 1 \\ A^c, 0 \leq \alpha < 1 \\ \emptyset, \alpha < 0 \end{cases}$ *all of whom are in* $\mathcal{F}$*. We call such R.V. also an* indicator function*.*

Example 1.2.3. *By same reasoning check that* $X(\omega) = \sum_{n=1}^{N} c_n I_{A_n}(\omega)$ *is a R.V. for any finite* N*, non-random* $c_n \in \mathbb{R}$ *and sets* $A_n \in \mathcal{F}$*. We call any such* X *a* simple function*, denoted by* $X \in \mathrm{SF}$*.*

Exercise 1.2.4. *Verify the following properties of indicator R.V.-s.*

(a) $I_\emptyset(\omega) = 0$ *and* $I_\Omega(\omega) = 1$
(b) $I_{A^c}(\omega) = 1 - I_A(\omega)$
(c) $I_A(\omega) \leq I_B(\omega)$ *if and only if* $A \subseteq B$
(d) $I_{\cap_i A_i}(\omega) = \prod_i I_{A_i}(\omega)$
(e) *If* A_i *are disjoint then* $I_{\cup_i A_i}(\omega) = \sum_i I_{A_i}(\omega)$

Though in our definition of a R.V. the σ-field $\mathcal{F}$ is implicit, the choice of $\mathcal{F}$ is very important (and we sometimes denote by $m\mathcal{F}$ the collection of all R.V. for a given σ-field $\mathcal{F}$). For example, there are non-trivial σ-fields $\mathcal{G}$ and $\mathcal{F}$ on $\Omega = \mathbb{R}$ such that $X(\omega) = \omega$ is measurable for $(\Omega, \mathcal{F})$, but not measurable for $(\Omega, \mathcal{G})$. Indeed, one such example is when $\mathcal{F}$ is the Borel σ-field $\mathcal{B}$ and $\mathcal{G} = \sigma(\{[a, b] : a, b \in \mathbf{Z}\})$ (for example, the set $\{\omega : \omega \leq \alpha\}$ is not in $\mathcal{G}$ whenever $\alpha \notin \mathbf{Z}$). To practice your understanding, solve the following exercise at this point.

Exercise 1.2.5. *Let* $\Omega = \{1, 2, 3\}$*. Find a* σ*-field* $\mathcal{F}$ *such that* $(\Omega, \mathcal{F})$ *is a measurable space, and a mapping* X *from* Ω *to* $\mathbb{R}$*, such that* X *is not a random variable on* $(\Omega, \mathcal{F})$*.*

Our next proposition explains why simple functions are quite useful in probability theory.

Proposition 1.2.6. *For every R.V.* $X(\omega)$ *there exists a sequence of simple functions* $X_n(\omega)$ *such that* $X_n(\omega) \to X(\omega)$ *as* $n \to \infty$*, for each fixed* $\omega \in \Omega$*.*

Proof. Let

$$f_n(x) = n\mathbf{1}_{x>n} + \sum_{k=0}^{n2^n-1} k2^{-n}\mathbf{1}_{(k2^{-n},(k+1)2^{-n}]}(x)\,,$$

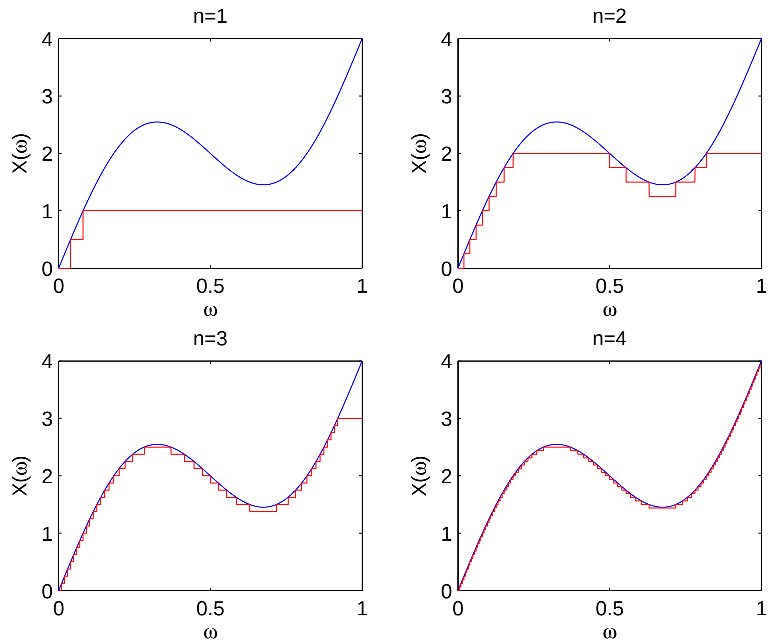

FIGURE 1. Illustration of approximation of a random variable using simple functions for different values of n.

noting that for R.V. $X \geq 0$, we have that $X_n = f_n(X)$ are simple functions. Since $X \geq X_{n+1} \geq X_n$ and $X(\omega) - X_n(\omega) \leq 2^{-n}$ whenever $X(\omega) \leq n$, it follows that $X_n(\omega) \to X(\omega)$ as $n \to \infty$, for each ω.

We write a general R.V. as $X(\omega) = X_+(\omega) - X_-(\omega)$ where $X_+(\omega) = \max(X(\omega), 0)$ and $X_-(\omega) = -\min(X(\omega), 0)$ are non-negative R.V.-s. By the above argument the simple functions $X_n = f_n(X_+) - f_n(X_-)$ have the convergence property we claimed. (See Figure 1 for an illustration.) ∎

The concept of almost sure prevails throughout probability theory.

DEFINITION 1.2.7. *We say that R.V. X and Y defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ are almost surely the same if $\mathbf{P}(\{\omega : X(\omega) \neq Y(\omega)\}) = 0$. This shall be denoted by $X \overset{a.s.}{=} Y$. More generally, the same notation applies to any property of a R.V. For example, $X(\omega) \geq 0$ a.s. means that $\mathbf{P}(\{\omega : X(\omega) < 0\}) = 0$. Hereafter, we shall consider such X and Y to be the same R.V. hence often omit the qualifier "a.s." when stating properties of R.V. We also use the terms* almost surely *(a.s.),* almost everywhere *(a.e.), and* with probability 1 *(w.p.1) interchangeably.*

The most important σ-fields are those "generated" by random variables, as defined next.

DEFINITION 1.2.8. *Given a R.V. X we denote by $\sigma(X)$ the smallest σ-field $\mathcal{G} \subseteq \mathcal{F}$ such that $X(\omega)$ is measurable on $(\Omega, \mathcal{G})$. One can show that $\sigma(X) = \sigma(\{\omega : X(\omega) \leq \alpha\})$. We call $\sigma(X)$ the* σ-field generated by X *and interchangeably use the notations $\sigma(X)$ and $\mathcal{F}_X$. Similarly, given R.V. $X_1, \ldots, X_n$ on the same measurable space*

$(\Omega, \mathcal{F})$, denote by $\sigma(X_k, k \leq n)$ the smallest σ-field $\mathcal{F}$ such that $X_k(\omega)$, $k = 1, \ldots, n$ are measurable on $(\Omega, \mathcal{F})$. That is, $\sigma(X_k, k \leq n)$ is the smallest σ-field containing $\sigma(X_k)$ for $k = 1, \ldots, n$.

Remark. One could also consider the possibly larger σ-field $\widehat{\sigma}(X) = \sigma(\{\omega : X(\omega) \in B\}$, for all Borel sets B), but it can be shown that $\sigma(X) = \widehat{\sigma}(X)$, a fact that we often use in the sequel (that $\sigma(X) \subseteq \widehat{\sigma}(X)$ is obvious, and with some effort one can also check that the converse holds).

Exercise 1.2.9. *Consider a sequence of two coin tosses, $\Omega = \{HH, HT, TH, TT\}$, $\mathcal{F} = 2^{\Omega}$, $\omega = (\omega_1 \omega_2)$. Specify $\sigma(X_0)$, $\sigma(X_1)$, and $\sigma(X_2)$ for the R.V.-s:*

$$\begin{aligned} X_0(\omega) &= 4, \\ X_1(\omega) &= 2X_0(\omega) I_{\{\omega_1 = H\}}(\omega) + 0.5 X_0(\omega) I_{\{\omega_1 = T\}}(\omega), \\ X_2(\omega) &= 2X_1(\omega) I_{\{\omega_2 = H\}}(\omega) + 0.5 X_1(\omega) I_{\{\omega_2 = T\}}(\omega). \end{aligned}$$

The concept of σ-field is needed in order to produce a rigorous mathematical theory. It further has the *crucial* role of quantifying the amount of information we have. For example, $\sigma(X)$ contains exactly those events A for which we can say whether $\omega \in A$ or not, based on the value of $X(\omega)$. Interpreting Example 1.1.13 as corresponding to sequentially tossing coins, the R.V. $X_n(\omega) = \omega_n$ gives the result of the n-th coin toss in our experiment Ω_∞ of infinitely many such tosses. The σ-field $\mathcal{F}_n = 2^{\Omega_n}$ of Example 1.1.4 then contains exactly the information we have upon observing the outcome of the first n coin tosses, whereas the larger σ-field $\mathcal{F}_c$ allows us to also study the limiting properties of this sequence. The sequence of R.V. $X_n(\omega)$ is an example of what we call a discrete time stochastic process.

1.2.2. Closure properties of random variables.

For the typical measurable space with uncountable Ω it is impractical to list all possible R.V. Instead, we state a few useful closure properties that often help us in showing that a given function $X(\omega)$ is indeed a R.V.

We start with closure with respect to taking limits.

Exercise 1.2.10. *Let $(\Omega, \mathcal{F})$ be a measurable space and let X_n be a sequence of random variables on it. Assume that for each $\omega \in \Omega$, the limit $X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$ exists and is finite. Prove that X_∞ is a random variable on $(\Omega, \mathcal{F})$.*
Hint: *Represent $\{\omega : X_\infty(\omega) > \alpha\}$ in terms of the sets $\{\omega : X_n(\omega) > \alpha\}$. Alternatively, check that $X_\infty(\omega) = \inf_m \sup_{n \geq m} X_n(\omega)$, or see* [**Bre92**, Proposition A.18] *for a detailed proof.*

We turn to deal with numerical operations involving R.V.s, for which we need first the following definition.

Definition 1.2.11. *A function $g : \mathbb{R} \mapsto \mathbb{R}$ is called* Borel *(measurable)* function *if g is a R.V. on $(\mathbb{R}, \mathcal{B})$.*

To practice your understanding solve the next exercise, in which you show that every continuous function g is Borel measurable. Further, every piecewise constant function g is also Borel measurable (where g piecewise constant means that g has at most countably many jump points between which it is constant).

Exercise 1.2.12. *Recall that a function $g : \mathbb{R} \to \mathbb{R}$ is continuous if $g(x_n) \to g(x)$ for every $x \in \mathbb{R}$ and any convergent sequence $x_n \to x$.*

(a) *Show that if g is a continuous function then for each $a \in \mathbb{R}$ the set $\{x : g(x) \leq a\}$ is closed. Alternatively, you may show instead that $\{x : g(x) < a\}$ is an open set for each $a \in \mathbb{R}$.*
(b) *Use whatever you opted to prove in (a) to conclude that all continuous functions are Borel measurable.*

We can and shall extend the notions of Borel sets and functions to $\mathbb{R}^n$, $n \geq 1$ by defining the Borel σ-field on $\mathbb{R}^n$ as $\mathcal{B}_n = \sigma(\{[a_1, b_1] \times \cdots \times [a_n, b_n] : a_i, b_i \in \mathbb{R}, i = 1, \ldots, n\})$ and calling $g : \mathbb{R}^n \mapsto \mathbb{R}$ Borel function if g is a R.V. on $(\mathbb{R}^n, \mathcal{B}_n)$. Convince yourself that these notions coincide for $n = 1$ with those of Example 1.1.7 and Definition 1.2.11.

PROPOSITION 1.2.13. *If $g : \mathbb{R}^n \to \mathbb{R}$ is a Borel function and $X_1, \ldots, X_n$ are R.V. on $(\Omega, \mathcal{F})$ then $g(X_1, \ldots, X_n)$ is also a R.V. on $(\Omega, \mathcal{F})$.*

If interested in the proof, c.f. [**Bre92**, Proposition 2.31].
This is the generic description of a R.V. on $\sigma(Y_1, \ldots, Y_n)$, namely:

THEOREM 1.2.14. *If Z is a R.V. on $(\Omega, \sigma(Y_1, \ldots, Y_n))$, then $Z = g(Y_1, \ldots, Y_n)$ for some Borel function $g : \mathbb{R}^n \to \mathbb{R}$*

For the proof of this result see [**Bre92**, Proposition A.21].
Here are some concrete special cases.

EXERCISE 1.2.15. *Choosing appropriate g in Proposition 1.2.13 deduce that given R.V. X_n, the following are also R.V.-s:*

$$\alpha X_n \text{ with } \alpha \in \mathbb{R}, \qquad X_1 + X_2, \qquad X_1 \cdot X_2 .$$

We consider next the effect that a Borel measurable function has on the amount of information quantified by the corresponding generated σ-fields.

PROPOSITION 1.2.16. *For any $n < \infty$, any Borel function $g : \mathbb{R}^n \to \mathbb{R}$ and R.V. $Y_1, \ldots, Y_n$ on the same measurable space we have the inclusion $\sigma(g(Y_1, \ldots, Y_n)) \subseteq \sigma(Y_1, \ldots, Y_n)$.*

In the following direct corollary of Proposition 1.2.16 we observe that the information content quantified by the respective generated σ-fields is invariant under invertible Borel transformations.

COROLLARY 1.2.17. *Suppose R.V. $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_m$ defined on the same measurable space are such that $Z_k = g_k(Y_1, \ldots, Y_n)$, $k = 1, \ldots, m$ and $Y_i = h_i(Z_1, \ldots, Z_m)$, $i = 1, \ldots, n$ for some Borel functions $g_k : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^m \to \mathbb{R}$. Then, $\sigma(Y_1, \ldots, Y_n) = \sigma(Z_1, \ldots, Z_m)$.*

PROOF. Left to the reader. Do it to practice your understanding of the concept of generated σ-fields. ∎

EXERCISE 1.2.18. *Provide example of a measurable space, a R.V. X on it, and:*

(a) *A function $g(x) \not\equiv x$ such that $\sigma(g(X)) = \sigma(X)$.*
(b) *A function f such that $\sigma(f(X))$ is strictly smaller than $\sigma(X)$ and is not the trivial σ-field $\{\emptyset, \Omega\}$.*

1.2.3. The (mathematical) expectation. A key concept in probability theory is the expectation of a R.V. which we define here, starting with the expectation of non-negative R.V.-s.

DEFINITION 1.2.19. *(see [**Bre92**, Appendix A.3]). The (mathematical)* expectation *of a R.V. $X(\omega)$ is denoted $\mathbf{E}X$. With $x_{k,n} = k2^{-n}$ and the intervals $I_{k,n} = (x_{k,n}, x_{k+1,n}]$ for $k = 0, 1, \ldots$, the expectation of $X(\omega) \geq 0$ is defined as:*

$$\mathbf{E}X = \lim_{n\to\infty} \Big[\sum_{k=0}^{\infty} x_{k,n} \mathbf{P}(\{\omega : X(\omega) \in I_{k,n}\}) \Big] . \tag{1.2.1}$$

For each value of n the non-negative series in the right-hand-side of (1.2.1) is well defined, though possibly infinite. For any k, n we have that $x_{k,n} = x_{2k,n+1} \leq x_{2k+1,n+1}$. Thus, the interval $I_{k,n}$ is the union of the disjoint intervals $I_{2k,n+1}$ and $I_{2k+1,n+1}$, implying that

$$x_{k,n}\mathbf{P}(X \in I_{k,n}) \leq x_{2k,n+1}\mathbf{P}(X \in I_{2k,n+1}) + x_{2k+1,n+1}\mathbf{P}(X \in I_{2k+1,n+1}) .$$

It follows that the series in Definition 1.2.19 are monotone non-decreasing in n, hence the limit as $n \to \infty$ exists as well.

We next discuss the two special cases for which the expectation can be computed explicitly, that of R.V. with countable range, followed by that of R.V. having a probability density function.

EXAMPLE 1.2.20. *Though not detailed in these notes, it is possible to show that for Ω countable and $\mathcal{F} = 2^{\Omega}$ our definition coincides with the well known elementary definition $\mathbf{E}X = \sum_{\omega} X(\omega)p_{\omega}$ (where $X(\omega) \geq 0$). More generally, the formula $\mathbf{E}X = \sum_i x_i \mathbf{P}(\{\omega : X(\omega) = x_i\})$ applies whenever the range of X is a bounded below countable set $\{x_1, x_2, \ldots\}$ of real numbers (for example, whenever $X \in \mathrm{SF}$).*

Here are few examples showing that we may have $\mathbf{E}X = \infty$ while $X(\omega) < \infty$ for all ω.

EXAMPLE 1.2.21. *Take $\Omega = \{1, 2, \ldots\}$, $\mathcal{F} = 2^{\Omega}$ and the probability measure corresponding to $p_k = ck^{-2}$, with $c = [\sum_{k=1}^{\infty} k^{-2}]^{-1}$ a positive, finite normalization constant. Then, the random variable $X(\omega) = \omega$ is finite but $\mathbf{E}X = c\sum_{k=1}^{\infty} k^{-1} = \infty$. For a more interesting example consider the infinite coin toss space $(\Omega_\infty, \mathcal{F}_c)$ of Example 1.1.13 with probability of a head on each toss equal to $1/2$ independently on all other tosses. For $k = 1, 2, \ldots$ let $X(\omega) = 2^k$ if $\omega_1 = \cdots = \omega_{k-1} = T$, $\omega_k = H$. This defines $X(\omega)$ for every sequence of coin tosses except the infinite sequence of all tails, whose probability is zero, for which we set $X(TTT\cdots) = \infty$. Note that X is finite a.s. However $\mathbf{E}X = \sum_k 2^k 2^{-k} = \infty$ (the latter example is the basis for a gambling question that give rise to what is known as the St. Petersburg paradox).*

REMARK. Using the elementary formula $\mathbf{E}Y = \sum_{m=1}^{N} c_m \mathbf{P}(A_m)$ for the simple function $Y(\omega) = \sum_{m=1}^{N} c_m I_{A_m}(\omega)$, as in Example 1.2.20, it can be shown that our definition of the expectation of $X \geq 0$ coincides with

$$\mathbf{E}X = \sup \{ \mathbf{E}Y : Y \in \mathrm{SF}, 0 \leq Y \leq X \}. \tag{1.2.2}$$

This alternative definition is useful for proving properties of the expectation, but less convenient for computing the value of $\mathbf{E}X$ for any specific X.

EXERCISE 1.2.22. *Write* $(\Omega, \mathcal{F}, \mathbf{P})$ *for a random experiment whose outcome is a recording of the results of* n *independent rolls of a balanced six-sided dice (including their order). Compute the expectation of the random variable* $D(\omega)$ *which counts the number of different faces of the dice recorded in these* n *rolls.*

DEFINITION 1.2.23. *We say that a R.V.* $X(\omega)$ *has a* probability density function f_X *if* $\mathbf{P}(a \le X \le b) = \int_a^b f_X(x)dx$ *for every* $a < b \in \mathbb{R}$. *Such density* f_X *must be a non-negative function with* $\int_{\mathbb{R}} f_X(x)dx = 1$.

PROPOSITION 1.2.24. *When a non-negative R.V.* $X(\omega)$ *has a probability density function* f_X*, our definition of the expectation coincides with the well known elementary formula* $\mathbf{E}X = \int_0^\infty x f_X(x)dx$.

PROOF. Fixing $X(\omega) \ge 0$ with a density $f_X(x)$, a finite m and a positive δ, let

$$\mathbf{E}_{\delta,m}(X) = \sum_{n=0}^{m} n\delta \int_{n\delta}^{(n+1)\delta} f_X(x)dx \, .$$

For any $x \in [n\delta, (n+1)\delta]$ we have that $x \ge n\delta \ge x - \delta$, implying that

$$\begin{aligned} \int_0^{(m+1)\delta} x f_X(x)dx &= \sum_{n=0}^{m} \int_{n\delta}^{(n+1)\delta} x f_X(x)dx \ge \mathbf{E}_{\delta,m}(X) \\ &\ge \sum_{n=0}^{m} \Big[\int_{n\delta}^{(n+1)\delta} x f_X(x)dx - \delta \int_{n\delta}^{(n+1)\delta} f_X(x)dx \Big] \\ &= \int_0^{(m+1)\delta} x f_X(x)dx - \delta \int_0^{(m+1)\delta} f_X(x)dx \, . \end{aligned}$$

Considering $m \to \infty$ we see that

$$\int_0^\infty x f_X(x)dx \ge \lim_{m\to\infty} \mathbf{E}_{\delta,m}(X) \ge \int_0^\infty x f_X(x)dx - \delta \, .$$

Taking $\delta \to 0$ the lower bounds converge to the upper bound, hence by (1.2.1),

$$\mathbf{E}X = \lim_{\delta\to 0} \lim_{m\to\infty} \mathbf{E}_{\delta,m}(X) = \int_0^\infty x f_X(x)dx \, ,$$

as stated. ∎

Definition 1.2.19 is called also *Lebesgue integral* of X with respect to the probability measure $\mathbf{P}$ and consequently denoted $\mathbf{E}X = \int X(\omega)d\mathbf{P}(\omega)$ (or $\int X(\omega)\mathbf{P}(d\omega)$). It is based on splitting the *range* of $X(\omega)$ to finitely many (small) intervals and approximating $X(\omega)$ by a constant on the corresponding set of ω values for which $X(\omega)$ falls into one such interval. This allows to deal with rather general domain Ω, in contrast to Riemann's integral where the *domain* of integration is split into finitely many (small) intervals – hence limited to $\mathbb{R}^d$. Even when $\Omega = [0, 1]$ it allows us to deal with measures $\mathbf{P}$ for which $\omega \mapsto \mathbf{P}([0, \omega])$ is not smooth (and hence Riemann's integral fails to exist). Though not done here, if the corresponding Riemann integral exists, then it necessarily coincides with our Lebesgue integral and with Definition 1.2.19 (as proved in Proposition 1.2.24 for R.V. X having a density).

We next extend the definition of the expectation from non-negative R.V.-s to general R.V.-s.

DEFINITION 1.2.25. *For a general R.V. X consider the non-negative R.V.-s $X_+ = \max(X,0)$ and $X_- = -\min(X,0)$ (so $X = X_+ - X_-$), and let* $\mathbf{E}\,X = \mathbf{E}\,X_+ - \mathbf{E}\,X_-$, *provided either* $\mathbf{E}\,X_+ < \infty$ *or* $\mathbf{E}\,X_- < \infty$.

DEFINITION 1.2.26. *We say that a random variable X is* integrable *(or has finite expectation) if* $\mathbf{E}|X| < \infty$, *that is, both* $\mathbf{E}\,X_+ < \infty$ *and* $\mathbf{E}\,X_- < \infty$.

EXERCISE 1.2.27. *Show that R.V. X is integrable if and only if* $\mathbf{E}[|X|I_{|X|>M}] \to 0$ *when* $M \to \infty$.

REMARK. Suppose $X = Y - Z$ for some non-negative R.V.-s Y and Z. Then, necessarily $Y = D + X_+$ and $Z = D + X_-$ for some $D \geq 0$ and all $\omega \in \Omega$. With some work it can be shown that the expectation is linear with respect to the addition of non-negative R.V. In particular, $\mathbf{E}Y = \mathbf{E}D + \mathbf{E}X_+$ and $\mathbf{E}Z = \mathbf{E}D + \mathbf{E}X_-$. Consequently,

$$\mathbf{E}Y - \mathbf{E}Z = \mathbf{E}X_+ - \mathbf{E}X_- = \mathbf{E}X$$

provided either $\mathbf{E}Y$ is finite or $\mathbf{E}Z$ is finite. We conclude that in this case $\mathbf{E}X = \mathbf{E}Y - \mathbf{E}Z$. However, it is possible to have X integrable while $\mathbf{E}D = \infty$ resulting with $\mathbf{E}Y = \mathbf{E}Z = \infty$.

EXAMPLE 1.2.28. *Using Proposition 1.2.24 it is easy to check that a R.V. X with a density f_X is integrable if and only if* $\int_{-\infty}^{\infty} |x| f_X(x)dx < \infty$, *in which case* $\mathbf{E}X = \int_{-\infty}^{\infty} x f_X(x)dx$.

Building on Example 1.2.28 we also have the following well known change of variables formula for the expectation.

PROPOSITION 1.2.29. *If a R.V. X has the probability density function f_X and $h : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function, then the R.V. $Y = h(X)$ is integrable if and only if* $\int_{-\infty}^{\infty} |h(x)| f_X(x)dx < \infty$, *in which case* $\mathbf{E}Y = \int_{-\infty}^{\infty} h(x) f_X(x)dx$.

DEFINITION 1.2.30. *A R.V. X with probability density function*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big), \qquad x \in \mathbb{R}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$, is called a non-degenerate Gaussian *(or* Normal*) R.V. with mean μ and variance σ^2, denoted by $X \sim N(\mu, \sigma^2)$.*

EXERCISE 1.2.31. *Verify that the* log-normal random variable Y, *that is* $Y = e^X$ *with* $X \sim N(\mu, \sigma^2)$, *has expected value* $\mathbf{E}Y = \exp(\mu + \sigma^2/2)$.

EXERCISE 1.2.32.

(a) *Find the values of $\alpha \in \mathbb{R}$ for which $f_X(x) = c_\alpha/(1 + |x|^\alpha)$ is a probability density function for some $c_\alpha \in \mathbb{R}$, and for such α write an expression for c_α (no need to compute the integral).*

(b) *Fixing $\alpha \in \mathbb{R}$ such that $f_X(x)$ is a probability density function, for which non-negative integer k is the random variable $Y = X^{2k+1}$ integrable? In case Y is integrable, compute its expectation* $\mathbf{E}Y$.

Typically, we cannot compute $\mathbf{E}X$ explicitly from the Definition 1.2.19. Instead, we either use the well known explicit formulas for discrete R.V.s and for R.V.s having a probability density function, or we appeal to properties of the expectation listed below.

Proposition 1.2.33. *The expectation has the following properties.*
(1). $\mathbf{E}I_A = \mathbf{P}(A)$ *for any* $A \in \mathcal{F}$.
(2). If $X(\omega) = \sum_{n=1}^{N} c_n I_{A_n}$ *is a simple function, then* $\mathbf{E}\ X = \sum_{n=1}^{N} c_n \mathbf{P}(A_n)$.
(3). If X *and* Y *are integrable R.V. then for any constants* α, β *the R.V.* $\alpha X + \beta Y$ *is integrable and* $\mathbf{E}\ (\alpha X + \beta Y) = \alpha(\mathbf{E}\ X) + \beta(\mathbf{E}\ Y)$.
(4). $\mathbf{E}\ X = c$ *if* $X(\omega) = c$ *with probability* 1.
(5). Monotonicity: If $X \geq Y$ *a.s., then* $\mathbf{E}X \geq \mathbf{E}Y$*. Further, if* $X \geq Y$ *a.s. and* $\mathbf{E}X = \mathbf{E}Y$*, then* $X = Y$ *a.s.*

In particular, property (3) of Proposition 1.2.33 tells us that the expectation is a linear functional on the vector space of all integrable R.V. (denoted below by L^1).

Exercise 1.2.34. *Using the definition of the expectation, prove the five properties detailed in Proposition 1.2.33.*

Indicator R.V.-s are often useful when computing expectations, as the following exercise illustrates.

Exercise 1.2.35. *A coin is tossed* n *times with the same probability* $0 < p < 1$ *for* H *showing on each toss, independently of all other tosses. A "run" is a sequence of tosses which result in the same outcome. For example, the sequence* $HHHTHTTH$ *contains five runs. Show that the expected number of runs is* $1 + 2(n-1)p(1-p)$.
Hint: *This is* [**GS01**, Exercise 3.4.1].

Since the explicit computation of the expectation is often not possible, we detail next a useful way to bound one expectation by another.

Proposition 1.2.36 (Jensen's inequality). *Suppose* $g(\cdot)$ *is a convex function, that is,*

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y) \qquad \forall\ x, y \in \mathbb{R}, \qquad 0 \leq \lambda \leq 1.$$

If X *is an integrable R.V. and* $g(X)$ *is also integrable, then* $\mathbf{E}(g(X)) \geq g(\mathbf{E}X)$.

Example 1.2.37. *Let* X *be a R.V. such that* $X(\omega) = xI_A + yI_{A^c}$ *and* $\mathbf{P}(A) = \lambda$*. Then,* $\mathbf{E}X = x\mathbf{P}(A) + y\mathbf{P}(A^c) = x\lambda + y(1-\lambda)$*. Further, then* $g(X(\omega)) = g(x)I_A + g(y)I_{A^c}$*, so for* g *convex,* $\mathbf{E}\ g(X) = g(x)\lambda + g(y)(1-\lambda) \geq g(x\lambda + y(1-\lambda)) = g(\mathbf{E}\ X)$.

The expectation is often used also as a way to bound tail probabilities, based on the following classical inequality.

Theorem 1.2.38 (Markov's inequality). *Suppose* f *is a non-decreasing, Borel measurable function with* $f(x) > 0$ *for any* $x > 0$*. Then, for any random variable* X *and all* $\varepsilon > 0$,

$$\mathbf{P}(|X(\omega)| > \varepsilon) \leq \frac{1}{f(\varepsilon)} \mathbf{E}(f(|X|)).$$

Proof. Let $A = \{\omega : |X(\omega)| > \varepsilon\}$. Since f is non-negative,

$$\mathbf{E}[f(|X|)] = \int f(|X(\omega)|) d\mathbf{P}(\omega) \geq \int I_A(\omega) f(|X(\omega)|) d\mathbf{P}(\omega).$$

Since f is non-decreasing, $f(|X(w)|) \geq f(\varepsilon)$ on the set A, implying that

$$\int I_A(\omega) f(|X(\omega)|) d\mathbf{P}(\omega) \geq \int I_A(\omega) f(\varepsilon) d\mathbf{P}(\omega) = f(\varepsilon)\mathbf{P}(A).$$

Dividing by $f(\varepsilon) > 0$ we get the stated inequality. ∎

We next specify the three most common instances of Markov's inequality.

EXAMPLE 1.2.39. *(a). Assuming* $X \geq 0$ *a.s. and taking* $f(x) = x$*, Markov's inequality is then,*

$$\mathbf{P}(X(\omega) \geq \varepsilon) \leq \frac{\mathbf{E}[X]}{\varepsilon}.$$

(b). Taking $f(x) = x^2$ *and* $X = Y - \mathbf{E}Y$*, Markov's inequality is then,*

$$\mathbf{P}(|Y - \mathbf{E}Y| \geq \varepsilon) = \mathbf{P}(|X(\omega)| \geq \varepsilon) \leq \frac{\mathbf{E}[X^2]}{\varepsilon^2} = \frac{\mathrm{Var}(Y)}{\varepsilon^2}.$$

(c). Taking $f(x) = e^{\theta x}$ *for some* $\theta > 0$*, Markov's Inequality is then,*

$$\mathbf{P}(X \geq \varepsilon) \leq e^{-\theta\varepsilon}\mathbf{E}[e^{\theta X}]\,.$$

In this bound we have an exponential decay in ε *at the cost of requiring* X *to have finite exponential moments.*

Here is an application of Markov's inequality for $f(x) = x^2$.

EXERCISE 1.2.40. *Show that if* $\mathbf{E}[X^2] = 0$ *then* $X = 0$ *almost surely.*

We conclude this section with *Schwarz inequality* (see also Proposition 2.2.4).

PROPOSITION 1.2.41. *Suppose* Y *and* Z *are random variables on the same probability space with both* $\mathbf{E}[Y^2]$ *and* $\mathbf{E}[Z^2]$ *finite. Then,* $\mathbf{E}|YZ| \leq \sqrt{\mathbf{E}Y^2\mathbf{E}Z^2}$*.*

1.3. Convergence of random variables

Asymptotic behavior is a key issue in probability theory and in the study of stochastic processes. We thus explore here the various notions of convergence of random variables and the relations among them. We start in Subsection 1.3.1 with the convergence for almost all possible outcomes versus the convergence of probabilities to zero. Then, in Subsection 1.3.2 we introduce and study the spaces of q-integrable random variables and the associated notions of convergence in q-means (that play an important role in the theory of stochastic processes).

Unless explicitly stated otherwise, throughout this section we assume that all R.V. are defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

1.3.1. Convergence almost surely and in probability. It is possible to find sequences of R.V. that have pointwise limits, that is $X_n(\omega) \to X(\omega)$ for all ω. One such example is $X_n(\omega) = a_n X(\omega)$, with non-random $a_n \to a$ for some $a \in \mathbb{R}$. However, this notion of convergence is in general not very useful, since it is sensitive to ill-behavior of the random variables on negligible sets of points ω. We provide next the more appropriate (and slightly weaker) alternative notion of convergence, that of almost sure convergence.

DEFINITION 1.3.1. *We say that random variables* X_n converge *to* X almost surely*, denoted* $X_n \stackrel{a.s.}{\to} X$*, if there exists* $A \in \mathcal{F}$ *with* $\mathbf{P}(A) = 1$ *such that* $X_n(\omega) \to X(\omega)$ *as* $n \to \infty$ *for each fixed* $\omega \in A$*.*

Just like the pointwise convergence, the convergence almost surely is invariant under application of a continuous function.

EXERCISE 1.3.2. *Show that if* $X_n \stackrel{a.s.}{\to} X$ *and* f *is a continuous function then* $f(X_n) \stackrel{a.s.}{\to} f(X)$ *as well.*

We next illustrate the difference between convergence pointwise and almost surely via a special instance of the Law of Large Numbers.

EXAMPLE 1.3.3. *Let $\overline{S}_n = \frac{1}{n}\sum_{i=1}^n I_{\{\omega_i=H\}}$ be the fraction of head counts in the first n independent fair coin tosses. That is, using the measurable space $(\Omega_\infty, \mathcal{F}_c)$ of Example 1.1.13 for which $\overline{S}_n$ are R.V. and endowing it with the probability measure $\mathbf{P}$ such that for each n, the restriction of $\mathbf{P}$ to the event space 2^{Ω_n} of first n tosses gives equal probability 2^{-n} to each of the 2^n possible outcomes. In this case, the* Law of Large Numbers *(L.L.N.) is the statement that $\overline{S}_n \overset{a.s.}{\to} \frac{1}{2}$. That is, as $n \to \infty$, the observed fraction of head counts in the first n independent fair coin tosses approach the probability of the coin landing Head, apart from a negligible collection of infinite sequences of coin tosses. However, note that there are $\omega \in \Omega_\infty$ for which the $\overline{S}_n(\omega)$ does not converge to $1/2$. For example, if $\omega_i = H$ for all i, then $\overline{S}_n = 1$ for all n.*

In principle, when dealing with almost sure convergence, one should check that the candidate limit X is also a random-variable. We always assume that the probability space is *complete* (as defined next), allowing us to hereafter ignore this technical point.

DEFINITION 1.3.4. *We say that $(\Omega, \mathcal{F}, \mathbf{P})$ is a* complete probability space *if any subset N of $B \in \mathcal{F}$ with $\mathbf{P}(B) = 0$ is also in $\mathcal{F}$.*
That is, a σ-field is made complete by adding to it all subsets of sets of zero probability (note that this procedure depends on the probability measure in use).

Indeed, it is possible to show that if X_n are R.V.-s such that $X_n(\omega) \to X(\omega)$ as $n \to \infty$ for each fixed $\omega \in A$ and $\mathbf{P}(A) = 1$, then there exists a R.V. $\overline{X}$ such that $N = \{\omega : X(\omega) \neq \overline{X}(\omega)\}$ is a subset of $B = A^c \in \mathcal{F}$ and $X_n \overset{a.s.}{\to} \overline{X}$. By assuming that the probability space is complete, we guarantee that N is in $\mathcal{F}$, consequently, $X \overset{a.s.}{=} \overline{X}$ is necessarily also a R.V.
A weaker notion of convergence is *convergence in probability* as defined next.

DEFINITION 1.3.5. *We say that X_n converge to X in probability, denoted $X_n \to_p X$, if $\mathbf{P}(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \to 0$ as $n \to \infty$, for any fixed $\varepsilon > 0$.*

THEOREM 1.3.6. *We have the following relations:*
(a) If $X_n \overset{a.s.}{\to} X$ then $X_n \to_p X$.
(b) If $X_n \to_p X$, then there exists a subsequence n_k such that $X_{n_k} \overset{a.s.}{\to} X$ for $k \to \infty$.

EXERCISE 1.3.7. *Let $B = \{\omega : \lim_n |X_n(\omega) - X(\omega)| = 0\}$ and fixing $\varepsilon > 0$ consider the sets $C_n = \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$, and the increasing sets $A_k = [\bigcup_{n\geq k} C_n]^c$. To practice your understanding at this point, explain why $X_n \overset{a.s.}{\to} X$ implies that $\mathbf{P}(A_k) \uparrow 1$, why this in turn implies that $\mathbf{P}(C_n) \to 0$ and why this verifies part (a) of Theorem 1.3.6.*

We will prove part (b) of Theorem 1.3.6 after developing the Borel-Cantelli lemmas, and studying a particular example, showing that convergence in probability does not imply in general convergence a.s.

PROPOSITION 1.3.8. *In general, $X_n \to_p X$ does not imply that $X_n \overset{a.s.}{\to} X$.*

PROOF. Consider the probability space $\Omega = (0, 1)$, with Borel σ-field and the Uniform probability measure U of Example 1.1.11. Suffices to construct an example of $X_n \to_p 0$ such that fixing each $\omega \in (0,1)$, we have that $X_n(\omega) = 1$ for infinitely many values of n. For example, this is the case when $X_n(\omega) = \mathbf{1}_{[t_n,t_n+s_n]}(\omega)$ with $s_n \downarrow 0$ as $n \to \infty$ slowly enough and $t_n \in [0, 1-s_n]$ are such that any $\omega \in [0,1]$ is in infinitely many intervals $[t_n, t_n + s_n]$. The latter property applies if $t_n = (i-1)/k$ and $s_n = 1/k$ when $n = k(k-1)/2 + i$, $i = 1, 2, \ldots, k$ and $k = 1, 2, \ldots$ (plot the intervals $[t_n, t_n + s_n]$ to convince yourself). ∎

Definition 1.3.9. *Let $\{A_n\}$ be a sequence of events, and $B_n = \bigcup_{k=n}^{\infty} A_k$. Define $A^\infty = \bigcap_{n=1}^{\infty} B_n$, so $\omega \in A^\infty$ if and only if $\omega \in A_k$ for infinitely many values of k.*

Our next result, called the first Borel-Cantelli lemma (see for example [**GS01**, Theorem 7.3.10, part (a)]), states that almost surely, A_k occurs for only finitely many values of k if the sequence $\mathbf{P}(A_k)$ converges to zero fast enough.

Lemma 1.3.10 (Borel-Cantelli I). *Suppose $A_k \in \mathcal{F}$ and $\sum_{k=1}^{\infty} \mathbf{P}(A_k) < \infty$. Then, necessarily $\mathbf{P}(A^\infty) = 0$.*

Proof. Let $b_n = \sum_{k=n}^{\infty} \mathbf{P}(A_k)$, noting that by monotonicity and countable sub-additivity of probability measures we have for all n that

$$\mathbf{P}(A^\infty) \le \mathbf{P}(B_n) \le \sum_{k=n}^{\infty} \mathbf{P}(A_k) := b_n$$

(see Exercise 1.1.3). Further, our assumption that the series b_1 is finite implies that $b_n \downarrow 0$ as $n \to \infty$, so considering the preceding bound $\mathbf{P}(A^\infty) \le b_n$ for $n \to \infty$, we conclude that $\mathbf{P}(A^\infty) = 0$. ∎

The following (strong) converse applies when in addition we know that the events $\{A_k\}$ are mutually independent (for a rigorous definition, see Definition 1.4.35).

Lemma 1.3.11 (Borel-Cantelli II). *If $\{A_k\}$ are independent and $\sum_{k=1}^{\infty} \mathbf{P}(A_k) = \infty$, then $\mathbf{P}(A^\infty) = 1$ (see* [**GS01**, Theorem 7.3.10, part (b)] *for the proof).*

We next use Borel-Cantelli Lemma I to prove part (b) of Theorem 1.3.6.

Proof of Theorem 1.3.6 part (*b*). Choose $a_l \to 0$. By the definition of convergence in probability there exist $n_l \uparrow \infty$ such that $\mathbf{P}(\{w : |X_{n_l}(\omega) - X(\omega)| > a_l\}) < 2^{-l}$. Define $A_l = \{w : |X_{n_l}(\omega) - X(\omega)| > a_l\}$. Then $\mathbf{P}(A_l) < 2^{-l}$, implying that $\sum_l \mathbf{P}(A_l) \le \sum_l 2^{-l} < \infty$. Therefore, by Borel-Cantelli I, $\mathbf{P}(A^\infty) = 0$. Now observe that $\omega \notin A^\infty$ amounts to $\omega \notin A_l$ for all but finitely many values of l, hence if $\omega \notin A^\infty$, then necessarily $X_{n_l}(\omega) \to X(\omega)$ as $l \to \infty$. To summarize, the measurable set $B := \{\omega : X_{n_l}(\omega) \to X(\omega)\}$ contains the set $(A^\infty)^c$ whose probability is one. Clearly then, $\mathbf{P}(B) = 1$, which is exactly what we set up to prove. ∎

Here is another application of Borel-Cantelli Lemma I which produces the a.s. convergence to zero of $k^{-1}X_k$ in case $\sup_n \mathbf{E}[X_n^2]$ is finite.

Proposition 1.3.12. *Suppose $\mathbf{E}[X_n^2] \le 1$ for all n. Then $n^{-1}X_n(\omega) \to 0$ a.s. for $n \to \infty$.*

PROOF. Fixing $\delta > 0$ let $A_k = \{\omega : |k^{-1}X_k(\omega)| > \delta\}$ for $k = 1, 2, \ldots$. Then, by part (b) of Example 1.2.39 and our assumption we have that

$$\mathbf{P}(A_k) = \mathbf{P}(\{\omega : |X_k(\omega)| > k\delta\}) \le \frac{\mathbf{E}(X_k^2)}{(k\delta)^2} \le \frac{1}{k^2}\frac{1}{\delta^2} .$$

Since $\sum_k k^{-2} < \infty$, it then follows by Borel Cantelli I that $\mathbf{P}(A^\infty) = 0$, where $A^\infty = \{\omega : |k^{-1}X_k(\omega)| > \delta$ for infinitely many values of $k\}$. Hence, for any fixed $\delta > 0$, with probability one $k^{-1}|X_k(\omega)| \le \delta$ for all large enough k, that is, $\limsup_{n\to\infty} n^{-1}|X_n(\omega)| \le \delta$ a.s. Considering a sequence $\delta_m \downarrow 0$ we conclude that $n^{-1}X_n \to 0$ for $n \to \infty$ and a.e. ω. ∎

A common use of Borel-Cantelli I is to prove convergence almost surely by applying the conclusion of the next exercise to the case at hand.

EXERCISE 1.3.13. *Suppose X and $\{X_n\}$ are random variables on the same probability space.*

(a) *Fixing $\varepsilon > 0$ and $A_n = \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$, show that $C_\varepsilon = \{\omega : \limsup_n |X_n(\omega) - X(\omega)| > \varepsilon\}$ is contained in A^∞ of Definition 1.3.9.*
(b) *Explain why if $\mathbf{P}(C_\varepsilon) = 0$ for any $\varepsilon > 0$, then $X_n \stackrel{a.s.}{\to} X$.*
(c) *Combining this and Borel-Cantelli I deduce that $X_n \stackrel{a.s.}{\to} X$ whenever $\sum_{n=1}^\infty \mathbf{P}(|X_n - X| > \varepsilon) < \infty$ for each $\varepsilon > 0$.*

We conclude this sub-section with a pair of exercises in which Borel-Cantelli lemmas help in finding certain asymptotic behavior for sequences of independent random variables.

EXERCISE 1.3.14. *Suppose T_n are independent* Exponential(1) *random variables (that is, $\mathbf{P}(T_n > t) = e^{-t}$ for $t \ge 0$).*

(a) *Using both Borel-Cantelli lemmas, show that*

$$\mathbf{P}(T_k(\omega) > \alpha \log k \textit{ for infinitely many values of } k) = \mathbf{1}_{\alpha \le 1} .$$

(b) *Deduce that* $\limsup_{n\to\infty}(T_n/\log n) = 1$ *almost surely.*

Hint: *See* [**Wil91**, Example 4.4] *for more details.*

EXERCISE 1.3.15. *Consider a two-sided infinite sequence of independent random variables $\{X_k, k \in \mathbf{Z}\}$ such that $\mathbf{P}(X_k = 1) = \mathbf{P}(X_k = 0) = 1/2$ for all $k \in \mathbf{Z}$ (for example, think of independent fair coin tosses). Let $\ell_m = \max\{i \ge 1 : X_{m-i+1} = \cdots = X_m = 1\}$ denote the length of the run of 1's going backwards from time m (with $\ell_m = 0$ in case $X_m = 0$). We are interested in the asymptotics of the longest such run during times $1, 2, \ldots, n$ for large n. That is,*

$$\begin{aligned} L_n &= \max\{\ell_m : m = 1, \ldots, n\} \\ &= \max\{m - k : X_{k+1} = \cdots = X_m = 1 \textit{ for some } m = 1, \ldots, n\} . \end{aligned}$$

(a) *Explain why $\mathbf{P}(\ell_m = k) = 2^{-(k+1)}$ for $k = 0, 1, 2, \ldots$ and any m.*
(b) *Applying the Borel-Cantelli I lemma for $A_n = \{\ell_n > (1+\varepsilon)\log_2 n\}$, show that for each $\varepsilon > 0$, with probability one, $\ell_n \le (1+\varepsilon)\log_2 n$ for all n large enough. Considering a countable sequence $\varepsilon_k \downarrow 0$ deduce that*

$$\limsup_{n\to\infty} \frac{L_n}{\log_2 n} \stackrel{a.s.}{\le} 1 .$$

(c) *Fixing* $\varepsilon > 0$ *let* $A_n = \{L_n < k_n\}$ *for* $k_n = [(1-\varepsilon)\log_2 n]$. *Explain why*

$$A_n \subseteq \bigcap_{i=1}^{m_n} B_i^c\,,$$

for $m_n = [n/k_n]$ *and the independent events* $B_i = \{X_{(i-1)k_n+1} = \cdots = X_{ik_n} = 1\}$. *Deduce from it that* $\mathbf{P}(A_n) \le \mathbf{P}(B_1^c)^{m_n} \le \exp(-n^\varepsilon/(2\log_2 n))$, *for all* n *large enough.*

(d) *Applying the Borel-Cantelli I lemma for the events* A_n *of part (c), followed by* $\varepsilon \downarrow 0$, *conclude that*

$$\liminf_{n\to\infty} \frac{L_n}{\log_2 n} \overset{a.s.}{\ge} 1$$

and consequently, that $L_n/(\log_2 n) \overset{a.s.}{\to} 1$.

1.3.2. L^q **spaces and convergence in** q**-mean.** Fixing $1 \le q < \infty$ we denote by $L^q(\Omega, \mathcal{F}, \mathbf{P})$ the collection of random variables X on $(\Omega, \mathcal{F})$ for which $\mathbf{E}\,(|X|^q) < \infty$ (and when the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is clear from the context, we often use the short notation L^q). For example, L^1 denotes the space of all integrable random-variables, and the random variables in L^2 are also called *square-integrable* .

We start by proving that these spaces are nested in terms of the parameter q.

PROPOSITION 1.3.16. *The sequence* $||X||_q = [\mathbf{E}\,(|X|^q)]^{1/q}$ *is non-decreasing in* q.

PROOF. Fix $q_1 > q_2$ and apply Jensen's inequality for the convex function $g(y) = |y|^{q_1/q_2}$ and the non-negative random variable $Y = |X|^{q_2}$, to get that $\mathbf{E}\left(|Y|^{q_1/q_2}\right) \ge (\mathbf{E}Y)^{q_1/q_2}$. Taking the $1/q_1$ power yields the stated result. ∎

Associated with the space L^q is the notion of convergence in q-mean, which we now define.

DEFINITION 1.3.17. *We say that* X_n converges in q-mean, *or in* L^q *to* X, *denoted* $X_n \overset{q.m.}{\to} X$, *if* $X_n, X \in L^q$ *and* $||X_n - X||_q \to 0$ *as* $n \to \infty$ *(i.e.,* $\mathbf{E}\,(|X_n - X|^q) \to 0$ *as* $n \to \infty$.

EXAMPLE 1.3.18. *For* $q = 2$ *we have the explicit formula*

$$||X_n - X||_2^2 = \mathbf{E}(X_n^2) - 2\mathbf{E}(X_n X) + \mathbf{E}(X^2).$$

Thus, it is often easiest to check convergence in 2*-mean.*

Check that the following claim is an immediate corollary of Proposition 1.3.16.

COROLLARY 1.3.19. *If* $X_n \overset{q.m.}{\to} X$ *and* $q \ge r$, *then* $X_n \overset{r.m.}{\to} X$.

Our next proposition details the most important general structural properties of the spaces L^q.

PROPOSITION 1.3.20. $L^q(\Omega, \mathcal{F}, \mathbf{P})$ *is a complete, normed (topological) vector space with the norm* $||\cdot||_q$; *That is,* $\alpha X + \beta Y \in L^q$ *whenever* $X, Y \in L^q$, $\alpha, \beta \in \mathbb{R}$, *with* $X \mapsto ||X||_q$ *a norm on* L^q *and if* $X_n \in L^q$ *are such that* $||X_n - X_m||_q \to 0$ *as* $n, m \to \infty$ *then* $X_n \overset{q.m.}{\to} X$ *for some* $X \in L^q$.

For example, check the following claim.

EXERCISE 1.3.21. *Fixing $q \geq 1$, use the triangle inequality for the norm $\|\cdot\|_q$ on L^q to show that if $X_n \overset{q.m.}{\longrightarrow} X$, then $\mathbf{E}|X_n|^q \to \mathbf{E}|X|^q$. Using Jensen's inequality for $g(x) = |x|$, deduce that also $\mathbf{E}X_n \to \mathbf{E}X$. Finally, provide an example to show that $\mathbf{E}X_n \to \mathbf{E}X$ does not necessarily imply $X_n \to X$ in L^1.*

As we prove below by an application of Markov's inequality, convergence in q-mean implies convergence in probability (for any value of q).

PROPOSITION 1.3.22. *If $X_n \overset{q.m.}{\to} X$, then $X_n \to_p X$.*

PROOF. For $f(x) = |x|^q$ Markov's inequality (i.e., Theorem 1.2.38) says that

$$\mathbf{P}(|Y| > \varepsilon) \leq \varepsilon^{-q}\mathbf{E}[|Y|^q].$$

Taking $Y = X_n - X$ gives $\mathbf{P}(|X_n - X| > \varepsilon) \leq \varepsilon^{-q}\mathbf{E}[|X_n - X|^q]$. Thus, if $X_n \overset{q.m.}{\to} X$ we necessarily also have that $X_n \to_p X$ as claimed. ∎

EXAMPLE 1.3.23. *The converse of Proposition 1.3.22 does not hold in general. For example, take the probability space of Example 1.1.11 and the R.V. $Y_n(\omega) = n\mathbf{1}_{[0,n^{-1}]}$. Since $Y_n(\omega) = 0$ for all $n \geq n_0$ and some finite $n_0 = n_0(\omega)$, it follows that $Y_n(\omega)$ converges a.s. to $Y(\omega) = 0$ and hence converges to zero in probability as well (see part (a) of Theorem 1.3.6). However, Y_n does not converge to zero in q-mean, even for the easiest case of $q = 1$ since $\mathbf{E}[Y_n] = nU([0, n^{-1}]) = 1$ for all n. The convergence in q-mean is in general not comparable to a.s. convergence. Indeed, the above example of $Y_n(\omega)$ with convergence a.s. but not in 1-mean is complemented by the example considered when proving Proposition 1.3.8 which converges to zero in q-mean but not almost surely.*

Your next exercise summarizes all that can go wrong here.

EXERCISE 1.3.24. *Give a counterexample to each of the following claims*

(a) *If $X_n \to X$ a.s. then $X_n \to X$ in L^q, $q \geq 1$;*
(b) *If $X_n \to X$ in L^q then $X_n \to X$ a.s.;*
(c) *If $X_n \to X$ in probability then $X_n \to X$ a.s.*

While neither convergence in 1-mean nor convergence a.s. are a consequence of each other, a sequence of R.V.s cannot have a 1-mean limit and a different a.s. limit.

PROPOSITION 1.3.25. *If $X_n \overset{q.m.}{\to} X$ and $X_n \overset{a.s.}{\to} Y$ then $X = Y$ a.s.*

PROOF. It follows from Proposition 1.3.22 that $X_n \to_p X$ and from part (a) of Theorem 1.3.6 that $X_n \to_p Y$. Note that for any $\varepsilon > 0$, and $\omega \in \Omega$, if $|Y(\omega) - X(\omega)| > 2\varepsilon$ then either $|X_n(\omega) - X(\omega)| > \varepsilon$ or $|X_n(\omega) - Y(\omega)| > \varepsilon$. Hence,

$$\begin{aligned}\mathbf{P}(\{\omega : |Y(\omega) - X(\omega)| > 2\varepsilon\}) &\leq \mathbf{P}(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \\ &+ \mathbf{P}(\{\omega : |X_n(\omega) - Y(\omega)| > \varepsilon\}).\end{aligned}$$

By Definition 1.3.5, both terms on the right hand side converge to zero as $n \to \infty$, hence $\mathbf{P}(\{\omega : |Y(\omega) - X(\omega)| > 2\varepsilon\}) = 0$ for each $\varepsilon > 0$, that is $X = Y$ a.s. ∎

REMARK. Note that $Y_n(\omega)$ of Example 1.3.23 are such that $\mathbf{E}Y_n = 1$ for every n, but certainly $\mathbf{E}[|Y_n - 1|] \not\to 0$ as $n \to \infty$. That is, $Y_n(\omega)$ does not converge to one in 1-mean. Indeed, $Y_n \overset{a.s.}{\to} 0$ as $n \to \infty$, so by Proposition 1.3.25 these Y_n simply do not have any 1-mean limit as $n \to \infty$.

