Lectures on
Brownian motion, martingales, and stochastic analysis

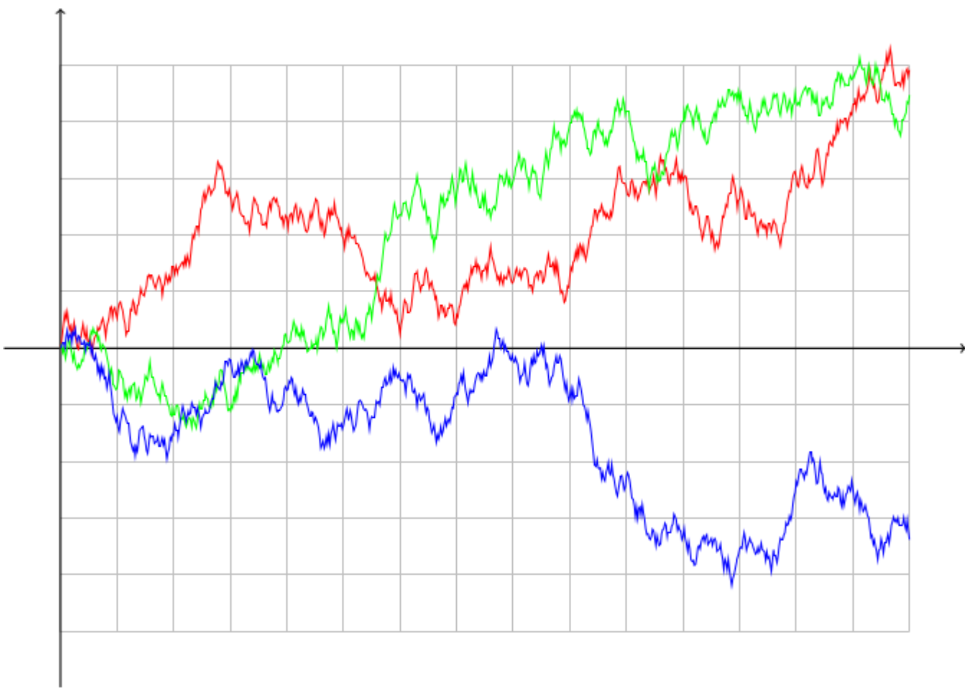
Eveliina Peltola*

April 27, 2025

Abstract

Brownian motion is a fundamentally important stochastic process, originally discovered in the 19th century (and in fact, in some forms even earlier) in the contexts of diffusion phenomena in statistical physics and as a speculative model for financial markets. It plays a key role in many modern mathematical topics, ranging from partial differential equations and potential theory to constructive quantum field theory and random geometry.

This lecture course introduces the mathematical foundations for Brownian motion, key concepts and techniques needed for working with it, and applications aiming at martingales, stochastic integration, and Itô's Formula. As supplementary material, there are many great textbooks on Brownian motion, for example [LeG16, MP10]. The readers are assumed to be familiar with basic notions of probability theory and stochastic processes, for instance as in Part A of Williams' textbook [Wil91], or Kytölä's lectures [Kyt20].



*Aalto University, Department of Mathematics and Systems Analysis. eveliina.peltola@aalto.fi
University of Bonn, Institute for Applied Mathematics. eveliina.peltola@hcm.uni-bonn.de

These **rough** lecture notes describe the content of the course "Brownian motion and stochastic analysis" at Aalto University given in spring 2025. **Please let me know if you find mistakes or misprints!**

Contents

1 Introduction

In this lecture course, we shall build some mathematical foundations for Brownian motion and its applications. After the puzzles it caused early on, Brownian motion has been set to rigorous basis and has been widely used to model phenomena not only in mathematics, physics, and chemistry, but also in economics and finance, biology, medicine, and other sciences. One might argue that the solid mathematical foundations for Brownian motion were laid by Norbert Wiener (1894–1964) and Andrey Kolmogorov (1903–1987), among others. There are many great textbooks on Brownian motion, for example [LeG16, MP10], which also some of the material here is based on. The readers are assumed to be familiar with basic notions of probability theory and stochastic processes, for instance as in Part A of Williams' textbook [Wil91], or Kytölä's lectures [Kyt20].

Some comments for reading these notes

- ▷ Sections whose titles are decorated with "$\star$"-symbols are meant as supplementary material.
- ▷ Some basic notions from probability theory are recalled in Appendix A.

1.1 Some historical remarks ($\star$)

The introduction of *Brownian motion* as a phenomenon in Nature is often attributed to the botanist Robert Brown (1773–1858). In 1828, he reported on the random movement of small pollen particles in water, undergoing an extremely erratic and never-ending motion [Bro28]. Thereafter, Brown and others performed further studies ruling out, e.g., the possibility of an organic force, interactions, convection currents, or evaporation, as a potential cause of the motion (cf. [Gou88]). Brown seemed to have correctly suspected that *heat* was playing a fundamental role for its origins. It was realized only much later, however, that the phenomenon underlying Brownian motion was fundamental and universal, leading to ubiquitous applications in sciences.

A few decades later, during his "annus mirabilis" year 1905, Albert Einstein demonstrated in his PhD thesis [Ein06] how thermal motion of the ambient molecules of the fluid was responsible for the random diffusion of the particles in the fluid. (Similar conclusions were drawn independently and simultaneously by William Sutherland [Sut05].) Einstein's theory provided convincing evidence that *atoms and molecules* do exist — this was indeed soon also verified experimentally by Jean Baptiste Perrin[1], who also wrote a famous book on the subject [Per13].

Later on, Brownian motion became a crucial piece of modern sciences, used for instance in important developments of *statistical mechanics*, *quantum field theory* (e.g. path integrals developed by Norbert Wiener in the 1920s [Wie23] and Richard Feynman in the 1940s [Fey48]), the theory of *stochastic processes* (pioneered by works of Mark Kac, e.g. [Kac47]), and *economics and finance*, as initiated by Louis Bachelier in 1900 for modelling option prices at the Paris stock market in his PhD thesis [Bac00] (supervised by Henri Poincaré). Let us also mention the option price model of Fischer Black & Myron Scholes [BS73], of whom the latter together with Robert C. Merton received the 1997 Nobel prize in economics. In mathematics, in addition to probability theory, Brownian motion is an important tool in analysis (especially *potential theory* and *partial differential equations*). Some quite recent developments include the general area of probability theory now termed as "*random geometry*" (which already has yielded 3 Fields medals in mathematics: Wendelin Werner 2006, Stanislav Smirnov 2010, Hugo Duminil-Copin 2022).

For interesting historical remarks and references concerning Brownian motion, check out Bertrand Duplantier's "Poincaré Seminar" Lectures [Dup05], and references therein.

[1] Perrin was awarded the 1926 Nobel Prize in Physics "for his work on the discontinuous structure of matter," while Einstein got his 1921 Nobel Prize in Physics – not really for explaining Brownian motion but – "for his services to theoretical physics, and especially for his discovery of the law of the photoelectric effect."

1.2 What is Brownian motion?

Already in the course of the 19th century, chemists and physicists made various conclusions on the properties of the perplexing Brownian motion. Not only the origin of the phenomenon was puzzling, but also the very nature of it. Some general features of Brownian motion are:

▷ *The motion is extremely irregular, and the trajectory seems not to have a tangent.*

Indeed, Brownian motion is a random path (continuous stochastic process), which is Hölder continuous for small enough Hölder exponent, but still nowhere differentiable, cf. Section 1.7.

Mathematically, Brownian motion is a continuous-time real-valued stochastic process usually denoted $B = (B_t)_{t\geq 0}$, defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. In particular,

* $\omega \mapsto B_t(\omega)$ is a *random variable* (Borel-measurable map from $(\Omega, \mathscr{F})$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$) for each fixed time $t \in [0, \infty)$;
* $t \mapsto B_t(\omega)$ is a *continuous function* from $[0, \infty)$ to $\mathbb{R}$ for almost every fixed $\omega \in \Omega$ (that is, $\mathbb{P}$-almost surely (a.s.): the event that B is continuous has probability one).

More precisely, one would want to say that Brownian motion is a random variable

$$B : \Omega \longrightarrow C([0, \infty), \mathbb{R}) := \{f : [0, \infty) \to \mathbb{R} \mid f \text{ is continuous}\},$$

where $C([0, \infty), \mathbb{R})$ is endowed with the sigma-algebra generated by the cylinder events

$$\big\{f(t_1) \in A_1, \ldots, f(t_n) \in A_n\big\}, \quad 0 \leq t_1 < \cdots < t_n \ \text{ and } \ A_1, \ldots, A_n \in \mathscr{B}(\mathbb{R}).$$

We will see that one can choose the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ judiciously in such a way that we have continuity for all $\omega \in \Omega$. (This is often called the Wiener space, cf. Section 1.8.2.) Meanwhile, the reader may ponder of why this is not obvious.

▷ *The motion of a Brownian particle does not remember its past.*

Indeed, Brownian motion is a *Markov process*, as discussed in detail in Section 6.

▷ *The motion never stops.*

This is the key property that led physicists to understand the origins of the motion as being caused by thermal fluctuations of molecules, thus also demonstrating that molecules exist!

1.3 From random walk to Brownian motion

When trying to rigorously formalize Brownian motion, a very natural idea is *discretization*: instead of trying to define a continuous-time stochastic process on the uncountable space $\mathbb{R}$, one can define *approximations* of it by discrete-time processes on finite or countable state spaces.

Consider real-valued random variables $\xi_1, \xi_2, \ldots$ defined on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Suppose that they are *independent and identically distributed* (i.i.d.). Define

$$S_0 = 0 \qquad \text{and} \qquad S_n := \sum_{j=1}^{n} \xi_j \qquad \text{for } n = 1, 2, \ldots.$$

Then, $S = (S_n)_{n\in\mathbb{N}_0}$ is called a *random walk* (indexed by $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$). It obviously satisfies

$$S_n = S_{n-1} + \xi_n \qquad \text{for } n = 1, 2, \ldots,$$

or in other words, $\xi_n = S_n - S_{n-1}$. These *increments* of S are i.i.d. by construction.

To construct Brownian motion from a random walk, the first idea would be to re-scale the integer times for S. However, in order to find a limiting object, we have to scale the space as well. As a starting point for seeking a suitable scaling, we take the *Central Limit Theorem*[2]:

Theorem 1.1. (Central Limit Theorem) *Let* $\xi_1, \xi_2, \ldots$ *and* S *be as above. Suppose that* $\mathbb{E}[\xi_1] = \mathfrak{m} < \infty$ *and* $\mathrm{Var}(\xi_1) = \mathfrak{s}^2 > 0$ [a]. *Then, we have*

$$\frac{S_n - n\,\mathfrak{m}}{\mathfrak{s}\sqrt{n}} \xrightarrow{\text{law}} N \sim N(0,1) \qquad \text{as } n \to \infty,$$

where the convergence takes place in distribution (or law), i.e., as weak convergence of probability measures, and N *is a Gaussian random variable with mean* $\mathbb{E}[N] = 0$ *and variance* $\mathrm{Var}(N) = 1$ *(also called a standard normal random variable). Recall that* N *has a continuous distribution with density*

$$\frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}.$$

[a] Hence, $\mathbb{E}[\xi_j] = \mathfrak{m}$ and $\mathrm{Var}(\xi_j) = \mathfrak{s}^2$ for all j, because ξ_j are i.i.d.

Let us now assume that $\xi_1, \xi_2, \ldots$ are chosen so that they are i.i.d. with

$$\mathbb{E}[\xi_1] = 0 \qquad \text{and} \qquad \mathrm{Var}(\xi_1) = 1, \tag{1.1}$$

e.g., as coin tosses (Bernoulli type random variables, see Exercise A.4). Building from these, consider a re-scaled random walk $S_t^{(m)}$ with times in $t = \frac{k}{m} \in \{0, \frac{1}{m}, \frac{2}{m}, \ldots\}$ of time-step size $1/m$,

$$S_{k/m}^{(m)} := c_m\, S_k, \qquad k = 0, 1, 2, \ldots,$$

with some constants $c_m \in \mathbb{R}$ to be determined. Since

$$\mathrm{Var}\big(S_{k/m}^{(m)}\big) = c_m^2\, \mathrm{Var}(S_k) = c_m^2\, k, \qquad k = 0, 1, 2, \ldots,$$

we see that the variance can only converge as $m \to \infty$ if we take c_m to be of order $1/\sqrt{m}$. Defining

$$S_{k/m}^{(m)} := \frac{S_k}{\sqrt{m}}, \qquad k = 0, 1, 2, \ldots,$$

and extending it by linear interpolation to a random function $S^{(m)} : [0, \infty) \to \mathbb{R}$, we find that

$$\mathbb{E}\big[S_t^{(m)}\big] = 0 \qquad \text{for all } t \geq 0, \qquad \text{and} \qquad \mathrm{Var}\big(S_{k/m}^{(m)}\big) = \frac{k}{m}$$

for all $m \in \mathbb{N} := \{1, 2, \ldots\}$. In particular, if we fix $n, m \in \mathbb{N}$ and times $0 \leq t_0 < t_1 < t_2 < \cdots < t_n$ such that each of them is a multiple of $1/m$, then we see that the increment

$$W_{m\,(t_{j+1} - t_j)} := S_{t_{j+1}}^{(m)} - S_{t_j}^{(m)} = \frac{1}{\sqrt{m}} \sum_{k = m\,t_j + 1}^{m\,t_{j+1}} \xi_k$$

is a sum of $m\,(t_{j+1} - t_j)$ i.i.d. random variables $\eta_k^{(m)} := \frac{\xi_k}{\sqrt{m}}$ satisfying

$$\mathbb{E}\big[\eta_k^{(m)}\big] = 0 \qquad \text{and} \qquad \mathrm{Var}\big(\eta_k^{(m)}\big) = \frac{1}{m} \qquad \text{for all } k.$$

[2] The Central Limit Theorem is proven e.g. in the Probability Theory course [Kyt20, Chapter XII.3].

Thus, the Central Limit Theorem yields

$$\frac{W_{m\,(t_{j+1}-t_j)}}{\sqrt{(t_{j+1}-t_j)}} \xrightarrow{\text{law}} N \sim N(0,1) \qquad \text{as } m \to \infty,$$

or in other words, $W_{m\,(t_{j+1}-t_j)} \xrightarrow{\text{law}} W \sim N(0, t_{j+1}-t_j)$, which has the Gaussian density

$$\frac{1}{\sqrt{2\pi(t_{j+1}-t_j)}} \exp\Big(- \frac{x}{2(t_{j+1}-t_j)} \Big).$$

We conclude that if S has a scaling limit process, it has Gaussian increments.

Donsker's theorem[3] says that the scaling limit exists and is a Brownian motion (regardless of the exact distribution of $\xi_1, \xi_2, \ldots$ as long as they are i.i.d. and normalized as in (1.1)).

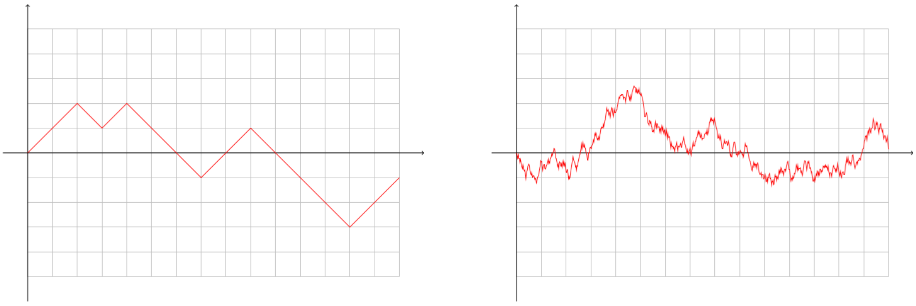


Illustration of the convergence of a random walk to Brownian motion.

1.4 Definition of Brownian motion

Motivated by discretization, it is natural that Brownian motion has Gaussian increments. We also expect it to satisfy the Markov Property, so we shall require in addition that the increments are independent. The following is suitable to determine the law of Brownian motion uniquely.

Definition 1.2. (One-dimensional Brownian motion) Fix $a \in \mathbb{R}$. A continuous-time real-valued stochastic process $B = (B_t)_{t \geq 0}$ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called a *Brownian motion* (starting at a) if and only if it satisfies the following properties:

BM0. (starting point): $\mathbb{P}$-almost surely, $B_0 = a$:

$$\mathbb{P}\big[\{\omega \in \Omega \mid B_0(\omega) = a\}\big] = 1.$$

BM1. (independent increments): for any partition $0 \leq t_0 < t_1 < t_2 < \cdots < t_n$, the increments $\{B_{t_{j+1}} - B_{t_j} \mid j = 0, 1, \ldots, n-1\}$ are independent random variables,

BM2. (stationary, Gaussian increments): for each $0 \leq s < t$, the increment $B_t - B_s$ has Gaussian distribution that only depends on the time difference:

$$B_t - B_s \sim N(0, t-s), \qquad 0 \leq s < t.$$

BM3. (a.s. continuous sample paths): $\mathbb{P}$-almost every sample path $t \mapsto B_t$ is continuous:

$$\mathbb{P}\big[\{\omega \in \Omega \mid t \mapsto B_t(\omega) \text{ is continuous}\}\big] = 1.$$

[3] Donsker's theorem is proven for example in the Large random systems course, see [KK19, Chapter VIiI].

- ▷ The property that the distribution of $B_t - B_s$ only depends on $t - s$ is called *stationarity*.
- ▷ When the starting point is $a = 0$, we call B a *standard* Brownian motion.
- ▷ Brownian motion in the Euclidean space $\mathbb{R}^n$ is defined as the process $(B_t^{(1)}, \ldots, B^{(n)})$, where the components are independent one-dimensional Brownian motions.

Remark 1.3. For the continuity property BM3 to make sense, one has to address the *measurability* of the event $\{\omega \in \Omega \mid t \mapsto B_t(\omega) \text{ is continuous}\}$. Note that we can only decide whether this event occurs if we know the values of $B_t(\omega)$ at all (thus, uncountably many) times $t \in [0, \infty)$.

For technical reasons, it is occasionally useful to relax the almost sure continuity property BM3 to the property[4] that there exists an event $E \in \mathscr{F}$ such that $\mathbb{P}[E] = 1$ and

$$E \subset \{\omega \in \Omega \mid t \mapsto B_t(\omega) \text{ is continuous}\}.$$

Note that the measurability of events depends on the choice of the probability space. We shall not delve into details concerning measurability issues at this point — but it is good to remember that for probability theory on uncountable spaces these issues have to be considered carefully.

Definition 1.4. (Pre-Brownian motion) A continuous-time real-valued stochastic process satisfying the properties BM0–BM2 is called a *pre-Brownian motion*.

One can show that a pre-Brownian motion has a *modification* that satisfies also property BM3 (see Theorem 1.23). However, not every modification of pre-Brownian motion has this property (see Exercise 1.12), so the requirement BM3 is essential for regarding (almost all) Brownian paths as random elements in the space $C([0, \infty), \mathbb{R})$ of *continuous* functions.

Remark 1.5. A continuous-time real-valued stochastic process having stationary independent increments is called a *Lévy process*. In discrete time, random walks have this property, and one would thus expect that scaling limits of random walks are Lévy processes. We get a Brownian motion as the scaling limit when the steps ξ_j of the random walk S_n have finite variance. Sample paths of a general Lévy process are not continuous, but it always admits a modification which is *càdlàg*:right-continuous with left limits. In fact, Brownian motion (possibly with drift) is the only Lévy process which admits a continuous modification, and Lévy showed that any a.s. continuous process with independent increments has Gaussian increments, see [Kal21, Theorem 13.4].

1.5 Brownian motion as a Gaussian process

Using the defining properties BM1–BM2 of a (pre-)Brownian motion, we can find explicitly all its *finite-dimensional marginal distributions* (FDDs), which are *Gaussian* by Lemma 1.9. For basics concerning Gaussian random variables and Gaussian processes, see [LeG16, Chapter 1].

Definition 1.6. Continuous-time real-valued stochastic process X is called a *Gaussian process* if any finite linear combination of the random variables X_t, $t \geq 0$, is Gaussian.

[4] In this case, abusing terminology we still say that $t \mapsto B_t$ is continuous almost surely, even if that event would not be $\mathscr{F}$-measurable. See Section 5.2 for more discussion on this.

Exercise 1.7. Show that the following statements are equivalent.

(i) $B = (B_t)_{t \geq 0}$ is a pre-Brownian motion on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

(ii) $B = (B_t)_{t \geq 0}$ is a centered[a] Gaussian process with covariance

$$\mathrm{Cov}(B_s, B_t) := \mathbb{E}[B_s B_t] - \mathbb{E}[B_s]\mathbb{E}[B_t] = \min\{s, t\}.$$

Hint: *Components of a Gaussian vector* $(X_1, \dots, X_n)$ *are independent if and only if* $\mathrm{Cov}(X_i, X_j) = 0$ *for all* $i \neq j$.

[a] The process X is called *centered* if $\mathbb{E}[X_t] = 0$ for all $t \geq 0$.

Definition 1.8. (FDDs) Consider a continuous-time stochastic process X on $(\Omega, \mathscr{F}, \mathbb{P})$ with values in a measurable space $(S, \mathscr{S})$. The *finite-dimensional marginal distributions* (FDDs) of X are the probabilities

$$\mathsf{P}_{(X_{t_1}, X_{t_2}, \dots, X_{t_n})}[A_1 \times \cdots \times A_n] = \mathbb{P}[X_{t_1} \in A_1, \dots, X_{t_n} \in A_n], \qquad \begin{cases} 0 \leq t_1 < \cdots < t_n, \\ A_1, \dots, A_n \in \mathscr{S}. \end{cases}$$

Recall that a density $f : \mathbb{R}^n \to [0, +\infty)$ for $(B_{t_1}, B_{t_2}, \dots, B_{t_n})$ satisfies in particular for each cylinder set $A_1 \times \cdots \times A_n$ of $\mathscr{B}(\mathbb{R}^n)$, where by definition each $A_j \in \mathscr{B}(\mathbb{R})$, the property

$$\mathsf{P}_{(B_{t_1}, B_{t_2}, \dots, B_{t_n})}[A_1 \times \cdots \times A_n] = \mathbb{P}[B_{t_1} \in A_1, \dots, B_{t_n} \in A_n] = \int_{A_1} \cdots \int_{A_n} f(x_1, \dots, x_n)\, \mathrm{d}x_n \cdots \mathrm{d}x_1.$$

Since the cylinder sets $A_1 \times \cdots \times A_n$ generate the Borel sigma-algebra $\mathscr{B}(\mathbb{R}^n)$ of the product space $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$ (they form a pi-system; cf. Appendix A), the above property implies the more general property for all Borel sets $A \in \mathscr{B}(\mathbb{R}^n)$:

$$\mathsf{P}_{(B_{t_1}, B_{t_2}, \dots, B_{t_n})}[A] = \mathbb{P}[(B_{t_1}, B_{t_2}, \dots, B_{t_n}) \in A] = \int_A f(x_1, \dots, x_n)\, \mathrm{d}x_n \cdots \mathrm{d}x_1.$$

The events $\{B_{t_1} \in A_1, \dots, B_{t_n} \in A_n\}$ indexed by $n \in \mathbb{N}$, $0 < t_1 < \cdots < t_n$, and $A_1, \dots, A_n \in \mathscr{B}(\mathbb{R})$ are called *cylinder events* and they form a pi-system too. Using Dynkin's Identification Theorem A.29 we can use them to generate a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ for pre-Brownian motion. The FDDs of pre-Brownian motion are given by the Gaussian density as follows.

Lemma 1.9. *Let B be a pre-Brownian motion. Then, for every $0 \leq t_1 < t_2 < \cdots < t_n$, the random vector $(B_{t_1}, B_{t_2}, \dots, B_{t_n})$ has a continuous distribution with density*

$$\frac{1}{(2\pi)^{n/2}\sqrt{t_1(t_2 - t_1)\cdots(t_n - t_{n-1})}} \exp\left(-\sum_{j=1}^{n} \frac{(x_j - x_{j-1})^2}{2(t_j - t_{j-1})} \right), \tag{1.2}$$

with the convention that $x_0 = 0$.

Exercise 1.10. Prove Lemma 1.9.

Corollary 1.11. *Brownian motion is a Gaussian process.*

Proof. This follows from Lemma 1.9, since Brownian motion is a pre-Brownian motion. ☐

One can show that these FDDs (1.2) determine the law of a pre-Brownian motion uniquely (see, e.g., [LeG16, Proposition 2.3]) up to a choice of starting point B_0 (via property BM0).

However, the continuity property BM3 of a Brownian motion cannot be expressed in terms of the FDDs — as it concerns uncountably many time instants simultaneously. Indeed, we have to require the continuity property BM3 separately. The next exercise illustrates a potential caveat.

Exercise 1.12. Let B be a Brownian motion and assume that the map $(\omega, t) \mapsto B_t(\omega)$ is jointly measurable[a] $\Omega \times [0, \infty) \to \mathbb{R}$. Let ξ be a random variable, independent of B, uniformly distributed on $[0, 1]$. Define

$$\tilde{B} = (\tilde{B}_t)_{t \geq 0}, \qquad \tilde{B}_t := \begin{cases} B_t, & \text{if } t \neq \xi, \\ 0, & \text{if } t = \xi. \end{cases}$$

Show that the FDDs of $\tilde{B}$ and B are the same, but the sample paths $t \mapsto \tilde{B}_t$ are discontinuous almost surely.

[a] For example, this follows from Lévy's construction of Brownian motion [MP10, Theorem 1.3].

Using, e.g., properties of Gaussians, it is not hard to check that (pre-)Brownian motion has the following useful symmetries. We will utilize them frequently.

Lemma 1.13. *(Symmetries of Brownian motion) Let $B = (B_t)_{t \geq 0}$ be a standard Brownian motion. The following processes are Brownian motions as well.*

1. $\widetilde{B}^{(1)} = (\widetilde{B}^{(1)}_t)_{t \geq 0}$ *defined by* $\widetilde{B}^{(1)}_t := -B_t$.
2. $\widetilde{B}^{(2)} = (\widetilde{B}^{(2)}_t)_{t \geq 0}$ *defined by* $\widetilde{B}^{(2)}_t := \frac{1}{\sqrt{\lambda}} B_{\lambda t}$, *where* $\lambda > 0$ *is fixed.*
3. $\widetilde{B}^{(3)} = (\widetilde{B}^{(3)}_t)_{t \geq 0}$ *defined by* $\widetilde{B}^{(3)}_t := B_{t+s} - B_s$, *where* $s > 0$ *is fixed.*
4. $\widetilde{B}^{(4)} = (\widetilde{B}^{(4)}_t)_{t \geq 0}$ *defined by* $\widetilde{B}^{(4)}_0 = 0$ *at time zero, and* $\widetilde{B}^{(4)}_t := t\, B_{1/t}$ *at times* $t > 0$.

Exercise 1.14. Prove Lemma 1.13.

Exercise 1.15. A *Brownian bridge* started at $a \in \mathbb{R}$ ending at $b \in \mathbb{R}$ is a continuous-time stochastic process $X = (X_t)_{t \in [0,1]}$ defined in terms of a Brownian motion B started at a as follows:

$$X_t := B_t - t\,(B_1 - b) \qquad \text{for all } 0 \leq t \leq 1, \qquad \text{where} \quad B_0 = a.$$

Morally, X can be thought of as Brownian motion on the time interval $[0, 1]$ started at a and *conditioned to end at b*. However, such a conditioning does not make literal sense, since the event that $B_1 = b$ has zero probability.

Solve the FFDs of Brownian bridge: that is, show that for each $n \in \mathbb{N}$, for every bounded continuous function $g : \mathbb{R}^n \to \mathbb{R}$, and for all $0 < t_1 < t_2 < \cdots < t_n < 1$, we have

$$\mathbb{E}\big[g(X_{t_1}, \ldots, X_{t_n})\big] = \int_{\mathbb{R}^n} g(x_1, \ldots, x_n)\, \frac{p_{t_1}(x, x_1)}{p_1(x, y)} \bigg(\prod_{j=2}^{n} p_{t_j - t_{j-1}}(x_j, x_{j-1}) \bigg) p_{1-t_n}(x_n, y)\, \mathrm{d}x_n \cdots \mathrm{d}x_1,$$

where p is the Gaussian density

$$p_t(x, y) = p_t(x - y) = \frac{1}{\sqrt{2\pi t}} \exp\Big(- \frac{(x-y)^2}{2t} \Big), \qquad x, y, \in \mathbb{R}.$$

Hint: *You can use Lemma 1.9.*

1.6 When are random variables essentially identical?

It is often useful in probability theory to be able to modify processes, for instance in order to rule out pathological behavior. Doing this pointwise in time yields to the following notions.

Definition 1.16. (Version) Let ξ and $\widetilde{\xi}$ be two random variables on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$. We say that $\widetilde{\xi}$ is a *version* of ξ if $\mathbb{P}[\widetilde{\xi} = \xi] = 1$.

It is customary in probability theory to consider random variables that agree $\mathbb{P}$-almost everywhere (as in Definition 1.16) as *identical*. (Exercise: Check that the identification up to passing to a different version is an equivalence relation on the set of all r.v.s on $(\Omega, \mathscr{F}, \mathbb{P})$.)

Definition 1.17. (Modification) Let X and $\widetilde{X}$ be two continuous-time stochastic processes on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$. We say that $\widetilde{X}$ is a *modification* of X if

$$\mathbb{P}[\widetilde{X}_t = X_t] = 1 \qquad \text{for all } t \geq 0. \tag{1.3}$$

Exercise 1.18. Show that, if $\widetilde{X}$ is a modification of X, then X and $\widetilde{X}$ have the same FDDs.

Exercise 1.19. Find an example where $\mathbb{P}$-almost every sample path $t \mapsto X_t$ is continuous, but $\mathbb{P}$-almost every sample path $t \mapsto \widetilde{X}_t$ is discontinuous. How could you rule out this phenomenon?

In Definition 1.17, for each fixed time $t \geq 0$, we allow the existence of an exceptional event

$$E(t) = \{\omega \in \Omega \mid \widetilde{X}_t(\omega) \neq X_t(\omega)\},$$

that may crucially depend on t, such that $\mathbb{P}[E(t)] = 0$. To say that the processes X and $\widetilde{X}$ are "the same," we would need the stronger property that they agree *almost surely at all times*:

$$\mathbb{P}[\widetilde{X}_t = X_t \text{ for all } t \geq 0] = 1. \tag{1.4}$$

Note that the event above is an *uncountable* intersection of events appearing in Definition 1.17,

$$\{\widetilde{X}_t = X_t \text{ for all } t \geq 0\} = \bigcap_{t \geq 0} \{\widetilde{X}_t = X_t\}.$$

It might not even be measurable — in that case, we consider an ambient event as in Remark 1.3.

Definition 1.20. (Indistinguishability) Two continuous-time stochastic processes X and $\widetilde{X}$ on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$ are *indistinguishable* if (1.4) holds.

It is customary in probability theory to consider stochastic processes as *identical up to indistinguishability.* We adopt this convention without always explicitly stating so.

If X and $\widetilde{X}$ are indistinguishable, then they clearly are modifications of each other. Exercise 1.21 below shows that the converse holds in particular for continuous processes. This is very useful for us, since we shall mainly be concerned with processes having a.s. continuous sample paths. It is important, however, to keep in mind that the notions in Definitions 1.17 and 1.20 are different in general — such subtleties arise in probability theory on uncountable spaces.

Exercise 1.21. Let X and $\widetilde{X}$ be continuous-time stochastic processes on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Prove that if the sample paths $t \mapsto X_t$ and $t \mapsto \widetilde{X}$ are almost surely (right-)continuous, then

$$\widetilde{X} \text{ is a modification of } X \qquad \Longleftrightarrow \qquad X \text{ and } \widetilde{X} \text{ are indistinguishable.}$$

1.7 Path properties of Brownian motion

Analytically, Brownian motion is a very interesting random path. It is continuous by definition, and we will show soon that it is Hölder continuous as well. Locally, Brownian paths roughly behave like $\sqrt{t}$ (see Exercise 1.22). However, this is not quite true globally — in particular, Brownian paths in fact barely fail to be $\frac{1}{2}$-Hölder continuous (see Theorem 1.23). The Law of the Iterated Logarithm (Theorem 1.29) gives the behavior of Brownian paths on this scale (see also Corollary 1.28). Brownian motion is nowhere differentiable (Theorem 1.27). Furthermore, its zero set is a Cantor type set: a closed uncountable set with no isolated points (see [MP10, Theorem 2.28]), which might be surprising. See the book [MP10] for more properties.

1.7.1 Hölder continuity and Kolmogorov's Continuity Criterion

Sample paths of Brownian motion behave locally like $\sqrt{t}$ in the following sense.

Exercise 1.22. Let B be a standard Brownian motion. Fix $t > 0$. For $n \in \mathbb{N}$, define[a]

$$\Delta_m^{(n)} := B(m\,2^{-n}\,t) - B((m-1)\,2^{-n}\,t) \qquad \text{for all } m = 1, 2, \ldots, 2^n,$$
$$Q^{(n)}(t) := \sum_{m=1}^{2^n} (\Delta_m^{(n)})^2.$$

1. What is the distribution of $\Delta_m^{(n)}$? What are $\mathbb{E}[\Delta_m^{(n)}]$ and $\mathrm{Var}(\Delta_m^{(n)})$?
2. Calculate $\mathbb{E}[Q^{(n)}(t)]$ and $\mathrm{Var}(Q^{(n)}(t))$.
   Hint: You may use the fact that for a Gaussian $\xi \sim N(0, \mathfrak{s}^2)$ and even $p \in \mathbb{N}$, we have $\mathbb{E}[X^p] = \mathfrak{s}^p (p-1)!!$.
3. Use Chebyshev's inequality (A.6) and Borel-Cantelli lemma A.13 to prove the almost sure convergence

$$Q^{(n)}(t) \quad \overset{\text{a.s.}}{\longrightarrow} \quad t \qquad \text{as } n \to \infty.$$

[a] We write here $B(t) = B_t$ to ease notation.

However, Brownian paths are not $\frac{1}{2}$-Hölder continuous, just α-Hölder for any smaller α.

Theorem 1.23. (Continuous modification of pre-Brownian motion) *Let B be a pre-Brownian motion. Then, there exists a modification $\widetilde{B}$ of B which is a.s. continuous (i.e. satisfies* BM3*). This $\widetilde{B}$ is α-Hölder continuous a.s, for any exponent $\alpha \in (0, 1/2)$:*

$$\mathbb{P}\left[\sup_{0 \le s \le t \le T} \frac{|\widetilde{B}_t - \widetilde{B}_s|}{|t-s|^\alpha} < \infty \ \text{ for all } T > 0 \right] \quad = \quad 1. \tag{1.5}$$

Proof sketch. The scaling property of pre-Brownian motion (see property 2 of Lemma 1.13) shows that if $T > 0$ is a fixed constant and B is a pre-Brownian motion, then the processes

$$(B_t)_{t\geq 0} \overset{(d)}{=} \big(\sqrt{T}\, B_{t/T}\big)_{t\geq 0}$$

are equal in distribution (i.e., in law). Thus, without loss of generality (cf. Exercise 1.24), we may assume that $T = 1$, that is, we scale $t \mapsto t/T =: t'$ and $s \mapsto s/T =: s'$. Note that then, we have

$$\frac{|B_t - B_s|}{|t-s|^\alpha} \overset{(d)}{=} \frac{\sqrt{T}\,|B_{t/T} - B_{s/T}|}{|t-s|^\alpha} \overset{(d)}{=} T^{\frac{1}{2}-\alpha}\,\frac{|B_{t/T} - B_{s/T}|}{|t/T - s/T|^\alpha} = T^{\frac{1}{2}-\alpha}\,\frac{|B_{t'} - B_{s'}|}{|t'-s'|^\alpha},$$

so the event that the ratio on the left-hand side is finite for all $0 \leq s \leq t \leq T$ is equivalent to the right-hand side being finite for all $0 \leq s' \leq t' \leq 1$. Next, since $B_t - B_s \sim N(0, t-s)$ by property BM2, we find (by an explicit computation using the Gaussian density) that

$$\mathbb{E}\big[|B_t - B_s|^p\big] \leq c_p\,|t-s|^{p/2} \qquad \text{for any } p \in \{2,4,6,\ldots\},$$

where $c_p = (p-1)(p-3)\cdots 3\cdot 1$. The assertion now follows from *Kolmogorov's Continuity Criterion* (stated below) together with Exercise 1.24. See also [LeG16, Theorem 2.9 & Corollary 2.11]. □

Exercise 1.24. Use Exercise 1.21 to check that patching copies of the above constructed modification $(\widetilde{B}_t)_{t\in[0,1]}$ for the unit interval $[0,1]$ to all time intervals $[0,1] \cup [1,2] \cup [2,3] \cup \cdots$ yields the desired modification $(\widetilde{B}_t)_{t\geq 0}$ of Brownian motion (which is well defined up to indistinguishability) that satisfies (1.5) in Theorem 1.23.

Theorem 1.25. (Kolmogorov's Continuity Criterion) *Let $X = (X_t)_{t\in[0,1]}$ be a continuous-time real-valued stochastic process. Suppose that there exist constants $p, c > 0$ and $\beta > 1$ such that for all $0 \leq s \leq t \leq 1$ we have the p:th moment bound*

$$\mathbb{E}\big[|X_t - X_s|^p\big] \leq c\,|t-s|^\beta. \tag{1.6}$$

Then, there exists a modification $\widetilde{X} = (\widetilde{X}_t)_{t\in[0,1]}$ of X which is almost surely α-Hölder continuous for any exponent $\alpha \in (0, \frac{\beta-1}{p})$.

Proof sketch. The idea is to construct a Hölder continuous modification of X by defining it on dyadic times and extending via a density argument. To begin, denote by

$$D = \bigsqcup_{n\in\mathbb{N}} D_n, \qquad D_n = \big\{k\,2^{-n} \mid k \in \{0,1,\ldots,2^n\}\big\},$$

the *dyadic times* on $[0,1]$; for instance,

$$D_1 = \big\{0, \tfrac{1}{2}, 1\big\}, \qquad D_2 = \big\{0, \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{3}{4}, 1\big\},$$

and $D_{n+1} \subset D_n$ for all n. Note that D is a *dense* subset of $[0,1]$, that is, $\overline{D} = [0,1]$. Therefore, for any time $t \in [0,1]$, we can find a sequence $t_j \to t$, with $j \to \infty$, such that $t_j \in D$ for all j. Define, for every $\omega \in \Omega$ and $t \in [0,1]$,

$$\widetilde{X}_t(\omega) := \lim_{\substack{t_j \to t \\ t_j \in D\ \forall\, j\in\mathbb{N}}} X_{t_j}(\omega), \quad \text{if the limit exists,} \tag{1.7}$$

and $\widetilde{X}_t(\omega) := 0$ otherwise. The key to the proof now is that the assumption (1.6) implies that

$$\mathbb{P}\,[t \mapsto X_t \text{ is } \alpha\text{-Hölder continuous on } D] = 1. \tag{1.8}$$

Step 1. We will sketch why (1.8) holds below — but before that, let us see how to conclude the proof knowing (1.8).

Step 2. Recalling that any function that is α-Hölder continuous on the dense set D has a *unique α-Hölder continuous extension* to $[0,1] = \overline{D}$, we see that $t \mapsto X_t(\omega)$ has a unique α-Hölder continuous extension to $[0,1]$ for almost all $\omega \in \Omega$ (that is, almost surely). In particular, the limit in (1.7) exists almost surely for all $t \in [0,1]$ (and, importantly, the exceptional set of $\omega \in \Omega$ for which the limit does not exist is independent of t, as we will see below).