1.4. Independence, weak convergence and uniform integrability

This section is devoted to independence (Subsection 1.4.3) and distribution (Subsection 1.4.1), the two fundamental aspects that differentiate probability from (general) measure theory. In the process, we consider in Subsection 1.4.1 the useful notion of convergence in law (or more generally, that of weak convergence), which is weaker than all notions of convergence of Section 1.3, and devote Subsection 1.4.2 to uniform integrability, a technical tool which is highly useful when attempting to exchange the order of a limit and expectation operations.

1.4.1. Distribution, law and weak convergence. As defined next, every R.V. X induces a probability measure on its range which is called the law of X.

Definition 1.4.1. *The* law *of a R.V. X, denoted $\mathcal{P}_X$, is the probability measure on $(\mathbb{R}, \mathcal{B})$ such that $\mathcal{P}_X(B) = \mathbf{P}(\{\omega : X(\omega) \in B\})$ for any Borel set B.*

Exercise 1.4.2. *For a R.V. defined on $(\Omega, \mathcal{F}, \mathbf{P})$ verify that $\mathcal{P}_X$ is a probability measure on $(\mathbb{R}, \mathcal{B})$.*
Hint: *First show that for $B_i \in \mathcal{B}$, $\{\omega : X(\omega) \in \cup_i B_i\} = \cup_i \{\omega : X(\omega) \in B_i\}$ and that if B_i are disjoint then so are the sets $\{\omega : X(\omega) \in B_i\}$.*

Note that the law $\mathcal{P}_X$ of a R.V. $X : \Omega \longrightarrow \mathbb{R}$, determines the values of the probability measure $\mathbf{P}$ on $\sigma(X) = \sigma(\{\omega : X(\omega) \leq \alpha\}, \alpha \in \mathbb{R})$. Further, recalling the remark following Definition 1.2.8, convince yourself that $\mathcal{P}_X$ carries exactly the same information as the restriction of $\mathbf{P}$ to $\sigma(X)$. We shall pick this point of view again when discussing stochastic processes in Section 3.1.

The law of a R.V. motivates the following "change of variables" formula which is useful in computing expectations (the special case of R.V. having a probability density function is given already in Proposition 1.2.29).

Proposition 1.4.3. *Let X be a R.V. on $(\Omega, \mathcal{F}, \mathbf{P})$ and let g be a Borel function on $\mathbb{R}$. Suppose either g is non-negative or $\mathbf{E}|g(X)| < \infty$. Then*

$$\mathbf{E}[g(X)] = \int_{\mathbb{R}} g(x) d\mathcal{P}_X(x)\,, \tag{1.4.1}$$

where the integral on the right hand side merely denotes the expectation of the random variable $g(x)$ on the (new) probability space $(\mathbb{R}, \mathcal{B}, \mathcal{P}_X)$.

A good way to practice your understanding of Definition 1.4.1 is by verifying that if $X = Y$ almost surely, then also $\mathcal{P}_X = \mathcal{P}_Y$ (that is, any two random variables we consider to be the same would indeed have the same law).

Exercise 1.4.4. *Convince yourself that Proposition 1.4.3 implies that if $\mathcal{P}_X = \mathcal{P}_Y$ then also $\mathbf{E}h(X) = \mathbf{E}h(Y)$ for any bounded Borel function $h : \mathbb{R} \to \mathbb{R}$ and prove the converse statement: If $\mathbf{E}h(X) = \mathbf{E}h(Y)$ for any bounded Borel function $h : \mathbb{R} \to \mathbb{R}$ then necessarily $\mathcal{P}_X = \mathcal{P}_Y$.*

The next concept we define, the distribution function, is closely associated with the law $\mathcal{P}_X$ of the R.V.

Definition 1.4.5. *The* distribution function *F_X of a real-valued R.V. X is*

$$F_X(\alpha) = \mathbf{P}(\{\omega : X(\omega) \leq \alpha\}) = \mathcal{P}_X((-\infty, \alpha]) \qquad \forall \alpha \in \mathbb{R}$$

As we have that $\mathbf{P}(\{\omega : X(\omega) \leq \alpha\}) = F_X(\alpha)$ for the generators $\{\omega : X(\omega) \leq \alpha\}$ of $\sigma(X)$, we are not at all surprised by the following proposition.

PROPOSITION 1.4.6. *The distribution function F_X uniquely determines the law $\mathcal{P}_X$ of X.*

Our next example highlights the possible shape of the distribution function.

EXAMPLE 1.4.7. *Consider Example 1.1.4 of n coin tosses, with σ-field $\mathcal{F}_n = 2^{\Omega_n}$, sample space $\Omega_n = \{H,T\}^n$, and the probability measure $\mathbf{P}_n(A) = \sum_{\omega \in A} p_\omega$, where $p_\omega = 2^{-n}$ for each $\omega \in \Omega_n$ (that is, $\omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$ for $\omega_i \in \{H,T\}$), corresponding to independent, fair, coin tosses. Let $Y(\omega) = I_{\{\omega_1 = H\}}$ measure the outcome of the first toss. The law of this random variable is,*

$$\mathcal{P}_Y(B) = \frac{1}{2}\mathbf{1}_{\{0 \in B\}} + \frac{1}{2}\mathbf{1}_{\{1 \in B\}}$$

and its distribution function is

$$F_Y(\alpha) = \mathcal{P}_Y((-\infty, \alpha]) = \mathbf{P}_n(Y(\omega) \le \alpha) = \begin{cases} 1, & \alpha \ge 1 \\ \frac{1}{2}, & 0 \le \alpha < 1 \\ 0, & \alpha < 0 \end{cases} . \tag{1.4.2}$$

Note that in general $\sigma(X)$ is a strict subset of the σ-field $\mathcal{F}$ (in Example 1.4.7 we have that $\sigma(Y)$ determines the probability measure for the first coin toss, but tells us nothing about the probability measure assigned to the remaining $n-1$ tosses). Consequently, though the law $\mathcal{P}_X$ determines the probability measure $\mathbf{P}$ on $\sigma(X)$ it usually does not completely determine $\mathbf{P}$.

Example 1.4.7 is somewhat generic, in the sense that if the R.V. X is a simple function, then its distribution function F_X is piecewise constant with jumps at the possible values that X takes and jump sizes that are the corresponding probabilities. In contrast, the distribution function of a R.V. with a density (per Definition 1.2.23) is almost everywhere differentiable, that is,

PROPOSITION 1.4.8. *A R.V. X has a (probability)* density *(function) f_X if and only if its distribution function F_X can be expressed as*

$$F_X(\alpha) = \int_{-\infty}^{\alpha} f_X(x)dx \,,$$

for all $\alpha \in \mathbb{R}$ (where $f_X \ge 0$ and $\int f_X(x)dx = 1$). Such F_X is continuous and almost everywhere differentiable with $\frac{dF_X}{dx}(x) = f_X(x)$ for almost every x.

EXAMPLE 1.4.9. *The distribution function for the R.V. $U(\omega) = \omega$ corresponding to Example 1.1.11 is*

$$F_U(\alpha) = \mathbf{P}(U \le \alpha) = \mathbf{P}(U \in [0, \alpha]) = \begin{cases} 1, & \alpha > 1 \\ \alpha, & 0 \le \alpha \le 1 \\ 0, & \alpha < 0 \end{cases} \tag{1.4.3}$$

and its density is $f_U(u) = \begin{cases} 1, & 0 \le u \le 1 \\ 0, & \textit{otherwise} \end{cases}$.

Every real-valued R.V. X has a distribution function but not necessarily a density. For example $X = 0$ w.p.1 has distribution function $F_X(\alpha) = \mathbf{1}_{\alpha \ge 0}$. Since F_X is discontinuous at 0 the R.V. X does not have a density.

The distribution function of any R.V. is necessarily non-decreasing. Somewhat surprisingly, there are continuous, non-decreasing functions that do not equal to the integral of their almost everywhere derivative. Such are the distribution functions of the non-discrete random variables that do not have a density. For the characterization of all possible distribution functions c.f. [**GS01**, Section 2.3].

Associated with the law of R.V.-s is the important concept of weak convergence which we define next.

Definition 1.4.10. *We say that R.V.-s X_n* converge in law (or weakly) *to a R.V. X, denoted by $X_n \xrightarrow{\mathcal{L}} X$ if $F_{X_n}(\alpha) \to F_X(\alpha)$ as $n \to \infty$ for each fixed α which is a continuity point of F_X (where α is a continuity point of $F_X(\cdot)$ if $F_X(\alpha_k) \to F_X(\alpha)$ whenever $\alpha_k \to \alpha$). In some books this is also called convergence in distribution, and denoted $X_n \xrightarrow{\mathcal{D}} X$.*

The next proposition, whose proof we do not provide, gives an *alternative definition of convergence in law*. Though not as easy to check as Definition 1.4.10, this alternative definition applies to more general R.V., whose range is not $\mathbb{R}$, for example to random vectors X_n with values in $\mathbb{R}^d$.

Proposition 1.4.11. *$X_n \xrightarrow{\mathcal{L}} X$ if and only if for each bounded h that is continuous on the range of X we have $\mathbf{E}h(X_n) \to \mathbf{E}h(X)$ as $n \to \infty$.*

Remark. Note that $F_X(\alpha) = \mathcal{P}_X((-\infty, \alpha]) = \mathbf{E}[I_{(-\infty,\alpha]}(X)]$ involves the discontinuous function $h(x) = 1_{(-\infty,\alpha]}(x)$. Restricting the convergence in law to continuity points of F_X is what makes Proposition 1.4.11 possible.

Exercise 1.4.12. *Show that if $X_n \xrightarrow{\mathcal{L}} X$ and f is a continuous function then $f(X_n) \xrightarrow{\mathcal{L}} f(X)$.*

We next illustrate the concept of convergence in law by a special instance of the Central Limit Theorem (one that is also called the Normal approximation for the Binomial distribution).

Example 1.4.13. *Let $\widehat{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (I_{\{\omega_i = H\}} - I_{\{\omega_i = T\}})$ be the normalized difference between head and tail counts in n independent fair coin tosses, that is, using the probability space $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$ of Example 1.4.7 (convince yourself that $\widehat{S}_n$ is a R.V. with respect to $(\Omega_n, \mathcal{F}_n)$). In this case, the* Central Limit Theorem *(C.L.T.) is the statement that $\widehat{S}_n \xrightarrow{\mathcal{L}} G$, where G is a Gaussian R.V. of zero mean and variance one, that is, having the distribution function*

$$F_G(g) = \int_{-\infty}^{g} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

(it is not hard to check that indeed $\mathbf{E}(G) = 0$ and $\mathbf{E}(G^2) = 1$; such R.V. is sometimes also called standard Normal*).*

By Proposition 1.4.11, this C.L.T. tells us that $\mathbf{E}[h(\widehat{S}_n)] \xrightarrow{n\to\infty} \mathbf{E}[h(G)]$ for each continuous and bounded function $h : \mathbb{R} \to \mathbb{R}$, where

$$\mathbf{E}h(\widehat{S}_n) = \sum_{\omega \in \Omega_n} 2^{-n} h(\widehat{S}_n(\omega)) \tag{1.4.4}$$

$$\mathbf{E}h(G) = \int_{-\infty}^{\infty} h(x) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx\,, \tag{1.4.5}$$

and the expression (1.4.5) is an instance of Proposition 1.2.29 (as G has a density $f_G(g) = (2\pi)^{-1/2}e^{-g^2/2}$). So, the C.L.T. allows us to approximate, for all large enough n, the non-computable sums (1.4.4) by the computable integrals (1.4.5).

Here is another example of convergence in law, this time in the context of extreme value theory.

EXERCISE 1.4.14. *Let $M_n = \max_{1\le i\le n}\{T_i\}$, where T_i, $i = 1, 2, \dots$ are independent* Exponential(λ) *random variables (i.e. $F_{T_i}(t) = 1 - e^{-\lambda t}$ for some $\lambda > 0$, all $t \ge 0$ and any i). Find non-random numbers a_n and a non-zero random variable M_∞ such that $(M_n - a_n)$ converges in law to M_∞.*
Hint: *Explain why $F_{M_n - a_n}(t) = (1 - e^{-\lambda t}e^{-\lambda a_n})^n$ and find $a_n \to \infty$ for which $(1 - e^{-\lambda t}e^{-\lambda a_n})^n$ converges per fixed t and its limit is strictly between 0 and 1.*

If the limit X has a density, or more generally whenever F_X is a continuous function, the convergence in law of X_n to X is equivalent to the pointwise convergence of the corresponding distribution functions. Such is the case in Example 1.4.13 since G has a density.

As we demonstrate next, in general, the convergence in law of X_n to X is *strictly* weaker than the pointwise convergence of the corresponding distribution functions and the rate of convergence of $\mathbf{E}h(X_n)$ to $\mathbf{E}h(X)$ depends on the specific function h we consider.

EXAMPLE 1.4.15. *The random variables $X_n = 1/n$ a.s. converge in law to $X = 0$, while $F_{X_n}(\alpha)$ does not converge to $F_X(\alpha)$ at the discontinuity point $\alpha = 0$ of F_X. Indeed, $F_{X_n}(\alpha) = \mathbf{1}_{[1/n,\infty)}(\alpha)$ converges to $F_X(\alpha) = \mathbf{1}_{[0,\infty)}(\alpha)$ for each $\alpha \ne 0$, since $F_X(\alpha) = 0 = F_{X_n}(\alpha)$ for all n and $\alpha < 0$, whereas for $\alpha > 0$ and all n large enough $\alpha > 1/n$ in which case $F_{X_n}(\alpha) = 1 = F_X(\alpha)$ as well. However, $F_X(0) = 1$ while $F_{X_n}(0) = 0$ for all n.*

Further, while $\mathbf{E}h(X_n) = h(\frac{1}{n}) \to h(0) = \mathbf{E}h(X)$ for each fixed continuous function h, the rate of convergence clearly varies with the choice of h.

In contrast to the preceding example, we provide next a very explicit necessary and sufficient condition for convergence in law of integer valued random variables.

EXERCISE 1.4.16. *Let $X_n, 1 \le n \le \infty$ be non-negative integer valued R.V.-s. Show that $X_n \to X_\infty$ in law if and only if $\mathbf{P}(X_n = m) \to_{n\to\infty} \mathbf{P}(X_\infty = m)$ for all m.*

The next exercise provides a few additional examples as well as the dual of Exercise 1.4.16 when each of the random variables X_n, $1 \le n \le \infty$ has a density.

EXERCISE 1.4.17.

(a) *Give an example of random variables X and Y on the same probability space, such that $\mathcal{P}_X = \mathcal{P}_Y$ while $\mathbf{P}(\{\omega : X(\omega) \ne Y(\omega)\}) = 1$.*
(b) *Give an example of random variables $X_n \xrightarrow{\mathcal{L}} X_\infty$ where each X_n has a probability density function, but X_∞ does not have such.*
(c) *Suppose Z_p denotes a random variable with a Geometric distribution of parameter $1 > p > 0$, that is $\mathbf{P}(Z_p = k) = p(1-p)^{k-1}$ for $k = 1, 2, \dots$. Show that $\mathbf{P}(pZ_p > t) \to e^{-t}$ as $p \to 0$, for each $t \ge 0$ and deduce that pZ_p converge in law to the* Exponential *random variable T, whose density is $f_T(t) = e^{-t}\mathbf{1}_{t\ge 0}$.*

(d) *Suppose R.V.-s X_n and X_∞ have (Borel measurable) densities $f_n(s)$ and $f_\infty(s)$, respectively, such that $f_n(s) \to f_\infty(s)$ as $n \to \infty$, for each fixed $s \in \mathbb{R}$ and further that f_∞ is strictly positive on $\mathbb{R}$. Let $g_n(s) = 2\max(0, 1 - f_n(s)/f_\infty(s))$. Explain why (recall Definition 1.2.23)*
$$\int_{\mathbb{R}} |f_n(s) - f_\infty(s)| ds = \int_{\mathbb{R}} g_n(s) f_\infty(s) ds,$$
why it follows from Corollary 1.4.28 that $\int_{\mathbb{R}} g_n(s) f_\infty(s) ds \to 0$ as $n \to \infty$ and how you deduce from this that $X_n \xrightarrow{\mathcal{L}} X_\infty$.

Our next proposition states that convergence in probability implies the convergence in law. This is perhaps one of the reasons we call the latter also *weak* convergence.

Proposition 1.4.18. *If $X_n \to_p X$, then $X_n \xrightarrow{\mathcal{L}} X$.*

The convergence in law does not require X_n to be defined in the same probability space (indeed, in Example 1.4.13 we have $\widehat{S}_n$ in a different probability space for each n. Though we can embed all of these R.V.-s in the same larger measurable space $(\Omega_\infty, \mathcal{F}_c)$ of Example 1.1.13, we still do not see their limit in law G in Ω_∞). Consequently, the converse of Proposition 1.4.18 cannot hold. Nevertheless, as we state next, when the limiting R.V. is a non-random constant, the convergence in law is equivalent to convergence in probability.

Proposition 1.4.19. *If $X_n \xrightarrow{\mathcal{L}} X$ and X is a non-random constant (almost surely), then $X_n \to_p X$ as well.*

We define next a natural extension of the convergence in law, which is very convenient for dealing with the convergence of stochastic processes.

Definition 1.4.20. *We say that a sequence of probability measures Q_n on a topological space $\mathbb{S}$ (i.e. a set $\mathbb{S}$ with a notion of open sets, or topology) and its Borel σ-field $\mathcal{B}_{\mathbb{S}}$ (= the σ-field generated by the open subsets of $\mathbb{S}$), converges weakly to a probability measure Q if for each fixed h continuous and bounded on $\mathbb{S}$,*
$$\int_{\mathbb{S}} h(\omega) dQ_n(\omega) \longrightarrow \int_{\mathbb{S}} h(\omega) dQ(\omega)\,,$$
as $n \to \infty$ (here $\int_{\mathbb{S}} h dQ$ denotes the expectation of the R.V. $h(\omega)$ in the probability space $(\mathbb{S}, \mathcal{B}_{\mathbb{S}}, Q)$, while $\int_{\mathbb{S}} h dQ_n$ corresponds to the space $(\mathbb{S}, \mathcal{B}_{\mathbb{S}}, Q_n)$). We shall use $Q_n \Rightarrow Q$ to denote weak convergence.

Example 1.4.21. *In particular, $X_n \xrightarrow{\mathcal{L}} X$ if and only if $\mathcal{P}_{X_n} \Rightarrow \mathcal{P}_X$. That is, when*
$$\int_{\mathbb{R}} h(\xi) d\mathcal{P}_{X_n}(\xi) \to \int_{\mathbb{R}} h(\xi) d\mathcal{P}_X(\xi),$$
for each fixed $h : \mathbb{R} \to \mathbb{R}$ continuous and bounded.

1.4.2. Uniform integrability, limits and expectation.

Recall that either convergence a.s. or convergence in q-mean imply convergence in probability, which in turn gives convergence in law. While we have seen that a.s. convergence and convergence in q-means are in general non-comparable, we next give an integrability condition that together with convergence in probability implies convergence in q-mean.

Definition 1.4.22. *A collection of R.V.-s* $\{X_\alpha, \alpha \in \mathcal{I}\}$ *is called* Uniformly Integrable *(U.I.) if*

$$\lim_{M\to\infty} \sup_\alpha \mathbf{E}[|X_\alpha| I_{|X_\alpha|>M}] = 0 .$$

Theorem 1.4.23. *(for proof see* [**GS01**, Theorem 7.10.3]*): If* $X_n \to_p X$ *and* $|X_n|^q$ *are U.I., then* $X_n \stackrel{q.m.}{\longrightarrow} X$.

We next detail a few special cases of U.I. collections of R.V.-s, showing among other things that any finite or uniformly bounded collection of integrable R.V.-s is U.I.

Example 1.4.24. *By Exercise 1.2.27 any* $X \in L^1$ *is U.I. Applying the same reasoning, if* $|X_\alpha| \le Y$ *for all* α *and some R.V.* Y *such that* $\mathbf{E}Y < \infty$, *then* X_α *are U.I. (indeed, in this case,* $|X_\alpha| I_{|X_\alpha|>M} \le Y I_{Y>M}$, *hence* $\mathbf{E}[|X_\alpha| I_{|X_\alpha|>M}] \le \mathbf{E}[Y I_{Y>M}]$, *which does not depend on* α *and as you have shown in Exercise 1.2.27, converges to zero when* $M \to \infty$*). In particular, a collection of R.V.-s* X_α *is U.I. when for some non-random* $C < \infty$ *and all* α *we have* $|X_\alpha| \le C$ *(take* $Y = C$ *in the preceding, or just try* $M > C$ *in Definition 1.4.22). Other such examples are the U.I. collection of R.V.* $\{c_\alpha Y\}$, *where* $Y \in L^1$ *and* $c_\alpha \in [-1, 1]$ *are non-random, and the U.I. collection* X_α *for countable* $\mathcal{I}$ *such that* $\mathbf{E}(\sup_\alpha |X_\alpha|) < \infty$ *(just take* $Y = \sup_\alpha |X_\alpha|$*). Finally, any finite collection of R.V.-s* X_i, $i = 1, \ldots, k$ *in* L^1 *is U.I. (simply take* $Y = |X_1| + |X_2| + \cdots + |X_k|$ *so* $Y \in L^1$ *and* $|X_i| \le Y$ *for* $i = 1, \ldots, k$*).*

In contrast, here is a concrete example of a sequence of R.V. X_n which is not U.I. Consider the probability space $(\Omega_\infty, \mathcal{F}_c, \mathbf{P})$ of fair coin tosses, as in Example 1.3.3 and let $X_n(\omega) = \inf\{i > n : \omega_i = H\}$, that is, the index of the first toss after the n-th one, for which the coin lands Head. Indeed, if $n \ge M$ then $X_n > n \ge M$, hence $\mathbf{E}[X_n I_{X_n>M}] = \mathbf{E}[X_n] > n$, implying that $\{X_n\}$ is not U.I.
In the next exercise you provide a criterion for U.I. that is very handy and often used by us (in future chapters).

Exercise 1.4.25. *A collection of R.V.* $\{X_\alpha : \alpha \in \mathcal{I}\}$ *is uniformly integrable if* $\mathbf{E}f(|X_\alpha|) \le C$ *for some finite* C *and all* $\alpha \in \mathcal{I}$, *where* $f \ge 0$ *is any function such that* $f(x)/x \to \infty$ *as* $x \to \infty$. *Verify this statement for* $f(x) = |x|^q$, $q > 1$.

Note that in Exercise 1.4.25 we must have $q > 1$, and thus $\sup_\alpha \mathbf{E}|X_\alpha| < \infty$ alone is not enough for the collection to be U.I. The following lemma shows what is required in addition; for a proof see [**GS01**, Lemma 7.10.6].

Lemma 1.4.26. *A collection of R.V.* $\{X_\alpha : \alpha \in \mathcal{I}\}$ *is uniformly integrable if and only if:*

(a) $\sup_\alpha \mathbf{E}|X_\alpha| < \infty$,
(b) *for any* $\varepsilon > 0$, *there exists a* $\delta > 0$ *such that* $\mathbf{E}(|X_\alpha| I_A) < \varepsilon$ *for all* $\alpha \in \mathcal{I}$ *and events* A *for which* $\mathbf{P}(A) < \delta$.

We now address the technical but important question of when can one exchange the order of limits and expectation operations. To see that this is not always possible, consider the R.V. $Y_n(\omega) = n\mathbf{1}_{[0,n^{-1}]}$ of Example 1.3.23 (defined on the probability space $(\mathbb{R}, \mathcal{B}, U)$). Since $\mathbf{E}[Y_n I_{Y_n>M}] = \mathbf{1}_{n>M}$, it follows that $\sup_n \mathbf{E}[Y_n I_{Y_n>M}] = 1$ for all M, hence Y_n are not a U.I. collection. Indeed, $Y_n \to 0$ a.s. while $\lim_{n\to\infty} \mathbf{E}(Y_n) = 1 > 0$.

In view of Example 1.4.24, our next result is merely a corollary of Theorem 1.4.23.

THEOREM 1.4.27 (Dominated Convergence). *If there exists a random variable Y such that $\mathbf{E}Y < \infty$, $|X_n| \le Y$ for all n and $X_n \to_p X$, then $\mathbf{E}X_n \longrightarrow \mathbf{E}X$.*

Considering non-random Y we get the following corollary,

COROLLARY 1.4.28 (Bounded Convergence). *Suppose $|X_n| \le C$ for some finite constant C and all n. If $X_n \to_p X$ then also $\mathbf{E}X_n \to \mathbf{E}X$.*

In case of monotone (upward) convergence of non-negative R.V. X_n to X we can exchange the order of limit and expectation even when X *is not integrable*, that is,

THEOREM 1.4.29 (Monotone Convergence). *If $X_n \ge 0$ and $X_n(\omega) \uparrow X(\omega)$ for a.e. ω, then $\mathbf{E}X_n \uparrow \mathbf{E}X$. This applies even if $X(\omega) = \infty$ for some $\omega \in \Omega$.*

Suppose $X_n \ge 0$ and $X_n(\omega) \downarrow X(\omega)$ for a.e. ω. If we assume in addition that $\mathbf{E}X_1 < \infty$, then we get the convergence $\mathbf{E}X_n \downarrow \mathbf{E}X$ already from the dominated convergence theorem. However, considering random variable $Z \ge 0$ such that $\mathbf{E}Z = \infty$ and $X_n(\omega) = n^{-1}Z(\omega) \downarrow 0$ as $n \to \infty$, we see that unlike monotonicity upwards, the monotonicity downwards does not really help much in convergence of expectations.

To practice your understanding, solve the following exercises.

EXERCISE 1.4.30. *Use Monotone Convergence to show that*

$$\mathbf{E}(\sum_{n=1}^{\infty} Y_n) = \sum_{n=1}^{\infty} \mathbf{E}Y_n\,,$$

for any sequence of non-negative R.V. Y_n. Deduce that if $X \ge 0$ and A_n are disjoint sets with $\mathbf{P}(\bigcup_n A_n) = 1$, then

$$\mathbf{E}(X) = \sum_{n=1}^{\infty} \mathbf{E}(X I_{A_n})\,.$$

Further, show that this applies also for any $X \in L^1$.

EXERCISE 1.4.31. *Prove Proposition 1.4.3, using the following four steps.*

(a) *Verify that the identity* (1.4.1) *holds for indicator functions $g(x) = I_B(x)$, $B \in \mathcal{B}$.*
(b) *Using linearity of the expectation, check that this identity holds whenever $g(x)$ is a (non-negative) simple function on* $(\mathbb{R}, \mathcal{B})$.
(c) *Combine the definition of the expectation via the identity* (1.2.2) *with Monotone Convergence to deduce that* (1.4.1) *is valid for any non-negative Borel function $g(x)$.*
(d) *Recall that $g(x) = g(x)_+ - g(x)_-$ for $g(x)_+ = \max(g(x), 0)$ and $g(x)_- = -\min(g(x), 0)$ non-negative Borel functions. Thus, using Definition 1.2.25 conclude that* (1.4.1) *holds whenever $\mathbf{E}|g(X)| < \infty$.*

REMARK. The four step procedure you just employed is a routine way for verifying identities about expectations. Not surprisingly it is called the *standard machine*.

As we next demonstrate, multiplication by an integrable positive random variable results with a *change of measure* that produces an *equivalent probability measure*.

EXERCISE 1.4.32 (Change of measure). *Suppose $Z \geq 0$ is a random variable on $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\mathbf{E}Z = 1$.*

(a) *Show that $\widetilde{\mathbf{P}} : \mathcal{F} \to \mathbb{R}$ given by $\widetilde{\mathbf{P}}(A) = \mathbf{E}[ZI_A]$ is a probability measure on $(\Omega, \mathcal{F})$.*

(b) *Denoting by $\widetilde{\mathbf{E}}[X]$ the expectation of a non-negative random variable X on $(\Omega, \mathcal{F})$ under the probability measure $\widetilde{\mathbf{P}}$, show that $\widetilde{\mathbf{E}}X = \mathbf{E}[XZ]$.* Hint: *Following the procedure outlined in Exercise 1.4.31, verify this identity first for $X = I_A$, then for non-negative $X \in \mathrm{SF}$, combining (1.2.2) and Monotone Convergence to extend it to all $X \geq 0$.*

(c) *Similarly, show that if $Z > 0$ then also $\mathbf{E}Y = \widetilde{\mathbf{E}}[Y/Z]$ for any random variable $Y \geq 0$ on $(\Omega, \mathcal{F})$. Deduce that in this case $\mathbf{P}$ and $\widetilde{\mathbf{P}}$ are* equivalent probability measures *on $(\Omega, \mathcal{F})$, that is $\mathbf{P}(A) = 0$ if and only if $\widetilde{\mathbf{P}}(A) = 0$.*

Here is an application of the preceding exercise, where by a suitable explicit change of measure you change the law of a random variable W without changing the function $W : \Omega \to \mathbb{R}$.

EXERCISE 1.4.33. *Suppose a R.V. W on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ has the $N(\mu, 1)$ law of Definition 1.2.30.*

(a) *Check that $Z = \exp(-\mu W + \mu^2/2)$ is a positive random variable with $\mathbf{E}Z = 1$.*

(b) *Show that under the corresponding equivalent probability measure $\widetilde{\mathbf{P}}$ of Exercise 1.4.32 the R.V. W has the $N(0, 1)$ law.*

1.4.3. Independence. We say that two events $A, B \in \mathcal{F}$ are $\mathbf{P}$-mutually *independent* if $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. For example, suppose two fair dice are thrown. The events $E_1 = \{\text{Sum of two is } 6\}$ and $E_2 = \{\text{first die is } 4\}$ are then not independent since

$$\mathbf{P}(E_1) = \mathbf{P}(\{(1,5)\ (2,4)\ (3,3)\ (4,2)\ (5,1)\}) = \frac{5}{36}, \quad \mathbf{P}(E_2) = \mathbf{P}(\{\omega : \omega_1 = 4\}) = \frac{1}{6}$$

and

$$\mathbf{P}(E_1 \cap E_2) = \mathbf{P}(\{(4,2)\}) = \frac{1}{36} \neq \mathbf{P}(E_1)\mathbf{P}(E_2).$$

However, one can check that E_2 and $E_3 = \{\text{sum of dice is } 7\}$ are independent.

In the context of independence we often do not explicitly mention the probability measure $\mathbf{P}$. However, events $A, B \in \mathcal{F}$ may well be independent with respect to one probability measure on $(\Omega, \mathcal{F})$ and not independent with respect to another.

EXERCISE 1.4.34. *Provide a measurable space $(\Omega, \mathcal{F})$, two probability measures $\mathbf{P}$ and $\mathbf{Q}$ on it and events A and B that are $\mathbf{P}$-mutually independent but not $\mathbf{Q}$-mutually independent.*

More generally we define the independence of events as follows.

DEFINITION 1.4.35. *Events $A_i \in \mathcal{F}$ are $\mathbf{P}$-mutually independent if for any $L < \infty$ and distinct indices $i_1, i_2, \ldots, i_L$,*

$$\mathbf{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_L}) = \prod_{k=1}^{L} \mathbf{P}(A_{i_k}).$$

In analogy with the mutual independence of events we define the independence of two R.V.-s and more generally, that of two σ-fields.

DEFINITION 1.4.36. *Two σ-fields $\mathcal{H}$, $\mathcal{G} \subseteq \mathcal{F}$ are $\mathbf{P}$-independent if*

$$\mathbf{P}(G \cap H) = \mathbf{P}(G)\mathbf{P}(H), \qquad \forall G \in \mathcal{G},\ \forall H \in \mathcal{H}$$

(see [**Bre92**, Definition 3.1] *for the independence of more than two σ-fields). The random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ are independent, if the corresponding σ-fields $\sigma(X_1, \ldots, X_n)$ and $\sigma(Y_1, \ldots, Y_m)$ are independent (see* [**Bre92**, Definition 3.2] *for the independence of $k > 2$ random vectors).*

To practice your understanding of this definition, solve the following exercise.

EXERCISE 1.4.37.

(a) *Verify that if σ-fields $\mathcal{G}$ and $\mathcal{H}$ are $\mathbf{P}$-independent, $\widetilde{\mathcal{G}} \subseteq \mathcal{G}$ and $\widetilde{\mathcal{H}} \subseteq \mathcal{H}$, then also $\widetilde{\mathcal{G}}$ and $\widetilde{\mathcal{H}}$ are $\mathbf{P}$-independent.*
(b) *Deduce that if random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ are $\mathbf{P}$-independent, then so are the R.V. $h(X_1, \ldots, X_n)$ and $g(Y_1, \ldots, Y_m)$ for any Borel functions $h : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}$.*

Beware that pairwise independence (of each pair A_k, A_j for $k \neq j$), does not imply mutual independence of *all* the events in question and the same applies to three or more random variables. Here is an illustrating example.

EXERCISE 1.4.38. *Consider the sample space $\Omega = \{0, 1, 2\}^2$ with probability measure on $(\Omega, 2^\Omega)$ that assigns equal probability (i.e. $1/9$) to each possible value of $\omega = (\omega_1, \omega_2) \in \Omega$. Then, $X(\omega) = \omega_1$ and $Y(\omega) = \omega_2$ are independent R.V. each taking the values $\{0, 1, 2\}$ with equal (i.e. $1/3$) probability. Define $Z_0 = X$, $Z_1 = (X + Y)\text{mod}3$ and $Z_2 = (X + 2Y)\text{mod}3$. Show that Z_0 is independent of Z_1, Z_0 is independent of Z_2, Z_1 is independent of Z_2, but if we know the value of Z_0 and Z_1, then we also know Z_2.*

See the next exercise for more in this direction.

EXERCISE 1.4.39. *Provide an example of three events A_1, A_2, A_3 that are not $\mathbf{P}$-mutually independent even though $\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i)\mathbf{P}(A_j)$ for any $i \neq j$ and an example of three events B_1, B_2, B_3 that are not $\mathbf{P}$-mutually independent even though $\mathbf{P}(B_1 \cap B_2 \cap B_3) = \mathbf{P}(B_1)\mathbf{P}(B_2)\mathbf{P}(B_3)$.*

We provide now an alternative criterion for independence of two random vectors which is often easier to verify than Definition 1.4.36, and which is to be expected in view of Theorem 1.2.14.

PROPOSITION 1.4.40. *For any finite $n, m \geq 1$, two random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ with values in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, are independent if and only if*

$$\mathbf{E}(h(X_1, \ldots, X_n)g(Y_1, \ldots, Y_m)) = \mathbf{E}(h(X_1, \ldots, X_n))\mathbf{E}(g(Y_1, \ldots, Y_m)), \tag{1.4.6}$$

for all bounded, Borel measurable functions $g : \mathbb{R}^m \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}$.

DEFINITION 1.4.41. *Square-integrable random variables X and Y defined on the same probability space are called* uncorrelated *if $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$.*

In particular, independent random variables are uncorrelated, but the converse is not necessarily true.

PROPOSITION 1.4.42. *Any two square-integrable independent random variables X and Y are also* uncorrelated.

EXERCISE 1.4.43. *Give an example of a pair of R.V. that are uncorrelated but not independent.*

The different definitions of independence we provided are consistent. For example, if the events $A, B \in \mathcal{F}$ are independent, then so are I_A and I_B. Indeed, we need to show that $\sigma(I_A) = \{\emptyset, \Omega, A, A^c\}$ and $\sigma(I_B) = \{\emptyset, \Omega, B, B^c\}$ are independent. Since $\mathbf{P}(\emptyset) = 0$ and $\emptyset$ is invariant under intersections, whereas $\mathbf{P}(\Omega) = 1$ and all events are invariant under intersection with Ω, it suffices to consider $G \in \{A, A^c\}$ and $H \in \{B, B^c\}$. We check independence first for $G = A$ and $H = B^c$. Noting that A is the union of the disjoint events $A \cap B$ and $A \cap B^c$ we have that

$$\mathbf{P}(A \cap B^c) = \mathbf{P}(A) - \mathbf{P}(A \cap B) = \mathbf{P}(A)[1 - \mathbf{P}(B)] = \mathbf{P}(A)\mathbf{P}(B^c)\,,$$

where the middle equality is due to the assumed independence of A and B. The proof for all other choices of G and H is very similar.

In this context verify that for any two measurable subsets A and B of Ω, if the indicator random variables I_A and I_B are uncorrelated, then they must also be independent of each other (in contrast with the more general case, see Exercise 1.4.43). The key to this is the fact that $\mathbf{E} I_A I_B = \mathbf{P}(A \cap B)$.

Independence helps when computing expectations. For example:

EXERCISE 1.4.44. *Let $\xi_1, \xi_2, \xi_3, \ldots$ be a sequence of square-integrable, non-negative, independent R.V.-s and $N(\omega) \in \{0, 1, 2, \ldots\}$ a R.V. on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that N is independent of ξ_m and $\mathbf{E}(\xi_m) = \mathbf{E}(\xi_1)$ for any m. Show that the non-negative R.V.*

$$X(\omega) = \sum_{i=1}^{N(\omega)} \xi_i(\omega) = \sum_{i=1}^{\infty} \xi_i(\omega) I_{N(\omega) \geq i}$$

Has the expectation

$$\mathbf{E}(X) = \sum_{i=1}^{\infty} \mathbf{E}(\xi_i)\mathbf{P}(N \geq i) = \mathbf{E}(\xi_1)\mathbf{E}(N)\,.$$

CHAPTER 2

Conditional expectation and Hilbert spaces

The most important concept in probability theory is the conditional expectation to which this chapter is devoted. In contrast with the elementary definition often used for a finite or countable sample space, the conditional expectation, as defined in Section 2.1, is itself a random variable. A brief introduction to the theory of Hilbert spaces is provided in Section 2.2. This theory is used here to show the existence of the conditional expectation. It is also applied in Section 5.1 to construct the Brownian motion, one of the main stochastic processes of interest to us. Section 2.3 details the important properties of the conditional expectation. Finally, in Section 2.4 we represent the conditional expectation as the expectation with respect to the *random* regular conditional probability distribution.

2.1. Conditional expectation: existence and uniqueness

After reviewing in Subsection 2.1.1 the elementary definition of the conditional expectation for discrete random variables we provide a definition that applies to any square-integrable R.V. We then show in Subsection 2.1.2 how to extend it to the general case of integrable R.V. and to the conditioning on any σ-field.

2.1.1. Conditional expectation in the discrete and L^2 cases. Suppose the random variables X and Y take on finitely many values. Then, we know that $\mathbf{E}(X|Y=y) = \sum_x x\mathbf{P}(X=x|Y=y)$, provided $\mathbf{P}(Y=y)>0$. Moreover, by the formula $\mathbf{P}(A|B) = \frac{\mathbf{P}(A\cap B)}{\mathbf{P}(B)}$, this amounts to

$$\mathbf{E}(X|Y=y) = \sum_x x\frac{\mathbf{P}(X=x, Y=y)}{\mathbf{P}(Y=y)}$$

With $f(y)$ denoting the function $\mathbf{E}[X|Y=y]$, we may now consider the random variable $\mathbf{E}(X|Y) = f(Y)$, which we call the *Conditional Expectation* (in short C.E.) of X given Y. For a general R.V. X we similarly define

$$f(y) := \mathbf{E}(X|Y=y) = \frac{\mathbf{E}(XI_{\{Y=y\}})}{\mathbf{P}(Y=y)}$$

for any y such that $\mathbf{P}(Y=y)>0$ and let $\mathbf{E}(X|Y) = f(Y)$ (see [**GS01**, Definition 3.7.3] for a similar treatment).

EXAMPLE 2.1.1. *For example, if R.V. $X=\omega_1$ and $Y=\omega_2$ for the probability space $\mathcal{F}=2^\Omega$, $\Omega=\{1,2\}^2$ and*

$$\mathbf{P}(1,1)=.5,\ \ \mathbf{P}(1,2)=.1,\ \ \mathbf{P}(2,1)=.1,\ \ \mathbf{P}(2,2)=.3,$$

then,

$$\mathbf{P}(X=1|Y=1) = \frac{\mathbf{P}(X=1, Y=1)}{\mathbf{P}(Y=1)} = \frac{5}{6}$$

implying that $\mathbf{P}(X=2|Y=1)=\frac{1}{6}$, *and*

$$\mathbf{E}(X|Y=1)=1\cdot\frac{5}{6}+2\cdot\frac{1}{6}=\frac{7}{6}$$

Likewise, check that $\mathbf{E}(X|Y=2)=\frac{7}{4}$, *hence* $\mathbf{E}(X|Y)=\frac{7}{6}I_{Y=1}+\frac{7}{4}I_{Y=2}$.

This approach works also for Y that takes countably many values. However, it is limited to this case, as it requires us to have $\mathbf{E}(I_{\{Y=y\}})=\mathbf{P}(Y=y)>0$. To bypass this difficulty, observe that $Z=\mathbf{E}(X|Y)=f(Y)$ as defined before, necessarily satisfies the identity

$$0=\mathbf{E}[(X-f(y))I_{\{Y=y\}}]=\mathbf{E}[(X-Z)I_{\{Y=y\}}]$$

for any $y\in\mathbb{R}$. Further, here also $\mathbf{E}[(X-Z)I_{\{Y\in B\}}]=0$ for any Borel set B, that is $\mathbf{E}[(X-Z)I_A]=0$ for any $A\in\sigma(Y)$. As we see in the sequel, demanding that the last identity holds for $Z=f(Y)$ provides us with a definition of $Z=\mathbf{E}(X|Y)$ that applies for any two R.V. X and Y on the same probability space $(\Omega,\mathcal{F},\mathbf{P})$, subject only to the mild condition that $\mathbf{E}(|X|)<\infty$ (For more details and proofs, see [**GS01**, Section 7.9]).

To gain better insight into this, more abstract definition, consider first $X\in L^2(\Omega,\mathcal{F},\mathbf{P})$ and the optimization problem

$$\mathbf{d}^2=\inf\{\mathbf{E}[(X-W)^2]:W\in\mathcal{H}_Y\} \tag{2.1.1}$$

for $\mathcal{H}_Y=L^2(\Omega,\sigma(Y),\mathbf{P})$. From the next proposition we see that the conditional expectation is the solution of this optimization problem.

PROPOSITION 2.1.2. *There exists a unique (a.s.) optimal* $Z\in\mathcal{H}_Y$ *such that* $\mathbf{d}^2=\mathbf{E}[(X-Z)^2]$. *Further, the optimality of* Z *is equivalent to the orthogonality property*

$$\mathbf{E}[(X-Z)V]=0,\qquad\forall\,V\in\mathcal{H}_Y. \tag{2.1.2}$$

PROOF. We state without proof the existence of $Z\in\mathcal{H}_Y$ having the orthogonality property (2.1.2), as we shall later see that this is a consequence of the more general theory of Hilbert spaces.

Suppose that both $Z_1,Z_2\in\mathcal{H}_Y$ are such that $\mathbf{E}[(X-Z_1)V]=0$ and $\mathbf{E}[(X-Z_2)V]=0$ for all $V\in\mathcal{H}_Y$. Taking the difference of the two equations, we see that $\mathbf{E}[(Z_1-Z_2)V]=0$ for any such V. Take $V=Z_1-Z_2\in\mathcal{H}_Y$ to get $\mathbf{E}[(Z_1-Z_2)^2]=0$, so $Z_1\overset{a.s.}{=}Z_2$.

Suppose now that $W,Z\in\mathcal{H}_Y$ and also that $\mathbf{E}[(X-Z)V]=0$ for any $V\in\mathcal{H}_Y$. It is easy to check the identity

$$\frac{1}{2}\{[(X-W)^2]-[(X-Z)^2]\}=(X-Z)(Z-W)+\frac{1}{2}(Z-W)^2\,, \tag{2.1.3}$$

which holds for any X,W,Z. Considering $V=Z-W\in\mathcal{H}_Y$, and taking expectation, we get by linearity of the expectation that

$$\mathbf{E}[(X-W)^2]-\mathbf{E}[(X-Z)^2]=\mathbf{E}(V^2)\geq 0,$$

with equality if only if $V=0$, that is, $W=Z$ a.s. Thus, $Z\in\mathcal{H}_Y$ that satisfies (2.1.2) is the a.s. unique minimizer of $\mathbf{E}[(X-W)^2]$ among all $W\in\mathcal{H}_Y$.

We finish the proof by demonstrating the converse. Namely, suppose $Z\in\mathcal{H}_Y$ is such that $\mathbf{E}[(X-W)^2]\geq\mathbf{E}[(X-Z)^2]$ for all $W\in\mathcal{H}_Y$. Taking the expectation in

(2.1.3), we thus have that

$$\mathbf{E}[(X-Z)(Z-W)] + \frac{1}{2}\mathbf{E}[(Z-W)^2] \geq 0\,.$$

Fixing a non-random $y \in \mathbb{R}$ and a R.V. $V \in \mathcal{H}_Y$, we apply this inequality for $W = Z - yV \in \mathcal{H}_Y$ to get $g(y) = y\mathbf{E}[(X-Z)V] + \frac{1}{2}y^2\mathbf{E}[V^2] \geq 0$. Since $\mathbf{E}[V^2] < \infty$, the requirement $g(y) \geq 0$ for *all* $y \in \mathbb{R}$ implies that $g'(0) = \mathbf{E}[(X-Z)V] = 0$. As $V \in \mathcal{H}_Y$ is arbitrary, we have just verified the stated orthogonality property (2.1.2). ∎

DEFINITION 2.1.3. *For $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ the* conditional expectation $Z = \mathbf{E}(X|Y)$ *is the unique R.V. in $\mathcal{H}_Y$ satisfying (2.1.2). By Proposition 2.1.2 this is also the minimizer of $\mathbf{E}[(X-W)^2]$ among all $W \in \mathcal{H}_Y$.*

To get some intuition into Definition 2.1.3, consider the special case of $Y \in \mathrm{SF}$ and suppose that $Y = \sum_{i=1}^n y_i I_{A_i}$ for some $n < \infty$, distinct y_i and disjoint sets $A_i = \{\omega : Y(\omega) = y_i\}$ such that $\mathbf{P}(A_i) > 0$ for all i.

Recall Theorem 1.2.14 that any $W \in \mathcal{H}_Y$ is of the form $W = g(Y)$ for some Borel function g such that $\mathbf{E}(g(Y)^2) < \infty$. As $Y(\omega)$ takes only finitely many possible values $\{y_1, \ldots, y_n\}$, every function $g(\cdot)$ on this set is bounded (hence square-integrable) and measurable. Further, the sets A_i are disjoint, resulting with $g(Y) = \sum_{i=1}^n g(y_i) I_{A_i}$. Hence, in this case, $\mathcal{H}_Y = \{\sum_{i=1}^n v_i I_{A_i} : v_i \in \mathbb{R}\}$ and the optimization problem (2.1.1) is equivalent to

$$\mathbf{d}^2 = \inf_{v_i} \mathbf{E}[(X - \sum_{i=1}^n v_i I_{A_i})^2] = \mathbf{E}(X^2) + \inf_{v_i} \Big\{ \sum_{i=1}^n \mathbf{P}(A_i) v_i^2 - 2 \sum_{i=1}^n v_i \mathbf{E}[X I_{A_i}] \Big\}\,,$$

the solution of which is obtained for $v_i = \mathbf{E}[X I_{A_i}]/\mathbf{P}(A_i)$. We conclude that

$$\mathbf{E}(X|Y) = \sum_i \frac{\mathbf{E}[X I_{Y=y_i}]}{\mathbf{P}(Y = y_i)} I_{Y=y_i}\,, \tag{2.1.4}$$

in agreement with our first definition for the conditional expectation in the case of discrete random variables.

EXERCISE 2.1.4. *Let $\Omega = \{a, b, c, d\}$, with event space $\mathcal{F} = 2^\Omega$ and probability measure such that $\mathbf{P}(\{a\}) = 1/2$, $\mathbf{P}(\{b\}) = 1/4$, $\mathbf{P}(\{c\}) = 1/6$ and $\mathbf{P}(\{d\}) = 1/12$.*

(a) *Find $\sigma(I_A)$, $\sigma(I_B)$ and $\sigma(I_A, I_B)$ for the subsets $A = \{a, d\}$ and $B = \{b, c, d\}$ of Ω.*

(b) *Let $\mathcal{H} = L^2(\Omega, \sigma(I_B), \mathbf{P})$. Find the conditional expectation $\mathbf{E}(I_A|I_B)$ and the value of $\mathbf{d}^2 = \inf\{\mathbf{E}[(I_A - W)^2] : W \in \mathcal{H}\}$.*

2.1.2. Conditional expectation: the general case. You should verify that the conditional expectation of Definition 2.1.3 is a R.V. on the measurable space $(\Omega, \mathcal{F}_Y)$ whose dependence on Y is only via $\mathcal{F}_Y$. This is consistent with our interpretation of $\mathcal{F}_Y$ as conveying the information content of the R.V. Y. In particular, $\mathbf{E}[X|Y] = \mathbf{E}[X|Y']$ whenever $\mathcal{F}_{Y'} = \mathcal{F}_Y$ (for example, if $Y' = h(Y)$ for some invertible Borel function h). By this reasoning we may and shall often use the notation $\mathbf{E}\,(X \mid \mathcal{F}_Y)$ for $\mathbf{E}(X|Y)$, and next define $\mathbf{E}(X|\mathcal{G})$ for $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ and arbitrary σ-field $\mathcal{G} \subseteq \mathcal{F}$. To this end, recall that Definition 2.1.3 is based on the orthogonality property (2.1.2). As we see next, in the general case where X is only in L^1, we have to consider the smaller class of almost surely bounded R.V. V on $(\Omega, \mathcal{G})$, in order to guarantee that the expectation of XV is well defined.

DEFINITION 2.1.5. *The* conditional expectation *of* $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ *given a* σ*-field* $\mathcal{G} \subseteq \mathcal{F}$*, is the R.V.* Z *on* $(\Omega, \mathcal{G})$ *such that*

$$\mathbf{E}\,[(X - Z)\; I_A] = 0, \qquad \forall\; A \in \mathcal{G} \tag{2.1.5}$$

and $\mathbf{E}(X|Y)$ *corresponds to the special case of* $\mathcal{G} = \mathcal{F}_Y$ *in (2.1.5).*

As we state next, in the special case of $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ this definition coincides with the C.E. we get by Definition 2.1.3.

THEOREM 2.1.6. *The C.E. of integrable R.V.* X *given any* σ*-field* $\mathcal{G}$ *exists and is a.s. unique. That is, there exists* Z *measurable on* $\mathcal{G}$ *that satisfies (2.1.5), and if* Z_1 *and* Z_2 *are both measurable on* $\mathcal{G}$ *satisfying (2.1.5), then* $Z_1 \overset{a.s}{=} Z_2$*.*
Further, if in addition $\mathbf{E}(X^2) < \infty$*, then such* Z *also satisfies*

$$\mathbf{E}\,[(X - Z)\; V] = 0, \qquad \forall\; V \in L^2(\Omega, \mathcal{G}, \mathbf{P}),$$

hence for $\mathcal{G} = \mathcal{F}_Y$ *the R.V.* Z *coincides with that of Definition 2.1.3.*

PROOF. (omit at first reading) We only outline the key ideas of the proof (for a detailed treatment see [**GS01**, Theorem 7.9.26]).

Step 1. Given $X \geq 0$, $\mathbf{E}X < \infty$, we define $X_n(\omega) = \min\{X(\omega), n\}$. Then, $X_n \uparrow X$ and X_n is bounded for each n. In particular, $X_n \in L^2$ so the C.E. $\mathbf{E}[X_n|\mathcal{G}]$ exists by orthogonal projection in Hilbert spaces (see Theorem 2.2.12, in analogy with Proposition 2.1.2). Further, the R.V. $\mathbf{E}[X_n|\mathcal{G}]$ are non-decreasing in n, so we define the C.E. of X given $\mathcal{G}$ as

$$\mathbf{E}[X|\mathcal{G}] := \lim_{n\to\infty} \mathbf{E}[X_n|\mathcal{G}]. \tag{2.1.6}$$

One then checks that

(1). The limit in (2.1.6) exists and is almost surely finite;

(2). This limit does not depend on the specific sequence X_n we have chosen in the sense that if $Y_n \uparrow X$ and $\mathbf{E}({Y_n}^2) < \infty$ then $\lim_{n\to\infty} \mathbf{E}[Y_n|\mathcal{G}] = \mathbf{E}[X|\mathcal{G}]$;

(3). The R.V. obtained by (2.1.6) agrees with Definition 2.1.5, namely, it satisfies the orthogonality property (2.1.5).

Step 2. If $\mathbf{E}|X| < \infty$ then $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X_+|\mathcal{G}] - \mathbf{E}[X_-|\mathcal{G}]$. ∎

We often use the notation $\mathbf{P}(A|\mathcal{G})$ for $\mathbf{E}(I_A|\mathcal{G})$ and any $A \in \mathcal{F}$. The reason for this is made clear in Section 2.4 when we relate the C.E. with the notion of (regular) conditional probability.

We detail in Section 2.3 few of the many properties of the C.E. that are easy to verify directly from Definition 2.1.5. However, some properties are much easier to see when considering Definition 2.1.3 and the proof of Theorem 2.1.6. For example,

PROPOSITION 2.1.7. *If* X *is a non-negative R.V., then a.s.* $\mathbf{E}(X|\mathcal{G}) \geq 0$.

PROOF. Suppose first that $X \in L^2$ is non-negative and $Z \in L^2(\Omega, \mathcal{G}, \mathbf{P})$. Note that $Z_+ = \max(Z, 0) \in L^2(\Omega, \mathcal{G}, \mathbf{P})$ as well. Moreover, in this case $(X - Z_+)^2 \leq (X - Z)^2$ a.s. hence also $\mathbf{E}[(X - Z_+)^2] \leq \mathbf{E}[(X - Z)^2]$. Consequently, if $Z = \mathbf{E}(X|\mathcal{G})$, then by Definition 2.1.3 $Z = Z_+$ a.s., that is, $Z \geq 0$ a.s. Following the proof of Theorem 2.1.6 we see that if $X \in L^1$ is non-negative then so are $X_n \in L^2$. We have already seen that $\mathbf{E}[X_n|\mathcal{G}]$ are non-negative, and so the same applies for their limit $\mathbf{E}[X|\mathcal{G}]$. ∎

As you show next, identities such as (2.1.4) apply for conditioning on any σ-field generated by a countable partition.

EXERCISE 2.1.8. *Let* $\mathcal{G} = \sigma(\{B_i\})$ *for a countable,* $\mathcal{F}$*-measurable partition* $B_1, B_2, \ldots$ *of* Ω *into sets of positive probability. That is,* $B_i \in \mathcal{F}$, $\cup_i B_i = \Omega$, $B_i \cap B_j = \emptyset, i \neq j$ *and* $\mathbf{P}(B_i) > 0$.

(a) *Show that* $G \in \mathcal{G}$ *if and only if* G *is a union of sets from* $\{B_i\}$.
(b) *Suppose* Y *is a random variable on* $(\Omega, \mathcal{G})$. *Show that* $Y(\omega) = \sum_i c_i I_{B_i}(\omega)$ *for some non-random* $c_i \in \mathbb{R}$.
(c) *Let* X *be an integrable random variable on* $(\Omega, \mathcal{F}, \mathbf{P})$. *Show that* $\mathbf{E}(X|\mathcal{G}) = \sum_i c_i I_{B_i}$ *with* $c_i = \mathbf{E}(X I_{B_i})/\mathbf{P}(B_i)$.
(d) *Deduce that for an integrable random variable* X *and* $B \in \mathcal{F}$ *such that* $0 < \mathbf{P}(B) < 1$,

$$\mathbf{E}(X|\sigma(B)) = \frac{\mathbf{E}(I_B X)}{\mathbf{P}(B)} I_B + \frac{\mathbf{E}(I_{B^c} X)}{\mathbf{P}(B^c)} I_{B^c} ,$$

and in particular $\mathbf{E}(I_A|I_B) = \mathbf{P}(A|B) I_B + \mathbf{P}(A|B^c) I_{B^c}$ *for any* $A \in \mathcal{F}$ *(where* $\mathbf{P}(A|B) = \mathbf{P}(A \cap B)/\mathbf{P}(B)$*).*

2.2. Hilbert spaces

The existence of the C.E. when $X \in L^2$ is a consequence of the more general theory of Hilbert spaces. We embark in this section on an exposition of this general theory, with most definitions given in Subsection 2.2.1 and their consequences outlined in Subsection 2.2.2. The choice of material is such that we state everything we need either for the existence of the C.E. or for constructing the Brownian motion in Section 5.1.

2.2.1. Definition of Hilbert spaces and subspaces. We start with the definition of a linear vector space.

DEFINITION 2.2.1. *A* linear vector space $\mathbb{H} = \{h\}$ *is a collection of vectors* h*, for which addition and scalar multiplication are well-defined:*
$h_1 + h_2 \in \mathbb{H}$ *- the addition of* $h_1 \in \mathbb{H}$, $h_2 \in \mathbb{H}$;
$\alpha h \in \mathbb{H}$ *- scalar multiplication* $\forall \alpha \in \mathbb{R}$, $h \in \mathbb{H}$;
and having the properties:
(i) $\alpha(h_1 + h_2) = \alpha h_1 + \alpha h_2$;
(ii) $(\alpha + \beta)h = \alpha h + \beta h$;
(iii) $\alpha(\beta h) = (\alpha\beta)h$;
(iv) $1h = h$.

The following examples are used throughout for illustration.

EXAMPLE 2.2.2. *The space* $\mathbb{R}^3$ *is a linear vector space with addition and scalar multiplication*

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{pmatrix}, \quad c \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} cx_1 \\ cy_1 \\ cz_1 \end{pmatrix}$$

Other examples are the spaces $L^q(\Omega, \mathcal{F}, \mathbf{P})$ *with the usual addition of R.V. and scalar multiplication of a R.V.* $X(\omega)$ *by a non-random constant.*

The key concept in the definition of a Hilbert space is the inner product which is defined next.

Definition 2.2.3. *An* inner product *in a linear vector space* $\mathbb{H}$ *is a real-valued function on pairs* $h_1, h_2 \in \mathbb{H}$, *denoted* (h_1, h_2), *having the following properties:*
1) $(h_1, h_2) = (h_2, h_1)$;
2) $(h_1, h_2 + \alpha h_3) = (h_1, h_2) + \alpha(h_1, h_3)$;
3) $(h, h) = 0$ *if and only if* $h = 0$, *otherwise* $(h, h) > 0$.

The canonical inner product on $\mathbb{R}^3$ is the *dot product* $(v_1, v_2) = x_1x_2 + y_1y_2 + z_1z_2$ and the canonical inner product on $L^2(\Omega, \mathcal{F}, \mathbf{P})$ is $(X, Y) = \mathbf{E}(XY)$.
The following inequality holds for any inner product (for the special case of most interest to us, see Proposition 1.2.41).

Proposition 2.2.4 (Schwarz inequality). *Let* $\|h\| = (h, h)^{\frac{1}{2}}$. *Then, for any* $h_1, h_2 \in \mathbb{H}$,

$$(h_1, h_2)^2 \leq \|h_1\|^2 \|h_2\|^2 .$$

Exercise 2.2.5. (a) *Note that* $(h_1 + \lambda h_2, h_1 + \lambda h_2) \geq 0$ *for all* $\lambda \in \mathbb{R}$. *By the bi-linearity of the inner product this is a quadratic function of* λ. *To prove Schwarz's inequality, find its coefficients and consider value of* λ *which minimizes the quadratic expression.*
(b) *Use Schwarz inequality to derive the triangle inequality* $\|h_1 + h_2\| \leq \|h_1\| + \|h_2\|$.
(c) *Using the bi-linearity of the inner product, check that the* parallelogram *law* $\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2$ *holds.*

Thus, each inner product induces a norm on the underlying linear vector space $\mathbb{H}$.

Definition 2.2.6. *The norm of* h *is* $\|h\| = (h, h)^{\frac{1}{2}}$. *In particular, the* triangle inequality $\|h_1 + h_2\| \leq \|h_1\| + \|h_2\|$ *is an immediate consequence of Schwarz inequality.*

In the space $\mathbb{R}^3$ we get the Euclidean norm $\|v\| = \sqrt{x^2 + y^2 + z^2}$, whereas for $\mathbb{H} = L^2(\Omega, \mathcal{F}, \mathbf{P})$ we have that $\|X\| = (\mathbf{E}X^2)^{\frac{1}{2}} = \|X\|_2$, and $\|X\| = 0$ if and only if $X \stackrel{a.s.}{=} 0$.
Equipping the space $\mathbb{H}$ with the norm $\|\cdot\|$ we have the corresponding notions of Cauchy and convergent sequences.

Definition 2.2.7. *We say that* $\{h_n\}$ *is a* Cauchy-sequence *in* $\mathbb{H}$ *if for any* $\varepsilon > 0$, *there exists* $N(\varepsilon) < \infty$ *such that* $\sup_{n,m \geq N} \|h_n - h_m\| \leq \varepsilon$. *We say that* $\{h_n\}$ *converges to* $h \in \mathbb{H}$ *if* $\|h_n - h\| \to 0$ *as* $n \to \infty$.

For example, in case of $\mathbb{H} = L^2(\Omega, \mathcal{F}, \mathbf{P})$, the concept of Cauchy-sequences is the same as what we had already seen in Section 1.3.2.
We are now ready to define the key concept of Hilbert space.

Definition 2.2.8. *A* Hilbert space *is a linear vector space* $\mathbb{H}$ *with inner product* (h, h), *which is complete with respect to the corresponding norm* $\|h\|$ *(that is, every Cauchy sequence in* $\mathbb{H}$ *converges to a point* h *in* $\mathbb{H}$*).*

By Proposition 1.3.20 we know that L^2 is a Hilbert space, since every Cauchy sequence in L^2 converges in quadratic mean to a point in L^2. To check your understanding you may wish to solve the next exercise at this point.

EXERCISE 2.2.9. *Check that the space $\mathbb{R}^3$ of three dimensional real-valued vectors, together with the inner product $(v,u) = v_1u_1 + v_2u_2 + v_3u_3$ is a Hilbert space.*

To define the conditional expectation we need the following additional structure.

DEFINITION 2.2.10. *A* Hilbert sub-space $\mathbb{K}$ *of* $\mathbb{H}$ *is a subset* $\mathbb{K} \subseteq \mathbb{H}$*, which when equipped with the same inner product is a Hilbert space. Specifically,* $\mathbb{K}$ *is a linear subspace (closed under addition, scalar multiplication), such that every Cauchy-sequence* $\{h_n\} \subseteq \mathbb{K}$ *has a limit* $h \in \mathbb{K}$*.*

EXAMPLE 2.2.11. *Note that in our setting $\mathcal{H}_Y = L^2(\Omega, \mathcal{F}_Y, \mathbf{P}) \subseteq L^2(\Omega, \mathcal{F}, \mathbf{P})$ is a Hilbert subspace. Indeed, if $Z_n \in \mathcal{H}_Y$ is Cauchy, then $Z_n = f_n(Y) \to Z \in \mathcal{H}_Y$ as well. More generally, the Hilbert subspace of $L^2(\Omega, \mathcal{F}, \mathbf{P})$ spanned by R.V.-s $\{X_1, X_2, \cdots, X_n\}$ is the set of $h = f(X_1, X_2, \cdots, X_n)$ with $\mathbf{E}\,(h^2) < \infty$.*

2.2.2. Orthogonal projection, separability and Fourier series. The existence of the C.E. is a special instance of the following important geometric theorem about Hilbert spaces.

THEOREM 2.2.12 (Orthogonal Projection). *Let $\mathbb{G} \subseteq \mathbb{H}$ be a given Hilbert subspace and $h \in \mathbb{H}$ given. Let $\mathbf{d} = \inf\{\|h-g\| : g \in \mathbb{G}\}$, then,*
(a). There exists $\widehat{h} \in \mathbb{G}$ such that $d = \|h - \widehat{h}\|$.
(b). Any such $\widehat{h}$ satisfies $(h - \widehat{h},\ g) = 0$ for all $g \in \mathbb{G}$.
(c). The vector $\widehat{h}$ that satisfies (b) (or (a)) is unique.
We call $\widehat{h}$ as above the orthogonal projection of h on $\mathbb{G}$.

Our next exercise outlines the hardest part of the proof of the orthogonal projection theorem.

EXERCISE 2.2.13. *Prove part (a) of Theorem 2.2.12 as follows: Consider $g_n \in \mathbb{G}$ for which $\|h - g_n\| \to \mathbf{d}$. By definition of $\mathbf{d}$ and the parallelogram law for $u = h - g_m$ and $v = g_n - h$, deduce that $\|g_n - g_m\| \to 0$ when both $n, m \to \infty$, hence (why?) g_n has a limit $\hat{h} \in \mathbb{G}$, such that $\|h - \hat{h}\| = \mathbf{d}$ (why?).*

In particular, Definition 2.1.3 amounts to saying that the conditional expectation of X given Y is the orthogonal projection of X in $\mathbb{H} = L^2(\Omega, \mathcal{F}, \mathbf{P})$ on $\mathbb{G} = L^2(\Omega, \mathcal{F}_Y, \mathbf{P})$.
Another example of orthogonal projection is the projection $(x_1, \dots, x_k, 0, \dots, 0)$ of $x \in \mathbb{R}^d$ on the hyper-space of lower dimension $\mathbb{G} = \{(y_1, \dots, y_k, 0, \dots, 0) : y_i \in \mathbb{R}\}$, some $k < d$.
An important ingredient of Hilbert space theory is the concept of linear functional, defined next.

DEFINITION 2.2.14. *A functional $f : \mathbb{H} \mapsto \mathbb{R}$ is* linear *if $f(h_1 + \alpha h_2) = f(h_1) + \alpha f(h_2)$. A linear functional $f(\cdot)$ is called* bounded *if $|f(h)| \le C\|h\|$ for some $C < \infty$ and all $h \in \mathbb{H}$.*

For example, each real function $f : \mathbb{R} \mapsto \mathbb{R}$ is also a functional in the Hilbert space $\mathbb{R}$. The functional $g(x) = 2x$ is then linear and bounded though $g(x)$ is not a *bounded function*. More generally,

EXAMPLE 2.2.15. *Fixing $h_0 \in \mathbb{H}$ a Hilbert space, let $f(h) = (h_0, h)$. This is a bounded linear functional on $\mathbb{H}$ since $\|h_0\| < \infty$ and $|f(h)| \le \|h_0\|\|h\|$ by Schwarz inequality.*

The next theorem (which we do not prove), states that any bounded linear functional is of the form of Example 2.2.15. It is often very handy when studying linear functionals that are given in a rather implicit form.

THEOREM 2.2.16 (Riesz representation). *If f is a bounded linear functional on a Hilbert space $\mathbb{H}$, then there exists a unique $h_0 \in \mathbb{H}$ such that $f(h) = (h_0, h)$ for all $h \in \mathbb{H}$.*

We next define the subclass of separable Hilbert spaces, which have particularly appealing properties. We use these later in Section 5.1.

DEFINITION 2.2.17. *Hilbert space $\mathbb{H}$ is* separable *if it has a* complete, countable, orthonormal basis *$\{h_m,\ m = 1, \cdots\}$, such that*

(a). $(h_m, h_l) = \begin{cases} 1, & m = l \\ 0, & otherwise \end{cases}$ (orthonormal).

(b). If $(h, h_m) = 0$ for all m, then $h = 0$ (complete).

The classical concept of Fourier series has an analog in the theory of separable Hilbert spaces.

DEFINITION 2.2.18. *We say that $h \in \mathbb{H}$ has a* Fourier series *$\sum_i (h_i, h)h_i$ if*

$$\|h - \sum_{i=1}^{N} (h_i, h)h_i\| \to 0, \qquad \text{as} \qquad N \to \infty.$$

PROPOSITION 2.2.19. *The completeness of a basis $\{h_m\}$ is equivalent to the existence of Fourier series (in $\{h_m\}$) for any $h \in \mathbb{H}$.*

A key ingredient in the theory of separable Hilbert spaces is the following approximation theorem.

THEOREM 2.2.20 (Parseval). *For any complete orthonormal basis $\{h_m\}$ of a separable Hilbert space H and any $f, g \in \mathbb{H}$,*

$$(f, g) = \lim_{N\to\infty} \sum_{i=1}^{N} (f, h_i)(g, h_i)\,.$$

In particular, $\|h\|^2 = \sum_{i=1}^{\infty} (h, h_i)^2$, for all $h \in \mathbb{H}$.

PROOF. Since both f and g have a Fourier series, it follows by Schwarz inequality that $a_N = (f - \sum_{i=1}^{N}(f, h_i)h_i, g - \sum_{i=1}^{N}(g, h_i)h_i) \to 0$ as $N \to \infty$. Since the inner product is bi-linear and $(h_i, h_j) = 1_{i=j}$ it follows that $a_N = (f, g) - \sum_{i=1}^{N}(f, h_i)(g, h_i)$, yielding the first statement of the theorem. The second statement corresponds to the special case of $f = g$. ∎

We conclude with a concrete example that relates the abstract theory of separable Hilbert spaces to the familiar classical Fourier series.

EXAMPLE 2.2.21. *Let $L^2((0,1), \mathcal{B}, U) = \{f : (0,1) \mapsto \mathbb{R}$ such that $\int_0^1 f^2(t)\, dt < \infty\}$ equipped with the inner product $(h, g) = \int_0^1 h(t)g(t)dt$. This is a separable Hilbert space. Indeed, the classical Fourier series of any such f is*

$$f(t) = \sum_{n=0}^{\infty} c_n \cos(2\pi nt) + \sum_{n=1}^{\infty} s_n \sin(2\pi nt),$$

where $c_n = 2\int_0^1 f(t)\cos(2\pi nt)dt$ and $s_n = 2\int_0^1 f(t)\sin(2\pi nt)dt$.

2.3. Properties of the conditional expectation

We start with two extreme examples for which we can easily compute the C.E.

EXAMPLE 2.3.1. *Suppose $\mathcal{G}$ and $\sigma(X)$ are independent (i.e. for all $A \in \mathcal{G}$ and $B \in \sigma(X)$ we have that $\mathbf{P}(A \cap B) = \mathbf{P}(A)\ \mathbf{P}(B)$). Then, $\mathbf{E}(XI_A) = \mathbf{E}(X)\ \mathbf{E}(I_A)$ for all $A \in \mathcal{G}$. So, the constant $Z = \mathbf{E}(X)$ satisfies (2.1.5), that is, in this case $\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X)$. In particular, if Y and V are independent random variables and f a Borel function such that $X = f(V)$ is integrable, then $\mathbf{E}(f(V)|Y) = \mathbf{E}(f(V))$.*

EXAMPLE 2.3.2. *Suppose $X \in L^1(\Omega, \mathcal{G}, \mathbf{P})$. Obviously, $Z = X$ satisfies (2.1.5) and by our assumption X is measurable on $\mathcal{G}$. Consequently, here $\mathbf{E}(X|\mathcal{G}) = X$ a.s. In particular, if $X = c$ a.s. (a constant R.V.), then we have that $\mathbf{E}(X|\mathcal{G}) = c$ a.s. (and conversely, if $\mathbf{E}(XI_G) = c\mathbf{P}(G)$ for all $G \in \mathcal{G}$ then $X = c$ a.s.).*

REMARK. Note that X may have the same law as Y while $\mathbf{E}(X|\mathcal{G})$ does not have the same law as $\mathbf{E}(Y|\mathcal{G})$. For example, take $\mathcal{G} = \sigma(X)$ with X and Y square integrable, independent, of the same distribution and positive variance. Then, $\mathbf{E}(X|\mathcal{G}) = X$ per Example 2.3.2 and the non-random constant $\mathbf{E}(Y|\mathcal{G}) = \mathbf{E}(Y)$ (from Example 2.3.1) have different laws.

You can solve the next exercise either directly, or by showing that $\mathcal{F}_0$ is independent of $\sigma(X)$ for any random variable X and then citing Example 2.3.1.

EXERCISE 2.3.3. *Let $\mathcal{F}_0 = \{\Omega, \emptyset\}$. Show that if $Z \in L^1(\Omega, \mathcal{F}_0, \mathbf{P})$ then Z is necessarily a non-random constant and deduce that $\mathbf{E}(X|\mathcal{F}_0) = \mathbf{E}X$ for any $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$.*

A yet another example consists of $X = Y + V$ with Y and V two independent R.V-s in $L^2(\Omega, \mathcal{F}, \mathbf{P})$. Let $\mu = \mathbf{E}(V)$. Taking a generic $W \in \mathcal{H}_Y$ we know that $W = g(Y)$ and $\mathbf{E}[g^2(Y)] < \infty$. By Definition 1.4.36 and Proposition 1.2.16 we know that the square-integrable random variables $Y + \mu - g(Y)$ and $V - \mu$ are independent, hence uncorrelated (see Proposition 1.4.42). Consequently,

$$\mathbf{E}[(V - \mu)(Y + \mu - g(Y))] = \mathbf{E}[Y + \mu - g(Y)]\mathbf{E}[V - \mu] = 0$$

(recall that $\mu = \mathbf{E}V$). Thus,

$$\mathbf{E}(X - W)^2 = \mathbf{E}[(V - \mu + Y + \mu - g(Y))^2] = \mathbf{E}[(Y + \mu - g(Y))^2] + \mathbf{E}[(V - \mu)^2]\,,$$

is minimal for $W = g(Y) = Y + \mu$. Comparing this result with Examples 2.3.1 and 2.3.2, we see that

$$\mathbf{E}[X|Y] = \mathbf{E}[Y + V|Y] = \mathbf{E}[Y|Y] + \mathbf{E}[V|Y] = Y + \mathbf{E}(V)\,.$$

With additional work (that we shall not detail) one can show that if $f(Y, V)$ is integrable then $\mathbf{E}(f(Y, V)|Y) = g(Y)$ where $g(y) = \mathbf{E}(f(y, V))$ (so the preceding corresponds to $f(y, v) = y + v$ and $g(y) = y + \mu$).

As we next show, for any probability space $(\Omega, \mathcal{F}, \mathbf{P})$ the conditional expectation of X given a σ-field $\mathcal{G} \subseteq \mathcal{F}$ is linear in the random variable X.

PROPOSITION 2.3.4. *Let $X, Y \in L^1(\Omega, \mathcal{F}, \mathbf{P})$. Then,*

$$\mathbf{E}(\alpha X + \beta Y|\mathcal{G}) = \alpha\mathbf{E}(X|\mathcal{G}) + \beta\mathbf{E}(Y|\mathcal{G})$$

for any $\alpha, \beta \in \mathbb{R}$.

PROOF. Let $Z = \mathbf{E}(X|\mathcal{G})$ and $V = \mathbf{E}(Y|\mathcal{G})$. Fixing $A \in \mathcal{G}$, it follows by linearity of the expectation operator and the definition of Z and V that

$$\mathbf{E}(I_A(\alpha Z+\beta V)) = \alpha\mathbf{E}(I_A Z)+\beta\mathbf{E}(I_A V) = \alpha\mathbf{E}(I_A X)+\beta\mathbf{E}(I_A Y) = \mathbf{E}(I_A(\alpha X+\beta Y)).$$

Since this applies for all $A \in \mathcal{G}$, it follows by definition that the R.V. $\alpha Z + \beta V$ on $(\Omega, \mathcal{G})$ is the C.E. of $\alpha X + \beta Y$ given $\mathcal{G}$, as stated. ∎

We next deal with the relation between the C.E. of X given two different σ-fields, one of which contains the other.

PROPOSITION 2.3.5 (Tower property). *Suppose $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ and $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$. Then, $\mathbf{E}(X|\mathcal{H}) = \mathbf{E}(\mathbf{E}(X|\mathcal{G})|\mathcal{H})$.*

PROOF. Let $Y = \mathbf{E}(X|\mathcal{G})$ and $Z = \mathbf{E}(Y|\mathcal{H})$. Fix $A \in \mathcal{H}$. Note that $A \in \mathcal{G}$ and consequently, by Definition 2.1.5 of $Y = \mathbf{E}(X|\mathcal{G})$ we have that

$$\mathbf{E}(I_A X) = \mathbf{E}(I_A Y) = \mathbf{E}(I_A Z)\,,$$

where the second identity holds by the Definition 2.1.5 of $Z = \mathbf{E}(Y|\mathcal{H})$ and the fact that $A \in \mathcal{H}$. Since $\mathbf{E}(I_A X) = \mathbf{E}(I_A Z)$ for any $A \in \mathcal{H}$, the proposition follows by the uniqueness of $\mathbf{E}(X|\mathcal{H})$ and its characterization in Definition 2.1.5. ∎

REMARK. Note that any σ-field $\mathcal{G}$ contains the *trivial σ-field* $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Consequently, when $\mathcal{H} = \mathcal{F}_0$ the tower property applies for any σ-field $\mathcal{G}$. Further, $\mathbf{E}(Y|\mathcal{F}_0) = \mathbf{E}Y$ for any integrable random variable Y (c.f. Exercise 2.3.3). Consequently, we deduce from the tower property that

$$\mathbf{E}(X) = \mathbf{E}(X|\mathcal{F}_0) = \mathbf{E}[\mathbf{E}(X|\mathcal{G})|\mathcal{F}_0] = \mathbf{E}(\mathbf{E}(X|\mathcal{G})), \tag{2.3.1}$$

for any $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ and any σ-field $\mathcal{G}$.

Your next exercise shows that we cannot dispense of the relationship between the two σ-fields in the tower property.

EXERCISE 2.3.6. *Give an example of a R.V. X and two σ-fields $\mathcal{F}_1$ and $\mathcal{F}_2$ on $\Omega = \{a, b, c\}$ in which $\mathbf{E}(\mathbf{E}(X|\mathcal{F}_1)|\mathcal{F}_2) \neq \mathbf{E}(\mathbf{E}(X|\mathcal{F}_2)|\mathcal{F}_1)$.*

Here are two important applications of the highly useful tower property (also called the law of iterated expectations).

EXERCISE 2.3.7. *Let $\mathsf{Var}(Y|\mathcal{G}) = \mathbf{E}(Y^2|\mathcal{G}) - \mathbf{E}(Y|\mathcal{G})^2$, where Y is a square-integrable random variable and $\mathcal{G}$ is a σ-field.*

(a) *Check that $\mathsf{Var}(Y|\mathcal{G}) = 0$ whenever $Y \in L^2(\Omega, \mathcal{G}, \mathbf{P})$.*
(b) *More generally, show that $\mathsf{Var}(Y) = \mathbf{E}(\mathsf{Var}(Y|\mathcal{G})) + \mathsf{Var}(\mathbf{E}(Y|\mathcal{G}))$.*
(c) *Show that if $\mathbf{E}(Y|\mathcal{G}) = X$ and $\mathbf{E}X^2 = \mathbf{E}Y^2 < \infty$ then $X = Y$ almost surely.*

EXERCISE 2.3.8. *Suppose that $\mathcal{G}_1$ and $\mathcal{G}_2$ are σ-fields and $\mathcal{G}_1 \subseteq \mathcal{G}_2$. Show that for any square-integrable R.V. X,*

$$\mathbf{E}[(X - \mathbf{E}(X|\mathcal{G}_2))^2] \leq \mathbf{E}[(X - \mathbf{E}(X|\mathcal{G}_1))^2].$$

Keeping the σ-field $\mathcal{G}$ fixed, we may think of the C.E. as an expectation in a different (conditional) probability space. Consequently, every property of the expectation has a corresponding extension to the C.E. For example, the extension of Proposition 1.2.36 is given next without proof (see [**GS01**, Exercise 7.9.4, page 349] for more details).

PROPOSITION 2.3.9 (Jensen's inequality). *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function (that is, $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$ for any x, y and $\lambda \in [0,1]$). Suppose $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ is such that $\mathbf{E}|f(X)| < \infty$. Then, $\mathbf{E}(f(X)|\mathcal{G}) \ge f(\mathbf{E}(X|\mathcal{G}))$.*

EXAMPLE 2.3.10. *Per $q \ge 1$, applying Jensen's inequality for the convex function $f(x) = |x|^q$ we have for all $X \in L^q(\Omega, \mathcal{F}, \mathbf{P})$, that $\mathbf{E}(|X|^q|\mathcal{G}) \ge |\mathbf{E}(X|\mathcal{G})|^q$.*

Combining this example with the tower property and monotonicity of the expectation operator, we find that

COROLLARY 2.3.11. *For each $q \ge 1$, the norm of the conditional expectation of $X \in L^q(\Omega, \mathcal{F}, \mathbf{P})$ given a σ-field $\mathcal{G}$ never exceeds the (L^q)-norm of X.*

PROOF. Recall Example 2.3.10 that $\mathbf{E}(|X|^q|\mathcal{G}) \ge |\mathbf{E}(X|\mathcal{G})|^q$ almost surely. Hence, by the tower property and monotonicity of the expectation operator,

$$\|X\|_q^q = \mathbf{E}(|X|^q) = \mathbf{E}[\mathbf{E}(|X|^q|\mathcal{G})] \ge \mathbf{E}[|\mathbf{E}(X|\mathcal{G})|^q] = \|\mathbf{E}(X|\mathcal{G})\|_q^q .$$

That is, $\|X\|_q \ge \|\mathbf{E}(X|\mathcal{G})\|_q$ for any $q \ge 1$. ∎

We also have the corresponding extensions of the Monotone and Dominated Convergence theorems of Section 1.4.2.

THEOREM 2.3.12 (Monotone Convergence for C.E.). *If $0 \le X_m \nearrow X$ and $\mathbf{E}(X) < \infty$ then $\mathbf{E}(X_m|\mathcal{G}) \nearrow \mathbf{E}(X|\mathcal{G})$ a.s. (see* [**Bre92**, Proposition 4.24] *for proof).*

THEOREM 2.3.13 (Dominated Convergence for C.E.). *If $|X_m| \le Y \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ and $X_m \stackrel{a.s}{\to} X$, then $\mathbf{E}(X_m|\mathcal{G}) \stackrel{a.s.}{\to} \mathbf{E}(X|\mathcal{G})$.*

REMARK. In contrast with Theorem 1.4.27, the convergence *in probability* to X by X_m that are dominated by an integrable R.V. Y does not imply the convergence almost surely of $\mathbf{E}(X_m|\mathcal{G})$ to $\mathbf{E}(X|\mathcal{G})$. Indeed, taking $\mathcal{G} = \mathcal{F}$ we see that such a statement would have contradicted Proposition 1.3.8.

Our next result shows that the C.E. operation is continuous on each of the L^q spaces.

THEOREM 2.3.14. *Suppose $X_n \stackrel{q.m.}{\to} X$, that is, $X_n, X \in L^q(\Omega, \mathcal{F}, \mathbf{P})$ with $\mathbf{E}(|X_n - X|^q) \to 0$. Then, $\mathbf{E}(X_n|\mathcal{G}) \stackrel{q.m.}{\to} \mathbf{E}(X|\mathcal{G})$.*

PROOF. By the linearity of conditional expectations and Jensen's Inequality,

$$\begin{aligned} \mathbf{E}(|\mathbf{E}(X_n|\mathcal{G}) - \mathbf{E}(X|\mathcal{G})|^q) &= \mathbf{E}(|\mathbf{E}(X_n - X|\mathcal{G})|^q) \\ &\le \mathbf{E}(\mathbf{E}(|X_n - X|^q|\mathcal{G})) = \mathbf{E}(|X_n - X|^q) \to 0, \end{aligned}$$

by hypothesis. ∎

We next show that one can *take out what is known* when computing the C.E.

PROPOSITION 2.3.15. *Suppose Y is bounded and measurable on $\mathcal{G}$, and that $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$. Then, $\mathbf{E}(XY|\mathcal{G}) = Y\mathbf{E}(X|\mathcal{G})$.*

PROOF. Note that $XY \in L^1$ and $Z = Y\mathbf{E}(X|\mathcal{G})$ is measurable on $\mathcal{G}$. Now, for any $A \in \mathcal{G}$,

$$\mathbf{E}[(XY - Z)I_A] = \mathbf{E}[(X - \mathbf{E}(X|\mathcal{G}))(YI_A)] = 0 ,$$

by the orthogonality property (2.1.5) of $\mathbf{E}(X|\mathcal{G})$ (since $V = YI_A$ is bounded and measurable on $\mathcal{G}$). Since $A \in \mathcal{G}$ is arbitrary, the proposition follows (see Definition 2.1.5 and Theorem 2.1.6). ∎

You can check your understanding by proving that Proposition 2.3.15 holds also whenever $Y \in L^2(\Omega, \mathcal{G}, \mathbf{P})$ and $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ (actually, this proposition applies as soon as Y is measurable on $\mathcal{G}$ and both X and XY are integrable).

EXERCISE 2.3.16. *Let $Z = (X, Y)$ be a uniformly chosen point in $(0, 1)^2$. That is, X and Y are independent random variables, each having the $U(0, 1)$ measure of Example 1.1.11. Set $T = I_A(Z) + 5I_B(Z)$ where $A = \{(x, y) : 0 < x < 1/4, 3/4 < y < 1\}$ and $B = \{(x, y) : 3/4 < x < 1, 0 < y < 1/2\}$.*

(a) *Find an explicit formula for the conditional expectation $W = \mathbf{E}(T|X)$ and use it to determine the conditional expectation $U = \mathbf{E}(TX|X)$.*
(b) *Find the value of $\mathbf{E}[(T - W) \sin(e^X)]$.*
(c) *Without any computation decide whether $\mathbf{E}(W^2) - \mathbf{E}(T^2)$ is negative, zero, or positive. Explain your answer.*

We conclude this section with a proposition extending Example 2.3.1 and a pair of exercises dealing with the connection (or lack thereof) between independence and C.E.

PROPOSITION 2.3.17. *If X is integrable and σ-fields $\mathcal{G}$ and $\sigma(\sigma(X), \mathcal{H})$ are independent, then*

$$\mathbf{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbf{E}[X|\,\mathcal{H}] \,.$$

EXERCISE 2.3.18. *Building on Exercise 1.4.38 provide a σ-field $\mathcal{G} \subseteq \mathcal{F}$ and a bounded random variable X which is independent of integrable Y, such that:*

(a) *$\mathbf{E}(X|\mathcal{G})$ is not independent of $\mathbf{E}(Y|\mathcal{G})$.*
(b) *$\mathbf{E}(XY|\mathcal{G}) \neq \mathbf{E}(X|\mathcal{G})\mathbf{E}(Y|\mathcal{G})$.*

EXERCISE 2.3.19. *Suppose that X and Y are square integrable random variables.*

(a) *Show that if $\mathbf{E}(X|Y) = \mathbf{E}(X)$ then X and Y are uncorrelated.*
(b) *Provide an example of uncorrelated X and Y for which $\mathbf{E}(X|Y) \neq \mathbf{E}(X)$.*
(c) *Provide an example where $\mathbf{E}(X|Y) = \mathbf{E}(X)$ but X and Y are not independent (this is also an example of uncorrelated but not independent R.V.).*

2.4. Regular conditional probability

Fixing two σ-fields $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$, let $\widetilde{\mathbf{P}}(A|\mathcal{G}) = \mathbf{E}(I_A|\mathcal{G})$ for any $A \in \mathcal{F}$. Obviously, $\widetilde{\mathbf{P}}(A|\mathcal{G}) \in [0, 1]$ exists for any such A and almost all $\omega \in \Omega$. Ideally, we would expect that the collection $\{\widetilde{\mathbf{P}}(A|\mathcal{G}) : A \in \mathcal{H}\}$ is also a probability measure on $(\Omega, \mathcal{H})$ for any fixed $\omega \in \Omega$. When this is the case, we call $\widetilde{\mathbf{P}}(\cdot|\mathcal{G})$ the Regular Conditional Probability (R.C.P.) on $\mathcal{H}$ given $\mathcal{G}$. Note that all conditional expectations can then be defined through the R.C.P. (see [**Bre92**, Proposition 4.28] for more details). Unfortunately, the R.C.P. fails to exist in general. Here is the reason for this unexpected difficulty. As the C.E. is only defined in the almost sure sense, for any countable collection of disjoint sets $A_n \in \mathcal{H}$ there is possibly a set of $\omega \in \Omega$ of probability zero for which

$$\widetilde{\mathbf{P}}(\bigcup_n A_n|\mathcal{G}) \neq \sum_n \widetilde{\mathbf{P}}(A_n|\mathcal{G}).$$

We may need to deal with an uncountable number of such collections to check the R.C.P. property throughout $\mathcal{H}$, causing the corresponding exceptional sets of ω to pile up to a non-negligible set.

The way we avoid this problem is by focusing on the R.C.P. for the special choice of $\mathcal{H} = \sigma(X)$ with X a specific real-valued R.V. of interest. That is, we restrict our attention to $A = \{\omega : X(\omega) \in B\}$ for a Borel set B. The existence of the R.C.P. is then equivalent to the question whether $\widetilde{\mathbf{P}}(X \in B|\mathcal{G}) = \widehat{\mathbf{P}}(B|\mathcal{G})$ is a probability measure on $(\mathbb{R}, \mathcal{B})$, as precisely defined next.

Definition 2.4.1. *Let X be a R.V. on $(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a σ-field. The collection $\widehat{\mathbf{P}}(B|\mathcal{G})$ is called the* regular conditional probability distribution *(R.C.P.D.) of X given $\mathcal{G}$ if:*
(a) $\widehat{\mathbf{P}}(B|\mathcal{G}) = \mathbf{E}[I_{X\in B}|\mathcal{G}]$ for any fixed Borel set $B \subseteq \mathbb{R}$. That is, $\widehat{\mathbf{P}}(B|\mathcal{G})$ is measurable on $(\Omega, \mathcal{G})$ such that $\mathbf{P}(\{\omega : X(\omega) \in B\} \cap A) = \mathbf{E}[I_A \widehat{\mathbf{P}}(B|\mathcal{G})]$ for any $A \in \mathcal{G}$.
(b) The collection $\widehat{\mathbf{P}}(\cdot|\mathcal{G})$ is a probability measure on $(\mathbb{R}, \mathcal{B})$ for any fixed $\omega \in \Omega$.

The R.C.P.D. is thus the analog of the law of X as in Definition 1.4.1, now given the information contained in $\mathcal{G}$ (which is why we added the word "distribution" to it). The essential point is of course that any real-valued R.V. has a R.C.P.D.

Proposition 2.4.2. *For any random variable X and any σ-field $\mathcal{G} \subseteq \mathcal{F}$, there exists a R.C.P.D. $\widehat{\mathbf{P}}(\cdot|\mathcal{G})$ (see* [**Bre92**, Theorem 4.30] *for a proof).*

We note in passing that the reason R.C.P.D. always exists is that the σ-field $\sigma(X)$ inherits the structure of $\mathcal{B}$ which in turn is "not too big" due to the fact the rational numbers are dense in $\mathbb{R}$.

To practice your understanding of Definition 2.4.1 solve the following exercise.

Exercise 2.4.3. *Show that $\sigma(X)$ and $\mathcal{G}$ are independent if and only if $\widehat{\mathbf{P}}(B|\mathcal{G}) = \mathbf{P}(X \in B)$ for all $B \in \mathcal{B}$.*

An alternative way of defining the R.C.P.D. is $\widehat{\mathbf{P}}_{X|\mathcal{G}}(\cdot,\cdot) : \mathcal{B} \times \Omega \to [0,1]$ such that
(a) For each $B \in \mathcal{B}$, almost surely $\widehat{\mathbf{P}}_{X|\mathcal{G}}(B,\cdot) = \mathbf{E}[I_{X\in B}|\mathcal{G}]$ and is measurable.
(b) For any ω, the function $\widehat{\mathbf{P}}_{X|\mathcal{G}}(\cdot,\omega)$ is a probability measure.

The R.C.P.D. gives us yet another alternative definition for the C.E.

Definition 2.4.4. *The* conditional expectation *of an integrable random variable X given a σ-field $\mathcal{G}$ is*

$$\mathbf{E}[X|\mathcal{G}] = \int_{-\infty}^{\infty} x \widehat{\mathbf{P}}_{X|\mathcal{G}}(dx,\omega),$$

where the right side is to be interpreted in the sense of Definition 1.2.19 of Lebesgue's integral *for the probability space $(\mathbb{R}, \mathcal{B}, \widehat{\mathbf{P}}_{X|\mathcal{G}}(\cdot,\omega))$.*

For a proof that this definition coincides with our earlier definition of the conditional expectation see for example [**Bre92**, Proposition 4.28].

Example 2.4.5. *The R.C.P.D. is rather explicit when $\mathcal{G} = \sigma(Y)$ and the random vector (X, Y) has a* probability density function*, denoted $f_{X,Y}$. That is, when for all $x, y \in \mathbb{R}$,*

$$\mathbf{P}(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\, du\, dv\,, \tag{2.4.1}$$

with $f_{X,Y}$ a non-negative Borel function on $\mathbb{R}^2$ such that $\int_{\mathbb{R}^2} f_{X,Y}(u,v)dudv = 1$. In this case, the R.C.P.D. of X given $\mathcal{G} = \sigma(Y)$ has the density function $f_{X|Y}(x|Y(\omega))$ where

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}\,,$$

and $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(v,y)dv$. Consequently, by Definition 2.4.4 we see that here

$$\mathbf{E}(X|Y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|Y)dx\,,$$

in agreement with the classical formulas provided in elementary (non-measure theoretic) probability courses.

To practice your understanding of the preceding example, solve the next exercise.

EXERCISE 2.4.6.

(a) *Suppose that the joint law of (X,Y,Z) has a density. Express the R.C.P.D. of Y given X,Z, in terms of this density.*

(b) *Using this expression show that if X is independent of the pair (Y,Z), then*

$$\mathbf{E}(Y|X,Z) = \mathbf{E}(Y|Z).$$

(c) *Give an example of random variables X,Y,Z, such that X is independent of Y and*

$$\mathbf{E}(Y|X,Z) \neq \mathbf{E}(Y|Z).$$

CHAPTER 3

Stochastic Processes: general theory

There are two ways to approach the definition of a stochastic process (S.P.) The easier approach, which we follow in Section 3.1, is to view a S.P. as a collection of R.V.-s $\{X_t(\omega), t \in \mathcal{I}\}$, focusing on the finite dimensional distributions and ignoring sample path properties (such as continuity). Its advantages are:
• The index set $\mathcal{I}$ can be arbitrary;
• The measurability of $t \mapsto X_t(\omega)$ as a function from $\mathcal{I}$ to $\mathbb{R}$ (per fixed ω), is not part of the definition of the process;
• There is no need to master functional analysis techniques and results.

The main disadvantages of this approach are:
• It only works when $X_t(\omega)$ is well-defined for each fixed t;
• Using it, we cannot directly determine important properties such as continuity or monotonicity of the sample path, or the law of $\sup_t X_t$.

A more sophisticated approach to S.P., mostly when $\mathcal{I}$ is an interval on the real line, views the S.P. as a function-valued R.V: $\Omega \to$ [Space of functions] (examples of such spaces might be $C[0,1]$ or $L^2[0,1]$). The main advantages of this approach are the complements of the disadvantages of our approach and vice verse.

Defining and using characteristic functions we study in Section 3.2 the important class of Gaussian stochastic processes. We conclude this chapter by detailing in Section 3.3 sufficient conditions for the continuity of the sample path $t \mapsto X_t(\omega)$, for almost all outcomes ω.

3.1. Definition, distribution and versions

Our starting point is thus the following definition of what a stochastic process is.

Definition 3.1.1. *Given* $(\Omega, \mathcal{F}, \mathbf{P})$*, a* stochastic process *(S.P.)* $\{X_t\}$ *is a collection* $\{X_t : t \in \mathcal{I}\}$ *of R.V.-s where the index* t *belongs to the index set* $\mathcal{I}$*. Typically,* $\mathcal{I}$ *is an interval in* $\mathbb{R}$ *(in which case we say that* $\{X_t\}$ *is a* continuous time stochastic process*), or a subset of* $\{1, 2, \ldots, n, \ldots\}$ *(in which case we say that* $\{X_t\}$ *is a* discrete time stochastic process. *We also call* $t \mapsto X_t(\omega)$ *the* sample function *(or* sample path*) of the S.P.*

Recall our notation $\sigma(X_t)$ for the σ-field generated by X_t. The discrete time stochastic processes are merely countable collections of R.V.-s $X_1, X_2, X_3, \ldots$ defined on the same probability space. All relevant information about such a process during a finite time interval $\{1, 2, \ldots, n\}$ is conveyed by the σ-field $\sigma(X_1, X_2, \ldots, X_n)$, namely, the σ-field generated by the "rectangle" sets $\bigcap_{i=1}^n \{\omega : X_i(\omega) \le \alpha_i\}$ for $\alpha_i \in \mathbb{R}$ (compare with Definition 1.2.8 of $\sigma(X)$). To deal with the full infinite time horizon we just take the σ-field $\sigma(X_1, X_2, \ldots)$ generated by the union of these sets

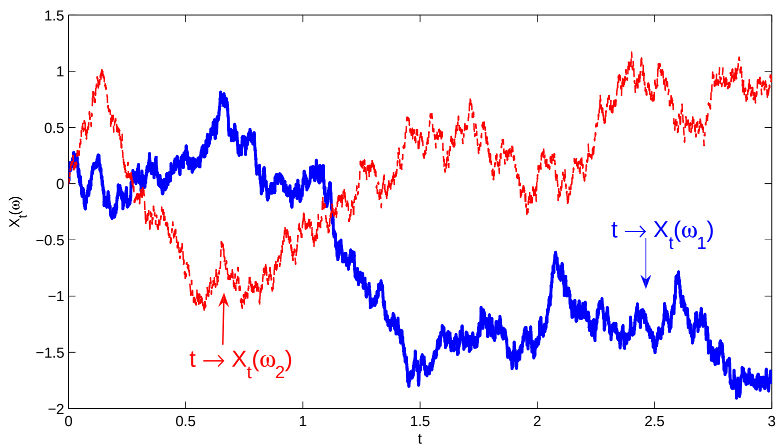

FIGURE 1. Two sample paths of a stochastic process, corresponding to two outcomes ω_1 and ω_2.

over $n = 1, 2, \dots$ (this is exactly what we did to define $\mathcal{F}_c$ of the coin tosses Example 1.1.13). Though we do not do so here, it is not hard to verify that in this setting the σ-field $\sigma(X_1, X_2, \dots)$ coincides with the smallest σ-field containing $\sigma(X_t)$ for all $t \in \mathcal{I}$, which we denote hereafter by $\mathcal{F}_{\mathbf{X}}$.

Perhaps the simplest example of a discrete time stochastic process is that of independent, identically distributed R.V.-s X_n (see Example 1.4.7 for one such construction). Though such a S.P. has little interesting properties worth study, it is the corner stone of the following, fundamental example, called the random walk.

DEFINITION 3.1.2. *A* random walk *is the sequence* $S_n = \sum_{i=1}^n \xi_i$, *where* ξ_i *are independent and identically distributed real-valued R.V.-s defined on the same probability space* $(\Omega, \mathcal{F}, \mathbf{P})$. *When* ξ_i *are integer-valued we say that this is a random walk on the integers, and call the special case of* $\xi_i \in \{-1, 1\}$ *a simple random walk.*

When considering a continuous-time S.P. we deal with uncountable collections of R.V.. As we soon see, this causes many difficulties. For example, when we talk about the *distribution* (also called the *law*) of such a S.P. we usually think of the restriction of $\mathbf{P}$ to a certain σ-field. But which σ-field to choose? That is, which one carries all the information we have in mind? Clearly, we require at least $\mathcal{F}_{\mathbf{X}}$ so we can determine the law $\mathcal{P}_{X_t}$ of $X_t(\omega)$ per fixed t. But is this enough? For example, we might be interested in sets such as $H_f = \{\omega : X_t(\omega) \le f(t), \forall 0 \le t \le 1\}$ for some functions $f : [0,1] \mapsto \mathbb{R}$. Indeed, the value of $\sup(X_t : 0 \le t \le 1)$ is determined by such sets for $f(t) = \alpha$ independent of t. Unfortunately, such sets typically involve an uncountable number of set operations and thus are usually not in $\mathcal{F}_{\mathbf{X}}$, and hence $\sup(X_t : 0 \le t \le 1)$ might not even be measurable on $\mathcal{F}_{\mathbf{X}}$. So, maybe we should take instead $\sigma(\{H_f, f : [0,1] \to \mathbb{R}\})$ which is quite different from $\mathcal{F}_{\mathbf{X}}$? We choose in these notes the smaller, i.e. simpler, σ-field $\mathcal{F}_{\mathbf{X}}$, but what is then the minimal information we need for specifying uniquely a probability measure on this space? Before tackling this issue, we provide some motivation to our interest in continuous-time S.P.

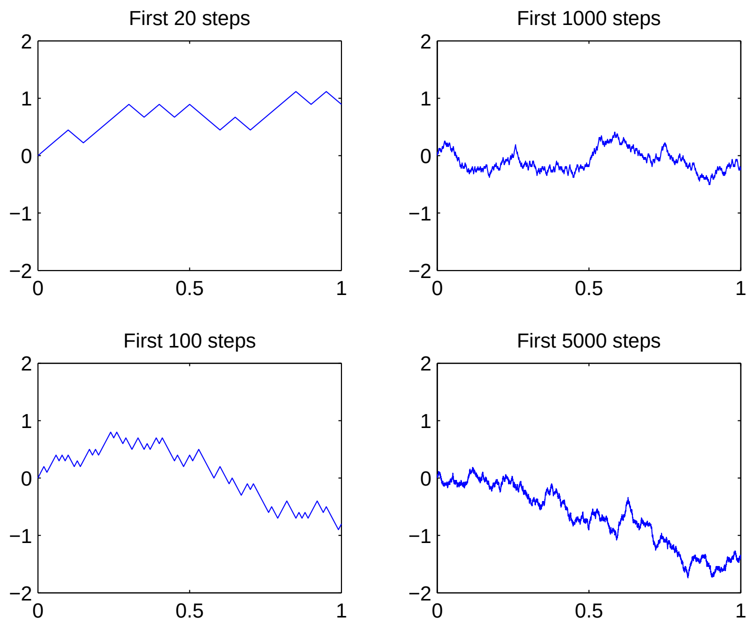

FIGURE 2. Illustration of scaled simple random walks for different values of n.

Assuming that $\mathbf{E}(\xi_1) = 0$ and $\mathbf{E}(\xi_1^2) = 1$ we have that $\mathbf{E}(S_n) = 0$ and

$$\mathbf{E}(S_n^2) = \mathbf{E}\Big[(\sum_{i=1}^n \xi_i)^2\Big] = \sum_{i,j=1}^n \mathbf{E}(\xi_i \xi_j) = n\,, \tag{3.1.1}$$

that is, $\mathbf{E}\Big[(n^{-1/2}S_n)^2\Big] = 1$. Further, in this case, by central limit theorem we have that

$$n^{-1/2}S_n \xrightarrow{\mathcal{L}} G = N(0,1)\,, \tag{3.1.2}$$

with $N(\mu, v)$ denoting a Gaussian R.V. of mean μ and variance v (c.f. Example 1.4.13). Replacing n by $[nt]$ (the integer part of nt), we get from (3.1.2) by rescaling that also $n^{-1/2}S_{[nt]} \xrightarrow{\mathcal{L}} N(0,t)$ for any fixed $0 \le t \le 1$. This leads us to state the functional C.L.T. where all values of $0 \le t \le 1$ are considered at once.

THEOREM 3.1.3. *(see* [**Bre92**, Section 12.2]*) Consider the random walk S_n when $\mathbf{E}(\xi_1) = 0$ and $\mathbf{E}(\xi_1^2) = 1$. Take the linear interpolation of the sequence S_n, scale space by $n^{-\frac{1}{2}}$ and time by n^{-1} (see Figure 2). Taking $n \to \infty$ we arrive at a limiting object which we call the* Brownian motion *on $0 \le t \le 1$. The convergence here is weak convergence in the sense of Definition 1.4.20 with $\mathbb{S}$ the set of continuous functions on $[0,1]$, equipped with the topology induced by the supremum norm.*

Even though it is harder to define the Brownian motion (being a continuous time S.P.), computations for it typically involve relatively simple Partial Differential Equations and are often more explicit than those for the random walk. In addition, unlike the random walk, the Brownian motion does not depend on the specific law of ξ_i (beyond having zero mean and unit variance). For a direct construction of the Brownian motion, to which we return in Section 5.1, see [**Bre92**, Section 12.7].

REMARK. The condition $\mathbf{E}(\xi_1^2) < \infty$ is almost necessary for the $n^{-1/2}$ scaling of space and the Brownian limiting process. Indeed, if $\mathbf{E}(\xi_1^\alpha) = \infty$ for some $0 < \alpha < 2$, then both scaling of space and the limit S.P. are changed with respect to Theorem 3.1.3.

In the next example, which is mathematically equivalent to Theorem 3.1.3, we replace the sum of independent and identically distributed R.V.-s by the product of such non-negative R.V.-s.

EXAMPLE 3.1.4. *Let $M_n = \prod_{i=1}^n Y_i$ where Y_i are positive, independent identically distributed random variables (for instance, the random daily return rates of a certain investment). Let $S_n = \log M_n$ and $\xi_i = \log Y_i$, noting that $S_n = \sum_{i=1}^n \xi_i$ is a random walk. Assuming $\mathbf{E}(\log Y_1) = 0$ and $\mathbf{E}[(\log Y_1)^2] = 1$ we know that $n^{-1/2} S_{[nt]} = n^{-1/2} \log M_{[nt]}$ converges to a Brownian motion W_t as $n \to \infty$. Hence, the limit behavior of M_n is related to that of the* Geometric Brownian motion *e^{W_t}.*

For continuous time stochastic processes we provide next few examples of events that are in the σ-field $\mathcal{F}_\mathbf{X}$ and few examples that in general may not belong to this σ-field (see [**Bre92**, Section 12.4] for more examples).

(1). $\{\omega : X_{t_1} \le \alpha\} \in \mathcal{F}_\mathbf{X}$.

(2). $\{\omega : X_{1/k} \le \alpha, k = 1, 2, \ldots, N\} \in \mathcal{F}_\mathbf{X}$.

(3). $\{\omega : X_{1/k} \le \alpha, \quad k = 1, 2, 3, \ldots\} = \{\omega : \sup_k X_{1/k}(\omega) \le \alpha\} \in \mathcal{F}_\mathbf{X}$.

(4). $\{\omega : \sup(X_t : 0 \le t \le 1) \le \alpha\}$ is not necessarily in $\mathcal{F}_\mathbf{X}$ since the supremum here is over an uncountable collection of R.V.-s. However, check that this event is in $\mathcal{F}_\mathbf{X}$ whenever all sample functions of the S.P. $\{X_t\}$ are right continuous.

(5). $\{\omega : X_t(\omega) : \mathcal{I} \mapsto \mathbb{R}$ is a measurable function$\}$ may also be outside $\mathcal{F}_\mathbf{X}$ (say for $\mathcal{I} = [0, 1]$).

We define next the stochastic process analog of the distribution function.

DEFINITION 3.1.5. *Given $N < \infty$ and a collection $t_1, t_2, \ldots, t_N$ in $\mathcal{I}$, we denote the (joint) distribution of $(X_{t_1}, \ldots, X_{t_N})$ by $F_{t_1, t_2, \cdots, t_N}(\cdot)$, that is,*

$$F_{t_1,t_2,\cdots,t_N}(\alpha_1, \alpha_2, \cdots, \alpha_N) = \mathbf{P}(X_{t_1} \le \alpha_1, \cdots, X_{t_N} \le \alpha_N),$$

for all $\alpha_1, \alpha_2, \cdots, \alpha_N \in \mathbb{R}$. We call the collection of functions $F_{t_1,t_2,\cdots,t_N}(\cdot)$, the finite dimensional distributions *(f.d.d.) of the S.P.*

Having independent increments is one example of a property that is determined by the f.d.d.

DEFINITION 3.1.6. *With $\mathcal{G}_t$ the smallest σ-field containing $\sigma(X_s)$ for any $0 \le s \le t$, we say that a S.P. $\{X_t\}$ has* independent increments *if $X_{t+h} - X_t$ is independent of $\mathcal{G}_t$ for any $h > 0$ and all $t \ge 0$. This property is determined by the f.d.d. That is, if the random variables $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are mutually independent, for all $n < \infty$ and $0 \le t_1 < t_2 < \cdots < t_n < \infty$ then the S.P. $\{X_t\}$ has independent increments.*

REMARK. For example, both the random walk and the Brownian motion are processes with independent increments.