Step 3. Before verifying (1.8), let us also note that the process $\widetilde{X} = (\widetilde{X}_t)_{t\in[0,1]}$ thus defined is a modification of X, because for each time $t \in [0,1]$, the assumption (1.6) shows that

$$\mathbb{E}\big[|X_t - \widetilde{X}_t|^p\big] = \mathbb{E}\Big[\lim_{\substack{t_j\to t\\ t_j\in D\,\forall j\in\mathbb{N}}} |X_t - X_{t_j}|^p\Big] \le \liminf_{\substack{t_j\to t\\ t_j\in D\,\forall j\in\mathbb{N}}} \mathbb{E}\big[|X_t - X_{t_j}|^p\big] \quad \text{[by Fatou's lemma A.17]}$$

$$\le c \lim_{\substack{t_j\to t\\ t_j\in D\,\forall j\in\mathbb{N}}} |t - t_j|^\beta = 0, \quad \text{[by (1.6)]}$$

which implies that (1.3) holds for each $t \ge 0$:

$$\mathbb{P}\big[|X_t - \widetilde{X}_t|^p = 0\big] = 1 \qquad \Longrightarrow \qquad \mathbb{P}[X_t = \widetilde{X}_t] = 1.$$

Back to Step 1. Let us now show how the assumption (1.6) implies that X is α-Hölder continuous on D with probability one. The idea is a Borel-Cantelli argument, frequently used in probability theory. Consider the increments[5] of X on dyadic times,

$$W_k^{(n)} := X(k\,2^{-n}) - X((k-1)\,2^{-n}).$$

Using the assumption (1.6), we obtain

$$\begin{aligned}\mathbb{P}\big[|W_k^{(n)}| > 2^{-n\alpha}\big] &= \mathbb{P}\big[|W_k^{(n)}|^p > 2^{-np\alpha}\big] \\ &\le 2^{np\alpha}\,\mathbb{E}\big[|W_k^{(n)}|^p\big] && \text{[by Markov's Inequality (A.5)]} \\ &\le c\,2^{np\alpha}\,|k\,2^{-n} - (k-1)\,2^{-n}|^\beta = c\,2^{n(p\alpha-\beta)}, && \text{[by (1.6)]}\end{aligned}$$

where $p\alpha - \beta < -1$. Using Union Bound (A.4), we thus obtain

$$\mathbb{P}\Big[\max_{k\in\{1,2,\ldots,2^n\}} |W_k^{(n)}| > 2^{-n\alpha}\Big] \le c \sum_{k=1}^{2^n} 2^{n(p\alpha-\beta)} \le c\,2^{n(1+p\alpha-\beta)},$$

and summing over $n \in \mathbb{N}$, because $1 + p\alpha - \beta < 0$, the following geometric series converges:

$$\sum_{n=1}^{\infty} \mathbb{P}\Big[\max_{k\in\{1,2,\ldots,2^n\}} |W_k^{(n)}| > 2^{-n\alpha}\Big] \le c \sum_{n=1}^{\infty} 2^{n(1+p\alpha-\beta)} < \infty.$$

Therefore, the (first) Borel-Cantelli lemma A.13 implies that almost surely, we have

$$\max_{k\in\{1,2,\ldots,2^n\}} |W_k^{(n)}| \le 2^{-n\alpha}$$

except for possibly finitely many n. In particular, we have

$$\sup_{n\in\mathbb{N}} \max_{k\in\{1,2,\ldots,2^n\}} \frac{|W_k^{(n)}|}{2^{-n\alpha}} = \sup_{n\in\mathbb{N}} \max_{k\in\{1,2,\ldots,2^n\}} \frac{|X(k\,2^{-n}) - X((k-1)\,2^{-n})|}{2^{-n\alpha}} \le M(\omega),$$

[5] For clarity, we write here $X_t = X(t)$.

where $M(\omega)$ is a random but almost surely finite constant. It is now straightforward to show that on the event $\{M(\omega) < \infty\}$, the map $t \mapsto X_t(\omega)$ is α-Hölder continuous for all dyadic times $t \in D$ (see [LeG16, Lemma 2.10]). This concludes Step 1, and together with Steps 2-3, the proof of Kolmogorov's Continuity Criterion. See [LeG16, Theorem 2.9] for more details. □

Remark 1.26. Since a continuous modification of pre-Brownian motion is unique up to indistinguishability, we will from now on speak of such a continuous modification as "the" Brownian motion and denote it as B. (So we will not consider pre-Brownian motions as separate objects.)

The proof of Theorem 1.23 shows that we can take in (1.5)

$$\alpha \;<\; \frac{\frac{p}{2}-1}{p} \;=\; \frac{1}{2}-\frac{1}{p} \;\xrightarrow{p\to\infty}\; \frac{1}{2}.$$

However, the exponent $\alpha < 1/2$ is the best we can hope for, see Theorems 1.27 and 1.29.

1.7.2 Non-Hölder continuity and Law of the Iterated Logarithm ($\star$)

Furthermore, Brownian motion B is in fact nowhere differentiable (it has no tangent anywhere).

Theorem 1.27. *For any exponent $\alpha \in (1/2, \infty)$, we have*

$$\mathbb{P}\left[\text{ for all } t \geq 0 \quad \limsup_{h\to 0+} \frac{|B_{t+h}-B_t|}{h^\alpha} = +\infty\right] \;=\; 1.$$

Proof. See [MP10, Theorem 1.30 & Remark 1.31]. □

Corollary 1.28. (Paley-Wiener-Zygmund theorem) *Almost surely, $t \mapsto B_t$ is nowhere differentiable, and we have a.s.*

$$\text{for all } t \geq 0 \quad \limsup_{h\to 0+} \frac{|B_{t+h}-B_t|}{\sqrt{h}} = +\infty \qquad \text{and} \qquad \liminf_{h\to 0+} \frac{|B_{t+h}-B_t|}{\sqrt{h}} = -\infty.$$

Proof. See [MP10, Theorem 1.30]. □

Theorem 1.29. (Khintchine's law of the iterated logarithm) *We have a.s.*

$$\limsup_{h\to 0+} \frac{|B_h|}{\sqrt{-2\,h\log\log h}} = +1 \qquad \text{and} \qquad \liminf_{h\to 0+} \frac{|B_h|}{\sqrt{-2\,h\log\log h}} = -1.$$

Proof. See [MP10, Theorem 5.1 & Corollary 5.3]. □

1.8 Wiener measure ($\star$)

We have defined Brownian motion (Definition 1.2) with the requirement BM3 that its sample paths $t \mapsto B_t$ are continuous *almost surely*. However, we did not yet address the basic question:

On which probability space $(\Omega, \mathscr{F}, \mathbb{P})$ *is Brownian motion defined?*

One way to answer this question is to construct Brownian motion explicitly, e.g, as a weak limit of simple random walk (cf. Section 1.3), or as in Lévy's construction (cf. [MP10, Theorem 1.3]), which directly yield $\Omega = C([0,\infty),\mathbb{R})$. However, the construction of *pre*-Brownian motion as a Gaussian process, say, gives a probability space where, a priori, the *continuity* of the sample paths is not guaranteed. In the next Section 1.8.1 we shall briefly address this problem.

In Section 1.8.2, we discuss the *Wiener space* of continuous functions, which gives a natural ambient probability space for Brownian motion. Note that different constructions of Brownian motion yield a priori different probability spaces $(\Omega, \mathscr{F}, \mathbb{P})$ for it. It turns out that for the Wiener space, the probability measure is independent of the construction of Brownian motion.

1.8.1 Measurability issues for Brownian motion ($\star$)

Let us consider more specifically the problem of the choice of an appropriate probability space $(\Omega, \mathscr{F}, \mathbb{P})$ for Brownian motion. In order for property BM3 in Definition 1.2 to make sense, the minimal requirement (cf. Remark 1.3) is that the set

$$\{\omega \in \Omega \mid t \mapsto B_t(\omega) \text{ is continuous}\}$$

contains an event $E \in \mathscr{F}$ such that $\mathbb{P}[E] = 1$. This gives some additional freedom for us.

As a natural candidate for the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, we could take

$$\Omega = \mathbb{R}^{[0,\infty)} := \{\text{functions } f : [0,\infty) \to \mathbb{R}\},$$

endowed with its cylinder sigma-algebra $\mathscr{F}(\mathbb{R}^{[0,\infty)})$ generated by the cylinder events

$$\{f(t_1) \in A_1, \ldots, f(t_n) \in A_n\}, \quad 0 \le t_1 < \cdots < t_n \text{ and } A_1, \ldots, A_n \in \mathscr{B}(\mathbb{R}).$$

However, one can show that the subset

$$C([0,\infty),\mathbb{R}) := \{f : [0,\infty) \to \mathbb{R} \mid f \text{ is continuous}\} \subset \mathbb{R}^{[0,\infty)}$$

comprising continuous functions is not $\mathscr{F}(\mathbb{R}^{[0,\infty)})$-measurable. To remedy this, let us consider a countable dense set $D \subset [0,\infty)$, e.g., $D = \mathbb{Q} \cap [0,\infty)$. Define

$$\begin{aligned} C^u_D([0,\infty),\mathbb{R}) &:= \{f \in \mathbb{R}^{[0,\infty)} \mid t \mapsto f(t) \text{ is uniformly continuous on bounded subsets of } D\} \\ &\in \mathscr{F}(\mathbb{R}^{[0,\infty)}). \end{aligned}$$

With this in mind, let us revisit Theorem 1.23 with a pre-Brownian motion B on the probability space $(\mathbb{R}^{[0,\infty)}, \mathscr{F}(\mathbb{R}^{[0,\infty)}), \mathbb{P})$ arising, for example, from constructing pre-Brownian motion as a Gaussian process. Then, Kolmogorov's Continuity Criterion (Theorem 1.25) gives rise to a modification $\widetilde{B}$ of the pre-Brownian motion which satisfies

$$\mathbb{P}\big[\widetilde{B} \in C^u_D([0,\infty),\mathbb{R})\big] = 1,$$

since Hölder continuous functions are uniformly continuous. Taking this event as E, we have

$$\{\widetilde{B} \in C^u_D([0,\infty),\mathbb{R})\} \subset \{\widetilde{B} \in C([0,\infty),\mathbb{R})\},$$

which in the context of Remark 1.3 shows that on the probability space $(\mathbb{R}^{[0,\infty)}, \mathscr{F}(\mathbb{R}^{[0,\infty)}), \mathbb{P})$, the sample paths of $\widetilde{B}$ are continuous almost surely.

1.8.2 Wiener measure and canonical Brownian motion ($\star$)

In fact, the first rigorous construction of Brownian motion is quite abstract, due to Norbert Wiener [Wie23]. His approach was later generalized to construct more general Gaussian measures on separable Hilbert spaces (called abstract Wiener spaces) [Jan08].

$\Omega = C([0,\infty),\mathbb{R})$ endowed with the sigma-algebra $\mathscr{F} = \mathscr{W}$ generated by the cylinder events

$$\{B_{t_1} \in A_1, \ldots, B_{t_n} \in A_n\}, \quad n \in \mathbb{N},\ 0 < t_1 < \cdots < t_n, \text{ and } A_1, \ldots, A_n \in \mathscr{B}(\mathbb{R})$$

is determined by the FDDs of B. Since every construction for a Brownian motion B gives the same FDDs, the following definition does not depend on the choice of the construction.

Definition 1.30. (Wiener measure) Let B be a (modification of) Brownian motion on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Define a probability measure $\mathbb{W}$ on the measurable space $(C([0,\infty),\mathbb{R}), \mathscr{W})$ as the law of B, that is,

$$\mathbb{W}[A] := \mathbb{P}[(t \mapsto B_t(\omega)) \in A], \qquad A \in \mathscr{W},$$

where $(t \mapsto B_t(\omega))$ stands for the random $\mathbb{P}$-a.s. continuous function $[0,\infty) \to \mathbb{R}$.

- ▷ The resulting probability space $(C([0,\infty),\mathbb{R}), \mathscr{W}, \mathbb{W})$ is called the *Wiener space*.
- ▷ The modification W of Brownian motion on the Wiener space induced by B is called the *Wiener process* (or the "canonical" Brownian motion).

By construction — regardless of how the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ for a modification of Brownian motion is defined — the Wiener measure $\mathbb{W}$ is supported on the measurable space $(C([0,\infty),\mathbb{R}), \mathscr{W})$ of continuous functions, so all (not just almost all) sample paths of the Wiener process W are continuous. This is useful, for example, because it guarantees that W is *progressively measurable* (see Corollary 5.20) in the sense of Definition 5.9, thanks to Proposition 5.19, with respect to any of its natural filtrations discussed in Section 6.1.

2 Conditional expected value

Throughout, we fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. In this section, we will consider real-valued random variables (r.v.) $\xi : \Omega \to \mathbb{R}$. The r.v. ξ is called[6]

- ▷ *integrable*, and denoted $\xi \in L^1(\mathbb{P})$, if $\mathbb{E}[|\xi|] < \infty$;
- ▷ *square-integrable*, and denoted $\xi \in L^2(\mathbb{P})$, if $\mathbb{E}[\xi^2] < \infty$.
- ▷ Recall that the Cauchy-Schwarz Inequality (Lemma 2.17) implies that $L^2(\mathbb{P}) \subset L^1(\mathbb{P})$, but there are elements in the latter space that do not belong to the former.

The space $L^2(\mathbb{P})$ has a very good structure: it is a *Hilbert space*. For this reason, one can do analysis on the space $L^2(\mathbb{P})$ much analogously to that on Euclidean spaces. Going beyond this space, the integrability condition for $L^1(\mathbb{P})$ is often times the minimal technical requirement for many arguments to carry through. One can extend, for example, the existence of conditional expected values for random variables from $L^2(\mathbb{P})$ to $L^1(\mathbb{P})$ via an approximation argument. (See [Kyt20, Appendix E] for details.) We shall discuss the space $L^2(\mathbb{P})$ further in Section 2.4.

2.1 Definition and uniqueness of conditional expected value

The idea of conditional expected value is to estimate the value of a random variable ξ with *partial information* available from a sub-sigma-algebra $\mathscr{G} \subset \mathscr{F}$. When seeking an optimal such estimate, one wants to minimize the committed error, which can be done (under certain assumptions) by projecting ξ to a random variable that is $\mathscr{G}$-measurable and represents ξ as well as possible given the information from $\mathscr{G}$ (as phrased in property CE3 below). This generalizes the familiar concept of conditional probability — see Exercise 2.4 for an illustration of the idea.

Before stating the formal definition, let us ponder what properties the expected value of ξ conditionally on $\mathscr{G} \subset \mathscr{F}$ should satisfy. Let us denote it by $\mathbb{E}[\xi \,|\, \mathscr{G}]$. Since $\mathscr{G}$ should give *additional and useful* information, the wish-list contains at least the following properties:

- ▷ $\mathbb{E}[\xi \,|\, \mathscr{G}]$ should be $\mathscr{F}$-measurable (that is, a random variable);
- ▷ $\mathbb{E}[\xi \,|\, \mathscr{G}]$ should be $\mathscr{G}$-measurable (it should "know" $\mathscr{G}$);
- ▷ $\mathbb{E}[\xi \,|\, \mathscr{G}]$ should be somehow uniquely determined (as stated below);
- ▷ $\mathbb{E}[\xi \,|\, \mathscr{G}]$ should be related to ξ in the "best possible way" given $\mathscr{G}$.

To investigate the last point, note that in $L^2(\mathbb{P})$, we have

$$\eta = 0 \quad \mathbb{P}\text{-a.s.} \qquad \Longleftrightarrow \qquad \eta^2 = 0 \quad \mathbb{P}\text{-a.s.} \qquad \Longleftrightarrow \qquad \mathbb{E}[\eta^2] = 0.$$

Hence, to estimate the error made in the comparison $\eta := \xi - \mathbb{E}[\xi \,|\, \mathscr{G}]$ is equivalent to studying

$$\mathbb{E}\big[\big(\xi - \mathbb{E}[\xi \,|\, \mathscr{G}]\big)^2\big] = \big\| \xi - \mathbb{E}[\xi \,|\, \mathscr{G}] \big\|_{L^2}^2,$$

which makes sense whenever $\eta := \xi - \mathbb{E}[\xi \,|\, \mathscr{G}] \in L^2(\mathbb{P})$. We will see later in Section 2.4 that the error is minimized if we define $\mathbb{E}[\xi \,|\, \mathscr{G}]$ as a suitable *orthogonal projection* (see Figure 2.1).

[6] Strictly speaking, elements in the spaces $L^1(\mathbb{P})$ and $L^2(\mathbb{P})$ should be defined as equivalence classes of measurable functions $\Omega \to \mathbb{R}$ which agree $\mathbb{P}$-almost everywhere in Ω (i.e., which agree almost surely, cf. Definition 1.16).

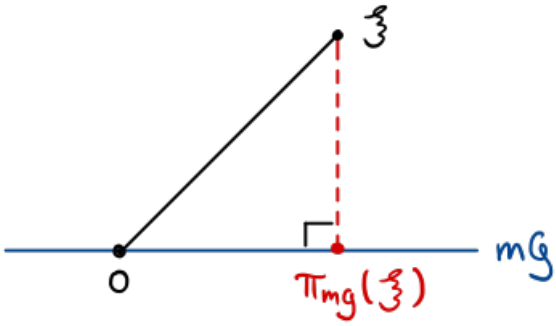


Figure 2.1. Illustration of the orthogonal projection of $\xi \in L^2(\mathbb{P})$ onto $\mathrm{m}\mathscr{G} \cap L^2(\mathbb{P})$. See also Proposition 2.24.

Definition 2.1. (Conditional expected value) Let $\xi \in L^1(\mathbb{P})$ and let $\mathscr{G} \subset \mathscr{F}$ be a sub-sigma-algebra. A random variable $\hat{\xi} : \Omega \to \mathbb{R}$ is (a version of) the *conditional expected value* of the random variable ξ *given* $\mathscr{G}$ if

CE1. (it is integrable): $\hat{\xi} \in L^1(\mathbb{P})$,

CE2. (it is $\mathscr{G}$-measurable): $\{\hat{\xi} \in A\} = \{\omega \in \Omega \mid \xi(\omega) \in A\} \in \mathscr{G}$ for all (Borel) $A \in \mathscr{B}(\mathbb{R})$,

CE3. (it represents "best" information): $\mathbb{E}[\hat{\xi}\, \mathbb{1}_G] = \mathbb{E}[\xi\, \mathbb{1}_G]$ for all $G \in \mathscr{G}$.

We denote the conditional expected value by $\hat{\xi} =: \mathbb{E}[\xi \,|\, \mathscr{G}]$.

It is not hard to see that a conditional expected value is (essentially) unique (Lemma 2.2). However, from its abstract Definition 2.1 it is not obvious at all that it actually exists.

Lemma 2.2. *Let $\xi \in L^1(\mathbb{P})$ and let $\mathscr{G} \subset \mathscr{F}$ be a sub-sigma-algebra. Suppose both $\hat{\xi}$ and ξ' are conditional expected values of ξ given $\mathscr{G}$. Then, we have $\hat{\xi} = \xi'$ almost surely.*

The lemma says that the conditional expected value is unique up to passing to a different *version*, as in Definition 1.16 (in other words, we regard it as only defined $\mathbb{P}$-almost everywhere).

Proof. Consider the sets $G_n := \left\{\omega \in \Omega \mid \hat{\xi}(\omega) - \xi'(\omega) \geq \frac{1}{n}\right\} \in \mathscr{G}$. For each fixed $n \in \mathbb{N}$, we have

$$\begin{aligned}
\tfrac{1}{n}\,\mathbb{P}[G_n] &\leq \mathbb{E}[(\hat{\xi} - \xi')\,\mathbb{1}_{G_n}] && \text{[by Markov's Inequality (A.5)]} \\
&= \mathbb{E}[\hat{\xi}\,\mathbb{1}_{G_n}] - \mathbb{E}[\xi'\,\mathbb{1}_{G_n}] && \text{[by linearity]} \\
&= \mathbb{E}[\xi\,\mathbb{1}_{G_n}] - \mathbb{E}[\xi\,\mathbb{1}_{G_n}] = 0 && \text{[by property (CE3)]} \\
\implies \quad \mathbb{P}[G_n] &= 0.
\end{aligned}$$

Using Union Bound (A.4), we thus obtain

$$\mathbb{P}[\hat{\xi} > \xi'] = \mathbb{P}\Big[\bigcup_{n\in\mathbb{N}} G_n\Big] \leq \sum_{n\in\mathbb{N}} \mathbb{P}[G_n] = 0 \qquad \implies \qquad \mathbb{P}[\hat{\xi} > \xi'] = 0.$$

Similarly, by symmetry, we find $\mathbb{P}[\hat{\xi} < \xi'] = 0$, so we conclude that $\mathbb{P}[\hat{\xi} = \xi'] = 1$. □

Before addressing its existence, we record a few motivating exercises that should give more intuition on the concept of conditional expected value. Recall that, if $\chi : \Omega \to \mathbb{R}$ is a random variable, the *sigma-algebra generated by* χ is the smallest sigma-algebra $\sigma(\chi)$ such that χ is $\sigma(\chi)$-measurable. We denote the conditional expected value of ξ given $\sigma(\chi)$ briefly by

$$\mathbb{E}[\xi\,|\,\chi] := \mathbb{E}[\xi\,|\,\sigma(\chi)].$$

Exercise 2.3. Let ξ and η be independent Bernoulli(p) distributed[a] random variables with parameter $p \in (0,1)$. Define $\chi = \mathbb{1}\{\xi + \eta = 0\}$. Calculate $\mathbb{E}[\xi\,|\,\chi]$ and $\mathbb{E}[\eta\,|\,\chi]$.

[a] Bernoulli distributed here means that $\mathbb{P}[\xi = 1] = \mathbb{P}[\eta = 1] = p$ and $\mathbb{P}[\xi = 0] = \mathbb{P}[\eta = 0] = 1 - p$.

Exercise 2.4. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $(\Omega_j)_{j\geq 1}$ be a finite or countably infinite partition of Ω into $\mathscr{F}$-measurable sets: $\Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$ and $\bigcup_{j\geq 1} \Omega_j = \Omega$. Assume that each Ω_j has positive probability. Let $\mathscr{G}$ be the sigma-algebra generated by $(\Omega_j)_{j\geq 1}$. Let $\xi : \Omega \to \mathbb{R}$ be an integrable random variable. Define

$$\eta : \Omega \to \mathbb{R}, \qquad \eta(\omega) := \frac{\mathbb{E}[\mathbb{1}_{\Omega_j}\xi]}{\mathbb{P}[\Omega_j]} \qquad \text{for } \omega \in \Omega_j.$$

Show that η is a version of $\mathbb{E}[\xi\,|\,\mathscr{G}]$.

Exercise 2.5. Let $\mathscr{G} \subset \mathscr{F}$ be a sigma-algebra such that for all $E \in \mathscr{G}$, we have $\mathbb{P}[E] \in \{0,1\}$. Show that for any integrable random variable $\xi : \Omega \to \mathbb{R}$, we have $\mathbb{E}[\xi\,|\,\mathscr{G}] = \mathbb{E}[\xi]$ almost surely.

Exercise 2.6. Suppose that ξ and η are real-valued random variables, which have a joint probability density $f_{\xi,\eta} : \mathbb{R}^2 \to [0,\infty)$. Recall that η then has a probability density given by

$$f_\eta(z) = \int_{\mathbb{R}} f_{\xi,\eta}(x,z)\,\mathrm{d}x.$$

Define a "conditional density" $f_{\xi\,|\,\eta} : \mathbb{R}^2 \to [0,\infty)$ by

$$f_{\xi\,|\,\eta}(x,z) := \begin{cases} \frac{f_{\xi,\eta}(x,z)}{f_\eta(z)}, & \text{if } f_\eta(z) > 0 \\ 0, & \text{if } f_\eta(z) = 0. \end{cases}$$

Prove that if $h : \mathbb{R} \to \mathbb{R}$ is a Borel function such that $h(\xi)$ is $\mathbb{P}$-integrable, then

$$g(z) := \int_{\mathbb{R}} h(x)\, f_{\xi\,|\,\eta}(x,z)\,\mathrm{d}x$$

defines a Borel function g such that $g(\eta)$ is a version of the conditional expected value $\mathbb{E}[h(\xi)\,|\,\sigma(\eta)]$ of $h(\xi)$ with respect to the sigma-algebra $\sigma(\eta)$ generated by η.

2.2 Existence of conditional expected value in $L^2(\mathbb{P})$

Let us now turn to the question whether the conditional expected value in Definition 2.1 actually exists. Indeed, thanks to the assumption $\xi \in L^1(\mathbb{P})$ in Definition 2.1, one can actually construct a version $\mathbb{E}[\xi\,|\,\mathscr{G}]$ by using the structure of the spaces $L^2(\mathbb{P}) \subset L^1(\mathbb{P})$. The construction is easier and more concrete if $\xi \in L^2(\mathbb{P})$, while the case of $\xi \in L^1(\mathbb{P}) \smallsetminus L^2(\mathbb{P})$ follows via a relatively standard truncation argument.

- ▷ If $\xi \in L^2(\mathbb{P})$, then $\mathbb{E}[\xi\,|\,\mathscr{G}]$ can be constructed explicitly as the *orthogonal projection* of ξ in $L^2(\mathbb{P})$ onto the subspace consisting of $\mathscr{G}$-measurable random variables, see Proposition 2.7.

▷ If $\xi \in L^1(\mathbb{P})$ but $\xi \notin L^2(\mathbb{P})$, then one can truncate it suitably to reduce to the $L^2(\mathbb{P})$ case.

To construct $\mathbb{E}[\xi \,|\, \mathscr{G}]$ for $\xi \in L^2(\mathbb{P})$, we will use properties of $L^2(\mathbb{P})$ from Section 2.4.

Proposition 2.7. (Conditional expected value in L^2) *Suppose that $\xi \in L^2(\mathbb{P})$. Let $\mathscr{G} \subset \mathscr{F}$ be a sub-sigma-algebra. Let $\pi_{\mathrm{m}\mathscr{G}}(\xi)$ be the orthogonal projection of ξ onto $\mathrm{m}\mathscr{G} \cap L^2(\mathbb{P})$,*

$$\mathrm{m}\mathscr{G} := \{\chi : \Omega \to \mathbb{R} \mid \chi \text{ is } \mathscr{G}\text{-measurable}\},$$

defined in Proposition 2.24. Then, property CE3 *holds for $\pi_{\mathrm{m}\mathscr{G}}(\xi)$:*

$$\mathbb{E}[\pi_{\mathrm{m}\mathscr{G}}(\xi)\, \mathbb{1}_G] = \mathbb{E}[\xi\, \mathbb{1}_G] \quad \text{for all } G \in \mathscr{G}.$$

Proof. Fix $G \in \mathscr{G}$. The indicator function $\mathbb{1}_G$ is $\mathscr{G}$-measurable and bounded, so $\mathbb{1}_G \in \mathrm{m}\mathscr{G} \cap L^2(\mathbb{P})$. Therefore, by item 1 of Proposition 2.24, we have $(\xi - \pi_{\mathrm{m}\mathscr{G}}(\xi)) \perp \mathbb{1}_G$. This gives property CE3:

$$0 \;=\; \langle \xi - \pi_{\mathrm{m}\mathscr{G}}(\xi), \mathbb{1}_G \rangle_{L^2} \;=\; \langle \xi, \mathbb{1}_G \rangle_{L^2} \;-\; \langle \pi_{\mathrm{m}\mathscr{G}}(\xi), \mathbb{1}_G \rangle_{L^2} \;=\; \mathbb{E}[\xi\, \mathbb{1}_G] - \mathbb{E}[\pi_{\mathrm{m}\mathscr{G}}(\xi)\, \mathbb{1}_G],$$

where $\langle \xi, \eta \rangle_{L^2} := \mathbb{E}[\xi\, \eta]$ for all $\xi, \eta \in L^2(\mathbb{P})$ (see Section 2.4). □

Upshot. Proposition 2.7 shows that $\pi_{\mathrm{m}\mathscr{G}}(\xi) = \mathbb{E}[\xi \,|\, \mathscr{G}] \in L^2(\mathbb{P})$. Recall also that it is *unique* by Lemma 2.2 — we have thus constructed $\mathbb{E}[\xi \,|\, \mathscr{G}]$ for all $\xi \in L^2(\mathbb{P})$. We will not discuss the case where $\xi \in L^1(\mathbb{P}) \smallsetminus L^2(\mathbb{P})$ here, but just refer to [Kyt20, Proposition E.13 in Appendix E].

2.3 Key properties of conditional expected value

Let us record here some important and useful properties of conditional expected value. We leave the verification of these properties as an exercise – they follow quite easily from the definitions (and the "standard machine").

Lemma 2.8. *Let $\xi \in L^1(\mathbb{P})$ and let $\mathscr{G} \subset \mathscr{F}$ be a sub-sigma-algebra. The conditional expected value $\mathbb{E}[\xi \,|\, \mathscr{G}]$ satisfies the following properties (almost surely).*

0. If $\xi \geq 0$ almost surely, then $\mathbb{E}[\xi \,|\, \mathscr{G}] \geq 0$ almost surely.

1. (Known quantity): *If ξ is itself $\mathscr{G}$-measurable, then $\mathbb{E}[\xi \,|\, \mathscr{G}] = \xi$.*

2. (Unbiased): $\mathbb{E}[\mathbb{E}[\xi \,|\, \mathscr{G}]] = \mathbb{E}[\xi]$.

3. (Linearity): *$\xi \mapsto \mathbb{E}[\xi \,|\, \mathscr{G}]$ is linear.*

4. (Tower property): *If $\mathscr{H} \subset \mathscr{G} \subset \mathscr{F}$ are sub-sigma-algebras, then*

$$\mathbb{E}\big[\mathbb{E}[\xi \,|\, \mathscr{G}] \,\big|\, \mathscr{H}\big] = \mathbb{E}[\xi \,|\, \mathscr{H}].$$

5. (Known can be taken out): *If η is $\mathscr{G}$-measurable and $\eta\,\xi \in L^1(\mathbb{P})$, then*

$$\mathbb{E}[\eta\, \xi \,|\, \mathscr{G}] = \eta\, \mathbb{E}[\xi \,|\, \mathscr{G}].$$

6. (Independent yields no information): *If ξ is independent of $\mathscr{G}$, then*

$$\mathbb{E}[\xi \,|\, \mathscr{G}] = \mathbb{E}[\xi].$$

Exercise 2.9. Prove Lemma 2.8.

Analogues of many results for the usual expected value also hold for conditional expected value.

Lemma 2.10. *Let $\xi \in L^1(\mathbb{P})$ and let $\mathscr{G} \subset \mathscr{F}$ be a sub-sigma-algebra. The conditional expected value $\mathbb{E}[\xi|\mathscr{G}]$ satisfies the following properties.*

1. (Conditional Jensen's inequality): *If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and $\varphi(\xi) \in L^1(\mathbb{P})$, then*
$$\varphi\big(\mathbb{E}[\xi|\mathscr{G}]\big) \le \mathbb{E}[\varphi(\xi)|\mathscr{G}] \qquad \text{almost surely.} \tag{2.1}$$

2. (Conditional Monotone Convergence Theorem): *Suppose that $\xi_1, \xi_2, \ldots \in L^1(\mathbb{P})$ is a sequence of non-negative random variables such that $\xi_n \le \xi_{n+1}$ for all $n \in \mathbb{N}$, and*
$$\xi_n \xrightarrow{\text{a.s.}} \xi.$$
Then, we have
$$\mathbb{E}[\xi_n|\mathscr{G}] \xrightarrow{\text{a.s.}} \mathbb{E}[\xi|\mathscr{G}].$$

3. (Conditional Dominated Convergence Theorem): *Suppose that $\xi_1, \xi_2, \ldots \in L^1(\mathbb{P})$ is a sequence of random variables for which there exists $\eta \in L^1(\mathbb{P})$ with $|\xi_n| \le \eta$ for all $n \in \mathbb{N}$, and*
$$\xi_n \xrightarrow{\text{a.s.}} \xi.$$
Then, we have
$$\mathbb{E}[\xi_n|\mathscr{G}] \xrightarrow{\text{a.s.}} \mathbb{E}[\xi|\mathscr{G}].$$

4. (Conditional Fatou's lemma): *Suppose that $\xi_1, \xi_2, \ldots \in L^1(\mathbb{P})$ is a sequence of non-negative random variables. Then, for any sub-sigma-algebra $\mathscr{G} \subset \mathscr{F}$, we have*
$$\mathbb{E}\Big[\liminf_{n\to\infty} \xi_n \,|\mathscr{G}\Big] \le \liminf_{n\to\infty} \mathbb{E}[\xi_n|\mathscr{G}] \qquad \text{almost surely.}$$

Exercise 2.11. Prove Lemma 2.10.

2.4 Background: Hilbert spaces and the space $L^2(\mathbb{P})$

In this section, we consider the space $L^2(\mathbb{P})$ of square-integrable random variables $(\Omega, \mathscr{F}, \mathbb{P})$ in more detail. In particular, we will show that $L^2(\mathbb{P})$ is a Hilbert space (see Theorem 2.20).

In fact, elements of $L^2(\mathbb{P})$ should be defined up to $\mathbb{P}$-almost sure equivalence: if $\mathbb{P}[\xi = \xi'] = 1$, then $\xi = \xi' \in L^2(\mathbb{P})$. In other words, elements in $L^2(\mathbb{P})$ are *identified* up to the equivalence relation induced by identifying[7] elements with the same value of the norm $\|\cdot\|_{L^2}$:
$$\eta = 0 \in L^2(\mathbb{P}) \qquad \Longleftrightarrow \qquad \|\eta\|_{L^2} := \sqrt{\mathbb{E}[\eta^2]} = 0 \qquad \Longleftrightarrow \qquad \eta = 0 \ \ \mathbb{P}\text{-a.s.},$$
which is equivalent to identifying them up to passing to a different version, cf. Definition 1.16.

[7] Strictly speaking $L^2(\mathbb{P})$ should be defined as the quotient space up to this equivalence relation.

2.4.1 Hilbert spaces and normed vector spaces

Definition 2.12. (Hilbert space) V is a *Hilbert space* if it has the following structures.

▷ It is a real *vector space*:

$$a\,v + b\,w \in V \quad \text{for all } a, b \in \mathbb{R} \text{ and } v, w \in V.$$

▷ It has an *inner product*, that is, a symmetric positive-definite $\mathbb{R}$-bilinear map

$$\langle \cdot, \cdot \rangle : V \times V \to [0, \infty),$$

* (symmetric): $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in V$;
* (positive-definite): $\langle v, v \rangle > 0$ for all $v \in V \setminus \{0\}$.
* ($\mathbb{R}$-linear[a]): $\langle a\,u + b\,v, w \rangle = a\,\langle u, w \rangle + b\,\langle v, w \rangle$ for all $a, b \in \mathbb{R}$ and for all $u, v, w \in V$.

▷ It has a *metric* (or *distance function*)

$$\mathrm{dist}(\cdot, \cdot) : V \times V \to [0, \infty)$$

induced by the inner product: $\mathrm{dist}(v, w) := \sqrt{\langle v - w, v - w \rangle}$.

▷ (V, dist) is a *complete metric space* (see Definition 2.16 and the remark below).

The property that (V, dist) is a complete metric space is equivalent[b] to the property that $(V, ||\cdot||)$ is a *complete normed vector space* (i.e., a Banach space) with respect to the *norm*

$$||\cdot|| : V \to [0, \infty)$$

induced by the inner product: $||\,v\,|| := \sqrt{\langle v, v \rangle}$.

[a] Note that $\mathbb{R}$-bilinearity follows from $\mathbb{R}$-linearity and symmetry.
[b] Note that $\mathrm{dist}(v, w) = ||\,v - w\,||$.

Definition 2.13. Let V be a Hilbert space. For two vectors $\xi, \eta \in V$, we write

$$\xi \perp \eta \qquad \Longleftrightarrow \qquad \langle \xi, \eta \rangle = 0.$$

In this case, we say that the two vectors are *orthogonal.*
If $W \subset V$ is a vector subspace, the vector $\xi \in V$ is said to be *orthogonal to* W if

$$\xi \perp \eta \qquad \text{for all } \eta \in W,$$

and in this case, we write $\xi \perp W$. We define the *orthogonal complement* of W in V as

$$W^{\perp} := \{\xi \in V \mid \xi \perp \eta \text{ for all } \eta \in W\} \subset V.$$

Exercise 2.14. Let V be a Hilbert space. For a subspace $W \subset V$, define its closure as

$$\overline{W} := \{\eta \in V \mid \exists \text{ sequence } (\eta_n)_{n\in\mathbb{N}} \text{ in } W \text{ such that } \|\eta_n - \eta\| \to 0 \text{ as } n \to 0\},$$

that is, the set of all vectors in V that can be approximated by vectors in W in the normed topology of V. Prove that the following orthogonal direct sum decomposition holds:

$$V = \overline{W} \oplus W^{\perp}, \qquad \text{where} \quad \overline{W} \cap W^{\perp} = \{0\}.$$

Definition 2.15. Let $(V, \|\cdot\|)$ be a normed vector space. A sequence $(v_n)_{n\in\mathbb{N}}$ in V is said to be a *Cauchy-sequence* in $(V, \|\cdot\|)$ if

$$\lim_{m\to\infty} \sup_{n,n'\geq m} \|v_n - v_{n'}\| = 0. \tag{2.2}$$

Definition 2.16. A normed vector space $(V, \|\cdot\|)$ is said to be *complete* if any Cauchy-sequence in $(V, \|\cdot\|)$ converges in $(V, \|\cdot\|)$.

2.4.2 $L^2(\mathbb{P})$ as a Hilbert space

Note that $L^2(\mathbb{P})$ has the following structures:

- ▷ It is a real vector space.
- ▷ It has an inner product: $\langle \xi, \eta \rangle_{L^2} := \mathbb{E}[\xi\eta]$ for all $\xi, \eta \in L^2(\mathbb{P})$.
- ▷ It has a norm: $\|\xi\|_{L^2} := \sqrt{\mathbb{E}[\xi^2]} = \sqrt{\langle \xi, \xi \rangle_{L^2}}$ for all $\xi \in L^2(\mathbb{P})$.
- ▷ It has a metric (distance function): $\text{dist}_{L^2}(\xi, \eta) := \|\xi - \eta\|_{L^2}$.

Lemma 2.17. (Cauchy-Schwarz Inequality) *For any* $\xi, \eta \in L^2(\mathbb{P})$, *we have*

$$\|\langle \xi, \eta \rangle_{L^2}\| \leq \|\xi\|_{L^2} \|\eta\|_{L^2}$$

Exercise 2.18. Prove Lemma 2.17.

Definition 2.19. Sequence $(\xi_n)_{n\in\mathbb{N}}$ in $L^2(\mathbb{P})$ is said to *converge* in L^2 to an element[a] $\xi \in L^2(\mathbb{P})$ if

$$\lim_{n\to\infty} \|\xi_n - \xi\|_{L^2} = 0.$$

In this case, we write $\xi_n \xrightarrow{L^2} \xi$.

[a] Note that the limit $\xi \in L^2(\mathbb{P})$ is only unique up to events of probability zero ($\mathbb{P}$-almost surely).

Theorem 2.20. *The space* $(L^2(\mathbb{P}), \|\cdot\|_{L^2})$ *is* complete *(cf. Definition 2.16).*

Proof. Suppose that $(\xi_n)_{n\in\mathbb{N}}$ is a Cauchy-sequence in $(L^2(\mathbb{P}), \|\cdot\|_{L^2})$, that is,

$$\lim_{m\to\infty} \sup_{n,n'\geq m} \|\xi_n - \xi_{n'}\|_{L^2} = 0. \tag{2.3}$$

Step 1. We will first argue that $(\xi_n)_{n\in\mathbb{N}}$ has a subsequence that converges in $(L^2(\mathbb{P}), \|\cdot\|_{L^2})$.

Property (2.3) implies that we can find indices $1 \leq m_1 < m_2 < \cdots$ such that

$$\|\xi_n - \xi_{n'}\|_{L^2} \leq 4^{-j} \qquad \text{for all } n, n' \geq m_j,\ j \in \mathbb{N}.$$

Using Cauchy-Schwarz Inequality (Lemma 2.17), we obtain

$$\mathbb{E}[|\xi_{m_{j+1}} - \xi_{m_j}|] \leq \|\xi_{m_{j+1}} - \xi_{m_j}\|_{L^2} \leq 4^{-j} \qquad \text{for all } n, n' \geq m_j,\ j \in \mathbb{N}.$$

Summing over $j \in \mathbb{N}$, this shows by Fubini's theorem A.19 that

$$\sum_{j=1}^{\infty} \mathbb{E}[|\xi_{m_{j+1}} - \xi_{m_j}|] \leq \sum_{j=1}^{\infty} 4^{-j} < \infty,$$

as a geometric series. Therefore, the next Lemma 2.21 applied to $\eta_j := |\xi_{m_{j+1}} - \xi_{m_j}|$ shows that

$$\sum_{j=1}^{\infty} |\xi_{m_{j+1}} - \xi_{m_j}| < \infty \qquad \text{almost surely.} \tag{2.4}$$

Lemma 2.21. *Consider a sequence* $(\eta_j)_{j\in\mathbb{N}}$ *of non-negative random variables. If*

$$\mathbb{E}\Big[\sum_{j=1}^{\infty} \eta_j\Big] = \sum_{j=1}^{\infty} \mathbb{E}[\eta_j] < \infty,$$

then we have

$$\mathbb{P}\Big[\sum_{j=1}^{\infty} \eta_j < \infty\Big] = 1 \qquad \text{and} \qquad \mathbb{P}\Big[\lim_{j\to\infty} \eta_j = 0\Big] = 1.$$

Proof of Lemma 2.21. See [Kyt20, Lemma VIII.6 & Proposition VIII.7]. □

Next, from (2.4), considering a telescoping sum, we see that

$$\begin{aligned}
\sum_{j=1}^{\infty} (\xi_{m_{j+1}} - \xi_{m_j}) &= \lim_{k\to\infty} \sum_{j=1}^{k} (\xi_{m_{j+1}} - \xi_{m_j}) \\
&= \lim_{k\to\infty} \Big((\xi_{m_{k+1}} - \xi_{m_k}) + (\xi_{m_k} - \xi_{m_{k-1}}) + \cdots + (\xi_{m_2} - \xi_{m_1}) \Big) \\
&= \lim_{k\to\infty} (\xi_{m_{k+1}} - \xi_{m_1}),
\end{aligned}$$

which shows that the *subsequential limit*

$$\xi := \lim_{k\to\infty} \xi_{m_k} \qquad \text{exists almost surely.} \tag{2.5}$$

Step 2. To conclude the proof, we promote the (almost sure) subsequential convergence in Step 1 to convergence of the whole sequence $(\xi_n)_{n\in\mathbb{N}}$ in $(L^2(\mathbb{P}), \|\cdot\|_{L^2})$. Indeed, we have

$$\begin{aligned}
\|\xi_n - \xi\|_{L^2}^2 = \mathbb{E}[(\xi_n - \xi)^2] &= \mathbb{E}\Big[\lim_{k\to\infty}(\xi_n - \xi_{m_k})^2\Big] && \text{[by (2.5)]} \\
&\le \liminf_{k\to\infty} \mathbb{E}\big[(\xi_n - \xi_{m_k})^2\big] && \text{[by Fatou's lemma A.17]} \\
&\le \liminf_{k\to\infty} 4^{-2j} = 4^{-2j}, \qquad \text{when } n \ge m_j.
\end{aligned}$$

Finally, taking $n \to \infty$, we find that $\xi_n \xrightarrow{L^2} \xi$, which is what we sought to prove. □

Definition 2.22. For two vectors $\xi, \eta \in L^2(\mathbb{P})$, we write

$$\xi \perp \eta \qquad \Longleftrightarrow \qquad \langle \xi, \eta \rangle_{L^2} = 0.$$

In this case, we say that the two vectors are *orthogonal.*
If $V \subset L^2(\mathbb{P})$ is a vector subspace, the vector $\xi \in L^2(\mathbb{P})$ is said to be *orthogonal to* V if

$$\xi \perp \eta \qquad \text{for all } \eta \in V,$$

and in this case, we write $\xi \perp V$.