The next example shows that the f.d.d. do not determine some other important properties of the stochastic process, such as *continuity* of its sample path. That is, knowing all f.d.d. is not enough for computing $\mathbf{P}(\{\omega : t \mapsto X_t(\omega)$ is continuous on $[0,1]\})$.

EXAMPLE 3.1.7. *Consider the probability space $\Omega = [0,1]$ with its Borel σ-field and the Uniform law on $[0,1]$ (that is, the probability of each interval equals its length, also known as Lebesgue measure restricted to $[0,1]$). Given $\omega \in \Omega$, we define two Stochastic Processes:*

$$Y_t(\omega) = 0, \quad \forall t, \omega \qquad\qquad X_t(\omega) = \begin{cases} 1, t = \omega \\ 0, \ otherwise \end{cases}$$

Let $A_t = \{\omega : X_t \neq Y_t\} = \{t\}$. Since $\mathbf{P}(A_t) = 0$, we have that $\mathbf{P}(X_t = Y_t) = 1$ for each fixed t. Moreover, let $A^N = \bigcup_{i=1}^N A_{t_i}$, then $\mathbf{P}(A^N) = 0$ (a finite union of negligible sets is negligible). Since this applies for any choice of N and $t_1, \dots, t_N$, we see that the f.d.d. of $\{X_t\}$ are the same as those of $\{Y_t\}$. Moreover, considering the set $A^\infty = \bigcup_{i=1}^\infty A_{t_i}$, involving a countable number of times, we see that $\mathbf{P}(A^\infty) = 0$, that is, almost surely, $X_t(\omega)$ agrees with $Y_t(\omega)$ at any fixed, countable, collection of times. But note that some global sample-path properties do not agree. For example,

$$\mathbf{P}(\{\omega : (\sup\{X_t(\omega) : 0 \le t \le 1\}) \neq 0\}) = 1,$$

$$\mathbf{P}(\{\omega : (\sup\{Y_t(\omega) : 0 \le t \le 1\}) \neq 0\}) = 0.$$

Also,

$$\mathbf{P}(\{\omega : t \mapsto X_t(\omega) \text{ is continuous}\}) = 0,$$

$$\mathbf{P}(\{\omega : t \mapsto Y_t(\omega) \text{ is continuous}\}) = 1.$$

While the maximal value and continuity of sample path are different for the two S.P. of Example 3.1.7, we should typically consider such a pair to be the same S.P., motivating our next two definitions.

DEFINITION 3.1.8. *Two S.P. $\{X_t\}$ and $\{Y_t\}$ are called* versions *of one another if they have the same finite-dimensional distributions.*

DEFINITION 3.1.9. *A S.P. $\{Y_t\}$ is called a* modification *of another S.P. $\{X_t\}$ if $\mathbf{P}(Y_t = X_t) = 1$ for all $t \in \mathcal{I}$.*

We consider next the relation between the concepts of modification and version, starting with:

EXERCISE 3.1.10. *Show that if $\{Y_t\}$ is a modification of $\{X_t\}$, then $\{Y_t\}$ is also a version of $\{X_t\}$.*

Note that a modification has to be defined on the same probability space as the original S.P. while this is not required of versions.
The processes in Example 3.1.7 are modifications of one another. In contrast, our next example is of two versions on the same probability space which are not modifications of each other.

EXAMPLE 3.1.11. *Consider* $\Omega_2 = \{HH, TT, HT, TH\}$ *with the uniform probability measure* $\mathbf{P}$*, corresponding to two independent fair coin tosses, whose outcome is* $\omega = (\omega_1, \omega_2)$*. Define on* $(\Omega_2, 2^{\Omega_2}, \mathbf{P})$ *the S.P*

$$X_t(\omega) = \mathbf{1}_{[0,1)}(t) I_H(\omega_1) + \mathbf{1}_{[1,2)}(t) I_H(\omega_2)\,, \qquad 0 \le t < 2\,,$$

and $Y_t(\omega) = 1 - X_t(\omega)$ *for* $0 \le t < 2$*.*

EXERCISE 3.1.12. *To practice your understanding you should at this point check that the processes* X_t *and* Y_t *of Example 3.1.11 are versions of each other but are not modifications of each other.*

DEFINITION 3.1.13. *We say that a collection of finite dimensional distributions is* consistent *if*

$$\lim_{\alpha_k \uparrow \infty} F_{t_1, \cdots, t_N}(\alpha_1, \cdots, \alpha_N) = F_{t_1, \cdots, t_{k-1}, t_{k+1}, t_N}(\alpha_1, \cdots, \alpha_{k-1}, \alpha_{k+1}, \cdots, \alpha_N)\,,$$

for any $1 \le k \le N$*,* $t_1 < t_2 < \cdots < t_N \in \mathcal{I}$ *and* $\alpha_i \in \mathbb{R}$*,* $i = 1, \ldots, N$*.*

Convince yourself that the f.d.d. of any S.P. must be consistent. Conversely,

PROPOSITION 3.1.14. *For any consistent collection of finite dimensional distributions, there exists a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *and a stochastic process* $\omega \mapsto X_t(\omega)$ *on it, whose f.d.d. are in agreement with the given collection (c.f.* [**Bre92**, Theorem 12.14]*, or* [**GS01**, Theorem 8.6.3]*). Further, the restriction of the probability measure* $\mathbf{P}$ *to the* σ*-field* $\mathcal{F}_{\mathbf{X}}$ *is uniquely determined by the given collection of f.d.d.*

We note in passing that the construction of Proposition 3.1.14 builds on the easier case of discrete time stochastic processes (which is treated for example in [**Bre92**, Section 2.4]).

We can construct a σ-field that is the image of $\mathcal{F}_{\mathbf{X}}$ on the range of $\omega \mapsto \{X_t(\omega) : t \in \mathcal{I}\} \subseteq \mathbb{R}^{\mathcal{I}}$.

DEFINITION 3.1.15. *For an interval* $\mathcal{I} \subseteq \mathbb{R}$*, let* $\mathbb{R}^{\mathcal{I}}$ *denote the set of all functions* $x : \mathcal{I} \to \mathbb{R}$*. A* finite dimensional rectangle *in* $\mathbb{R}^{\mathcal{I}}$ *is any set of the form* $\{x : x(t_i) \in J_i, i = 1, \ldots, n\}$ *for a non-negative integer* n*, intervals* $J_i \subseteq \mathbb{R}$ *and times* $t_i \in \mathcal{I}$*,* $i = 1, \ldots, n$*. The* cylindrical σ-field $\mathcal{B}^{\mathcal{I}}$ *is the* σ*-field generated by the collection of finite dimensional rectangles.*

• *What follows may be omitted at first reading.*

While f.d.d. do not determine important properties of the sample path $t \mapsto X_t(\omega)$ of the S.P. (see Example 3.1.7), they uniquely determine the probabilities of events in $\mathcal{F}_{\mathbf{X}}$ (hence each property of the sample path that can be expressed as an element of the cylindrical σ-field $\mathcal{B}^{\mathcal{I}}$).

PROPOSITION 3.1.16. *For an interval* $\mathcal{I} \subseteq \mathbb{R}$ *and any S.P.* $\{X_t\}$ *on* $t \in \mathcal{I}$*, the* σ*-field* $\mathcal{F}_{\mathbf{X}}$ *consists of the events* $\{\omega : X_\cdot(\omega) \in \Gamma\}$ *for* $\Gamma \in \mathcal{B}^{\mathcal{I}}$*. Further, if a S.P.* $\{Y_t\}$ *is a version of* $\{X_t\}$*, then* $\mathbf{P}(X_\cdot \in \Gamma) = \mathbf{P}(Y_\cdot \in \Gamma)$ *for all such* Γ *(see* [**Bre92**, Corollary 12.9 and Proposition 12.12] *for proofs).*

3.2. Characteristic functions, Gaussian variables and processes

Subsection 3.2.1 is about the fundamental concept of characteristic function and its properties. Using it, we study in Subsection 3.2.2 the Gaussian random vectors and stochastic processes. Subsection 3.2.3 deals with *stationarity*, an important concept in the general theory of stochastic processes which is simpler to check for the Gaussian processes.

3.2.1. Characteristic function. We start with the definition of the characteristic function of a random vector. In doing so we adopt the convention that a complex valued random variable Z is a function from Ω to $\mathbb{C}$ such that both the real and imaginary parts of Z are Borel measurable, and if $Z = X + iY$ with $X, Y \in \mathbb{R}$ integrable random variables (and $i = \sqrt{-1}$), then $\mathbf{E}(Z) = \mathbf{E}(X) + i\mathbf{E}(Y) \in \mathbb{C}$.

DEFINITION 3.2.1. *A* random vector $\underline{X} = (X_1, X_2, \cdots, X_n)$ *with values in* $\mathbb{R}^n$ *has the* characteristic function

$$\Phi_{\underline{X}}(\underline{\theta}) = \mathbf{E}[e^{i\sum_{k=1}^n \theta_k X_k}]\,,$$

where $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_n) \in \mathbb{R}^n$ *and* $i = \sqrt{-1}$.

REMARK. The characteristic function $\Phi_{\underline{X}} : \mathbb{R}^n \to \mathbb{C}$ exists for any $\underline{X}$ since

$$e^{i\sum_{k=1}^n \theta_k X_k} = \cos(\sum_{k=1}^n \theta_k X_k) + i\sin(\sum_{k=1}^n \theta_k X_k)\,, \tag{3.2.1}$$

with both real and imaginary parts being bounded (hence integrable) random variables. Actually, we see from (3.2.1) that $\Phi_{\underline{X}}(\underline{0}) = 1$ and $|\Phi_{\underline{X}}(\underline{\theta})| \leq 1$ for all $\underline{\theta} \in \mathbb{R}^n$ (see [**Bre92**, Proposition 8.27] or [**GS01**, Section 5.7] for other properties of the characteristic function).

Our next proposition justifies naming $\Phi_{\underline{X}}$ the characteristic function of (the law of) $\underline{X}$.

PROPOSITION 3.2.2. *The characteristic function determines the* law *of a random vector. That is, if* $\Phi_{\underline{X}}(\underline{\theta}) = \Phi_{\underline{Y}}(\underline{\theta})$ *for all* $\underline{\theta}$ *then* $\underline{X}$ *has the same law (= probability measure on* $\mathbb{R}^n$*) as* $\underline{Y}$ *(for proof see* [**Bre92**, Theorems 11.4 and 8.24] *or* [**GS01**, Corollary 5.9.3]*).*

REMARK. The law of a non-negative random variable X is also determined by its moment generating function $M_X(s) = \mathbf{E}[e^{sX}]$ at $s < 0$ (see [**Bre92**, Proposition 8.51] for a proof). While the real-valued function $M_X(s)$ is a simpler object, it is unfortunately useless for the many random variables X which are neither non-negative nor non-positive and for which $M_X(s) = \infty$ for all $s \neq 0$.

The characteristic function is very useful in connection with convergence in law. Indeed,

EXERCISE 3.2.3. *Show that if* $X_n \xrightarrow{\mathcal{L}} X$ *then* $\Phi_{X_n}(\theta) \to \Phi_X(\theta)$ *for any* $\theta \in \mathbb{R}$.

REMARK. Though much harder to prove, the converse of Exercise 3.2.3 is also true, namely if $\Phi_{X_n}(\theta) \to \Phi_X(\theta)$ for each $\theta \in \mathbb{R}$ then $X_n \xrightarrow{\mathcal{L}} X$.

We continue with a few explicit computations of the characteristic function.

EXAMPLE 3.2.4. *Consider X a* Bernoulli*(p) random variable, that is,* $\mathbf{P}(X=1) = p$ *and* $\mathbf{P}(X=0) = 1-p$. *Its characteristic function is by definition*

$$\Phi_X(\theta) = \mathbf{E}[e^{i\theta X}] = pe^{i\theta} + (1-p)e^{i0\theta} = pe^{i\theta} + 1 - p \,.$$

The same type of explicit formula applies to any R.V. $X \in$ SF. *Moreover, such formulas apply for any discrete valued R.V. For example, if $X \sim$* Poisson*(λ) then*

$$\Phi_X(\theta) = \mathbf{E}[e^{i\theta X}] = \sum_{k=0}^{\infty} \frac{(\lambda e^{i\theta})^k}{k!} e^{-\lambda} = e^{\lambda(e^{i\theta}-1)} \,. \tag{3.2.2}$$

The characteristic function has an explicit form also when the R.V. X has a probability density function f_X as in Definition 1.2.23. Indeed, then by Proposition 1.2.29 we have that

$$\Phi_X(\theta) = \int_{-\infty}^{\infty} e^{i\theta x} f_X(x) dx \,, \tag{3.2.3}$$

which is merely the Fourier transform of the density f_X. For example, applying this formula we see that the Uniform random variable U of Example 1.1.11 has characteristic function $\Phi_U(\theta) = (e^{i\theta}-1)/(i\theta)$. Assuming that the density f_X is bounded and continuous, we also have the explicit inversion formula

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\theta x} \Phi_X(\theta) d\theta \,, \tag{3.2.4}$$

as a way to explain Proposition 3.2.2 ([**Bre92**, Theorem 8.39] shows that this inversion formula is valid whenever $\int |\Phi_X(\theta)| d\theta < \infty$, see also [**GS01**, Theorem 5.9.1]).

We next recall the extension of the notion of density as in Definition 1.2.23 to a random vector (as done already in (2.4.1)).

DEFINITION 3.2.5. *We say that a random vector $\underline{X} = (X_1, \ldots, X_n)$ has a* probability density function *$f_{\underline{X}}$ if*

$$\mathbf{P}(\{\omega : a_i \le X_i(\omega) \le b_i, i = 1, \ldots, n\}) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_{\underline{X}}(x_1, \ldots, x_n) dx_n \cdots dx_1 \,,$$

for every $a_i < b_i$, $i = 1, \ldots, n$. Such density $f_{\underline{X}}$ must be a non-negative Borel measurable function with $\int_{\mathbb{R}^n} f_{\underline{X}}(\underline{x}) d\underline{x} = 1$ ($f_{\underline{X}}$ is sometimes called the joint density of $X_1, \ldots, X_n$ as in [**GS01**, Definition 4.5.2]*).*

Adopting the notation $(\underline{\theta}, \underline{x}) = \sum_{k=1}^n \theta_k x_k$ we have the following extension of the Fourier transform formula (3.2.3) to random vectors $\underline{X}$ with density,

$$\Phi_{\underline{X}}(\underline{\theta}) = \int_{\mathbb{R}^n} e^{i(\underline{\theta}, \underline{x})} f_{\underline{X}}(\underline{x}) d\underline{x}$$

(this is merely a special case of the extension of Proposition 1.2.29 to $h : \mathbb{R}^n \to \mathbb{R}$). Though we shall not do so, we can similarly extend the explicit inversion formula of (3.2.4) to $\underline{X}$ having bounded continuous density, or alternatively, having an absolutely integrable characteristic function.

The computation of the characteristic function is much simplified in the presence of independence, as shown by the following alternative of Proposition 1.4.40.

PROPOSITION 3.2.6. *If* $\underline{X} = (X_1, X_2, \ldots, X_n)$ *with* X_i *mutually* independent *R.V., then clearly,*

$$\Phi_{\underline{X}}(\underline{\theta}) = \mathbf{E}[\prod_{k=1}^{n} e^{i\theta_k X_k}] = \prod_{k=1}^{n} \Phi_{X_k}(\theta_k) \qquad \forall\ \underline{\theta} \in \mathbb{R}^n \tag{3.2.5}$$

Conversely, if (3.2.5) holds then the random variables X_i*,* $i = 1, \ldots, n$ *are mutually independent of each other.*

3.2.2. Gaussian variables, vectors and processes. We start by recalling some linear algebra concepts we soon need.

DEFINITION 3.2.7. *An* $n \times n$ *matrix* $\mathbf{A}$ *with entries* A_{jk} *is called* non-negative definite *(or positive semidefinite) if* $A_{jk} = A_{kj}$ *for all* j, k*, and for any* $\underline{\theta} \in \mathbb{R}^n$

$$(\underline{\theta}, \mathbf{A}\underline{\theta}) = \sum_{j=1}^{n}\sum_{k=1}^{n} \theta_j A_{jk} \theta_k \geq 0.$$

We next define the Gaussian random vectors via their characteristic functions.

DEFINITION 3.2.8. *We say that a* random vector $\underline{X} = (X_1, X_2, \cdots, X_n)$ *has a* Gaussian *(or multivariate Normal) distribution if*

$$\Phi_{\underline{X}}(\underline{\theta}) = e^{-\frac{1}{2}(\underline{\theta}, \Sigma\underline{\theta})} e^{i(\underline{\theta}, \underline{\mu})},$$

for some non-negative definite $n \times n$ *matrix* Σ*, some* $\underline{\mu} = (\mu_1, \mu_2, \cdots, \mu_n) \in \mathbb{R}^n$ *and all* $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_n) \in \mathbb{R}^n$*.*

REMARK. In the special case of $n = 1$, we say that a random variable X is Gaussian if for some $\mu \in \mathbb{R}$, some $\sigma^2 \geq 0$ and all $\theta \in \mathbb{R}$,

$$\mathbf{E}[e^{i\theta X}] = e^{-\frac{1}{2}\theta^2\sigma^2 + i\theta\mu}.$$

As we see next, the classical definition of Gaussian distribution via its density amounts to a strict subset of the distributions we consider in Definition 3.2.8.

DEFINITION 3.2.9. *We say that* $\underline{X}$ *has a* non-degenerate *Gaussian distribution if the matrix* Σ *is invertible, or alternatively, when* Σ *is (strictly) positive definite matrix, that is* $(\underline{\theta}, \Sigma\underline{\theta}) > 0$ *whenever* $\underline{\theta}$ *is a non-zero vector (for an equivalent definition see* [**GS01**, Section 4.9]*).*

PROPOSITION 3.2.10. *A random vector* $\underline{X}$ *with a non-degenerate* Gaussian distribution *has the* density

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2}(det\Sigma)^{1/2}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu}, \Sigma^{-1}(\underline{x}-\underline{\mu}))}$$

(see also [**GS01**, Definition 4.9.4]*). In particular, if* $\sigma^2 > 0$*, then a Gaussian random variable* X *has the density*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

(for example, see [**GS01**, Example 4.4.4]*).*

Our next proposition links the vector $\underline{\mu}$ and the matrix Σ to the first two moments of the Gaussian distribution.

Proposition 3.2.11. *The* parameters *of the* Gaussian distribution *are* $\mu_j = \mathbf{E}(X_j)$ *and* $\Sigma_{jk} = \mathbf{E}[(X_j - \mu_j)(X_k - \mu_k)]$, $j, k = 1, \dots, n$ *(c.f.* [**GS01**, Theorem 4.9.5]*). Thus* μ *is the mean vector and* Σ *is the covariance matrix of* X*.*

As we now demonstrate, there is more to a Gaussian random vector than just having coordinates that are Gaussian random variables.

Exercise 3.2.12. *Let* X *be a Gaussian R.V. independent of* S*, with* $\mathbf{E}(X) = 0$ *and* $\mathbf{P}(S = 1) = \mathbf{P}(S = -1) = 1/2$*.*

(a) *Check that* SX *is Gaussian.*
(b) *Give an example of uncorrelated, zero-mean, Gaussian R.V.* X_1 *and* X_2 *such that the vector* $\underline{X} = (X_1, X_2)$ *is not Gaussian and where* X_1 *and* X_2 *are not independent.*

Exercise 3.2.13. *Suppose* (X, Y) *has a bi-variate Normal distribution (per Definition 3.2.8), with mean vector* $\underline{\mu} = (\mu_X, \mu_Y)$ *and the covariance matrix* $\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$*, with* σ_X*,* $\sigma_Y > 0$ *and* $|\rho| \le 1$*.*

(a) *Show that* (X, Y) *has the same law as* $(\mu_X + \sigma_X \rho U + \sigma_X \sqrt{1-\rho^2} V, \mu_Y + \sigma_Y U)$*, where* U *and* V *are independent Normal R.V.-s of mean zero and variance one. Explain why this implies that* $Z = X - (\rho\sigma_X/\sigma_Y)Y$ *is independent of* Y*.*
(b) *Explain why such* X *and* Y *are independent whenever they are uncorrelated (hence also whenever* $\mathbf{E}(X|Y) = \mathbf{E}X$*).*
(c) *Verify that* $\mathbf{E}(X|Y) = \mu_X + \frac{\rho\sigma_X}{\sigma_Y}(Y - \mu_Y)$*.*

Part (b) of Exercise 3.2.13 extends to any Gaussian random vector. That is,

Proposition 3.2.14. *If a Gaussian random vector* $\underline{X} = (X_1, \dots, X_n)$ *has uncorrelated coordinates, then its coordinates are also mutually independent.*

Proof. Since the coordinates X_k are uncorrelated, the corresponding matrix Σ has zero entries except at the main-diagonal $j = k$ (see Proposition 3.2.11). Hence, by Definition 3.2.8, the characteristic function $\Phi_{\underline{X}}(\underline{\theta})$ is of the form of $\prod_{k=1}^n \Phi_{X_k}(\theta_k)$. This in turn implies that the coordinates X_k of the random vector $\underline{X}$ are mutually independent (see Proposition 3.2.6). ∎

Definition 3.2.8 allows for Σ that is non-invertible, so for example the random variable $X = \mu$ a.s. is considered a Gaussian variable though it obviously does not have a density (hence does not fit Definition 3.2.9). The reason we make this choice is to have the collection of Gaussian distributions closed with respect to convergence in 2-mean, as we prove below to be the case.

Proposition 3.2.15. *Suppose a sequence of* n*-dimensional Gaussian random vectors* $\underline{X}^{(k)}$*,* $k = 1, 2, \dots$ *converges in* 2*-mean to an* n*-dimensional random vector* $\underline{X}$*, that is,* $\mathbf{E}[(X_i - X_i^{(k)})^2] \to 0$ *as* $k \to \infty$*, for* $i = 1, 2, \dots, n$*. Then,* $\underline{X}$ *is a Gaussian random vector, whose parameters* $\underline{\mu}$ *and* Σ *are the limits of the corresponding parameters* $\underline{\mu}^{(k)}$ *and* $\Sigma^{(k)}$ *of* $\underline{X}^{(k)}$*.*

Proof. We start by verifying the convergence of the parameters of $\underline{X}^{(k)}$ to those of $\underline{X}$. To this end, fixing $1 \le i, j \le n$ and applying the inequality

$$|(a + x)(b + y) - ab| \le |ay| + |bx| + |xy|$$

for $a = X_i$, $b = X_j$, $x = X_i^{(k)} - X_i$ and $y = X_j^{(k)} - X_j$, we get by monotonicity of the expectation that

$$\begin{aligned} \mathbf{E}|X_i^{(k)} X_j^{(k)} - X_i X_j| &\leq \mathbf{E}|X_i(X_j^{(k)} - X_j)| + \mathbf{E}|X_j(X_i^{(k)} - X_i)| \\ &+ \mathbf{E}|(X_i^{(k)} - X_i)(X_j^{(k)} - X_j)| \,. \end{aligned}$$

Thus, by the Schwarz inequality, c.f. Proposition 1.2.41, we see that

$$\begin{aligned} \mathbf{E}|X_i^{(k)} X_j^{(k)} - X_i X_j| &\leq \|X_i\|_2 \|X_j^{(k)} - X_j\|_2 + \|X_j\|_2 \|X_i^{(k)} - X_i\|_2 \\ &+ \|X_i^{(k)} - X_i\|_2 \|X_j^{(k)} - X_j\|_2 \,. \end{aligned}$$

So, the assumed convergence in 2-mean of $\underline{X}^{(k)}$ to $\underline{X}$ implies the convergence in 1-mean of $X_i^{(k)} X_j^{(k)}$ to $X_i X_j$ as $k \to \infty$. This in turn implies that $\mathbf{E}X_i^{(k)} X_j^{(k)} \to \mathbf{E}X_i X_j$ (c.f. Exercise 1.3.21). Further, the assumed convergence in 2-mean of $X_l^{(k)}$ to X_l (as $k \to \infty$) implies the convergence of $\mu_l^{(k)} = \mathbf{E}X_l^{(k)}$ to $\mu_l = \mathbf{E}X_l$, for $l = 1, 2, \ldots, n$, and hence also that

$$\Sigma_{ij}^{(k)} = \mathbf{E}X_i^{(k)} X_j^{(k)} - \mu_i^{(k)} \mu_j^{(k)} \to \mathbf{E}X_i X_j - \mu_i \mu_j = \Sigma_{ij} \,.$$

In conclusion, we established the convergence of the mean vectors $\underline{\mu}^{(k)}$ and the covariance matrices $\Sigma^{(k)}$ to the mean vector $\underline{\mu}$ and the covariance matrix Σ of $\underline{X}$, respectively.

Fixing $\underline{\theta} \in \mathbb{R}^n$ the assumed convergence in 2-mean of $\underline{X}^{(k)}$ to $\underline{X}$ also implies the convergence in 2-mean, and hence in probability, of $(\underline{\theta}, \underline{X}^{(k)})$ to $(\underline{\theta}, \underline{X})$. Hence, by bounded convergence, $\Phi_{\underline{X}^{(k)}}(\underline{\theta}) \to \Phi_{\underline{X}}(\underline{\theta})$ for each fixed $\underline{\theta} \in \mathbb{R}^n$ (see Corollary 1.4.28). Since $\Phi_{\underline{X}^{(k)}}(\underline{\theta}) = e^{-\frac{1}{2}(\underline{\theta}, \Sigma^{(k)} \underline{\theta})} e^{i(\underline{\theta}, \underline{\mu}^{(k)})}$ for each k, the convergence of the parameters $(\underline{\mu}^{(k)}, \Sigma^{(k)})$ implies that the function $\Phi_{\underline{X}}(\underline{\theta})$ must also be of such form. That is, necessarily $\underline{X}$ has a Gaussian distribution, whose parameters are the limits of the corresponding parameters of $\underline{X}^{(k)}$, as claimed. ∎

The next proposition provides an alternative to Definition 3.2.8.

Proposition 3.2.16. *A random vector $\underline{X}$ has the Gaussian distribution if and only if $(\sum_{i=1}^n a_{ji} X_i, j = 1, \ldots, m)$ is a Gaussian random vector for any non-random coefficients $a_{11}, a_{12}, \ldots, a_{mn} \in \mathbb{R}$ (c.f.* [**GS01**, Definition 4.9.7]*).*

It is usually much easier to check Definition 3.2.8 than to check the conclusion of Proposition 3.2.16. However, it is often very convenient to use the latter en-route to the derivation of some other property of $\underline{X}$.

We are finally ready to define the class of Gaussian stochastic processes.

Definition 3.2.17. *A stochastic process (S.P.) $\{X_t,\ t \in \mathcal{I}\}$ is* Gaussian *if for all $n < \infty$ and all $t_1, t_2, \cdots, t_n \in \mathcal{I}$, the random vector $(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$ has a Gaussian distribution, that is, all finite dimensional distributions of the process are Gaussian.*

To see that you understood well the definitions of Gaussian vectors and processes, convince yourself that the following corollary holds.

Corollary 3.2.18. *All distributional properties of Gaussian processes are determined by* the mean $\mu(t) = \mathbf{E}(X_t)$ *of the process and its* auto-covariance function $\rho(t, s) = \mathbf{E}[(X_t - \mu(t))\ (X_s - \mu(s))]$.

Applying Proposition 3.2.14 for the Gaussian vector $\underline{X} = (Y_{t_2} - Y_{t_1}, \dots, Y_{t_n} - Y_{t_{n-1}})$ of increments of a Gaussian stochastic process $\{Y_t\}$ (with arbitrary finite n and $0 \le t_1 < t_2 < \cdots < t_n$), we conclude from Definition 3.1.6 that

PROPOSITION 3.2.19. *If* $\mathrm{Cov}(Y_{t+h} - Y_t, Y_s) = 0$ *for a Gaussian stochastic process* $\{Y_t\}$*, all* $t \ge s$ *and* $h > 0$*, then the S.P.* $\{Y_t\}$ *has uncorrelated hence* independent increments *(which is thus also equivalent to* $\mathbf{E}(Y_{t+h} - Y_t | \sigma(Y_s, s \le t)) = \mathbf{E}(Y_{t+h} - Y_t)$ *for any* $t \ge 0$ *and* $h > 0$*).*

The special class of Gaussian processes plays a key role in our construction of the Brownian motion. When doing so, we shall use the following extension of Proposition 3.2.15.

PROPOSITION 3.2.20. *If the S.P.* $\{X_t,\ t \in \mathcal{I}\}$ *and the Gaussian S.P.* $\{X_t^{(k)},\ t \in \mathcal{I}\}$ *are such that* $\mathbf{E}[(X_t - X_t^{(k)})^2] \to 0$ *as* $k \to \infty$*, for each fixed* $t \in \mathcal{I}$*, then* X_t *is a Gaussian S.P. with mean and auto-covariance functions that are the pointwise limits of those for* $X_t^{(k)}$*.*

PROOF. Fixing $n < \infty$ and $t_1, t_2, \cdots, t_n \in \mathcal{I}$, apply Proposition 3.2.15 for the sequence of Gaussian random vectors $(X_{t_1}^{(k)}, X_{t_2}^{(k)}, \cdots, X_{t_n}^{(k)})$ to see that the distribution of $(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$ is Gaussian. Since this applies for all finite dimensional distributions of the S.P. $\{X_t,\ t \in \mathcal{I}\}$ we are done (see Definition 3.2.17). ∎

Here is the derivation of the C.L.T. statement of Example 1.4.13 and its extension towards a plausible construction of the Brownian motion.

EXERCISE 3.2.21. *Consider the random variables* $\widehat{S}_k$ *of Example 1.4.13.*

(a) *Applying Proposition 3.2.6 verify that the corresponding characteristic functions are*

$$\Phi_{\widehat{S}_k}(\theta) = [\cos(\theta/\sqrt{k})]^k .$$

(b) *Recalling that* $\delta^{-2}\log(\cos\delta) \to -0.5$ *as* $\delta \to 0$*, find the limit of* $\Phi_{\widehat{S}_k}(\theta)$ *as* $k \to \infty$ *while* $\theta \in \mathbb{R}$ *is fixed.*

(c) *Suppose random vectors* $\underline{X}^{(k)}$ *and* $\underline{X}$ *in* $\mathbb{R}^n$ *are such that* $\Phi_{\underline{X}^{(k)}}(\underline{\theta}) \to \Phi_{\underline{X}}(\underline{\theta})$ *as* $k \to \infty$*, for any fixed* $\underline{\theta}$*. It can be shown that then the laws of* $\underline{X}^{(k)}$*, as probability measures on* $\mathbb{R}^n$*, must converge weakly in the sense of Definition 1.4.20 to the law of* $\underline{X}$*. Explain how this fact allows you to verify the C.L.T. statement* $\widehat{S}_n \xrightarrow{\mathcal{L}} G$ *of Example 1.4.13.*

EXERCISE 3.2.22. *Consider the random vectors* $\underline{X}^{(k)} = (\frac{1}{\sqrt{k}} S_{k/2}, \frac{1}{\sqrt{k}} S_k)$ *in* $\mathbb{R}^2$*, where* $k = 2, 4, 6, \dots$ *is even, and* S_k *is the* simple random walk *of Definition 3.1.2, with* $\mathbf{P}(\xi_1 = -1) = \mathbf{P}(\xi_1 = 1) = 0.5$.

(a) *Verify that*

$$\Phi_{\underline{X}^{(k)}}(\underline{\theta}) = [\cos((\theta_1 + \theta_2)/\sqrt{k})]^{k/2}[\cos(\theta_2/\sqrt{k})]^{k/2} ,$$

where $\underline{\theta} = (\theta_1, \theta_2)$.

(b) *Find the mean vector* $\underline{\mu}$ *and the covariance matrix* Σ *of a Gaussian random vector* $\underline{X}$ *for which* $\Phi_{\underline{X}^{(k)}}(\underline{\theta})$ *converges to* $\Phi_{\underline{X}}(\underline{\theta})$ *as* $k \to \infty$.

(c) *Upon appropriately generalizing what you did in part (b), I claim that the* Brownian motion *of Theorem 3.1.3 must be a Gaussian stochastic process. Explain why, and guess what is the mean $\mu(t)$ and auto-covariance function $\rho(t,s)$ of this process (if needed take a look at Chapter 5).*

We conclude with a concrete example of a Gaussian stochastic process.

EXERCISE 3.2.23. *Let $Y_n = \sum_{k=1}^n \xi_k V_k$ for i.i.d. random variables $\{\xi_k\}$ such that $p = \mathbf{P}(\xi_k = 1) = 1 - \mathbf{P}(\xi_k = -1)$ and i.i.d. Gaussian random variables $\{V_k\}$ of zero mean and variance one that are independent of the collection $\{\xi_k\}$.*

(a) *Compute the mean $\mu(n)$ and auto-covariance function $\rho(\ell, n)$ for the discrete time stochastic process $\{Y_n\}$.*
(b) *Find the law of $\xi_1 V_1$, explain why $\{Y_n\}$ is a Gaussian process and provide the joint density $f_{Y_n,Y_{2n}}(x,y)$ of Y_n and Y_{2n}.*

3.2.3. Stationary processes. We conclude this section with a brief discussion of the important concept of stationarity, that is, invariance of the law of the process to translation of time.

DEFINITION 3.2.24. *A stochastic process $\{X_t\}$ indexed by $t \in \mathbb{R}$ is called* (strong sense) stationary *if its f.d.d. satisfy*

$$\begin{aligned} F_{t_1,t_2,\ldots,t_N}(\alpha_1,\alpha_2,\ldots,\alpha_N) &= \mathbf{P}(X_{t_1} \le \alpha_1,\ldots,X_{t_N} \le \alpha_N) \\ &= \mathbf{P}(X_{t_1+\tau} \le \alpha_1,\ldots,X_{t_N+\tau} \le \alpha_N) \\ &= F_{t_1+\tau,t_2+\tau,\ldots,t_N+\tau}(\alpha_1,\alpha_2,\ldots,\alpha_N)\,, \end{aligned}$$

for all $\tau \in \mathbb{R}$, $N < \infty$, $\alpha_i \in \mathbb{R}$, $i = 1,\ldots,N$ and any monotone $t_1 < t_2 \cdots < t_N \in \mathbb{R}$. A similar definition applies to discrete time S.P. indexed by t on the integers, just then t_i and τ take only integer values.

It is particularly easy to verify the stationarity of Gaussian S.P. since

PROPOSITION 3.2.25. *A Gaussian S.P. is stationary if and only if $\mu(t) = \mu$ (a constant) and $\rho(t,s) = r(|t-s|)$, where $r : \mathbb{R} \to \mathbb{R}$ is a function of the time difference $|t-s|$. (A stochastic process whose mean and auto-covariance function satisfy these two properties is called* weak sense (or covariance) stationary. *In general, a weak sense stationary process is not a strong sense stationary process (for example, see* [**GS01**, Example 8.2.5]*). However, as the current Proposition shows, the two notions of stationarity are equivalent in the Gaussian case.)*

Convince yourself that Proposition 3.2.25 is an immediate consequence of Corollary 3.2.18 (alternatively, use directly Proposition 3.2.11). For more on stationary Gaussian S.P. solve the following exercise and see [**GS01**, Section 9.6] (or [**Bre92**, Section 11.5] for the case of discrete time).

EXERCISE 3.2.26. *Suppose $\{X_t\}$ is a zero-mean, (weak sense) stationary process with auto-covariance function $r(t)$.*

(a) *Show that $|r(h)| \le r(0)$ for all $h > 0$.*
(b) *Show that if $r(h) = r(0)$ for some $h > 0$ then $X_{t+h} \overset{a.s.}{=} X_t$ for each t.*
(c) *Explain why part (c) of Exercise 3.2.13 implies that if $\{X_t\}$ is a zero-mean, stationary, Gaussian process with auto-covariance function $r(t)$ such that $r(0) > 0$, then $\mathbf{E}(X_{t+h}|X_t) = \frac{r(h)}{r(0)} X_t$ for any t and $h \ge 0$.*

(d) *Conclude that there is no zero-mean, stationary, Gaussian process of independent increments other than the trivial process* $X_t \equiv X_0$.

DEFINITION 3.2.27. *We say that a process* $\{X_t, t \geq 0\}$ *has* stationary increments *if* $X_{t+h} - X_t$ *and* $X_{s+h} - X_s$ *have the same law for all* $s, t, h, \geq 0$. *The same definition applies to discrete time S.P. indexed by* t *on the integers, just with* t, s *and* h *taking only integer values (and if the S.P. is indexed by a non-negative integer time, then so are the values of* s, t *and* h*).*

EXAMPLE 3.2.28. *Clearly, a sequence of independent and identically distributed random variables* $\ldots, X_{-2}, X_{-1}, X_0, X_1, \ldots$ *is a discrete time stationary process. However, many processes are not stationary. For example, the random walk* $S_n = \sum_{i=1}^{n} X_i$ *of Definition 3.1.2 is a non-stationary S.P. when* $\mathbf{E}X_1 = 0$ *and* $\mathbf{E}X_1^2 = 1$. *Indeed, if* $\{S_n\}$ *was a stationary process then the law of* S_n, *and in particular its second moment, would not depend on* n *– in contradiction with (3.1.1). Convince yourself that every stationary process has stationary increments, but note that the random walk* S_n *has stationary increments, thus demonstrating that stationary increments are not enough for stationarity.*

For more on stationary discrete time S.P. see [**Bre92**, Section 6.1], or see [**GS01**, Chapter 9] for the general case.

3.3. Sample path continuity

As we have seen in Section 3.1, the distribution of the S.P. does not specify uniquely the probability of events outside the rather restricted σ-field $\mathcal{F}_{\mathbf{X}}$ and in particular provides insufficient information about the behavior of its supremum, as well as about the continuity of its sample path.

Our goal in this section is thus to find relatively easy to check sufficient conditions for the existence of a modification of the S.P. that has a somewhat "nice" sample paths. The following definition of sample path continuity is the first step in this direction.

DEFINITION 3.3.1. *We say that* $\{X_t\}$ *has* continuous sample path *w.p.1 if* $\mathbf{P}(\{\omega : t \mapsto X_t(\omega) \text{ is continuous}\}) = 1$. *Similarly, we use the term* continuous modification *to denote a modification* $\{\widetilde{X}_t\}$ *of a given S.P.* $\{X_t\}$ *such that* $\{\widetilde{X}_t\}$ *has continuous sample path w.p.1.*

The next definition of *Hölder continuity* provides a quantitative refinement of this notion of continuity, by specifying the maximal possible smoothness of the sample path of X_t.

DEFINITION 3.3.2. *A S.P.* Y_t *is* locally Hölder continuous *with exponent* γ *if for some* $c < \infty$ *and a R.V.* $h(\omega) > 0$,

$$\mathbf{P}(\{\omega : \sup_{0 \leq s,t \leq T,\ |t-s| \leq h(\omega)} \frac{|Y_t(\omega) - Y_s(\omega)|}{|t-s|^{\gamma}} \leq c\}) = 1.$$

REMARK. The word "locally" in the above definition refers to the R.V. $h(\omega)$. When it holds for unrestricted $t, s \in [0, T]$ we say that Y_t is *globally (or uniformly) Hölder continuous* with exponent γ. A particular important special case is that of $\gamma = 1$, corresponding to *Lipschitz continuous* functions.

Equipped with Definition 3.3.2 our next theorem gives a very useful criterion for the existence of a continuous modification (and even yields a further "degree of smoothness" in terms of the Hölder continuity of the sample path).

THEOREM 3.3.3 (Kolmogorov's continuity theorem). *Given a S.P.* $\{X_t, t \in [0,T]\}$, *suppose there exist* $\alpha, \beta, c, h_0 > 0$ *such that*

$$\mathbf{E}(|X_{t+h} - X_t|^\alpha) \le ch^{1+\beta}, \qquad \text{for all} \quad 0 \le t,\ t+h \le T,\ \ 0 < h < h_0\,. \tag{3.3.1}$$

Then, there exists a continuous modification Y_t *of* X_t *such that* Y_t *is also locally Hölder continuous with exponent* γ *for any* $0 < \gamma < \beta/\alpha$.

REMARK. In case you have wondered why exponent γ near $(1+\beta)/\alpha$ does not work in Theorem 3.3.3, read its proof, and in particular, the derivation of [**KS97**, inequality (2.9), page 54]. Or, see [**Oks03**, Theorem 2.6, page 10], for a somewhat weaker result.

It is important to note that condition (3.3.1) of Theorem 3.3.3 involves only the joint distribution of (X_t, X_{t+h}) and as such is verifiable based on the f.d.d. of the process. In particular, either all versions of the given S.P. satisfy (3.3.1) or none of them does.

The following example demonstrates that we must have $\beta > 0$ in (3.3.1) to deduce the existence of a continuous modification.

EXAMPLE 3.3.4. *Consider the stochastic process* $X_t(\omega) = I_{\{U \le t\}}(\omega)$, $t \in [0,1]$, *where* U *is a Uniform*$[0,1]$ *random variable (that is,* $U(\omega) = \omega$ *on the probability space* $(U, \mathbb{R}, \mathcal{B})$ *of Example 1.1.11). Note that* $|X_{t+h} - X_t| = 1$ *if* $0 \le t < U \le t+h$, *and* $|X_{t+h} - X_t| = 0$ *otherwise. So,* $\mathbf{E}(|X_{t+h} - X_t|^\alpha) = U((t, t+h]) \le h$ *for any* $h > 0$ *and* $t \ge 0$. *That is,* $\{X_t, t \ge 0\}$ *satisfies (3.3.1) with* $c = 1$, $\beta = 0$ *any* α *and* h_0. *However, clearly the sample path of* $X_t(\omega)$ *is discontinuous at* $t = U(\omega)$ *whenever* $\omega \ne 0$. *That is, almost surely* $\{X_t\}$ *has discontinuous sample paths (and it is further possible to show that this S.P. has no continuous modification).*

The following application of Kolmogorov's continuity theorem demonstrates the importance of choosing wisely the free parameter α in this theorem.

EXERCISE 3.3.5. *Suppose the stochastic process* X_t *is such that* $\mathbf{E}(X_t) = 0$ *and* $\mathbf{E}(X_t^2) = 1$ *for all* $t \in [0,T]$.

(a) *Show that* $|\mathbf{E}(X_t X_{t+h})| \le 1$ *for any* $h > 0$ *and* $t \in [0, T-h]$.

(b) *Suppose that for some* $\lambda < \infty$, $p > 1$ *and* $h_0 > 0$,

$$\mathbf{E}(X_t X_{t+h}) \ge 1 - \lambda h^p \qquad \text{for all } \ 0 < h \le h_0\,. \tag{3.3.2}$$

Using Kolmogorov's continuity theorem show that then X_t *has a continuous modification.*

(c) *Suppose* X_t *is a Gaussian stochastic process such that* $\mathbf{E}(X_t) = 0$ *and* $\mathbf{E}(X_t^2) = 1$ *for all* $t \in [0,T]$. *Show that if* X_t *satisfies the inequality (3.3.2) for some* $\lambda < \infty$, $p > 0$ *and* $h_0 > 0$, *then for any* $0 < \gamma < p/2$, *the process* X_t *has a modification which is locally Hölder continuous with exponent* γ.

Hint: *see Section 5.1 for the moments of Gaussian random variable.*

As we show next, there exist non-Gaussian S.P.-s satisfying (3.3.2) with $p = 1$ for which there is no continuous modification.

EXAMPLE 3.3.6. *One such example is the* "random telegraph signal" R_t *which is defined as follows. Let* τ_i*,* $i = 1, 2, \ldots$ *be independent random times, each having the* Exponential(1) distribution*, that is,* $\mathbf{P}(\tau_i \leq x) = 1 - e^{-x}$ *for all* i *and* $x \geq 0$*. Starting at* $R_0 \in \{-1, 1\}$ *such that* $\mathbf{P}(R_0 = 1) = 1/2$*, the S.P.* R_t *alternately jumps between* -1 *and* $+1$ *at the random times* $s_k = \sum_{i=1}^{k} \tau_i$ *for* $k = 1, 2, 3, \ldots$*, so* $R_t \in \{-1, 1\}$ *keeps the same value in each of the intervals* (s_k, s_{k+1})*. Since almost surely* $s_1 < \infty$*, this S.P. does not have a continuous modification. However,*

$$\mathbf{E}(R_t R_{t+\varepsilon}) = \mathbf{P}(R_t = R_{t+\varepsilon}) - \mathbf{P}(R_t \neq R_{t+\varepsilon}) = 1 - 2\mathbf{P}(R_t \neq R_{t+\varepsilon}) \,,$$

and since for any $t \geq 0$ *and* $\varepsilon > 0$*,*

$$\varepsilon^{-1}\mathbf{P}(R_t \neq R_{t+\varepsilon}) \leq \varepsilon^{-1}\mathbf{P}(\tau_i \leq \varepsilon) = \varepsilon^{-1}(1 - e^{-\varepsilon}) \leq 1 \,,$$

we see that R_t *indeed satisfies (3.3.2) with* $p = 1$ *and* $\lambda = 2$*.*

The stochastic process R_t of Example 3.3.6 is a special instance of the continuous-time *Markov jump processes*, which we study in Section 6.3. Though the sample path of this process is almost never continuous, it has the right-continuity property of the following definition, as is the case for all continuous-time Markov jump processes of Section 6.3.

DEFINITION 3.3.7. *We say that a S.P.* X_t *has* right-continuous with left limits *(in short, RCLL) sample path, if for a.e.* $\omega \in \Omega$*, the sample path* $X_t(\omega)$ *is right-continuous and of left-limits at any* $t \geq 0$ *(that is, for* $h \downarrow 0$ *both* $X_{t+h}(\omega) \to X_t(\omega)$ *and the limit of* $X_{t-h}(\omega)$ *exists). Similarly, a modification having RCLL sample path with probability one is called* RCLL modification *of the S.P.*

REMARK. To practice your understanding, check that any S.P. having continuous sample path also has RCLL sample path (in particular, the Brownian motion of Section 5.1 is such). The latter property plays a major role in continuous-time martingale theory, as we shall see in Sections 4.2 and 4.3.2. For more on RCLL sample path see [**Bre92**, Section 14.2].

Perhaps you expect any two S.P.-s that are modifications of each other to have (a.s.) *indistinguishable* sample path, i.e. $\mathbf{P}(X_t = Y_t \text{ for all } t \in \mathcal{I}) = 1$. This is indeed what happens for discrete time, but in case of continuous time, in general such property may fail, though it holds when both S.P.-s have right-continuous sample paths (a.s.).

EXERCISE 3.3.8.

(a) *Let* $\{X_n\}$*,* $\{Y_n\}$ *be discrete time S.P.-s that are modifications of each other. Show that* $\mathbf{P}(X_n = Y_n$ *for all* $n \geq 0) = 1$*.*
(b) *Let* $\{X_t\}$*,* $\{Y_t\}$ *be continuous time S.P.-s that are modifications of each other. Suppose that both processes have right-continuous sample paths a.s. Show that* $\mathbf{P}(X_t = Y_t$ *for all* $t \geq 0) = 1$*.*
(c) *Provide an example of two S.P.-s which are modifications of one another but which are not indistinguishable.*

We conclude with a hierarchy of the sample path properties which have been considered here.

PROPOSITION 3.3.9. *The following implications apply for the sample path of any stochastic process:*

$$\text{Hölder continuity} \;\Rightarrow\; \text{Continuous w.p.1} \;\Rightarrow\; \text{RCLL} \,.$$

Proof. The stated relations between local Hölder continuity, continuity and right-continuity with left limits hold for any function $f : [0, \infty) \to \mathbb{R}$. Considering $f(t) = X_t(\omega)$ for a fixed $\omega \in \Omega$ leads to the corresponding relation for the sample path of the S.P. ∎

All the S.P. of interest to us shall have at least a RCLL modification, hence with all properties implied by it, and unless explicitly stated otherwise, hereafter we always assume that we are studying the RCLL modification of the S.P. in question.

The next theorem shows that objects like $Y_t = \int_0^t X_s ds$ are well defined under mild conditions and is crucial for the successful rigorous development of stochastic calculus.

Theorem 3.3.10 (Fubini's theorem). *If X_t has RCLL sample path and for some interval I and σ-field $\mathcal{H}$, almost surely $\int_I \mathbf{E}[|X_t|\,|\mathcal{H}]dt$ is finite then almost surely $\int_I X_t dt$ is finite and*

$$\int_I \mathbf{E}[X_t|\mathcal{H}]\ dt = \mathbf{E}\ [\int_I X_t\ dt|\mathcal{H}]\,.$$

Remark. Taking $\mathcal{H} = \{\emptyset, \Omega\}$ we get as a special case of Fubini's theorem that $\int_I \mathbf{E}X_t\ dt = \mathbf{E}\ (\int_I X_t\ dt)$ whenever $\int_I \mathbf{E}|X_t|dt$ is finite.

Here is a converse of Fubini's theorem, where the differentiability of the sample path $t \mapsto X_t$ implies the differentiability of the mean $t \mapsto \mathbf{E}X_t$ of a S.P.

Exercise 3.3.11. *Let $\{X_t, t \geq 0\}$ be a stochastic process such that for each $\omega \in \Omega$ the sample path $t \mapsto X_t(\omega)$ is differentiable at any $t \geq 0$.*

(a) *Verify that $\frac{\partial}{\partial t} X_t$ is a random variable for each fixed $t \geq 0$.*
(b) *Suppose further that there is an integrable random variable Y such that $|X_t - X_s| \leq |t - s|Y$ for almost every $\omega \in \Omega$ and all $t, s \geq 0$. Using the dominated convergence theorem, show that $t \mapsto \mathbf{E}X_t$ is then differentiable with a finite derivative such that for all $t \geq 0$,*

$$\frac{d}{dt}\mathbf{E}(X_t) = \mathbf{E}\Big(\frac{\partial}{\partial t} X_t\Big).$$

CHAPTER 4

Martingales and stopping times

In this chapter we study a collection of stochastic processes called martingales. Among them are some of the S.P. we already met, namely the random walk (in discrete time) and the Brownian motion (in continuous time). Many other S.P. found in applications are also martingales. We start in Section 4.1 with the simpler setting of discrete time martingales and filtrations (also called discrete parameter martingales and filtrations). The analogous theory of continuous time (or parameter) filtrations and martingales is introduced in Section 4.2, taking care also of sample path (right)-continuity. As we shall see in Section 4.3, martingales play a key role in computations involving stopping times. Martingales share many other "nice" properties, chiefly among which are tail bounds and convergence theorems. Section 4.4 deals with martingale representations and tail inequalities (for both discrete and continuous time). These lead to the various convergence theorems we present in Section 4.5. To demonstrate the power of martingale theory we also analyze in Section 4.6 the extinction probabilities of *branching processes*.

4.1. Discrete time martingales and filtrations

Subsection 4.1.1 introduces the concepts of filtration and martingale with some illustrating examples. As shown in Subsection 4.1.2, square-integrable martingales are analogous to zero-mean, orthogonal increments. The related super-martingales and sub-martingales are the subject of Subsection 4.1.3. For additional examples see [**GS01**, Sections 12.1, 12.2 and 12.8] or [**KT75**, Section 6.1] (for discrete time MGs), [**KT75**, Section 6.7] (for filtrations), and [**Ros95**, Section 7.3] (for applications to random walks).

4.1.1. Martingales and filtrations, definition and examples. A filtration represents any procedure of collecting more and more information as time goes on. Our starting point is the following rigorous mathematical definition of a (discrete time) filtration.

Definition 4.1.1. *A* filtration *is a non-decreasing family of of sub-σ-fields $\{\mathcal{F}_n\}$ of our measurable space $(\Omega, \mathcal{F})$. That is, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_n \cdots \subseteq \mathcal{F}$ and $\mathcal{F}_n$ is a σ-field for each n.*

Given a filtration, we are interested in S.P. such that for each n the information gathered by that time suffices for evaluating the value of the n-th element of the process. That is,

Definition 4.1.2. *A S.P. $\{X_n, n = 0, 1, \ldots\}$ is* adapted *to a filtration $\{\mathcal{F}_n\}$ if $\omega \mapsto X_n(\omega)$ is a R.V. on $(\Omega, \mathcal{F}_n)$ for each n, that is, if $\sigma(X_n) \subseteq \mathcal{F}_n$ for each n.*

At this point you should convince yourself that $\{X_n\}$ is adapted to the filtration $\{\mathcal{F}_n\}$ if and only if $\sigma(X_0, X_1, \ldots, X_n) \subseteq \mathcal{F}_n$ for all n. That is, DEFINITION 4.1.3. *The filtration* $\{\mathcal{G}_n\}$ *with* $\mathcal{G}_n = \sigma(X_0, X_1, \cdots, X_n)$ *is the minimal filtration with respect to which* $\{X_n\}$ *is adapted. We therefore call it the* canonical filtration *for the S.P.* $\{X_n\}$.

Whenever clear from the context what it means, we shall use the notation X_n both for the whole S.P. $\{X_n\}$ and for the n-th R.V. of this process, and likewise we may sometimes use $\mathcal{F}_n$ to denote the whole filtration $\{\mathcal{F}_n\}$.

A martingale consists of a filtration and an adapted S.P. which has the property of being a "fair game", that is, the expected future reward given current information is exactly the current value of the process. We now make this into a rigorous definition.

DEFINITION 4.1.4. *A* martingale *(denoted MG) is a pair* $(X_n, \mathcal{F}_n)$*, where* $\{\mathcal{F}_n\}$ *is a filtration and* X_n *an integrable (i.e.* $\mathbf{E}|X_n| < \infty$*), S.P. adapted to this filtration such that*

$$\mathbf{E}[X_{n+1}|\mathcal{F}_n] = X_n \qquad \forall n, \quad a.s. \tag{4.1.1}$$

The slower a filtration $n \mapsto \mathcal{F}_n$ grows, the easier it is for an adapted S.P. to be a martingale. That is, if $\mathcal{H}_n \subseteq \mathcal{F}_n$ for all n and S.P. $\{X_n\}$ adapted to filtration $\{\mathcal{H}_n\}$ is such that $(X_n, \mathcal{F}_n)$ is a martingale, then you can show using the tower property that $(X_n, \mathcal{H}_n)$ is also a martingale. To check your understanding of the preceding, and more generally, of what a martingale is, solve the next exercise.

EXERCISE 4.1.5. *Suppose* $(X_n, \mathcal{F}_n)$ *is a martingale. Show that then* $\{X_n\}$ *is also a martingale with respect to its canonical filtration and that a.s.* $\mathbf{E}[X_\ell|\mathcal{F}_n] = X_n$ *for all* $\ell > n$.

In view of Exercise 4.1.5, unless explicitly stated otherwise, when we say that $\{X_n\}$ is a MG we mean using the canonical filtration $\sigma(X_k, k \le n)$.

EXERCISE 4.1.6. *Provide an example of a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, a filtration* $\{\mathcal{F}_n\}$ *and a stochastic process* $\{X_n\}$ *adapted to* $\{\mathcal{F}_n\}$ *such that:*

(a) $\{X_n\}$ *is a martingale with respect to its canonical filtration but* $(X_n, \mathcal{F}_n)$ *is not a martingale.*
(b) *Provide a probability measure* $\mathbf{Q}$ *on* $(\Omega, \mathcal{F})$ *under which* $\{X_n\}$ *is not a martingale even with respect to its canonical filtration.*

Hint: *Go for a simple construction. For example,* $\Omega = \{a, b\}$, $\mathcal{F}_0 = \mathcal{F} = 2^\Omega$, $X_0 = 0$ *and* $X_n = X_1$ *for all* $n \ge 1$.

We next provide a convenient alternative characterization of the martingale property in terms of the *martingale differences* .

PROPOSITION 4.1.7. *If* $X_n = \sum_{i=0}^n D_i$ *then the canonical filtration for* $\{X_n\}$ *is the same as the canonical filtration for* $\{D_n\}$*. Further,* $(X_n, \mathcal{F}_n)$ *is a martingale if and only if* $\{D_n\}$ *is an integrable S.P., adapted to* $\{\mathcal{F}_n\}$*, such that* $\mathbf{E}(D_{n+1}|\mathcal{F}_n) = 0$ *a.s. for all* n.

PROOF. Since the transformation from $(X_0, \dots, X_n)$ to $(D_0, \dots, D_n)$ is continuous and invertible, it follows from Corollary 1.2.17 that $\sigma(X_k, k \le n) = \sigma(D_k, k \le n)$ for each n. By Definition 4.1.3 we see that $\{X_n\}$ is adapted to a filtration $\{\mathcal{F}_n\}$ if and only if $\{D_n\}$ is adapted to this filtration. It is very easy to show by induction on n that $\mathbf{E}|X_k| < \infty$ for $k = 0, \dots, n$ if and only if $\mathbf{E}|D_k| < \infty$ for

$k = 0, \ldots, n$. Hence, $\{X_n\}$ is an integrable S.P. if and only if $\{D_n\}$ is. Finally, with X_n measurable on $\mathcal{F}_n$ it follows from the linearity of the C.E. that

$$\mathbf{E}[X_{n+1}|\mathcal{F}_n] - X_n = \mathbf{E}[X_{n+1} - X_n|\mathcal{F}_n] = \mathbf{E}[D_{n+1}|\mathcal{F}_n]\,,$$

and the alternative expression for the martingale property follows from (4.1.1). ∎

In view of Proposition 4.1.7 we call $D_n = X_n - X_{n-1}$ for $n \geq 1$ and $D_0 = X_0$ the martingale differences associated with a martingale $\{X_n\}$. We now detail a few simple examples of martingales, starting with the random walk of Definition 3.1.2.

Example 4.1.8. *The random walk $S_n = \sum_{k=1}^n \xi_k$, with ξ_k independent, identically distributed, such that $\mathbf{E}|\xi_1| < \infty$ and $\mathbf{E}\xi_1 = 0$, is a MG (for its canonical filtration). More generally, $\{S_n\}$ is a MG even when the independent and integrable R.V. ξ_k of zero mean have non-identical distributions. Further, the canonical filtration may be replaced by the filtration $\sigma(\xi_1, \cdots, \xi_n)$. Indeed, this is just an application of Proposition 4.1.7 for the case where the differences $D_k = S_k - S_{k-1} = \xi_k$, $k \geq 1$ (and $D_0 = 0$), are independent, integrable and $\mathbf{E}[D_{n+1}|D_0, D_1, \ldots, D_n] = \mathbf{E}(D_{n+1}) = 0$ for all $n \geq 0$ by our assumption that $\mathbf{E}\xi_k = 0$ for all k. Alternatively, for a direct proof recall that $\mathbf{E}|S_n| \leq \sum_{k=1}^n \mathbf{E}|\xi_k| < \infty$ for all n, that is, the S.P. S_n is integrable. Moreover, since ξ_{n+1} is independent of $\sigma(S_1, \ldots, S_n)$ and $\mathbf{E}(\xi_{n+1}) = 0$, we have that*

$$\mathbf{E}[S_{n+1}|S_1, \cdots, S_n] = \mathbf{E}[S_n|S_1, \cdots, S_n] + \mathbf{E}[\xi_{n+1}|S_1, \cdots, S_n] = S_n + \mathbf{E}(\xi_{n+1}) = S_n\,,$$

implying that $\{S_n\}$ is a MG for its canonical filtration.

Exercise 4.1.9. *Let $\{\mathcal{F}_n\}$ be a filtration and X an integrable R.V. Define $Y_n = \mathbf{E}(X|\mathcal{F}_n)$ and show that $(Y_n, \mathcal{F}_n)$ is a martingale. How do you interpret Y_n?*

Our next example takes an arbitrary S.P. $\{V_n\}$ and creates a MG by considering $\sum_{k=1}^n V_k \xi_k$ for an appropriate auxiliary sequence $\{\xi_k\}$ of R.V.

Example 4.1.10. *Let $Y_n = \sum_{k=1}^n V_k \xi_k$, where $\{V_n\}$ is an arbitrary bounded S.P. and $\{\xi_n\}$ is a sequence of integrable R.V. such that for $n = 0, 1, \ldots$ both $\mathbf{E}(\xi_{n+1}) = 0$ and ξ_{n+1} is independent of $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n, V_1, \ldots, V_{n+1})$. Then, $\{Y_n\}$ is a MG for its canonical filtration and even for the possibly larger filtration $\{\mathcal{F}_n\}$. This is yet another application of Proposition 4.1.7, now with the differences $D_k = Y_k - Y_{k-1} = V_k \xi_k$, $k \geq 1$ (and $D_0 = 0$). Indeed, we assumed ξ_k are integrable and $|V_k| \leq C_k$ for some non-random finite constants C_k, resulting with $\mathbf{E}|D_k| \leq C_k \mathbf{E}|\xi_k| < \infty$, whereas*

$$\begin{aligned}
\mathbf{E}[D_{n+1}|\mathcal{F}_n] &= \mathbf{E}[V_{n+1}\xi_{n+1}|\mathcal{F}_n] = V_{n+1}\mathbf{E}[\xi_{n+1}|\mathcal{F}_n] \quad \text{(take out what is known)} \\
&= V_{n+1}\mathbf{E}[\xi_{n+1}] = 0 \quad \text{(zero mean } \xi_{n+1} \text{ is independent of } \mathcal{F}_n\text{)},
\end{aligned}$$

giving us the martingale property.

A special case of Example 4.1.10 is when the auxiliary sequence $\{\xi_k\}$ is independent of the given S.P. $\{V_n\}$ and consists of zero-mean, independent, identically distributed R.V. For example, random i.i.d. signs $\xi_k \in \{-1, 1\}$ (with $\mathbf{P}(\xi_k = 1) = \frac{1}{2}$) are commonly used in discrete mathematics applications (for other martingale applications c.f. [**AS00**, Chapter 7]).

Example 4.1.10 is a special case of the powerful martingale transform method. To explore this further, we first extract what was relevant in this example about the relation between the sequence $\{V_n\}$ and the filtration $\{\mathcal{F}_n\}$.

Definition 4.1.11. *We call a sequence* $\{V_n\}$ previsible *(or predictable) for the filtration* $\{\mathcal{F}_n\}$ *if* V_n *is measurable on* $\mathcal{F}_{n-1}$ *for all* $n \geq 1$.

The other relevant property of Example 4.1.10 is the fact that $X_n = \sum_{k=1}^n \xi_k$ is a MG for the filtration $\{\mathcal{F}_n\}$ (which is merely a re-run on Example 4.1.8, now with the excess but irrelevant information carried by the S.P. $\{V_n\}$). Having understood the two key properties of Example 4.1.10, we are ready for the more general martingale transform.

Theorem 4.1.12. *Let* $(X_n, \mathcal{F}_n)$ *be a MG and* $\{V_n\}$ *be a previsible sequence for the same filtration. The sequence of R.V.*

$$Y_n = \sum_{k=1}^{n} V_k(X_k - X_{k-1}) \ ,$$

called the martingale transform *of V. with respect to X., is then a MG with respect to the filtration* $\{\mathcal{F}_n\}$*, provided* $|V_n| \leq C_n$ *for some non-random constants* $C_n < \infty$*, or more generally* $\mathbf{E}|V_n|^q < \infty$ *and* $\mathbf{E}|X_n|^p < \infty$ *for all* n *and some* $1 \leq p, q < \infty$ *such that* $\frac{1}{q} + \frac{1}{p} = 1$.

Remark. The integrability conditions imposed in Theorem 4.1.12 ensure that $\mathbf{E}|V_k||X_k| < \infty$, hence that the MG transform $\{Y_n\}$ is an integrable S.P. Once this is established, $\{Y_n\}$ would be a MG, so we can state different versions of the theorem by further varying our integrability conditions.

Here is a direct example of the martingale transform method.

Example 4.1.13. *The S.P.* $Y_n = \sum_{k=1}^n X_{k-1}(X_k - X_{k-1})$ *is a MG whenever* $X_n \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ *is a MG (just note that the sequence* $V_n = X_{n-1}$ *is previsible for any filtration* $\{\mathcal{F}_n\}$ *with respect to which* $\{X_n\}$ *is adapted and take* $p = q = 2$ *in Theorem 4.1.12).*

A classical martingale is derived from the random products of Example 3.1.4.

Example 4.1.14. *Consider the integrable S.P.* $M_n = \prod_{i=1}^n Y_i$ *for strictly positive R.V.* Y_i*. By Corollary 1.2.17 its canonical filtration coincides with* $\mathcal{G}_n = \sigma(Y_1, \cdots, Y_n)$ *and since* $\mathbf{E}(M_{n+1}|\mathcal{G}_n) = \mathbf{E}(Y_{n+1}M_n|\mathcal{G}_n) = M_n\mathbf{E}(Y_{n+1}|\mathcal{G}_n)$*, the MG condition for* $\{M_n\}$ *is just* $\mathbf{E}(Y_{n+1}|Y_1, \cdots, Y_n) = 1$ *a.s. for all* n*. In the context of i.i.d.* Y_i *as in Example 3.1.4 we see that* $\{M_n\}$ *is a MG if and only if* $\mathbf{E}(Y_1) = 1$*, corresponding to "neutral return rate". Note that this is not the same as the condition* $\mathbf{E}(\log Y_1) = 0$ *under which the associated random walk* $S_n = \log M_n$ *is a MG.*

4.1.2. Orthogonal increments and square-integrable martingales. In case $\mathbf{E}(X_n^2) < \infty$ for all n, we have an alternative definition of being a MG, somewhat reminiscent of our definition of the conditional expectation.

Proposition 4.1.15. *A S.P.* $X_n \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ *adapted to the filtration* $\{\mathcal{F}_n\}$ *is a MG if and only if* $\mathbf{E}[(X_{n+1} - X_n)Z] = 0$ *for any* $Z \in L^2(\Omega, \mathcal{F}_n, \mathbf{P})$.

Proof. This follows from the definition of MG, such that $\mathbf{E}[X_{n+1}|\mathcal{F}_n] = X_n$, together with the definition of the C.E. via orthogonal projection on $L^2(\Omega, \mathcal{F}_n, \mathbf{P})$ as in Definition 2.1.3. ∎

A martingale $\{X_n\}$ such that $\mathbf{E}(X_n^2) < \infty$ for all n is called L^2-MG, or *square-integrable* MG. We gain more insight to the nature of such MGs by reformulating Proposition 4.1.15 in terms of the martingale differences $D_n = X_n - X_{n-1}$. To this end, we have the following definition.

DEFINITION 4.1.16. *We say that* $D_n \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ *is an* orthogonal sequence *of R.V. if* $\mathbf{E}[D_n h(D_0, D_1, \ldots, D_{n-1})] = \mathbf{E}[D_n]\mathbf{E}[h(D_0, \ldots, D_{n-1})]$ *for any* $n \geq 1$ *and every Borel function* $h : \mathbb{R}^n \to \mathbb{R}$ *such that* $\mathbf{E}[h(D_0, \ldots, D_{n-1})^2] < \infty$. *Alternatively, in view of Definition 2.1.3 and Proposition 1.2.14 this is merely the statement that* $\mathbf{E}[D_n | D_0, D_1, \ldots, D_{n-1}] = \mathbf{E}[D_n]$ *for all* n.

It is possible to extend Proposition 1.4.40 to the statement that the random variables $D_k \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ are mutually independent if and only if

$$\mathbf{E}[g(D_n)h(D_0, D_1, \ldots, D_{n-1})] = \mathbf{E}[g(D_n)]\mathbf{E}[h(D_0, \ldots, D_{n-1})]$$

for every n and any Borel functions $h : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ such that $\mathbf{E}[h(D_0, \ldots, D_{n-1})^2] < \infty$ and $\mathbf{E}[g(D_n)^2] < \infty$. In particular, independence implies orthogonality (where $g(\cdot)$ has to be linear), which in turn implies that the R.V. D_n are uncorrelated, that is, $\mathbf{E}(D_n D_i) = \mathbf{E}(D_n)\mathbf{E}(D_i)$ for all $0 \leq i < n < \infty$ (i.e. both $g(\cdot)$ and $h(\cdot)$ have to be linear functions).
As we show next, for square-integrable S.P. the martingale property amounts to having zero-mean, orthogonal differences. It is thus just the right compromise between the perhaps too restrictive requirement of having zero-mean, independent differences, and the ineffective property of just having zero-mean, uncorrelated differences.

PROPOSITION 4.1.17. *A S.P.* $X_n \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ *is a MG for its canonical filtration if and only if it has an orthogonal, zero-mean differences sequence* $D_n = X_n - X_{n-1}$, $n \geq 1$.

PROOF. In view of Definition 4.1.16, this is just a simple reformulation of Proposition 4.1.15. ∎

The MG property is especially simple to understand for a *Gaussian* S.P. As we have seen a necessary condition for the MG property is to have $\mathbf{E}D_n = 0$ and $\mathbf{E}(D_n D_i) = 0$ for all $0 \leq i < n$. With the Gaussian vector $\underline{D} = (D_0, \cdots, D_n)$ having uncorrelated coordinates, we know that the corresponding matrix Σ has zero entries except at the main-diagonal $j = k$ (see Proposition 3.2.11). Hence, by Definition 3.2.8, the characteristic function $\Phi_{\underline{D}}(\underline{\theta})$ is of the form of $\prod_{k=0}^{n} \Phi_{D_k}(\theta_k)$. This in turn implies that the coordinates D_k of this random vector are mutually independent (see Proposition 3.2.6). In conclusion, for a Gaussian S.P. having independent, orthogonal or uncorrelated differences are equivalent properties, which together with each of these differences having a zero mean is also equivalent to the MG property.

4.1.3. Sub-martingales and super-martingales. Often when operating on a MG, we naturally end up with a sub or super martingale, as defined below. Moreover, these processes share many of the properties of martingales, so it is useful to develop a unified theory for them.

DEFINITION 4.1.18. *A* sub-martingale *(denoted subMG) is an integrable S.P.* $\{X_n\}$, *adapted to the filtration* $\{\mathcal{F}_n\}$, *such that*

$$\mathbf{E}[X_{n+1} | \mathcal{F}_n] \geq X_n \quad \forall n, \quad a.s.$$

A super-martingale *(denoted supMG) is an integrable S.P. $\{X_n\}$, adapted to the filtration $\{\mathcal{F}_n\}$ such that*

$$\mathbf{E}[X_{n+1}|\mathcal{F}_n] \le X_n \quad \forall n, \quad a.s.$$

(A typical S.P. $\{X_n\}$ is neither a subMG nor a supMG, as the sign of the R.V. $\mathbf{E}[X_{n+1}|\mathcal{F}_n] - X_n$ may well be random, or possibly dependent upon the time index n).

Remark 4.1.19. Note that $\{X_n\}$ is a subMG if and only if $\{-X_n\}$ is a supMG. By this identity, all results about subMG-s have dual statements for supMG-s and vice verse. We often state only one out of each such pair of statements. Further, $\{X_n\}$ is a MG if and only if $\{X_n\}$ is both a subMG and a supMG. As a result, every statement holding for either subMG-s or supMG-s, also hold for MG-s.

Example 4.1.20. *In the context of Example 4.1.14, if $\mathbf{E}(Y_{n+1}|Y_1, \cdots, Y_n) \ge 1$ a.s. for all n then $\{M_n\}$ is a subMG, and if $\mathbf{E}(Y_{n+1}|Y_1, \cdots, Y_n) \le 1$ a.s. for all n then $\{M_n\}$ is a supMG. Such martingales appear for example in mathematical finance, where Y_i denotes the random proportional change in the value of a risky asset at the i-th trading round. So, positive conditional mean return rate yields a subMG while negative conditional mean return rate gives a supMG.*

Remark 4.1.21. If $\{X_n\}$ a subMG, then necessarily $n \mapsto \mathbf{E}X_n$ is non-decreasing, since by the tower property of (conditional) expectation

$$\mathbf{E}[X_n] = \mathbf{E}[\mathbf{E}[X_n|\mathcal{F}_{n-1}]] \ge \mathbf{E}[X_{n-1}]\,,$$

for all $n \ge 1$. Convince yourself that by Remark 4.1.19 this implies that if $\{X_n\}$ is a supMG then $n \mapsto \mathbf{E}X_n$ is non-increasing, hence for a MG $\{X_n\}$ we have that $\mathbf{E}(X_n) = \mathbf{E}(X_0)$ for all n.

We next detail a few examples in which subMG-s or supMG-s naturally appear, starting with an immediate consequence of Jensen's inequality

Exercise 4.1.22. *Suppose $(X_n, \mathcal{F}_n)$ is a martingale and $\Psi : \mathbb{R} \mapsto \mathbb{R}$ is a convex function such that $\mathbf{E}[|\Psi(X_n)|] < \infty$. Show that $(\Psi(X_n), \mathcal{F}_n)$ is a sub-martingale.*
Hint: *Use Proposition 2.3.9.*

Remark. Some examples of convex functions for which the above exercise is commonly applied are $\Psi(x) = |x|^p$, $p \ge 1$, $\Psi(x) = e^x$ and $\Psi(x) = x \log x$ (the latter only for $x > 0$). Taking instead a concave function $\Psi(\cdot)$ leads to a supMG, as for example when $\Psi(x) = x^p$, $p \in (0,1)$ or $\Psi(x) = \log x$, both restricted to $x > 0$.

Here is a concrete application of Exercise 4.1.22.

Exercise 4.1.23. *Let $\xi_1, \xi_2, \ldots$ be independent with $\mathbf{E}\xi_i = 0$ and $\mathbf{E}\xi_i^2 = \sigma_i^2$.*

(a) *Let $S_n = \sum_{i=1}^n \xi_i$ and $s_n^2 = \sum_{i=1}^n \sigma_i^2$. Show that $\{S_n^2\}$ is a sub-martingale and $\{S_n^2 - s_n^2\}$ is a martingale.*

(b) *Suppose also that $m_n = \prod_{i=1}^n \mathbf{E}(e^{\xi_i}) < \infty$. Show that $\{e^{S_n}\}$ is a sub-martingale and $M_n = e^{S_n}/m_n$ is a martingale.*

A special case of Exercise 4.1.23 is the random walk $S_n = \sum_{k=1}^n \xi_k$. Assuming in addition to $\mathbf{E}(\xi_1) = 0$ that also $\mathbf{E}\xi_1^2 < \infty$ we see that $(S_n^2, \sigma(\xi_1, \cdots, \xi_n))$ is a subMG. Further, $s_n^2 = \mathbf{E}(S_n^2) = n\mathbf{E}(\xi_1^2)$ (see (3.1.1)), hence $S_n^2 - n\mathbf{E}(\xi_1^2)$ is a MG. Likewise, e^{S_n} is a subMG for the same filtration whenever $\mathbf{E}(\xi_1) = 0$ and $\mathbf{E}(e^{\xi_1}) < \infty$. Though e^{S_n} is not a MG (unless $\xi_i = 0$ a.s.), the normalized $M_n = e^{S_n}/[\mathbf{E}(e^{\xi})]^n$ is merely the MG of Example 4.1.14 for the product of i.i.d. $Y_i = e^{\xi_i}/\mathbf{E}(e^{\xi})$.

4.2. Continuous time martingales and right continuous filtrations

In duality with Definitions 4.1.1 and 4.1.4 we define the continuous time filtrations and martingales as follows.

DEFINITION 4.2.1. *The pair $(X_t, \mathcal{F}_t)$, $t \geq 0$ real-valued, is called a* continuous time martingale *(in short MG), if:*

(a). The σ-fields $\mathcal{F}_t \subseteq \mathcal{F}$, $t \geq 0$ form a continuous time filtration*, that is, $\mathcal{F}_t \subseteq \mathcal{F}_{t+h}$, for all $t \geq 0$ and $h > 0$.*

(b). The continuous time S.P. $\{X_t\}$ is integrable and adapted *to this filtration. That is, $\mathbf{E}|X_t| < \infty$ and $\sigma(X_t) \subseteq \mathcal{F}_t$ for all $t \geq 0$.*

(c). For any fixed $t \geq 0$ and $h > 0$, the identity $\mathbf{E}(X_{t+h}|\mathcal{F}_t) = X_t$ holds a.s.

Similar definitions apply to S.P. $\{X_t\}$ indexed by $t \in I$ for an interval $I \subseteq \mathbb{R}$, just requiring also that in (a)-(c) both t and $t+h$ be in I.

• Replacing the equality in (c) of Definition 4.2.1 with $\geq$ or with $\leq$, defines the *continuous time subMG, or supMG*, respectively. These three groups of S.P. are related in the same manner as in the discrete time setting (c.f. Remark 4.1.19).

• Similar to Remark 4.1.21 we have that if $\{X_t\}$ is a subMG then $\mathbf{E}X_t \geq \mathbf{E}X_s$ for all $t \geq s$, and if $\{X_t\}$ is a supMG then $\mathbf{E}X_t \leq \mathbf{E}X_s$ for all $t \geq s$, so $\mathbf{E}X_t = \mathbf{E}X_0$ for all t when $\{X_t\}$ is a MG.

• Let $\sigma(X_s, 0 \leq s \leq t)$ denote the smallest σ-field containing $\sigma(X_s)$ for each $s \leq t$ (compare with $\mathcal{F}_{\mathbf{X}}$ of Section 3.1). In analogy with Definition 4.1.3 and Exercise 4.1.5, the *canonical filtration* for a continuous time S.P. $\{X_t\}$ is $\sigma(X_s, 0 \leq s \leq t)$. Further, as you check below, if $(X_t, \mathcal{F}_t)$ is a MG, then $(X_t, \sigma(X_s, 0 \leq s \leq t))$ is also a MG. Hence, as in the discrete time case, our default is to use the canonical filtration when studying MGs (or sub/supMGs).

EXERCISE 4.2.2. *Let $\mathcal{G}_t = \sigma(X_s, s \leq t)$. Show that*

(a) *If $(X_t, \mathcal{F}_t)$, $t \geq 0$ is a continuous time martingale for some filtration $\{\mathcal{F}_t\}$, then $(X_t, \mathcal{G}_t)$, $t \geq 0$ is also a martingale.*
(b) *If $(X_t, \mathcal{G}_t)$ is a continuous time sub-martingale and $\mathbf{E}(X_t) = \mathbf{E}(X_0)$ for all $t \geq 0$, then $(X_t, \mathcal{G}_t)$ is also a martingale.*

To practice your understanding, verify that the following identity holds for any square-integrable MG $(X_t, \mathcal{F}_t)$

$$\mathbf{E}[X_t^2|\mathcal{F}_s] - X_s^2 = \mathbf{E}[(X_t - X_s)^2|\mathcal{F}_s] \qquad \text{for any } \ t \geq s \geq 0 \tag{4.2.1}$$

and upon taking expectations deduce that for such MG the function $\mathbf{E}X_t^2$ is non-decreasing in t.

We have the following analog of Example 4.1.8, showing that in particular, the Brownian motion is a (continuous time) MG.

PROPOSITION 4.2.3. *Any integrable S.P. $\{M_t\}$ of independent increments (see Definition 3.1.6), and constant mean (i.e. $\mathbf{E}M_t = \mathbf{E}M_0$), is a MG.*

PROOF. Recall that a process M_t has independent increments if $M_{t+h} - M_t$ is independent of $\mathcal{G}_t = \sigma(M_s, 0 \leq s \leq t)$, for all $h > 0$ and $t \geq 0$. We assume that in addition to having independent increments, also $\mathbf{E}|M_t| < \infty$ and $\mathbf{E}M_t = \mathbf{E}M_0$

for all t. Then, $\mathbf{E}(M_{t+h} - M_t|\mathcal{G}_t) = \mathbf{E}(M_{t+h} - M_t) = 0$. Since M_t is measurable on $\mathcal{G}_t$, we deduce that $\mathbf{E}(M_{t+h}|\mathcal{G}_t) = M_t$, that is, $\{M_t\}$ is a MG (for its canonical filtration $\mathcal{G}_t$). ∎

Conversely, as you check below, any Gaussian martingale M_t is square-integrable and of independent increments, in which case $M_t^2 - \mathbf{E}M_t^2$ is also a martingale.

EXERCISE 4.2.4.

(a) *Deduce from the identity (4.2.1) that if the MG $\{M_t, t \geq 0\}$ of Proposition 4.2.3 is square-integrable, then $(M_t^2 - A_t, \mathcal{G}_t)$ is a MG for $\mathcal{G}_t = \sigma(M_s, s \leq t)$ and the non-random, non-decreasing function $A_t = \mathbf{E}M_t^2 - \mathbf{E}M_0^2$.*

(b) *Show that if a Gaussian S.P. $\{M_t\}$ is a MG, then it is square-integrable and of independent increments.*

(c) *Conclude that $(M_t^2 - A_t, \mathcal{G}_t)$ is a MG for $A_t = \mathbf{E}M_t^2 - \mathbf{E}M_0^2$ and any Gaussian MG $\{M_t\}$.*

Some of the many MGs associated with the Brownian motion are provided next (see [**KT75**, Section 7.5] for more).

EXERCISE 4.2.5. *Let $\mathcal{G}_t$ denote the canonical filtration of a Brownian motion W_t.*

(a) *Show that for any $\lambda \in \mathbb{R}$, the S.P. $M_t(\lambda) = \exp(\lambda W_t - \lambda^2 t/2)$, is a continuous time martingale with respect to $\mathcal{G}_t$.*

(b) *Explain why $\frac{d^k}{d\lambda^k} M_t(\lambda)$ are also martingales with respect to $\mathcal{G}_t$.*

(c) *Compute the first three derivatives in λ of $M_t(\lambda)$ at $\lambda = 0$ and deduce that the S.P. $W_t^2 - t$ and $W_t^3 - 3tW_t$ are also MGs.*

Our next example is of continuous time subMG-s and supMG-s that are obtained in the context of Example 3.1.4 and play a major role in modeling the prices of risky assets.

EXAMPLE 4.2.6. *For each $n \geq 1$ let $\{S_k^{(n)}\}$ denote the random walk corresponding to i.i.d. increments $\xi_i^{(n)}$ each having the $N(r/n, 1/n)$ law (i.e. $\xi_i^{(n)}$ is a Gaussian random variable of mean r/n and variance $1/n$). The risky asset value at time $t = k/n$ is $\exp(S_k^{(n)})$, modeling discrete trading rounds held each $1/n$ units of time with r denoting the mean return rate per unit time. So, consider the linear interpolation between the points $(k/n, \exp(S_k^{(n)}))$ in the plane, for $k = 0, 1, \ldots, n$, which in the limit $n \to \infty$ converge weakly to the continuous time stochastic process $Y_t = \exp(W_t + rt)$, called the* Geometric Brownian motion. *The Geometric Brownian motion $\{Y_t\}$ is a martingale for $r = -1/2$, a subMG for $r \geq -1/2$ and a supMG for $r \leq -1/2$ (compare with Example 3.1.4).*

As we explain next, each result derived for continuous time MGs implies the corresponding result for discrete time MGs.

EXAMPLE 4.2.7. *Any discrete time MG $(X_n, \mathcal{F}_n)$ is made into a continuous time MG $(X_t, \mathcal{F}_t)$ by the* interpolation *$\mathcal{F}_t = \mathcal{F}_n$ and $X_t = X_n$ for $n = 0, 1, 2, \ldots$ and all $t \in [n, n+1)$. Indeed, this interpolation keeps the integrable $\{X_t\}$ adapted to $\{\mathcal{F}_t\}$ and the latter a filtration, so it remains only to check the MG condition (c), that is, fixing $t \geq 0$ and $h > 0$,*

$$\mathbf{E}[X_{t+h}|\mathcal{F}_t] = \mathbf{E}[X_{t+h}|\mathcal{F}_\ell] = \mathbf{E}[X_m|\mathcal{F}_\ell] = X_\ell \ ,$$

where ℓ and $m \geq \ell$ are the integer parts of t and $t+h$, respectively, and the rightmost equality follows since $(X_n, \mathcal{F}_n)$ is a (discrete time) MG.

While we do not pursue this subject further here, observe that the continuous time analog of the martingale transform of Theorem 4.1.12 is the "stochastic integral"

$$Y_t = \int_0^t V_s \, dX_s \ ,$$

which for X_t a Brownian motion, is the main object of study in *Stochastic calculus* or *Stochastic differential equations* (to which the texts [**Oks03, KS97, Mik98**] are devoted). For example, the analog of Example 4.1.13 for the Brownian motion is $Y_t = \int_0^t W_s dW_s$, which for the appropriate definition of the stochastic integral (due to Ito), is merely the martingale $Y_t = \frac{1}{2}(W_t^2 - t)$ for the filtration $\mathcal{F}_t = \sigma(W_s, s \le t)$. Note that this Ito stochastic integral coincides with martingale theory at the price of deviating from the standard integration by parts formula. Indeed, the latter would apply if the Brownian sample path $t \mapsto W_t(\omega)$ was differentiable with probability one. As we see in Section 5.2, this is definitely not true, so the breakup of standard integration by parts should not come as a surprise.

Similarly to what happens for discrete-time (see Exercise 4.1.22), the subMG property of X_t is inherited by $\Psi(X_t)$ for convex, non-decreasing functions Ψ.

EXERCISE 4.2.8. *Suppose $(X_t, \mathcal{F}_t)$ is a MG, Ψ is convex and $\mathbf{E}|\Psi(X_t)| < \infty$. Using Jensen's inequality for the C.E. check that $(\Psi(X_t), \mathcal{F}_t)$ is a subMG. Moreover, same applies even when $(X_t, \mathcal{F}_t)$ is only a subMG, provided Ψ is also non-decreasing.*

Here are additional closure properties of subMGs, supMGs and MGs.

EXERCISE 4.2.9. *Suppose $(X_t, \mathcal{F}_t)$ and $(Y_t, \mathcal{F}_t)$ are subMGs and $t \mapsto f(t)$ is a non-decreasing, non-random function.*

(a) *Verify that $(X_t + Y_t, \mathcal{F}_t)$ is a subMG and hence so is $(X_t + f(t), \mathcal{F}_t)$.*
(b) *Rewrite this, first for supMGs X_t and Y_t, then in case of MGs.*

Building on Exercise 1.4.32 we next note that each positive martingale $(Z_t, \mathcal{F}_t)$ induces a collection of probability measures $\widetilde{\mathbf{P}}_t$ that are equivalent to the restrictions of $\mathbf{P}$ to $\mathcal{F}_t$ and satisfy a certain "martingale Bayes rule".

EXERCISE 4.2.10. *Given a positive MG $(Z_t, \mathcal{F}_t)$ with $\mathbf{E}Z_0 = 1$ consider for each $t \ge 0$ the probability measure $\widetilde{\mathbf{P}}_t : \mathcal{F}_t \to \mathbb{R}$ given by $\widetilde{\mathbf{P}}_t(A) = \mathbf{E}[Z_t I_A]$.*

(a) *Show that $\widetilde{\mathbf{P}}_t(A) = \widetilde{\mathbf{P}}_s(A)$ for any $A \in \mathcal{F}_s$ and $0 \le s \le t$.*
(b) *Fixing $0 \le u \le s \le t$ and $Y \in L^1(\Omega, \mathcal{F}_s, \widetilde{\mathbf{P}}_t)$, set $X_{s,u} = \mathbf{E}(Y Z_s | \mathcal{F}_u)/Z_u$. With $\widetilde{\mathbf{E}}_t$ denoting the expectation under $\widetilde{\mathbf{P}}_t$, deduce that $\widetilde{\mathbf{E}}_t(Y|\mathcal{F}_u) = X_{s,u}$ almost surely under $\widetilde{\mathbf{P}}_t$ (hence also under $\mathbf{P}$, by Exercise 1.4.32).*

Hint: *Show in part (b) that $\widetilde{\mathbf{E}}_t[Y I_A] = \mathbf{E}[Z_u X_{s,u} I_A] = \widetilde{\mathbf{E}}_t[X_{s,u} I_A]$ for any $A \in \mathcal{F}_u$.*

In the theory of continuous time S.P. it makes sense to require that each new piece of information has a definite first time of arrival. From a mathematical point of view, this is captured by the following concept of right-continuous filtration.

DEFINITION 4.2.11. *A filtration is called* right-continuous *if for any $t \ge 0$,*

$$\bigcap_{h>0} \mathcal{F}_{t+h} = \mathcal{F}_t \, .$$

(To avoid unnecessary technicalities we further assume the "usual conditions" of a complete probability space per Definition 1.3.4 with $N \in \mathcal{F}_0$ whenever $\mathbf{P}(N) = 0$).

For example, check that the filtration $\mathcal{F}_t = \mathcal{F}_n$ for $t \in [n, n+1)$, as in the interpolation of discrete-time MG-s (Example 4.2.7), is right-continuous. Like many other right-continuous filtrations, this filtration has jumps, that is sometimes $\sigma(\mathcal{F}_s, 0 \leq s < t)$ is a strict subset of $\mathcal{F}_t$ (i.e, for integer t), accounting for a new piece of information arriving at time t.

As we show next, not all filtrations are right-continuous and the continuity of the sample path of a S.P. does not guarantee that its canonical filtration is right-continuous.

EXAMPLE 4.2.12. *Consider the uniform probability measure on* $\Omega = \{-1, 1\}$ *and* $\mathcal{F} = 2^{\Omega}$. *The process* $X_t(\omega) = \omega t$ *clearly has continuous sample path. It is easy to see that its canonical filtration has* $\mathcal{G}_0 = \{\emptyset, \Omega\}$ *while* $\mathcal{G}_h = \mathcal{F}$ *for all* $h > 0$ *and is evidently not right-continuous at* $t = 0$.

As we next see, right continuity of the filtration translates into a RCLL sample path for any MG with respect to that filtration. We shall further see in Subsection 4.3.2 how right continuity of the filtration helps in the treatment of stopping times.

THEOREM 4.2.13. *If* $(X_t, \mathcal{F}_t)$ *is a MG with a right-continuous filtration* $\{\mathcal{F}_t\}$, *then* $\{X_t\}$ *has a RCLL modification as in Definition 3.3.7 (see* [**Bre92**, Theorem 14.7] *for a proof of a similar result).*

REMARK. You should check directly that the interpolated MGs of Example 4.2.7 have RCLL sample path with probability one. In view of Proposition 3.3.9, the same applies for the Brownian motion and any other MG with continuous sample path. Further, by Theorem 4.2.13, any MG with respect to the Brownian canonical filtration, or to the interpolated filtration of Example 4.2.7, has a RCLL modification. As we see in Sections 4.3–4.5 right continuity of the sample path allows us to deduce many important martingale properties, such as tail bounds, convergence results and optional stopping theorems.

4.3. Stopping times and the optional stopping theorem

This section is about stopping times and the relevance of martingales to their study, mostly via Doob's optional stopping theorem. The simpler concept of stopping time for a discrete filtration is considered first in Subsection 4.3.1, where we also provide a few illustrating examples of how Doob's theorem is used. For additional material and examples, see [**GS01**, Section 12.5], [**Ros95**, Section 6.2] or [**KT75**, Section 6.4]. The more delicate issue of stopping time for a continuous parameter filtration is dealt with in Subsection 4.3.2, with applications to hitting times for the Brownian motion given in Section 5.2.

4.3.1. Stopping times for discrete parameter filtrations. One of the advantages of MGs is in providing information about the law of stopping times, which we now define in the context of a discrete parameter filtration.

DEFINITION 4.3.1. *A random variable* τ *taking values in* $\{0, 1, \ldots, n, \ldots, \infty\}$ *is a* stopping time *for the filtration* $\{\mathcal{F}_n\}$ *if the event* $\{\tau \leq n\}$ *is in* $\mathcal{F}_n$ *for each finite* $n \geq 0$ *(c.f.* [**Bre92**, Section 5.6]*).*

Intuitively speaking a stopping time is such that the decision whether to stop or not by a given time is based on the information available at that time, and not on future, yet unavailable information. Some examples to practice your understanding are provided in the next exercises.

EXERCISE 4.3.2. *Show that* τ *is a stopping time for the discrete time filtration* $\{\mathcal{F}_n\}$ *if and only if* $\{\tau < t\} \in \mathcal{F}_t$ *for all* $t \geq 0$*, where* $\mathcal{F}_t = \mathcal{F}_{[t]}$ *is the interpolated filtration of Example 4.2.7.*

EXERCISE 4.3.3. *Suppose* $\theta, \tau, \tau_1, \tau_2, \ldots$ *are stopping times for the same (discrete time) filtration* $\mathcal{F}_n$*. Show that so are* $\min(\theta, \tau)$*,* $\max(\theta, \tau)$*,* $\theta + \tau$*,* $\sup_n \tau_n$ *and* $\inf_n \tau_n$*.*

EXERCISE 4.3.4. *Show that the* first hitting time $\tau(\omega) = \min\{k \geq 0 : X_k(\omega) \in B\}$ *of a Borel set* $B \subseteq \mathbb{R}$ *by a sequence* $\{X_k\}$*, is a stopping time for the canonical filtration* $\mathcal{F}_n = \sigma(X_k, k \leq n)$*. Provide an example where the* last hitting time $\theta = \sup\{k \geq 0 : X_k \in B\}$ *of a set* B *by the sequence, is not a stopping time (not surprising, since we need to know the whole sequence* $\{X_k\}$ *in order to verify that there are no visits to* B *after a given time* n*).*

As we see in Theorem 4.3.6, the MG, subMG and subMG properties are inherited by the stopped S.P. defined next.

DEFINITION 4.3.5. *Using the notation* $n \wedge \tau = \min(n, \tau(\omega))$*, the* stopped at τ stochastic process $\{X_{n\wedge\tau}\}$ *is given by*

$$X_{n\wedge\tau}(\omega) = \begin{cases} X_n(\omega), & n \leq \tau(\omega) \\ X_{\tau(\omega)}(\omega), & n > \tau(\omega) \end{cases}$$

THEOREM 4.3.6. *If* $(X_n, \mathcal{F}_n)$ *is a subMG (or supMG or a MG) and* τ *is a stopping time for* $\{\mathcal{F}_n\}$*, then* $(X_{n\wedge\tau}, \mathcal{F}_n)$ *is also a subMG, or supMG or MG, respectively (for proof see* [**Bre92**, Proposition 5.26]*).*

COROLLARY 4.3.7. *If* $(X_n, \mathcal{F}_n)$ *is a subMG and* τ *is a stopping time for* $\{\mathcal{F}_n\}$*, then* $\mathbf{E}(X_{n\wedge\tau}) \geq \mathbf{E}(X_0)$ *for all* n*. If in addition* $(X_n, \mathcal{F}_n)$ *is a MG, then* $\mathbf{E}(X_{n\wedge\tau}) = \mathbf{E}(X_0)$*.*

The main result of this subsection is the following theorem, where the uniform integrability of Subsection 1.4.2 comes handy (see also [**GS01**, Theorem 12.5.1] for a similar result).

THEOREM 4.3.8 (Doob's optional stopping). *If* $(X_n, \mathcal{F}_n)$ *is a subMG and* $\tau < \infty$ *a.s. is a stopping time for the filtration* $\{\mathcal{F}_n\}$ *such that the sequence* $\{X_{n\wedge\tau}\}$ *is uniformly integrable, then* $\mathbf{E}(X_\tau) \geq \mathbf{E}(X_0)$*. If in addition* $(X_n, \mathcal{F}_n)$ *is a MG, then* $\mathbf{E}(X_\tau) = \mathbf{E}(X_0)$*.*

PROOF. Note that $X_{n\wedge\tau(\omega)}(\omega) \to X_{\tau(\omega)}(\omega)$ as $n \to \infty$, whenever $\tau(\omega) < \infty$. Since $\tau < \infty$ a.s. we deduce that $X_{n\wedge\tau} \stackrel{a.s}{\to} X_\tau$, hence also $X_{n\wedge\tau} \to_p X_\tau$ (see part (b) of Theorem 1.3.6). With $\{X_{n\wedge\tau}\}$ uniformly integrable, combining Theorem 1.4.23 with the preceding corollary, we obtain that in case of a subMG $(X_n, \mathcal{F}_n)$,

$$\mathbf{E}[X_\tau] = \mathbf{E}[\lim_{n\to\infty} X_{n\wedge\tau}] = \lim_{n\to\infty} \mathbf{E}[X_{n\wedge\tau}] \geq \mathbf{E}(X_0)\,,$$

as stated in the theorem. The same argument shows that $\mathbf{E}(X_\tau) \leq \mathbf{E}(X_0)$ in case of a supMG $(X_n, \mathcal{F}_n)$, hence $\mathbf{E}(X_\tau) = \mathbf{E}(X_0)$ for a MG $(X_n, \mathcal{F}_n)$. ∎

There are many examples in which MGs are applied to provide information about specific stopping times. We detail below one such example, pertaining to the *symmetric simple random walk*.

EXAMPLE 4.3.9. *Consider the simple random walk (SRW)* $S_n = \sum_{k=1}^n \xi_k$, *with* $\xi_k \in \{-1, 1\}$ *independent and identically distributed R.V. such that* $\mathbf{P}(\xi_k = 1) = 1/2$, *so* $\mathbf{E}\xi_k = 0$. *Fixing positive integers a and b, let* $\tau_{a,b} = \inf\{n \geq 0 : S_n \geq b$, *or* $S_n \leq -a\}$, *that is,* $\tau_{a,b}$ *is the first time the SRW exits the interval* $(-a, b)$. *Recall that* $\{S_n\}$ *is a MG and the first hitting time* $\tau_{a,b}$ *is a stopping time. Moreover,* $|S_{n\wedge\tau_{a,b}}| \leq \max(a, b)$ *is a bounded sequence, hence uniformly integrable. Further,* $\mathbf{P}(\tau_{a,b} > k) \leq \mathbf{P}(-a < S_k < b) \to 0$ *as* $k \to \infty$, *hence* $\tau_{a,b} < \infty$ *almost surely. We also have that* $S_0 = 0$, *so by Doob's optional stopping theorem we know that* $\mathbf{E}(S_{\tau_{a,b}}) = 0$. *Note that using only increments* $\xi_k \in \{-1, 1\}$, *necessarily* $S_{\tau_{a,b}} \in \{-a, b\}$. *Consequently,*

$$\mathbf{E}(S_{\tau_{a,b}}) = -a\mathbf{P}(S_{\tau_{a,b}} = -a) + b\mathbf{P}(S_{\tau_{a,b}} = b) = 0\,.$$

Since $\mathbf{P}(S_{\tau_{a,b}} = -a) = 1 - \mathbf{P}(S_{\tau_{a,b}} = b)$ *we have two equations with two unknowns, which we easily solve to get that* $\mathbf{P}(S_{\tau_{a,b}} = -a) = \frac{b}{b+a}$ *is the probability that the SRW hits* $-a$ *before hitting* $+b$. *This probability is sometimes called the gambler's ruin, for a gambler with initial capital of* $+a$, *betting on the outcome of independent fair games, a unit amount per game, gaining (or losing)* ξ_k *and stopping when either all his capital is lost (ruin), or his accumulated gains reach the amount* $+b$ *(c.f.* [**GS01**, Example 1.7.4]*).*

REMARK. You should always take the time to verify the conditions of Doob's optional stopping theorem before using it. To see what might happen otherwise, consider Example 4.3.9 without specifying a lower limit. That is, let $\tau_b = \inf\{n : S_n \geq b\}$. Though we shall not show it, $\mathbf{P}(\tau_b > n) = \mathbf{P}(\max_{k\leq n} S_k < b) \to 0$ as $n \to \infty$. Consequently, with probability one, $\tau_b < \infty$ in which case clearly $S_{\tau_b} = b$. Hence, $\mathbf{E}(S_{\tau_b}) = b \neq \mathbf{E}(S_0) = 0$. This shows that you cannot apply Doob's optional stopping theorem here and that the sequence of random variables $\{S_{n\wedge\tau_b}\}$ is not uniformly integrable.

There is no general recipe of how to go about finding the appropriate martingale for each example, making this some sort of an "art" work. We illustrate this in the context of the *asymmetric simple random walk*.

EXAMPLE 4.3.10. *Consider the stopping times* $\tau_{a,b}$ *for the SRW* $\{S_n\}$ *of Example 4.3.9, now with* $p := \mathbf{P}(\xi_k = 1) \neq \frac{1}{2}$ *(excluding also the trivial cases of* $p = 0$ *and* $p = 1$*). We wish to compute the ruin probability* $r = \mathbf{P}(S_{\tau_{a,b}} = -a)$ *in this case. Note that* $\{S_n\}$ *is no longer a MG. However,* $M_n = \prod_{i=1}^n Y_i$ *is a MG for* $Y_i = e^{\theta\xi_i}$ *provided the constant* $\theta \neq 0$ *is such that* $\mathbf{E}[e^{\theta\xi}] = 1$ *(see Example 4.1.14). Further,* $M_n = e^{\theta S_n}$ *is non-negative with* $M_{n\wedge\tau_{a,b}} \leq \exp(|\theta||S_{n\wedge\tau_{a,b}}|)$ *bounded hence uniformly integrable (see Example 4.3.9). Thus, applying Doob's optional stopping theorem to* M_n *at* $\tau_{a,b}$ *we have that*

$$1 = \mathbf{E}(M_0) = \mathbf{E}(M_{\tau_{a,b}}) = \mathbf{E}(e^{\theta S_{\tau_{a,b}}}) = re^{-\theta a} + (1 - r)e^{\theta b}\,,$$

giving an explicit formula for r *in terms of* a, b *and* θ. *To complete the solution check that* $\mathbf{E}[e^{\theta\xi}] = pe^{\theta} + (1 - p)e^{-\theta} = 1$ *for* $\theta = \log[(1 - p)/p]$, *allowing you to find* r *in terms of* a, b *and* p. *Note that* $\theta > 0$ *if* $0 < p < \frac{1}{2}$, *that is,* $\mathbf{E}(\xi_k) = 2p - 1 < 0$, *while* $\theta < 0$ *if* $\frac{1}{2} < p < 1$ *(i.e.* $\mathbf{E}(\xi_k) > 0$*).*

We could have alternatively tried to use the perhaps more natural martingale $X_n = S_n - (2p - 1)n$. *Once we show that* $\tau_{a,b}$ *has a finite expectation, it follows that so* *does* $\sup_n |X_{n\wedge\tau_{a,b}}|$*. Thus,* $X_{n\wedge\tau_{a,b}}$ *is a U.I. sequence (see Example 1.4.24), and by Doob's optional stopping for* X_n *at* $\tau_{a,b}$ *we have that*

$$0 = \mathbf{E}(X_0) = \mathbf{E}(X_{\tau_{a,b}}) = \mathbf{E}[S_{\tau_{a,b}} - (2p-1)\tau_{a,b}] .$$

Hence, by linearity of the expectation, we see that $(2p-1)\mathbf{E}(\tau_{a,b}) = \mathbf{E}(S_{\tau_{a,b}}) = -ra + (1-r)b$*. Having only one equation in two unknowns,* r *and* $\mathbf{E}(\tau_{a,b})$*, we cannot find* r *directly this way.*

However, having found r *already using the martingale* M_n*, we can deduce now an explicit formula for* $\mathbf{E}(\tau_{a,b})$*.*

To practice your understanding, find another martingale that allows you to explicitly compute $\mathbf{E}(\tau_{a,b})$ for the symmetric (i.e. $p = 1/2$), SRW of Example 4.3.9.

4.3.2. Stopping times for continuous parameter filtrations. We start with the definition of stopping times for a continuous parameter filtration $\{\mathcal{F}_t\}$ in a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, analogous to Definition 4.3.1.

Definition 4.3.11. *A non-negative random variable* $\tau(\omega)$ *is called* stopping time *with respect to the continuous time filtration* $\{\mathcal{F}_t\}$ *if* $\{\omega : \tau(\omega) \le t\} \in \mathcal{F}_t$ *for all* $t \ge 0$*.*

Exercise 4.3.12.

(a) *Show that* τ *is a stopping time for the right-continuous filtration* $\{\mathcal{F}_t\}$ *if and only if* $\{\tau < t\} \in \mathcal{F}_t$ *for all* $t \ge 0$*.*
(b) *Redo Exercise 4.3.3 in case* $\theta, \tau, \tau_1, \tau_2, \ldots$ *are stopping times for a right-continuous filtration* $\{\mathcal{F}_t\}$*.*

Deterministic times are clearly also stopping times (as then $\{\omega : \tau(\omega) \le t\}$ is either Ω or $\emptyset$, both of which are in any σ-field). First hitting times of the type of the next proposition are another common example of stopping times.

Proposition 4.3.13. *If right-continuous S.P.* $\{X_t\}$ *is adapted to the filtration* $\{\mathcal{F}_t\}$ *then* $\tau_B(\omega) = \inf\{t \ge 0 : X_t(\omega) \in B\}$ *is a stopping time for* $\mathcal{F}_t$ *when either:*

(a) B *is an open set and* $\mathcal{F}_t$ *is a right continuous filtration;*
(b) B *is a closed set and the sample path* $t \mapsto X_t(\omega)$ *is continuous for all* $\omega \in \Omega$*.*

Proof. (omit at first reading) If $\mathcal{F}_t$ is a right continuous filtration then by Exercise 4.3.12 it suffices to show that $A_{t-} := \{\omega : \tau_B < t\} \in \mathcal{F}_t$. By definition of τ_B we have that A_{t-} is the union of the sets $\Gamma_s(B) := \{\omega : X_s(\omega) \in B\}$ over all $s < t$. Further, $\Gamma_s(B) \in \mathcal{F}_t$ for $s \le t$ (since $\{X_t\}$ is adapted to $\mathcal{F}_t$). This is however an *uncountable union* and indeed in general A_{t-} may not be in $\mathcal{F}_t$. Nevertheless, when $u \mapsto X_u$ is right continuous and B is an open set, it follows by elementary real analysis that the latter union equals the countable union of $\Gamma_s(B)$ over all rational values of $s < t$. Consequently, in this case A_{t-} is in $\mathcal{F}_t$, completing the proof of part (a).

Without right continuity of $\mathcal{F}_t$ it is no longer enough to consider A_{t-}. Instead, we should show that $A_t := \{\omega : \tau_B \le t\} \in \mathcal{F}_t$. Unfortunately, the identity $A_t = \bigcup_{s\le t} \Gamma_s(B)$ is not true in general (and having B open is not of much help). However, by definition of τ_B we have that $\tau_B(\omega) \le t$ if and only if there exist non-increasing $t_n(\omega) \le t + 1/n$ such that $X_{t_n(\omega)}(\omega) \in B$ for all n. Clearly, $t_n \downarrow s$ for some $s = s(\omega) \le t$. With $u \mapsto X_u(\omega)$ right continuous, it follows that $X_s(\omega)$ is the limit as $n \to \infty$ of $X_{t_n(\omega)}(\omega) \in B$ so our assumption that B is closed leads to $X_s(\omega) \in B$. Thus, under these two assumptions we have the identity $A_t = \bigcup_{s \le t} \Gamma_s(B)$. With B no longer an open set, the countable union of $\Gamma_s(B)$ over $\mathcal{Q}_t = \{s : s = t$ or $s \in \mathcal{Q}$ and $s < t\}$ is in general a *strict subset* of the union of $\Gamma_s(B)$ over all $s \le t$. Fortunately, for a closed set B and $u \mapsto X_u$ *continuous* we have by a real analysis argument (which we do not provide here), that

$$\bigcup_{s \le t} \Gamma_s(B) = \bigcap_{k=1}^{\infty} \bigcup_{s \in \mathcal{Q}_t} \Gamma_s(B_k)$$

(for the open sets $B_k = \bigcup_{y \in B}(y - k^{-1}, y + k^{-1})$), thus concluding that $A_t \in \mathcal{F}_t$ and completing the proof of part (b). ∎

Beware that unlike discrete time, the first hitting time τ_B might *not be* a stopping time for a Borel set B which is not closed, even when considering the canonical filtration $\mathcal{G}_t$ of a process X_t with continuous sample path.

EXAMPLE 4.3.14. *Indeed, consider the open set $B = (0, \infty)$ and the S.P. $X_t(\omega) = \omega t$ of Example 4.2.12 of continuous sample path and canonical filtration that is not right continuous. It is easy to check that in this case $\tau_B(1) = 0$ while $\tau_B(-1) = \infty$. As the event $\{\omega : \tau_B(\omega) \le 0\} = \{1\}$ is not in $\mathcal{G}_0 = \{\emptyset, \Omega\}$, we see that in this case τ_B is not a stopping time for $\mathcal{G}_t$.*

Stochastic processes are typically defined only up to a set of zero probability, and the continuity with probability one of the sample path $t \mapsto X_t(\omega)$ is not enough to assure that τ_B of Proposition 4.3.13 is a stopping time for $\sigma(X_s, s \le t)$ whenever B is closed. This problem is easily fixed by modifying the value of such S.P. on a set of zero probability, so as to make its sample path continuous for *all* $\omega \in \Omega$.