The Hilbert space structure of $L^2(\mathbb{P})$ allows us to define orthogonal projections very analogously to the case of Euclidean spaces. In Euclidean spaces one usually consider projections onto lines, planes, or other closed subspaces. Similarly, in the case of a Hilbert space, an orthogonal projection can be defined[8] onto any closed subspace (see Proposition 2.24).

Definition 2.23. (Closed subspace) A vector subspace $V \subset L^2(\mathbb{P})$ is said to be *closed* if it is a closed set in the topology induced by the norm $\|\cdot\|_{L^2}$. This is equivalent to the property that for all sequences $(\eta_n)_{n\in\mathbb{N}}$ in V that converge, the limit belongs to V:

$$\eta_n \xrightarrow{L^2} \eta \qquad \Longrightarrow \qquad \eta \in V.$$

Proposition 2.24. (Orthogonal projection) *Let $V \subset L^2(\mathbb{P})$ be a closed subspace. Consider a vector $\xi \in L^2(\mathbb{P})$. Then, the following hold (draw a picture!).*

▷ *Consider a vector $\eta \in V$. The following are equivalent:*

1. *(orthogonality): $(\xi - \eta) \perp V$;*
2. *(minimal distance to V): $\|\xi - \eta\|_{L^2} = \inf_{\chi\in V} \|\xi - \chi\|_{L^2}$.*

▷ *If either of these equivalent properties holds, then η exists and is almost surely unique.*

The vector $\eta \in V$ satisfying these properties is called the orthogonal projection of ξ onto V.

Proof. See [Kyt20, Proposition E.8 in Appendix E]. □

[8] Although not evident in these notes, the completeness of the space $L^2(\mathbb{P})$ is needed for the existence of the orthogonal projection in Proposition 2.24 — see [Kyt20, Appendix E] for a detailed exposition.

3 Martingales in discrete time

Throughout this section, we work with a fixed probability space $(\Omega, \mathscr{F}, \mathbb{P})$. We consider discrete-time stochastic processes $X = (X_n)_{n\in\mathbb{N}_0}$ taking real values: $X_n : \Omega \to \mathbb{R}$ for all n. We are interested in the mathematical notion of *information accumulating over time* (termed "filtration") and processes that have particularly useful properties regarding how their conditional expected value changes over time as information accumulates (termed "martingales"). The main result of this section is *Doob's Optional Stopping Theorem* 3.20, which is an extremely useful tool in martingale theory (and in stochastic analysis based on the theory of Brownian motion).

3.1 Filtrations

We begin with mathematically formalizing the notion of "information."

Definition 3.1. (Filtration) A collection $\mathcal{F}_\bullet = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ comprising an increasing family $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathscr{F}$ of sub-sigma-algebras of $\mathscr{F}$ is called a *filtration*. The tuple $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is called a *filtered probability space*.

Often, when considering a given stochastic process X, the filtration $\mathcal{F}_\bullet$ is implicitly taken to be the natural filtration *generated* by X, defined as

$$\mathcal{F}_n^X := \sigma(X_0, X_1, \dots, X_n).$$

The idea is that the information accumulated over time is gathered in the elements $\mathcal{F}_n^X$ of the filtration: $\mathcal{F}_n^X$ contains the information available at time n.

Definition 3.2. Stochastic process $X = (X_n)_{n\in\mathbb{N}_0}$ is *adapted to a filtration* $\mathcal{F}_\bullet$ if

$$X_n \in \mathrm{m}\mathcal{F}_n := \{\chi : \Omega \to \mathbb{R} \mid \chi \text{ is } \mathcal{F}_n\text{-measurable}\} \quad \text{for all } n \in \mathbb{N}_0.$$

For example, X is always adapted to its natural filtration $\mathcal{F}_\bullet^X$.

3.2 Martingales

As a warm-up example, let us consider the simple symmetric[9] random walk $S = (S_n)_{n\in\mathbb{N}_0}$,

$$S_0 = 0 \qquad \text{and} \qquad S_n := \sum_{j=1}^n \xi_j \qquad \text{for } n = 1, 2, \dots, \tag{3.1}$$

where the steps $\xi_1, \xi_2, \dots$ are i.i.d. coin tosses: $\mathbb{P}[\xi_1 = +1] = \frac{1}{2} = \mathbb{P}[\xi_1 = -1]$. The natural filtration of S is just generated by its steps:

▷ $\mathcal{F}_0^S := \{\varnothing, \Omega\}$;

▷ $\mathcal{F}_n^S := \sigma(S_0, S_1, \dots, S_n) = \sigma(\xi_1, \dots, \xi_n)$, for $n \geq 1$.

The process S has the following further important properties:

▷ $S_n \in \mathrm{m}\mathcal{F}_n^S$ for all $n \in \mathbb{N}_0$ (adapted);

[9]The random walk is said to be *symmetric* if its steps are centered: $\mathbb{E}[\xi_j] = 0$ for all j.

- ▷ $|S_n| \le n$ for all $n \in \mathbb{N}_0$ (integrable);
- ▷ its increments are centered:

$$\begin{aligned}\mathbb{E}[S_{n+1} - S_n \,|\, \mathcal{F}_n^S] &= \mathbb{E}[\xi_{n+1} \,|\, \mathcal{F}_n^S] \\ &= \mathbb{E}[\xi_{n+1}] && \text{[by item 6 of Lemma 2.8, as } \xi_{n+1} \text{ indep. of } \mathcal{F}_n^S] \\ &= 0.\end{aligned}$$

The last property also implies the following *martingale property*:

$$\begin{aligned} 0 &= \mathbb{E}[S_{n+1} - S_n \,|\, \mathcal{F}_n^S] \\ &= \mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n^S] - \mathbb{E}[S_n \,|\, \mathcal{F}_n^S] && \text{[by linearity]} \\ &= \mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n^S] - S_n && \text{[by item 5 of Lemma 2.8, as } S_n \in \mathrm{m}\mathcal{F}_n^S] \\ \Longrightarrow \qquad \mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n^S] &= S_n. \end{aligned}$$

In words, this means that the "best prediction" for the value of S_{n+1} given the information $\mathcal{F}_n^S$ up to time n is the current value S_n (that is measurable with respect to $\mathcal{F}_n^S$).

Definition 3.3. (Martingale) Stochastic process $M = (M_n)_{n \in \mathbb{N}_0}$ on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is called a *martingale* if

MG1. (it is adapted to $\mathcal{F}_\bullet$): for every $n \in \mathbb{N}_0$, the random variable M_n is $\mathcal{F}_n$-measurable;

MG2. (it is integrable): for every $n \in \mathbb{N}_0$, the random variable $M_n \in L^1(\mathbb{P})$ is integrable;

MG3. (it has the martingale property): almost surely, we have

$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] = M_n \qquad \text{for every } n \in \mathbb{N}_0.$$

If instead $\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] \le M_n$ or $\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] \ge M_n$ in property MG3, M is said to be a *supermartingale* or a *submartingale*, respectively.

Using item 1 of Lemma 2.8 and linearity, we see that[10] M is a martingale if and only if M satisfies MG1 & MG2 and its increments satisfy

$$\mathbb{E}[M_{n+1} - M_n \,|\, \mathcal{F}_n] \; = \; 0. \tag{3.2}$$

- ▷ Note that the question whether a process is a martingale depends on the choice of the filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ (which is, often, implicit).
- ▷ M is a martingale if and only if M is both a supermartingale and a submartingale.
- ▷ M is a supermartingale if and only if $-M$ is a submartingale. Thus, a statement about one automatically implies a converse statement for the other.

Example 3.4. Simple random walk of the form (3.1) with i.i.d. steps $\xi_1, \xi_2, \ldots \in L^1(\mathbb{P})$

- ▷ a martingale if $\mathbb{E}[\xi_1] = 0$;
- ▷ a supermartingale if $\mathbb{E}[\xi_1] \le 0$;
- ▷ a submartingale if $\mathbb{E}[\xi_1] \ge 0$.

[10] The analogous equivalence holds for super- or submartingales with "≤" or "≥".

The (super)martingale property propagates to the future in the following sense.

Lemma 3.5. *Let $M = (M_n)_{n\in\mathbb{N}_0}$ be a supermartingale. Then, we have*

$$\mathbb{E}[M_{n+k} \mid \mathcal{F}_n] \le M_n \qquad \text{for all } n \in \mathbb{N}_0,\ k \in \mathbb{N}. \tag{3.3}$$

In particular, we have

$$\mathbb{E}[M_n] \le \mathbb{E}[M_0] \qquad \text{for all } n \in \mathbb{N}_0. \tag{3.4}$$

If M is a martingale, then equality holds in (3.3) *and* (3.4).
If M is a submartingale, then the inequalities in (3.3) *and* (3.4) *read "$\ge$" instead.*

Exercise 3.6. Prove Lemma 3.5. *Hint: You can use properties of conditional expected value from Lemma 2.8.*

In particular, (3.3) shows that a super/submartingale is *monotone* in expectation. Martingales were introduced to model a "fair" game — however, (super)martingales actually represent a quite unfavorable gambling scheme, see Proposition 3.11 (you cannot beat the house!).

Example 3.7. Let $\xi \in L^1(\mathbb{P})$ be a random variable on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$. The conditional expected values $\big(\mathbb{E}[\xi \mid \mathcal{F}_n]\big)_{n\in\mathbb{N}_0}$ define a martingale[11]:

- ▷ (MG1): for every $n \in \mathbb{N}_0$, we have $\mathbb{E}[\xi \mid \mathcal{F}_n] \in \mathrm{m}\mathcal{F}_n$ by Definition 2.1;
- ▷ (MG2): for every $n \in \mathbb{N}_0$, we have $\mathbb{E}[\xi \mid \mathcal{F}_n] \in L^1(\mathbb{P})$ by Definition 2.1 (item CE1);
- ▷ (MG3): for every $n \in \mathbb{N}_0$, we have $\mathbb{E}\big[\mathbb{E}[\xi \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n\big] = \mathbb{E}[\xi \mid \mathcal{F}_n]$ by the Tower Property (item 4 of Lemma 2.8).

Example 3.8. Let us consider a gambling scheme (doubling strategy) based on random walk, whose steps $\xi_1, \xi_2, \ldots$ are i.i.d. coin tosses: $\mathbb{P}[\xi_1 = +1] = \frac{1}{2} = \mathbb{P}[\xi_1 = -1]$. Consider the natural filtration $\mathcal{F}^S_n := \sigma(\xi_1, \ldots, \xi_n)$, for $n \ge 1$, and $\mathcal{F}^S_0 := \{\varnothing, \Omega\}$. Set $M_0 = 0$, $M_1 := \xi_1$, and for $n \ge 1$,

$$M_{n+1} := \begin{cases} M_n + 2^n\, \xi_{n+1}, & M_n \le 0, \\ M_n, & M_n > 0, \end{cases} \tag{3.5}$$

that is, on each round, we bet on a fair coin toss so that (see also the below table)

- ▷ if we lose, then we *double the bet*, and bet again,
- ▷ if we win, we stop (this happens when M_n becomes positive).

Then the process $M = (M_n)_{n\in\mathbb{N}_0}$ is a martingale with respect to $\mathcal{F}^S_\bullet$.

round (n)	1	2	3	4	5	6
coin toss (ξ_n)	lose	lose	lose	lose	lose	win!
cumulative profit ($(H \bullet X)_n$)	−1	−3	−7	−15	−31	+1

[11] This process is often called the "prediction martingale," or the "tautological martingale."

Is this a perfect winning strategy? Well, it can take a really long time before we win a round. Thus, we would need a huge budget in order to be able to play as long as needed.

Exercise 3.9. Prove that the process (3.5) is a martingale.

3.3 Predictable processes and the first example of a stochastic integral

Definition 3.10. (Predictability) Stochastic process $H = (H_n)_{n \in \mathbb{N}_0}$ on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is called *predictable* (or *previsible*) if for every integer[a] $n \in \mathbb{N}$, the random variable H_n is $\mathcal{F}_{n-1}$-measurable, i.e., $H_n \in \mathrm{m}\mathcal{F}_{n-1}$.

[a] Note that we don't care much about the initial value H_0, which is some random variable: $H_0 \in \mathrm{m}\mathscr{F}$.

The idea is that the value of H_n is known already *before* time n. This will be important in the definition of stochastic integrals (especially later when we consider continuous-time processes).

Let us consider a simple example of an investment strategy. Let

- ▷ X_n be the price of a given stock in the evening of day n; and
- ▷ H_n be the number of these stocks in our portfolio in the morning of day n.

Then, the change in the stock price during day n is $X_n - X_{n-1}$, and the profit in our portfolio during day n is $H_n(X_n - X_{n-1})$. The cumulative profit up to day n is

$$\sum_{k=1}^{n} H_k(X_k - X_{k-1}) =: (H \bullet X)_n. \tag{3.6}$$

Setting also $(H \bullet X)_0 = 0$, we obtain a stochastic process $H \bullet X = ((H \bullet X)_n)_{n \in \mathbb{N}_0}$. This is an example of a "discrete stochastic integral" of H with respect to X. The following result shows that a strategy given by a (super)martingale is not expected to be profitable.

Proposition 3.11. *Consider a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$. Let $M = (M_n)_{n \in \mathbb{N}_0}$ be a supermartingale and consider a process $H = (H_n)_{n \in \mathbb{N}}$. Suppose that*

- ▷ *H is non-negative: $H_n \geq 0$ for all $n \in \mathbb{N}$;*
- ▷ *H is predictable (Definition 3.10);*
- ▷ *H is bounded: for all $n \in \mathbb{N}$, there exists a constant $c_n \in (0, \infty)$ such that $H_n \leq c_n$.*

Then, the discrete integral process $H \bullet M$ is also a supermartingale.
Assuming instead that M is a (sub)martingale, $H \bullet M$ is instead a (sub)martingale.

Proposition 3.11 also shows together with Lemma 3.5 that under the above conditions, the discrete integral (cumulative profit) process $H \bullet M$ is non-increasing in expectation:

$$\mathbb{E}[(H \bullet M)_m \,|\, \mathcal{F}_n] \leq (H \bullet M)_n \qquad \text{for all } 0 \leq n \leq m.$$

Proof of Proposition 3.11. We check the defining properties of a supermartingale for $H \bullet M$:

▷ (MG1): Clearly the constant $(H \bullet M)_0 = 0 \in \mathrm{m}\mathcal{F}_0$, and for every $n \in \mathbb{N}$, since $H_k, M_k \in \mathrm{m}\mathcal{F}_k$ for all $1 \le k \le n$, we also have

$$(H \bullet M)_n := \sum_{k=1}^{n} H_k (M_k - M_{k-1}) \ \in\ \mathrm{m}\mathcal{F}_n.$$

▷ (MG2): Clearly the constant $(H \bullet M)_0 = 0$ is integrable. For every $n \in \mathbb{N}$, since H_k are bounded by c_k for all $1 \le k \le n$, and M_k are integrable for all $0 \le k \le n$, we also have

$$\mathbb{E}\big[|(H \bullet M)_n|\big] \ \le\ \sum_{k=1}^{n} \mathbb{E}\,[H_k|M_k - M_{k-1}|] \ \le\ \sum_{k=1}^{n} c_k \big(\mathbb{E}\,[|M_k|] + \mathbb{E}\,[|M_{k-1}|]\big) \ <\ \infty.$$

▷ (MG3): For every $n \in \mathbb{N}_0$, we have

$$\begin{aligned}
&\mathbb{E}\,[(H \bullet M)_{n+1} - (H \bullet M)_n \,|\, \mathcal{F}_n] \\
&= \mathbb{E}\,[\underbrace{H_{n+1}}_{\in\, \mathrm{m}\mathcal{F}_n} (M_{n+1} - M_n) \,|\, \mathcal{F}_n] && \text{[telescoping sum]} \\
&= H_{n+1}\, \mathbb{E}\,[(M_{n+1} - M_n) \,|\, \mathcal{F}_n] && \text{[by item 5 of Lemma 2.8, as } H \text{ is predictable]} \\
&= \underbrace{H_{n+1}}_{\ge 0} \big(\underbrace{\mathbb{E}\,[M_{n+1} \,|\, \mathcal{F}_n]}_{\le M_n} - \underbrace{\mathbb{E}\,[M_n \,|\, \mathcal{F}_n]}_{= M_n \,\in\, \mathrm{m}\mathcal{F}_n}\big) && \text{[by linearity]} \\
&\le 0. && \text{[since } H \text{ is non-negative, } M \text{ is a supermartingale,} \\
& && \text{and using item 1 of Lemma 2.8, as } M \text{ is adapted]}
\end{aligned}$$

From (3.2) (with "$\le$"), we conclude that $H \bullet M$ is a supermartingale.

The case of a (sub)martingale follows by a similar computation. □

Exercise 3.12. Show that Proposition 3.11 also holds also under the following *alternative* assumptions:

▷ M is a (super/sub)martingale;

▷ H is non-negative and predictable;

▷ $M_n, H_m \in L^2(\mathbb{P})$ for all $n \in \mathbb{N}_0$ and $m \in \mathbb{N}$.

That is, check that under these assumptions, $H \bullet M$ is a (super/sub)martingale (recall Cauchy-Schwarz Inequality).

3.4 Stopping times and Optional Stopping Theorem

Recall from Lemma 3.5 that the best prediction for the value of a martingale M at any future time ($m \ge n$) given the information up to the present time (n) is its value at the present time:

$$\mathbb{E}\,[M_m \,|\, \mathcal{F}_n] = M_n, \qquad 0 \le n \le m. \tag{3.7}$$

Often in applications one is interested in the conditional expected value of M at a *random time.*

As an example, one might be interested in the value of the discrete integral process $H \bullet M$, representing cumulative profit, when the stock price M reaches some target value. Roughly speaking, determining whether this has happened should be possible based on the current available information. This motivates the definition of a "stopping time."

Definition 3.13. (Stopping time) Random variable $\tau : \Omega \to \mathbb{N}_0 \cup \{+\infty\}$ on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is called a *stopping time* (or *optional time*) if for every $n \in \mathbb{N}_0$, the event $\{\tau \le n\}$ is $\mathcal{F}_n$-measurable:

$$\{\tau \le n\} := \{\omega \in \Omega \mid \tau(\omega) \le n\} \in \mathcal{F}_n.$$

Exercise 3.14. Consider a random variable $\tau : \Omega \to \mathbb{N}_0 \cup \{+\infty\}$ on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$. Show that the following are equivalent.

1. τ is a stopping time with respect to $\mathcal{F}_\bullet$, that is, $\{\tau \le n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}_0$.
2. We have $\{\tau = n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}_0$.
3. The indicator process $H_n = \mathbb{1}\{n \le \tau\}$ is predictable, that is, $H_n \in \mathrm{m}\mathcal{F}_{n-1}$ for all $n \in \mathbb{N}$.

Exercise 3.15. Let σ, τ be stopping times with respect to (w.r.t.) a filtration $\mathcal{F}_\bullet$.

1. Show that also $\sigma \wedge \tau := \min\{\sigma, \tau\}$ and $\sigma \vee \tau := \max\{\sigma, \tau\}$ are stopping times w.r.t. $\mathcal{F}_\bullet$.
2. Is $\sigma + \tau$ a stopping time w.r.t. $\mathcal{F}_\bullet$?

Definition 3.16. (Stopped process) Let τ be an a.s. finite stopping time ($\mathbb{P}[\tau < \infty] = 1$), and let X be a process adapted to the filtration $\mathcal{F}_\bullet$. Then, we set

$$X_\tau := \sum_{k=0}^{\infty} \mathbb{1}\{\tau = k\} X_k.$$

We thus define the *stopped process*

$$(X_{n\wedge\tau})_{n\in\mathbb{N}_0}, \qquad \text{where } n \wedge \tau := \min\{n, \tau\}.$$

Theorem 3.17. *Let τ be a stopping time, and let M be a (super)martingale. Then, the stopped process $(M_{n\wedge\tau})_{n\in\mathbb{N}_0}$ is also a (super)martingale and*

$$\mathbb{E}[M_{n\wedge\tau}] \le \mathbb{E}[M_0]. \tag{3.8}$$

If M is a martingale, then equality holds in (3.8).

Remark 3.18. In Theorem 3.17, we only require that M is integrable (via property MG2 in Definition 3.3) — we impose no requirements for the stopping time τ.

In the statement of Theorem 3.17, one might wonder whether we could omit the cutoff n in $n \wedge \tau$ and deduce that $\mathbb{E}[M_\tau] \le \mathbb{E}[M_0]$. In order to define the value M_τ (cf. Definition 3.16). we would at least want to require the stopping time τ to be almost surely finite. However, this is not a sufficient condition, as the next example shows.

Example 3.19. Consider the simple symmetric random walk S (as in Example 3.4) started at $S_0 = 0$. It is a martingale, and one can check (see Exercise 3.28) that the hitting time at level 1,

$$\tau := \inf\{n \in \mathbb{N} \mid S_n = 1\},$$

is a stopping time such that $\mathbb{P}[\tau < \infty] = 1$. However, we have

$$\mathbb{E}[S_{n\wedge\tau}] = \mathbb{E}[S_0] = 0 \qquad \text{for all } n \in \mathbb{N}_0, \qquad \text{but} \qquad \mathbb{E}[S_\tau] = 1 \neq 0 = \mathbb{E}[S_0].$$

Hence, we have $\mathbb{E}[S_\tau] > \mathbb{E}[S_0]$ even though τ is almost surely finite.

From Example 3.19, we see that additional assumptions are needed in order to retain the property $\mathbb{E}[M_\tau] \leq \mathbb{E}[M_0]$ even for almost surely finite stopping times τ. We shall list a few of those in Theorem 3.20 below. Before this, let us first prove Theorem 3.17.

Proof of Theorem 3.17. Note that for any $n \in \mathbb{N}_0$, the random variable $n \wedge \tau$ is a *finite* stopping time. Moreover, by Exercise 3.14, the process $H = (H_n)_{n\in\mathbb{N}_0}$ defined as $H_n := \mathbb{1}\{n \leq \tau\}$ is non-negative, bounded, and predictable. Therefore, Proposition 3.11 shows that the discrete integral process $H \bullet M$ is a (super)martingale. Using its definition (3.6), we obtain

$$\begin{aligned}(H \bullet M)_n \; &:= \; \sum_{k=1}^{n} H_k(M_k - M_{k-1}) \; = \; \sum_{k=1}^{n} \mathbb{1}\{n \leq \tau\}(M_k - M_{k-1}) \; = \; \sum_{k=1}^{n\wedge\tau}(M_k - M_{k-1}) \\ &= \; (M_{n\wedge\tau} - M_{n\wedge\tau-1}) + \cdots + (M_1 - M_0) \\ &= \; M_{n\wedge\tau} - M_0\end{aligned}$$

by cancellations in the telescoping sum. Now, since both $H \bullet M$ and the constant process M_0 are (super)martingales, so is the stopped process $(M_{n\wedge\tau})_{n\in\mathbb{N}_0} = \big((H \bullet M)_n + M_0\big)_{n\in\mathbb{N}_0}$.

Asserted formula (3.8) then follows from (3.4) in Lemma 3.5. □

Theorem 3.20. ((Doob's) Optional Stopping Theorem) *Let M be a martingale. For a stopping time τ, under any of the following conditions* OST1, OST2, *or* OST3, *we have*

$$M_\tau \in L^1(\mathbb{P}) \qquad \text{and} \qquad \mathbb{E}[M_\tau] = \mathbb{E}[M_0]. \tag{3.9}$$

OST1. (τ is a.s. bounded): *There exists a constant $C \in (0, \infty)$ such that $\mathbb{P}[\tau \leq C] = 1$.*

OST2. (τ is a.s. finite and $|M|$ is a.s. bounded): *There exists a constant $C \in (0, \infty)$ such that $\mathbb{P}[|M_n| \leq C] = 1$ for all $n \in \mathbb{N}_0$, and $\mathbb{P}[\tau < \infty] = 1$.*

OST3. (τ is integrable and M has a.s. bounded increments): *There exists a constant $C \in (0, \infty)$ such that $\mathbb{P}[|M_n - M_{n-1}| \leq C] = 1$ for all*[a] *$n \in \mathbb{N}_0$, and $\mathbb{E}[\tau] < \infty$.*

[a] Here, we use the convention that $M_0 - M_{-1} = M_0$.

Proof. Note that $\mathbb{E}[M_{n\wedge\tau}] = \mathbb{E}[M_0]$ for any $n \in \mathbb{N}$ by Theorem 3.17.

▷ Assuming OST1, using Theorem 3.17 with $n = \lfloor C \rfloor$ being the integer part of C gives

$$\mathbb{E}[M_\tau] \; = \; \mathbb{E}\big[M_{\tau\wedge\lfloor C\rfloor}\big] \; = \; \mathbb{E}[M_0].$$

From the proof of Theorem 3.17 (with $H_n := \mathbb{1}\{n \leq \tau\}$), we also find that

$$|M_\tau| \; = \; |M_{\tau\wedge\lfloor C\rfloor}| \; = \; |(H \bullet M)_{\lfloor C\rfloor} + M_0| \; \leq \; |(H \bullet M)_{\lfloor C\rfloor}| + |M_0|$$

which shows that $M_\tau \in L^1(\mathbb{P})$, because $(H \bullet M)_{\lfloor C\rfloor}, M_0 \in L^1(\mathbb{P})$ by property MG2.

▷ Assuming OST2, since τ is a.s. finite, we have $n \wedge \tau \to \tau$ almost surely as $n \to \infty$. This shows that $M_{n\wedge\tau} \to M_\tau$ almost surely as $n \to \infty$. Since we also assumed that $|M|$ is a.s. bounded, using the Bounded Convergence Theorem (BCT) [Kyt20, Corollary VII.21], we obtain

$$\mathbb{E}[M_\tau] = \mathbb{E}\Big[\lim_{n\to\infty} M_{n\wedge\tau}\Big] \overset{\text{(BCT)}}{=} \lim_{n\to\infty} \underbrace{\mathbb{E}[M_{n\wedge\tau}]}_{=\mathbb{E}[M_0] \text{ for all } n} = \mathbb{E}[M_0].$$

Moreover, we have $M_\tau \in L^1(\mathbb{P})$, because $|M_\tau| \le C$ almost surely.

▷ Assuming OST3, note that the assumption $\mathbb{E}[\tau] < \infty$ implies that $\mathbb{P}(\tau < \infty) = 1$. Hence, similarly as above, we have $M_{n\wedge\tau} \to M_\tau$ almost surely as $n \to \infty$. Moreover, we have

$$\begin{aligned} |M_{n\wedge\tau}| &= |M_0 + \sum_{k=1}^{n\wedge\tau} (M_k - M_{k-1})| \\ &\le |M_0| + C|n \wedge \tau| \le |M_0| + C\tau, \qquad n \in \mathbb{N}. \end{aligned}$$

Hence, similarly as above, we can use the Dominated Convergence Theorem (DCT) [Kyt20, Theorem VII.19] to obtain

$$\mathbb{E}[M_\tau] = \mathbb{E}\Big[\lim_{n\to\infty} M_{n\wedge\tau}\Big] \overset{\text{(DCT)}}{=} \lim_{n\to\infty} \underbrace{\mathbb{E}[M_{n\wedge\tau}]}_{=\mathbb{E}[M_0] \text{ for all } n} = \mathbb{E}[M_0],$$

and similarly,

$$\mathbb{E}[|M_\tau|] = \mathbb{E}\Big[\lim_{n\to\infty} |M_{n\wedge\tau}|\Big] \overset{\text{(DCT)}}{=} \lim_{n\to\infty} \mathbb{E}[|M_{n\wedge\tau}|] \le \mathbb{E}[|M_0|] + C\,\mathbb{E}[\tau].$$

As $M_0 \in L^1(\mathbb{P})$ by property MG2 and $\tau \in L^1(\mathbb{P})$ by assumption, we see that $M_\tau \in L^1(\mathbb{P})$. □

We leave it as an exercise to extend Theorem 3.20 to super- and submartingales.

One can also compare values of martingales at different stopping times.

Proposition 3.21. *Let M be a (super)martingale. Let σ, τ be two stopping times such that $\mathbb{P}[\sigma \le \tau < \infty] = 1$ and $\mathbb{P}[\tau \le C] = 1$ for some constant $C \in (0, \infty)$. Then, we have*

$$\mathbb{E}[M_\tau] \le \mathbb{E}[M_\sigma]. \tag{3.10}$$

If M is a martingale, then equality holds in (3.10).

Proof. This result is an extension of part OST1 of Theorem 3.20, and the proof proceeds similarly as for Theorem 3.17. We have

$$M_\tau - M_\sigma = \sum_{n=\sigma+1}^{\tau} (M_n - M_{n-1}) = \sum_{n=1}^{\lfloor C \rfloor} H_n(M_n - M_{n-1}) = (H \bullet M)_{\lfloor C \rfloor},$$

where $H_n := \mathbb{1}\{\sigma < n \le \tau\}$. The process $H = (H_n)_{n\in\mathbb{N}_0}$ is clearly non-negative and bounded. It is predictable by Exercise 3.14, because σ, τ are stopping times, so

$$\{n \le \tau\} = \{\tau < n\}^c = \{\tau \le n-1\}^c \in \mathcal{F}_{n-1} \qquad \text{and} \qquad \{\sigma < n\} = \{\sigma \le n-1\}^c \in \mathcal{F}_{n-1}.$$

Thus, Proposition 3.11 shows that the discrete integral process $H \bullet M$ is a (super)martingale, so

$$0 = \mathbb{E}[(H \bullet M)_0] \ge \mathbb{E}[(H \bullet M)_{\lfloor C \rfloor}] = \mathbb{E}[M_\tau] - \mathbb{E}[M_\sigma],$$

which proves (3.10). □

As a converse of item OST1 of Theorem 3.20, martingales can be in fact *characterized* in terms of their *expected values at all bounded stopping times* — see Proposition 3.22 and Lemma 3.26.

Proposition 3.22. *Let M be an integrable, adapted stochastic process. Prove that M is a martingale if and only if for all bounded stopping times τ, we have $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$.*

Exercise 3.23. Prove Proposition 3.22.
Hint: *The direction "⇒" follows from item* OST1 *of Theorem 3.20. To show the converse direction "⇐," prove first the following observation: for any integers $n, m \in \mathbb{N}_0$ such that $n < m$, and an event $A \in \mathcal{F}_n$, the process $\tau := n\,\mathbb{1}_A + m\,\mathbb{1}_{\Omega\setminus A}$ is a stopping time.*

For a stopping time τ, the sigma-algebra

$$\mathcal{F}_\tau := \{A \in \mathcal{F}_\infty \mid A \cap \{\tau \le n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}_0\} \subset \mathscr{F}, \tag{3.11}$$

where $\mathcal{F}_\infty = \sigma\big(\bigcup_{n\in\mathbb{N}_0} \mathcal{F}_n\big)$, represents information available up to the random time τ.

Exercise 3.24. Consider a stopping time τ with respect to a filtration $\mathcal{F}_\bullet$.

1. Check that $\mathcal{F}_\tau$ defined in (3.11) is a sigma-algebra.
2. Show that τ is $\mathcal{F}_\tau$-measurable.
3. Suppose that τ is almost surely finite. Show that X_τ is $\mathcal{F}_\tau$-measurable (cf. Definition 3.16).

Exercise 3.25. Consider two stopping times σ, τ with respect to a filtration $\mathcal{F}_\bullet$.

1. Show that if $\sigma \le \tau$, then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.
2. Show that $\mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$.
3. Show that $\{\sigma \le \tau\} \in \mathcal{F}_{\sigma\wedge\tau}$ and $\{\sigma = \tau\} \in \mathcal{F}_{\sigma\wedge\tau}$.

Lemma 3.26. *Let M be an integrable, adapted stochastic process. Prove that M is a martingale if and only if for any two bounded stopping times σ and τ, we have*

$$\mathbb{E}[M_\tau \,|\, \mathcal{F}_\sigma] = M_{\tau\wedge\sigma} \qquad \textit{almost surely.}$$

Exercise 3.27. Prove Lemma 3.26. *Hint*: *Use a similar strategy as for Proposition 3.22.*

Using Doob's Optional Stopping Theorem 3.20, one can compute hitting probabilities for random walks, as the next exercise shows. The formula is often referred to as "Gambler's ruin."

Exercise 3.28. (Gambler's ruin) Fix $a, b, x \in \mathbb{Z}$ with $a < b$ and $a \leq x \leq b$. Let $\xi_1, \xi_2, \dots$ be i.i.d. random variables with

$$\mathbb{P}[\xi_1 = +1] = p, \qquad \mathbb{P}[\xi_1 = -1] = 1 - p.$$

Define a process $X = (X_n)_{n \in \mathbb{N}_0}$ by

$$X_0 = x \qquad \text{and} \qquad X_n := x + \sum_{k=1}^{n} \xi_k \qquad \text{for } n \in \mathbb{N},$$

and define the *first hitting time to a or b by X* (with the convention that $\inf \varnothing = +\infty$) as

$$\tau := \inf\{n \in \mathbb{N}_0 \mid X_n = a \text{ or } X_n = b\}.$$

1. Show that τ is a stopping time with respect to (w.r.t) the natural filtration $\mathcal{F}^X_\bullet$ generated by X.
2. Assume that $p = \frac{1}{2}$. Show that $\mathbb{P}[\tau < +\infty] = 1$. Define a function

$$f(z) = \frac{z - a}{b - a}, \quad z \in \mathbb{Z}.$$

   Show that the process $f(X) = (f(X_n))_{n \in \mathbb{N}_0}$ is a martingale w.r.t. $\mathcal{F}^X_\bullet$.
3. Assume that $p = \frac{1}{2}$ as above. Show that $\mathbb{P}[X_\tau = b] = f(x)$. *Hint*: *Optional Stopping Theorem 3.20.*
4. Assume that $p \neq \frac{1}{2}$. Show that $\mathbb{P}[\tau < +\infty] = 1$. Define a function

$$g(z) = \alpha^z, \quad z \in \mathbb{Z},\ \alpha \in \mathbb{R}.$$

   Find a value of $\alpha \in \mathbb{R} \setminus \{0, 1\}$ such that the process $g(X) = (g(X_n))_{n \in \mathbb{N}_0}$ is a martingale w.r.t. $\mathcal{F}^X_\bullet$ and show that

$$\mathbb{P}[X_\tau = b] = \frac{g(x) - g(a)}{g(b) - g(a)}.$$

   Hint: *Optional Stopping Theorem 3.20.*

3.5 Toolbox: Useful convex transformations ($\star$)

Let us record here useful consequences of conditional Jensen's inequality (item 1 of Lemma 2.10).

Lemma 3.29.

1. *Let $M = (M_n)_{n \in \mathbb{N}_0}$ be a submartingale.*
   Let $\phi : \mathbb{R} \to \mathbb{R}$ be a non-decreasing convex function such that $\mathbb{E}[|\phi(M_n)|] < \infty$ for all $n \in \mathbb{N}_0$. Then $\phi(M) = (\phi(M_n))_{n \in \mathbb{N}_0}$ is a submartingale.
2. *Let $M = (M_n)_{n \in \mathbb{N}_0}$ be a supermartingale.*
   Let $\phi : \mathbb{R} \to \mathbb{R}$ be a non-decreasing concave function such that $\mathbb{E}[|\phi(M_n)|] < \infty$ for all $n \in \mathbb{N}_0$. Then $\phi(M) = (\phi(M_n))_{n \in \mathbb{N}_0}$ is a supermartingale.

Proof. To prove item 1, we check the defining properties of a submartingale for $\phi(M)$:

- ▷ (MG1): for every $n \in \mathbb{N}_0$, we have $\phi(M_n) = \phi \circ M_n \in \mathrm{m}\mathcal{F}_n$ since ϕ is a Borel function;
- ▷ (MG2): for every $n \in \mathbb{N}_0$, we have $\phi(M_n) \in L^1(\mathbb{P})$ by assumption;

▷ (MG3): for every $n \in \mathbb{N}_0$, we have

$$\begin{aligned}\mathbb{E}[\phi(M_{n+1}) \mid \mathcal{F}_n] &\geq \phi\big(\mathbb{E}[M_{n+1} \mid \mathcal{F}_n]\big) && \text{[by conditional Jensen's inequality (2.1)]} \\ &\geq \phi(M_n). && \text{[since } \phi \text{ is non-decreasing]}\end{aligned}$$

This proves item 1. Item 2 then follows easily by considering the process $-M$. □

Remark 3.30. If M is a martingale, then item 1 of Lemma 3.29 can be formulated in the following form: if $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function such that $\mathbb{E}[|\phi(M_n)|] < \infty$ for all $n \in \mathbb{N}_0$, then $\phi(M) = (\phi(M_n))_{n \in \mathbb{N}_0}$ is a submartingale. It is evident from the proof that in this case, we do not need to assume any monotonicity for ϕ. However, $\phi(M)$ might still not be a martingale.

Corollary 3.31. *Let $M = (M_n)_{n \in \mathbb{N}_0}$ be a martingale. Then, for any $p \geq 1$, the process $(|M_n|^p)_{n \in \mathbb{N}_0}$ is a submartingale.*

Proof. Because $x \mapsto |x|^p$ is convex for any $p \geq 1$, the claim follows from item 1 of Lemma 3.29. □

Corollary 3.32. *Let $M = (M_n)_{n \in \mathbb{N}_0}$ be a submartingale. Fix a constant $a \in \mathbb{R}$ and define $\phi(x) := \max\{0, x - a\}$. Then, $\phi(M) = (\phi(M_n))_{n \in \mathbb{N}_0}$ is a submartingale.*

Proof. Because $x \mapsto \max\{0, x - a\}$ is non-decreasing and convex for any $a \in \mathbb{R}$, the claim follows from item 1 of Lemma 3.29. □

Corollary 3.33. *Let $M = (M_n)_{n \in \mathbb{N}_0}$ be a supermartingale. Fix a constant $a \in \mathbb{R}$ and define $\phi(x) := \min\{0, x - a\}$. Then, $\phi(M) = (\phi(M_n))_{n \in \mathbb{N}_0}$ is a supermartingale.*

Proof. Because $x \mapsto \min\{0, x - a\}$ is non-decreasing and concave for any $a \in \mathbb{R}$, the claim follows from item 2 of Lemma 3.29. □

4 Martingale convergence theorems

Consider a (super/sub)martingale M on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$. Our next goal is to investigate the *long-time behavior* of M_n as $n \to \infty$. Recall from Lemma 3.5 that M is *monotone* in expectation. Thus, in analogy with the Monotone Convergence Theorem, one would expect that (super/sub)martingales would have a (possibly infinite) limit $\lim_{n\to\infty} M_n =: M_\infty$.

- ▷ Under which conditions could we hope for the existence of a *finite* limit? It is reasonable to expect that a finite limit would exist if

  - ⋆ the absolute value $|M_n|$ does not grow too much as $n \to \infty$;
  - ⋆ the value of M_n does not oscillate too much as $n \to \infty$.

  It turns out that (super/sub)martingales never oscillate too wildly, see Lemma 4.4 (Doob's Upcrossing Lemma) — essentially because they are monotone in expectation.

- ▷ If we believe that there should be a limit, we should address in which sense the limit exists (almost surely, in L^1, in L^2, in probability...).

Doob's Martingale Convergence Theorems give sufficient conditions for the existence of the limit M_∞. The first (Theorem 4.1 in Section 4.1) concerns *almost sure convergence*, and the second (Theorem 4.13 in Section 4.3) *convergence in* L^1. For the latter, we need to address the growth of $|M_n|$ more carefully, and a key assumption is "uniform integrability," see Section 4.2.