EXERCISE 4.3.15. *Let $\mathcal{G}_t$ denote the canonical filtration of the S.P. $\{X_t\}$.*

(a) *Verify that $\mathcal{G}_{t^+} = \bigcap_{h>0} \mathcal{G}_{t+h}$ is a right-continuous filtration.*
(b) *Considering part (a) of Proposition 4.3.13 for filtration $\mathcal{G}_{t^+}$, deduce that for any fixed $b > 0$ and $\delta > 0$ the random variable $\tau_b^{(\delta)} = \inf\{t \ge \delta : X_{t-\delta} > b\}$ is a stopping time for $\{\mathcal{G}_t\}$, provided that $\{X_t\}$ has right-continuous sample path.*
(c) *With $Y_t = \int_0^t X_s^2 ds$ use part (b) of Proposition 4.3.13 to show that $\theta_1 = \inf\{t \ge 0 : Y_t = b\}$ is another stopping time for $\{\mathcal{G}_t\}$. Then explain why $\theta_2 = \inf\{t \ge 0 : Y_{2t} = b\}$, is in general not a stopping time for this filtration.*

We next extend to the continuous parameter setting the notion of stopped submartingale (as in Theorem 4.3.6), focusing on processes with right-continuous sample path.

THEOREM 4.3.16. *If τ is a stopping time for the filtration $\{\mathcal{F}_t\}$ and the S.P. $\{X_t\}$ of right-continuous sample path is a subMG (or supMG or a MG) for $\{\mathcal{F}_t\}$, then $X_{t \wedge \tau} = X_{t \wedge \tau(\omega)}(\omega)$ is also a subMG (or supMG or MG, respectively), for this filtration.*

Equipped with Theorem 4.3.16 we extend also Doob's optional stopping theorem to the continuous parameter setting (compare with Theorem 4.3.8).

THEOREM 4.3.17 (Doob's optional stopping). *If $(X_t, \mathcal{F}_t)$ is a subMG with right-continuous sample path and $\tau < \infty$ a.s. is a stopping time for the filtration $\{\mathcal{F}_t\}$ such that $\{X_{t\wedge\tau}\}$ is U.I. then $\mathbf{E}(X_\tau) \geq \mathbf{E}(X_0)$. If in addition $(X_t, \mathcal{F}_t)$ is a MG, then $\mathbf{E}(X_\tau) = \mathbf{E}(X_0)$.*

Here are three representative applications of Doob's optional stopping theorem in the context of first hitting times (for the Brownian motion, the Geometric Brownian motion and a Brownian motion with drift). When solving these and any other exercise with first hitting times, assume that the Brownian motion W_t has been modified on a set of zero probability so that $t \mapsto W_t(\omega)$ is continuous for *all* $\omega \in \Omega$.

EXERCISE 4.3.18. *Let W_t be a Brownian motion. Fixing $a > 0$ and $b > 0$ let $\tau_{a,b} = \inf\{t \geq 0 : W_t \notin (-a, b)\}$. We will see in Section 5.2 that $\tau_{a,b}$ is finite with probability one.*

(a) *Check that $\tau_{a,b}$ is a stopping time and that $W_{t\wedge\tau_{a,b}}$ is uniformly integrable.*
(b) *Applying Doob's optional stopping theorem for this stopped martingale, compute the probability that W_t reaches level b before it reaches level $-a$.*
(c) *Justify using the optional stopping theorem for $\tau_{b,b}$ and the martingales $M_t(\lambda)$ of Exercise 4.2.5. Deduce from it the value of $\mathbf{E}(e^{-\theta\tau_{b,b}})$ for $\theta > 0$.*

Hint: *In part (c) you may use the fact that the S.P. $\{-W_t\}$ has the law as $\{W_t\}$.*

EXERCISE 4.3.19. *Consider the Geometric Brownian motion $Y_t = e^{W_t}$. Fixing $b > 1$, compute the Laplace transform $\mathbf{E}(e^{-\theta\bar{\tau}_b})$ of the stopping time $\bar{\tau}_b = \inf\{t > 0 : Y_t = b\}$ at arbitrary $\theta \geq 0$ (see* [**KT75**, page 364] *for a related option pricing application).*

EXERCISE 4.3.20. *Consider $M_t = \exp(\lambda Z_t)$ for non-random constants λ and r, where $Z_t = W_t + rt$, $t \geq 0$, and W_t is a Brownian motion.*

(a) *Compute the conditional expectation $\mathbf{E}(M_{t+h}|\mathcal{G}_t)$ for $\mathcal{G}_t = \sigma(Z_u, u \leq t)$ and $t, h \geq 0$.*
(b) *Find the value of $\lambda \neq 0$ for which $(M_t, \mathcal{G}_t)$ is a martingale.*
(c) *Fixing $a, b > 0$, apply Doob's optional stopping theorem to find the law of $Z_{\tau_{a,b}}$ for $\tau_{a,b} = \inf\{t \geq 0 : Z_t \notin (-a, b)\}$.*

An important concept associated with each stopping time is the *stopped σ-field* defined next (see also [**Bre92**, Definition 12.41] or [**KS97**, Definition 1.2.12] or [**GS01**, Problem 12.4.7]).

DEFINITION 4.3.21. *The* stopped σ-field *$\mathcal{F}_\tau$ associated with the stopping time τ for a filtration $\{\mathcal{F}_t\}$ is the collection of events $A \in \mathcal{F}$ such that $A \cap \{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$ for each $t \geq 0$.*

EXERCISE 4.3.22. *Suppose θ and τ are stopping times for a filtration $\{\mathcal{F}_t\}$.*

(a) *Verify that $\mathcal{F}_\tau$ is a σ-field and that τ is $\mathcal{F}_\tau$-measurable.*
(b) *Show that if $A \in \mathcal{F}_\theta$ then $A \cap \{\omega : \theta(\omega) \leq \tau(\omega)\} \in \mathcal{F}_\tau$ and deduce that $\mathcal{F}_\theta \subseteq \mathcal{F}_\tau$ when $\theta(\omega) \leq \tau(\omega)$ for all $\omega \in \Omega$.*

You should interpret $\mathcal{F}_\tau$ as quantifying the information given upon stopping at τ. Our next example illustrates which events are in $\mathcal{F}_\tau$.

EXAMPLE 4.3.23. *For the measurable space $(\Omega_2, \mathcal{F}_2)$ corresponding to two coin tosses, consider the stopping time τ such that $\tau = 1$ if the first coin shows H and $\tau = 2$ otherwise. Convince yourself that the corresponding stopped σ-field is $\mathcal{F}_\tau = \sigma(\{HT, HH\}, \{TH\}, \{TT\})$.*

Here is a more sophisticated example to the same effect.

EXAMPLE 4.3.24. *Suppose the S.P. $\{X_t : 0 \le t \le T\}$ has continuous sample path, in which case $\tau_b := \inf\{t \ge 0 : X_t = b\}$ is by Proposition 4.3.13 a stopping time with respect to the canonical filtration $\mathcal{G}_t$ of $\{X_t\}$, for any non-random $b \in \mathbb{R}$. Denote the corresponding stopped σ-field by $\mathcal{G}_{\tau_b}$. Fixing non-random $c \in \mathbb{R}$ consider the event*

$$A_c = \{\omega : \sup_{s\in[0,T]} X_s(\omega) > c\} .$$

Since the sample path $s \mapsto X_s(\omega)$ is continuous, $\omega \in A_c$ if and only if $X_s(\omega) > c$ for some s in the countable set $\mathcal{Q}_T$ of all rational numbers in $[0,T]$. Consequently, the event $A_c = \bigcup_{s\in\mathcal{Q}_T} \{\omega : X_s(\omega) > c\}$ is in $\mathcal{G}_T$. However, in general $A_c \notin \mathcal{G}_t$ for $t < T$ since if $X_s(\omega) \le c$ for all $s \le t$, then we are not sure whether $\omega \in A_c$ or not, till we observe $X_s(\omega)$ also for $s \in (t,T]$. Nevertheless, when $c < b$, if $\tau_b(\omega) \le t$ for some non-random $t < T$ then clearly $\omega \in A_c$. Hence, for $c < b$ and any $t < T$,

$$A_c \bigcap \{\omega : \tau_b(\omega) \le t\} = \{\omega : \tau_b(\omega) \le t\} \in \mathcal{G}_t$$

(recall that τ_b is a stopping time for $\mathcal{G}_t$). Since this applies for all t, we deduce in view of Definition 4.3.21 that $A_c \in \mathcal{G}_{\tau_b}$ for each $c < b$. In contrast, when $c \ge b$, for stochastic processes X_s of "sufficiently generic" sample path the event $\{\tau_b \le t\}$ does not tell us enough about A_c to conclude that $A_c \cap \{\tau_b \le t\} \in \mathcal{G}_t$ for all t. Consequently, for such stochastic processes $A_c \notin \mathcal{G}_{\tau_b}$ when $c \ge b$.

4.4. Martingale representations and inequalities

In Subsection 4.4.1 we show that martingales are at the core of all adapted processes. We further explore there the structure of certain sub-martingales, giving rise to fundamental objects such as the innovation process and the increasing process associated with square-integrable martingales. This is augmented in Subsection 4.4.2 by the study of maximal inequalities for sub-martingales (and martingales). Such inequalities are a key technical tool in many applications of probability theory.

4.4.1. Martingale decompositions.

To demonstrate the relevance of martingales to the study of many S.P., we start with a representation of any adapted, integrable, discrete-time S.P. as the sum of a martingale and a previsible process.

THEOREM 4.4.1 (Doob's decomposition). *Given an integrable S.P. $\{X_n\}$, adapted to a discrete parameter filtration $\{\mathcal{F}_n\}$, $n \ge 0$, there exists a decomposition $X_n = Y_n + A_n$ such that $(Y_n, \mathcal{F}_n)$ is a MG and $\{A_n\}$ is a previsible S.P. This decomposition is unique up to the value of Y_0, a R.V. measurable on $\mathcal{F}_0$.*

PROOF. Let $A_0 = 0$ and for all $n \ge 1$,

$$A_n = A_{n-1} + \mathbf{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}).$$

By definition of the conditional expectation (C.E.) we see that $A_k - A_{k-1}$ is measurable on $\mathcal{F}_{k-1}$ for any $k \ge 1$. Since $\mathcal{F}_{k-1} \subseteq \mathcal{F}_{n-1}$ for all $k \le n$ and $A_n = A_0 + \sum_{k=1}^n (A_k - A_{k-1})$, it follows that $\{A_n\}$ is previsible for the filtration $\{\mathcal{F}_n\}$. We next check that $Y_n = X_n - A_n$ is a MG. To this end, recall that $\{X_n\}$ integrable implies that so is $\{X_n - X_{n-1}\}$ whereas the C.E. only reduces the L^1 norm (see Corollary 2.3.11). Therefore, $\mathbf{E}|A_n - A_{n-1}| \le \mathbf{E}|X_n - X_{n-1}| < \infty$. So, A_n is integrable, as is X_n, implying (by the triangle inequality for the L^1 norm) that Y_n is integrable as well. With $\{X_n\}$ adapted and $\{A_n\}$ previsible (hence adapted), we see that $\{Y_n\}$ is also adapted. It remains to check that almost surely, $\mathbf{E}(Y_n|\mathcal{F}_{n-1}) - Y_{n-1} = 0$ for all $n \geq 1$. Indeed, for any $n \geq 1$,

$$\begin{aligned}
&\mathbf{E}(Y_n|\mathcal{F}_{n-1}) - Y_{n-1} = \mathbf{E}(X_n - A_n|\mathcal{F}_{n-1}) - (X_{n-1} - A_{n-1}) \\
&= \mathbf{E}[X_n - X_{n-1} - (A_n - A_{n-1})|\mathcal{F}_{n-1}] \quad \text{(since } Y_{n-1} \text{ is measurable on } \mathcal{F}_{n-1}\text{)} \\
&= \mathbf{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] - (A_n - A_{n-1}) \quad \text{(since } A_n \text{ is previsible)} \\
&= 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{(by the definition of } A_n\text{)}.
\end{aligned}$$

We finish the proof by checking the stated uniqueness of the decomposition. To this end, suppose that we have two such decompositions, $X_n = Y_n + A_n = \widetilde{Y}_n + \widetilde{A}_n$. Then, $\widetilde{Y}_n - Y_n = A_n - \widetilde{A}_n$. Since $\{A_n\}$ and $\{\widetilde{A}_n\}$ are previsible for the filtration $\{\mathcal{F}_n\}$,

$$\begin{aligned}
A_n - \widetilde{A}_n &= \mathbf{E}(A_n - \widetilde{A}_n|\mathcal{F}_{n-1}) = \mathbf{E}(\widetilde{Y}_n - Y_n|\mathcal{F}_{n-1}) \\
&= \widetilde{Y}_{n-1} - Y_{n-1} \quad \text{(since } \widetilde{Y}_n \text{ and } Y_n \text{ are MG-s)} \\
&= A_{n-1} - \widetilde{A}_{n-1}.
\end{aligned}$$

We thus conclude that $A_n - \widetilde{A}_n$ is independent of n. If $Y_0 = \widetilde{Y}_0$ we deduce further that $A_n - \widetilde{A}_n = A_0 - \widetilde{A}_0 = \widetilde{Y}_0 - Y_0 = 0$ for all n. In conclusion, as soon as we determine Y_0, a R.V. measurable on $\mathcal{F}_0$, both sequences $\{A_n\}$ and $\{Y_n\}$ are uniquely determined. ∎

Definition 4.4.2. *When using Doob's decomposition for the canonical filtration $\sigma(X_k, k \leq n)$, the MG $\{Y_n\}$ is called the* innovation process *associated with $\{X_n\}$. The reason for this name is that $X_{n+1} = (A_{n+1} + Y_n) + (Y_{n+1} - Y_n)$, where $A_{n+1} + Y_n$ is measurable on $\sigma(X_k, k \leq n)$ while $Y_{n+1} - Y_n$ describes the "new" part of X_{n+1}. Indeed, assuming $X_n \in L^2$ we have that $Y_{n+1} - Y_n$ is orthogonal to $\sigma(X_k, k \leq n)$ in the sense of (2.1.2) (see Proposition 4.1.17). The innovation process is widely used in prediction, estimation and control of time series, where the fact that it is a MG is very handy.*

As we see next, Doob's decomposition is very attractive when $(X_n, \mathcal{F}_n)$ is a subMG.

Exercise 4.4.3. *Check that the previsible part of Doob's decomposition of a submartingale $(X_n, \mathcal{F}_n)$ is a non-decreasing sequence, that is, $A_n \leq A_{n+1}$ for all n. What can you say about the previsible part of Doob's decomposition of a supermartingale?*

We next illustrate Doob's decomposition for two classical subMGs.

Example 4.4.4. *Consider the subMG $\{S_n^2\}$ for the random walk $S_n = \sum_{k=1}^n \xi_k$, where ξ_k are independent and identically distributed R.V. such that $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}\xi_1^2 = 1$. We already saw that $Y_n = S_n^2 - n$ is a MG. Since Doob's decomposition, $S_n^2 = Y_n + A_n$ is unique, in this special case the non-decreasing previsible part of the decomposition is $A_n = n$.*

In Example 4.4.4 we have a non-random A_n. However, for most subMGs the corresponding non-decreasing A_n is a random sequence, as is the case in our next two examples.

EXAMPLE 4.4.5. *Consider the subMG* $(M_n, \mathcal{G}_n)$ *where* $M_n = \prod_{i=1}^n Z_i$ *for i.i.d.* $Z_i > 0$ *such that* $\mathbf{E}Z_1 > 1$ *and* $\mathcal{G}_n = \sigma(Z_i : i \leq n)$ *(see Example 4.1.14). The non-decreasing previsible part of its Doob's decomposition is such that for* $n \geq 1$

$$\begin{aligned} A_{n+1} - A_n &= \mathbf{E}[M_{n+1} - M_n|\mathcal{G}_n] = \mathbf{E}[Z_{n+1}M_n - M_n|\mathcal{G}_n] \\ &= M_n\mathbf{E}[Z_{n+1} - 1|\mathcal{G}_n] = M_n(\mathbf{E}Z_1 - 1) \end{aligned}$$

(since Z_{n+1} *is independent of* $\mathcal{G}_n$*). In this case* $A_n = (\mathbf{E}Z_1 - 1)\sum_{k=1}^{n-1} M_k + A_1$*, where we are free to choose for* A_1 *any non-random constant. We see that* A_n *is a random sequence (assuming the R.V.* Z_i *are not a.s. constant).*

EXERCISE 4.4.6. *Consider the stochastic process* $Z_t = W_t + rt$*,* $t \geq 0$*, with* W_t *a Brownian motion and* r *a non-random constant. Compute the previsible part* A_n *in Doob's decomposition of* $X_n = Z_n^2$*,* $n = 0, 1, 2, \ldots$ *with respect to the discrete time filtration* $\mathcal{F}_n = \sigma(Z_k, k = 0, \ldots, n)$*, starting at* $A_0 = 0$*.*

Doob's decomposition is particularly useful in connection with square-integrable martingales $\{X_n\}$, where one can relate the limit of X_n as $n \to \infty$ with that of the non-decreasing sequence A_n in the decomposition of $\{X_n^2\}$.

The continuous-parameter analog of Doob's decomposition is a fundamental ingredient of stochastic integration. Here we provide this decomposition for a special but very important class of subMG associated with *square-integrable martingales* (see also [**KS97**, page 30]).

THEOREM 4.4.7 (Doob-Meyer decomposition). *Suppose* $\{\mathcal{F}_t\}$ *is a right-continuous filtration and the martingale* $(M_t, \mathcal{F}_t)$ *of continuous sample path is such that* $\mathbf{E}M_t^2 < \infty$ *for each* $t \geq 0$*. Then, there exists a unique (integrable) S.P.* $\{A_t\}$ *such that*
(a). $A_0 = 0$*.*
(b). $\{A_t\}$ *has continuous sample path.*
(c). $\{A_t\}$ *is adapted to* $\{\mathcal{F}_t\}$*.*
(d). $t \mapsto A_t$ *is non-decreasing.*
(e). $(M_t^2 - A_t, \mathcal{F}_t)$ *is a MG.*

REMARK. This is merely the decomposition of the subMG $X_t = M_t^2$, where property (a) resolves the issue of uniqueness of the R.V. A_0 measurable on $\mathcal{F}_0$, property (b) specifies the smoothness of the sample-path of the continuous-time S.P. $\{A_t\}$ and property (d) is in analogy with the monotonicity you saw already in Exercise 4.4.3.

DEFINITION 4.4.8. *The S.P.* $\{A_t\}$ *in the Doob-Meyer decomposition (of* $\{M_t^2\}$*) is called the* increasing part *or the* increasing process *associated with the MG* $(M_t, \mathcal{F}_t)$*.*

EXAMPLE 4.4.9. *Starting with the Brownian motion* $\{W_t\}$ *(which a martingale), the Doob-Meyer decomposition of* W_t^2 *gives the increasing part* $A_t = t$ *and the martingale* $W_t^2 - t$ *(compare to Example 4.4.4 of the random walk). Indeed, recall Exercise 4.2.4 that the increasing part is non-random for any Gaussian martingale. Also, in Section 5.3 we show that the increasing process associated with* W_t *coincides with its* quadratic variation*, as is the case for all square-integrable MGs of continuous sample path.*

EXERCISE 4.4.10. *Find a non-random* $f(t)$ *such that* $X_t = e^{W_t - f(t)}$ *is a martingale, and for this value of* $f(t)$ *find the increasing process associated with the martingale* X_t *via the Doob-Meyer decomposition.*

Hint: *Try an increasing process* $A_t = \int_0^t e^{2W_s - h(s)} ds$ *and use Fubini's theorem to find the non-random* $h(s)$ *for which* $M_t = X_t^2 - A_t$ *is a martingale with respect to the filtration* $\mathcal{G}_t = \sigma(W_s, s \le t)$.

EXERCISE 4.4.11. *Suppose a stopping time* τ *for a right-continuous filtration* $\{\mathcal{F}_t\}$ *and the increasing part* A_t *for a square-integrable martingale* $(M_t, \mathcal{F}_t)$ *of continuous sample path are such that* $A_\tau = 0$.

(a) *Applying Theorem 4.3.16 for* $X_t = M_t$ *and* $X_t = M_t^2 - A_t$ *deduce that both* $(M_{t\wedge\tau}, \mathcal{F}_t)$ *and* $(M^2_{t\wedge\tau}, \mathcal{F}_t)$ *are martingales.*
(b) *Applying (4.2.1) verify that* $\mathbf{E}[(M_{t\wedge\tau} - M_0)^2 | \mathcal{F}_0] = 0$, *and deduce from it that* $\mathbf{P}(M_{t\wedge\tau} = M_0) = 1$ *for any* $t \ge 0$.
(c) *Explain why the assumed continuity of* $t \mapsto M_t$ *implies that* $\mathbf{P}(M_{t\wedge\tau} = M_0$ *for all* $t \ge 0) = 1$.

REMARK. Taking non-random $\tau \to \infty$ we conclude from the preceding exercise that if a a square-integrable martingale of continuous sample path has a zero increasing part, then it is almost surely constant (in time).

EXERCISE 4.4.12. *Suppose* $\{\mathcal{F}_t\}$ *is a right-continuous filtration and* $(X_t, \mathcal{F}_t)$ *and* $(Y_t, \mathcal{F}_t)$ *are square-integrable martingales of continuous sample path.*

(a) *Verify that* $(X_t + Y_t, \mathcal{F}_t)$ *and* $(X_t - Y_t, \mathcal{F}_t)$ *are then square-integrable martingales of continuous sample path.*
(b) *Let* $Z_t = (A_t^{X+Y} - A_t^{X-Y})/4$, *where* $A_t^{X\pm Y}$ *denote the increasing part of the MG* $(X_t \pm Y_t, \mathcal{F}_t)$. *Show that* $(X_t Y_t - Z_t, \mathcal{F}_t)$ *is a martingale of continuous sample path and verify that for all* $0 \le s < t$,

$$\mathbf{E}[(X_t - X_s)(Y_t - Y_s)|\mathcal{F}_s] = \mathbf{E}[(X_t Y_t - X_s Y_s)|\mathcal{F}_s] = \mathbf{E}[Z_t - Z_s|\mathcal{F}_s].$$

REMARK. The S.P. $\{Z_t\}$ of the preceding exercise is called the *cross variation* for the martingales $(X_t, \mathcal{F}_t)$ and $(Y_t, \mathcal{F}_t)$, with two such martingales considered orthogonal if and only if their cross variation is zero (c.f. [**KS97**, Definition 1.5.5]). Note that the cross variation for a martingale $(M_t, \mathcal{F}_t)$ and itself is merely its quadratic variation (i.e. the associated increasing process A_t).

4.4.2. Maximal inequalities for martingales. Sub-martingales (and super-martingales) are rather tame stochastic processes. In particular, as we see next, the tail of $\max X_n$ over $1 \le n \le N$, is bounded by moments of X_N. This is a major improvement over Markov's inequality, relating the typically much smaller tail of the R.V. X_N to its moments (see Example 1.2.39).

THEOREM 4.4.13 (Doob's inequality).
(a). Suppose $\{X_n\}$ *is a subMG. Then, for all* $x > 0$ *and* $N < \infty$,

$$\mathbf{P}(\max_{0\le n\le N} X_n > x) \le x^{-1}\mathbf{E}|X_N|. \tag{4.4.1}$$

(b). Suppose $\{X_n, n \le \infty\}$ *is a subMG (that is,* $\mathbf{E}(X_m|X_k, k \le \ell) \ge X_\ell$ *for all* $0 \le \ell < m \le \infty$*). Then, for all* $x > 0$,

$$\mathbf{P}(\sup_{0\le n<\infty} X_n > x) \le x^{-1}\mathbf{E}|X_\infty|. \tag{4.4.2}$$

(c). Suppose $\{X_t\}$, $t \in [0, T]$ *is a continuous-parameter, right continuous subMG (that is, each sample path* $t \mapsto X_t(\omega)$ *is right continuous). Then, for all* $x > 0$,

$$\mathbf{P}(\sup_{0\le t\le T} X_t > x) \le x^{-1}\mathbf{E}|X_T|. \tag{4.4.3}$$

PROOF. (omit at first reading) Following [**GS01**, Theorem 12.6.1], we prove (4.4.1) by considering the stopping time $\tau_x = \min\{n \ge 0 : X_n > x\}$. Indeed,

$$x\mathbf{P}(\max_{0\le n\le N} X_n > x) = x\mathbf{P}(\tau_x \le N) \le \mathbf{E}(X_{\tau_x} I_{\{\tau_x \le N\}}) .$$

Since $S = \min(N, \tau_x) \le N$ is a bounded stopping time, by another version of Doob's optional stopping (not detailed in these notes), we have that

$$\mathbf{E}(X_{\tau_x} I_{\{\tau_x \le N\}}) + \mathbf{E}(X_N I_{\{\tau_x > N\}}) = \mathbf{E}(X_{\min(N,\tau_x)}) \le \mathbf{E}(X_N) .$$

Noticing that,

$$\mathbf{E}(X_N) - \mathbf{E}(X_N I_{\{\tau_x > N\}}) = \mathbf{E}(X_N I_{\{\tau_x \le N\}}) \le \mathbf{E}|X_N| ,$$

we thus get the inequality (4.4.1).

The assumptions of part (b) result with the sequence $(X_0, X_1, \ldots, X_{N-1}, X_\infty)$ being also a subMG, hence by part (a) we have that for any $N < \infty$,

$$\mathbf{P}(\max_{0\le n\le N-1} X_n > x) \le x^{-1}\mathbf{E}|X_\infty|.$$

The events $A_N = \{\omega : \max_{n\le N} X_n(\omega) > x\}$ monotonically increase in N to $A_\infty = \{\omega : \max_{n<\infty} X_n(\omega) > x\}$. Therefore, we get the inequality (4.4.2) by the continuity of each probability measure under such an operation (see the remark following Definition 1.1.2).

Fixing an integer N, let I_N denote the (finite) set of times in $[0, T]$ that are given as a ratio of two integers from $\{0, 1, \ldots, N\}$. For a given continuous time subMG X_t, $t \in [0, T]$, consider the discrete-time subMG X_{t_n}, where t_n is the n-th smallest point in I_N for $n = 1, 2, \ldots$, and $t_\infty = T$. Applying part (b) to this subMG we see that

$$\mathbf{P}(\sup_{t\in I_N} X_t > x) \le x^{-1}\mathbf{E}|X_T|.$$

With the sets I_N monotonically increasing in N to the set $\mathcal{Q}_T$ of all rational numbers in $[0, T]$, we may and shall change the range of t-s in the preceding supremum to $\mathcal{Q}_T$. Assuming that X_t has right continuous sample path,

$$\mathbf{P}(A_T) := \mathbf{P}(\sup_{0\le t\le T} X_t > x) = \mathbf{P}(\sup_{t\in\mathcal{Q}_T} X_t > x) ,$$

yielding the inequality (4.4.3). ∎

REMARK. Part (c) of Doob's inequality is [**KS97**, Theorem 1.3.8(i), page 14]. Many other martingale inequalities exist. For example, see [**KS97**, Theorem 1.3.8, parts (ii) and (iii)].

The following easy refinements of Theorem 4.4.13 are left to the reader.

EXERCISE 4.4.14. *Considering Exercise 4.2.8 with $\Psi(\cdot) = |\cdot|$ verify that when $\{X_n\}$ or $\{X_t\}$ is also a MG, the inequalities (4.4.1)–(4.4.3) apply with $|X_n|$ (or $|X_t|$, respectively), in their left-hand-side.*

Here are a few consequences of Doob's inequality for the random walk S_n.

EXAMPLE 4.4.15. *Consider the random walk $S_n = \sum_{i=1}^{n} \xi_i$ where the i.i.d. ξ_i are integrable and of zero mean. Applying Doob's inequality (4.4.1) to the sub-martingale $|S_n|$ we get that $\mathbf{P}(U_n > x) \le x^{-1}\mathbf{E}[|S_n|]$ for $U_n = \max\{|S_k| : 0 \le k \le n\}$ and any $x > 0$. If in addition $\mathbf{E}\xi_i^2 = 1$, then applying Doob's inequality to the sub-MG S_n^2*

yields the bound $\mathbf{P}(U_n > x) \le x^{-2}\mathbf{E}[S_n^2] = nx^{-2}$ *so* $U_n \le c\sqrt{n}$ *with high probability (for large* c *and all* n*). Next, let* $Z_n = \max\{S_k : 0 \le k \le n\}$ *denote the running maximum associated with the random walk, with* $Z_\infty = \max_k S_k$ *the global maximum of its sample path (possibly infinite). Replacing the assumption* $\mathbf{E}\xi_i = 0$ *with* $\mathbf{E}(e^{\xi_i}) = 1$ *results with the* exponential martingale $M_n = e^{S_n}$ *(see Example 4.1.14). Applying Doob's inequality (4.4.1) to the latter, we have for each* $x > 0$ *the bound*

$$\mathbf{P}(Z_n > x) = \mathbf{P}(\max_{0\le k\le n} S_k > x) = \mathbf{P}(\max_{0\le k\le n} M_k > e^x) \le e^{-x}\mathbf{E}(M_n) = e^{-x} .$$

This is an example of large-deviations, or exponential tail bounds. Since the events $\{Z_n > x\}$ *increase monotonically in* n *to* $\{Z_\infty > x\}$*, we conclude that* Z_∞ *is finite a.s. and has exponential upper tail, that is,* $\mathbf{P}(Z_\infty > x) \le e^{-x}$ *for all* $x \ge 0$*.*

As we have just seen, for sub-martingales with p-finite moments, $p > 1$, we may improve the rate of decay in terms of $x > 0$, from x^{-1} provided by Doob's inequality, to x^{-p}. Moreover, combining it with the formula

$$\mathbf{E}(Z^p) = \int_0^\infty px^{p-1}\mathbf{P}(Z > x)dx \tag{4.4.4}$$

which holds for any $p > 1$ and any non-negative R.V. Z, allows us next to efficiently bound the moments of the maximum of a subMG.

THEOREM 4.4.16. *Suppose* $\{X_t\}$ *is a right continuous subMG for* $t \in [0, T]$ *such that* $\mathbf{E}[(X_t)_+^p] < \infty$ *for some* $p > 1$ *and all* $t \ge 0$*. Then, for* $q = p/(p-1)$*, any* $x > 0$ *and* $t \le T$*,*

$$\mathbf{P}(\sup_{0\le u\le t} X_u > x) \le x^{-p}\mathbf{E}\Big[(X_t)_+^p\Big], \tag{4.4.5}$$

$$\mathbf{E}\big[(\sup_{0\le u\le t} X_u)_+^p\big] \le q^p\mathbf{E}[(X_t)_+^p] , \tag{4.4.6}$$

where $(y)_+^p$ *denotes the function* $(\max(y, 0))^p$*.*

EXAMPLE 4.4.17. *For the Brownian motion* W_t *we get from (4.4.5) that for any* $x > 0$ *and* $p > 1$*,*

$$\mathbf{P}(\sup_{0\le t\le T} W_t > x) \le x^{-p}\mathbf{E}[(W_T)_+^p]$$

and the right-hand side can be explicitly computed since W_T *is a Gaussian R.V. of zero-mean and variance* T*. We do not pursue this further since in Section 5.2 we explicitly compute the probability density function of* $\sup_{0\le t\le T} W_t$*.*

In the next exercise you are to restate the preceding theorem for the discrete parameter subMG $X_n = |Y_n|$.

EXERCISE 4.4.18. *Show that if* $\{Y_n\}$ *is a MG and* $\mathbf{E}[|Y_n|^p] < \infty$ *for some* $p > 1$ *and all* $n \le N$*, then for* $q = p/(p-1)$*, any* $y > 0$ *and all* $n \le N$*,*

$$\mathbf{P}(\max_{k\le n} |Y_k| > y) \le y^{-p}\mathbf{E}\Big[|Y_n|^p\Big] ,$$

$$\mathbf{E}\big[\big(\max_{k\le n} |Y_k|\big)^p\big] \le q^p\mathbf{E}\big[|Y_n|^p\big] .$$

4.5. Martingale convergence theorems

The fact that the maximum of a subMG does not grow too rapidly is closely related to convergence properties of subMG (also of supMG-s and of martingales).

The next theorem is stated in [**KS97**, Theorem 1.3.15 and Problem 1.3.19]. See [**GS01**, Theorem 12.3.1] for the corresponding discrete-time results and their proof.

THEOREM 4.5.1 (Doob's convergence theorem). *Suppose $(X_t, \mathcal{F}_t)$ is a right continuous subMG.*
(a). If $\sup_{t\geq 0}\{\mathbf{E}[(X_t)_+]\} < \infty$, then $X_\infty = \lim_{t\to\infty} X_t$ exists w.p.1. *Further, in this case $\mathbf{E}|X_\infty| \leq \lim_{t\to\infty} \mathbf{E}|X_t| < \infty$.*
(b). If $\{X_t\}$ is uniformly integrable then $X_t \to X_\infty$ also in L^1. Further, the L^1 convergence $X_t \to X_\infty$ implies that $X_t \leq \mathbf{E}(X_\infty|\mathcal{F}_t)$ for any fixed $t \geq 0$.

REMARK. Doob's convergence theorem is stated here for a continuous parameter subMG (of right continuous sample path). The corresponding theorem is also easy to state for $(X_t, \mathcal{F}_t)$ which is a (right continuous) supMG since then $(-X_t, \mathcal{F}_t)$ is a subMG. Similarly, to any discrete parameter subMG (or supMG) corresponds the right continuous interpolated subMG (or supMG, respectively), as done in Example 4.2.7 (for the MG case). Consequently, Doob's convergence theorem applies also when replacing $(X_t, \mathcal{F}_t)$ by $(X_n, \mathcal{F}_n)$ throughout its statement (with right-continuity then irrelevant).

Further, Doob's convergence theorem provides the following characterization of right continuous, *Uniformly Integrable* (U.I.) martingales.

COROLLARY 4.5.2. *If $(X_t, \mathcal{F}_t)$ is a right continuous MG and $\sup_t \mathbf{E}|X_t| < \infty$, then $X_\infty = \lim\limits_{t\to\infty} X_t$ exists w.p.1. and is integrable. If $\{X_t\}$ is also U.I. then $X_t = \mathbf{E}(X_\infty|\mathcal{F}_t)$ for all t (such a martingale, namely $X_t = \mathbf{E}(X|\mathcal{F}_t)$ for an integrable R.V. X and a filtration $\{\mathcal{F}_t\}$, is called* Doob's martingale of X with respect to *$\{\mathcal{F}_t\}$).*

REMARK. To understand the difference between parts (a) and (b) of Doob's convergence theorem, recall that if $\{X_t\}$ is uniformly integrable then $\mathbf{E}[(X_t)_+] \leq C$ for some $C < \infty$ and all t (see Definition 1.4.22). Further, by Theorem 1.4.23, the uniform integrability together with convergence almost surely imply convergence in L^1. So, the content of part (b) is that the L^1 convergence $X_t \to X_\infty$ implies also that $X_t \leq \mathbf{E}(X_\infty|\mathcal{F}_t)$ for any $t \geq 0$.

Keep in mind that many important martingales do not converge. For example, as we see in Section 5.2, the path of the Brownian motion W_t exceeds any level $\alpha > 0$ within finite time τ_α. By symmetry, the same applies to any negative level $-\alpha$. Thus, almost surely, $\limsup_{t\to\infty} W_t = \infty$ and $\liminf_{t\to\infty} W_t = -\infty$, that is, the magnitude of oscillations of the Brownian sample path grows indefinitely. Indeed, note that $\mathbf{E}[(W_t)_+] = \sqrt{t/(2\pi)}$ is unbounded, so Doob's convergence theorem does not apply to the martingale W_t.

An important family of U.I. martingales are those with bounded second moment (see Exercise 1.4.25). For example, the next proposition is a direct consequence of Doob's convergence theorem.

PROPOSITION 4.5.3. *If the right continuous MG $\{Y_t\}$ is such that $\mathbf{E}Y_t^2 \le C$ for some $C < \infty$ and all $t \ge 0$, then there exists a R.V. Y_∞ such that $Y_t \to Y_\infty$ almost surely and in L^2. Moreover, $\mathbf{E}Y_\infty^2 \le C < \infty$ and the corresponding result holds in the context of discrete parameter MGs.*

PROOF. (omit at first reading)
Considering part (a) of Doob's martingale convergence theorem for the non-negative right continuous subMG $X_t = Y_t^2$ and its a.s. limit $X_\infty = Y_\infty^2$ we deduce that $Y_t \stackrel{a.s.}{\to} Y_\infty$ for a square-integrable Y_∞ such that

$$C \ge \lim_{t\to\infty} \mathbf{E}Y_t^2 = \lim_{t\to\infty} \mathbf{E}X_t \ge \mathbf{E}X_\infty = \mathbf{E}Y_\infty^2 .$$

To complete the proof it suffices to show that $\mathbf{E}(|Y_t - Y_\infty|^2) \to 0$ as $t \to \infty$. To this end, considering (4.4.6) for the non-negative right-continuous subMG $X_u = |Y_u|$ and $p = 2$ we have that $\mathbf{E}(Z_t) \le 4C$ for $Z_t = \sup_{0\le u\le t} Y_u^2$ and any $t < \infty$. Since $0 \le Z_t \uparrow Z = \sup_{0\le u} Y_u^2$, we have by monotone convergence that $\mathbf{E}(Z) \le 4C < \infty$. With Y_∞ the a.s. limit of Y_t, it follows that $Y_\infty^2 \le Z$ as well. Hence, $V_t = |Y_t - Y_\infty|^2 \le 2Y_t^2 + 2Y_\infty^2 \le 4Z$. We see that $V_t \to 0$ a.s. and $V_t \le Z$ with $\mathbf{E}Z < \infty$, so applying the dominated convergence theorem to V_t, we get that $\mathbf{E}(|Y_t - Y_\infty|^2) \to 0$ as $t \to \infty$, completing the proof of the proposition. ∎

REMARK. Beware that Proposition 4.5.3 does not have an L^1 analog. Namely, there exists a non-negative MG $\{Y_n\}$ such that $\mathbf{E}Y_n = 1$ for all n and $Y_n \to Y_\infty = 0$ almost surely, so obviously, Y_n does not converge to Y_∞ in L^1. One such example is given in Proposition 4.6.5.

Here are a few applications of Doob's convergence theorem.

EXERCISE 4.5.4. *Consider an urn that at stage 0 contains one red ball and one blue ball. At each stage a ball is drawn at random from the urn, with all possible choices being equally likely, and it and one more ball of the same color are then returned to the urn. Let R_n denote the number of red balls at stage n and $M_n = R_n/(n+2)$ the corresponding fraction of red balls.*

(a) *Find the law of R_{n+1} conditional on $R_n = k$ and use it to compute $\mathbf{E}(R_{n+1}|R_n)$.*
(b) *Check that $\{M_n\}$ is a martingale with respect to its canonical filtration.*
(c) *Applying Proposition 4.5.3 conclude that $M_n \to M_\infty$ in L^2 and that $\mathbf{E}(M_\infty) = \mathbf{E}(M_0) = 1/2$.*
(d) *Using Doob's (maximal) inequality show that $\mathbf{P}(\sup_{k\ge1} M_k > 3/4) \le 2/3$.*

EXAMPLE 4.5.5. *Consider the martingale $S_n = \sum_{k=1}^n \xi_k$ for independent, square-integrable, zero-mean random variables ξ_k. Since $\mathbf{E}S_n^2 = \sum_{k=1}^n \mathbf{E}\xi_k^2$, it follows from Proposition 4.5.3 that the random series $S_n(\omega) \to S_\infty(\omega)$ almost surely and in L^2 provided $\sum_k \mathbf{E}\xi_k^2 < \infty$.*

EXERCISE 4.5.6. *Deduce from part (a) of Doob's convergence theorem that if $\{X_t\}$ is a non-negative right continuous martingale, then $X_t \stackrel{a.s.}{\to} X_\infty$ and $\mathbf{E}X_\infty \le \mathbf{E}X_0 < \infty$. Further, $X_n \stackrel{a.s.}{\to} X_\infty$ and $\mathbf{E}X_\infty \le \mathbf{E}X_0 < \infty$ for any non-negative, discrete-time martingale $\{X_n\}$.*

4.6. Branching processes: extinction probabilities

We use martingales to study the extinction probabilities of branching processes (defined next). See [**KT75**, Chapter 8] or [**GS01**, Sections 5.4-5.5] for more on these processes.

DEFINITION 4.6.1 (Branching Process). *The* Branching process *is a discrete time S.P.* $\{Z_n\}$ *taking non-negative integer values, such that* $Z_0 = 1$ *and for any* $n \geq 1$,

$$Z_n = \sum_{j=1}^{Z_{n-1}} N_j^{(n)},$$

where N *and* $N_j^{(n)}$ *for* $j = 1, 2, \ldots$ *are independent, identically distributed, non-negative integer valued R.V. with finite mean* $m = \mathbf{E}(N) < \infty$*, and where we use the convention that if* $Z_{n-1} = 0$ *then also* $Z_n = 0$*.*

The S.P. $\{Z_n\}$ is interpreted as counting the size of an evolving population, with $N_j^{(n)}$ being the number of offspring of j^{th} individual of generation $(n-1)$ and Z_n being the size of the n-th generation. Associated with the branching process is the family tree with the root denoting the 0-th generation and having $N_j^{(n)}$ edges from vertex j at distance $(n-1)$ from the root to vertices of distance n from the root. Random trees generated in such a fashion are called *Galton-Watson trees* and are the subject of much research. We focus on the simpler S.P. $\{Z_n\}$ and shall use throughout the filtration $\mathcal{F}_n = \sigma(\{N_j^{(k)}, k \leq n, j = 1, 2, \ldots\})$. We note in passing that in general $\mathcal{G}_n = \sigma(\{Z_k, k \leq n\})$ is a strict subset of $\mathcal{F}_n$ (since in general one can not recover the number of offspring of each individual knowing only the total population sizes at the different generations).

PROPOSITION 4.6.2. *The S.P.* $X_n = m^{-n} Z_n$ *is a martingale for the filtration* $\{\mathcal{F}_n\}$*.*

PROOF. Note that the value of Z_n is a non-random function of the values of $\{N_j^{(k)}, k \leq n, j = 1, 2, \ldots\}$. Hence, it follows that $\{Z_n\}$ is adapted to the filtration $\{\mathcal{F}_n\}$. It suffices to show that

$$\mathbf{E}[Z_{n+1} | \mathcal{F}_n] = m Z_n. \tag{4.6.1}$$

Indeed, then by the tower property and induction $\mathbf{E}[Z_{n+1}] = m\mathbf{E}[Z_n] = m^{n+1}$, providing the integrability of $\{Z_n\}$. Moreover, the identity (4.6.1) reads also as $\mathbf{E}[X_{n+1} | \mathcal{F}_n] = X_n$ for $X_n = m^{-n} Z_n$, which is precisely the stated martingale property of $(X_n, \mathcal{F}_n)$. To prove (4.6.1), note that the random variables $N_j^{(n+1)}$ are independent of $\mathcal{F}_n$ on which Z_n is measurable. Hence, by the linearity of the expectation

$$\mathbf{E}[Z_{n+1} | \mathcal{F}_n] = \sum_{j=1}^{Z_n} \mathbf{E}[N_j^{(n+1)} | \mathcal{F}_n] = \sum_{j=1}^{Z_n} \mathbf{E}(N_j^{(n+1)}) = m Z_n,$$

as claimed. ∎

While proving Proposition 4.6.2 we showed that $\mathbf{E} Z_n = m^n$ for all $n \geq 0$. Thus, the mean total population size of a sub-critical branching process (i.e. $m < 1$) is

$$\mathbf{E}(\sum_{n=0}^{\infty} Z_n) = \sum_{n=0}^{\infty} m^n = \frac{1}{1-m} < \infty\,.$$

In particular, $\sum_n Z_n(\omega)$ is then finite w.p.1. Since $Z_n \geq 0$ are integer valued, this in turn implies that the *extinction probability* is one, namely,

$$p_{ex} := \mathbf{P}(\{\omega : Z_n(\omega) = 0 \text{ for all } n \text{ large enough } \}) = 1 \,.$$

We provide next another proof of this result, using the martingale $X_n = m^{-n} Z_n$ of Proposition 4.6.2.

PROPOSITION 4.6.3 (sub-critical process dies off). *If $m < 1$ then $p_{ex} = 1$, that is, w.p.1. the population eventually dies off.*

PROOF. Recall that $X_n = m^{-n} Z_n$ is a non-negative MG. Thus, by Doob's martingale convergence theorem, $X_n \stackrel{a.s.}{\to} X_\infty$, where the R.V. X_∞ is almost surely finite (see Exercise 4.5.6). With $Z_n = m^n X_n$ and $m^n \to 0$ (for $m < 1$), this implies that $Z_n \stackrel{a.s.}{\to} 0$. Because Z_n are integer valued, $Z_n \to 0$ only if $Z_n = 0$ eventually, completing the proof of the proposition. ∎

As we see next, more can be done in the special case where each individual has at most one offspring.

EXERCISE 4.6.4. *Consider the sub-critical branching process $\{Z_n\}$ where independently each individual has either one offspring (with probability $0 < p < 1$) or no offspring as all (with probability $q = 1 - p$). Starting with $Z_0 = N \geq 1$ individuals, compute the law $\mathbf{P}(T = n)$ of the extinction time $T = \min\{k \geq 1 : Z_k = 0\}$.*

We now move to critical branching processes, namely, those $\{Z_n\}$ for which $m = \mathbf{E}N = 1$. Excluding the trivial case in which each individual has exactly one offspring, we show that again the population eventually dies off w.p.1.

PROPOSITION 4.6.5 (critical process dies off). *If $m = 1$ and $\mathbf{P}(N = 1) < 1$ then $p_{ex} = 1$.*

PROOF. As we have already seen, when $m = 1$ the branching process $\{Z_n\}$ is a non-negative martingale. So, by Exercise 4.5.6, $Z_n \stackrel{a.s.}{\to} Z_\infty$, with Z_∞ finite almost surely. Since Z_n are integer valued this can happen only if for almost every ω there exist non-negative integers k and ℓ (possibly depending on ω), such that $Z_n = k$ for all $n \geq \ell$. Our assumption that $\mathbf{P}(N = 1) < 1$ and $\mathbf{E}(N) = 1$ for the non-negative integer valued R.V. N implies that $\mathbf{P}(N = 0) > 0$. Note that given $Z_n = k > 0$, if $N_j^{(n+1)} = 0$ for $j = 1, \ldots, k$, then $Z_{n+1} = 0 \neq k$. By the independence of $\{N_j^{(n+1)}, j = 1, \ldots\}$ and $\mathcal{F}_n$ deduce that for each n and $k > 0$,

$$\mathbf{P}(Z_{n+1} = k | Z_n = k, \mathcal{F}_n) \leq 1 - \mathbf{P}(Z_{n+1} = 0 | Z_n = k, \mathcal{F}_n) = 1 - \mathbf{P}(N = 0)^k := \eta_k < 1.$$

For non-random integers $m > \ell \geq 0$ and $k \geq 0$ let $A_{\ell,m,k}$ denote the event $\{Z_n = k$, for $n = \ell, \ldots, m\}$. Note that $I_{A_{\ell,m,k}} = I_{Z_m = k} I_{A_{\ell,m-1,k}}$, with $A_{\ell,m-1,k} \in \mathcal{F}_{m-1}$ implying that $Z_{m-1} = k$. So, first "taking out what is known" (per Proposition 2.3.15), then applying the preceding inequality for $n = m - 1$, we deduce that for any $m > \ell$ and $k > 0$,

$$\mathbf{E}(I_{A_{\ell,m,k}} | \mathcal{F}_{m-1}) \leq \eta_k I_{A_{\ell,m-1,k}} \,.$$

Hence, by the tower property we have that $\mathbf{P}(A_{\ell,m,k}) \leq \eta_k \mathbf{P}(A_{\ell,m-1,k})$. With $\mathbf{P}(A_{\ell,\ell,k}) \leq 1$, it follows that $\mathbf{P}(A_{\ell,m,k}) \leq \eta_k^{m-\ell}$. We deduce that if $k > 0$, then $\mathbf{P}(A_{\ell,k}) = 0$ for $A_{\ell,k} = A_{\ell,\infty,k} = \{Z_n = k$, for all $n \geq \ell\}$ and any $\ell \geq 0$. We have seen that $\mathbf{P}(\cup_{\ell,k \geq 0} A_{\ell,k}) = 1$ while $\mathbf{P}(A_{\ell,k}) = 0$ whenever $k > 0$. So, necessarily $\mathbf{P}(\cup_\ell A_{\ell,0}) = 1$, which amounts to $p_{ex} = 1$. ∎

REMARK. Proposition 4.6.5 shows that in case $m = 1$, the martingale $\{Z_n\}$ converges to 0 w.p.1. If this sequence was U.I. then by part (b) of Doob's convergence theorem, necessarily $Z_n \to 0$ also in L^1, i.e. $\mathbf{E}Z_n \to 0$. However, $\mathbf{E}Z_n = 1$ for all n, so we conclude that the sequence $\{Z_n\}$ is not uniformly integrable. Further, note that either $Z_n = 0$ or $Z_n \geq 1$, so $1 = \mathbf{E}(Z_n I_{Z_n \geq 1}) = \mathbf{E}(Z_n | Z_n \geq 1)\mathbf{P}(Z_n \geq 1)$. With $p_{ex} = 1$, the probability $\mathbf{P}(Z_n \geq 1)$ of survival for n generations decays to zero as $n \to \infty$ and consequently, conditional upon survival, the mean population size $\mathbf{E}(Z_n|Z_n \geq 1) = 1/\mathbf{P}(Z_n \geq 1)$ grows to infinity as $n \to \infty$.

The martingale of Proposition 4.6.2 does not provide information on the value of p_{ex} for a super-critical branching process, that is, when $m > 1$. However, as we see next, when N is also square-integrable, it implies that $m^{-n}Z_n$ converges in law to a non-zero random variable.

EXERCISE 4.6.6. *Let Z_n be the population size for the n-th generation of a (super-critical) branching process, with the number of offspring having mean $m = \mathbf{E}(N) > 1$ and finite variance $\sigma^2 = \mathrm{Var}(N) < \infty$.*

(a) *Check that $\mathbf{E}(Z_{n+1}^2) = m^2\mathbf{E}(Z_n^2) + \sigma^2\mathbf{E}(Z_n)$.*
(b) *Compute $\mathbf{E}(X_n^2)$ for $X_n = m^{-n}Z_n$.*
(c) *Show that X_n converges in law to some R.V. X_∞ with $\mathbf{P}(X_\infty > 0) > 0$*
Hint: *Use Proposition 4.5.3.*

For a super-critical branching process, if $\mathbf{P}(N = 0) = 0$ extinction is of course impossible. Otherwise, $p_{ex} \geq \mathbf{P}(Z_1 = 0) = \mathbf{P}(N = 0) > 0$. Nevertheless, even in this case $p_{ex} < 1$ for any super-critical branching process. That is, with positive probability the population survives forever. Turning to compute p_{ex}, consider the function

$$\varphi(p) = \mathbf{P}(N = 0) + \sum_{k=1}^{\infty} \mathbf{P}(N = k)p^k \,. \tag{4.6.2}$$

Note that the convex continuous function $\varphi(p)$ is such that $\varphi(0) = \mathbf{P}(N = 0) > 0$, $\varphi(1) = 1$ and $\varphi'(1) = \mathbf{E}N > 1$. It is not hard to verify that for any such function there exists a unique solution $\rho \in (0, 1)$ of the equation $p = \varphi(p)$. Upon verifying that ρ^{Z_n} is then a martingale, you are to show next that $p_{ex} = \rho$ (for a different derivation without martingales, see [**GS01**, Section 5.4] or [**KT75**, Section 8.3]).

EXERCISE 4.6.7. *Consider a super-critical branching process $\{Z_n\}$ with $Z_0 = 1$, $\mathbf{P}(N = 0) > 0$ and let ρ denote the unique solution of $p = \varphi(p)$ in $(0, 1)$.*

(a) *Fixing $p \in (0, 1)$ check that $\mathbf{E}(p^{Z_{n+1}}|\mathcal{F}_n) = \varphi(p)^{Z_n}$ and verify that thus $M_n = \rho^{Z_n}$ is a uniformly bounded martingale with respect to the filtration $\mathcal{F}_n$.*
(b) *Use Doob's convergence theorem to show that $M_n \to M_\infty$ almost surely and in L^1 with $\mathbf{E}(M_\infty) = \rho$.*
(c) *Check that since Z_n are non-negative integers, the random variable M_∞ can only take the values ρ^k for $k = 0, 1, \ldots$ or the value 0 that corresponds to $Z_n(\omega) \to \infty$.*
(d) *Adapting the proof of Proposition 4.6.5 show that actually $M_\infty(\omega) \in \{0, 1\}$ with probability one.*
(e) *Noting the $M_\infty = 1$ if and only if the branching process is eventually extinct, conclude that $p_{ex} = \rho$.*

EXERCISE 4.6.8. *Suppose* $\{Z_n\}$ *is a branching process with* $\mathbf{P}(N = 1) < 1$ *and* $Z_0 = 1$. *Show that*

$$\mathbf{P}(\lim_{n\to\infty} Z_n = \infty) = 1 - p_{ex}\,,$$

first in case $m \le 1$, *then in case* $\mathbf{P}(N = 0) = 0$ *and finally using the preceding exercise, for* $m > 1$ *and* $\mathbf{P}(N = 0) > 0$.

EXERCISE 4.6.9. *Let* $\{Z_n\}$ *be a branching process with* $Z_0 = 1$. *Compute* p_{ex} *in each of the following situations and specify for which values of the various parameters extinction is certain.*

(a) *The offspring distribution satisfies, for some* $0 < p < 1$,

$$\mathbf{P}(N = 0) = p\ ,\ \ \mathbf{P}(N = 2) = 1 - p.$$

(b) *The offspring distribution is (shifted) Geometric, i.e. for some* $0 < p < 1$,

$$\mathbf{P}(N = k) = p(1 - p)^k\ ,\ \ k = 0, 1, 2, \ldots$$

CHAPTER 5

The Brownian motion

The Brownian motion is the most fundamental continuous time stochastic process. It is both a martingale of the type considered in Section 4.2 and a Gaussian process as considered in Section 3.2. It also has continuous sample path, independent increments, and the strong Markov property of Section 6.1. Having all these beautiful properties allows for a rich mathematical theory. For example, many probabilistic computations involving the Brownian motion can be made explicit by solving partial differential equations. Further, the Brownian motion is the corner stone of diffusion theory and of stochastic integration. As such it is the most fundamental object in applications to and modeling of natural and man-made phenomena.

In this chapter we define and construct the Brownian motion (in Section 5.1), then deal with a few of the many interesting properties it has. Specifically, in Section 5.2 we use stopping time and martingale theory to study the hitting times and the running maxima of this process, whereas in Section 5.3 we consider the smoothness and variation of its sample path.

5.1. Brownian motion: definition and construction

Our starting point is an axiomatic definition of the Brownian motion via its Gaussian property.

Definition 5.1.1. *A stochastic process* (W_t, $0 \le t \le T$) *is called a* Brownian motion *(or a Wiener Process) if:*
(a) W_t *is a Gaussian process*
(b) $\mathbf{E}(W_t) = 0$, $\mathbf{E}(W_t W_s) = \min(t, s)$
(c) For almost every ω, *the sample path,* $t \mapsto W_t(\omega)$ *is continuous on* $[0, T]$.

Note that (a) and (b) of Definition 5.1.1 completely characterize the finite dimensional distributions of the Brownian motion (recall Corollary 3.2.18 that Gaussian processes are characterized by their mean and auto-covariance functions). Adding property (c) to Definition 5.1.1 allows us to characterize its sample path as well. We shall further study the Brownian sample path in Sections 5.2 and 5.3. We next establish the independence of the zero-mean Brownian increments, implying that the Brownian motion is an example of the martingale processes of Section 4.2 (see Proposition 4.2.3). Note however that the Brownian motion is a non-stationary process (see Proposition 3.2.25), though it does have stationary increments.

Proposition 5.1.2. *The Brownian motion has independent increments of zero-mean.*

Proof. From part (b) of Definition 5.1.1, we obtain that for $t \ge s$ and $h > 0$,

$$\mathrm{Cov}(W_{t+h} - W_t, W_s) = \mathbf{E}[(W_{t+h} - W_t)W_s] = \mathbf{E}(W_{t+h}W_s) - E(W_t W_s) = s - s = 0\,.$$

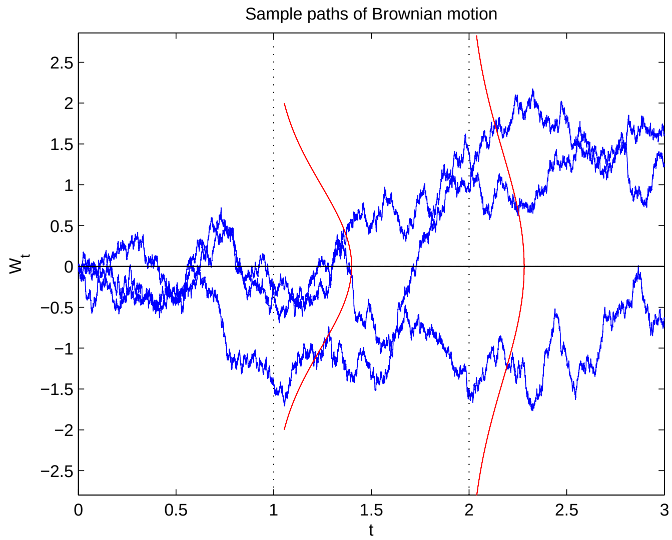

FIGURE 1. Three sample paths of Brownian motion. The density curves illustrate that the random variable W_1 has a $N(0,1)$ distribution, while W_2 has a $N(0,2)$ distribution.

Thus, $W_{t+h} - W_t$ and W_s are uncorrelated for any fixed $h > 0$ and $s \le t$. Fixing $n < \infty$ and $0 \le s_1 < s_2 < \ldots < s_n \le t$, since $\{W_t\}$ is a Gaussian process, we know that $(W_{t+h}, W_t, W_{s_1}, \ldots, W_{s_n})$ is a Gaussian random vector, and hence so is $\underline{X} = (W_{t+h} - W_t, W_{s_1}, \ldots, W_{s_n})$ (recall Proposition 3.2.16). The vector $\underline{X}$ has mean $\underline{\mu} = \underline{0}$ and covariance matrix Σ such that $\Sigma_{0k} = \mathbf{E}(W_{t+h} - W_t)W_{s_k} = 0$ for $k = 1, \ldots, n$. In view of Definition 3.2.8 this results with the characteristic function $\Phi_{\underline{X}}(\underline{\theta})$ being the product of the characteristic function of $W_{t+h} - W_t$ and that of $(W_{s_1}, \ldots, W_{s_n})$. Consequently, $W_{t+h} - W_t$ is independent of $(W_{s_1}, \ldots, W_{s_n})$ (see Proposition 3.2.6). Since this applies for any $0 \le s_1 < s_2 < \ldots < s_n \le t$, it can be shown that $W_{t+h} - W_t$ is also independent of $\sigma(W_s, s \le t)$.

In conclusion, the Brownian motion is an example of a zero mean S.P. with independent increments. That is, $(W_{t+h} - W_t)$ is independent of $\{W_s, s \in [0,t]\}$, as stated. ∎

We proceed to construct the Brownian motion as in [**Bre92**, Section 12.7]. To this end, consider

$$L^2([0,T]) = \{f(u) : \int_0^T f^2(u)du < \infty\},$$

equipped with the inner product, $(f,g) = \int_0^T f(u)g(u)du$, where we identify f, g such that $f(t) = g(t)$ for almost every $t \in [0,T]$, as being the same function. As we have seen in Example 2.2.21, this is a separable *Hilbert space*, and there exists a non-random sequence of functions, $\{\phi_i(t)\}_{i=1}^{\infty}$ in $L^2([0,T])$, such that for any

$f, g \in L^2([0,T])$,

$$\lim_{n\to\infty} \sum_{i=1}^{n} (f, \phi_i)(g, \phi_i) = (f, g) \tag{5.1.1}$$

(c.f. Definition 2.2.17 and Theorem 2.2.20). Let X_i be i.i.d., Gaussian R.V.-s with $\mathbf{E}X_i = 0$ and $\mathbf{E}X_i^2 = 1$, all of which are defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$. For each positive integer N define the stochastic process

$$V_t^N = \sum_{i=1}^{N} X_i \int_0^t \phi_i(u) du \,.$$

Since $\phi_i(t)$ are non-random and any linear combination of the coordinates of a Gaussian random vector gives a Gaussian random vector (see Proposition 3.2.16), we see that V_t^N is a Gaussian process.

We shall show that the random variables V_t^N converge in $L^2(\Omega, \mathcal{F}, \mathbf{P})$ to some random variable V_t, for any fixed, non-random, $t \in [0,T]$. Moreover, we show that the S.P. V_t has properties (a) and (b) of Definition 5.1.1. Then, applying Kolmogorov's continuity theorem, we deduce that the continuous modification of the S.P. V_t is the Brownian motion.

Our next result provides the first part of this program.

Proposition 5.1.3. *Fixing $t \in [0,T]$, the sequence $N \mapsto V_t^N$ is a Cauchy sequence in the Hilbert space $L^2(\Omega, \mathcal{F}, \mathbf{P})$. Consequently, there exists a S.P. $V_t(\omega)$ such that $\mathbf{E}[(V_t - V_t^N)^2] \to 0$ as $N \to \infty$, for any $t \in [0,T]$. The S.P. V_t is Gaussian with $\mathbf{E}(V_t) = 0$ and $\mathbf{E}(V_t V_s) = \min(t, s)$.*

Proof. Fix $t \in [0,T]$, noting that for any i,

$$\int_0^t \phi_i(u) du = \int_0^T \mathbf{1}_{[0,t]}(u) \phi_i(u) du = (\mathbf{1}_{[0,t]}, \phi_i) \,. \tag{5.1.2}$$

Set $V_t^0 = 0$ and let

$$\psi_n(t) = \sum_{i=n+1}^{\infty} (\mathbf{1}_{[0,t]}, \phi_i)^2 \,.$$

Since $\mathbf{E}(X_i X_j) = \mathbf{1}_{i=j}$ we have for any $N > M \geq 0$,

$$\begin{aligned}
\mathbf{E}\big[(V_t^N - V_t^M)^2\big] &= \sum_{i=M+1}^{N} \sum_{j=M+1}^{N} \mathbf{E}[X_i X_j] \Big(\int_0^t \phi_i(u) du\Big)\Big(\int_0^t \phi_j(u) du\Big) \\
&= \sum_{i=M+1}^{N} \Big(\int_0^t \phi_i(u) du\Big)^2 = \psi_M(t) - \psi_N(t) \qquad (5.1.3)
\end{aligned}$$

(using (5.1.2) for the rightmost equality). Applying (5.1.1) for $f = g = \mathbf{1}_{[0,t]}(\cdot)$ we have that for all M,

$$\psi_M(t) \leq \psi_0(t) = \sum_{i=1}^{\infty} (\mathbf{1}_{[0,t]}, \phi_i)^2 = (\mathbf{1}_{[0,t]}, \mathbf{1}_{[0,t]}) = t < \infty.$$

In particular, taking $M = 0$ in (5.1.3) we see that $\mathbf{E}[(V_t^N)^2]$ are finite for all N. It further follows from the finiteness of the infinite series $\psi_0(t)$ that $\psi_n(t) \to 0$ as $n \to \infty$. In view of (5.1.3) we deduce that V_t^N is a Cauchy sequence in

$L^2(\Omega, \mathcal{F}, \mathbf{P})$, converging to some random variable V_t by the completeness of this space (see Proposition 1.3.20).

Being the pointwise (in t) limit in 2-mean of Gaussian processes, the S.P. V_t is also Gaussian, with the mean and auto-covariance functions for V_t being the (pointwise in t) limits of those for V_t^N (c.f. Proposition 3.2.20). Recall that $\mathbf{E}(V_t^N) = \sum_{i=1}^{N} \mathbf{E}X_i \int_0^t \phi_i(u)du = 0$, for all N, hence $\mathbf{E}(V_t) = 0$ as well.

Repeating the argument used when deriving (5.1.3) we see that for any $s, t \in [0, T]$,

$$\mathbf{E}(V_t^N V_s^N) = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{E}[X_i X_j] (\int_0^t \phi_i(u)du)(\int_0^s \phi_j(u)du) = \sum_{i=1}^{N} (\mathbf{1}_{[0,t]}, \phi_i)(\mathbf{1}_{[0,s]}, \phi_i) .$$

Applying (5.1.1) for $f = \mathbf{1}_{[0,t]}(\cdot)$ and $g = \mathbf{1}_{[0,s]}(\cdot)$, both in $L^2([0,T])$, we now have that

$$\begin{aligned} \mathbf{E}(V_t V_s) &= \lim_{N\to\infty} \mathbf{E}(V_t^N V_s^N) = \lim_{N\to\infty} \sum_{i=1}^{N} (\mathbf{1}_{[0,t]}, \phi_i)(\mathbf{1}_{[0,s]}, \phi_i) \\ &= (\mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]}) = \min(t, s) , \end{aligned}$$

as needed to conclude the proof of the proposition. ∎

Having constructed a Gaussian stochastic process V_t with the same distribution as a Brownian motion, we next apply *Kolmogorov's continuity theorem*, so as to obtain its continuous modification. This modification is then a Brownian motion. To this end, recall that a *Gaussian* R.V. Y with $\mathbf{E}Y = 0, \mathbf{E}Y^2 = \sigma^2$ has *moments* $\mathbf{E}(Y^{2n}) = \frac{(2n)!}{2^n n!}\sigma^{2n}$. In particular, $\mathbf{E}(Y^4) = 3(\mathbf{E}(Y^2))^2$. Since V_t is Gaussian with

$$\mathbf{E}[(V_{t+h} - V_t)^2] = \mathbf{E}[(V_{t+h} - V_t)V_{t+h}] - \mathbf{E}[(V_{t+h} - V_t)V_t] = h ,$$

for all t and $h > 0$, we get that

$$\mathbf{E}[(V_{t+h} - V_t)^4] = 3[\mathbf{E}(V_{t+h} - V_t)^2]^2 = 3h^2,$$

as needed to apply Kolmogorov's theorem (with $\alpha = 4$, $\beta = 1$ and $c = 3$ there).

REMARK. There is an alternative direct construction of the Brownian motion as the limit of time-space rescaled random walks (see Theorem 3.1.3 for details). Further, though we constructed the Brownian motion W_t as a stochastic process on $[0, T]$ for some finite $T < \infty$, it easily extends to a process on $[0, \infty)$, which we thus take hereafter as the index set of the Brownian motion.

The Brownian motion has many interesting scaling properties, some of which are summarized in your next two exercises.

EXERCISE 5.1.4. *Suppose W_t is a Brownian motion and $\alpha, s, T > 0$ are non-random constants. Show the following.*

(a) *(Symmetry) $\{-W_t, t \geq 0\}$ is a Brownian motion.*
(b) *(Time homogeneity) $\{W_{s+t} - W_s, t \geq 0\}$ is a Brownian motion.*
(c) *(Time reversal) $\{W_T - W_{T-t}, 0 \leq t \leq T\}$ is a Brownian motion.*
(d) *(Scaling, or self-similarity) $\{\sqrt{\alpha} W_{t/\alpha}, t \geq 0\}$ is a Brownian motion.*
(e) *(Time inversion) If $\tilde{W}_0 = 0$ and $\tilde{W}_t = tW_{1/t}$, then $\{\tilde{W}_t, t \geq 0\}$ is a Brownian motion.*

(f) *With W_t^i denoting independent Brownian motions find the constants c_n such that $c_n \sum_{i=1}^n W_t^i$ are also Brownian motions.*

EXERCISE 5.1.5. *Fix $\rho \in [-1,1]$. Let $\widetilde{W}_t = \rho W_t^1 + \sqrt{1-\rho^2} W_t^2$ where W_t^1 and W_t^2 are two independent Brownian motions. Show that $\widetilde{W}_t$ is a Brownian motion and find the value of $\mathbf{E}(W_t^1 \widetilde{W}_t)$.*

EXERCISE 5.1.6. *Fixing $s > 0$ show that the S.P. $\{W_s - W_{s-t}, 0 \le t \le s\}$ and $\{W_{s+t} - W_s, t \ge 0\}$ are two independent Brownian motions and for $0 < t \le s$ evaluate $q_t = \mathbf{P}(W_s > W_{s-t} > W_{s+t})$.*

Applying Doob's inequality you are to prove next the *law of large numbers* for Brownian motion, namely, that almost surely $t^{-1} W_t \to 0$ for $t \to \infty$ (compare with the more familiar law of large numbers, $n^{-1}[S_n - \mathbf{E}S_n] \to 0$ for a random walk S_n).

EXERCISE 5.1.7. *Let W_t be a Brownian motion.*

(a) *Use the inequality* (4.4.6) *to show that for any $0 < u < v$,*

$$\mathbf{E}\Big[\big(\sup_{u \le t \le v} |W_t|/t\big)^2\Big] \le \frac{4v}{u^2}.$$

(b) *Taking $u = 2^n$ and $v = 2^{n+1}$, $n \ge 1$ in part (a), apply Markov's inequality to deduce that for any $\epsilon > 0$,*

$$\mathbf{P}\Big(\sup_{2^n \le t \le 2^{n+1}} |W_t|/t > \epsilon\Big) \le 8\epsilon^{-2} 2^{-n}.$$

(c) *Applying Borel-Cantelli lemma I conclude that almost surely $t^{-1} W_t \to 0$ for $t \to \infty$.*

Many important S.P. are derived from the Brownian motion W_t. Our next two exercises introduce a few of these processes, the *Brownian bridge* $B_t = W_t - \min(t,1) W_1$, the *Geometric Brownian motion* $Y_t = e^{W_t}$, and the *Ornstein-Uhlenbeck process* $U_t = e^{-t/2} W_{e^t}$. We also define $X_t = x + \mu t + \sigma W_t$, a *Brownian motion with drift* $\mu \in \mathbb{R}$ and diffusion coefficient $\sigma > 0$ starting from $x \in \mathbb{R}$. (See Figure 2 for illustrations of sample paths associated with these processes.)

EXERCISE 5.1.8. *Compute the mean and the auto-covariance functions of the processes B_t, Y_t, U_t, and X_t. Justify your answers to:*

(a) *Which of the processes W_t, B_t, Y_t, U_t, X_t is Gaussian?*
(b) *Which of these processes is stationary?*
(c) *Which of these processes has continuous sample path?*
(d) *Which of these processes is adapted to the filtration $\sigma(W_s, s \le t)$ and which is also a sub-martingale for this filtration?*

EXERCISE 5.1.9. *Show that for $0 \le t \le 1$ each of the following S.P. has the same distribution as the Brownian bridge and explain why both have continuous modifications.*

(a) $\widehat{B}_t = (1-t) W_{t/(1-t)}$ *for $t < 1$ with $\widehat{B}_1 = 0$.*
(b) $Z_t = t W_{1/t - 1}$ *for $t > 0$ with $Z_0 = 0$.*

EXERCISE 5.1.10. *Let $X_t = \int_0^t W_s ds$ for a Brownian motion W_t.*

(a) *Verify that X_t is a well defined stochastic process. That is, check that $\omega \mapsto X_t(\omega)$ is a random variable for each fixed $t \ge 0$.*

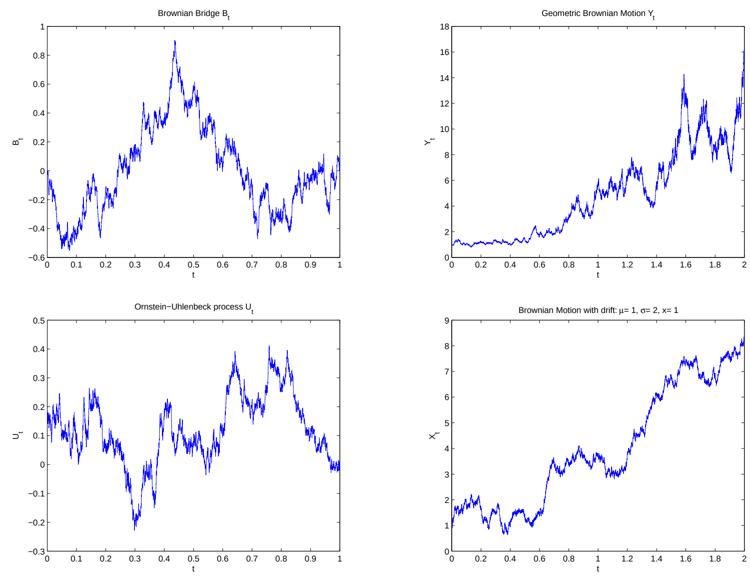

FIGURE 2. Illustration of sample paths for processes in Exercise 5.1.8.

(b) *Using Fubini's theorem 3.3.10 find* $\mathbf{E}(X_t)$ *and* $\mathbf{E}(X_t^2)$.
(c) *Is* X_t *a Gaussian process? Does it have continuous sample paths a.s.? Does it have stationary increments? Independent increments?*

EXERCISE 5.1.11. *Suppose* W_t *is a Brownian motion.*

(a) *Compute the* probability density function *of the random vector* (W_s, W_t) *for positive* $t \neq s$. *Then compute* $\mathbf{E}(W_s|W_t)$ *and* $\mathrm{Var}(W_s|W_t)$, *first for* $s > t$, *then for* $s < t$.
Hint: *Consider Example 2.4.5.*
(b) *Explain why the Brownian Bridge* $\{B_t, 0 \leq t \leq 1\}$ *has the same distribution as* $\{W_t, 0 \leq t \leq 1$, *conditioned upon* $W_1 = 0\}$ *(which is the reason for naming* B_t *a Brownian bridge).*
Hint: *Both Exercise 2.4.6 and parts of Exercise 5.1.8 may help here.*

We conclude with the *fractional Brownian motion*, another Gaussian S.P. of considerable interest in financial mathematics and analysis of computer network traffic.

EXERCISE 5.1.12. *Fix* $H \in (0,1)$. *A Gaussian stochastic process* $\{X_t,\ t \geq 0\}$, *is called a* fractional Brownian motion *(or in short, fBM), of Hurst parameter* H *if* $\mathbf{E}(X_t) = 0$ *and*

$$\mathbf{E}(X_t X_s) = \frac{1}{2}[|t|^{2H} + |s|^{2H} - |t-s|^{2H}]\ ,\ s,t \geq 0.$$

(a) *Show that an fBM of Hurst parameter* H *has a continuous modification that is also locally Hölder continuous with exponent* γ *for any* $0 < \gamma < H$.
(b) *Verify that in case* $H = 1/2$ *such modification yields the (standard) Brownian motion.*
(c) *Show the self-similarity property, whereby for any non-random* $\alpha > 0$ *the process* $\{\alpha^H X_{t/\alpha}\}$ *is an fBM of the same Hurst parameter* H.

(d) *For which values of H is the fBM a process of stationary increments and for which values of H is it a process of independent increments?*

5.2. The reflection principle and Brownian hitting times

We start with Paul Lévy's martingale characterization of the Brownian motion, stated next.

THEOREM 5.2.1 (Lévy's martingale characterization). *Suppose square-integrable MG $(X_t, \mathcal{F}_t)$ of right-continuous filtration and continuous sample path is such that $(X_t^2 - t, \mathcal{F}_t)$ is also a MG. Then, X_t is a Brownian motion.*

REMARK. The continuity of X_t is *essential* for Lévy's martingale characterization. For example, the square-integrable martingale $X_t = N_t - t$, with N_t the Poisson process of rate one (per Definition 6.2.1), is such that $X_t^2 - t$ is also a MG (see Exercise 6.2.2). Of course, almost all sample path of the Poisson process are discontinuous.

A consequence of this characterization is that a square-integrable MG with continuous sample path and unbounded increasing part is merely a *time changed Brownian motion* (c.f. [**KS97**, Theorem 3.4.6]).

PROPOSITION 5.2.2. *Suppose $(X_t, \mathcal{F}_t)$ is a square-integrable martingale with $X_0 = 0$, right-continuous filtration and continuous sample path. If the increasing part A_t in the corresponding Doob-Meyer decomposition of Theorem 4.4.7 is almost surely unbounded then $W_s = X_{\tau_s}$ is a Brownian motion, where $\tau_s = \inf\{t \geq 0 : A_t > s\}$ are $\mathcal{F}_t$-stopping times such that $s \mapsto \tau_s$ is non-decreasing and right-continuous mapping of $[0, \infty)$ to $[0, \infty)$, with $A_{\tau_s} = s$ and $X_t = W_{A_t}$.*

Our next proposition may be viewed as yet another application of Lévy's martingale characterization. In essence it states that each stopping time acts as a *regeneration point* for the Brownian motion. In particular, it implies that the Brownian motion is a strong Markov process (in the sense of Definition 6.1.21). As we soon see, this "regeneration" property is very handy for finding the distribution of certain Brownian hitting times and running maxima.

PROPOSITION 5.2.3. *If τ is a stopping time for the canonical filtration $\mathcal{G}_t$ of the Brownian motion W_t then the S.P. $X_t = W_{t+\tau} - W_\tau$ is also a Brownian motion, which is independent of the stopped σ-field $\mathcal{G}_\tau$.*

REMARK. This result is stated as [**Bre92**, Theorem 12.42], with a proof that starts with a stopping time τ taking a countable set of values and moves to the general case by approximation, using sample path continuity. Alternatively, with the help of some amount of stochastic calculus one may verify the conditions of Lévy's theorem for X_t and the filtration $\mathcal{F}_t = \sigma(W_{s+\tau} - W_\tau, 0 \leq s \leq t)$. We will detail neither approach here.

We next apply Proposition 5.2.3 for computing the probability density function of the *first hitting time* $\tau_\alpha = \inf\{t > 0 : W_t = \alpha\}$ for any fixed $\alpha > 0$. Since the Brownian motion has continuous sample path, we know that $\tau_\alpha = \min\{t > 0 : W_t = \alpha\}$ and that the maximal value of W_t for $t \in [0, T]$ is always achieved at some $t \leq T$. Further, since $W_0 = 0 < \alpha$, if $W_s \geq \alpha$ for some $s > 0$, then $W_u = \alpha$ for some $u \in [0, s]$, that is, $\tau_\alpha \leq s$ with $W_{\tau_\alpha} = \alpha$. Consequently,

$$\{\omega : W_T(\omega) \geq \alpha\} \subseteq \{\omega : \max_{0 \leq s \leq T} W_s(\omega) \geq \alpha\} = \{\omega : \tau_\alpha(\omega) \leq T\} .$$

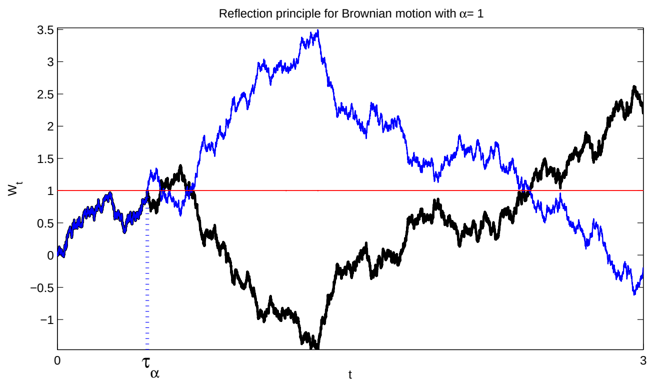

FIGURE 3. Illustration of the reflection principle for Brownian motion.

Recall that $X_t = W_{t+\tau_\alpha} - W_{\tau_\alpha}$ is a Brownian motion, independent of the random variable τ_α (which is measurable on $\mathcal{G}_{\tau_\alpha}$). In particular, the law of X_t is invariant to a sign-change, so we have the *reflection principle* for the Brownian motion, stating that

$$\begin{aligned}\mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha,\ W_T \ge \alpha) &= \mathbf{P}(\tau_\alpha \le T, X_{T-\tau_\alpha} \ge 0)\\ = \mathbf{P}(\tau_\alpha \le T, X_{T-\tau_\alpha} \le 0) &= \mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha,\ W_T \le \alpha)\,.\end{aligned}$$

Also, $\mathbf{P}(W_T = \alpha) = 0$ and we have that

$$\begin{aligned}\mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha) &= \mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha, W_T \ge \alpha) + \mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha, W_T \le \alpha)\\ &= 2\mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha, W_T \ge \alpha) = 2\mathbf{P}(W_T \ge \alpha) \qquad (5.2.1)\\ &= 2\int_{\alpha T^{-1/2}}^{\infty} e^{\frac{-x^2}{2}} \frac{dx}{\sqrt{2\pi}}\end{aligned}$$

Among other things, this shows that $\mathbf{P}(\tau_\alpha > T) \to 0$ as $T \to \infty$, hence $\tau_\alpha < \infty$ with probability one. Further, we have that the probability density function of τ_α at T is given by

$$p_{\tau_\alpha}(T) = \frac{\partial[\mathbf{P}(\tau_\alpha \le T)]}{\partial T} = 2\frac{\partial}{\partial T}\int_{\alpha T^{-1/2}}^{\infty} e^{\frac{-x^2}{2}} \frac{dx}{\sqrt{2\pi}} = \frac{\alpha}{\sqrt{2\pi}T^{3/2}} e^{-\frac{\alpha^2}{2T}}\,. \qquad (5.2.2)$$

This computation demonstrates the power of the reflection principle and more generally, that many computations for stochastic processes are the most explicit when they are done for the Brownian motion.

Our next exercise provides yet another example of a similar nature.

EXERCISE 5.2.4. *Let W_t be a Brownian motion.*

(a) *Show that $-\min_{0\le t\le T} W_t$ and $\max_{0\le t\le T} W_t$ have the same distribution which is also the distribution of $|W_T|$.*

(b) *Show that the probability α that the Brownian motion W_u attains the value zero at some $u \in (s, s+t)$ is given by $\alpha = \int_{-\infty}^{\infty} p_t(|x|)\phi_s(x)dx$, where $p_t(x) = \mathbf{P}(|W_t| \ge x)$ for $x, t > 0$ and $\phi_s(x)$ denotes the probability density of the R.V. W_s.*

Remark: *The explicit formula* $\alpha = (2/\pi)\arccos(\sqrt{s/(s+t)})$ *is obtained in* [**KT75**, page 348] *by computing this integral.*

Remark. Using a *reflection principle* type argument one gets the discrete time analog of (5.2.1), whereby the simple random walk S_n of Definition 3.1.2 satisfies for each integer $r > 0$ the identity

$$\mathbf{P}(\max_{0\le k\le n} S_k \ge r) = 2\mathbf{P}(S_n > r) + \mathbf{P}(S_n = r)\,.$$

Fixing $\alpha > 0$ and $\beta > 0$ consider the stopping time $\tau_{\beta,\alpha} = \inf\{t : W_t \ge \alpha$ or $W_t \le -\beta\}$ (for the canonical filtration of the Brownian motion W_t). By continuity of the Brownian sample path we know that $W_{\tau_{\beta,\alpha}} \in \{\alpha, -\beta\}$. Applying Doob's optional stopping theorem for the uniformly integrable stopped martingale $W_{t\wedge\tau_{\beta,\alpha}}$ of continuous sample path we get that $\mathbf{P}(W_{\tau_{\beta,\alpha}} = \alpha) = \beta/(\alpha+\beta)$ (for more details see Exercise 4.3.18).

Exercise 5.2.5. *Show that* $\mathbf{E}(\tau_{\beta,\alpha}) = \alpha\beta$ *by applying Doob's optional stopping theorem for the uniformly integrable stopped martingale* $W^2_{t\wedge\tau_{\beta,\alpha}} - t\wedge\tau_{\beta,\alpha}$.

We see that the expected time it takes the Brownian motion to exit the interval $(-\beta,\alpha)$ is finite for any finite α and β. As $\beta\uparrow\infty$, these exit times $\tau_{\beta,\alpha}$ converge monotonically to the time of reaching level α, namely $\tau_\alpha = \inf\{t > 0 : W_t = \alpha\}$. Exercise 5.2.5 implies that τ_α has infinite expected value (we can see this also directly from the formula (5.2.2) for its probability density function).

To summarize, the Brownian motion eventually reach any level, the expected time it takes for doing so is infinite, while the exit time of any finite interval has finite mean (and moreover, all its moments are finite).

Building on Exercises 4.2.9 and 5.2.5 here is an interesting fact about the *planar Brownian motion*.

Exercise 5.2.6. *The* planar Brownian motion *is an* $\mathbb{R}^2$*-valued stochastic process* $\underline{W}_t = (X_t, Y_t)$ *consisting of two independent Brownian motions* $\{X_t\}$ *and* $\{Y_t\}$*. Let* $R_t = \sqrt{X_t^2 + Y_t^2}$ *denote its distance from the origin and* $\theta_r = \inf\{t : R_t \ge r\}$ *the corresponding first hitting time for a sphere of radius* $r > 0$ *around the origin.*

(a) *Show that* $M_t = R_t^2 - 2t$ *is a martingale for* $\mathcal{F}_t = \sigma(X_s, Y_s, s \le t)$.
Hint: *Consider Proposition 2.3.17.*
(b) *Check that* $\theta_r \le \tau_{r,r} = \inf\{t : |X_t| \ge r\}$ *and that* θ_r *is a stopping time for the filtration* $\mathcal{F}_t$.
(c) *Verify that* $\{M_{t\wedge\theta_r}\}$ *is uniformly integrable and deduce from Doob's optional stopping theorem that* $\mathbf{E}[\theta_r] = r^2/2$.

5.3. Smoothness and variation of the Brownian sample path

We start with a definition of the q-th variation of a function $f(t)$ on a finite interval $t \in [a,b]$, $a < b$ of the real line, where $q \ge 1$. We shall study here only the *total variation*, corresponding to $q = 1$ and the *quadratic variation*, corresponding to $q = 2$.

Definition 5.3.1. *For any finite partition* π *of* $[a,b]$*, that is,* $\pi = \{a = t_0^{(\pi)} < t_1^{(\pi)} < \ldots < t_k^{(\pi)} = b\}$*, let* $\|\pi\| = \max_i\{t_{i+1}^{(\pi)} - t_i^{(\pi)}\}$ *denote the length of the longest* *interval in π and*

$$V_{(\pi)}^{(q)}(f) = \sum_i |f(t_{i+1}^{(\pi)}) - f(t_i^{(\pi)})|^q$$

denote the q-th variation of $f(\cdot)$ on π. The q-th variation *of $f(\cdot)$ on $[a,b]$ is then*

$$V^{(q)}(f) = \lim_{\|\pi\|\to 0} V_{(\pi)}^{(q)}(f)\,, \tag{5.3.1}$$

provided such limit exists.

We next extend this definition to continuous time stochastic processes.

Definition 5.3.2. *The* q-th variation *of a S.P. X_t on the interval $[a,b]$ is the random variable $V^{(q)}(X)$ obtained when replacing $f(t)$ by $X_t(\omega)$ in the above definition, provided the limit (5.3.1) exists (in some sense).*

The quadratic variation is affected by the smoothness of the sample path. For example, suppose that a S.P. $X(t)$ has Lipschitz sample path with probability one. Namely, there exists a random variable $L(\omega)$ which is finite almost surely, such that $|X(t) - X(s)| \le L|t-s|$ for all $t, s \in [a,b]$. Then,

$$\begin{aligned} V_{(\pi)}^{(2)}(X) &\le L^2 \sum_i (t_{i+1}^{(\pi)} - t_i^{(\pi)})^2 \\ &\le L^2\|\pi\| \sum_i (t_{i+1}^{(\pi)} - t_i^{(\pi)}) = L^2\|\pi\|(b-a)\,, \end{aligned} \tag{5.3.2}$$

converges to zero almost surely as $\|\pi\| \to 0$. So, such a S.P. has zero quadratic variation on $[a,b]$.

By considering different time intervals we view the quadratic variation as yet another stochastic process.

Definition 5.3.3. *The* quadratic variation *of a stochastic process X, denoted $V_t^{(2)}(X)$ is the non-decreasing, non-negative S.P. corresponding to the quadratic variation of X on the intervals* $[0,t]$.

Focusing hereafter on the Brownian motion, we have that,

Proposition 5.3.4. *For a Brownian motion $W(t)$, as $\|\pi\| \to 0$ we have that $V_{(\pi)}^{(2)}(W) \to (b-a)$ in 2-mean.*

Proof. Fixing a finite partition π, note that

$$\begin{aligned} \mathbf{E}[V_{(\pi)}^{(2)}(W)] &= \sum_i \mathbf{E}[(W(t_{i+1}) - W(t_i))^2] \\ &= \sum_i \mathrm{Var}(W(t_{i+1}) - W(t_i)) = \sum_i (t_{i+1} - t_i) = b - a\,. \end{aligned}$$

Similarly, by the independence of increments,

$$\begin{aligned} \mathbf{E}[V_{(\pi)}^{(2)}(W)^2] &= \sum_{i,j} \mathbf{E}[(W(t_{i+1}) - W(t_i))^2\,(W(t_{j+1}) - W(t_j))^2] \\ &= \sum_i \mathbf{E}[(W(t_{i+1}) - W(t_i))^4] \\ &\quad + \sum_{i\neq j} \mathbf{E}[(W(t_{i+1}) - W(t_i))^2]\,\mathbf{E}[(W(t_{j+1}) - W(t_j))^2] \end{aligned}$$

Since $W(t_{j+1}) - W(t_j)$ is Gaussian of mean zero and variance $(t_{j+1} - t_j)$, it follows that

$$\mathbf{E}[V^{(2)}_{(\pi)}(W)^2] = 3\sum_i (t_{i+1}-t_i)^2 + \sum_{i\neq j} (t_{i+1}-t_i)(t_{j+1}-t_j) = 2\sum_i (t_{i+1}-t_i)^2 + (b-a)^2 \,.$$

So, $\mathrm{Var}(V^{(2)}_{(\pi)}(W)) = \mathbf{E}(V^{(2)}_{(\pi)}(W)^2) - (b-a)^2 \le 2\|\pi\|(b-a) \to 0$ as $\|\pi\| \to 0$. With the mean of $V^{(2)}_{(\pi)}(W)$ being $(b-a)$ and its variance converging to zero, we have the stated convergence in 2-mean. ∎

Here are two consequences of Proposition 5.3.4.

COROLLARY 5.3.5. *The quadratic variation of the Brownian motion is the S.P. $V^{(2)}_t(W) = t$, which is the same as the* increasing process *in the Doob-Meyer decomposition of W_t^2. More generally, the quadratic variation equals the increasing process for any* square-integrable martingale *of continuous sample path and right-continuous filtration (as shown for example in* [**KS97**, Theorem 1.5.8, page 32]*).*

REMARK. Since $V^{(2)}_{(\pi)}$ are observable on the sample path, considering finer and finer partitions π_n, one may numerically estimate the quadratic variation for a given sample path of a S.P. The quadratic variation of the Brownian motion is non-random, so if this numerical estimate significantly deviates from t, we conclude that Brownian motion is not a good model for the given S.P.

COROLLARY 5.3.6. *With probability one, the sample path of the Brownian motion $W(t)$ is* not Lipschitz continuous *in any interval $[a, b]$, $a < b$.*

PROOF. Fix a finite interval $[a, b]$, $a < b$ and let Γ_L denote the set of outcomes ω for which $|W(t) - W(s)| \le L|t - s|$ for all $t, s \in [a, b]$. From (5.3.2) we see that if $\|\pi\| \le 1/(2L^2)$ then

$$\mathrm{Var}(V^{(2)}_{(\pi)}(W)) \ge \mathbf{E}[(V^{(2)}_{(\pi)}(W) - (b-a))^2 I_{\Gamma_L}] \ge \frac{(b-a)^2}{4}\mathbf{P}(\Gamma_L)\,.$$

By Proposition 5.3.4 we know that $\mathrm{Var}(V^{(2)}_{(\pi)}(W)) \to 0$ as $\|\pi\| \to 0$, hence necessarily $\mathbf{P}(\Gamma_L) = 0$. As the set Γ of outcomes for which the sample path of $W(t)$ is Lipschitz continuous is just the (countable) union of Γ_L over positive integer values of L, it follows that $\mathbf{P}(\Gamma) = 0$, as stated. ∎

We can even improve upon this negative result as following.

EXERCISE 5.3.7. *Fixing $\gamma > \frac{1}{2}$ check that by the same type of argument as above, with probability one, the sample path of the Brownian motion is* not globally Hölder continuous *of exponent γ in any interval $[a, b]$, $a < b$.*
In contrast, applying Theorem 3.3.3 verify that with probability one the sample path of the Brownian motion is locally Hölder continuous for any exponent $\gamma < 1/2$ (see part (c) of Exercise 3.3.5 for a similar derivation).

The next exercise shows that one can strengthen the convergence of the quadratic variation for $W(t)$ by imposing some restrictions on the allowed partitions.

EXERCISE 5.3.8. *Let $V^{(2)}_{(\pi)}(W)$ denote the approximation of the quadratic variation of the Brownian motion for a finite partition π of $[a, a+t]$. Combining Markov's* *inequality (for $f(x) = x^2$) and Borel-Cantelli I show that for the Brownian motion $V^{(2)}_{(\pi_n)}(W) \xrightarrow{a.s.} t$ if the finite partitions π_n are such that $\sum_n \|\pi_n\| < \infty$.*

In the next exercise, you are to follow a similar procedure, en-route to finding the quadratic variation for a Brownian motion with drift.

EXERCISE 5.3.9. *Let $Z(t) = W(t) + rt$, $t \geq 0$, where $W(t)$ is a Brownian motion and r a non-random constant.*

(a) *What is the law of $Y = Z(t+h) - Z(t)$?*
(b) *For which values of $t' < t$ and $h, h' > 0$ are the variables Y and $Y' = Z(t'+h') - Z(t')$ independent?*
(c) *Find the quadratic variation $V^{(2)}_t(Z)$ of the stochastic process $\{Z(t)\}$.* Hint: *See Exercise 5.3.15.*

Typically, the stochastic integral $I_t = \int_0^t X_s dW_s$ is first constructed in case X_t is a "simple" process (that is having sample path that are piecewise constant on non-random intervals), exactly as you do next.

EXERCISE 5.3.10. *Suppose $(W_t, \mathcal{F}_t)$ satisfies Lévy's characterization of the Brownian motion. Namely, it is a square-integrable martingale of right-continuous filtration and continuous sample path such that $(W_t^2 - t, \mathcal{F}_t)$ is also a martingale. Suppose X_t is a bounded $\mathcal{F}_t$-adapted* simple process. *That is,*

$$X_t = \eta_0 \mathbf{1}_{\{0\}}(t) + \sum_{i=0}^{\infty} \eta_i \mathbf{1}_{(t_i, t_{i+1}]}(t)\,,$$

where the non-random sequence $t_k > t_0 = 0$ is strictly increasing and unbounded (in k), while the (discrete time) S.P. $\{\eta_n\}$ is uniformly (in n and ω) bounded and adapted to $\mathcal{F}_{t_n}$. Provide an explicit formula for $A_t = \int_0^t X_u^2 du$, then show that both

$$I_t = \sum_{j=0}^{k-1} \eta_j (W_{t_{j+1}} - W_{t_j}) + \eta_k (W_t - W_{t_k}), \text{ when } t \in [t_k, t_{k+1})\,,$$

and $I_t^2 - A_t$ are martingales with respect to $\mathcal{F}_t$ and explain why this implies that $\mathbf{E} I_t^2 = \mathbf{E} A_t$ and $V^{(2)}_t(I) = A_t$.

We move from the quadratic variation $V^{(2)}$ to the *total variation* $V^{(1)}$. Note that, when $q = 1$, the limit in (5.3.1) always exists and equals the supremum over all finite partitions π.

EXAMPLE 5.3.11. *The total variation is particularly simple for* monotone functions. *Indeed, it is easy to check that if $f(t)$ is monotone then its total variation is $V^{(1)}(f) = \max_{t\in[a,b]}\{f(t)\} - \min_{t\in[a,b]}\{f(t)\}$. In particular, the total variation of monotone functions is finite on finite intervals even though the functions may well be discontinuous.*

In contrast we have that

PROPOSITION 5.3.12. *The total variation of the Brownian motion $W(t)$ is infinite with probability one.*

PROOF. Let $\alpha(h) = \sup_{a \le t \le b-h} |W(t+h) - W(t)|$. With probability one, the sample path $W(t)$ is continuous hence uniformly continuous on the closed, bounded interval $[a, b]$. Therefore, $\alpha(h) \xrightarrow{a.s.} 0$ as $h \to 0$. Let π_n divide $[a, b]$ to 2^n equal parts, so $\|\pi_n\| = 2^{-n}(b-a)$. Then,

$$\begin{aligned} V^{(2)}_{(\pi_n)}(W) &= \sum_{i=0}^{2^n-1} [W(a + (i+1)\|\pi_n\|) - W(a + i\|\pi_n\|)]^2 \\ &\le \alpha(\|\pi_n\|) \sum_{i=0}^{2^n-1} |W(a + (i+1)\|\pi_n\|) - W(a + i\|\pi_n\|)| \,. \end{aligned} \tag{5.3.3}$$

Recall Exercise 5.3.8, that almost surely $V^{(2)}_{(\pi_n)}(W) \to (b-a) < \infty$. This, together with (5.3.3) and the fact that $\alpha(\|\pi_n\|) \xrightarrow{a.s.} 0$, imply that

$$V^{(1)}_{(\pi_n)}(W) = \sum_{i=0}^{2^n-1} |W(a + (i+1)\|\pi_n\|) - W(a + i\|\pi_n\|)| \xrightarrow{a.s.} \infty,$$

implying that $V^{(1)}(W) = \infty$ with probability one, as stated. ∎

REMARK. Comparing Example 5.3.11 and Proposition 5.3.12 we have that the sample path of the Brownian motion is almost surely non-monotone on each non-empty open interval. Here is an alternative, direct proof of this result (c.f. [**KS97**, Theorem 2.9.9]).

EXERCISE 5.3.13. *Let W_t be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$.*

(a) *Let $A_n = \bigcap_{i=1}^{n} \{\omega \in \Omega : W_{i/n}(\omega) - W_{(i-1)/n}(\omega) \ge 0\}$ and $A = \{\omega \in \Omega : t \mapsto W_t(\omega)$ is non-decreasing on $[0, 1]\}$. Explain why $A = \cap_n A_n$ why $\mathbf{P}(A_n) = 2^{-n}$ and why it implies that $A \in \mathcal{F}$ and $\mathbf{P}(A) = 0$.*

(b) *Use the symmetry of the Brownian motion's sample path (per Exercise 5.1.4) to deduce that the probability that it is monotone on $[0, 1]$ is 0. Verify that the same applies for any interval $[s, t]$ with $0 \le s < t$ non-random.*

(c) *Show that, for almost every ω, the sample path $t \mapsto W_t(\omega)$ is non-monotone on any non-empty open interval.*
Hint: *Let F denote the set of ω such that $t \mapsto W_t(\omega)$ is monotone on some non-empty open interval, observing that*

$$F = \bigcup_{s,t \in \mathbb{Q}, 0 \le s < t} \{\omega \in \Omega : t \mapsto W_t(\omega) \text{ is monotone on } [s, t]\}.$$

To practice your understanding, solve the following exercises.

EXERCISE 5.3.14. *Consider the stochastic process $Y(t) = W(t)^2$, for $0 \le t \le 1$, with $W(t)$ a Brownian motion.*

(a) *Show that for any $\gamma < 1/2$ the sample path of $Y(t)$ is locally Hölder continuous of exponent γ with probability one.*

(b) *Compute $\mathbf{E}[V^{(2)}_{(\pi)}(Y)]$ for a finite partition π of $[0, t]$ to k intervals, and find its limit as $\|\pi\| \to 0$.*

(c) *Show that the total variation of $Y(t)$ on the interval $[0, 1]$ is infinite.*

EXERCISE 5.3.15.

(a) *Show that if functions* $f(t)$ *and* $g(t)$ *on* $[a,b]$ *have zero and finite quadratic variations, respectively (i.e.* $V^{(2)}(f) = 0$ *and* $V^{(2)}(g) < \infty$ *exists), then* $V^{(2)}(g+f) = V^{(2)}(g)$.

(b) *Show that if a (uniformly) continuous function* $f(t)$ *has finite total variation then* $V^{(q)}(f) = 0$ *for any* $q > 1$.

(c) *Suppose both* X_t *and* $\widetilde{A}_t$ *have continuous sample path, such that* $t \mapsto \widetilde{A}_t$ *has finite total variation on any bounded interval and* X_t *is a square-integrable martingale. Deduce that then* $V_t^{(2)}(X + \widetilde{A}) = V_t^{(2)}(X)$.

- *What follows should be omitted at first reading.*

We saw that the sample path of the Brownian motion is rather irregular, for it is neither monotone nor Lipschitz continuous at any open interval. [**Bre92**, Theorem 12.25] somewhat refines the latter conclusion, showing that with probability one the sample path is nowhere differentiable.

We saw that almost surely the sample path of the Brownian motion is Hölder continuous of any exponent $\gamma < \frac{1}{2}$ (see Exercise 5.3.7), and of no exponent $\gamma > \frac{1}{2}$. The exact *modulus of continuity* of the Brownian path is provided by P. Lévy's (1937) theorem (see [**KS97**, Theorem 2.9.25, page 114]):

$$\mathbf{P}(\limsup_{\delta \downarrow 0} \frac{1}{g(\delta)} \sup_{\substack{0 \le s,t \le 1 \\ |t-s| \le \delta}} |W(t) - W(s)| = 1) = 1,$$

where $g(\delta) = \sqrt{2\delta \log(\frac{1}{\delta})}$ for any $\delta > 0$. This means that $|W(t) - W(s)| \le Cg(\delta)$ for any $C > 1$, $\delta > 0$ small enough (possibly depending on ω), and $|t - s| < \delta$.

Many other "irregularity" properties of the Brownian sample path are known. For example ([**KS97**, Theorem 2.9.12]), for almost every ω, the set of points of local maximum for the path is countable and dense in $[0, \infty)$, and all *local maxima* are strict (recall that t is a point of local maximum of $f(\cdot)$ if $f(s) \le f(t)$ for all s in some open interval around t, and it is strict if in this interval also $f(s) < f(t)$ except at $s = t$). Moreover, almost surely, the *zero set* of points t where $W(t) = 0$, is closed, unbounded, of zero Lebesgue measure, with accumulation point at zero and no isolated points (this is [**KS97**, Theorem 2.9.6], or [**Bre92**, Theorem 12.35]). These properties further demonstrate just how wildly the Brownian path change its direction. Try to visualize a path having such properties!

We know that W_t is a Gaussian R.V. of variance t. As such it has the law of $\sqrt{t}W_1$, suggesting that the Brownian path grows like $\sqrt{t}$ as $t \to \infty$. While this is true when considering fixed, non-random times, it ignores the random fluctuations of the path. Accounting for these we obtain the following *Law of the Iterated Logarithm*,

$$\limsup_{t \to \infty} \frac{W_t(\omega)}{\sqrt{2t \log(\log t)}} = 1, \qquad \text{almost surely.}$$

Since $-W_t$ is also a Brownian motion, this is equivalent to

$$\liminf_{t \to \infty} \frac{W_t(\omega)}{\sqrt{2t \log(\log t)}} = -1, \qquad \text{almost surely.}$$

Recall that $tW_{1/t}$ is also a Brownian motion (see Exercise 5.1.4), so the law of the iterated logarithm is equivalent to

$$\limsup_{t\to 0} \frac{W_t(\omega)}{\sqrt{2t\log(\log(\frac{1}{t}))}} = 1 \quad \& \quad \liminf_{t\to 0} \frac{W_t(\omega)}{\sqrt{2t\log(\log(\frac{1}{t}))}} = -1, \qquad \text{almost surely,}$$

providing information on the behavior of W_t for *small* t (for proof, see [**Bre92**, Theorem 12.29]). An immediate consequence of the law of the iterated logarithm is the law of large numbers for Brownian motion (which you have already proved in Exercise 5.1.7).

CHAPTER 6

Markov, Poisson and Jump processes

We briefly explore in this chapter three important families of stochastic processes. We start in Section 6.1 with Markov chains (in discrete time) and processes (in continuous time). Section 6.2 deals with the particular example of the Poisson process and its relation to Exponential inter-arrivals and order statistics. As we see in Section 6.3 Markov jump processes are the natural extension of the Poisson process, sharing many of its interesting mathematical properties.

6.1. Markov chains and processes

The rich theory of Markov processes is the subject of many text books and one can easily teach a full course on this subject alone. Thus, we limit ourselves here to the definition of Markov processes and to their most basic properties. For more on Markov chains and processes, see [**Bre92**, Section 7] and [**Bre92**, Section 15], respectively. Alternatively, see [**GS01**, Chapter 6] for Markov chains and processes with countable state space.

As usual we start with the simpler case of discrete time stochastic processes.

Definition 6.1.1. *A discrete time stochastic process* $\{X_n,\ n = 0, 1, \ldots\}$ *with each R.V.* X_n *taking values in a measurable space* $(\mathbb{S}, \mathcal{B})$ *is called a* Markov chain *if for every non-negative integer* n *and any set* $A \in \mathcal{B}$*, almost surely* $\mathbf{P}(X_{n+1} \in A|X_0, \ldots, X_n) = \mathbf{P}(X_{n+1} \in A|X_n)$*. The set* $\mathbb{S}$ *is called the* state space *of the Markov chain.*

Remark. Definition 6.1.1 is equivalent to the identity

$$\mathbf{E}(f(X_{n+1})|X_1, \ldots, X_n) = \mathbf{E}(f(X_{n+1})|X_n)$$

holding almost surely for each bounded measurable function $f(\cdot)$.

Definition 6.1.2. *A* homogeneous Markov chain *is a Markov chain that has a modification for which* $\mathbf{P}(X_{n+1} \in A|X_n)$ *does not depend on* n *(except via the value of* X_n*).*

To simplify the presentation we assume hereafter that $\mathbb{S}$ is a closed subset of $\mathbb{R}$ and $\mathcal{B} = \mathcal{B}_{\mathbb{S}}$ is the corresponding restriction of the Borel σ-field to $\mathbb{S}$.
The distribution of a homogeneous Markov chain is determined by its stationary transition probabilities as stated next.

Definition 6.1.3. *To each homogeneous Markov chain* $\{X_n\}$ *with values in a closed subset* $\mathbb{S}$ *of* $\mathbb{R}$ *correspond its* stationary transition probabilities $p(A|x)$ *such that* $p(\cdot|x)$ *is a probability measure on* $(\mathbb{S}, \mathcal{B})$ *for any* $x \in \mathbb{S}$*;* $p(A|\cdot)$ *is measurable on* $\mathcal{B}$ *for any* $A \in \mathcal{B}$*, and almost surely* $p(A|X_n) = \mathbf{P}(X_{n+1} \in A|X_n)$ *for all* $n \geq 0$*.*

Many S.P. of interest are homogeneous Markov chains. One example that we have already seen is the random walk $S_n = \sum_{i=1}^{n} \xi_i$, where ξ_i are i.i.d. random variables. Indeed, $S_{n+1} = S_n + \xi_{n+1}$ with ξ_{n+1} independent of $\mathcal{G}_n = \sigma(S_0, \dots, S_n)$, hence $\mathbf{P}(S_{n+1} \in A|\mathcal{G}_n) = \mathbf{P}(S_n + \xi_{n+1} \in A|S_n)$. With ξ_{n+1} having the same law as ξ_1, we thus get that $\mathbf{P}(S_n + \xi_{n+1} \in A|S_n) = p(A|S_n)$ for the stationary transition probabilities $p(A|x) = \mathbf{P}(\xi_1 \in \{y - x : y \in A\})$. By similar reasoning we see that another example of a homogeneous Markov chain is the branching process Z_n of Definition 4.6.1, with stationary transition probabilities $p(A|x) = \mathbf{P}(\sum_{j=1}^{x} N_j \in A)$ for integer $x \geq 1$ and $p(A|0) = \mathbf{1}_{0\in A}$.