4.1 Martingale Convergence Theorem

One possible way to ensure the existence of the limit M_∞ (almost surely) is to require that the martingale $M = (M_n)_{n\in\mathbb{N}_0}$ is *uniformly* L^1*-bounded*, that is,

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}[|M_n|] < \infty. \tag{4.1}$$

Note that (4.1) is stronger than M just being integrable (i.e., $\mathbb{E}[|M_n|] < \infty$ for all $n \in \mathbb{N}_0$, which holds for any martingale by property MG2) — see also Exercise 4.6.

Theorem 4.1. ((Doob's) Martingale Convergence Theorem (a.s.)) *Let M be a uniformly L^1-bounded martingale. Then, there exists a random variable $M_\infty \in L^1(\mathbb{P})$ such that*

- ▷ *it is measurable: $M_\infty \in \mathrm{m}\mathcal{F}_\infty$ where $\mathcal{F}_\infty := \sigma\big(\bigcup_{n\in\mathbb{N}_0} \mathcal{F}_n\big)$;*
- ▷ *we have $\lim_{n\to\infty} M_n = M_\infty$ almost surely.*

- ▷ Note that convergence almost surely does not automatically imply convergence in L^1.
- ▷ We know that the limit is finite, but we do not know whether the limit is trivial (constant). Indeed, the proof of Theorem 4.1 discussed in the rest of this section is not constructive, so it does not provide us with any means to find the limit M_∞ concretely.

Remark 4.2. A similar result also holds for super- and submartingales, with (4.1) replaced by

- ▷ $\sup_{n\in\mathbb{N}_0} \mathbb{E}[\max\{-M_n, 0\}] < \infty$ for a supermartingale;
- ▷ $\sup_{n\in\mathbb{N}_0} \mathbb{E}[\max\{M_n, 0\}] < \infty$ for a submartingale.

Indeed, if M is a martingale, then it is also a supermartingale, and $-M$ is a submartingale. Hence, the above condition for submartingales follows from that for supermartingales by considering $-M$, while the condition for supermartingales follows by the following bound:

$$\begin{aligned}\mathbb{E}[|M_n|] &= \mathbb{E}[M_n] + 2\,\mathbb{E}[\max\{-M_n, 0\}] \\ &\le \mathbb{E}[\underbrace{M_0}_{\in L^1(\mathbb{P})}] + 2\,\mathbb{E}[\max\{-M_n, 0\}]. \qquad \text{[by Lemma 3.5]}\end{aligned}$$

Thus, we see that M is uniformly L^1-bounded if and only if $\sup_{n \in \mathbb{N}_0} \mathbb{E}[\max\{-M_n, 0\}] < \infty$.

The above discussion shows that Theorem 4.1 is a consequence of the following result.

Proposition 4.3. *Let M be a submartingale on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$. Assume that*

$$\sup_{n \in \mathbb{N}_0} \mathbb{E}[\max\{M_n, 0\}] < \infty. \tag{4.2}$$

Then, there exists a random variable $M_\infty \in L^1(\mathbb{P})$ such that

- ▷ *it is measurable: $M_\infty \in \mathrm{m}\mathcal{F}_\infty$ where $\mathcal{F}_\infty := \sigma\big(\bigcup_{n \in \mathbb{N}_0} \mathcal{F}_n\big)$;*
- ▷ *we have $\lim_{n \to \infty} M_n = M_\infty$ almost surely.*

With a boundedness assumption (4.2) in place, the key tool to investigate the existence of the limit of M_n is to study its *oscillation*. This is controlled by *Doob's Upcrossing Lemma* 4.4. It also has an interpretation in terms of a model for investments in a stock market. The idea is that — in a potentially profitable investment strategy — one buys stocks with a low price (valleys of the graph of M) and sells them with a high price (peaks of the graph of M). Analytically, we consider time intervals where M crosses from a given value a (buying threshold) to a given higher value b (selling threshold). (The expected gain at each round is $b - a$.)

To formalize this, fix $a < b$. Define the stopping times $\tau_0 < \tau_1 < \cdots$ as

- ▷ $\tau_0 := -1$ (by convention);
- ▷ $\tau_{2k-1} := \inf\{n > \tau_{2k-2} \mid M_n \le a\}$ ("buying times"), $k = 1, 2, \ldots$;
- ▷ $\tau_{2k} := \inf\{n > \tau_{2k-1} \mid M_n \ge b\}$ ("selling times"), $k = 1, 2, \ldots$.

See also Figure 4.1 for an illustration. Then, between each pair of times τ_{2k-1} and τ_{2k}, termed an *upcrossing*, the "stock price process" M changes at least by the amount $b - a$. In this model, our portfolio will be defined as $H_0 := 0$ and for $n \ge 1$,

$$H_n := \begin{cases} 1 \text{ (keep)}, & \tau_{2k-1} < n \le \tau_{2k} \text{ for some } k, \\ 0 \text{ (sell and wait)}, & \text{else.} \end{cases}$$

Let us estimate the profit up to time n. At each upcrossing, the profit is at least $b - a$, and the number of upcrossings up to time n is

$$U_n = U_n^{[a,b]} := \max\{k \in \mathbb{N} \mid \tau_{2k} \le n\}.$$

Doob's Upcrossing Lemma says that on average, the profit obtained from the upcrossings is at least the number of upcrossings times the minimum profit $b - a$ per each upcrossing.

Write $X_n := a + \max\{M_n - a, 0\}$ (i.e., $X_n = a$ when M_n is below a, and $X_n = M_n$ otherwise).

Lemma 4.4. (Doob's Upcrossing Lemma) *Let M be a submartingale. Then, we have*

$$(b-a)\,\mathbb{E}[U_n^{[a,b]}] \le \mathbb{E}[\max\{M_n - a, 0\}] - \mathbb{E}[\max\{M_0 - a, 0\}] = \mathbb{E}[X_n - X_0].$$

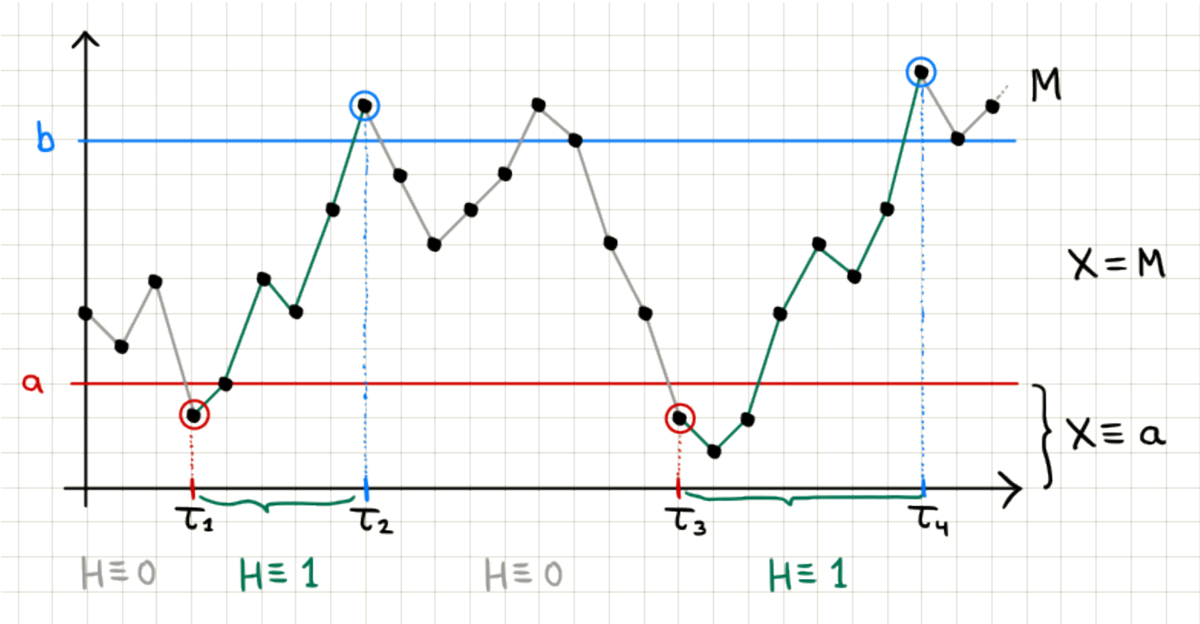


Figure 4.1. Illustration of the setup in Doob's Upcrossing Lemma 4.4.

Proof. By Corollary 3.32, X is a submartingale. Therefore, the discrete integral process (3.6),

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

is also a submartingale by Proposition 3.11, as H is non-negative, bounded, and predictable. Recall that $H_k = 1$ during upcrossings and $H_k = 0$ otherwise.

▷ On the one hand, we obtain

$$(H \bullet X)_n = \sum_{k=1}^{U_n} \sum_{j=\tau_{2k-1}+1}^{\tau_{2k}} (X_j - X_{j-1}) = \sum_{k=1}^{U_n} (\underbrace{X_{\tau_{2k}}}_{\ge b} - \underbrace{X_{\tau_{2k-1}}}_{= a}) \ge (b-a)\,U_n. \tag{4.3}$$

▷ On the other hand, we obtain

$$X_n - X_0 = \sum_{k=1}^{n} (X_k - X_{k-1}) = (H \bullet X)_n + ((1-H) \bullet X)_n. \tag{4.4}$$

Here, both $H \bullet X$ and $(1-H) \bullet X$ are submartingales by Proposition 3.11, as H and $1-H$ are non-negative, bounded, and predictable. Combining (4.3, 4.4) and taking expected value, we see that

$$\mathbb{E}[X_n - X_0] \ge (b-a)\,\mathbb{E}[U_n] + \underbrace{\mathbb{E}[((1-H) \bullet X)_n]}_{\ge \mathbb{E}[((1-H)\bullet X)_0] = 0} \ge (b-a)\,\mathbb{E}[U_n],$$

which is what we sought to prove. □

Proof of Proposition 4.3. Let us consider the limit of $M_n(\omega)$ for $\omega \in \Omega$ as $n \to \infty$.

Step 1. For fixed $\omega \in \Omega$, the sequence $(M_n(\omega))_{n\in\mathbb{N}_0}$ is a sequence of real numbers. If it fails to converge (to a finite or infinite value), then it upcrosses infinitely often some interval $[p, q]$ given by two rational numbers $p, q \in \mathbb{Q}$ such that $p < q$, i.e.,

$$\liminf_{n\to\infty} M_n(\omega) \;<\; p \;<\; q \;<\; \limsup_{n\to\infty} M_n(\omega).$$

In other words, we have

$$\begin{aligned}
&\mathbb{P}\Big[\text{the limit } \lim_{n\to\infty} M_n(\omega) \text{ does not exist in } [-\infty, +\infty]\Big]\\
&\le \mathbb{P}\Big[\text{there exists } p, q \in \mathbb{Q} \text{ s.t. } p < q \text{ and } \liminf_{n\to\infty} M_n(\omega) \;<\; p \;<\; q \;<\; \limsup_{n\to\infty} M_n(\omega)\Big].
\end{aligned}$$

Step 2. For each $a < b$, the number $U_n^{[a,b]} := \sup\{k \in \mathbb{N} \mid \tau_{2k} \le n\}$ of upcrossings across $[a, b]$ is a non-decreasing sequence, so it has a limit $\lim_{n\to\infty} U_n =: U_\infty$ almost surely. Moreover, we have

$$\begin{aligned}
\mathbb{E}[U_n] \;&\le\; \frac{\mathbb{E}[X_n - X_0]}{b-a} && \text{[by Lemma 4.4]}\\
&\le\; \frac{\mathbb{E}\big[\max\{M_n, 0\}\big] + |a|}{b-a} \;<\; \infty \qquad \text{for all } n \in \mathbb{N}. && \text{[by assumption (4.2)]}
\end{aligned}$$

Hence, we see that $U_\infty < \infty$ almost surely, since

$$\mathbb{E}[U_\infty] \;=\; \mathbb{E}\Big[\lim_{n\to\infty} U_n\Big] \;\le\; \liminf_{n\to\infty} \mathbb{E}[U_n] \;<\; \infty$$

using Fatou's lemma A.17. (Recall that $\mathbb{E}[U_\infty] < \infty$ implies that $\mathbb{P}[U_\infty < \infty] = 1$.)

Step 3. Applying Steps 1 & 2 for all pairs $p, q \in \mathbb{Q}$ of rationals such that $p < q$, we find that

$$\begin{aligned}
&\mathbb{P}\Big[\text{the limit } \lim_{n\to\infty} M_n(\omega) \text{ does not exist in } [-\infty, +\infty]\Big]\\
&\le \mathbb{P}\Big[\exists\, p, q \in \mathbb{Q} \text{ s.t. } p < q \text{ and } \liminf_{n\to\infty} M_n(\omega) \;<\; p \;<\; q \;<\; \limsup_{n\to\infty} M_n(\omega)\Big] && \text{[by Step 1]}\\
&\le \sum_{p,q\in\mathbb{Q}} \mathbb{P}\Big[\liminf_{n\to\infty} M_n(\omega) \;<\; p \;<\; q \;<\; \limsup_{n\to\infty} M_n(\omega)\Big] && \text{[by Union Bound (A.4)]}\\
&= 0. && \text{[by Step 2]}
\end{aligned}$$

This shows that the limit $\lim_{n\to\infty} M_n = M_\infty$ exists almost surely, with $M_\infty : \Omega \to [-\infty, \infty]$.

Step 4. Next, we show that $M_\infty \in L^1(\mathbb{P})$. Using the assumed bound (4.2), we obtain

$$\begin{aligned}
\mathbb{E}[\max\{M_\infty, 0\}] \le \liminf_{n\to\infty} \mathbb{E}[\max\{M_n, 0\}] & \qquad \text{[by Fatou's lemma A.17]}\\
< \infty, & \qquad \text{[by assumption (4.2)]}
\end{aligned}$$

and similarly, because $\max\{-M_n, 0\} = \max\{M_n, 0\} - M_n$, we obtain

$$\mathbb{E}[\max\{-M_\infty, 0\}] \;\le\; \liminf_{n\to\infty} \mathbb{E}[\max\{-M_n, 0\}] \;=\; \liminf_{n\to\infty} \Big(\mathbb{E}[\max\{M_n, 0\}] - \mathbb{E}[M_n]\Big) \;<\; \infty,$$

since $\mathbb{E}[M_n] \ge \mathbb{E}[M_0]$ (as M is a submartingale). Thus, we have $\mathbb{E}[|M_\infty|] < \infty$, as desired.

Step 5. Lastly, since the limit exists almost surely and is unique up to indistinguishability, we can define (pick a modification of) it to be

$$M_\infty(\omega) := \limsup_{n\to\infty} M_n(\omega), \qquad \omega \in \Omega.$$

Then, it is evident that the limit is measurable in the sense that $M_\infty \in \mathrm{m}\mathcal{F}_\infty$. □

Corollary 4.5. *Let M be a non-negative supermartingale. Then, there exists an integrable random variable $M_\infty \in L^1(\mathbb{P})$ such that $M_\infty \geq 0$ almost surely and*

$$M_n \xrightarrow{\text{a.s.}} M_\infty \qquad \text{as } n \to \infty.$$

Proof. Consider the submartingale $N = -M$. It is non-positive: $N_n \leq 0$ for all $n \in \mathbb{N}_0$. Thus, N satisfies $\sup_{n\in\mathbb{N}_0} \mathbb{E}[\max\{N_n, 0\}] = 0$, so the Proposition 4.3 implies the assertion. □

Exercise 4.6. (A wild martingale) Let $\xi_1, \xi_2, \ldots$ be independent random variables such that

$$\mathbb{P}\Big[\xi_k = \frac{2^k}{2^k - 1}\Big] = \frac{2^k - 1}{2^k} \qquad \text{and} \qquad \mathbb{P}\Big[\xi_k = -2^k\Big] = \frac{1}{2^k} \qquad \text{for each } k \in \mathbb{N}.$$

Define a process $M = (M_n)_{n\in\mathbb{N}_0}$ by

$$M_0 = 0 \qquad \text{and} \qquad M_n := \sum_{k=1}^{n} \xi_k \qquad \text{for } n \in \mathbb{N}.$$

1. Show that M is a martingale w.r.t. its natural filtration $\mathcal{F}^M_\bullet$.
2. Prove that, almost surely, we have $\xi_k > 1$ except for finitely many $k \in \mathbb{N}$.
   Hint: You can use the First Borel-Cantelli Lemma A.13.
   Conclude that $\mathbb{P}\big[\lim_{n\to\infty} M_n = +\infty\big] = 1$.
3. Why does this not contradict the Martingale Convergence Theorem 4.1?

4.2 Toolbox: Uniform integrability

To change the almost sure convergence in Doob's Martingale Convergence Theorem 4.1 to *convergence in L^1*, the key assumption is *uniform integrability* (UI). Under this assumption, one obtains a stronger convergence result (Theorem 4.13 in Section 4.3). This is also often helpful when trying to verify that the limit M_∞ is not trivial and to *construct* M from its limit M_∞.

Definition 4.7. (Uniform integrability) A collection $(\xi_i)_{i\in I}$ of real-valued random variables is said to be *Uniformly Integrable* (UI) if

$$\lim_{R\to\infty} \sup_{i\in I} \mathbb{E}\Big[|\xi_i|\, \mathbb{1}\{|\xi_i| \geq R\}\Big] = 0. \tag{4.5}$$

Stochastic process $X = (X_n)_{n\in\mathbb{N}_0}$ is UI if (4.5) holds with $I = \mathbb{N}_0$.

▷ The next exercise shows that an *UI collection is uniformly L^1-bounded.*

> **Exercise 4.8.** Show that if $(\xi_i)_{i\in I}$ is UI, then there exists a constant $C < \infty$ such that $\mathbb{E}[|\xi_i|] \le C$ for all $i \in I$.

▷ The converse does not necessarily hold: uniformly L^1-bounded does *not* imply UI.

Example 4.9. On the probability space $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}([0,1]), \mu_{\text{Leb}})$ given by the Lebesgue measure μ_{Leb}, consider the random variables

$$\xi_n := n\,\mathbb{1}\{(0,\tfrac{1}{n})\}, \qquad n \in \mathbb{N}.$$

Then, we have

$$\mathbb{E}[|\xi_n|] \;=\; \mathbb{E}\big[n\,\mathbb{1}\{(0,\tfrac{1}{n})\}\big] \;=\; n\tfrac{1}{n} \;=\; 1 \qquad \text{for all } n \in \mathbb{N},$$

so the collection $(\xi_n)_{n\in\mathbb{N}}$ is uniformly L^1-bounded. However, if $n \ge R$, then also

$$\mathbb{E}\big[|\xi_n|\,\mathbb{1}\{|\xi_n| \ge R\}\big] \;=\; \mathbb{E}\big[n\,\mathbb{1}\{(0,\tfrac{1}{n})\}\,\mathbb{1}\{n \ge R\}\big] \;=\; \mathbb{E}\big[n\,\mathbb{1}\{(0,\tfrac{1}{n})\}\big] \;=\; n\tfrac{1}{n} \;=\; 1,$$

which does not tend to zero as $R \to \infty$. Hence, the collection $(\xi_n)_{n\in\mathbb{N}}$ is not UI.

▷ However, $L^1(\mathbb{P})$-domination is sufficient to guarantee UI, as the next exercise shows.

> **Exercise 4.10.** Suppose that if there exists an integrable random variable $\eta \in L^1(\mathbb{P})$ such that $|\xi_i| \le \eta$ for all $i \in I$, then $(\xi_i)_{i\in I}$ is UI.

▷ Alternatively, uniform boundedness in $L^p(\mathbb{P})$ *for some* $p > 1$ does imply UI.

> **Exercise 4.11.** Suppose that the collection $(\xi_i)_{i\in I}$ is uniformly L^p-bounded for some $p > 1$, i.e., there exists a constant $C < \infty$ such that $\mathbb{E}[|\xi_i|^p] \le C$ for all $i \in I$. Show that $(\xi_i)_{i\in I}$ is UI.

Proposition 4.12. *Consider a sequence $\xi_1, \xi_2, \ldots \in L^1(\mathbb{P})$ of random variables. Also, let $\xi \in L^1(\mathbb{P})$. The following are equivalent:*

1. *The sequence $(\xi_n)_{n\in\mathbb{N}}$ converges in $L^1(\mathbb{P})$:*

$$\xi_n \xrightarrow{L^1} \xi, \qquad \text{that is,} \qquad \lim_{n\to\infty} \mathbb{E}[|\xi_n - \xi|] \;=\; 0.$$

2. *The sequence $(\xi_n)_{n\in\mathbb{N}}$ is UI and converges to ξ in probability.*

Proof. See, for instance, [Wil91, Theorem 13.7]. □

4.3 Martingale Reconstruction Theorem — convergence in L^1

Recall that the conditional expected value of a random variable $\xi \in L^1(\mathbb{P})$ defines a "tautological" martingale $M_n := \mathbb{E}[\xi \,|\, \mathcal{F}_n]$. When $n \to \infty$, if M satisfies the assumptions of Theorem 4.1, we know that it has an almost sure limit M_∞. In fact, one can then prove that

$$M_n \;:=\; \mathbb{E}[\xi \,|\, \mathcal{F}_n] \;=\; \mathbb{E}[M_\infty \,|\, \mathcal{F}_n].$$

Such an identity shows in particular that if the limit is trivial, i.e. $M_\infty \equiv 0$, then the whole martingale is almost surely constant, i.e. $\mathbb{P}[M_n = 0 \text{ for all } n \in \mathbb{N}_0] = 1$.

The following result could be termed as "Martingale Reconstruction Theorem," since it allows to construct the martingale M from its limit M_∞. (This is not standard terminology, though.) The key assumption is *uniform integrability* (UI), discussed in Section 4.2.

Theorem 4.13. ((Doob's) Martingale Convergence Theorem (L^1)) *Let M be a martingale on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$. Assume that M is Uniformly Integrable* (UI):

$$\lim_{R\to\infty} \sup_{n\in\mathbb{N}_0} \mathbb{E}\big[|M_n|\, \mathbb{1}\{|M_n| > R\}\big] = 0. \tag{4.6}$$

Then, there exists a random variable $M_\infty \in L^1(\mathbb{P})$ such that

- ▷ *it is measurable: $M_\infty \in \mathrm{m}\mathcal{F}_\infty$ where $\mathcal{F}_\infty := \sigma\big(\bigcup_{n\in\mathbb{N}_0} \mathcal{F}_n\big)$;*
- ▷ *we have $\lim_{n\to\infty} M_n = M_\infty$ almost surely;*
- ▷ *we have $M_n \xrightarrow{L^1} M_\infty$, that is,*

$$\lim_{n\to\infty} \mathbb{E}[|M_n - M_\infty|] \;=\; 0.$$

Moreover, in this case, almost surely

$$M_n = \mathbb{E}[M_\infty \,|\, \mathcal{F}_n], \qquad n \in \mathbb{N}_0.$$

Proof. By Exercise 4.8, the assumption (4.6) implies (4.1), so Theorem 4.1 readily shows that the limit $\lim_{n\to\infty} M_n = M_\infty$ exists almost surely, with $M_\infty \in L^1(\mathbb{P})$ satisfying $M_\infty \in \mathrm{m}\mathcal{F}_\infty$.

Recall that almost sure convergence implies convergence in probability (Exercise A.21). Moreover, *uniform integrability combined with convergence in probability* (cf. Proposition 4.12) *implies convergence in L^1*. Hence, we conclude that

$$M_n \xrightarrow{L^1} M_\infty, \qquad \text{that is,} \qquad \lim_{n\to\infty} \mathbb{E}[|M_n - M_\infty|] \;=\; 0.$$

It thus remains to be proven that $M_n = \mathbb{E}[M_\infty \,|\, \mathcal{F}_n]$ almost surely. To this end, by (3.7),

$$\mathbb{E}[M_m \,|\, \mathcal{F}_n] = M_n, \qquad 0 \le n \le m,$$

which shows that for any fixed $n \in \mathbb{N}_0$ and $G_n \in \mathcal{F}_n$, we have

$$\begin{aligned} \mathbb{E}[M_n \mathbb{1}_{G_n}] \;=\; \mathbb{E}[\mathbb{E}[M_m \,|\, \mathcal{F}_n] \mathbb{1}_{G_n}] \;&=\; \mathbb{E}[\mathbb{E}[M_m \mathbb{1}_{G_n} \,|\, \mathcal{F}_n]] \qquad &&\text{[by item 5 of Lemma 2.8]} \\ &=\; \mathbb{E}[M_m \mathbb{1}_{G_n}]. &&\text{[by item 4 of Lemma 2.8]} \end{aligned}$$

On the other hand, for any fixed $n \in \mathbb{N}_0$ and $G_n \in \mathcal{F}_n$, we also have

$$|\mathbb{E}[M_m \mathbb{1}_{G_n}] - \mathbb{E}[M_\infty \mathbb{1}_{G_n}]| = |\mathbb{E}[(M_m - M_\infty)\mathbb{1}_{G_n}]| \leq \mathbb{E}[|M_m - M_\infty|] \overset{m\to\infty}{\longrightarrow} 0,$$

since $M_m \overset{L^1}{\longrightarrow} M_\infty$. Thus, for any $0 \leq n \leq m$, we see that

$$\mathbb{E}[M_n \mathbb{1}_{G_n}] = \mathbb{E}[M_m \mathbb{1}_{G_n}] \overset{m\to\infty}{\longrightarrow} \mathbb{E}[M_\infty \mathbb{1}_{G_n}]$$

so

$$\mathbb{E}[M_n \mathbb{1}_{G_n}] = \mathbb{E}[M_\infty \mathbb{1}_{G_n}], \qquad n \in \mathbb{N}_0.$$

This shows that the random variable M_n satisfies property CE3 in Definition 2.1. Because M_n also satisfies property CE1 (thanks to property MG2) and property CE2 (thanks to property MG1), we see from Lemma 2.2 that

$$M_n = \mathbb{E}[M_\infty \,|\, \mathcal{F}_n].$$

This concludes the proof. □

As a converse to Theorem 4.13, if $\xi \in L^1(\mathbb{P})$ is a random variable on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$, then the conditional expected values $\big(\mathbb{E}[\xi \,|\, \mathcal{F}_n]\big)_{n\in\mathbb{N}_0}$ define a UI martingale by Lemma 4.14 stated below. In particular,

$$\mathbb{E}[\xi \,|\, \mathcal{F}_n] \overset{L^1}{\longrightarrow} \xi \qquad \text{and} \qquad \mathbb{E}[\xi \,|\, \mathcal{F}_n] \overset{\text{a.s.}}{\longrightarrow} \xi.$$

This essentially yields, on a given filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$, a *bijective correspondence* between UI martingales and random variables $\xi \in L^1(\mathbb{P})$ (up to indistinguishability).

Lemma 4.14. *Let $\xi \in L^1(\mathbb{P})$ be an integrable random variable on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $(\mathscr{G}_j)_{j\in J}$ be any collection of sub-sigma-algebras $\mathscr{G}_j \subset \mathscr{F}$. The conditional expected values $\big(\mathbb{E}[\xi \,|\, \mathscr{G}_j]\big)_{j\in J}$ form a UI collection of random variables on $(\Omega, \mathscr{F}, \mathbb{P})$:*

$$\lim_{R\to\infty} \sup_{j\in J} \mathbb{E}\Big[|\mathbb{E}[\xi \,|\, \mathscr{G}_j]| \, \mathbb{1}\{|\mathbb{E}[\xi \,|\, \mathscr{G}_j]| \geq R\}\Big] = 0.$$

Exercise 4.15. Prove Lemma 4.14 using Lemma 4.16.

Lemma 4.16. *For any integrable random variable $\xi \in L^1(\mathbb{P})$, we have*

$$\sup\Big\{\mathbb{E}\big[|\xi| \, \mathbb{1}_A\big] \,\big|\, A \in \mathscr{F} \text{ and } \mathbb{P}[A] \leq \delta\Big\} \overset{\delta\to 0}{\longrightarrow} 0.$$

Exercise 4.17. Prove Lemma 4.16.
Hint: You can use the First Borel-Cantelli Lemma A.14 and Markov's Inequality (A.5).

4.4 Toolbox: Doob's maximal inequalities ($\star$)

For later use, we record here useful inequalities for (sub)martingales, also due to Doob.

- ▷ The first one (Proposition 4.18) states, roughly speaking, that the expected value of a (non-negative sub)martingale M at time n can be used to control the probability that M reaches some given value λ before time n, somewhat analogously to Markov's Inequality (A.5).
- ▷ The second one (Proposition 4.21) is a similar maximal inequality for the square of a (non-negative sub)martingale.

Proposition 4.18. (Doob's maximal inequality) *Let M be a martingale or a non-negative submartingale. Then, we have*

$$\begin{aligned} \mathbb{P}\Big[\max_{0\le k\le n} |M_k| \ge \lambda\Big] &\le \tfrac{1}{\lambda}\,\mathbb{E}\Big[|M_n|\,\mathbb{1}\Big\{\max_{0\le k\le n}|M_k|\ge\lambda\Big\}\Big] \\ &\le \tfrac{1}{\lambda}\,\mathbb{E}[|M_n|], \qquad n\in\mathbb{N},\ \lambda>0. \end{aligned} \tag{4.7}$$

We present a proof relying directly on properties of submartingales and conditional expected value (Lemmas 2.8 and 3.5). See Exercise 4.19 for an alternative proof using Optional Stopping.

Proof. If M is a martingale, then $|M|$ is a submartingale by Corollary 3.31. Hence it suffices to verify (4.7) for a non-negative submartingale M. Fix $n\in\mathbb{N}$ and $\lambda>0$, and consider the event

$$E_n(\lambda) := \Big\{\max_{0\le k\le n} M_k \ge \lambda\Big\} := \bigsqcup_{k=1}^{n} A_k(\lambda), \tag{4.8}$$

where

$$\begin{aligned} &A_0(\lambda) := \{M_0\ge\lambda\}, \\ &A_k(\lambda) := \{M_k\ge\lambda\}\cap\{M_j<\lambda \text{ for all } 0\le j\le k-1\}, \qquad k\in\{1,2,\ldots,n\}. \end{aligned}$$

Note that $M_k(\omega)\ge\lambda$ when $\omega\in A_k(\lambda)$. Hence, we have

$$\begin{aligned} \lambda\,\mathbb{P}[A_k(\lambda)] &= \lambda\int_{A_k(\lambda)} \mathrm{d}\mathbb{P}(\omega) \\ &\le \int_{A_k(\lambda)} M_k(\omega)\,\mathrm{d}\mathbb{P}(\omega) && [\text{since } M_k(\omega)\ge\lambda] \\ &\le \int_{A_k(\lambda)} \mathbb{E}[M_n\,|\,\mathcal{F}_k](\omega)\,\mathrm{d}\mathbb{P}(\omega) && [\text{by Lemma 3.5}] \\ &= \mathbb{E}\big[\mathbb{1}_{A_k(\lambda)}\,\mathbb{E}[M_n\,|\,\mathcal{F}_k]\big] \\ &= \mathbb{E}\big[\mathbb{E}[M_n\,\mathbb{1}_{A_k(\lambda)}\,|\,\mathcal{F}_k]\big] && [\text{by item 5 of Lemma 2.8, as } A_k(\lambda)\in\mathcal{F}_k] \\ &= \mathbb{E}\big[M_n\,\mathbb{1}_{A_k(\lambda)}\big]. && [\text{by Tower property (item 4 of Lemma 2.8)}] \end{aligned}$$

Using the disjoint union (4.8), we find that

$$\mathbb{P}[E_n(\lambda)] = \sum_{k=0}^{n}\mathbb{P}[A_k(\lambda)] \le \tfrac{1}{\lambda}\sum_{k=0}^{n}\mathbb{E}\big[M_n\,\mathbb{1}_{A_k(\lambda)}\big] \le \tfrac{1}{\lambda}\,\mathbb{E}\big[M_n\,\mathbb{1}_{E_n(\lambda)}\big],$$

which gives the first asserted inequality (4.7), while second is immediate, since $\mathbb{1}_{E_n(\lambda)}\le 1$. □

Exercise 4.19. Prove (4.7) for a non-negative submartingale M using Optional Stopping (Proposition 3.21). *Hint: Consider a stopping time* $\tau = \inf\{k \geq 0 \mid M_k \geq \lambda\} \wedge n$ *and the two events* $E_n(\lambda)$ *and* $\Omega \smallsetminus E_n(\lambda)$ *from* (4.8).

Corollary 4.20. *Let M be a martingale or a non-negative submartingale. Then, we have*

$$\mathbb{P}\Big[\sup_{k\geq 0} |M_k| \geq \lambda\Big] \leq \tfrac{1}{\lambda} \sup_{k\geq 0} \mathbb{E}[|M_k|], \qquad \lambda > 0. \tag{4.9}$$

Proof. Proposition 4.18 gives

$$\mathbb{P}\Big[\max_{0\leq k\leq n} |M_k| \geq \lambda\Big] \leq \tfrac{1}{\lambda} \mathbb{E}[|M_n|] \leq \tfrac{1}{\lambda} \sup_{k\geq 0} \mathbb{E}[|M_k|], \qquad n \in \mathbb{N},\ \lambda > 0.$$

We obtain (4.9) by Monotone Convergence Theorem [Kyt20, Theorem VII.8] with $n \to \infty$. □

Proposition 4.21. (Doob's L^2-maximal inequality) *Let M be a martingale or a non-negative submartingale. Then, we have*

$$\mathbb{E}\Big[\max_{0\leq k\leq n} M_k^2\Big] \leq 4\,\mathbb{E}[M_n^2], \qquad n \in \mathbb{N}. \tag{4.10}$$

Proof. If M is a martingale, then $|M|$ is a submartingale by Corollary 3.31. Hence, it suffices to verify (4.10) for a non-negative submartingale M.

Note that the claim is clear if $\mathbb{E}[M_k^2] = \infty$ for some $0 \leq k \leq n$, since the submartingale property (Lemma 3.5) then shows that $\mathbb{E}[M_n^2] = \infty$ as well. Hence, we assume that $\mathbb{E}[M_n^2] < \infty$, so both sides of (4.10) are finite. Let us simplify the notation by denoting

$$Y_n := \max_{0\leq k\leq n} M_k^2 \qquad \text{and} \qquad Z_n := \max_{0\leq k\leq n} M_k.$$

Note that since $M_k \geq 0$ for all k, we have $Y_n = Z_n^2$. Note also that (cf. Exercise 4.22)

$$\mathbb{E}[Z_n^2] = 2\int_0^\infty \lambda\,\mathbb{P}[Z_n \geq \lambda]\,\mathrm{d}\lambda. \tag{4.11}$$

Thus, Doob's maximal inequality (4.7) implies that

$$\begin{aligned}
\mathbb{E}[Y_n] &= \mathbb{E}[Z_n^2] && \text{[since } M_k \geq 0 \text{ for all } k\text{, so } Y_n = Z_n^2] \\
&= 2\int_0^\infty \lambda\,\mathbb{P}[Z_n \geq \lambda]\,\mathrm{d}\lambda && \text{[by (4.11)]} \\
&\leq 2\int_0^\infty \lambda\,\tfrac{1}{\lambda}\,\mathbb{E}[M_n\,\mathbb{1}_{E_n(\lambda)}]\,\mathrm{d}\lambda && \text{[by (4.7)]} \\
&\leq 2\,\mathbb{E}\Big[M_n \underbrace{\int_0^\infty \mathbb{1}_{E_n(\lambda)}\,\mathrm{d}\lambda}_{=\int_0^{Z_n}\mathrm{d}\lambda = Z_n}\Big] && \text{[by Fubini's theorem A.19]} \\
&= 2\,\mathbb{E}[M_n Z_n] \leq 2\big(\mathbb{E}[M_n^2]\,\mathbb{E}[Z_n^2]\big)^{1/2} && \text{[by Cauchy-Schwarz Inequality (Lemma 2.17)]} \\
&= 2\big(\mathbb{E}[M_n^2]\,\mathbb{E}[Y_n]\big)^{1/2}, && \text{[since } M_k \geq 0 \text{ for all } k\text{, so } Y_n = Z_n^2]
\end{aligned}$$

using the notation from the proof of Proposition 4.18. Now, we consider two cases:

1. If $\mathbb{E}[Y_n] = 0$, then the left-hand side of (4.10) equals zero, so the claim is clear.
2. If $\mathbb{E}[Y_n] > 0$, then we find from the above analysis that $\big(\mathbb{E}[Y_n]\big)^{1/2} \le 2\big(\mathbb{E}[M_n^2]\big)^{1/2}$, and squaring both sides gives asserted inequality (4.10). □

Exercise 4.22. Show that, for a square-integrable random variable $\xi \in L^2(\mathbb{P})$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, we have

$$\mathbb{E}[\xi^2] = 2\int_0^\infty \lambda\, \mathbb{P}[\xi \ge \lambda]\, \mathrm{d}\lambda.$$

Hint: You can use Fubini's theorem A.19.

Corollary 4.23. *Let M be a martingale or a non-negative submartingale. Then, we have*

$$\mathbb{E}\Big[\sup_{k\ge 0} M_k^2\Big] \;\le\; 4 \sup_{k\ge 0} \mathbb{E}[M_k^2]. \tag{4.12}$$

Proof. Proposition 4.21 gives

$$\mathbb{E}\Big[\max_{0\le k\le n} M_k^2\Big] \;\le\; 4\,\mathbb{E}[M_n^2] \;\le\; 4 \sup_{k\ge 0} \mathbb{E}[M_k^2], \qquad n \in \mathbb{N}.$$

We obtain (4.12) by Monotone Convergence Theorem [Kyt20, Theorem VII.8] with $n \to \infty$. □

5 On continuous-time processes and measurability issues

In order to discuss Brownian motion as a *continuous-time* Markov process and a martingale, we next turn to generalizing some concepts from the discrete-time theory to continuous time. One of the main caveats here is that the time index set $[0,\infty)$ is *uncountable*, while $\mathbb{N}_0$ is countable. Thus, one easily runs into issues concerning measurability. We try to avoid delving too much into these issues here, but we shall highlight the main ideas and important concepts.

5.1 Filtrations and stopping times

Throughout this section, we work with a fixed probability space $(\Omega, \mathscr{F}, \mathbb{P})$. We consider continuous-time stochastic processes $X = (X_t)_{t\geq 0}$ taking real values: $X_t : \Omega \to \mathbb{R}$ for all t.

Definition 5.1. (Filtration) A *filtration* is a collection $\mathcal{F}_\bullet = (\mathcal{F}_t)_{t\geq 0}$ of sub-sigma-algebras of $\mathscr{F}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ for all $0 \leq s \leq t$. The tuple $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is called a *filtered probability space.*

Definition 5.2. Filtration $\mathcal{F}_\bullet$ is called *right-continuous* if

$$\mathcal{F}_t = \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon} \qquad \text{for all } t \geq 0.$$

Right-continuity of filtrations is sometimes needed, e.g., when considering stopping times (see Lemma 5.4 for the first example). While perhaps being just a curiosity at this point, right-continuity becomes quite important later on when constructing stochastic integrals.

Definition 5.3. (Stopping time) Let $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ be a filtered probability space. Random variable $\tau : \Omega \to [0,+\infty]$ is called a *stopping time* w.r.t. $\mathcal{F}_\bullet$ if

$$\{\tau \leq t\} := \{\omega \in \Omega \mid \tau(\omega) \leq t\} \in \mathcal{F}_t \qquad \text{for all } t \geq 0.$$

Lemma 5.4. *Let $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ be a filtered probability space with* right-continuous *filtration $\mathcal{F}_\bullet$. Then, we have*

$$\{\tau \leq t\} \in \mathcal{F}_t \qquad \Longleftrightarrow \qquad \{\tau < t\} \in \mathcal{F}_t.$$

Proof.

"⇒" We can write

$$\{\tau < t\} = \bigcup_{n\in\mathbb{N}} \underbrace{\left\{\tau \leq t - \tfrac{1}{n}\right\}}_{\in\, \mathcal{F}_{t-1/n}} \in \mathcal{F}_t.$$

"⇐" If $\{\tau < t\} \in \mathcal{F}_t$ for all $t \geq 0$, then we have

$$\{\tau \leq t\} = \bigcap_{n=m}^{\infty} \underbrace{\left\{\tau < t + \tfrac{1}{n}\right\}}_{\in\, \mathcal{F}_{t+1/n} \subset \mathcal{F}_{t+1/m}} \in \mathcal{F}_{t+\frac{1}{m}}, \qquad m \in \mathbb{N},$$

which implies that

$$\{\tau \le t\} \in \bigcap_{m\in\mathbb{N}} \mathcal{F}_{t+\frac{1}{m}} = \mathcal{F}_t,$$

where in the last equality we used the *right-continuity* of the filtration. □

5.2 Continuous-time processes and usual conditions

Throughout this section, we consider a real-valued stochastic process $X = (X_t)_{t\ge 0}$ on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$, and discuss properties of X and of the filtration $\mathcal{F}_\bullet$.

Definition 5.5. Stochastic process X is said to be *adapted to a filtration* $\mathcal{F}_\bullet$ if

$$X_t \in \mathrm{m}\mathcal{F}_t := \{\chi : \Omega \to \mathbb{R} \mid \chi \text{ is } \mathcal{F}_t\text{-measurable}\} \quad \text{for all } t \ge 0.$$

Any process X generates its *natural filtration* $\mathcal{F}^X_\bullet := (\mathcal{F}^X_t)_{t\ge 0}$ as the history up to time t,

$$\mathcal{F}^X_t := \sigma(\{X_s \mid 0 \le s \le t\}).$$

- ▷ To rule out pathologies, we might want all subsets of zero-probability events[12] to be *measurable*. This is accomplished by simply *adding* them to the filtration. The resulting construction is called the *natural augmented filtration* $\mathcal{F}^a_\bullet := (\mathcal{F}^a_t)_{t\ge 0}$ defined as

$$\mathcal{F}^a_t := \sigma(\mathcal{F}^X_t \cup \mathscr{N}), \qquad \mathscr{N} := \big\{N \subset \Omega \mid N \subset E \in \mathscr{F} \text{ for some } E \text{ with } \mathbb{P}[E] = 0\big\}. \tag{5.1}$$

- ▷ The natural augmented filtration is not always right-continuous. Thus, it is often convenient to consider the *natural right-continuous augmented filtration* $\mathcal{F}^+_\bullet := (\mathcal{F}^+_t)_{t\ge 0}$ defined as

$$\mathcal{F}^+_t := \bigcap_{\varepsilon>0} \mathcal{F}^a_{t+\varepsilon} = \bigcap_{\varepsilon>0} \sigma(\mathcal{F}^X_{t+\varepsilon} \cup \mathscr{N}), \qquad t \ge 0.$$

We will use these natural filtrations for Brownian motion in the next Section 6.

In general, a given filtration is said to satisfy the *usual conditions* if the above hold. These conditions are needed, e.g., when dealing with some properties of stopping times (see Lemma 5.7).

Definition 5.6. (Usual conditions) Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Filtration $\mathcal{F}_\bullet$ is said to satisfy the *usual conditions* if it is

UC1. *right-continuous*, that is,

$$\mathcal{F}_t = \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon} \qquad \text{for all } t \ge 0;$$

UC2. *complete*, that is, it contains all negligible sets (5.1): $\mathscr{N} \subset \mathcal{F}_0$.

Unless there are specific reasons not to, it is conventional to exclusively work with filtrations satisfying the usual conditions UC1 & UC2. One can then usually deal with almost sure properties and properties concerning open sets without worrying about measurability issues.

[12] Subsets of zero-probability events are sometimes called *negligible* sets.

From now on, we will consider a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ where the filtration $\mathcal{F}_\bullet$ satisfies the usual conditions (cf. Definition 5.6). We then say briefly that *the filtered probability space* $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ *satisfies the usual conditions.*

Lemma 5.7. (Entering times) *Let* $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ *be a filtered probability space satisfying the usual conditions. Consider a stochastic process* $X = (X_t)_{t\geq 0}$ *adapted to* $\mathcal{F}_\bullet$ *which has a.s. continuous sample paths,*

$$\mathbb{P}\Big[\{\omega \in \Omega \mid t \mapsto X_t(\omega) \text{ is continuous}\}\Big] = 1.$$

Then, the following hold.

1. *For any closed set* $F \in \mathscr{B}(\mathbb{R})$*, the following random variable is a stopping time*:

$$\tau_F := \inf\{t \geq 0 \mid X_t \in F\}.$$

2. *For any open set* $O \in \mathscr{B}(\mathbb{R})$*, the following random variable is a stopping time*:

$$\tau_O := \inf\{t \geq 0 \mid X_t \in O\}.$$

Proof. Note that by assumption, $\mathbb{P}[E_{\text{cont}}] = 1$ and $\mathbb{P}[E^c_{\text{cont}}] = 0$, where

$$E_{\text{cont}} := \{\omega \in \Omega \mid t \mapsto X_t(\omega) \text{ is continuous}\} \qquad \text{and} \qquad E^c_{\text{cont}} := \Omega \setminus E_{\text{cont}}.$$

By property UC2, $\mathcal{F}_\bullet$ contains all negligible sets (5.1), so we have

$$E \cap E^c_{\text{cont}} \in \mathcal{F}_0 \qquad \text{for all } E \subset \Omega, \tag{5.2}$$

which also implies that $(\Omega \setminus (E \cap E^c_{\text{cont}})) \in \mathcal{F}_0$ and thus,

$$E \cap E_{\text{cont}} = E \cap \underbrace{(\Omega \setminus (E \cap E^c_{\text{cont}}))}_{\in \mathcal{F}_0 \subset \mathcal{F}_t} \in \mathcal{F}_t \qquad \text{for all } E \in \mathcal{F}_t. \tag{5.3}$$

Hence, restricting attention to the set E_{cont}, or to the set E^c_{cont}, retains measurability.

1. Fix $t \geq 0$ and consider $\{\tau_F \leq t\} = (\{\tau_F \leq t\} \cap E^c_{\text{cont}}) \cup (\{\tau_F \leq t\} \cap E_{\text{cont}})$. We have
   - ▷ $\{\tau_F \leq t\} \cap E^c_{\text{cont}} \in \mathcal{F}_0 \subset \mathcal{F}_t$ by (5.2), and
   - ▷ since the map $x \mapsto \text{dist}(x, F)$ is continuous, we have by (5.3)

$$\{\tau_F \leq t\} \cap E_{\text{cont}} = \Big\{\inf_{q \in [0,t] \cap \mathbb{Q}} \underbrace{\text{dist}(X_q, F)}_{\in \text{m}\mathcal{F}_q \subset \text{m}\mathcal{F}_t} = 0\Big\} \cap E_{\text{cont}} \in \mathcal{F}_t.$$

2. Fix $t \geq 0$ and consider $\{\tau_O < t\} = (\{\tau_O < t\} \cap E^c_{\text{cont}}) \cup (\{\tau_O < t\} \cap E_{\text{cont}})$. We have
   - ▷ $\{\tau_O < t\} \cap E^c_{\text{cont}} \in \mathcal{F}_0 \subset \mathcal{F}_t$ by (5.2), and
   - ▷ since on the event E_{cont} the set $\{t \geq 0 \mid X_t \in O\}$ is open, we have by (5.3)

$$\{\tau_O < t\} \cap E_{\text{cont}} = \Big(\bigcup_{q \in [0,t) \cap \mathbb{Q}} \underbrace{\{X_q \in O\}}_{\in \mathcal{F}_q \subset \mathcal{F}_t}\Big) \cap E_{\text{cont}} \in \mathcal{F}_t.$$

Assertion 2 then follows from Lemma 5.4 (which uses right-continuity UC1 of $\mathcal{F}_\bullet$). □

5.3 Stopped processes and progressive measurability

Throughout this section, we consider a real-valued stochastic process $X = (X_t)_{t\geq 0}$ on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$. In Sections 5.1–5.2, we have discussed the notion of stopping times and usual conditions for $\mathcal{F}_\bullet$. Often we would like to consider the *value* of X at a stopping time τ. However, two problems immediately arise:

- ▷ a stopping time $\tau : \Omega \to [0, +\infty]$ can have positive probability of being *infinite*;
- ▷ the map $\omega \mapsto X_{\tau(\omega)}(\omega)$ defined as the composition

$$\begin{array}{ccccc} \Omega & \to & \Omega \times [0, +\infty] & \to & \mathbb{R}, \\ \omega & \mapsto & (\omega, \tau(\omega)) & \mapsto & X_{\tau(\omega)}(\omega) \end{array}$$

  might fail to be *measurable*, in which case $X_{\tau(\omega)}(\omega)$ is *not a random variable.*

On the event $\{\tau = \infty\}$ we set $X_{\tau(\omega)} := 0$ by convention. If this event has probability zero, since we anyway consider processes up to indistinguishability, no harm is done. To address the measurability of $X_{\tau(\omega)}(\omega)$, we need to introduce some terminology.

5.3.1 Progressive measurability

Definition 5.8. Real-valued stochastic process X on $(\Omega, \mathscr{F}, \mathbb{P})$ is said to be *measurable* if

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

is a (Borel-)measurable map from $(\Omega \times [0, \infty), \mathscr{F} \otimes \mathscr{B}([0, \infty)))$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Definition 5.9. Real-valued stochastic process X on $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is said to be *progressively measurable* with respect to the filtration $\mathcal{F}_\bullet$ if for each fixed time $T \in [0, \infty)$,

$$(\omega, t) \mapsto X_t(\omega), \qquad t \in [0, T],$$

is a (Borel-)measurable map from $(\Omega \times [0, T], \mathcal{F}_T \otimes \mathscr{B}([0, T]))$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Definition 5.10. The *progressive sigma-algebra* $\mathscr{P}$ on $\Omega \times [0, \infty)$ is defined by

$$G \in \mathscr{P} \quad \Longleftrightarrow \quad \text{for each } T \geq 0 \text{ we have } G \cap (\Omega \times [0, T]) \in \mathcal{F}_T \otimes \mathscr{B}([0, T]), \tag{5.4}$$

i.e., $\mathscr{P}$ is the collection of all sets $G \in \mathscr{F} \otimes \mathscr{B}([0, \infty))$ such that $(\omega, t) \mapsto \mathbb{1}_G(\omega, t)$ is progressively measurable. The collection

$$\Pi_T := \{E \times (u, v] \mid E \in \mathcal{F}_T \text{ and } 0 \leq u < v \leq T\} \tag{5.5}$$

is a *pi-system* on $\Omega \times [0, T]$ *generating* $\mathcal{F}_T \otimes \mathscr{B}([0, T])$.

Thus, X is progressively measurable if and only if the map $(\omega, t) \mapsto X_t(\omega)$ is measurable on $(\Omega \times [0, \infty), \mathscr{P})$. (This will become important when defining stochastic integrals in Section 9.)

Note that a progressively measurable process is both adapted and measurable. Theorem 5.16 gives a converse statement, but one has to possibly pass to a different modification. We also discuss other criteria for progressive measurability in Section 5.3.3. In particular, we will see in a couple of different ways that Brownian motion has a progressively measurable modification.

5.3.2 Stopped processes in continuous time

We are now ready to define the stopped process properly.

Definition 5.11. Let τ be a stopping time, and let X be a progressively measurable process. Then, we set

$$X_\tau(\omega) := \mathbb{1}\{\tau(\omega) < \infty\} \, X_{\tau(\omega)}(\omega). \tag{5.6}$$

We also define the *stopped process* as $X^\tau := (X_{t\wedge\tau})_{t\geq 0}$.

We need to check that the stopped process is actually a stochastic process.

Lemma 5.12. *Let τ be a stopping time, and let X be a progressively measurable process. Then, X_τ defined in* (5.6) *is a random variable measurable with respect to*

$$\mathcal{F}_\tau := \{A \in \mathcal{F}_\infty \mid A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\} \;\subset\; \mathscr{F}, \tag{5.7}$$

where $\mathcal{F}_\infty = \sigma\big(\bigcup_{t\geq 0} \mathcal{F}_t\big)$.

The sigma-algebra $\mathcal{F}_\tau$ represents information available up to the random time τ.

Exercise 5.13. Consider a stopping time τ with respect to a filtration $\mathcal{F}_\bullet$.

1. Check that $\mathcal{F}_\tau$ defined in (5.7) is a sigma-algebra.
2. Show that τ is $\mathcal{F}_\tau$-measurable.
3. Fix $s \geq 0$. Show that events $\{\tau > s\}$, $\{\tau < s\}$, and $\{\tau = s\}$ belong to $\mathcal{F}_\tau$.
4. Let σ be a stopping time w.r.t. $\mathcal{F}_\bullet$. Show that the events $\{\tau > \sigma\}$, $\{\tau < \sigma\}$, and $\{\tau = \sigma\}$ belong to $\mathcal{F}_\tau$.

Proof. The idea is to truncate the stopping time τ by some finite time $t \geq 0$ and then take $t \to \infty$.

Step 0. Since X is progressively measurable, the map $(\omega, s) \mapsto X_s(\omega)$ is a (Borel-)measurable map from $(\Omega \times [0,t], \mathcal{F}_t \otimes \mathscr{B}([0,t]))$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ for each fixed time $t \in [0,\infty)$.

Step 1. Consider first the restriction of the map $\omega \mapsto X_\tau(\omega)$ to the event $\{\tau \leq t\}$, which is $\mathcal{F}_t$-measurable because τ is a stopping time. This map is a composition of the following maps:

$$\begin{array}{ccccc} \underbrace{\{\tau \leq t\}}_{\in \mathcal{F}_t} & \to & \Omega \times [0,t] & \to & \mathbb{R}, \\ \omega & \mapsto & (\omega, \tau(\omega) \wedge t) & \mapsto & X_{\tau(\omega)\wedge t}(\omega). \end{array}$$

▷ The first map is measurable because $\tau \wedge t \in \mathrm{m}\mathcal{F}_t$. Indeed, since τ is a stopping time, we have

$$\{\tau \wedge t \le s\} = \Big(\underbrace{\{\tau \le t\}}_{\in \mathcal{F}_t} \cap \underbrace{\{\tau \le s\}}_{\in \mathcal{F}_s \subset \mathcal{F}_t} \Big) \cup \Big(\underbrace{\{t < \tau\}}_{\substack{= \{\tau \le t\}^c \\ \in \mathcal{F}_t}} \cap \underbrace{\{\tau \le s\}}_{\in \mathcal{F}_s \subset \mathcal{F}_t} \Big) \in \mathcal{F}_t \qquad \text{for all } 0 \le s \le t.$$

As the sets $\{[0,s] \mid 0 \le s \le t\}$ generate the sigma-algebra $\mathscr{B}([0,t])$, this implies $\tau \wedge t \in \mathrm{m}\mathcal{F}_t$.

▷ The second map is measurable from $(\Omega \times [0,t], \mathcal{F}_t \otimes \mathscr{B}([0,t]))$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ by Step 0.

Step 2. We then take $t \to \infty$ to deduce that $\{X_\tau \in A\} \in \mathcal{F}_\tau$ for all $A \in \mathscr{B}(\mathbb{R})$. Indeed, we have

$$\underbrace{\{X_\tau \in A\} \cap \{\tau \le t\}}_{\in \mathcal{F}_t} \quad \xrightarrow{t\to\infty} \quad \underbrace{\{X_\tau \in A\}}_{\in \mathcal{F}_\infty},$$

which shows that $\{X_\tau \in A\} \in \mathcal{F}_\tau$. This also implies that $X_\tau \in \mathscr{F}$, so it is a random variable. □

Corollary 5.14. *Let τ a stopping time, and let X be a progressively measurable process. Then, the stopped process $X^\tau := (X_{t\wedge\tau})_{t\ge 0}$ is also a progressively measurable process.*

Proof. This follows from the proof of Lemma 5.12. We leave the details as an exercise. □

Exercise 5.15. Prove Corollary 5.14.

5.3.3 Criteria for progressive measurability (⋆)

Note that a progressively measurable process is both adapted and measurable. The next result gives a converse statement, but one has to possibly pass to a different modification.

Theorem 5.16. *Every measurable adapted process on $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ has a progressively measurable modification.*

Proof. For an elementary but complicated proof, see [OS13]. See also references therein. □

Exercise 5.17. Find an example of a process which is measurable and adapted, but not progressively measurable. Construct a progressively measurable modification for it.

Corollary 5.18. *Brownian motion has a progressively measurable modification.*

Proof. This follows by constructing an adapted and measurable modification of Brownian motion. One construction is by Lévy's construction of Brownian motion [MP10, Theorem 1.3]. □

There is, in fact, a much easier way to deduce that Brownian motion has a progressively measurable modification. Indeed, we will next show that *continuity of sample paths implies progressive measurability*. This would be, of course, a direct consequence of Theorem 5.16 (as continuous functions are measurable), but proving it directly is very elementary. This result holds for processes taking values in metric spaces — for ease, we consider it for real-valued processes.

Proposition 5.19. *Consider an adapted stochastic process $X = (X_t)_{t\geq 0}$ on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$. Suppose that for every $\omega \in \Omega$, the sample path $t \mapsto X_t(\omega)$ is* right-continuous*. Then, X is progressively measurable.*

We leave it as an exercise to check the following modifications of the above result.

- ▷ The same conclusion also holds after replacing right-continuous by *left-continuous*.
- ▷ If the process X has *almost surely* (a.s.) right-continuous sample paths, and the filtration $\mathcal{F}_\bullet$ is *complete* (e.g., satisfies the usual conditions), then the same conclusion also holds.

Proof. Fix $t > 0$. We need to show that $(\omega, s) \mapsto X_s(\omega)$ is a measurable map

$$(\Omega \times [0,t], \mathcal{F}_t \otimes \mathscr{B}([0,t])) \quad \longrightarrow \quad (\mathbb{R}, \mathscr{B}(\mathbb{R})).$$

For $n \in \mathbb{N}$ and $s \in [0,t]$, define a random variable

$$X_s^{(n)} := \begin{cases} X_{kt/n}, & \text{when } s \in \big[\frac{(k-1)t}{n}, \frac{kt}{n}\big) \text{ with } k \in \{1,2,\ldots,n\} \\ X_t. \end{cases}$$

Then, since the sample paths of X are right-continuous, we have

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

for every $\omega \in \Omega$ and $s \in [0,t]$. Also, for each $A \in \mathscr{B}(\mathbb{R})$, we have

$$\begin{aligned} &\big\{(\omega,s) \in \Omega \times [0,t] \,\big|\, X_s^{(n)}(\omega) \in A\big\} \\ &= \Big(\underbrace{\{X_t \in A\}}_{\in \mathcal{F}_t} \times \{t\}\Big) \cup \bigcup_{k=1}^{n} \Big(\underbrace{\{X_{kt/n} \in A\}}_{\in \mathcal{F}_{kt/n} \subset \mathcal{F}_t} \times \big[\tfrac{(k-1)t}{n}, \tfrac{kt}{n}\big)\Big) \in \mathcal{F}_t \otimes \mathscr{B}([0,t]). \end{aligned}$$

Thus, for each $n \in \mathbb{N}$, the map $(\omega, s) \mapsto X_s^{(n)}(\omega)$ is a measurable map

$$(\Omega \times [0,t], \mathcal{F}_t \otimes \mathscr{B}([0,t])) \quad \longrightarrow \quad (\mathbb{R}, \mathscr{B}(\mathbb{R})).$$

Since $X_s(\omega)$ is a pointwise limit of these measurable maps, it is also measurable. □

Corollary 5.20. *Brownian motion has a progressively measurable modification with respect to any of the natural filtrations $\mathcal{F}_\bullet^+$, $\mathcal{F}_\bullet^a$, and $\mathcal{F}_\bullet^B$.*

Proof. The Wiener process W in Definition 1.30 is a modification of Brownian motion that satisfies the assumptions in Proposition 5.19. Hence, the claim follows immediately. □

We call a process X which has a.s. continuous sample paths, i.e.,

$$\mathbb{P}\Big[\{\omega \in \Omega \mid t \mapsto X_t(\omega) \text{ is continuous}\}\Big] = 1,$$

briefly a *continuous process*. As we work with processes up to indistinguishability and assume that the filtration $\mathcal{F}_\bullet$ satisfies the usual conditions, we omit "almost surely (a.s.)."

Corollary 5.21. *Let $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ be a filtered probability space satisfying the usual conditions. Let τ a stopping time, and consider an* adapted continuous *process $X = (X_t)_{t\geq 0}$. Then, the stopped process $X^\tau := (X_{t\wedge\tau})_{t\geq 0}$ is also an adapted continuous process.*

Proof. By Proposition 5.19, the process X is progressively measurable, so by Corollary 5.14, the process X^τ is well-defined and adapted. The continuity of X^τ is clear from its definition. □

One could relax the continuity assumption to the requirement that X has a.s. *càdlàg* sample paths (X is a càdlàg process). In that case, the stopped process X^τ is càdlàg as well.

6 Brownian motion as a Markov process

After the background knowledge from Sections 5.1–5.2, we are ready to begin to consider basic properties of Brownian motion as a continuous-time process.

6.1 Markov and martingale property for Brownian motion

We shall mainly work with the *natural right-continuous augmented filtration* of Brownian motion,

$$\mathcal{F}_\bullet^+ := (\mathcal{F}_t^+)_{t\geq 0}, \qquad \mathcal{F}_t^+ := \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}^a, \qquad t \geq 0, \tag{6.1}$$

where $\mathcal{F}_t^a := \sigma(\mathcal{F}_t^B \cup \mathscr{N})$ is the *augmented* filtration obtained from Brownian motion's *natural* filtration $\mathcal{F}_t^B := \sigma(\{B_s \mid 0 \leq s \leq t\})$. Keeping mind readers interested in these nuances, we shall indicate the various filtrations with the different notations $\mathcal{F}_\bullet^B \subset \mathcal{F}_\bullet^a \subset \mathcal{F}_\bullet^+$. However, it is safe and convenient to always work with $\mathcal{F}_\bullet^+$ (which satisfies the usual conditions, cf. Definition 5.6).

The Markov Property of Brownian motion shows that for predicting what happens in the future, any past information besides the current state is irrelevant.

Proposition 6.1. (Markov Property) *Let B be a (standard) Brownian motion. For any $s > 0$, the process $(B_{t+s} - B_s)_{t\geq 0}$ is also a (standard) Brownian motion, which is independent of $(B_t)_{t\in[0,s]}$ and of the sigma-algebras $\mathcal{F}_s^+$, $\mathcal{F}_s^a$, and $\mathcal{F}_s^B$.*

Exercise 6.2. Prove Proposition 6.1 for the natural filtration $\mathcal{F}_\bullet^B$.

We prove Proposition 6.1 for the natural right-continuous augmented filtration $\mathcal{F}_\bullet^+$ (which is its strongest one, since $\mathcal{F}_\bullet^B \subset \mathcal{F}_\bullet^a \subset \mathcal{F}_\bullet^+$) slightly later.

Corollary 6.3. (Martingale property) *Let B be a (standard) Brownian motion. It has the martingale property* (MG3) *for any of the filtrations $\mathcal{F}_\bullet^+$, $\mathcal{F}_\bullet^a$, and $\mathcal{F}_\bullet^B$: almost surely,*

$$\mathbb{E}[M_t \,|\, \mathcal{F}_s^\star] = M_s \qquad \text{for every } 0 \leq s < t, \tag{MG3}$$

with $\star \in \{+, a, B\}$.

Proof. Because $\mathcal{F}_\bullet^B \subset \mathcal{F}_\bullet^a \subset \mathcal{F}_\bullet^+$, basic properties of conditional expected value from Lemma 2.8 guarantee that it suffices to prove the martingale property (MG3) for the largest filtration $\mathcal{F}_\bullet^+$:

$$\mathbb{E}[B_t \,|\, \mathcal{F}_s^+] = B_s, \qquad 0 \leq s < t. \tag{6.2}$$

Indeed, assuming (6.2), since $\mathcal{F}_s^B \subset \mathcal{F}_s^+$, the Tower Property (item 4 of Lemma 2.8) gives

$$\begin{aligned} \mathbb{E}[B_t \,|\, \mathcal{F}_s^B] &= \mathbb{E}\big[\mathbb{E}[B_t \,|\, \mathcal{F}_s^+] \,|\, \mathcal{F}_s^B\big] && \text{[by item 4 of Lemma 2.8]} \\ &= \mathbb{E}\big[B_s \,|\, \mathcal{F}_s^B\big] && \text{[by (6.2)]} \\ &= B_s, && \text{[by item 1 of Lemma 2.8]} \end{aligned}$$

and similarly for $\mathcal{F}_s^a \subset \mathcal{F}_s^+$. The derivation of (6.2) is item 1 of Exercise 6.4. □

Exercise 6.4. (Martingales from Brownian motion) Let B be a standard Brownian motion.

1. Show that the following processes are martingales w.r.t. $\mathcal{F}^+_\bullet$ (see Definition 7.1 in Section 7):
$$(B_t)_{t\geq 0}, \qquad (B_t^2 - t)_{t\geq 0}.$$
Hint: *You can use the Markov Property of Brownian motion* (*Proposition 6.1*).

2. Fix a constant $\alpha \in \mathbb{R}$ and suppose that $\xi \sim N(0,1)$ is a standard Gaussian random variable. Calculate $\mathbb{E}[e^{\alpha\xi}]$.

3. Fix a constant $\alpha \in \mathbb{R}$. Find a constant $\gamma \in \mathbb{R}$ such that the following process is a martingale w.r.t. $\mathcal{F}^+_\bullet$:
$$\big(\exp(\alpha B_t + \gamma t)\big)_{t\geq 0}.$$
Hint: *You can use the Markov Property of Brownian motion* (*Proposition 6.1*).

Proof of Proposition 6.1. Fix $s > 0$. Item 3 of Lemma 1.13 shows that the process $\widetilde{B} = (\widetilde{B}_t)_{t\geq 0}$ defined by $\widetilde{B}_t := B_{t+s} - B_s$ is a standard Brownian motion. It thus remains to prove that $\widetilde{B}$ is independent of $(B_t)_{t\in[0,s]}$ and of the sigma-algebras $\mathcal{F}^+_s$, $\mathcal{F}^a_s$, and $\mathcal{F}^B_s$. Since $(B_t)_{t\in[0,s]}$ is $\mathcal{F}^B_s$-measurable and $\mathcal{F}^B_s \subset \mathcal{F}^a_s \subset \mathcal{F}^+_s$, *it suffices to prove that* $\widetilde{B}$ *is independent of* $\mathcal{F}^+_s$. This can be done by showing that its FDDs are independent of $\mathcal{F}^+_s$ in the sense detailed in Exercise 6.5.

To this end, we fix times $0 \leq t_1 < t_2 < \cdots < t_n$ and a bounded continuous function $g : \mathbb{R}^n \to \mathbb{R}$, and to finish the proof, we aim to prove the following factorization property:

$$\mathbb{E}\big[\mathbb{1}_A\, g(\widetilde{B}_{t_1}, \widetilde{B}_{t_2}, \ldots, \widetilde{B}_{t_n})\big] = \mathbb{P}[A]\, \mathbb{E}\big[g(\widetilde{B}_{t_1}, \widetilde{B}_{t_2}, \ldots, \widetilde{B}_{t_n})\big] \qquad \text{for all } A \in \mathcal{F}^+_s. \tag{6.3}$$

Note that if we would consider $\mathcal{F}^B_s$ or $\mathcal{F}^a_s$ instead of $\mathcal{F}^+_s$, because B has independent increments by its defining property BM1, we would be done (augmenting the filtration by negligible sets does not alter independence properties).

To upgrade the independence to $\mathcal{F}^+_s$, we can use a limiting argument and the right-continuity of $\mathcal{F}^+_\bullet$. For this, fix $\varepsilon > 0$. Then (writing $B(t) = B_t$ to ease notation), we already know that

$$\begin{aligned} &\mathbb{E}\big[\mathbb{1}_A\, g(\underbrace{B(t_1+s+\varepsilon) - B(s+\varepsilon)}_{\text{indep. of } \mathcal{F}^a_{s+\varepsilon}}, \ldots, \underbrace{B(t_n+s+\varepsilon) - B(s+\varepsilon)}_{\text{indep. of } \mathcal{F}^a_{s+\varepsilon}})\big] \\ &= \mathbb{P}[A]\, \mathbb{E}\big[g(B(t_1+s+\varepsilon) - B(s+\varepsilon), \ldots, B(t_n+s+\varepsilon) - B(s+\varepsilon))\big], \end{aligned} \tag{6.4}$$

because $A \in \mathcal{F}^a_{s+\varepsilon}$ (by definition of the right-continuous filtration $\mathcal{F}^+_\bullet$). Then, we take $\varepsilon \to 0$ and use Bounded Convergence Theorem (BCT) [Kyt20, Corollary VII.21]:

$$\begin{aligned} &\mathbb{E}\big[\mathbb{1}_A\, g(\widetilde{B}(t_1), \ldots, \widetilde{B}(t_n))\big] \\ &= \mathbb{E}\big[\mathbb{1}_A\, g(B(t_1+s) - B(s), \ldots, B(t_n+s) - B(s))\big] \\ &= \mathbb{E}\big[\mathbb{1}_A \lim_{\varepsilon\to 0} g(B(t_1+s+\varepsilon) - B(s+\varepsilon), \ldots, B(t_n+s+\varepsilon) - B(s+\varepsilon))\big] && \text{[by continuity]} \\ &= \lim_{\varepsilon\to 0} \mathbb{E}\big[\mathbb{1}_A\, g(B(t_1+s+\varepsilon) - B(s+\varepsilon), \ldots, B(t_n+s+\varepsilon) - B(s+\varepsilon))\big] && \text{[by BCT]} \\ &= \mathbb{P}[A] \lim_{\varepsilon\to 0} \mathbb{E}\big[g(B(t_1+s+\varepsilon) - B(s+\varepsilon), \ldots, B(t_n+s+\varepsilon) - B(s+\varepsilon))\big] && \text{[by (6.4)]} \\ &= \mathbb{P}[A]\, \mathbb{E}\big[\lim_{\varepsilon\to 0} g(B(t_1+s+\varepsilon) - B(s+\varepsilon), \ldots, B(t_n+s+\varepsilon) - B(s+\varepsilon))\big] && \text{[by BCT]} \\ &= \mathbb{P}[A]\, \mathbb{E}\big[g(B(t_1+s) - B(s), \ldots, B(t_n+s) - B(s))\big] && \text{[by continuity]} \\ &= \mathbb{P}[A]\, \mathbb{E}\big[g(\widetilde{B}(t_1), \ldots, \widetilde{B}(t_n))\big], \end{aligned}$$

which is the sought identity (6.3). This finishes the proof of of Proposition 6.1. □

Exercise 6.5. Consider a continuous-time real-valued stochastic process $X = (X_t)_{t\geq 0}$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\mathscr{G} \subset \mathscr{F}$ be a sub-sigma-algebra. Prove that the following are equivalent.

1. For all $0 \leq t_1 < t_2 < \cdots < t_n$, for all bounded continuous functions $g : \mathbb{R}^n \to \mathbb{R}$, for all $n \in \mathbb{N}$, and for all $A \in \mathscr{G}$, we have

$$\mathbb{E}[\mathbb{1}_A \, g(X_{t_1}, X_{t_2}, \ldots, X_{t_n})] = \mathbb{P}[A] \, \mathbb{E}[g(X_{t_1}, X_{t_2}, \ldots, X_{t_n})].$$

2. X is independent of $\mathscr{G}$.

6.2 Applications of the Markov Property

A useful consequence of the Markov Property is that *germ sigma-algebra*

$$\mathcal{F}_0^+ := \bigcap_{\varepsilon>0} \mathcal{F}_\varepsilon^a$$

is trivial. This is known as Blumenthal's 0-1 law.

Proposition 6.6. (Blumenthal's 0-1 law) *Let B be a (standard) Brownian motion. Then, the* germ sigma-algebra $\mathcal{F}_0^+$ *is trivial: for any $A \in \mathcal{F}_0^+$, we have $\mathbb{P}[A] \in \{0, 1\}$.*

Proof. Consider $A \in \mathcal{F}_0^+$. We will show that A is *independent of itself*, which implies that

$$\mathbb{P}[A] = \mathbb{P}[A \cap A] = (\mathbb{P}[A])^2 \qquad \Longrightarrow \qquad \mathbb{P}[A] \in \{0, 1\}.$$

Note that by the Markov Property (Proposition 6.1), the process B — and hence its natural sigma-algebra $\mathcal{F}_\infty^B := \sigma\big(\bigcup_{t\geq 0} \mathcal{F}_t^B\big)$ — is independent of $\mathcal{F}_0^+$. Since adding negligible sets does not alter independence properties, also $\mathcal{F}_\infty^a = \sigma(\mathcal{F}_\infty^B \cup \mathscr{N})$ is independent of $\mathcal{F}_0^+$. However,

$$\mathcal{F}_0^+ := \bigcap_{\varepsilon>0} \mathcal{F}_\varepsilon^a \subset \sigma\Big(\bigcup_{t\geq 0} \mathcal{F}_t^a\Big) := \mathcal{F}_\infty^a,$$

which shows that $\mathcal{F}_0^+$ — and hence $A \in \mathcal{F}_0^+$ — is independent of itself. □

Exercise 6.7. Using the time-inversion symmetry from Item 4 of Lemma 1.13, prove that the *tail sigma-algebra*

$$\mathcal{T} := \bigcap_{t\geq 0} \sigma(\{B_s \mid s \geq t\})$$

is trivial: for any $A \in \mathcal{T}$, we have $\mathbb{P}[A] \in \{0, 1\}$.

As another interesting consequence, we see that Brownian motion oscillates infinitely often in any time interval, in the following precise sense.

Proposition 6.8. *Let B be a (standard) Brownian motion. Then, almost surely, we have*

$$\sup_{s\in(0,\varepsilon]} B_s > 0 \qquad \text{and} \qquad \inf_{s\in(0,\varepsilon]} B_s < 0 \qquad \text{for all } \varepsilon > 0. \tag{6.5}$$

In particular, B has almost surely a zero on every interval:

$$\mathbb{P}\Big[\text{ for all } \varepsilon > 0, \text{ there exists } t \in (0, \varepsilon] \text{ such that } B_t = 0\Big] = 1. \tag{6.6}$$

Proof. Consider a sequence $s_n \to 0$ as $n \to \infty$. Consider the event that $B_{s_n} > 0$ infinitely often:

$$\left\{\limsup_{s_n \to 0} \{B_{s_n} > 0\}\right\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{B_{s_k} > 0\} \in \mathcal{F}_0^+.$$

The reverse Fatou's lemma A.18 shows that

$$\mathbb{P}\left[\left\{\limsup_{s_n \to 0} \{B_{s_n} > 0\}\right\}\right] \geq \limsup_{s_n \to 0} \underbrace{\mathbb{P}[B_{s_n} > 0]}_{=1/2} \geq \frac{1}{2},$$

where $\mathbb{P}[B_{s_n} > 0] = 1/2$, as $B_{s_n} \sim N(0, s_n)$. Blumenthal's 0-1 law (Proposition 6.6) then implies

$$\mathbb{P}\left[\left\{\limsup_{s_n \to 0} \{B_{s_n} > 0\}\right\}\right] = 1.$$

Thus, almost surely, $B_{s_n} > 0$ for infinitely many $s_n \to 0$, which shows the left-hand side of (6.5). The case of the infimum (the right-hand side of (6.5)) is completely symmetric.

Lastly, since B is almost surely continuous, the mean value theorem implies (6.6). □

Corollary 6.9. *Almost surely, the zeros of B on $[0,\infty)$ accumulate to* 0.

Proof. Indeed, otherwise there would exist with a positive probability an interval $(0, \varepsilon]$ containing no zeros, which contradicts Proposition 6.8. □

Remark 6.10. With some further topological work, one can in fact show that the set of zeros of B on any interval is *uncountable.* In fact,

$$Z := \{t \in [0,1] \mid B_t = 0\}$$

has the following properties almost surely:

- ▷ it is uncountable [MP10, Theorem 2.28 and Exercise 2.9];
- ▷ it has Lebesgue measure zero [MP10, Theorem 4.24];
- ▷ it has Hausdorff dimension equal to 1/2 [MP10, Theorem 4.24];
- ▷ it has no isolated points [MP10, Theorem 2.28];

Corollary 6.11. *Almost surely, we have*

$$\sup_{s \geq 0} B_s = +\infty \qquad \text{and} \qquad \inf_{s \geq 0} B_s = -\infty.$$

Proof sketch. Using the scaling property of Brownian motion (see property 2 of Lemma 1.13), we see that the supremum process (see [LeG16, Proposition 2.14] for details why it is measurable)

$$S_t := \sup_{s \in [0,t]} B_s$$

satisfies

$$S_\infty \overset{(d)}{=} \frac{1}{\lambda} S_\infty \qquad \text{for all } \lambda > 0.$$

Hence, $S_\infty \in \{0, \pm\infty\}$ almost surely. Since we already know from Proposition 6.8 that $S_\infty > 0$ almost surely, we see that $S_\infty = +\infty$. The case of the infimum is completely symmetric. □

6.3 Strong Markov Property for Brownian motion

The *strong* Markov Property is an enhanced version of Proposition 6.1, where instead of a *fixed* time s, we take a *stopping time* (recall Definition 5.3). To discuss the Strong Markov Property rigorously, we will use the background knowledge from Section 5.3.

Proposition 6.12. (Strong Markov Property) *Let B be a (standard) Brownian motion and τ an almost surely finite stopping time with respect to $\mathcal{F}^+_\bullet$. Define*

$$\widetilde{B}_t := B_{t+\tau} - B_\tau, \qquad t \geq 0,$$

on the event $\{\tau < \infty\}$ and $\widetilde{B}_t \equiv 0$ on the event $\{\tau = \infty\}$. Then, the process $(\widetilde{B}_t)_{t\geq 0}$ is also a (standard) Brownian motion, which is independent of the sigma-algebra $\mathcal{F}^+_\tau$.

Remark 6.13. A similar conclusion also holds if we assume that $\mathbb{P}[\tau < \infty] > 0$ (not necessarily $\mathbb{P}[\tau < \infty] = 1$), under the conditional probability measure $\hat{\mathbb{P}}[\cdot] := \mathbb{P}[\cdot \,|\, \tau < \infty] > 0$, by defining

$$\widetilde{B}_t := \begin{cases} B_{t+\tau} - B_\tau, & \text{on the event } \{\tau < \infty\}, \\ 0, & \text{on the event } \{\tau = \infty\}, \end{cases} \qquad t \geq 0.$$

See [LeG16, Theorem 2.20] for more details.

Proof of Proposition 6.12. We need to show two claims for the process $\widetilde{B} = (\widetilde{B}_t)_{t\geq 0}$:

- ▷ **Claim 1.** $\widetilde{B}$ is a (standard) Brownian motion.
- ▷ **Claim 2.** $\widetilde{B}$ is independent of the sigma-algebra $\mathcal{F}^+_\tau$.

We shall aim to prove Claim 2, and then Claim 1 will follow as a by-product. As in the proof of Proposition 6.1, we will show that the FDDs of $\widetilde{B}$ are independent of $\mathcal{F}^+_\tau$ in the sense detailed in Exercise 6.5. To this end, we fix times $0 \leq t_1 < t_2 < \cdots < t_m$ and a bounded continuous function $g : \mathbb{R}^m \to \mathbb{R}$. We aim to prove the following factorization property:

$$\mathbb{E}\big[\mathbb{1}_A \, g(\widetilde{B}_{t_1}, \widetilde{B}_{t_2}, \ldots, \widetilde{B}_{t_m})\big] = \mathbb{P}[A]\, \mathbb{E}\big[g(\widetilde{B}_{t_1}, \widetilde{B}_{t_2}, \ldots, \widetilde{B}_{t_m})\big] \qquad \text{for all } A \in \mathcal{F}^+_\tau. \tag{6.7}$$

<u>**Step 0.**</u> By assumption, we have $\mathbb{P}[\tau < \infty] = 0$ and $\mathbb{P}[\tau = \infty] = 0$. As $\mathcal{F}^+_\bullet$ contains all negligible sets (5.1), we may consider events intersected with $\{\tau < \infty\}$ and retain measurability:

$$E \cap \{\tau = \infty\} \;\in\; \mathcal{F}_0 \qquad \text{for all } E \subset \Omega, \tag{6.8}$$

which also implies that $(\Omega \smallsetminus (E \cap \{\tau = \infty\})) \in \mathcal{F}_0$ and thus,

$$E \cap \{\tau < \infty\} = E \cap \underbrace{(\Omega \smallsetminus (E \cap \{\tau = \infty\}))}_{\in\, \mathcal{F}_0 \,\subset\, \mathcal{F}_t} \;\in\; \mathcal{F}_t \qquad \text{for all } E \in \mathcal{F}_t. \tag{6.9}$$

Therefore, we may replace $A \in \mathcal{F}^+_\tau$ with $\hat{A} := A \cap \{\tau < \infty\} \in \mathcal{F}^+_\tau$ in (6.7) – note that $\mathbb{P}[\hat{A}] = \mathbb{P}[A]$.

<u>**Step 1.**</u> We apply an approximation argument for the stopping time τ (see also Exercise 6.14). Define τ_n as the *smallest number of the form $k\,2^{-n}$, with some $k \in \mathbb{N}_0$, such that $\tau \leq \tau_n$*:

$$\tau_n := 2^{-n}\left(\lfloor 2^n \tau \rfloor + 1\right), \qquad n \in \mathbb{N}_0,$$

where $\lfloor x \rfloor$ denotes the largest integer less or equal to x. Then, each τ_n is a stopping time and we have[13] $\tau_n \downarrow \tau$ as $n \uparrow \infty$ almost surely. Similarly as in the proof of Proposition 6.1, we obtain

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{1}_{\hat{A}}\, g(\widetilde{B}(t_1),\ldots,\widetilde{B}(t_m))\big] \\
&= \mathbb{E}\big[\mathbb{1}_{\hat{A}}\, g(B(t_1+\tau)-B(\tau),\ldots,B(t_m+\tau)-B(\tau))\big] \\
&= \mathbb{E}\big[\mathbb{1}_{\hat{A}} \lim_{n\to\infty} g(B(t_1+\tau_n)-B(\tau_n),\ldots,B(t_m+\tau_n)-B(\tau_n))\big] && \text{[by continuity]} \\
&= \lim_{n\to\infty} \mathbb{E}\big[\mathbb{1}_{\hat{A}}\, g(B(t_1+\tau_n)-B(\tau_n),\ldots,B(t_m+\tau_n)-B(\tau_n))\big]. && \text{[by DCT]} \qquad (6.10)
\end{aligned}$$

where we also used the Dominated Convergence Theorem (DCT) [Kyt20, Theorem VII.19].