Remark. Many Markov chains are not martingales. For example, a sequence of independent variables X_n is a Markov chain, but unless $\mathbf{P}(X_n = c) = 1$ for some c non-random and all n, it is not a martingale (for $\mathbf{E}[X_{n+1}|X_0, \dots, X_n] = \mathbf{E}X_{n+1} \neq X_n(\omega)$). Likewise, many martingales do not have the Markov property. For example, the sequence $X_n = X_0(1 + S_n)$ with X_0 uniformly chosen in $\{1, 3\}$ independently of the simple random walk S_n of zero-mean, is a martingale, since

$$\mathbf{E}[X_{n+1}|X_0, \dots, X_n] = X_n + X_0\mathbf{E}[\xi_{n+1}|X_0, \dots, X_n] = X_n \ ,$$

but not a Markov chain, since X_0^2 is not a measurable on $\sigma(X_n)$, hence

$$\mathbf{E}[X_{n+1}^2|X_0, \dots, X_n] = X_n^2 + X_0^2 \neq X_n^2 + \mathbf{E}[X_0^2|X_n] = \mathbf{E}[X_{n+1}^2|X_n] \, .$$

Let $\mathbf{P}_x$ denote the law of the homogeneous Markov chain starting at $X_0 = x$. For example, in the context of the random walk we normally start at $S_0 = 0$, i.e. consider the law $\mathbf{P}_0$, whereas for the branching process example we normally take $Z_0 = 1$, hence in this context consider the law $\mathbf{P}_1$.

Whereas it suffices to know the stationary transition probability p to determine $\mathbf{P}_x$ for any given $x \in \mathbb{S}$, we are often interested in settings in which X_0 is a R.V. To this end, we next define the initial distribution associated with the Markov chain.

Definition 6.1.4. *The* initial distribution *of a Markov chain is the probability measure* $\pi(A) = \mathbf{P}(X_0 \in A)$ *on* $(\mathbb{S}, \mathcal{B})$.

Indeed, by the tower property (with respect to $\sigma(X_0)$), it follows from our definitions that for any integer $k \geq 0$ and a (measurable) set $B \in \mathcal{B}^{k+1}$,

$$\mathbf{P}((X_0, X_1, \dots, X_k) \in B) = \int \mathbf{P}_x((x, X_1, \dots, X_k) \in B)\pi(dx) \, ,$$

where $\int h(x)\pi(dx)$, or more generally $\int h(x)p(dx|y)$, denote throughout the Lebesgue integral with respect to x (with y, if present, being an argument of the resulting function). For example, the preceding formula means the expectation (as in Definition 1.2.19), of the measurable function $h(x) = \mathbf{P}_x((x, X_1, \dots, X_k) \in B)$ when x is an $\mathbb{S}$-valued R.V. of law $\pi(\cdot)$.

While we do not pursue this direction further, computations of probabilities of events of interest for a Markov chain are easier when $\mathbb{S}$ is a countable set. In this case, $\mathcal{B} = 2^{\mathbb{S}}$ and for any $A \subseteq \mathbb{S}$

$$p(A|x) = \sum_{y \in A} p(y|x) \, ,$$

so the stationary transition probabilities are fully determined by $p(y|x) \geq 0$ such that $\sum_{y\in\mathbb{S}} p(y|x) = 1$ for each $x \in \mathbb{S}$, and all our Lebesgue integrals are merely sums. For example, this is what we have for the simple random walk (with $\mathbb{S}$ the set of integers, and $p(x+1|x) = \mathbf{P}(\xi_1 = 1) = 1 - p(x-1|x)$ for any integer x), or the branching process (with $\mathbb{S}$ the set on non-negative integers). Things are even simpler when $\mathbb{S}$ is a finite set, in which case identifying $\mathbb{S}$ with the set $\{1, \dots, m\}$ for some $m < \infty$, we view $p(y|x)$ as the (x, y)-th entry of an $m \times m$ dimensional *transition probability matrix*, and use matrix theory for evaluating all probabilities of interest.

EXERCISE 6.1.5. *Consider the probability space corresponding to a sequence of independent rolls of a fair six-sided die. Determine which of the following S.P. is then a homogeneous Markov chain and for each of these specify its state space and stationary transition probabilities.*

(a) X_n*: the largest number rolled in* n *rolls.*
(b) N_n*: the number of 6's in* n *rolls.*
(c) C_n*: at time* n*, the number of rolls since the most recent 6.*
(d) B_n*: at time* n*, the number of rolls until the next 6.*

Here is a glimpse of the rich theory of Markov chains of countable state space (which is beyond the scope of this course).

EXERCISE 6.1.6. *Suppose* $\{X_n\}$ *is a homogeneous Markov chain of countable state space* $\mathbb{S}$ *and for each* $x \in \mathbb{S}$ *let* $\rho_x = \mathbf{P}_x(T_x < \infty)$ *for the (first) return time* $T_x = \inf\{n \geq 1 : X_n = x\}$*. We call* $x \in \mathbb{S}$ *a* recurrent state *if* $\rho_x = 1$*. That is, starting at* x *the Markov chain returns to state* x *in a finite time w.p.1. (and we call* $x \in \mathbb{S}$ *a* transient state *if* $\rho_x < 1$*). Let* $N_x = \sum_{n=1}^{\infty} I_{\{X_n = x\}}$ *count the number of visits to state* x *by the Markov chain (excluding time zero).*

(a) *Show that* $\mathbf{P}_x(N_x = k) = \rho_x^k(1 - \rho_x)$ *and hence* $\mathbf{E}_x(N_x) = \rho_x/(1 - \rho_x)$*.*
(b) *Deduce that* $x \in \mathbb{S}$ *is a recurrent state if and only if* $\sum_n \mathbf{P}_x(X_n = x) = \infty$*.*

EXERCISE 6.1.7. *Let* $\underline{X}_n$ *be a symmetric simple random walk in* $\mathbb{Z}^d$*. That is,* $\underline{X}_n = (X_n^{(1)}, \dots, X_n^{(d)})$ *where* $X_n^{(i)}$ *are for* $i = 1, \dots, d$ *independent symmetric simple one-dimensional random walks. That is, from the current state* $x \in \mathbb{Z}^d$ *the walk moves to one of the* 2^d *possible neighboring states* $(x_1 \pm 1, \dots, x_d \pm 1)$ *with probability* 2^{-d}*, independent of all previous moves. Show that for* $d = 1$ *or* $d = 2$ *the origin* $(0, \dots, 0)$ *is a recurrent state of this Markov chain, while for any* $d \geq 3$ *it is a transient state.*
Hint: *Observe that* $\underline{X}_n$ *can only return to the origin when* n *is even. Then combine Exercise 6.1.6 with the approximation* $\binom{2n}{n}2^{-2n} = (4\pi n)^{-1/2}(1 + o(1))$*.*

A second setting in which computations are more explicit is when $\mathbb{S} = \mathbb{R}$ (or a closed interval thereof), and for each x the stationary transition probability $p(\cdot|x)$ has a density $p(y|x) \geq 0$ (such that $\int p(y|x)dy = 1$). Such density $p(y|x)$ is often called the *(stationary) transition probability kernel*, as in this case for any bounded Borel function $h(\cdot)$

$$\mathbf{E}_x(h(X_1)) = \int h(y)dp(y|x) = \int h(y)p(y|x)dy\,,$$

so probabilities are computed by iterated one-dimensional Riemann integrals involving the integration kernel $p(\cdot|\cdot)$.

Turning hereafter to the general setting, with $\mathcal{G}_n = \sigma(X_0, \dots, X_n)$ denoting its canonical filtration, it is not hard to check that any homogeneous Markov chain $\{X_n\}$ has the *Markov property*,

$$\mathbf{P}((X_\tau, X_{\tau+1}, \dots, X_{\tau+k}) \in B|\mathcal{G}_\tau) = \mathbf{P}_{X_\tau}((X_0, \dots, X_k) \in B)\,, \tag{6.1.1}$$

holding almost surely for any integer $k \geq 0$, non-random $\tau \geq 0$ and $B \in \mathcal{B}^{k+1}$.

We state next the strong Markov property, one of the most important features of Markov chains.

PROPOSITION 6.1.8 (Strong Markov Property). *Let X_n be a homogeneous Markov chain. Then, (6.1.1) holds for any almost surely finite stopping time τ with respect to its canonical filtration $\mathcal{G}_n$, with $\mathcal{G}_\tau$ denoting the corresponding stopped σ-field of Definition 4.3.21 (c.f.* [**Bre92**, Proposition 7.8]*).*

The regular Markov property corresponds to the special case of τ non-random. Another special case is $\mathbf{P}(X_{\tau+1} \in A|\mathcal{G}_\tau) = p(A|X_\tau)$ almost surely (take $k = 1$ in the proposition).

We conclude with another example of a Markov chain, the first-order auto-regressive (AR) process, common in time series modeling.

EXERCISE 6.1.9. *Let ξ_i, $i = 0, 1, 2, \dots$ be a Gaussian stochastic process with ξ_i independent, each having zero-mean and variance one. Fixing non-random constants α and β, let $X_0 = \beta\xi_0$ and $X_n = \alpha X_{n-1} + \xi_n$ for any $n \geq 1$.*

(a) *Check that X_n is a homogeneous Markov chain with state space $\mathbb{S} = \mathbb{R}$. Provide its stationary transition probabilities and the characteristic function of its initial distribution.*
(b) *Check that X_n is adapted to the filtration $\mathcal{F}_n = \sigma(\xi_k, k = 0, \dots, n)$.*
(c) *Show that X_n is a zero-mean, discrete time, Gaussian stochastic process.*
(d) *Show that the auto-correlation $\rho(n, m) = \mathbf{E}(X_n X_m)$ is such that $\rho(n, n+\ell) = \alpha^\ell \rho(n, n)$ for any $n, \ell \geq 0$.*
(e) *Find a recursion formula for $\rho(n, n)$ in terms of $\rho(n-1, n-1)$ for $n \geq 1$. Fixing $-1 < \alpha < 1$, find the value of $\beta > 0$ for which the stochastic process $\{X_n\}$ is stationary. What would happen if $|\alpha| \geq 1$?*

We move now to deal with continuous time, starting with the definition of a Markov process.

DEFINITION 6.1.10. *A stochastic process $X(t)$ indexed by $t \in [0, \infty)$ and taking values in a measurable space $(\mathbb{S}, \mathcal{B})$ is called a* Markov process *if for any $t, u \geq 0$ and $A \in \mathcal{B}$ we have that almost surely*

$$\mathbf{P}(X(t+u) \in A|\sigma(X(s), s \leq t)) = \mathbf{P}(X(t+u) \in A|X(t)).$$

Equivalently, we call $X(t)$ a Markov process if for any $t, u \geq 0$ and any bounded measurable function $f(\cdot)$ on $(\mathbb{S}, \mathcal{B})$, almost surely,

$$\mathbf{E}(f(X(t+u))|\sigma(X(s), s \leq t)) = \mathbf{E}(f(X(t+u))|X(t)).$$

The set $\mathbb{S}$ is called the state space *of the Markov process.*

Similar to Definition 6.1.4 the *initial distribution* of a Markov process is the probability measure $\pi(A) = \mathbf{P}(X(0) \in A)$. Taking hereafter $\mathbb{S}$ as a closed subset of $\mathbb{R}$ and $\mathcal{B} = \mathcal{B}_{\mathbb{S}}$, as in the discrete time setting, the law $\mathbf{P}_x$ of the process starting at non-random $X(0) = x$ is determined by its transition probabilities, which we define next.

DEFINITION 6.1.11. *For each $t > s$ and fixed $x \in \mathbb{S}$ there exists a probability measure $p_{t,s}(\cdot|x)$ on $(\mathbb{S}, \mathcal{B})$ such that for each fixed $A \in \mathcal{B}$, the function $p_{t,s}(A|\cdot)$ is measurable and*

$$\mathbf{P}(X(t) \in A|X(s)) = \mathbf{E}(I_{X(t)\in A}|X(s)) = p_{t,s}(A|X(s)) \qquad a.s. \tag{6.1.2}$$

Such a collection $p_{t,s}(A|x)$ is called the transition probabilities *for the Markov process $\{X(t)\}$.*

The transition probabilities $p_{t,s}(\cdot|\cdot)$, together with the initial distribution $\pi(\cdot)$ determine the *finite dimensional distributions* of the Markov process. Indeed, for each non-negative integer k, every $0 = t_0 < t_1 < \cdots < t_k$ and $A_0, \ldots, A_k \in \mathcal{B}$ we have that

$$\begin{aligned} &\mathbf{P}(X(t_k) \in A_k, \ldots, X(t_0) \in A_0) \\ &= \int_{A_0} \cdots \int_{A_k} p_{t_k,t_{k-1}}(dx_k|x_{k-1}) \cdots p_{t_1,t_0}(dx_1|x_0)\pi(dx_0) \end{aligned} \tag{6.1.3}$$

(using Lebesgue's integrals in the above). We note in passing that as for Markov chains, computationally things are simpler when $\mathbb{S}$ is a finite or countable set, for which we can replace the integrals by sums in (6.1.3) and everywhere else, or when $\mathbb{S} = \mathbb{R}$ and $p_{t,s}(\cdot|x)$ have densities, in which case we can use the standard Riemann integrals.

We turn to deal with the constraints imposed on the collection $p_{t,s}(\cdot|\cdot)$ for different values of t and s. To this end, combining the tower property for $\mathcal{G}_u = \sigma(X(v), v \le u)$ at a fixed $u \in (s,t)$, with the definition of a Markov process, convince yourself that (6.1.2) implies that the *Chapman-Kolmogorov equations*

$$p_{t,s}(A|x) = \int p_{t,u}(A|y)p_{u,s}(dy|x)\,, \tag{6.1.4}$$

hold almost surely in x chosen according to the law of $X(s)$, for each $A \in \mathcal{B}$ and $t > u > s \ge 0$ (c.f. [**Bre92**, Proposition 15.9] where these equations are derived from (6.1.3)). In case you wonder, the integral on the right side of (6.1.4) stands for the expectation (as in Definition 1.2.19), of the measurable function $h(y) = p_{t,u}(A|y)$ when y is an $\mathbb{S}$-valued R.V. of law $p_{u,s}(\cdot|x)$ (and with A and t, u, s, x being fixed parameters). As before, such integrals are more explicit in case $\mathbb{S}$ is finite or countable, or when the relevant transition probabilities $p_{u,s}(\cdot|x)$ have densities.

For modeling purposes it is more effective to specify the collection $p_{t,s}(A|x)$ rather than starting with the Markov process $X(t)$ and its probability space. When doing so, we obviously wish to know that there exists a Markov process whose transition probabilities coincide with those we specify. As we already saw, a necessary condition is that the Chapman-Kolmogorov equations hold almost surely in x. Things are much nicer when these equations hold for all x, motivating our next definition.

DEFINITION 6.1.12. *We say that* $p_{t,s}(A|x)$ *for* $t > s \geq 0$, $x \in \mathbb{S}$ *and* $A \in \mathcal{B}$ *are* regular transition probabilities *if* $p_{t,s}(\cdot|x)$ *are probability measures on* $(\mathbb{S}, \mathcal{B})$*, the functions* $p_{t,s}(A|\cdot)$ *are Borel measurable and the Chapman-Kolmogorov equations (6.1.4) hold for every* $t > u > s \geq 0$, $x \in \mathbb{S}$ *and* $A \in \mathcal{B}$.

Indeed, as we state next, for any regular transition probabilities and any probability measure π on $(\mathbb{S}, \mathcal{B})$ there exists a Markov process with these transition probabilities and having initial distribution π (c.f. [**Bre92**, Theorem 5.11] for the proof). In this respect, the Chapman-Kolmogorov equations characterize all structural relations between the different probability measures $p_{t,s}(\cdot|\cdot)$.

THEOREM 6.1.13. *Given regular transition probabilities* $p_{t,s}(\cdot|\cdot)$ *and a probability measure* $\pi(\cdot)$ *on* $(\mathbb{S}, \mathcal{B})$*, the identities (6.1.3) define the finite dimensional distributions of a Markov process* $\{X(t)\}$ *having the specified transition probabilities and the initial distribution* π.

Similarly to Markov chains, we focus next on the (time) homogeneous Markov processes, associated with stationary regular transition probabilities.

DEFINITION 6.1.14. *A* homogeneous Markov process *is a Markov process with regular transition probabilities of the form* $p_{t,s}(\cdot|\cdot) = p_{t-s}(\cdot|\cdot)$*, which in turn are called* stationary regular transition probabilities.

Setting $\tilde{t} = t - u$ and $\tilde{s} = u - s$ it is easy to verify that for a homogeneous Markov process the Chapman-Kolmogorov equations (6.1.4) simplify to

$$p_{\tilde{t}+\tilde{s}}(A|x) = \int p_{\tilde{t}}(A|y) p_{\tilde{s}}(dy|x) \quad \forall \tilde{t}, \tilde{s} \geq 0.$$

Further, any such process satisfies the *regular Markov property*.

PROPOSITION 6.1.15. *Suppose* $X(t)$ *is a homogeneous Markov process, with* $\mathcal{G}_t$ *denoting its canonical filtration* $\sigma(X(s), s \leq t)$ *and* $\mathbf{P}_x$ *the law of the process starting at* $X(0) = x$*. Then, any such process has the* regular Markov property*. That is,*

$$\mathbf{P}_x(X(\cdot + \tau) \in \Gamma | \mathcal{G}_\tau) = \mathbf{P}_{X(\tau)}(X(\cdot) \in \Gamma), \qquad a.s. \tag{6.1.5}$$

for any $x \in \mathbb{R}$, non-random $\tau \geq 0$ *and* Γ *in the cylindrical* σ*-field* $\mathcal{B}^{[0,\infty)}$ *of Definition 3.1.15.*

As you show next, being a Markov process is a property that is invariant under invertible non-random mappings of the state space as well as under invertible monotone time mappings.

EXERCISE 6.1.16. *Let* $\{X_t, t \geq 0\}$ *be a Markov process of state space* $\mathbb{S}$*. Suppose* $h_t : \mathbb{S} \to \mathbb{S}'$ *are measurable and invertible for any fixed* $t \geq 0$ *and* $g : [0, \infty) \to [0, \infty)$ *is invertible and strictly increasing.*

(a) *Verify that* $Y_t = h_t(X_{g(t)})$ *is a Markov process.*
(b) *Show that if* $\{X_t\}$ *is a homogeneous Markov process then so is* $\{h(X_t)\}$.

A host of particularly simple Markov processes is provided by our next proposition.

PROPOSITION 6.1.17. *Every continuous time stochastic process of independent increments is a Markov process. Further, every continuous time S.P. of stationary independent increments is a homogeneous Markov process.*

For example, Proposition 6.1.17 implies that the Brownian motion is a homogeneous Markov process. It corresponds to initial distribution $X(0) = 0$ and having the stationary regular transition probabilities

$$p_t(A|x) = \int_A \frac{e^{-(y-x)^2/2t}}{\sqrt{2\pi t}} dy \tag{6.1.6}$$

(note that we have here a stationary transition kernel $e^{-(y-x)^2/2t}/\sqrt{2\pi t}$).

Remark. To re-cap, we have seen three main ways of showing that a S.P. $\{X_t, t \geq 0\}$ is a Markov process:

(a) Computing $\mathbf{P}(X_{t+h} \in A|\mathcal{G}_t)$ directly and checking that it only depends on X_t (and not on X_s for $s < t$), per Definition 6.1.10.
(b) Showing that the process has independent increments and applying Proposition 6.1.17.
(c) Showing that it is an invertible function of another Markov process, and appealing to Exercise 6.1.16.

Our choice of name for $p_t(\cdot|\cdot)$ is motivated in part by the following fact.

Proposition 6.1.18. *If a Markov process or a Markov chain is also a stationary process, then it is a homogeneous Markov process, or Markov chain, respectively.*

Note however that many homogeneous Markov processes and Markov chains are not stationary processes. Convince yourself that among such examples are the Brownian motion (in continuous time) and the random walk (in discrete time).

Solve the next two exercises to practice your understanding of the definition of Markov processes.

Exercise 6.1.19. *Let $B_t = W_t - \min(t,1)W_1$, $Y_t = e^{W_t}$ and $U_t = e^{-t/2}W_{e^t}$, where W_t is a Brownian motion.*

(a) *Determine which of the S.P. W_t, B_t, U_t and Y_t is a Markov process for $t \leq 1$ and among those, which are also time homogeneous.*
Hint: *Consider part (a) of Exercise 5.1.9.*
(b) *Provide an example (among these S.P.) of a homogeneous Markov process whose increments are neither independent nor stationary.*
(c) *Provide an example (among these S.P.) of a Markov process of stationary increments, which is not a homogeneous Markov process.*

Exercise 6.1.20. *Explain why $Z_t = W_t + rt$, $t \geq 0$, with W_t a Brownian motion and r a non-random constant, is a homogeneous Markov process. Provide the state space $\mathbb{S}$, the initial distribution $\pi(\cdot)$ and the stationary regular transition probabilities $p_t(A|x)$ for this process.*

The homogeneous Markov chains are fully characterized by the initial distributions and the (one-step) transition probabilities $p(\cdot|\cdot)$. In contrast, we need to specify the transition probabilities $p_t(\cdot|\cdot)$ for all $t > 0$ in order to determine all distributional properties of the associated homogeneous Markov process. While we shall not do so, in view of the Chapman-Kolmogorov relationship, using functional analysis one may often express $p_t(\cdot|\cdot)$ in terms of a single operator, which is called the "generator" of the Markov process. For example, the generator of the Brownian motion is closely related to the heat equation, hence the reason that many computations can then be explicitly done via the theory of PDE.

Similar to the case of Markov chains, we seek to have the *strong* Markov property for a homogeneous Markov process. Namely,

DEFINITION 6.1.21. *A homogeneous Markov process is called a strong Markov process if (6.1.5) holds also for any almost surely finite stopping time τ with respect to its canonical filtration $\mathcal{G}_t$.*

Recall that we have already seen in Proposition 5.2.3 that

COROLLARY 6.1.22. *The Brownian motion is a strong Markov process.*

Unfortunately, the theory of Markov processes is more involved than that of Markov chains and in particular, not all homogeneous Markov processes have the *strong* Markov property. Indeed, as we show next, even having also continuous sample path does not imply the strong Markov property.

EXAMPLE 6.1.23. *With X_0 independent of the Brownian motion W_t, consider the S.P. $X_t = X_0 + W_t I_{\{X_0 \neq 0\}}$ of continuous sample path. Noting that $I_{X_0=0} = I_{X_t=0}$ almost surely (as the difference occurs on the event $\{\omega : W_t(\omega) = -X_0(\omega) \neq 0\}$ which is of zero probability), by the independence of increments of W_t, hence of X_t in case $X_0 \neq 0$, we have that almost surely,*

$$\begin{aligned} \mathbf{P}(X_{t+u} \in A|\sigma(X_s, s \leq t)) &= I_{0\in A} I_{X_0=0} + \mathbf{P}(W_{t+u} - W_t + X_t \in A|X_t) I_{X_0 \neq 0} \\ &= I_{0\in A} I_{X_t=0} + p_t(A|X_t) I_{X_t \neq 0}\,, \end{aligned}$$

for $p_t(\cdot|x)$ of (6.1.6). Consequently, X_t is a homogeneous Markov process (regardless of the distribution of X_0), whose stationary regular transition probabilities are given by $p_t(\cdot|x)$ of (6.1.6) for $x \neq 0$ while $p_t(A|0) = I_{0\in A}$. By Proposition 6.1.15 X_t satisfies the regular Markov. However, this process does not satisfy the strong Markov property. For example, (6.1.5) does not hold for the almost surely finite stopping time $\tau = \inf\{t \geq 0 : X_t = 0\}$ and $\Gamma = \{x(\cdot) : x(1) > 0\}$ (in which case its left side is $\frac{1}{2}\mathbf{1}_{x\neq 0}$ whereas its right side is zero).

Our next proposition further helps in clarifying where the extra difficulty comes from.

PROPOSITION 6.1.24. *The Markov property (6.1.5) holds for any stopping time τ (with respect to the canonical filtration of the homogeneous Markov process $\{X(t)\}$), provided τ assumes at most a countable number of non-random values (for a proof, see* [**Bre92**, Proposition 15.19]*).*

Indeed, any stopping time for a Markov chain assumes at most a countable number of values $\{0, 1, 2, \ldots\}$, hence the reason that every homogeneous Markov chain has the strong Markov property.

In the following exercise you use the strong Markov property to compute the probability that a Brownian motion that starts at $x \in (c, d)$ reaches level d before it reaches level c (i.e., the event $W_{\tau_{a,b}} = b$ for $\tau_{a,b}$ of Exercise 4.3.18 with $b = d - x$ and $a = x - c$).

EXERCISE 6.1.25. *Consider the stopping time $\tau = \inf\{t \geq 0 : X_t \geq d \text{ or } X_t \leq c\}$ and the law $\mathbf{P}_x$ of the Markov process $X_t = W_t + x$, where W_t is a Brownian motion and $x \in (c, d)$ non-random.*

(a) *Using the strong Markov property of X_t show that $u(x) = \mathbf{P}_x(X_\tau = d)$ is an* harmonic function, *namely $u(x) = (u(x+r) + u(x-r))/2$ for any $c \leq x - r < x < x + r \leq d$, with boundary conditions $u(c) = 0$ and $u(d) = 1$.*

(b) *Check that $v(x) = (x-c)/(d-c)$ is an harmonic function satisfying the same boundary conditions as $u(x)$. Since boundary conditions at $x = c$ and $x = d$ uniquely determine the value of harmonic function in (c, d) (a fact you do not need to prove), you thus showed that $\mathbf{P}_x(X_\tau = d) = (x-c)/(d-c)$.*

6.2. Poisson process, Exponential inter-arrivals and order statistics

Following [**Bre92**, Chapter 14.6], we are going to consider throughout continuous time stochastic processes N_t, $t \geq 0$ satisfying the following condition.

CONDITION. $\boxed{C_0}$ *Each sample path $N_t(\omega)$ is piecewise constant, non-decreasing, right continuous, with $N_0(\omega) = 0$, all jump discontinuities are of size one, and there are infinitely many of them.*

Associated with each sample path $N_t(\omega)$ satisfying $\boxed{C_0}$ are the jump times $0 = T_0 < T_1 < \cdots < T_n < \cdots$ such that $T_k = \inf\{t \geq 0 : N_t \geq k\}$ for each k, or equivalently

$$N_t = \sup\{k \geq 0 : T_k \leq t\}.$$

In applications we find such N_t as counting the number of discrete events occurring in the interval $[0, t]$ for each $t \geq 0$, with T_k denoting the arrival or occurrence time of the k-th such event. For this reason processes like N_t are also called *counting processes*.

Recall Example 1.1.4 that a random variable N has the Poisson(μ) law if

$$\mathbf{P}(N = k) = \frac{\mu^k}{k!} e^{-\mu}, \quad k \geq 0, \quad \text{integer},$$

and that a stochastic process N_t has *independent increments* if the random variable $N_{t+h} - N_t$ is independent of $\sigma(N_s : 0 \leq s \leq t)$ for any $h > 0$ and $t \geq 0$.

We define next the *Poisson process*. To this end, we set the following condition:

CONDITION. $\boxed{C_1}$ *For any k and any $0 < t_1 < \cdots < t_k$, the increments N_{t_1}, $N_{t_2} - N_{t_1}$, $\ldots$, $N_{t_k} - N_{t_{k-1}}$, are independent random variables and for some $\lambda > 0$ and all $t > s \geq 0$, the increment $N_t - N_s$ has the* Poisson($\lambda(t-s)$) *law.*

Equipped with Conditions $\boxed{C_0}$ and $\boxed{C_1}$ we have

DEFINITION 6.2.1. *Among the processes satisfying* $\boxed{C_0}$ *the* Poisson Process *is the unique S.P. having also the property* $\boxed{C_1}$.

Thus, the Poisson process has independent increments, each having a Poisson law, where the parameter of the count $N_t - N_s$ is proportional to the length of the corresponding interval $[s, t]$. The constant of proportionality λ is called the *rate* or *intensity* of the Poisson process.

In particular, it follows from our definition that $M_t = N_t - \lambda t$ is a MG (see Proposition 4.2.3). As you show next, this provides an example of a Doob-Meyer

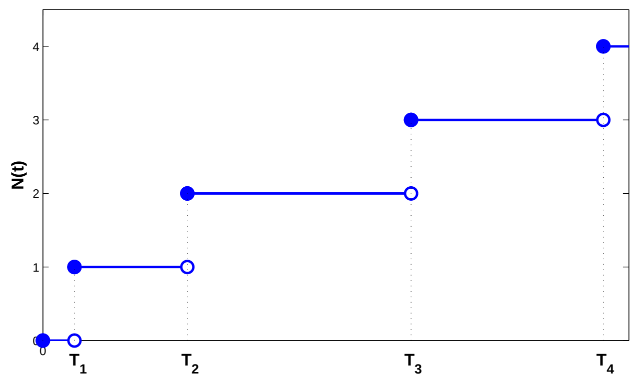

FIGURE 1. A sample path of a Poisson process.

decomposition with a continuous increasing process for a martingale of discontinuous sample path. It also demonstrates that continuity of sample path is essential for the validity of both *Lévy's martingale characterization* and Proposition 5.2.2 about the *time changed Brownian motion*.

EXERCISE 6.2.2. *Show that the martingale $M_t = N_t - \lambda t$ is square-integrable, and such that $M_t^2 - \lambda t$ is a martingale for the (right-continuous) filtration $\sigma(N_s, s \leq t)$.*

In a similar manner, the independence of increments of the Poisson process provides us with the corresponding exponential martingales (as it did for the Brownian motion). Here we get the martingales $L_t = e^{\theta N_t - \beta t}$, for some $\beta = \beta(\theta, \lambda)$ and all θ (with $\beta(\cdot)$ computed by you in part (a) of Exercise 6.2.9).

REMARK 6.2.3. It is possible to define in a similar manner the counting process for discrete events on $\mathbb{R}^d$, any integer $d \geq 2$. This is done by assigning random integer counts N_A to Borel subsets A of $\mathbb{R}^d$ in an additive manner (that is $N_{A\cup B} = N_A + N_B$ whenever A and B are disjoint). Such processes are called *point processes*, so the Poisson process is perhaps the simplest example of a point process.

Just as the Brownian motion has many nice properties, so does the Poisson process, starting with the following characterization of this process (for its proof, see [**Bre92**, Theorem 14.23]).

PROPOSITION 6.2.4. *The Poisson processes are the only stochastic processes with* stationary independent increments *(see Definition 3.2.27), that satisfy* $\boxed{C_0}$.

Next, note that the Poisson process is a homogeneous Markov process, whose state space $\mathbb{S}$ is $\{0, 1, 2, \ldots\}$ (see Proposition 6.1.17), with the initial distribution $N_0 = 0$ and the stationary regular transition probabilities

$$p_t(x+k|x) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k, x \geq 0, \quad \text{integers.}$$

Furthermore, we have, similarly to Corollary 6.1.22, that

PROPOSITION 6.2.5. *The Poisson process is a strong Markov process.*

REMARK. Note that the Poisson process is yet another example of a process with stationary independent increments that is clearly not a stationary process.

Another way to characterize the Poisson process is by the joint distribution of the jump (arrival) times $\{T_k\}$. To this end, recall that

PROPOSITION 6.2.6 (Memoryless property of the Exponential law). *We say that a random variable T has* Exponential(λ) *law if* $\mathbf{P}(T > t) = e^{-\lambda t}$*, for all $t \geq 0$ and some $\lambda > 0$. Except for the trivial case of $T = 0$ w.p.1. these are the only laws for which* $\mathbf{P}(T > x + y | T > y) = \mathbf{P}(T > x)$*, for all $x, y \geq 0$.*

We check only the easy part of the proposition, that is, taking T of Exponential(λ) law we have for any $x, y \geq 0$,

$$\mathbf{P}(T > x + y | T > y) = \frac{\mathbf{P}(T > x + y)}{\mathbf{P}(T > y)} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} = e^{-\lambda x} = \mathbf{P}(T > x)\,.$$

EXERCISE 6.2.7. *For T of Exponential(λ) law use integration by parts to show that the S.P. $X_t = I_{[T,\infty)}(t) - \lambda \min(t, T)$ is of zero mean. Then use the memoryless property of T to deduce that X_t is a martingale.*

In view of Proposition 6.2.6, N_t having independent increments is related to the following condition of having Exponentially distributed inter-arrival times.

CONDITION. $\boxed{C_2}$ *The gaps between jump times $T_k - T_{k-1}$ for $k = 1, 2, \ldots$ are i.i.d. random variables, each of Exponential(λ) law.*

Indeed, an equivalent definition of the Poisson process is:

PROPOSITION 6.2.8. *A stochastic process N_t that satisfies $\boxed{C_0}$ is a Poisson process of rate λ if and only if it satisfies $\boxed{C_2}$ (c.f.* [**Bre92**, page 309]*).*

Proposition 6.2.8 implies that for each k, the k-th arrival time T_k of the Poisson process of rate λ has the Gamma(k, λ) law corresponding to the sum of k i.i.d. Exponential(λ) random variables. In the next exercise we arrive at the same conclusion by an application of Doob's optional stopping theorem.

EXERCISE 6.2.9. *Let N_t, $t \geq 0$ be a Poisson process of rate λ.*

(a) *Show that $L_t = \exp(-\theta N_t - \beta t)$ is a martingale for the canonical filtration $\mathcal{G}_t = \sigma(N_s, 0 \leq s \leq t)$, whenever $\beta = -\lambda(1 - e^{-\theta})$.*
(b) *Check that $\{T_k \leq t\} = \{N_t \geq k\} \in \mathcal{G}_t$ and deduce that $T_r = \min(t \geq 0 : N_t = r)$ is a stopping time with respect to $\mathcal{G}_t$, for each positive integer r.*
(c) *Using Doob's optional stopping theorem for the martingale $(L_t, \mathcal{G}_t)$ (see Theorem 4.3.17), compute the value of $\mathbf{E}(e^{-\beta T_r})$ for $\beta > 0$.*
(d) *Check that the preceding evaluation of $\mathbf{E}(e^{-\beta T_r})$ equals what you get by assuming that T_r has the Gamma distribution of parameters r and λ. That is, by taking T_r with the probability density function*

$$f_{T_r}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{r-1}}{(r-1)!} \qquad \forall t \geq 0\,.$$

REMARK. Obviously the sample path of the Poisson process are never continuous. However, $\mathbf{P}(N_{t+h} - N_t \geq 1) = 1 - e^{-\lambda h} \to 0$ as $h \downarrow 0$, implying that $\mathbf{P}(T_k = t) = 0$ for all $t \geq 0$ and $k = 1, 2, \ldots$, so this S.P. has no fixed discontinuities (i.e. occurring at non-random times). The Poisson process is a special case of the family of Markov jump processes of Section 6.3, all of whom have discontinuous sample path but no fixed discontinuities.

Use the next exercise to check your understanding of the various characterizations and properties of the Poisson process.

EXERCISE 6.2.10. *Let $\{N_t\}$ be a Poisson process with rate $\lambda > 0$ and τ a finite stopping time for its canonical filtration. State which of the four stochastic processes $N_t^{(1)} = 2N_t$, $N_t^{(2)} = N_{2t}$, $N_t^{(3)} = N_{t^2}$ and $N_t^{(4)} = N_{\tau+t} - N_\tau$ is a Poisson process and if so, identify its rate.*

The Poisson process is related not only to the Exponential distribution but also to the Uniform measure, as we state next (and for a proof c.f. [**KT75**, Theorem 4.2.3]).

PROPOSITION 6.2.11. *Fixing positive t and a positive integer n let U_i be i.i.d. random variables, each uniformly distributed in $[0, t]$ and consider their* order statistics *U_i^* for $i = 1, \dots, n$. That is, permute the order of U_i for $i = 1, \dots, n$ such that $U_1^* \le U_2^* \le \cdots \le U_n^*$ while $\{U_1, \dots, U_n\} = \{U_1^*, \dots, U_n^*\}$ (for example $U_1^* = \min(U_i : i = 1, \dots, n)$ and $U_n^* = \max(U_i : i = 1, \dots, n)$). The joint distribution of $(U_1^*, \dots, U_n^*)$ is then precisely that of the first n arrival times $(T_1, \dots, T_n)$ of a Poisson process, conditional on the event $N_t = n$. Alternatively, fixing n and $0 \le t_1 \le t_2 \le \cdots \le t_n \le t$, we have that*

$$\mathbf{P}(T_k \le t_k,\ k = 1, \dots, n | N_t = n) = \frac{n!}{t^n} \int_0^{t_1} \int_{x_1}^{t_2} \cdots \int_{x_{n-1}}^{t_n} dx_1 dx_2 \cdots dx_n \,.$$

Here are a few applications of Proposition 6.2.11.

EXERCISE 6.2.12. *Let $T_k = \sum_{i=1}^k \xi_i$ for independent Exponential(λ) random variables ξ_i, $i = 1, 2, \dots$ and $N_t = \sup\{k \ge 0 : T_k \le t\}$ the corresponding Poisson process.*

(a) *Express $v = \mathbf{E}(\sum_{k=1}^{N_t} (t - T_k))$ in terms of $g(n) = \mathbf{E}(\sum_{k=1}^n T_k | N_t = n)$ and the law of N_t.*
(b) *Compute the values of $g(n) = \mathbf{E}(\sum_{k=1}^n T_k | N_t = n)$.*
(c) *Compute the value of v.*
(d) *Suppose that T_k is the arrival time to the train station of the k-th passenger on a train that departs the station at time t. What is the meaning of N_t and of v in this case?*

EXERCISE 6.2.13. *Let N_t be a Poisson process of rate $\lambda > 0$.*

(a) *Fixing $0 < s \le t$ show that conditional on $N_t = n$, the R.V. N_s has the Binomial(n, p) law for $p = s/t$. That is,*

$$\mathbf{P}(N_s = k | N_t = n) = \binom{n}{k} \big(\frac{s}{t}\big)^k \big(1 - \frac{s}{t}\big)^{n-k}, \qquad k = 0, 1, \dots, n.$$

(b) *What is the probability that the first jump of this process occurred before time $s \in [0, t]$ given that precisely n jumps have occurred by time t?*

EXERCISE 6.2.14. *Let N_t be a Poisson process of rate $\lambda > 0$. Compute $v(n) = \mathbf{E}[N_s | N_t = n]$ and $\mathbf{E}[N_s | N_t]$, first for $s > t$, then for $0 \le s \le t$.*

EXERCISE 6.2.15. *Commuters arrive at a bus stop according to a Poisson process with rate $\lambda > 0$. Suppose that the bus is large enough so that anyone who is waiting for the bus when it arrives is able to board the bus.*

(a) *Suppose buses arrive every T units of time (for non-random $T > 0$). Immediately after a bus arrives and all the waiting commuters board it, you uniformly and independently select a commuter that just boarded the bus. Find the expected amount of time that this commuter waited for the bus to arrive.*

(b) *Now supposes buses arrive according to a Poisson process with rate $1/T$. Assume that arrivals of commuters and buses are independent. What is the expected waiting time of a randomly selected commuter on the bus in this case? Is it different than the expected waiting time of the first commuter to arrive at the bus stop?*

We have the following interesting approximation property of the Poisson distribution (see also [**GS01**, Section 4.12]).

THEOREM 6.2.16 (Poisson approximation). *Suppose that for each n, the random variables $Z_l^{(n)}$ are independent, non-negative integers where $\mathbf{P}(Z_l^{(n)} = 1) = p_l^{(n)}$ and $\mathbf{P}(Z_l^{(n)} \geq 2) = \varepsilon_l^{(n)}$ are such that as $n \to \infty$,*

(i). $\displaystyle\sum_{l=1}^{n} p_l^{(n)} \to \lambda \in (0, \infty)$,

(ii). $\displaystyle\sum_{l=1}^{n} \varepsilon_l^{(n)} \to 0$,

(iii). $\displaystyle\max_{l=1,\cdots,n} \{p_l^{(n)}\} \to 0$.

Then, $S_n = \displaystyle\sum_{l=1}^{n} Z_l^{(n)}$ *converges in distribution to* Poisson(λ) *when* $n \to \infty$.

For example, consider $Z_l^{(n)} = \{0, 1\}$ with $\mathbf{P}(Z_l^{(n)} = 1) = \frac{\lambda}{n}$, resulting with S_n having the Binomial$(n, \frac{\lambda}{n})$ law. In this case, $\varepsilon_l^{(n)} = 0$ and $p_l^{(n)} = \frac{\lambda}{n}$, with $\displaystyle\sum_{l=1}^{n} p_l^{(n)} = \lambda$. Hence, applying Theorem 6.2.16 we have that the Binomial$(n, \frac{\lambda}{n})$ probability measures converge weakly as $n \to \infty$ to the Poisson(λ) probability measure. This is the classical Poisson approximation of the Binomial, often derived in elementary probability courses.

The Poisson approximation theorem relates the Poisson process to the following condition.

CONDITION. $\boxed{C_3}$ *The S.P. N_t has no fixed discontinuities, that is $\mathbf{P}(T_k = t) = 0$ for all k and $t \geq 0$. Also, for any fixed k, $0 < t_1 < t_2 < \cdots < t_k$ and non-negative integers $n_1, n_2, \cdots, n_k$,*

$$\begin{aligned}
\mathbf{P}(N_{t_k+h} - N_{t_k} = 1 | N_{t_j} = n_j,\ j \leq k) &= \lambda h + o(h), \\
\mathbf{P}(N_{t_k+h} - N_{t_k} \geq 2 | N_{t_j} = n_j,\ j \leq k) &= o(h),
\end{aligned}$$

where $o(h)$ denotes a function $f(h)$ such that $h^{-1} f(h) \to 0$ as $h \downarrow 0$.

Indeed, as we show next, Theorem 6.2.16 plays for the Poisson process the same role that the Central Limit Theorem plays for the Brownian motion, that is, providing a characterization of the Poisson process that is very attractive for the purpose of modeling real-world phenomena.

Proposition 6.2.17. *A stochastic process N_t that satisfies* $\boxed{C_0}$ *is a Poisson process of rate λ if and only if it satisfies condition* $\boxed{C_3}$.

Proof. (omit at first reading) Fixing k, the t_j and the n_j, denote by A the event $\{N_{t_j} = n_j,\ j \le k\}$. For a Poisson process of rate λ the random variable $N_{t_k+h} - N_{t_k}$ is independent of A with $\mathbf{P}(N_{t_k+h} - N_{t_k} = 1) = e^{-\lambda h}\lambda h$ and $\mathbf{P}(N_{t_k+h} - N_{t_k} \ge 2) = 1 - e^{-\lambda h}(1 + \lambda h)$. Since $e^{-\lambda h} = 1 - \lambda h + o(h)$ we see that the Poisson process satisfies $\boxed{C_3}$. To prove the converse, we start with a S.P. N_t that satisfies both $\boxed{C_0}$ and $\boxed{C_3}$. Fixing A as above, let $D_t = N_{t_k+t} - N_{t_k}$ and $p_n(t) = \mathbf{P}(D_t = n|A)$. To show that N_t satisfies $\boxed{C_1}$, hence is a Poisson process, it suffices to show that $p_n(t) = e^{-\lambda t}(\lambda t)^n/n!$ for any $t \ge 0$ and non-negative integer n (the independence of increments then follows by induction on k). It is trivial to check the case of $t = 0$, that is, $p_n(0) = \mathbf{1}_{n=0}$. Fixing $u > 0$ and $s \ge 0$ we have that $|p_n(u) - p_n(s)| \le \mathbf{P}(D_u \ne D_s|A) \le \mathbf{P}(N_{t_k+u} \ne N_{t_k+s})/\mathbf{P}(A)$. By $\boxed{C_3}$ we know that N_t is continuous in probability so taking $u \to s$ yields that $s \mapsto p_n(s)$ is continuous on $[0, \infty)$. If $D_{t+h} = n$ then necessarily $D_t = m$ for some $m \le n$. Fixing n and $t > 0$, by the rules of conditional probability

$$p_n(t+h) = \sum_{m=0}^{n} p_m(t)\mathbf{P}(D_{t+h} - D_t = n - m|A_m)\,,$$

where $A_m = A \cap \{D_t = m\}$. Applying $\boxed{C_3}$ for each value of m separately (with $t_{k+1} = t_k + t$ and $n_{k+1} = n_k + m$ there), we thus get that $p_n(t+h) = p_n(t)(1 - \lambda h) + p_{n-1}(t)\lambda h + o(h)$. Taking $h \downarrow 0$ gives the system of differential equations

$$p_n'(t) = -\lambda p_n(t) + \lambda p_{n-1}(t)\,,$$

with boundary conditions $p_{-1}(t) = 0$ for all t and $p_n(0) = \mathbf{1}_{n=0}$ (a-priori, $t > 0$ and $p_n'(t)$ stands for the right-hand derivative, but since $t \mapsto p_n(t)$ is continuous, the differential equations apply also at $t = 0$ and $p_n'(t)$ can be taken as a two-sided derivative when $t > 0$). It is easy to check that $p_n(t) = e^{-\lambda t}(\lambda t)^n/n!$ satisfies these equations. It is well known that the solution of such a differential system of equations is unique, so we are done. ∎

The collection of all Poisson processes is closed with respect to the merging and thinning of their streams.

Proposition 6.2.18. *If $N_t^{(1)}$ and $N_t^{(2)}$ are two independent Poisson processes of rates λ_1 and λ_2 respectively, then $N_t^{(1)} + N_t^{(2)}$ is a Poisson process of rate $\lambda_1 + \lambda_2$. Conversely, the sub-sequence of jump times obtained by independently keeping with probability p each of the jump times of a Poisson process of rate λ corresponds to a Poisson process of rate λp.*

We conclude with the law of large numbers for the Poisson process.

Exercise 6.2.19. *Redo Exercise 5.1.7 with W_t replaced by the MG $\frac{1}{\sqrt{\lambda}}(N_t - \lambda t)$ for a Poisson process N_t of rate $\lambda > 0$, to arrive at the* law of large numbers *for the Poisson process. That is, $t^{-1}N_t \overset{a.s.}{\to} \lambda$ for $t \to \infty$.*

Remark. A *inhomogeneous Poisson process* X_t with rate function $\lambda(t) \ge 0$ for $t \ge 0$ is a counting process of independent increments, such that for all $t > s \ge 0$, the increment $X_t - X_s$ has the Poisson($\Lambda(t) - \Lambda(s)$) law with $\Lambda(t) = \int_0^t \lambda(s)ds$. This is merely a non-random *time change* $X_t = N_{\Lambda(t)}$ of a Poisson process N_t of rate one (see also Proposition 5.2.2 for the time change of a Brownian motion). Condition $\boxed{C_3}$ is then inherited by $\{X_t\}$ (upon replacing λ by $\lambda(t_k)$), but condition $\boxed{C_2}$ does not hold, as the gaps between jump times S_k of $\{X_t\}$ are neither i.i.d. nor of Exponential law. Nevertheless, we have much information about these jump times as $T_r = \Lambda(S_r)$ for the jump times T_r of the homogeneous Poisson process of rate one (for example, $\mathbf{P}(S_1 \leq s | X_t = 1) = \Lambda(s)/\Lambda(t)$ for all $0 \leq s \leq t$).

6.3. Markov jump processes, compound Poisson processes

We begin with the definition of the family of Markov jump processes that are the natural extension of the Poisson process. Whereas the Brownian motion and the related diffusion processes are commonly used to model processes with continuous sample path, Markov jump processes are the most common object in modeling situations with inherent jumps.

DEFINITION 6.3.1. *Let T_i denote the jump times of a Poisson process $\{Y(t)\}$ of rate $\lambda > 0$. We say that a stochastic process $X(t)$ is a* Markov jump process *if its sample path are constant apart from jumps of size X_i at times T_i, and the sequence $Z_n = \sum_{j=1}^n X_j$ is a Markov chain which is independent of $\{T_i\}$. That is, $X(t) = Z_{Y(t)}$. The Markov jump processes with i.i.d. jump sizes X_i are called* compound Poisson processes.

In particular, taking constant $X_i = 1$ in Definition 6.3.1 leads to the Poisson process, while $Z_{n+1} = -Z_n \in \{-1, 1\}$ gives the random telegraph signal of Example 3.3.6. The latter is also an example of a Markov jump process which is not a compound Poisson process.

Clearly, the sample paths of the Markov jump process $X(t)$ inherit the RCLL property of the Poisson process $Y(t)$. We further show next that compound Poisson processes inherit many other properties of the underlying Poisson process.

PROPOSITION 6.3.2. *Any compound Poisson process $X(t)$ (that is, a Markov jump process of i.i.d. jump sizes X_i), has stationary, independent increments and the characteristic function $\mathbf{E}(e^{iuX(t)}) = \exp\{\lambda t \int (e^{iux} - 1) dF_X(x)\}$ (where $F_X(x)$ denotes the distribution function of the jump size X_1).*

PROOF OUTLINE. To prove the proposition one has to check that for any $h > 0$ and $t > 0$, the random variable $X(t+h) - X(t)$ is independent of $\sigma(X(s), 0 \leq s \leq t)$, and its law does not depend on t, neither of which we do here (c.f. [**Bre92**, Section 14.7] where both steps are detailed).

Thus, we only prove here the stated formula for the characteristic function of $X(t)$. To this end, we first condition on the event $\{Y(t) = n\}$. Then, $X(t) = \sum_{i=1}^{n} X_i$, implying by the independence of X_j that

$$\mathbf{E}[e^{iuX(t)} | Y(t) = n] = \mathbf{E}[e^{iu \sum_{j=1}^n X_j}] = h(u)^n ,$$

where $h(u) = \int_{\mathbb{R}} e^{iux} dF_X(x)$ is the characteristic function of the jump size X. Since $Y(t)$ is a Poisson(λt) random variable, we thus have by the tower property of the expectation that

$$\begin{aligned}\mathbf{E}[e^{iuX(t)}] &= \mathbf{E}[\mathbf{E}[e^{iuX(t)}|Y(t)]] = \mathbf{E}[h(u)^{Y(t)}] \\ &= \sum_{n=0}^{\infty} h(u)^n \mathbf{P}(Y(t)=n) = \sum_{n=0}^{\infty} (h(u))^n \frac{(\lambda t)^n}{n!} e^{-\lambda t} = e^{\lambda t(h(u)-1)}\,,\end{aligned}$$

as claimed. ∎

REMARK. Similarly to Proposition 6.3.2 it is not hard to check that if $X(t)$ is a compound Poisson process for i.i.d. X_i that are square integrable, then $\mathbf{E}(X(t)) = \lambda t \mathbf{E}(X_1)$ and $\mathsf{Var}(X(t)) = \lambda t \mathbf{E}(X_1^2)$.

Consequences of Proposition 6.3.2:

• One of the implications of Proposition 6.3.2 is that any compound Poisson process is a homogeneous Markov process (see Proposition 6.1.17).

• Another implication is that if in addition $\mathbf{E}X_1 = 0$, then $X(t)$ is a martingale (see Proposition 4.2.3) . Like in the case of a Poisson process, many other martingales can also be derived out of $X(t)$.

Fixing any disjoint finite partition of $\mathbb{R}\backslash\{0\}$ to Borel sets B_k, $k = 1, \ldots, m$ we have the decomposition of any Markov jump process $X(t) = \sum_{k=1}^{m} X(B_k, t)$, in terms of the contributions

$$X(B_k, t) = \sum_{\tau \le t} (X(\tau) - X(\tau-)) \mathbf{1}_{\{X(\tau)-X(\tau-) \in B_k\}}$$

to $X(t)$ by jumps whose size belong to B_k. What is so interesting about this decomposition is the fact that for any compound Poisson process, these components are in turn independent compound Poisson processes, as we state next.

PROPOSITION 6.3.3. *If the Markov jump process $X(t)$ is a compound Poisson process, then the S.P. $X(B_k, t)$ for $k = 1, \ldots, m$ are independent compound Poisson processes, with $\int_{B_k}(e^{iux}-1)dF_X(x)$ replacing $\int_{\mathbb{R}}(e^{iux}-1)dF_X(x)$ in the formula for the characteristic function for $X(B_k, t)$ (c.f.* [**Bre92**, Proposition 14.25]*).*

Here is a concrete example that give rise to a compound Poisson process.

EXERCISE 6.3.4. *A basketball team scores baskets according to a Poisson process with rate 2 baskets per minute. Each basket is worth either 1, 2, or 3 points; the team attempts shots according to the following percentages: 20% for 1 point, 50% for 2 points, and 30% for 3 points.*

(a) *What is the expected amount of time until the team scores its first basket?*
(b) *Given that at the five minute mark of the game the team has scored exactly one basket, what is the probability that the team scored the basket in the first minute?*
(c) *What is the probability that the team scores exactly three baskets in the first five minutes of the game?*
(d) *What is the team's expected score at the five minute mark of the game?*
(e) *Let $Z(t)$ count the number of 2-point baskets by the team up to time t. What type of S.P. is $\{Z(t)\}$.*

Bibliography

[AS00] Noga Alon and Joel H. Spencer, *The probabilistic method*, 2nd ed., Wiley-Interscience, 2000.

[Bre92] Leo Breiman, *Probability*, Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, 1992.

[GS01] Geoffrey Grimmett and David Stirzaker, *Probability and random processes*, 3rd ed., Oxford University Press, 2001.

[KS97] Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, 2nd ed., Graduate Texts in Mathematics, vol. 113, Springer Verlag, 1997.

[KT75] Samuel Karlin and Howard M. Taylor, *A first course in stochastic processes*, 2nd ed., Academic Press, 1975.

[Mik98] Thomas Mikosch, *Elementary stochastic calculus*, World Scientific, 1998.

[Oks03] Bernt Oksendal, *Stochastic differential equations: An introduction with applications*, 6th ed., Universitext, Springer Verlag, 2003.

[Ros95] Sheldon M. Ross, *Stochastic processes*, 2nd ed., John Wiley and Sons, 1995.

[Wil91] David Williams, *Probability with martingales*, Cambridge University Press, 1991.

[Zak] Moshe Zakai, *Stochastic processes in communication and control theory*.

Index