Step 2. Next, we are going apply the simple Markov Property from Proposition 6.1. To this end, we first need to replace τ_n by a finite deterministic time. Since τ_n takes *countably many values*, and we have $\hat{A} := A \cap \{\tau < \infty\} \in \mathcal{F}^+_\tau$, we can simply consider the different values separately:

$$\mathbb{1}_{\hat{A}} = \sum_{k=0}^{\infty} \mathbb{1}\{\hat{A} \cap \{\tau_n = k\,2^{-n}\}\}.$$

▷ One the one hand, by definition (5.7) we have

$$\begin{aligned}
\hat{A} \cap \{\tau_n = k\,2^{-n}\} &= \hat{A} \cap \{(k-1)\,2^{-n} < \tau_n \le k\,2^{-n}\} \\
&= \big(\underbrace{\hat{A} \cap \{\tau_n \le k\,2^{-n}\}}_{\in \mathcal{F}^+_{k\,2^{-n}}}\big) \smallsetminus \big(\underbrace{\hat{A} \cap \{\tau_n \le (k-1)\,2^{-n}\}}_{\in \mathcal{F}^+_{(k-1)\,2^{-n}} \subset \mathcal{F}^+_{k\,2^{-n}}}\big) \;\in\; \mathcal{F}^+_{k\,2^{-n}}.
\end{aligned}$$

▷ On the other hand, Proposition 6.1 shows that for any fixed $j \in \{1,2,\ldots,m\}$ and $k \in \mathbb{N}_0$,

$$B(t_j + k\,2^{-n}) - B(k\,2^{-n}) \overset{(\mathrm{d})}{=} B(t_j) \qquad \text{is independent of} \qquad \mathcal{F}^+_{k\,2^{-n}}.$$

Now, (e.g., using Fubini's theorem A.19) we see that (6.10) equals

$$\begin{aligned}
&\lim_{n\to\infty} \underbrace{\sum_{k=0}^{\infty} \mathbb{P}[\hat{A} \cap \{\tau_n = k\,2^{-n}\}]}_{=\mathbb{P}[\hat{A}]}\; \mathbb{E}\big[g(\underbrace{B(t_1+k\,2^{-n}) - B(k\,2^{-n})}_{\overset{(\mathrm{d})}{=} B(t_1)},\ldots,\underbrace{B(t_m+k\,2^{-n}) - B(k\,2^{-n})}_{\overset{(\mathrm{d})}{=} B(t_m)})\big] \\
&= \mathbb{P}[\hat{A}] \lim_{n\to\infty} \mathbb{E}\big[g(B(t_1),\ldots,B(t_m))\big] \\
&= \mathbb{P}[\hat{A}]\, \mathbb{E}\big[g(B(t_1),\ldots,B(t_m))\big].
\end{aligned}$$

This shows that

$$\mathbb{E}\big[\mathbb{1}_A\, g(\widetilde{B}(t_1),\widetilde{B}(t_2),\ldots,\widetilde{B}(t_m))\big] = \mathbb{P}[A]\,\mathbb{E}\big[g(B(t_1),\ldots,B(t_m))\big] \qquad \text{for all } A \in \mathcal{F}^+_\tau. \quad (6.11)$$

Step 3. We can now prove Claim 1 (that $\widetilde{B}$ is Brownian motion). Taking $A = \Omega$ in (6.11) yields

$$\mathbb{E}\big[g(\widetilde{B}(t_1),\widetilde{B}(t_2),\ldots,\widetilde{B}(t_m))\big] = \mathbb{E}\big[g(B(t_1),\ldots,B(t_m))\big]. \qquad (6.12)$$

Because the test function $g : \mathbb{R}^m \to \mathbb{R}$ is arbitrary, this implies by Exercise 6.15 that $\widetilde{B}$ *and* B *have the same FDDs*. Therefore, as the sample paths $\widetilde{B}$ are almost surely continuous, we may conclude that $\widetilde{B}$ is a Brownian motion.

[13] Here, the notation $\tau_n \downarrow \tau$ means that $\tau_{n+1} \le \tau_n$ for all $n\mathbb{N}$ and $\lim_{n\to\infty} \tau_n = \tau$.

Step 4. To complete the proof of Claim 2 (that $\widetilde{B}$ is independent of $\mathcal{F}_\tau^+$), it remains to observe that (6.12) shows in particular that

$$\begin{aligned}\mathbb{E}\big[\mathbb{1}_A\, g(\widetilde{B}(t_1),\widetilde{B}(t_2),\ldots,\widetilde{B}(t_m))\big] &= \mathbb{P}[A]\,\mathbb{E}\big[g(B(t_1),\ldots,B(t_m))\big] &&\text{[by (6.11)]}\\ &= \mathbb{P}[A]\,\mathbb{E}\big[g(\widetilde{B}(t_1),\widetilde{B}(t_2),\ldots,\widetilde{B}(t_m))\big] &&\text{[by (6.12)]}\end{aligned}$$

for all $A \in \mathcal{F}_\tau^+$, which proves the sought factorization property (6.7). □

Exercise 6.14. Let $\mathcal{F}_\bullet = (\mathcal{F}_t)_{t\geq 0}$ be a filtration and τ an a.s. finite stopping time w.r.t. $\mathcal{F}_\bullet$. Define

$$\tau_n := 2^{-n}\big(\lfloor 2^n\,\tau\rfloor + 1\big), \qquad n \in \mathbb{N}_0,$$

where $\lfloor x \rfloor$ denotes the largest integer less or equal to x.

1. Show that τ_n is a stopping time w.r.t. $\mathcal{F}_\bullet$.
2. Show that[a], almost surely, $\tau_n \downarrow \tau$ as $n \uparrow \infty$.

[a] Here, the notation $\tau_n \downarrow \tau$ means that $\tau_{n+1} \leq \tau_n$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} \tau_n = \tau$.

Exercise 6.15. Let μ and ν be two finite Borel measures on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. Suppose that

$$\int_{\mathbb{R}^n} g\,\mathrm{d}\mu = \int_{\mathbb{R}^n} g\,\mathrm{d}\nu \quad \text{for all bounded continuous functions } g : \mathbb{R}^n \to \mathbb{R}. \qquad (\heartsuit)$$

Prove that $\mu = \nu$.

Hint: Recall Dynkin's Identification Theorem A.29 and the fact that $\mathscr{B}(\mathbb{R}^n)$ is generated by a pi-system comprising the closed subsets. Note also that an approximation argument (using monotone convergence) shows that $(\heartsuit)$ implies $\mu[A] = \nu[A]$ for all closed subsets $A \subset \mathbb{R}^n$. These facts yield the assertion.

6.4 Reflection principle for Brownian motion

Using the Strong Markov Property (Proposition 6.12), we find the law of the maximum process

$$S_t := \sup_{s\in[0,t]} B_s, \qquad t \geq 0. \tag{6.13}$$

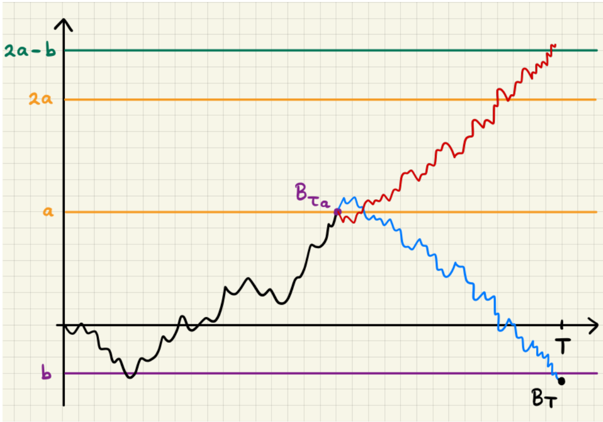
Recall that from the simple Markov Property (Proposition 6.1), we already derived that almost surely, $S_t > 0$ for any $t > 0$, and that $S_\infty = +\infty$.

Proposition 6.16. (Reflection principle) *Let B be a (standard) Brownian motion and S its maximum process* (6.13)*. Fix $a > 0$, $b \leq a$, and $T > 0$. Then, we have*

$$\mathbb{P}\,[S_T \geq a \text{ and } B_T \leq b] \;=\; \mathbb{P}\,[B_T \geq 2a - b] \;=\; \frac{1}{\sqrt{2\pi}} \int_{\frac{2a-b}{\sqrt{T}}}^{\infty} \exp\big(-\tfrac{s^2}{2}\big)\,\mathrm{d}s. \tag{6.14}$$

Proof. The idea to derive the probability is to reflect the Brownian path after it hits level a. For this purpose, we define the stopping time (recall Lemma 5.7) with respect to the filtration $\mathcal{F}_\bullet^+$:

$$\tau_a := \inf\{t \geq 0 \mid B_t = a\}.$$



From Corollary 6.9, we see that $\tau_a < \infty$ almost surely (see [LeG16, Proposition 2.14] for details). Moreover, note that

$$\{\tau_a \le T\} = \{S_T \ge a\} \qquad \text{and} \qquad \mathbb{P}[B_{\tau_a} = a] = 1.$$

Now, we consider the processes B and $\widetilde{B}$ defined as $\widetilde{B}_t := B_{t+\tau_a} - B_{\tau_a} = B_{t+\tau_a} - a$ for $t \ge 0$, appearing in the Strong Markov Property (Proposition 6.12) with $\tau = \tau_a$. Note that

$$\widetilde{B}_{T-\tau_a} = B_T - a,$$

and recall from the Strong Markov Property that $\widetilde{B}$ is independent of $\mathcal{F}^+_{\tau_a}$. Thus, we obtain

$$\begin{aligned}\mathbb{P}[S_T \ge a \text{ and } B_T \le b] \;=\; \mathbb{P}[\tau_a \le T \text{ and } B_T \le b] \;&=\; \mathbb{P}[\tau_a \le T \text{ and } B_T - a \le b - a] \\ &=\; \mathbb{P}[\tau_a \le T \text{ and } \widetilde{B}_{T-\tau_a} \le b - a].\end{aligned}$$

Now, note that by symmetry (item 1 in Lemma 1.13), we have

$$(\tau_a, \widetilde{B}) \overset{(\mathrm{d})}{=} (\tau_a, -\widetilde{B}), \qquad \Longrightarrow \qquad \mathbb{P}[\tau_a \le T \text{ and } \widetilde{B}_{T-\tau_a} \le b - a] \;=\; \mathbb{P}[\tau_a \le T \text{ and } \widetilde{B}_{T-\tau_a} \ge a - b].$$

Therefore, we conclude that

$$\begin{aligned}\mathbb{P}[S_T \ge a \text{ and } B_T \le b] \;&=\; \mathbb{P}[\tau_a \le T \text{ and } \underbrace{\widetilde{B}_{T-\tau_a}}_{= B_T - a} \ge a - b] \\ &=\; \mathbb{P}[\tau_a \le T \text{ and } B_T \ge \underbrace{2a - b}_{\ge a}] \\ &=\; \mathbb{P}[B_T \ge 2a - b]. \qquad\qquad [\text{since } \{B_T \ge 2a - b\} \subset \{\tau_a \le T\}]\end{aligned}$$

The asserted formula for $\mathbb{P}[B_T \ge 2a - b]$ follows using the Gaussian distribution (A.3). □

We can now explicitly write down the law of S.

Corollary 6.17. *For the maximum process* (6.13), *we have*

$$\mathbb{P}[S_T \ge a] \;=\; \sqrt{\tfrac{2}{\pi}} \int_{\frac{a}{\sqrt{T}}}^{\infty} \exp\big(-\tfrac{s^2}{2}\big)\,\mathrm{d}s, \qquad t > 0,\; a \ge 0.$$

Proof. We obtain the asserted formula by using the observation

$$\mathbb{P}[S_T \ge a] \;=\; \mathbb{P}[S_T \ge a \text{ and } B_T \ge a] \;+\; \underbrace{\mathbb{P}[S_T \ge a \text{ and } B_T \le a]}_{\mathbb{P}[B_T \ge 2a - a = a]} \;=\; 2\,\mathbb{P}[B_T \ge a],$$

and the Gaussian distribution (A.3) for the right-hand side. □

Exercise 6.18. Consider standard Brownian motion B. For each $y \in \mathbb{R}$, let $\tau_y := \inf\{t \geq 0 \mid B_t = y\}$, and write[a] $h(y,t) := \mathbb{P}[\tau_y > t]$. Also, let $T^- := \sup\{t < 1 \mid B_t = 0\}$ be the last time when B visits zero before time one.

1. Show that τ_y is a stopping time.
2. Is T^- a stopping time?
3. Use the Markov Property (at time u) to show that for each $u \in [0,1]$,

$$\mathbb{P}[T^- \leq u] = \int_{\mathbb{R}} p_u(0,y)\, h(y, 1-u)\, \mathrm{d}y,$$

where p is the Gaussian density

$$p_t(x,y) = p_t(x-y) = \frac{1}{\sqrt{2\pi t}} \exp\Big(- \frac{(x-y)^2}{2t} \Big), \qquad x, y, \in \mathbb{R}.$$

Hint: *You may use the reflection principle* (*Proposition 6.16*).

[a] You do not need to calculate $h(y;t)$ explicitly in this problem.

Exercise 6.19. (Wald's lemma) Consider standard Brownian motion B. Let τ be any stopping time such that $\mathbb{E}[\tau] < \infty$ and $\mathbb{E}[\tau^4] < \infty$[a]. Define an increasing sequence $(\tau_n)_{n \in \mathbb{N}}$ of stopping times recursively by

$$\tau_1 := \tau \qquad \text{and} \qquad \tau_n := \tilde{\tau}^{(n)} + \tau_{n-1} \quad \text{for } n \geq 2,$$

where $\tilde{\tau}^{(n)}$ is the same function as τ but associated to Brownian motion $(\tilde{B}^{(n)}(t))_{t \geq 0}$ defined via the Strong Markov Property at time τ_{n-1} as

$$\tilde{B}^{(n)}(t) := \tilde{B}(t + \tau_{n-1}) - \tilde{B}(\tau_{n-1}).$$

1. Prove the *Law of Large Numbers* for Brownian motion: almost surely, we have

$$\lim_{t \to \infty} \frac{B(t)}{t} = 0.$$

Hint: *Recall the time-inversion symmetry from item 4 of Lemma 1.13.*

2. Show that, almost surely, we have

$$\lim_{n \to \infty} \frac{B(\tau_n)}{n} = 0.$$

Hint: *Use the Law of Large Numbers* (*Theorem A.27*) *on the one hand for usual i.i.d sequences and on the other hand for B.*

3. Show that $\mathbb{E}[B(\tau)] < \infty$. *Hint*: *You can use, e.g., the Second Borel-Cantelli Lemma A.14. Recall also that* $\mathbb{E}[|\xi|] = \int_0^\infty \mathbb{P}[\xi \geq t]\, \mathrm{d}t$ *for a non-negative random variable ξ.*

4. Show that, almost surely, we have

$$\lim_{n \to \infty} \frac{B(\tau_n)}{n} = \mathbb{E}[B(\tau)].$$

5. Conclude that $\mathbb{E}[B(\tau)] = 0$.

[a] In fact, since the general Kolmogorov's Law of Large Numbers for i.i.d. random variables holds without the assumption $\mathbb{E}[\tau^4] < \infty$, we could just assume $\mathbb{E}[\tau] < \infty$ here.

7 Continuous continuous-time martingales

In this section, we summarize some important properties of *continuous-time martingales*, which will be needed in Sections 8–9. Many of them are analogous to the discrete-time case. The measurability issues arising from the fact that the index set $[0, \infty)$ is uncountable we can overcome, for instance, by assuming *continuity* of the sample paths. Throughout, we will consider a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ satisfying the *usual conditions* (Definition 5.6).

Definition 7.1. (Martingale) Stochastic process $M = (M_t)_{t\geq 0}$ on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is called a *martingale* if

MG1. (it is adapted to $\mathcal{F}_\bullet$): for every $t \geq 0$, the random variable M_t is $\mathcal{F}_t$-measurable;

MG2. (it is integrable): for every $t \geq 0$, the random variable $M_t \in L^1(\mathbb{P})$ is integrable;

MG3. (it has the martingale property): almost surely, we have

$$\mathbb{E}[M_t \mid \mathcal{F}_s] = M_s \qquad \text{for every } 0 \leq s < t. \tag{MG3}$$

If instead $\mathbb{E}[M_t \mid \mathcal{F}_s] \leq M_s$ or $\mathbb{E}[M_t \mid \mathcal{F}_s] \geq M_s$ in property MG3, M is said to be a *supermartingale* or a *submartingale*, respectively.

The set of all *continuous martingales* on $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is denoted as

$$\mathcal{M}_c := \{M = (M_t)_{t\geq 0} \mid M \text{ is a martingale and } t \mapsto M_t(\omega) \text{ is almost surely continuous}\}.$$

Exercise 7.2. Show that $\mathcal{M}_c$ is a real vector space.

7.1 Optional Stopping Theorems

The assumption that M is *continuous* is used to guarantee that M is progressively measurable. Then, the stopped process is also well-defined, continuous, and adapted (cf. Corollary 5.21).

Theorem 7.3. (Optional Stopping Theorem – Stopping Time Characterization of mgles) *Consider a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ satisfying the usual conditions.*
Let $M = (M_t)_{t\geq 0}$ be a continuous process satisfying properties MG1 (*adapted to* $\mathcal{F}_\bullet$) *and* MG2 (*integrable*). *Then, the following are equivalent*:

1. *M is a martingale*: $M \in \mathcal{M}_c$.
2. *For any two bounded stopping times σ and τ, we have $\mathbb{E}[M_\tau \mid \mathcal{F}_\sigma] = M_{\tau\wedge\sigma}$ a.s.*
3. *For any bounded stopping time τ, we have $\mathbb{E}[M_\tau] = M_0$.*
4. *For any bounded stopping time τ, the process $M^\tau := (M_{t\wedge\tau})_{t\geq 0}$ is a martingale.*

Proof sketch. We use approximation arguments and the discrete-time results from Section 3.4.

(1 ⇒ 2) For any two bounded stopping times $\sigma, \tau \le C$, define τ_n and σ_n as in Exercise 6.14:

$$\tau_n := 2^{-n}\big(\lfloor 2^n \tau \rfloor + 1\big) \qquad \text{and} \qquad \sigma_n := 2^{-n}\big(\lfloor 2^n \sigma \rfloor + 1\big), \qquad n \in \mathbb{N}_0,$$

so that $\tau_n \downarrow \tau$ and $\sigma_n \downarrow \sigma$ as $n \uparrow \infty$ almost surely. For any fixed $n \in \mathbb{N}$, the process $(M_{k2^{-n}})_{k\in\mathbb{N}_0}$ is a discrete-time martingale and τ_n and σ_n are stopping times with respect to the discrete-time filtration $(\mathcal{F}_{k2^{-n}})_{k\in\mathbb{N}_0}$. Hence, from Lemma 3.26 we obtain

$$\mathbb{E}\left[M_{\tau_n} \,|\, \mathcal{F}_{\sigma_n}\right] = M_{\tau_n \wedge \sigma_n} \qquad \text{almost surely for all } n \in \mathbb{N}_0.$$

Similarly, since $\tau, \sigma \le C$, we have $\tau_n, \sigma_n \le 2^{-n}\big(\lfloor 2^n C \rfloor + 1\big) =: C_n \le C_0$, and thus, Lemma 3.26 also shows that

$$M_{\tau_n} = \mathbb{E}\left[M_{C_0} \,|\, \mathcal{F}_{\tau_n}\right], \quad M_{\sigma_n} = \mathbb{E}\left[M_{C_0} \,|\, \mathcal{F}_{\sigma_n}\right], \quad \text{and} \quad M_{\tau_n \wedge \sigma_n} = \mathbb{E}\left[M_{C_0} \,|\, \mathcal{F}_{\tau_n \wedge \sigma_n}\right]$$

almost surely for all $n \in \mathbb{N}_0$, and Lemma 4.14 then implies that the collections $(M_{\tau_n})_{n\in\mathbb{N}_0}$ and $(M_{\sigma_n})_{n\in\mathbb{N}_0}$ and $(M_{\tau_n \wedge \sigma_n})_{n\in\mathbb{N}_0}$ are Uniformly Integrable. By continuity, we have

$$M_{\tau_n} \to M_\tau, \qquad M_{\sigma_n} \to M_\sigma, \qquad \text{and} \qquad M_{\tau_n \wedge \sigma_n} \to M_{\tau \wedge \sigma}$$

almost surely as $n \to \infty$, and by the uniform integrability, Proposition 4.12, and Exercise A.21, the convergence also holds in L^1. In particular, all M_τ, M_σ, and $M_{\tau\wedge\sigma}$ are integrable — so $M_{\tau\wedge\sigma}$ satisfies property CE1 of the conditional expected value in Definition 2.1. We leave it as an exercise to verify the measurability property CE2 for $M_{\tau\wedge\sigma}$. To verify property CE3, note that if $A \in \mathcal{F}_\sigma \subset \mathcal{F}_{\sigma_n}$, then by property CE3 applied in the discrete case, we have

$$\mathbb{E}\left[M_{\tau_n \wedge \sigma_n} \mathbb{1}_A\right] = \mathbb{E}\left[\mathbb{E}\left[M_{\tau_n} \,|\, \mathcal{F}_{\sigma_n}\right] \mathbb{1}_A\right] = \mathbb{E}\left[M_{\tau_n} \mathbb{1}_A\right] \qquad \text{almost surely for all } n \in \mathbb{N}_0.$$

By the convergence in L^1, taking $n \to \infty$ we obtain

$$\mathbb{E}\left[M_{\tau\wedge\sigma} \mathbb{1}_A\right] = \mathbb{E}\left[M_\tau \mathbb{1}_A\right],$$

which shows by property CE3 that $\mathbb{E}\left[M_\tau \,|\, \mathcal{F}_\sigma\right] = M_{\tau\wedge\sigma}$ almost surely.

(1 ⇒ 3) This follows from the above by taking $\sigma \equiv 0$.

(1 ⇒ 4) This is proved, e.g., in [LeG16, Corollary 3.24].

(3 ⇒ 1) This is very similar as Proposition 3.22, and we leave its details as an exercise. □

In most applications, the stopping times of interest are not bounded. Thus, it is desirable to have a stronger version of Optional Stopping. Analogously to the discrete case, *uniform integrability* (UI, Definition 4.7) is a useful condition to guarantee sufficient control of M.

Definition 7.4. Stochastic process $X = (X_t)_{t\ge0}$ is *Uniformly Integrable* (UI) if

$$\lim_{R\to\infty} \sup_{t\ge0} \mathbb{E}\Big[\,|X_t|\, \mathbb{1}\{|X_t| \ge R\}\Big] = 0.$$

Recall also the various conditions for UI from Exercises 4.8–4.11.

The continuous-time analogue of the Martingale Reconstruction Theorem 4.13 is the following result, that is usually referred to as the "Optional Stopping Theorem for UI martingales."

Theorem 7.5. (Optional Stopping and Martingale Convergence Theorem for UI mgles) *Consider a filtered probability space* $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ *satisfying the usual conditions.*

Let $M \in \mathcal{M}_c$ *be a continuous* Uniformly Integrable (UI) *martingale. Then, there exists a random variable* $M_\infty \in L^1(\mathbb{P})$ *such that*

- ▷ *we have* $\lim_{t\to\infty} M_t = M_\infty$ *almost surely*;
- ▷ *we have* $M_t \xrightarrow{L^1} M_\infty$*, that is,*

$$\lim_{t\to\infty} \mathbb{E}[|M_t - M_\infty|] = 0.$$

Moreover, in this case, if σ *and* τ *are two stopping times such that* $\mathbb{P}[\sigma \le \tau] = 1$*, then*

$$\mathbb{E}[M_\tau \mid \mathcal{F}_\sigma] = M_\sigma \qquad \text{almost surely.} \tag{OST}$$

In particular, for any stopping time τ*, we have*

$$\mathbb{E}[M_\tau] = M_0 \qquad \text{and} \qquad \mathbb{E}[M_\infty \mid \mathcal{F}_\tau] = M_\tau$$

Note that (OST) in particular shows the Martingale Reconstruction Property

$$M_t = \mathbb{E}[M_\infty \mid \mathcal{F}_t], \qquad t \ge 0.$$

Proof. [LeG16, Theorems 3.19 and 3.21] shows the convergence via a similar argument as in the discrete case, based on controlling upcrossings. [LeG16, Theorem 3.22] shows (OST). □

There is also a version of Theorem 7.5 for *uniformly* L^2*-bounded continuous martingales*, which we discuss in the next Section 7.2.1.

7.2 The space of uniformly L^2-bounded continuous martingales

The set of all *uniformly* L^2*-bounded continuous martingales* on $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is denoted as

$$\mathcal{M}_c^2 := \{M = (M_t)_{t\ge0} \in \mathcal{M}_c \mid M \text{ is uniformly } L^2\text{-bounded: } \sup_{t\ge0} \mathbb{E}[M_t^2] < \infty\}.$$

- ▷ We will by convention identify elements of $\mathcal{M}_c^2$ up to indistinguishability (cf. Definition 1.20).
- ▷ $\mathcal{M}_c^2$ is a vector subspace of the space $\mathcal{M}_c$ of all continuous martingales (cf. Exercises 7.2 & 7.6).
- ▷ Crucially, we will prove in Theorems 7.16 & 7.17 that $\mathcal{M}_c^2$ is also a *Hilbert space*, analogously to $L^2(\mathbb{P})$ (recall Section 2.4, especially Theorem 2.20). The *completeness* of $\mathcal{M}_c^2$ will be a key tool for the construction of stochastic integrals in Sections 8–9.

Exercise 7.6. Show that $\mathcal{M}_c^2$ is a real vector space.

Remark 7.7. Note that the property

$$\sup_{t\ge0} \mathbb{E}[M_t^2] < \infty \tag{7.1}$$

of being *uniformly* L^2-bounded is stronger than square-integrability: a process $X = (X_t)_{t\ge0}$ is called *square-integrable* if $\mathbb{E}[X_t^2] < \infty$ for all $t \ge 0$. For instance, Brownian motion B is square-integrable, because $\mathbb{E}[B_t^2] = t < \infty$ for all $t \ge 0$, but it is not uniformly L^2-bounded, since $\mathbb{E}[B_t^2] = t \to \infty$ as $t \to \infty$. Moreover, we know from Corollary 6.11 that B_t has no limit as $t \to \infty$.

7.2.1 Optional Stopping and Martingale Convergence Theorem in $\mathcal{M}_c^2$

Theorem 7.8. (Optional Stopping and Martingale Convergence Theorem for L^2 mgles) *Consider a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ satisfying the usual conditions. Let $M \in \mathcal{M}_c^2$. Then, there exists a random variable $M_\infty \in L^2(\mathbb{P})$ such that*

- ▷ *we have* $\lim_{t\to\infty} M_t = M_\infty$ *almost surely;*
- ▷ *we have* $M_t \xrightarrow{L^2} M_\infty$, *that is,*

$$\lim_{t\to\infty} \mathbb{E}[|M_t - M_\infty|^2] = 0;$$

- ▷ *we have*

$$\tfrac{1}{2}\Big(\mathbb{E}\Big[\sup_{t\geq 0} M_t^2\Big]\Big)^{1/2} \leq \|M_\infty\|_{L^2} \leq \sup_{t\geq 0}\|M_t\|_{L^2} < \infty. \tag{7.2}$$

Moreover, in this case, if σ and τ are two stopping times such that $\mathbb{P}[\sigma \leq \tau] = 1$, then

$$\mathbb{E}[M_\tau \,|\, \mathcal{F}_\sigma] = M_\sigma \qquad \text{almost surely.} \tag{OST}$$

In particular, for any stopping time τ, we have

$$\mathbb{E}[M_\tau] = M_0 \qquad \text{and} \qquad \mathbb{E}[M_\infty \,|\, \mathcal{F}_\tau] = M_\tau \tag{7.3}$$

Proof. Recall from Exercise 4.11 that a uniformly $L^2(\mathbb{P})$-bounded martingale is UI. Hence, it readily follows from Theorem 7.5 that M_t converges to M_∞ almost surely and in $L^1(\mathbb{P})$, and that (OST) and (7.3) hold for the limit. It remains to verify that the convergence $M_t \to M_\infty$ as $t \to \infty$ also takes place in $L^2(\mathbb{P})$ and that the inequalities (7.2) hold for the limit. For this, we use a continuous-time analogue of *Doob's L^2-maximal inequality*, stated below in Proposition 7.14 (it can be derived from the corresponding Proposition 4.21 & Corollary 4.23 via discretization).

▷ Note first that since $M \in \mathcal{M}_c^2$, using the triangle inequality we find that for any $T \geq 0$,

$$\mathbb{E}[|M_T - M_\infty|^2] \leq \mathbb{E}\Big[\Big(2\sup_{t\geq 0}|M_t|\Big)^2\Big] = 4\,\mathbb{E}\Big[\sup_{t\geq 0} M_t^2\Big],$$

and by Doob's L^2-maximal inequality (7.8), we have

$$\mathbb{E}\Big[\sup_{t\geq 0} M_t^2\Big] \leq 4\sup_{s\geq 0}\mathbb{E}[M_s^2] < \infty,$$

which implies that $M - M_\infty \in \mathcal{M}_c^2$. Moreover, applying Doob's L^2-maximal inequality (7.8) to the continuous martingale $t \mapsto M_t - M_\infty$, we obtain

$$\mathbb{E}\Big[\sup_{t\geq 0}|M_t - M_\infty|^2\Big] \leq 4\sup_{T\geq 0}\mathbb{E}[|M_T - M_\infty|^2] < \infty,$$

so the Dominated Convergence Theorem (DCT) [Kyt20, Theorem VII.19] gives

$$\lim_{s\to\infty}\mathbb{E}[|M_s - M_\infty|^2] = \mathbb{E}\Big[\underbrace{\lim_{s\to\infty}|M_s - M_\infty|^2}_{=0}\Big] = 0$$

by the almost sure convergence $M_t \to M_\infty$. This shows that $M_t \xrightarrow{L^2} M_\infty$.

▷ The second inequality in (7.2) follows from the L^2-convergence: we have

$$\| M_\infty \|_{L^2} = \lim_{T\to\infty} \| M_T \|_{L^2} \le \sup_{t\ge 0} \| M_t \|_{L^2}.$$

▷ Lastly, the first inequality in (7.2) follows from the L^2-convergence together with Doob's L^2-maximal inequality (7.7) in Proposition 7.14:

$$\mathbb{E}\Big[\sup_{0\le t\le T} M_t^2\Big] \le 4\| M_T \|_{L^2}^2 \xrightarrow{L^2} 4\| M_\infty \|_{L^2}^2 < \infty,,$$

so again, Dominated Convergence Theorem [Kyt20, Theorem VII.19] shows that

$$\mathbb{E}\Big[\sup_{t\ge 0} M_t^2\Big] \le 4\| M_\infty \|_{L^2}^2,$$

and the first inequality in (7.2) follows by taking square root. □

7.2.2 $\mathcal{M}_c^2$ as a normed vector space

To study the space $\mathcal{M}_c^2$, we will use the L^2-Martingale Convergence Theorem 7.8 throughout: any $M \in \mathcal{M}_c^2$ converges in the long run to a random variable $M_\infty \in L^2(\mathbb{P})$ (a.s. and in L^2):

$$\mathbb{P}\Big[\lim_{t\to\infty} M_t = M_\infty\Big] = 1 \qquad \text{and} \qquad M_t \xrightarrow{L^2} M_\infty.$$

In particular, to each $M \in \mathcal{M}_c^2$ we will always associate its final value M_∞.

Lemma 7.9. *The space $\mathcal{M}_c^2$ has the following equivalent norms: for $M \in \mathcal{M}_c^2$,*

$$\| M \|_{\mathcal{M}_c^2} := \big(\mathbb{E}[M_\infty^2]\big)^{1/2} = \| M_\infty \|_{L^2},$$
$$|||M|||_{\sup} := \Big(\mathbb{E}\Big[\big(\sup_{t\ge 0} | M_t |\big)^2\Big]\Big)^{1/2} = \Big\| \sup_{t\ge 0} | M_t | \Big\|_{L^2}.$$

In particular, we have

$$\| M \|_{\mathcal{M}_c^2} \le |||M|||_{\sup} \le 2\| M \|_{\mathcal{M}_c^2}. \tag{7.4}$$

For the proof of this result, we use continuous-time analogues of *Doob's maximal inequalities* (check out Section 4.4). We state the relevant inequalities in Proposition 7.12 and Proposition 7.14 below (their proofs are left as exercises for a motivated reader).

Remark 7.10. Recall that elements in $\mathcal{M}_c^2$ are identified up to *indistinguishability* (cf. Definition 1.20). The Martingale Reconstruction Property (7.3) from Theorem 7.8 (or alternatively, Lemma 7.9) shows that this is equivalent to defining $\mathcal{M}_c^2$ as the *quotient space* up to the equivalence relation induced by identifying elements with the same value of the norm $\| \cdot \|_{\mathcal{M}_c^2}$:

$$\begin{aligned} M = 0 \in \mathcal{M}_c^2 \quad &\Longleftrightarrow \quad \| M \|_{\mathcal{M}_c^2} = 0 \\ &\Longleftrightarrow \quad \mathbb{E}[M_\infty^2] = 0 \\ &\Longleftrightarrow \quad M_\infty = 0 \ \ \mathbb{P}\text{-a.s.} \\ &\Longleftrightarrow \quad \mathbb{P}[M_t = 0 \text{ for all } t \ge 0] = 1. \end{aligned}$$

Proof of Lemma 7.9. It is straightforward to check that $\|\cdot\|_{\mathcal{M}_c^2}$ and $|||\cdot|||_{\sup}$ are norms on $\mathcal{M}_c^2$. We obtain the right-hand side of the claimed (7.4) from (7.2) in Theorem 7.8:

$$\begin{aligned}|||M|||_{\sup}^2 := \mathbb{E}\Big[\big(\sup_{t\geq 0}|M_t|\big)^2\Big] &= \mathbb{E}\Big[\sup_{t\geq 0} M_t^2\Big] \\ &\leq 4\|M_\infty\|_{L^2}^2 =: 4\|M\|_{\mathcal{M}_c^2}^2. \qquad \text{[by (7.2)]}\end{aligned}$$

To derive the left-hand side of (7.4), note that the L^2-convergence in Theorem 7.8 gives

$$\begin{aligned}\|M\|_{\mathcal{M}_c^2} &:= \|M_\infty\|_{L^2} \\ &\leq \sup_{t\geq 0}\|M_t\|_{L^2} = \sup_{t\geq 0}\sqrt{\mathbb{E}[M_t^2]} \qquad \text{[by (7.2)]} \\ &= \Big(\sup_{t\geq 0}\mathbb{E}[M_t^2]\Big)^{1/2} \qquad \text{[since } x\mapsto\sqrt{x} \text{ is increasing on } [0,\infty)] \\ &\leq \Big(\mathbb{E}\Big[\big(\sup_{t\geq 0}|M_t|\big)^2\Big]\Big)^{1/2} \qquad \Big[\text{as } M_T^2 \leq \sup_{t\geq 0} M_t^2 = \big(\sup_{t\geq 0}|M_t|\big)^2 \text{ for all } T\geq 0\Big] \\ &=: |||M|||_{\sup},\end{aligned}$$

which gives the left-hand side of the claimed (7.4) and concludes the proof. □

The next exercise is a sanity check that ensures that the middle quantity in Doob's maximal inequality (7.5) in Proposition 7.12 is well-defined.

Exercise 7.11. Prove that, for a continuous martingale or a non-negative continuous submartingale M on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ satisfying the usual conditions, the set

$$\Big\{\omega\in\Omega \,\Big|\, \sup_{0\leq t\leq T}|M_t(\omega)|\geq\lambda\Big\}, \qquad T>0,\ \lambda>0,$$

is an event, that is, it belongs to the sigma-algebra $\mathscr{F}$. *Hint*: *Use a countable dense subset of* $[0,T]$.

Proposition 7.12. (Doob's maximal inequality) *Let M be a continuous martingale or a non-negative continuous submartingale. Then, we have*

$$\begin{aligned}\mathbb{P}\Big[\sup_{0\leq t\leq T}|M_t|\geq\lambda\Big] &\leq \tfrac{1}{\lambda}\mathbb{E}\Big[|M_T|\,\mathbb{1}\Big\{\sup_{0\leq t\leq T}|M_t|\geq\lambda\Big\}\Big] \\ &\leq \tfrac{1}{\lambda}\mathbb{E}[|M_T|], \qquad T>0,\ \lambda>0.\end{aligned} \tag{7.5}$$

Moreover, we have

$$\mathbb{P}\Big[\sup_{t\geq 0}|M_t|\geq\lambda\Big] \leq \frac{1}{\lambda}\sup_{T\geq 0}\mathbb{E}[|M_T|], \qquad \lambda>0. \tag{7.6}$$

Exercise 7.13. Prove Proposition 7.12. *Hint*:

▷ *Discretize time, use Doob's discrete-time maximal inequality* (*Proposition 4.18*), *and pass to the limit with ever finer discretizations to obtain* (7.5).

▷ *Use Monotone Convergence Theorem to get* (7.6).

Proposition 7.14. (Doob's L^2-maximal inequality) *Let M be a continuous martingale or a non-negative continuous submartingale. Then, we have*

$$\mathbb{E}\Big[\sup_{0\le t\le T} M_t^2\Big] \le 4\,\mathbb{E}[M_T^2], \qquad T>0. \tag{7.7}$$

Moreover, we have

$$\mathbb{E}\Big[\sup_{t\ge 0} M_t^2\Big] \le 4\sup_{T\ge 0}\mathbb{E}[M_T^2]. \tag{7.8}$$

Exercise 7.15. Prove Proposition 7.14. *Hint*:

▷ *Discretize time, use Doob's discrete-time L^2-inequality* (*Proposition 4.21*), *and pass to the limit with ever finer discretizations to obtain* (7.7).

▷ *Use Monotone Convergence Theorem to get* (7.8).

7.2.3 Completeness of $\mathcal{M}_c^2$ ($\star$)

Consider a *Cauchy-sequence* $(M^{(n)})_{n\in\mathbb{N}}$ in $(\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$:

$$\lim_{m\to\infty}\sup_{n,n'\ge m}\|M^{(n)}-M^{(n')}\|_{\mathcal{M}_c^2} = 0. \tag{7.9}$$

We will show that there exists $M\in\mathcal{M}_c^2$ such that

$$\lim_{n\to\infty}\|M^{(n)}-M\|_{\mathcal{M}_c^2} = 0,$$

that is, $(M^{(n)})_{n\in\mathbb{N}}$ converges in $(\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$. This yields the completeness of $(\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$:

Theorem 7.16. *The space $(\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$ is complete.*

Proof sketch. Consider a Cauchy-sequence $(M^{(n)})_{n\in\mathbb{N}}$ in $(\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$. With the L^2-Martingale Convergence Theorem 7.8, it is reasonable to expect that the limit M could be constructed as

$$M_t := \mathbb{E}[M_\infty\,|\,\mathcal{F}_t], \qquad t\ge 0, \tag{7.10}$$

where

(Cvg1): for each fixed $n\in\mathbb{N}$, we have $M_t^{(n)}\to M_\infty^{(n)}$ as $t\to\infty$ a.s. and in $L^2(\mathbb{P})$ (by Theorem 7.8);

(Cvg2): $M_\infty^{(n)}\xrightarrow{L^2} M_\infty\in L^2(\mathbb{P})$ as $n\to\infty$ by completeness of $L^2(\mathbb{P})$ (Theorem 2.20): indeed, $(M_\infty^{(n)})_{n\in\mathbb{N}}$ is a Cauchy-sequence in $(L^2(\mathbb{P}), \|\cdot\|_{L^2})$ since $(M^{(n)})_{n\in\mathbb{N}}$ is a Cauchy-sequence in $(\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$ — directly by definition of the norm $\|\cdot\|_{\mathcal{M}_c^2}$ (cf. Lemma 7.9).

It remains to be verified that

▷ the martingale M given by formula (7.10) belongs to the space

$$\mathcal{M}_c^2 = \Big\{M=(M_t)_{t\ge0}\in\mathcal{M}_c \;\Big|\; M \text{ is uniformly } L^2\text{-bounded: } \sup_{t\ge0}\|M_t\|_{L^2}<\infty\Big\}$$

of uniformly L^2-bounded continuous martingales; and

▷ the sequence $(M^{(n)})_{n\in\mathbb{N}}$ indeed converges to M in $(\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$.

Step 1. We first verify that M is uniformly L^2-bounded: $\sup_{t\ge0} \| M_t \|_{L^2} < \infty$.

Recall from Proposition 2.7 that the conditional expected value of a square-integrable random variable $\xi \in L^2(\mathbb{P})$ given $\mathcal{F}_t$ is constructed as the *orthogonal projection* of ξ onto $\mathrm{m}\mathcal{F}_t \cap L^2(\mathbb{P})$. In particular, for each $t \ge 0$, the map

$$\xi \;\mapsto\; \mathbb{E}\,[\,\xi\,|\,\mathcal{F}_t]$$

is a *contraction* $L^2(\mathbb{P}) \to L^2(\mathbb{P})$:

$$\|\,\mathbb{E}\,[\,\xi\,|\,\mathcal{F}_t]\,\|_{L^2} \;\le\; \|\,\xi\,\|_{L^2}, \qquad \xi \in L^2(\mathbb{P}),\; t \ge 0.$$

Taking $\xi = M_\infty$ as in (7.10) gives

$$\sup_{t\ge0} \|\,M_t\,\|_{L^2} \;=\; \sup_{t\ge0} \|\,\mathbb{E}\,[M_\infty\,|\,\mathcal{F}_t]\,\|_{L^2} \;\le\; \|\,M_\infty\,\|_{L^2} \;<\; \infty.$$

Step 2. We then show that the above two limits (Cvg1) and (Cvg2) commute:

$$M_t^{(n)} \quad \xrightarrow{L^2} \quad M_t \;\in\; L^2(\mathbb{P}) \qquad \text{for each fixed } t \ge 0.$$

Indeed, from definition (7.10), we have

$$\begin{aligned}
M_t \;:=\; \mathbb{E}\,[M_\infty\,|\,\mathcal{F}_t] \;&=\; \mathbb{E}\Big[\lim_{\substack{n\to\infty\\ \text{in } L^2}} M_\infty^{(n)}\,|\,\mathcal{F}_t\Big] && \text{[by (Cvg2)]}\\
&=\; \lim_{\substack{n\to\infty\\ \text{in } L^2}} \mathbb{E}\Big[M_\infty^{(n)}\,|\,\mathcal{F}_t\Big] && \text{[since orthogonal projection is continuous in } L^2(\mathbb{P})\text{]}\\
&=\; \lim_{\substack{n\to\infty\\ \text{in } L^2}} M_t^{(n)}, && \text{[by (7.11) below]}
\end{aligned}$$

since *Martingale Reconstruction Property* (7.3) from Theorem 7.8 gives

$$M_t^{(n)} = \mathbb{E}\,[M_\infty^{(n)}\,|\,\mathcal{F}_t], \qquad t \ge 0,\; n \in \mathbb{N}. \tag{7.11}$$

Step 3. We lastly prove that $t \mapsto M_t$ is a.s. continuous (i.e., has a continuous modification).

Property (7.9) implies that we can find indices $1 \le m_1 < m_2 < \cdots$ such that

$$\|\,M^{(n)} - M^{(n')}\,\|_{\mathcal{M}_c^2} \le 2^{-j} \qquad \text{for all } n, n' \ge m_j,\; j \in \mathbb{N}.$$

Then, using the other norm $|||\cdot|||_{\sup}$ from Lemma 7.9, we can estimate

$$\begin{aligned}
&\sum_{j=1}^{\infty} \mathbb{E}\Big[\sup_{t\ge0} \big|\,M_t^{(m_{j+1})} - M_t^{(m_j)}\,\big|\Big]\\
\le\;& \sum_{j=1}^{\infty} \Big\|\sup_{t\ge0} \big|\,M_t^{(m_{j+1})} - M_t^{(m_j)}\,\big|\Big\|_{L^2} && \text{[by Cauchy-Schwarz (Lemma 2.17)]}\\
=\;& \sum_{j=1}^{\infty} |||\,M^{(m_{j+1})} - M^{(m_j)}\,|||_{\sup} && \text{[by definition of } |||\cdot|||_{\sup}\text{]}\\
\le\;& 2\sum_{j=1}^{\infty} \|\,M^{(m_{j+1})} - M^{(m_j)}\,\|_{\mathcal{M}_c^2} && \text{[by Lemma 7.9]}\\
\le\;& 2\sum_{j=1}^{\infty} 2^{-j} \;<\; \infty.
\end{aligned}$$

Lemma 2.21 now shows that

$$\sup_{t\geq 0}\big|M_t^{(m_{j+1})}-M_t^{(m_j)}\big| \quad \overset{j\to\infty}{\longrightarrow} \quad 0 \qquad \text{almost surely.}$$

This shows that $M^{(m_j)}$ has almost surely[14] a limit in the space $C([0,\infty),\mathbb{R})$ uniformly on compacts, and this limit is a continuous function:

$$M^{(m_j)} \overset{j\to\infty}{\longrightarrow} \widetilde{M} \in C([0,\infty),\mathbb{R}).$$

On the other hand, we also know from Step 2 that the pointwise-in-time convergence

$$M_t^{(m_j)} \quad \underset{j\to\infty}{\overset{L^2}{\longrightarrow}} \quad M_t \in L^2(\mathbb{P}) \qquad \text{for each fixed } t\geq 0$$

holds in $L^2(\mathbb{P})$, so it follows that $M=\widetilde{M}$ (up to indistinguishability). In particular, the limit M is a.s. continuous (i.e., has a continuous modification — recall Definition 1.17). ☐

Theorem 7.17. *The space* $\mathcal{M}_c^2$ *is a Hilbert space with inner product*

$$\langle M,N\rangle_{\mathcal{M}_c^2} := \mathbb{E}[M_\infty N_\infty], \qquad M,N\in\mathcal{M}_c^2.$$

Proof. It is straightforward to check that $\langle\cdot,\cdot\rangle_{\mathcal{M}_c^2}$ defines an inner product. Note also that

$$\|M\|_{\mathcal{M}_c^2} := \big(\mathbb{E}[M_\infty^2]\big)^{1/2} = \big(\langle M,M\rangle_{\mathcal{M}_c^2}\big)^{1/2}.$$

The completeness of $(\mathcal{M}_c^2,\|\cdot\|_{\mathcal{M}_c^2})$ follows from Theorem 7.16. ☐

7.2.4 Toolbox: Sophomore's dream trick – useful identities for squares ($\star$)

Let us record here an identity which is used frequently in stochastic analysis. We give a special case of it in Lemma 7.18 and a more general version in Corollary 7.20. It applies to square-integrable martingales — in particular, martingales in $\mathcal{M}_c^2$.

Lemma 7.18. (Sophomore's dream trick) *Let M be a square-integrable martingale:* $\mathbb{E}[M_t^2]<\infty$ *for all $t\geq 0$. Then, for any $0\leq s\leq t$, we have*

$$\mathbb{E}[(M_t-M_s)^2\,|\,\mathcal{F}_s] = \mathbb{E}[M_t^2\,|\,\mathcal{F}_s]-M_s^2 = \mathbb{E}[M_t^2-M_s^2\,|\,\mathcal{F}_s]. \tag{7.12}$$

If furthermore $M\in\mathcal{M}_c^2$, then for any $t\geq 0$, we have

$$\mathbb{E}[(M_\infty-M_t)^2] = \mathbb{E}[M_\infty^2]-\mathbb{E}[M_t^2], \tag{7.13}$$

where $M_\infty=\lim_{t\to\infty}M_t\in L^2(\mathbb{P})$ *is given by Theorem 7.8.*

[14] Note that since the filtration $\mathcal{F}_\bullet$ is assumed to be complete, by considering processes up to indistinguishability we can ignore the event of probability zero where the uniform convergence in $C([0,\infty),\mathbb{R})$ does not hold.

Proof. By linearity (item 3 of Lemma 2.8), we have

$$\mathbb{E}\big[(M_t - M_s)^2 \,|\, \mathcal{F}_s\big] \;=\; \mathbb{E}\big[M_t^2 \,|\, \mathcal{F}_s\big] \;-\; 2\,\mathbb{E}\big[M_t\, M_s \,|\, \mathcal{F}_s\big] \;+\; \mathbb{E}\big[M_s^2 \,|\, \mathcal{F}_s\big].$$

By item 5 of Lemma 2.8 and the martingale property MG3 of M, the middle term equals

$$-2\, M_s \underbrace{\mathbb{E}\big[M_t \,|\, \mathcal{F}_s\big]}_{= M_s} \;=\; -2\, M_s^2,$$

while by item 1 of Lemma 2.8, the third term equals M_s^2. This shows the first claim (7.12).

We leave the proof of the second claim (7.13) as Exercise 7.19. □

Exercise 7.19. Show that if $M \in \mathcal{M}_c^2$, then (7.13) holds for any $t \geq 0$.

A generalization of the Sophomore's dream trick gives a useful identity for sums of squares of increments of a square-integrable martingale. We will use it in the proof of Theorem 8.28.

Corollary 7.20. *Let M be a square-integrable martingale: $\mathbb{E}[M_t^2] < \infty$ for all $t \geq 0$. Fix $0 \leq s \leq t$. Then, for any partition $s = s_0 < s_1 < \cdots < s_m = t$ of the time interval $[s,t]$, we have*

$$\sum_{j=1}^{m} \mathbb{E}\big[(M_{s_j} - M_{s_{j-1}})^2 \,|\, \mathcal{F}_s\big] \;=\; \mathbb{E}[M_t^2 \,|\, \mathcal{F}_s] - M_s^2 \;=\; \mathbb{E}[(M_t - M_s)^2 \,|\, \mathcal{F}_s]. \tag{7.14}$$

In particular,

$$\sum_{j=1}^{m} \mathbb{E}\big[(M_{s_j} - M_{s_{j-1}})^2\big] \;=\; \mathbb{E}[M_t^2 - M_s^2]. \tag{7.15}$$

Proof. Applying Lemma 7.18 and the Tower property (item 4 of Lemma 2.8), we have

$$\mathbb{E}\big[(M_{s_j} - M_{s_{j-1}})^2 \,|\, \mathcal{F}_s\big] \;=\; \mathbb{E}\big[\mathbb{E}\big[M_{s_j}^2 - M_{s_{j-1}}^2 \,|\, \mathcal{F}_{s_{j-1}}\big] \,|\, \mathcal{F}_s\big] \;=\; \mathbb{E}\big[M_{s_j}^2 - M_{s_{j-1}}^2 \,|\, \mathcal{F}_s\big].$$

▷ The first equality in asserted formula (7.14) follows from linearity and a telescoping sum:

$$\begin{aligned}
\sum_{j=1}^{m} \mathbb{E}\big[(M_{s_j} - M_{s_{j-1}})^2 \,|\, \mathcal{F}_s\big] \;&=\; \sum_{j=1}^{m} \mathbb{E}\big[M_{s_j}^2 - M_{s_{j-1}}^2 \,|\, \mathcal{F}_s\big] \\
&=\; \sum_{j=1}^{m} \Big(\mathbb{E}\big[M_{s_j}^2 \,|\, \mathcal{F}_s\big] - \mathbb{E}\big[M_{s_{j-1}}^2 \,|\, \mathcal{F}_s\big]\Big) \\
&=\; \mathbb{E}\big[M_{s_m}^2 \,|\, \mathcal{F}_s\big] - \mathbb{E}\big[M_{s_0}^2 \,|\, \mathcal{F}_s\big] \\
&=\; \mathbb{E}\big[M_t^2 \,|\, \mathcal{F}_s\big] - \mathbb{E}\big[M_s^2 \,|\, \mathcal{F}_s\big] \;=\; \mathbb{E}\big[M_t^2 \,|\, \mathcal{F}_s\big] - M_s^2,
\end{aligned}$$

using also item 1 of Lemma 2.8 to evaluate $\mathbb{E}\big[M_s^2 \,|\, \mathcal{F}_s\big] = M_s^2$.

▷ The second equality in asserted formula (7.14) follows from (7.12) in Lemma 7.18.

▷ Lastly, formula (7.15) follows by taking expected value of both sides of (7.14) and using linearity and the Tower property (items 3 and 4 of Lemma 2.8). □

8 Towards stochastic integration: continuous semimartingales

Our next aim is to develop a theory of stochastic integrals for continuous-time processes. Recall from Section 3.3 the process (3.6), thought of as a "discrete stochastic integral"

$$\sum_{k=1}^{n} H_k(Y_k - Y_{k-1}) =: (H \bullet Y)_n, \qquad n \in \mathbb{N},$$

of H with respect to Y (e.g., "cumulative profit"), where

▷ the *integrator* $(Y_n)_{n\in\mathbb{N}_0}$ is a suitable discrete-time process (e.g., stock price) and

▷ the *integrand* $(H_n)_{n\in\mathbb{N}}$ is another suitable discrete-time process (e.g., number of stocks).

Somewhat analogously to the Lebesgue (or Riemann) integral in real analysis, an idea for constructing a stochastic integral of H with respect to Y for two *continuous-time* processes $(Y_t)_{t\geq 0}$ and $(H_t)_{t\geq 0}$ could be to *approximate* it using a suitable discrete version, like the one above. With such a heuristic idea on the back of our minds, in Sections 8–9 we develop the theory of stochastic integration w.r.t. *continuous semimartingales*: these are defined as linear combinations of continuous *local martingales* (space $\mathcal{M}_c^{\mathrm{loc}}$, see Section 8.2) and continuous *finite-variation processes* (space $\mathcal{V}_c$, see Section 8.1). These two classes of processes are separate in the sense that *the only process that belongs to both classes is the zero process* (see Theorem 8.28 in the end of Section 8.2). Hence, we shall construct the integral of suitable H with respect to Y separately with $Y = X \in \mathcal{M}_c^{\mathrm{loc}}$ being a local martingale and with $Y = A \in \mathcal{V}_c$ being a finite-variation process. In conclusion, the following *Doob-Meyer decomposition* is a key motivation for us to develop the theory of finite-variation processes on the one hand, and local martingales on the other hand. It will make much more sense after reading Sections 8.1–8.2.

Definition 8.1. (Doob-Meyer decomposition) Consider a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ satisfying the usual conditions.
A continuous *semimartingale* on $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is an adapted continuous process

$$Y = Y_0 + X + A, \qquad X_0 = A_0 = 0, \tag{8.1}$$

with initial value $Y_0 \in \mathrm{m}\mathcal{F}_0$, where $X \in \mathcal{M}_c^{\mathrm{loc}}$ is a continuous local martingale and $A \in \mathcal{V}_c$ is a continuous finite-variation process. This decomposition is unique.

In real analysis, while the *Riemann integral* does produce a meaningful notion, the *Lebesgue integral* gives a construction which is much more amenable to applications (e.g., integrating functions with respect to other functions, or exchanging limits and integrals). Thus, it is reasonable to expect that as the first attempt for a construction of a *stochastic* integral, one would try out Lebesgue's method (though Riemann sum approximations will also play a role later).

In the next Section 8.1, we construct a *stochastic pathwise Lebesgue-Stieltjes type integral*

$$\int_0^t H_s \,\mathrm{d}A_s, \qquad t \geq 0,$$

defined for each $\omega \in \Omega$ and $t \geq 0$ as

$$\omega \quad \longmapsto \quad \int_0^t H_s(\omega)\,\mathrm{d}A_s(\omega),$$

where $A \in \mathcal{V}_c$ is a process whose sample paths $A(\omega) : [0,\infty) \to \mathbb{R}$ are continuous and have finite variation (see Definition 8.11), and H is a suitable real-valued process (e.g. bounded and measurable). See Proposition 8.17 for the precise setup.

Note, however, that even for a very well behaved H, the usual Lebesgue-Stieltjes integral

$$\int_0^t H_s \,\mathrm{d}B_s, \qquad t \geq 0,$$

with respect to sample paths $B(\omega) : [0, \infty) \to \mathbb{R}$ of Brownian motion is not well-defined, because B has almost surely infinite variation on any time interval $[0, t]$ (recall Exercise 1.22 and check out Exercise 8.5). However, one can show that Brownian motion has a finite *quadratic variation* (see Definition 8.36), which plays a key role in the construction of a stochastic integral w.r.t. it. Indeed, B is a local martingale, a prototypical example of an integrator in the space $\mathcal{M}_c^{\text{loc}}$. We discuss local martingales in Section 8.2 and develop applicable integration theory in Section 9.

8.1 Finite-variation processes and pathwise Lebesgue-Stieltjes integral

One possible generalization of Lebesgue integral (with respect to Lebesgue measure) to an integral with respect to a given well enough behaved function is known as the *Lebesgue-Stieltjes integral*. In essence, any function $f : [0, \infty) \to \mathbb{R}$ with *finite total variation* (cf. Definition 8.2) gives rise to a *signed measure* μ_f that can be used to define integral of the form

$$\int_0^t g(s) \,\mathrm{d}f(s) = \int_0^t g(s) \,\mathrm{d}\mu_f(s)$$

where $g : [0, \infty) \to \mathbb{R}$ is a suitable (integrable) function. In this section, we discuss the construction of Lebesgue-Stieltjes integral for processes whose sample paths have (a.s.) finite variation.

8.1.1 Functions of finite total variation and associated Borel measures

Definition 8.2. (Finite variation) Function $f : [0, \infty) \to \mathbb{R}$ has *finite variation* if for each $t \geq 0$, its *total variation* $\mathrm{var}_f(t)$ up to time t is finite:

$$\mathrm{var}_f(t) \;:=\; \lim_{n\to\infty} \sum_{k=1}^{\lfloor 2^n t \rfloor} |f(k2^{-n}) - f((k-1)2^{-n})| \;<\; \infty. \tag{8.2}$$

We say that $f : [0, \infty) \to \mathbb{R}$ is a (continuous) *finite-variation function* and write $f \in \mathrm{FV}$ if

- ▷ f is continuous;
- ▷ f has finite variation;
- ▷ and $f(0) = 0$.

The value $f(0) = 0$ is fixed just by convention (it is needed for uniqueness statements, and makes sense in view of Figure 8.1.1). As a sanity check, note that the definition makes sense:

Lemma 8.3. *Let $f : [0, \infty) \to \mathbb{R}$ be a continuous function, $f(0) = 0$. The following hold.*

1. *The limit*

$$\mathrm{var}_f(t) \;:=\; \lim_{n\to\infty} \sum_{k=1}^{\lfloor 2^n t \rfloor} |f(k2^{-n}) - f((k-1)2^{-n})| \;\in\; [0, +\infty], \qquad t \geq 0,$$

exists and defines a non-decreasing function $\mathrm{var}_f : [0, \infty) \to [0, +\infty]$ *with* $\mathrm{var}_f(0) = 0$.

2. *If* $\mathrm{var}_f(t) < \infty$ *for all* $t \geq 0$, *then* $\mathrm{var}_f : [0, \infty) \to [0, \infty)$ *is continuous (and* $f \in \mathrm{FV}$*).*

Proof. Write $\mathrm{var}_f^{(n)}(t) := \sum_{k=1}^{\lfloor 2^n t\rfloor} |f(k2^{-n}) - f((k-1)2^{-n})|$.

1. Fix $t \geq 0$. Note that the sequence $(\mathrm{var}_f^{(n)}(t))_{n\in\mathbb{N}}$ is non-decreasing as $n \uparrow \infty$:

$$\begin{aligned}\mathrm{var}_f^{(n)}(t) &:= \sum_{k=1}^{\lfloor 2^n t\rfloor} |f(k2^{-n}) - f((k-1)2^{-n})| \\ &\leq \sum_{k=1}^{\lfloor 2^n t\rfloor} \Big(|f(2k2^{-n-1}) - f((2k-1)2^{-n-1})| + |f((2k-1)2^{-n-1}) - f((2k-2)2^{-n-1})|\Big) \\ &\leq \sum_{\ell=1}^{2\lfloor 2^n t\rfloor} |f(\ell 2^{-n-1}) - f((\ell-1)2^{-n-1})| = \mathrm{var}_f^{n+1}(t),\end{aligned}$$

where we changed the summation index from k to $\ell = 2k$ in the last line. Hence, the sequence $(\mathrm{var}_f^{(n)}(t))_{n\in\mathbb{N}}$ has a limit in $[0, +\infty]$ for each fixed $t \geq 0$.
Moreover, since $t \mapsto \mathrm{var}_f^{(n)}(t)$ is non-decreasing on $[0, \infty)$ for each fixed $n \in \mathbb{N}$, the limit

$$t \quad\mapsto\quad \mathrm{var}_f(t) := \lim_{n\to\infty} \mathrm{var}_f^{(n)}(t)$$

is also non-decreasing on $[0, \infty)$. Clearly, we have $\mathrm{var}_f(0) = 0$. This proves item 1.

2. We leave this as Exercise 8.4. □

One could relax the continuity assumption to the requirement that $f : [0, \infty) \to \mathbb{R}$ is *càdlàg*, i.e., right-continuous with left limits. In that case, the total variation var_f is càdlàg as well.

Exercise 8.4. Show that if the continuous function $f : [0, \infty) \to \mathbb{R}$ has finite variation, then the limit

$$t \quad\mapsto\quad \mathrm{var}_f(t) := \lim_{n\to\infty} \mathrm{var}_f^{(n)}(t)$$

defines a continuous function $\mathrm{var}_f : [0, \infty) \to [0, \infty)$.

Exercise 8.5. Show that almost surely, Brownian motion does not have finite variation. *Hint*: *Exercise 1.22.*

Lemma 8.6. *Let $f \in \mathrm{FV}$. Then, we have*

$$|f(t) - f(s)| \leq \mathrm{var}_f(t) - \mathrm{var}_f(s) \qquad \text{for all } 0 \leq s \leq t. \tag{8.3}$$

In particular, for any partition $0 = t_0 < t_1 < \cdots < t_n = t$, we have

$$\sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| \leq \mathrm{var}_f(t). \tag{8.4}$$

Property (8.4) shows that the dyadic subdivision of $[0, t]$ used in Definition 8.2 in a sense maximizes the oscillation of f. In fact, the total variation $\mathrm{var}_f(t)$ could be defined as the limit over any refining subdivisions of $[0, t]$, see [LeG16, Proposition 4.2].

Proof. Using continuity, we approximate t from below by $t_n^- := 2^{-n}\lfloor 2^n t\rfloor \uparrow t$ and s from above by $s_n^+ := 2^{-n}\lceil 2^n s\rceil \downarrow s$ as $n \uparrow \infty$. Then, we have

$$|f(t_n^-) - f(s_n^+)| \leq \sum_{k=2^n s_n^+ + 1}^{2^n t_n^-} |f(k2^{-n}) - f((k-1)2^{-n})|$$

$$\begin{aligned} &\le \mathrm{var}_f^{(n)}(t_n^-) - \mathrm{var}_f^{(n)}(s_n^+) \\ &\le \mathrm{var}_f^{(n)}(t) - \mathrm{var}_f^{(n)}(s). \qquad\qquad [\text{by monotonicity of } \mathrm{var}_f^{(n)}(\cdot)] \end{aligned}$$

As $n \to \infty$, the right-hand side converges to $\mathrm{var}_f(t) - \mathrm{var}_f(s)$, while the left-hand side converges to $|f(t) - f(s)|$ as $n \to \infty$. This shows (8.3). To prove (8.4), note that

$$\begin{aligned} \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| &\le \sum_{k=1}^{n} |\mathrm{var}_f(t_k) - \mathrm{var}_f(t_{k-1})| = \sum_{k=1}^{n} \big(\mathrm{var}_f(t_k) - \mathrm{var}_f(t_{k-1})\big) \\ &= \mathrm{var}_f(t_n) - \underbrace{\mathrm{var}_f(t_0)}_{=0} = \mathrm{var}_f(t) \end{aligned}$$

by non-negativity of var_f and cancellations in the telescoping sum. □

The key for constructing integrals with respect to finite-variation processes is the following characterization of finite-variation functions f in terms of differences of two non-negative non-decreasing parts $f^{\pm}$. See also Figure 8.1.1.

Proposition 8.7. (Characterization of FV) *Let $f : [0, \infty) \to \mathbb{R}$ be a continuous function with $f(0) = 0$. Then, $f \in \mathrm{FV}$ if and only if we have*

$$f(t) = f^+(t) - f^-(t)$$

for two non-decreasing continuous functions $f^{\pm} : [0, \infty) \to [0, \infty)$ with $f^{\pm}(0) = 0$.

Remark 8.8. There are many choices for the decomposition $f^{\pm}$. The choice

$$f^{\pm}(t) := \tfrac{1}{2}(\mathrm{var}_f(t) \pm f(t)) \tag{8.5}$$

is optimal in the sense of Exercise 8.9.

Exercise 8.9. Check that if $f^{\pm}, \widetilde{f}^{\pm}$ both satisfy the properties in Proposition 8.7, then

$$f^+(t) \le \widetilde{f}^+(t) \qquad \text{and} \qquad f^-(t) \le \widetilde{f}^-(t) \qquad \text{for all } t \ge 0.$$

Hint: *Show first that the space* FV *of finite-variation functions is a real vector space* (*cf. Exercise 8.12*), *and that*

$$\mathrm{var}_{af}(t) = |a|\, \mathrm{var}_f(t) \qquad \text{and} \qquad \mathrm{var}_{f_1+f_2}(t) \le \mathrm{var}_{f_1}(t) + \mathrm{var}_{f_2}(t), \qquad a \in \mathbb{R},\ t \ge 0.$$

In the sequel, we will always implicitly use the optimal choice (8.5).

Proof of Proposition 8.7.

"⇒" If the total variation var_f is finite at all times, then we can define

$$f^{\pm}(t) := \tfrac{1}{2}(\mathrm{var}_f(t) \pm f(t)). \tag{8.6}$$

By Lemma 8.3, these functions $f^{\pm} : [0, \infty) \to [0, \infty)$ are continuous and satisfy $f^{\pm}(0) = 0$. To see that they are non-decreasing, using Lemma 8.6, we have

$$f^+(t) - f^+(s) = \tfrac{1}{2}\underbrace{(\mathrm{var}_f(t) - \mathrm{var}_f(s))}_{\ge |f(t) - f(s)|} + \tfrac{1}{2}(f(t) - f(s)) \ge 0 \qquad \text{for all } 0 \le s \le t,$$

and similarly,

$$f^-(t) - f^-(s) = \tfrac{1}{2}\underbrace{(\mathrm{var}_f(t) - \mathrm{var}_f(s))}_{\ge |f(t) - f(s)|} - \tfrac{1}{2}(f(t) - f(s)) \ge 0 \qquad \text{for all } 0 \le s \le t.$$

"⇐" Conversely, we estimate the variation of $f(t) := f^+(t) - f^-(t)$, assuming that $f^\pm$ satisfy the above properties. We have

$$\begin{aligned} \mathrm{var}_f^{(n)}(t) &:= \sum_{k=1}^{\lfloor 2^n t \rfloor} |f(k2^{-n}) - f((k-1)2^{-n})| \\ &\leq \sum_{k=1}^{\lfloor 2^n t \rfloor} |f^+(k2^{-n}) - f^+((k-1)2^{-n})| + \sum_{k=1}^{\lfloor 2^n t \rfloor} |f^-(k2^{-n}) - f^-((k-1)2^{-n})| \\ &= \sum_{k=1}^{\lfloor 2^n t \rfloor} \big(f^+(k2^{-n}) - f^+((k-1)2^{-n})\big) + \sum_{k=1}^{\lfloor 2^n t \rfloor} \big(f^-(k2^{-n}) - f^-((k-1)2^{-n})\big) \end{aligned}$$

since $f^\pm$ are non-decreasing. Thus, due to cancellations in the telescoping sums, we see that the right-hand side equals

$$\begin{aligned} &\underbrace{f^+(2^{-n}\lfloor 2^n t \rfloor)}_{\leq f^+(t)} - \underbrace{f^+(0)}_{=0} + \underbrace{f^-(2^{-n}\lfloor 2^n t \rfloor)}_{\leq f^-(t)} - \underbrace{f^-(0)}_{=0} \\ \leq\ & f^+(t) + f^-(t) < \infty \qquad \text{for all } n \in \mathbb{N} \text{ and } t \geq 0. \end{aligned}$$

This shows that $\mathrm{var}_f(t) := \lim_{n\to\infty} \mathrm{var}_f^{(n)}(t) < \infty$, as claimed. □

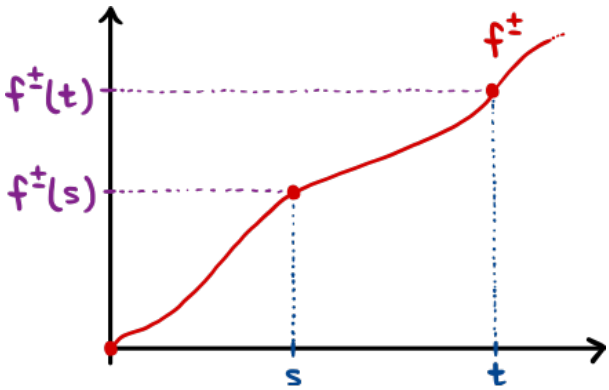


Figure 8.1.1. Illustration of the determination of the Borel measures $\mu_f^\pm$.

The key idea arising from the decomposition in Proposition 8.7 is that the two non-decreasing functions $f^\pm : [0,\infty) \to [0,\infty)$ represent *cumulative distribution functions* of two Borel measures (see Figure 8.1.1 for an illustration) on $([0,\infty), \mathscr{B}([0,\infty)))$, which we denote as $\mu_f^\pm$:

$$\mu_f^+([s,t]) := f^+(t) - f^+(s) \qquad \text{and} \qquad \mu_f^-([s,t]) := f^-(t) - f^-(s) \qquad \text{for all } 0 \leq s \leq t.$$

Then,

$$\mu_f := \mu_f^+ - \mu_f^-$$

is a signed Borel measure on $([0,\infty), \mathscr{B}([0,\infty)))$, uniquely[15] determined by f as

$$\mu([0,t]) = f(t) \qquad \text{for all } t \geq 0,$$

see Lemma 8.10. For convenience, we also write

$$|\mu|_f := \mu_f^+ + \mu_f^-$$

(this positive measure is also sometimes called the total variation of f [LeG16]).

[15] (8.5) gives the unique decomposition $\mu := \mu^+ - \mu^-$ for which $\mu^\pm$ are supported on *disjoint* Borel sets [LeG16].

8.1.2 Lebesgue-Stieltjes integral with respect to finite-variation functions

We are now ready to formulate the notion of an integral with respect to $f \in \mathrm{FV}$. We set

$$\begin{aligned}\int_0^\infty g(s)\,\mathrm{d}f(s) &:= \int_0^\infty g(s)\,\mathrm{d}\mu_f(s)\\ &:= \int_0^\infty g(s)\,\mathrm{d}\mu_f^+(s) - \int_0^\infty g(s)\,\mathrm{d}\mu_f^-(s),\end{aligned}$$

for any integrable function

$$g \in L^1(\mu_f^+) \cap L^1(\mu_f^-) = \left\{ g : [0,\infty) \to \mathbb{R} \,\middle|\, \int_0^\infty |g(s)|\,\mathrm{d}\mu_f^+(s) < \infty \text{ and } \int_0^\infty |g(s)|\,\mathrm{d}\mu_f^-(s) < \infty \right\},$$

and analogously, we set

$$\int_0^t g(s)\,\mathrm{d}f(s) := \int_0^\infty g(s)\,\mathbb{1}_{[0,t]}(s)\,\mathrm{d}f(s), \qquad t \geq 0.$$

We also denote

$$\begin{aligned}\int_0^\infty g(s)|\,\mathrm{d}f(s)| &:= \int_0^\infty g(s)\,\mathrm{d}|\mu|_f(s)\\ &:= \int_0^\infty g(s)\,\mathrm{d}\mu_f^+(s) + \int_0^\infty g(s)\,\mathrm{d}\mu_f^-(s),\end{aligned}$$

Lemma 8.10. *For each $f \in \mathrm{FV}$, we have*

$$\mu([0,t]) = \int_0^t \mathrm{d}f(s) = f(t) \qquad \text{and} \qquad |\mu|([0,t]) = \int_0^t |\,\mathrm{d}f(s)| = \mathrm{var}_f(t), \qquad t \geq 0.$$

Proof. By definition (using the choice (8.5)), we have

$$\begin{aligned}\int_0^t \mathrm{d}f(s) = \mu([0,t]) &:= \mu^+([0,t]) - \mu^-([0,t])\\ &:= (f^+(t) - f^+(0)) - (f^-(t) - f^-(0)) = f^+(t) - f^-(t)\\ &:= \tfrac{1}{2}(\mathrm{var}_f(t) + f(t)) - \tfrac{1}{2}(\mathrm{var}_f(t) - f(t)) = f(t),\end{aligned}$$

and similarly,

$$\begin{aligned}\int_0^t |\,\mathrm{d}f(s)| = |\mu|([0,t]) &:= \mu^+([0,t]) + \mu^-([0,t])\\ &:= (f^+(t) - f^+(0)) + (f^-(t) - f^-(0)) = f^+(t) + f^-(t)\\ &:= \tfrac{1}{2}(\mathrm{var}_f(t) + f(t)) + \tfrac{1}{2}(\mathrm{var}_f(t) - f(t)) = \mathrm{var}_f(t).\end{aligned}$$

□

8.1.3 Finite-variation processes

Let us now consider adapted stochastic processes on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ satisfying the usual conditions (Definition 5.6). Observe that the condition UC2 of the underlying filtration $\mathcal{F}_\bullet$ enables us to assume continuity and finite variation property only almost surely, as we can get rid of the complementary event by defining our processes to equal zero there, without losing measurability.

Definition 8.11. A (continuous) *finite-variation process* is an adapted stochastic process $A = (A_t)_{t\geq 0}$ whose sample paths are almost surely finite-variation functions: $A \in \mathrm{FV}$ a.s.:

$$\mathbb{P}\Big[\{\omega \in \Omega \mid t \mapsto A_t(\omega) \text{ is continuous, has finite variation, and } A_0(\omega) = 0\}\Big] = 1.$$

Its *total variation process* is defined as

$$V_t^A(\omega) := \begin{cases} \mathrm{var}_{A(\omega)}(t) := \lim_{n\to\infty} \sum_{k=1}^{\lfloor 2^n t\rfloor} |A_{k2^{-n}}(\omega) - A_{(k-1)2^{-n}}(\omega)|, & \omega \in E^A, \\ 0 & \omega \notin E^A, \end{cases} \qquad t \geq 0,$$

$$E^A := \big\{\omega \in \Omega \mid t \mapsto A_t(\omega) \text{ is continuous, has finite variation, and } A_0(\omega) = 0\big\}. \quad (8.7)$$

The set of all (continuous) *finite-variation processes* on $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is denoted as

$$\mathcal{V}_c := \big\{A = (A_t)_{t\geq 0} \mid A \text{ is a continuous finite-variation process, } A_0 = 0\big\}.$$

The requirement $A_0 = 0$ is imposed for convenience, and it is useful later.

Exercise 8.12. Show that $\mathcal{V}_c$ is a real vector space.

Lemma 8.13. *Let $A \in \mathcal{V}_c$ be a finite-variation process w.r.t a filtration $\mathcal{F}_\bullet$ satisfying the usual conditions (Definition 5.6). Then, the total variation process V^A is adapted to $\mathcal{F}_\bullet$ and, almost surely, $t \mapsto V_t^A$ is continuous, non-decreasing, and $V_0^A = 0$.*

Proof sketch. It suffices to study the measurability on the event E^A, by the usual condition UC2. On this event, Lemma 8.10 gives the identity

$$V_t^A(\omega) = \int_0^t |\,\mathrm{d}A_s(\omega)|, \qquad t \geq 0,\ \omega \in E^A,$$

which shows that $\omega \mapsto V_t^A(\omega)$ is $\mathcal{F}_t$-measurable on E^A for all $t \geq 0$. Thus, V^A is adapted to $\mathcal{F}_\bullet$. The other asserted properties of V_t^A follow from Lemma 8.3. □

A particularly important class of finite-variation processes are *increasing* (non-decreasing) processes. Assuming that $A_0 = 0$, any non-decreasing continuous process is also non-negative.

Lemma 8.14. *Let V be an adapted process whose sample paths are almost surely*

- ▷ *continuous;*
- ▷ *non-decreasing;*
- ▷ *and $V_0 = 0$.*

Then, V is a finite-variation process: $V \in \mathcal{V}_c$.

Proof. This follows immediately from Proposition 8.7 (with $f(t) = f^+(t) = V_t$). □

Definition 8.15. We call processes in Lemma 8.14 *increasing* and we denote $V \in \mathcal{V}_c^+$.

Corollary 8.16. *Let $A \in \mathcal{V}_c$ be a finite-variation process. Then, the total variation process V^A is an increasing process: $V^A \in \mathcal{V}_c^+$.*

8.1.4 Pathwise integral with respect to finite-variation processes

We can now apply the construction from Section 8.1.2 pathwise. Indeed, for each (continuous) finite-variation process $A \in \mathcal{V}_c$ on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ satisfying the usual conditions (Definition 5.6), we can define the measures $\mu^{\pm}_{A(\omega)}$ and the pathwise integral

$$\omega \quad \longmapsto \quad \int_0^\infty g(s)\, \mathrm{d}A_s(\omega) := \int_0^\infty g(s)\, \mathrm{d}\mu^+_{A(\omega)}(s) - \int_0^\infty g(s)\, \mathrm{d}\mu^-_{A(\omega)}(s), \qquad \omega \in \Omega,$$

provided that the function $g : [0, \infty) \to \mathbb{R}$ is integrable. More generally, we can take $g = H$ to be a stochastic process too:

$$\omega \quad \longmapsto \quad \int_0^\infty H_s(\omega)\, \mathrm{d}A_s(\omega) := \int_0^\infty H_s(\omega)\, \mathrm{d}\mu^+_{A(\omega)}(s) - \int_0^\infty H_s(\omega)\, \mathrm{d}\mu^-_{A(\omega)}(s), \qquad \omega \in \Omega.$$

<u>**Issues.**</u> A few natural questions now arise:

- ▷ Does $(\omega, t) \mapsto \int_0^t H_s(\omega)\, \mathrm{d}A_s(\omega) =: (H \bullet A)_t(\omega)$ define a stochastic process $H \bullet A$?
- ▷ Is $H \bullet A$ adapted to $\mathcal{F}_\bullet$?

To ensure desired properties, we need to make assumptions on the integrand process H.

- ▷ In order for the integral to be finite, we assume (almost sure) *integrability*, see (8.8).
- ▷ If H is *progressively measurable* (cf. Definition 5.9)[16], then $H \bullet A$ is adapted to $\mathcal{F}_\bullet$.

Proposition 8.17. (Integral) *Let $A \in \mathcal{V}_c$ be a finite-variation process w.r.t a filtration $\mathcal{F}_\bullet$ satisfying the usual conditions (Definition 5.6). Suppose that H is a progressively measurable process such that*[a] *almost surely (i.e. for $\mathbb{P}$-almost every $\omega \in \Omega$),*

$$\int_0^t |H_s(\omega)||\,\mathrm{d}A_s(\omega)| < \infty \qquad \text{for all } t \geq 0. \tag{8.8}$$

Then, the integral process $H \bullet A = ((H \bullet A)_t)_{t \geq 0}$ defined as

$$(H \bullet A)_t(\omega) := \begin{cases} \int_0^t H_s(\omega)\, \mathrm{d}A_s(\omega), & \omega \in E^A \cap E^H, \\ 0, & \omega \notin E^A \cap E^H, \end{cases} \qquad t \geq 0,$$

where E^A is the event (8.7) and $E^H := \{\omega \in \Omega \mid (8.8) \text{ holds}\}$, is a finite-variation process.

[a] Recall that $|\,\mathrm{d}A_s(\omega)|$ means $\mathrm{d}\mu^+_{A(\omega)}(s) + \mathrm{d}\mu^-_{A(\omega)}(s)$.

Proof sketch. See, e.g. [LeG16, Proposition 4.5] for more details.

[16] Recall that for example, an adapted process with (a.s.) continuous sample paths is progressively measurable.

Step 1. ($H \bullet A$ is well-defined): Write $H^{\pm} := \max\{\pm H, 0\}$ for the positive and negative parts and

$$\begin{aligned} H \bullet A &= H^+ \bullet A^+ - H^+ \bullet A^- - H^- \bullet A^+ + H^- \bullet A^- \\ &= \big(H^+ \bullet A^+ + H^- \bullet A^-\big) - \big(H^+ \bullet A^- + H^- \bullet A^+\big). \end{aligned} \tag{8.9}$$

The integral $(H \bullet A)_t$ is finite for all $t \geq 0$, since the assumption (8.8) implies that all terms on the right-hand side of (8.9) are finite.

Step 2. ($H \bullet A$ is continuous): This is clear for $\omega \notin E^A \cap E^H$. For $\omega \in E^A \cap E^H$, we have

$$(H \bullet A)_t(\omega) := \int_0^t H_s(\omega)\, \mathrm{d}A_s(\omega) = \int_0^\infty H_s(\omega)\, \mathbb{1}_{[0,t]}(s)\, \mathrm{d}A_s(\omega).$$

Since the convergence

$$\begin{aligned} \mathbb{1}_{[0,u]}(s) &\to \mathbb{1}_{[0,t]}(s) \qquad \text{as } u \downarrow t, \\ \mathbb{1}_{[0,u]}(s) &\to \mathbb{1}_{[0,t)}(s) \qquad \text{as } u \uparrow t \end{aligned}$$

holds pointwise for $s \geq 0$, Dominated Convergence Theorem [Kyt20, Theorem VII.19] shows that the integral is continuous:

$$\lim_{s \to t} (H \bullet A)_s(\omega) = (H \bullet A)_t(\omega), \qquad \omega \in E^A \cap E^H.$$

Step 3. ($H \bullet A$ has finite variation): For each $\omega \in \Omega$, the last line of (8.9) is a difference of two non-decreasing, non-negative, continuous functions, which both vanish at $t = 0$, so Proposition 8.7 readily shows that the left-hand side of (8.9) is a finite-variation function $t \mapsto (H \bullet A)_t(\omega)$.

Step 4. ($H \bullet A$ is adapted to $\mathcal{F}_\bullet$): This can be proven via the "standard machine" and a monotone class argument as follows. By (8.9), we may restrict our attention to *non-negative* H and *increasing* $A \in \mathcal{V}_c^+$. Moreover, it suffices to prove that for each *fixed* $T > 0$, the quantity $(H \bullet A)_T$ is $\mathcal{F}_T$-measurable whenever (8.8) holds at time $t = T$ and the following map is measurable:

$$H : (\Omega \times [0,T], \mathcal{F}_T \otimes \mathscr{B}([0,T])) \to (\mathbb{R}, \mathscr{B}(\mathbb{R})). \tag{8.10}$$

Step 4a). Consider the pi-system (5.5) associated to the progressive sigma-algebra $\mathscr{P}$,

$$\Pi_T := \{E \times (u,v] \mid E \in \mathcal{F}_T \text{ and } 0 \leq u < v \leq T\}.$$

Note that, for each $E \in \mathcal{F}_T$ and $0 \leq u < v \leq T$, the random variable

$$\omega \longmapsto \int_0^T \mathbb{1}_E(\omega) \mathbb{1}_{(u,v]}(s)\, \mathrm{d}A_s(\omega) = \mathbb{1}_E(\omega)(A_v(\omega) - A_u(\omega)) = (\mathbb{1}_{E \times (u,v]} \bullet A)_T$$

is $\mathcal{F}_T$-measurable since A is adapted. This shows the claim for the *indicator* function $H = \mathbb{1}_G$ of every set $G = E \times (u,v] \in \Pi_T$ in the pi-system (5.5).

Step 4b). We use *Monotone Class Theorem* A.30 to argue that $(H \bullet A)_T$ is $\mathcal{F}_T$-measurable for any *bounded* process (8.10). To this end, we verify that the collection

$$\mathcal{H} := \big\{H : (\Omega \times [0,T], \mathcal{F}_T \otimes \mathscr{B}([0,T])) \to (\mathbb{R}, \mathscr{B}(\mathbb{R})) \mid H \text{ is bounded and } (H \bullet A)_T \in \mathrm{m}\mathcal{F}_T\big\}$$

is a *monotone class*, that is, satisfies the following properties.

- ▷ The constant process $1 \in \mathcal{H}$; this is clear.
- ▷ $\mathcal{H}$ is an $\mathbb{R}$-vector space; this is clear.
- ▷ Suppose that $(H^{(n)})_{n\in\mathbb{N}}$ is a sequence of elements of $\mathcal{H}$ such that

$$0 \le H_s^{(n)}(\omega) \uparrow H_s(\omega) \qquad \text{as } n \uparrow \infty$$

pointwise for $(\omega, s) \in \Omega \times [0,T]$, where the limit $H : \Omega \times [0,T] \to \mathbb{R}$ is bounded. Then,

  - ⁎ H is a measurable map from $(\Omega \times [0,T], \mathcal{F}_T \otimes \mathscr{B}([0,T]))$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ as a pointwise limit of such processes;
  - ⁎ H is *bounded* by assumption; and
  - ⁎ $(H \bullet A)_T \in \mathrm{m}\mathcal{F}_T$ as a pointwise limit of elements in $\mathrm{m}\mathcal{F}_T$: by the Monotone Convergence Theorem [Kyt20, Theorem VII.8], we have (on the event $\{E^{H^{(n)}}$ infinitely often$\}$)

$$\underbrace{(H^{(n)} \bullet A)_T(\omega)}_{\in\, \mathrm{m}\mathcal{F}_T} = \int_0^T H_s^{(n)}(\omega)\, \mathrm{d}A_s(\omega) \overset{n\to\infty}{\longrightarrow} \underbrace{(H \bullet A)_T(\omega)}_{\in\, \mathrm{m}\mathcal{F}_T}.$$

  Hence, $H \in \mathcal{H}$, which verifies that $\mathcal{H}$ is a monotone class.

Thus, because $\mathcal{H}$ contains the indicator function $\mathbb{1}_G$ of every set $G \in \Pi_T$ in the pi-system (5.5) of Step 4a, Monotone Class Theorem A.30 implies that $\mathcal{H}$ contains *all bounded measurable functions* (8.10). This shows that $(H \bullet A)_T \in \mathrm{m}\mathcal{F}_T$ for any *bounded* process (8.10).

Step 4c). It remains to argue (via the "standard machine") that $(H \bullet A)_T$ is $\mathcal{F}_T$-measurable for any non-negative process (8.10). We can approximate such H by linear combinations of simple (piecewise constant) bounded processes $H^{(n)} \in \mathcal{H}$ as follows:

$$H_s^{(n)}(\omega) := \sum_{k=1}^{n2^n} (k-1)2^{-n}\, \mathbb{1}_{\{(k-1)2^{-n} < H_s(\omega) \le k2^{-n}\}} + n\, \mathbb{1}_{\{H_s(\omega) > n\}}, \qquad s \in [0,T],\ \omega \in \Omega$$

which converge pointwise from below to H: we have $H_s^{(n)}(\omega) \uparrow H_s(\omega)$ as $n \uparrow \infty$. Hence, again by the Monotone Convergence Theorem [Kyt20, Theorem VII.8], $(H \bullet A)_T \in \mathrm{m}\mathcal{F}_T$ as a pointwise limit of elements in $\mathrm{m}\mathcal{F}_T$: by Step 4b, we have (on the event $\{E^{H^{(n)}}$ infinitely often$\}$)

$$\underbrace{(H^{(n)} \bullet A)_T(\omega)}_{\in\, \mathrm{m}\mathcal{F}_T} = \int_0^T H_s^{(n)}(\omega)\, \mathrm{d}A_s(\omega) \overset{n\to\infty}{\longrightarrow} \underbrace{(H \bullet A)_T(\omega)}_{\in\, \mathrm{m}\mathcal{F}_T}.$$

This shows that $(H \bullet A)_T \in \mathrm{m}\mathcal{F}_T$ for any non-negative process (8.10) (concluding by (8.9)). □

Remark 8.18. Instead[17] of the pi-system (5.5), we could also use the more restrictive *predictable* (*or previsible*) *pi-system*

$$\Pi^+ := \{E \times (u,v] \mid E \in \mathcal{F}_u \text{ and } 0 \le u < v\}. \tag{8.11}$$

If $H : \Omega \times (0,\infty) \to \mathbb{R}$ is a *predictable* process, that is, a measurable map from $(\Omega \times (0,\infty), \mathscr{P}^+)$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with respect to the sigma-algebra $\mathscr{P}^+ := \sigma(\Pi^+)$ generated by (8.11), then H is also progressively measurable. Note that the converse does not hold: e.g., the Poisson process with respect to its natural filtration is progressively measurable but not predictable.

[17] This makes little difference when considering continuous processes, but it does with discontinuous processes.

Analogously to Proposition 5.19, (left-)continuous processes are predictable. However, as the example of the Poisson process shows, right-continuous processes are not necessarily predictable.

Proposition 8.19. *Consider an adapted real-valued stochastic process $H = (H_t)_{t \geq 0}$ on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$. Suppose that for every $\omega \in \Omega$, the sample path $t \mapsto H_t(\omega)$ is left-continuous. Then H is predictable.*

Proof. For $n \in \mathbb{N}$ and $t \geq 0$, define a random variable $H_t^{(n)} := H_{(k-1)2^{-n}}$ when $t \in ((k-1)2^{-n}, k2^{-n}]$ with $k \in \mathbb{N}$, and $H_0^{(n)} := H_0$. Then, $H^{(n)}$ are predictable and, since the sample paths of H are left-continuous, $(\omega, t) \mapsto H_t(\omega)$ is a limit of the measurable maps $(\omega, t) \mapsto H_t^{(n)}(\omega)$: we have

$$\lim_{n\to\infty} H_t^{(n)}(\omega) = \lim_{n\to\infty} \sum_{k=1}^{\infty} H_{(k-1)2^{-n}}(\omega)\, \mathbb{1}_{\{(k-1)2^{-n} < H_t(\omega) \leq k2^{-n}\}} = H_t(\omega)$$

for every $\omega \in \Omega$ and $t \geq 0$. Hence, H is predictable. □

Remark 8.20. Comparison of predictable and progressively measurable processes:

▷ Any predictable process is progressively measurable.

▷ Any continuous, right-continuous, or left-continuous process is progressively measurable.

▷ Any continuous or left-continuous process is predictable.

▷ Processes taking discrete values (e.g. Poisson process) are not in general predictable.

▷ Right-continuous processes are not necessarily predictable: for example the Poisson process has a right-continuous modification but is not predictable.

8.2 Local martingales

Throughout, as before, we consider a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ satisfying the *usual conditions*. Recall the definition of a continuous-time martingale from Definition 7.1: process $M = (M_t)_{t\geq 0}$ satisfying MG1 (adapted): M_t is $\mathcal{F}_t$-measurable for every t; MG2 (integrable): $M_t \in L^1(\mathbb{P})$ for every t; and MG3 (martingale property): a.s., $\mathbb{E}[M_t \,|\, \mathcal{F}_s] = M_s$ for every $0 \leq s < t$.

We frequently use the following notation for important spaces of martingales.

▷ The set of all *continuous martingales* on $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is denoted as

$$\mathcal{M}_c := \big\{ M = (M_t)_{t\geq 0} \,|\, M \text{ is a martingale and } t \mapsto M_t(\omega) \text{ is almost surely continuous} \big\}.$$

▷ The set of all *uniformly L^2-bounded continuous martingales* on $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is denoted as

$$\mathcal{M}_c^2 := \big\{ M = (M_t)_{t\geq 0} \in \mathcal{M}_c \,|\, M \text{ is uniformly } L^2\text{-bounded: } \sup_{t\geq 0} \mathbb{E}[M_t^2] < \infty \big\} \subset \mathcal{M}_c.$$

By Exercises 7.2 & 7.6, both $\mathcal{M}_c^2$ and $\mathcal{M}_c$ are vector spaces.

It is often useful to relax especially the integrability condition MG2 slightly. This can be accomplished by considering the process stopped before a blow-up, which can be accomplished by the notion of *stopped processes* (as in Section 5.3 — here, one needs to be mindful about measurability issues). This idea leads to the following notion of a "local" martingale (where "local" refers to locality in time, and is standard terminology in the literature).

Definition 8.21. (Local martingale) Stochastic process $X = (X_t)_{t\geq 0}$ on a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is called a *local martingale* if

loc-MG1. (it is adapted to $\mathcal{F}_\bullet$): for every $t \geq 0$, the random variable X_t is $\mathcal{F}_t$-measurable;

loc-MG2. there exists a sequence[a] $(\tau_n)_{n\in\mathbb{N}}$ of stopping times such that $\tau_n(\omega) \uparrow \infty$ as $n \uparrow \infty$ almost surely, and the stopped processes $X^{\tau_n} := (X_{t\wedge\tau_n})_{t\geq 0}$ are martingales.

The set of all *continuous local martingales* on $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is denoted as

$\mathcal{M}_c^{\text{loc}} := \{X = (X_t)_{t\geq 0} \mid X \text{ is a local martingale and } t \mapsto X_t(\omega) \text{ is almost surely continuous}\}.$

[a] One sometimes says that the sequence $(\tau_n)_{n\in\mathbb{N}}$ *reduces*, or *localizes* X.

Exercise 8.22. Show that $\mathcal{M}_c^{\text{loc}}$ is a real vector space.

- ▷ Here, we only consider *continuous* local martingales to ensure measurability. One could relax the continuity assumption and consider càdlàg processes instead.
- ▷ It follows from Definition 8.21 that the initial value $X_0 \in L^1(\mathbb{P})$ of X is integrable. Sometimes in the literature (e.g. [LeG16]) it is assumed either that $X_0 = 0$ for simplicity, or X_0 is allowed to be an arbitrary $\mathcal{F}_0$-measurable random variable (not necessarily integrable).
- ▷ Any continuous martingale is also a continuous local martingale (by applying item 4 of Theorem 7.3 to $\tau_n := n$). Therefore, we have

$$\mathcal{M}_c^2 \subset \mathcal{M}_c \subset \mathcal{M}_c^{\text{loc}}.$$

- ▷ If $X \in \mathcal{M}_c^{\text{loc}}$ is bounded, then it is also a martingale: $X \in \mathcal{M}_c$ (see Exercise 8.25).
- ▷ If $X \in \mathcal{M}_c^{\text{loc}}$ and τ is a stopping time, then the stopped process $X^\tau := (X_{t\wedge\tau})_{t\geq 0}$ is also a continuous local martingale (see Exercise 8.26).

Exercise 8.23. Let $X = (X_t)_{t\geq 0}$ be a local martingale such that X is non-negative, i.e., $X_t \geq 0$ for each $t \geq 0$. Prove that X is a supermartingale. *Hint: Conditional Fatou's lemma (item 4 of Lemma 2.10).*

Exercise 8.24. Let $X = (X_t)_{t\geq 0}$ be a continuous local martingale. Suppose the collection

$$\big(\{X_\tau \mid \tau \text{ is a stopping time such that } \tau(\omega) \leq T \text{ for each } \omega \in \Omega\}\big)_{T\geq 0}$$

is UI. Show that $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ for all bounded stopping times τ. Conclude that X is a martingale.
Hint: Optional Stopping Theorem 7.3.

Exercise 8.25. Let $X = (X_t)_{t\geq 0}$ be a continuous local martingale which is uniformly L^1-*dominated*: there exists an integrable random variable $\eta \in L^1(\mathbb{P})$ such that $|X_t| \leq \eta$ for all $t \geq 0$. Prove that X is a UI martingale.

Exercise 8.26. Let X be a continuous local martingale and τ a stopping time. Prove that the stopped process $X^\tau := (X_{t\wedge\tau})_{t\geq0}$ is also a continuous local martingale: $X^\tau \in \mathcal{M}_c^{\text{loc}}$.

In the definition of a local martingale, we can also without loss of generality require that each X^{τ_n} is UI by modifying the localizing sequence τ_n to, e.g., $\sigma_n := \tau_n \wedge n$ for each n.

Proposition 8.27. (Local martingale) *Let $\mathcal{F}_\bullet = (\mathcal{F}_t)_{t\geq0}$ be a filtration. The following are equivalent for a continuous real-valued stochastic process $X = (X_t)_{t\geq0}$ adapted to $\mathcal{F}_\bullet$.*

loc-MG'1. *X is a local martingale w.r.t. $\mathcal{F}_\bullet$.*

loc-MG'2. *There exists a sequence $(\sigma_n)_{n\in\mathbb{N}}$ of stopping times w.r.t. $\mathcal{F}_\bullet$ such that $\sigma_n \uparrow \infty$ as $n \uparrow \infty$ a.s., and the stopped processes $X^{\sigma_n} := (X_{t\wedge\sigma_n})_{t\geq0}$ are UI martingales.*

Proof. The implication loc-MG'2 ⇒ loc-MG'1 is clear. To show loc-MG'1 ⇒ loc-MG'2, let X be a continuous local martingale and $(\tau_n)_{n\in\mathbb{N}}$ a localizing sequence as in loc-MG2 of the original Definition 8.21. Define $\sigma_n := \tau_n \wedge n$ for each n. Then, they are stopping times and $\sigma_n \uparrow \infty$ as $n \uparrow \infty$ a.s. Moreover, since $(X_{t\wedge\tau_n})_{t\geq0}$ is a continuous local martingale (see Corollary 5.21), so is $X^{\sigma_n} = (X_{t\wedge\tau_n\wedge n})_{t\geq0}$ for any fixed $n \in \mathbb{N}$. As the time-interval $[0, n]$ is compact, the process X^{σ_n} is furthermore uniformly bounded in t by some constant that depends on n. It thus follows that each X^{σ_n} is UI (by Exercise 4.10). (It suffices that they are UI separately[18] for each n.) □

Any finite-variation continuous local martingale is (essentially) zero.

Theorem 8.28. *Let $X \in \mathcal{M}_c^{\text{loc}}$. If $X \in \mathcal{V}_c$ is also a finite-variation process, then*

$$\mathbb{P}\left[X_t = 0 \text{ for all } t \geq 0\right] = 1.$$

Exercise 8.29. Prove that the Doob-Meyer decomposition (8.1) is unique by using Theorem 8.28 and linearity.

Proof. Since by Lemmas 8.13 and 8.10, the variation V^X of X,

$$t \quad \mapsto \quad \int_0^t |\,\mathrm{d}X_s\,| = V_t^X,$$

is a continuous process, Lemma 5.7 shows that the first time when it becomes large,

$$\tau_n := \inf\Big\{t \geq 0 \mid \int_0^t |\,\mathrm{d}X_s\,| \geq n\Big\},$$

is a stopping time, as the hitting time of the closed set $[n, \infty)$. Note that, as X has finite variation, we have $\tau_n \uparrow \infty$ as $n \uparrow \infty$ almost surely. The *rough goal* is to show that $\mathbb{E}\left[(X_t^{\tau_n})^2\right] = 0$, which implies that $X_t^{\tau_n} = 0$ almost surely, and then take the limit $n \to \infty$. (Note that a density argument will also be needed to conclude that $X_t = 0$ almost surely *simultaneously* for all t.)

[18] A possible source of confusion here is that the uniform integrability (UI) concerns the processes indexed by the time-index $t \in [0, \infty)$, while the integers $n \in \mathbb{N}$ indexing the stopping times have a different role.

Step 1. We first argue that, for each $n \in \mathbb{N}$, the stopped process $X^{\tau_n} := (X_{t\wedge\tau_n})_{t\geq 0}$ is a continuous *bounded* local martingale, and thus, a *continuous bounded martingale* (by Exercise 8.25).

▷ Indeed, Exercise 8.26 shows that $X^{\tau_n} \in \mathcal{M}_c^{\mathrm{loc}}$.

▷ To see why X^{τ_n} is bounded, note that

$$|X_t^{\tau_n}| = |X_{t\wedge\tau_n}| \leq \int_0^{t\wedge\tau_n} |\,\mathrm{d}X_s| \leq n, \qquad t \geq 0.$$

Note also that since $X \in \mathcal{V}_c$, also $X^{\tau_n} \in \mathcal{V}_c$ is a finite-variation process and $V_t^{X^{\tau_n}} = \int_0^t |\,\mathrm{d}X_s^{\tau_n}| \leq n$.

Step 2. Fix $n \in \mathbb{N}$. Next, we apply the Sophomore's dream trick from Corollary 7.20 to a partition $0 = s_0 < s_1 < \cdots < s_m = t$ of the time interval $[0,t]$ with fixed $t > 0$ to obtain

$$\begin{aligned}\mathbb{E}\big[(X_t^{\tau_n})^2\big] &= \sum_{j=1}^m \mathbb{E}\Big[\big(X_{s_j}^{\tau_n} - X_{s_{j-1}}^{\tau_n}\big)^2\Big] \qquad \text{[using Corollary 7.20 and } X_{s_0}^{\tau_n} = X_0^{\tau_n} = 0\text{]}\\ &= \mathbb{E}\Bigg[O_{s_0,s_1,\ldots,s_m}^{(n)} \sum_{j=1}^m |X_{s_j}^{\tau_n} - X_{s_{j-1}}^{\tau_n}|\Bigg],\end{aligned}$$

where we estimated one term in the square by the oscillation

$$O_{s_0,s_1,\ldots,s_m}^{(n)} := \sup_{1\leq j\leq m} |X_{s_j}^{\tau_n} - X_{s_{j-1}}^{\tau_n}| \leq 2n < \infty \qquad \text{for each fixed } n \in \mathbb{N},$$

which is uniformly bounded for each fixed n since $|X_s^{\tau_n}| \leq n$ for all $s \geq 0$. By applying (8.4) from Lemma 8.6 to $X \in \mathcal{V}_c$ (for which, in particular, $X \in \mathrm{FV}$ a.s.), we see that

$$\sum_{j=1}^m |X_{s_j}^{\tau_n} - X_{s_{j-1}}^{\tau_n}| \leq \mathrm{var}_X(t) =: V_t^{X^{\tau_n}} \leq n \qquad \text{for each fixed } n \in \mathbb{N},$$

which is also uniformly bounded for each fixed n. This shows that

$$\mathbb{E}\big[(X_t^{\tau_n})^2\big] \leq \mathbb{E}\Bigg[O_{s_0,s_1,\ldots,s_m}^{(n)} \sum_{j=1}^m |X_{s_j}^{\tau_n} - X_{s_{j-1}}^{\tau_n}|\Bigg] \leq \mathbb{E}\big[\underbrace{O_{s_0,s_1,\ldots,s_m}^{(n)} V_t^{X^{\tau_n}}}_{\leq 2n^2}\big].$$

Step 3. Because X is (a.s.) continuous, it is (a.s.) uniformly continuous on the compact time interval $[0,t]$, so taking finer and finer partitions, we have

$$O_{s_0,s_1,\ldots,s_m}^{(n)} \overset{m\to\infty}{\longrightarrow} 0 \qquad \text{almost surely, for each fixed } n \in \mathbb{N}.$$

Together with Step 2, this shows that for each fixed $n \in \mathbb{N}$, we have

$$\mathbb{E}\big[(X_t^{\tau_n})^2\big] \leq \mathbb{E}\big[O_{s_0,s_1,\ldots,s_m}^{(n)} V_t^{X^{\tau_n}}\big] \overset{m\to\infty}{\longrightarrow} 0, \tag{8.12}$$

applying the Bounded Convergence Theorem [Kyt20, Corollary VII.21]. Since the right-hand side of (8.12) does not depend on n, this implies that $\mathbb{E}\,[(X_t^{\tau_n})^2] = 0$, so

$$\mathbb{P}\,[X_t^{\tau_n} = 0] = 1, \qquad t \geq 0.$$

Step 4. Since X has finite variation, we have $\tau_n \uparrow \infty$ as $n \uparrow \infty$. Thus, Step 3 shows that

$$X_t = \lim_{n\to\infty} X_t^{\tau_n} = 0 \qquad \text{almost surely, for each fixed } t \geq 0.$$

In other words, we have $\mathbb{P}\,[X_t = 0] = 1$ for any $t \geq 0$.

Step 5. Lastly, it remains to argue that $X_t = 0$ almost surely *simultaneously* for all $t \geq 0$. To this end, we take a countable dense set of times and use Union Bound (A.4):

$$\mathbb{P}\left[X_q = 0 \text{ for all } q \in \mathbb{Q} \cap [0,\infty)\right] \;\geq\; 1 \;-\; \sum_{q \in \mathbb{Q}\cap[0,\infty)} \underbrace{\mathbb{P}\left[X_q \neq 0\right]}_{=0 \text{ by Step 4}} \;=\; 1.$$

Because X is (a.s.) continuous, similarly as in the proof of Kolmogorov's Continuity Criterion (Theorem 1.25), there exists a modification of X which is identically zero, so

$$\mathbb{P}\left[X_t = 0 \text{ for all } t \geq 0\right] = 1.$$

This concludes the proof. □

Corollary 8.30. *Brownian motion is not a finite-variation process.*

Proof. Brownian motion is a non-trivial (non-zero) continuous martingale (thus, a local martingale). Hence, Theorem 8.28 shows that it cannot be a finite-variation process. □

8.3 Quadratic variation process for local martingales

While non-trivial local martingales have infinite total variation, they have finite *quadratic variation.* This is the correction needed in order to derive a version of the "Fundamental Theorem of Calculus," or *Itô's Formula*, for stochastic integrals, to be discussed more in Section 10.

8.3.1 Motivation — stochastic integration by parts

Let us consider the following stochastic integration by parts for a finite-variation process.

Lemma 8.31. *Let $A \in \mathcal{V}_c$ be a finite-variation process. We have*

$$A_t^2 = 2\int_0^t A_s \,\mathrm{d}A_s, \qquad t \geq 0.$$

Proof sketch. Consider the following telescoping sum:

$$\begin{aligned} A^2_{2^{-n}\lfloor 2^n t\rfloor} &= \sum_{k=1}^{\lfloor 2^n t\rfloor} \left(A^2_{k2^{-n}} - A^2_{(k-1)2^{-n}}\right) \\ &= \sum_{k=1}^{\lfloor 2^n t\rfloor} 2\,A_{(k-1)2^{-n}}\left(A_{k2^{-n}} - A_{(k-1)2^{-n}}\right) + \sum_{k=1}^{\lfloor 2^n t\rfloor}\left(A_{k2^{-n}} - A_{(k-1)2^{-n}}\right)^2. \end{aligned}$$

Taking $n \to \infty$, by continuity the left-hand side tends to A_t^2 almost surely (and thus, in probability), while the first term on the right-hand side is a Riemann sum tending to

$$2\int_0^t A_s \,\mathrm{d}A_s,$$

in probability (cf. Exercise 8.32), and the second term on the right-hand side tends to zero in probability, because A has finite variation (cf. Exercise 8.33). □

Exercise 8.32. Let $A \in \mathcal{V}_c$ be a finite-variation process. Fix $t \geq 0$. Show that

$$\sum_{k=1}^{\lfloor 2^n t \rfloor} 2\, A_{(k-1)2^{-n}} \big(A_{k2^{-n}} - A_{(k-1)2^{-n}} \big) \quad \xrightarrow{\mathbb{P}} \quad 2 \int_0^t A_s \, \mathrm{d}A_s .$$

Exercise 8.33. Let $A \in \mathcal{V}_c$ be a finite-variation process. Fix $t \geq 0$. Show that

$$\sum_{k=1}^{\lfloor 2^n t \rfloor} \big(A_{k2^{-n}} - A_{(k-1)2^{-n}} \big)^2 \quad \xrightarrow{\mathbb{P}} \quad 0.$$

Consider now a continuous local martingale $X \in \mathcal{M}_c^{\mathrm{loc}}$. We would still expect to be able to define a stochastic integral (which we will indeed define in Section 9) that we would denote as

$$\int_0^t X_s \, \mathrm{d}X_s, \qquad t \geq 0.$$

However, if we believe that the integral should still respect the Riemann sum approximation,

$$\sum_{k=1}^{\lfloor 2^n t \rfloor} 2\, X_{(k-1)2^{-n}} \big(X_{k2^{-n}} - X_{(k-1)2^{-n}} \big) \quad \xrightarrow{\mathbb{P}} \quad 2 \int_0^t X_s \, \mathrm{d}X_s ,$$

we see that in the integration by parts formula, there is a possibly non-trivial second order error term, namely the limit (in a suitable sense, see Definition 8.34 below) as $n \to \infty$ of

$$\sum_{k=1}^{\lfloor 2^n t \rfloor} \big(X_{k2^{-n}} - X_{(k-1)2^{-n}} \big)^2 .$$

Evaluating this limit motivates the definition of the quadratic variation process, see (8.14) in Theorem 8.35. Thus, we expect the following type of integration by parts to hold for $X \in \mathcal{M}_c^{\mathrm{loc}}$:

$$X_t^2 \; = \; 2 \int_0^t X_s \, \mathrm{d}X_s \; + \; \lim_{n \to \infty} \sum_{k=1}^{\lfloor 2^n t \rfloor} \big(X_{k2^{-n}} - X_{(k-1)2^{-n}} \big)^2, \qquad t \geq 0. \tag{8.13}$$

While the limits in Exercises 8.32 & 8.33 for a finite-variation process hold in probability, the limit in (8.13) turns out to hold in a weaker sense, as detailed in Definition 8.34 & Theorem 8.35.

Definition 8.34. A sequence $(X^{(n)})_{n \in \mathbb{N}}$ of real-valued continuous-time stochastic processes converges *in probability uniformly on compacts* to the process X if

$$\lim_{n \to \infty} \mathbb{P}\Big[\sup_{t \in [0,T]} | X_t^{(n)} - X_t | \; > \; \varepsilon \Big] = 0 \qquad \text{for all } \varepsilon > 0 \text{ and for all } T > 0.$$

We will see in Section 10 that the integration by parts formula (8.13) indeed holds with this second order correction — see in particular Theorem 10.1. This is the gist of Itô's integration theory. See also Proposition 9.14 & Exercise 9.16 for the Riemann sum approximations.

8.3.2 Definition of quadratic variation

Recall from Definition 8.15 that we call (continuous) non-decreasing finite-variation processes just simply *increasing processes*, and we denote them as

$$\mathcal{V}_c^+ := \big\{ X = (X_t)_{t \geq 0} \mid X \text{ is a continuous increasing process, } X_0 = 0 \big\} \; \subset \; \mathcal{V}_c .$$

By the convention $X_0 = 0$, any (continuous) increasing process is non-negative.

Theorem 8.35. (Quadratic variation) *Let $X \in \mathcal{M}_c^{\text{loc}}$. There exists a unique process (up to indistinguishability) denoted $\langle X, X\rangle = (\langle X, X\rangle_t)_{t\geq 0}$ such that the following hold:*

QV1. $\langle X, X\rangle \in \mathcal{V}_c^+$ *is a (continuous)* increasing *process*;

QV2. $t \mapsto X_t^2 - \langle X, X\rangle_t$ *is a continuous local martingale, i.e., it belongs to $\mathcal{M}_c^{\text{loc}}$.*

Moreover, the following convergence of processes holds:

$$\sum_{k=1}^{\lfloor 2^n t\rfloor} \big(X_{k2^{-n}} - X_{(k-1)2^{-n}}\big)^2 \quad \xrightarrow[n\to\infty]{\mathbb{P}} \quad \langle X, X\rangle_t, \qquad t \geq 0, \tag{8.14}$$

in probability uniformly on compacts (*cf. Definition 8.34), and in particular, in probability pointwise in time.*

Definition 8.36. The process $\langle X, X\rangle$ is called the *quadratic variation* of X.

Let us make some quick remarks:

- ▷ Note that the process $\langle X, X\rangle$ is independent of the initial value X_0, but only depends on the increments of X. Thus, it is often useful to assume that $X_0 = 0$.
- ▷ If M is a martingale, then $M^2 - \langle M, M\rangle$ is a martingale (cf. Proposition 8.38 and Lemma 8.43).
- ▷ Brownian motion has $\langle B, B\rangle_t = t$, and $t \mapsto B_t^2 - t$ is a martingale (cf. item 1 of Exercise 6.4).

The proof of Theorem 8.35 is very elaborate.

- ▷ The uniqueness part follows from Theorem 8.28 (see Lemma 8.37 in Section 8.3.3).
- ▷ Proving the existence of the quadratic variation $\langle X, X\rangle$ will be done in several stages.
- ▷ First, one can construct it for *bounded* continuous martingales making use the Hilbert space structure of $\mathcal{M}_c^2$ from Section 7.2. (This will be the content of Section 8.3.4.)
- ▷ With the construction done for bounded continuous martingales, a *localization* (that is, truncation by suitable stopping times and a limiting procedure) argument gives the construction of $\langle X, X\rangle$ for general $X \in \mathcal{M}_c^{\text{loc}}$. (This will be the content of Section 8.3.5.)

During first reading, readers are advised to still read the next Section 8.3.3 (uniqueness), and thereafter skip Sections 8.3.4–8.3.5 (existence) and jump directly to Sections 8.3.7 for now. It is recommended to rather come back to Sections 8.3.4–8.3.5 after reading Section 9.

8.3.3 Uniqueness of the quadratic variation

Lemma 8.37. *Let A and $\widetilde{A}$ be two processes satisfying properties* QV1 *and* QV2*. Then, A and $\widetilde{A}$ are indistinguishable:* $\mathbb{P}[\widetilde{A}_t = A_t$ for all $t \geq 0] = 1$.

Proof. Consider the difference $A - \widetilde{A}$. We aim to show that it is the zero process.

▷ Since both $A \in \mathcal{V}_c$ and $\widetilde{A} \in \mathcal{V}_c$ are finite-variation processes by property QV1, so is $A - \widetilde{A} \in \mathcal{V}_c$ (this follows by linearity, Exercise 8.12).

▷ Since both $X^2 - A$ and $X^2 - \widetilde{A}$ are continuous local martingales by property QV2, so is

$$A - \widetilde{A} = \underbrace{(X^2 - \widetilde{A})}_{\in \mathcal{M}_c^{\mathrm{loc}}} - \underbrace{(X^2 - A)}_{\in \mathcal{M}_c^{\mathrm{loc}}} \quad \in \quad \mathcal{M}_c^{\mathrm{loc}}$$

(this follows by linearity, Exercise 8.22).

Hence, Theorem 8.28 applied to $A - \widetilde{A}$ shows that, almost surely, $\widetilde{A}_t = A_t$ for all $t \geq 0$. □

8.3.4 Existence of the quadratic variation for bounded continuous martingales (⋆)

Terminology. We call a process X which is a.s. uniformly bounded, i.e.,

$$\mathbb{P}\Big[\{\omega \in \Omega \mid \text{there exists a constant } C \in (0,\infty) \text{ such that } |X_t| \leq C \text{ for all } t \geq 0\}\Big] = 1,$$

briefly a *bounded process.* As we work with processes up to indistinguishability and assume that the filtration $\mathcal{F}_\bullet$ satisfies the usual conditions, we omit "almost surely (a.s.)."

The set of all (uniformly) *bounded continuous martingales* on $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ is denoted as

$$\begin{aligned}\mathcal{M}_c^b \; &:= \; \Big\{M = (M_t)_{t\geq 0} \in \mathcal{M}_c \mid M \text{ is uniformly bounded: } \exists\, C \in (0,\infty) \text{ s.t. } \sup_{t\geq 0} |M_t| \;\leq\; C\Big\} \\ &\subset \; \mathcal{M}_c^2 \;\subset\; \mathcal{M}_c.\end{aligned}$$

Proposition 8.38. *Suppose that $M \in \mathcal{M}_c^b$ is a bounded continuous martingale. Then, there exists a process $\langle M, M\rangle = (\langle M, M\rangle_t)_{t\geq 0}$ such that the following hold:*

QV'1. $\langle M, M\rangle \in \mathcal{V}_c^+$ *is a (continuous)* increasing *process;*

QV'2. $M^2 - \langle M, M\rangle \in \mathcal{M}_c^b$ *is a bounded continuous martingale.*

In particular, the process $\langle M, M\rangle$ is the quadratic variation of M.

To construct the process $\langle M, M\rangle$, we use the definition and properties of simple (elementary) integral processes from Section 9.2 — in brief, they are integrals over piecewise constant processes. As a piece of intuition, we would expect that approximation by simple integrals would yield a stochastic integral $\int_0^t M_s \, \mathrm{d}M_s$ for M, and the quadratic variation process $\langle M, M\rangle$ would arise as a correction term in the integration by parts for M as in (8.13):

$$M_t^2 \;=\; 2\int_0^t M_s \, \mathrm{d}M_s \;+\; \lim_{n\to\infty} \sum_{k=1}^{\lfloor 2^n t\rfloor} \big(M_{k2^{-n}} - M_{(k-1)2^{-n}}\big)^2, \qquad t \geq 0.$$

Thus, along the proof of Proposition 8.38 we shall not only construct $\langle M, M\rangle$ but also verify:

Corollary 8.39. *Suppose that $M \in \mathcal{M}_c^b$ is a bounded continuous martingale, and let $\langle M, M\rangle$ be its quadratic variation. Then, the following convergence of processes holds:*

$$\sum_{k=1}^{\lfloor 2^n t\rfloor} \big(M_{k2^{-n}} - M_{(k-1)2^{-n}}\big)^2 \quad \xrightarrow[n\to\infty]{\mathbb{P}} \quad \langle M, M\rangle_t, \qquad t \geq 0, \tag{8.15}$$

in probability uniformly on compacts (*cf. Definition 8.34), and in particular, in probability pointwise in time.*

Proof. This follows from Step 2 of the proof of Proposition 8.38. □

Proof sketch of Proposition 8.38. Without loss of generality, we consider the case where $M_0 = 0$. Fix $N \in \mathbb{N}$ throughout. Consider the *dyadic partitions* of the time interval $[0, N]$,

$$\{k\,2^{-n} \mid k \in \{0, 1, \ldots, 2^n N\}\}, \qquad n \in \mathbb{N}.$$

- ▷ Define for each $n \in \mathbb{N}$ the *simple processes approximating M* (see also Definition 9.4) as

  $$H_t^{(n,N)}(\omega) = \sum_{k=1}^{2^n N} M_{(k-1)2^{-n}}(\omega)\, \mathbb{1}_{((k-1)2^{-n}, k2^{-n}]}(t), \qquad t \geq 0,$$

  where each $M_{(k-1)2^{-n}} : \Omega \to \mathbb{R}$ is a bounded $\mathcal{F}_{(k-1)2^{-n}}$-measurable random variable, since M is a bounded martingale. Note that by Proposition 5.19, since $t \mapsto H_t^{(n,N)}$ is left-continuous, it is progressively measurable — in particular, adapted.

- ▷ Define for each $n \in \mathbb{N}$ the *simple integral process* (see also Definition 9.5) as

  $$\begin{aligned} Y_0^{(n,N)}(\omega) &:= (H^{(n,N)} \bullet M)_0(\omega) := 0, \\ Y_t^{(n,N)}(\omega) &:= (H^{(n,N)} \bullet M)_t(\omega) \\ &:= \sum_{k=1}^{2^n N} M_{(k-1)2^{-n}}(\omega)\big(M_{t\wedge k2^{-n}}(\omega) - M_{t\wedge(k-1)2^{-n}}(\omega)\big), \quad t > 0,\ \omega \in \Omega. \end{aligned} \tag{8.16}$$

  By Lemma 9.6 (Exercise 9.7), we have $Y^{(n,N)} \in \mathcal{M}_c^2$. Remembering the motivation for the search of the quadratic variation from the discussion after Lemma 8.31, we *expect* that

  $$Y_t^{(n,N)} \quad \overset{n\to\infty}{\longrightarrow} \quad \int_0^t M_s \,\mathrm{d}M_s, \qquad t \in [0, N], \tag{8.17}$$

  and that the error in the integration by parts formula, which can be written in the form $A_t^{(n,N)} := M_t^2 - 2\,Y_t^{(n,N)}$, converges to the desired quadratic variation:

  $$A_t^{(n,N)} \quad \overset{n\to\infty}{\longrightarrow} \quad \langle M, M\rangle_t, \qquad t \in [0, N], \tag{8.18}$$

  in a suitable sense (namely, the integral (8.17) will converge in the space $\mathcal{M}_c^2$ of martingales, which is a Hilbert space by Theorem 7.17 — this will become clearer later in the proof of Theorem 9.12, and we do not need it here — and the error (8.18) converges in probability (uniformly on compacts) as in Corollary 8.39, which will follow from the proof below).

Step 1. Lemma 8.40 shows that, for each fixed $N \in \mathbb{N}$, the sequence $(Y^{(n,N)})_{n \in \mathbb{N}}$ is a *Cauchy-sequence* in $\mathcal{M}_c^2$. Hence, by the completeness of $(\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$ from Theorem 7.16, we have

$$Y^{(n,N)} \quad \overset{n\to\infty}{\longrightarrow} \quad Y^{(N)} \;\in\; \mathcal{M}_c^2,$$

that is, $\| Y^{(n,N)} - Y^{(N)} \|_{\mathcal{M}_c^2} = \| Y_\infty^{(n,N)} - Y_\infty^{(N)} \|_{L^2} \to 0$ as $n \to \infty$.

Lemma 8.40. *Fix $N \in \mathbb{N}$ and consider the sequence $(Y^{(n,N)})_{n \in \mathbb{N}}$ in $\mathcal{M}_c^2$ defined in (8.16). It is a Cauchy-sequence:*

$$\lim_{m\to\infty} \sup_{n,n' \geq m} \| Y^{(n,N)} - Y^{(n',N)} \|_{\mathcal{M}_c^2} \;=\; 0. \tag{8.19}$$

We prove Lemma 8.40 in the end of this section.

Step 2. Next, we define a process $A_t^{(N)} := M_t^2 - 2\,Y_t^{(N)}$ for $t \in [0, N]$. We will argue that it satisfies properties QV'1 and QV'2 on the time window $t \in [0, N]$ (and later take $N \to \infty$).

- ▷ By Step 1, we have $M^2 - A^{(N)} = 2Y^{(N)} \in \mathcal{M}_c^b$, which shows property QV'2.
- ▷ Moreover, since

$$\begin{aligned} M_{k2^{-n}}^2 - 2\,Y_{k2^{-n}}^{(n,N)} \;&=\; \sum_{j=1}^{k} \big(M_{j2^{-n}}^2 - M_{(j-1)2^{-n}}^2 \big) \;-\; 2 \sum_{j=1}^{k} M_{(j-1)2^{-n}} \big(M_{j2^{-n}} - M_{(j-1)2^{-n}} \big) \\ &=\; \sum_{j=1}^{k} \big(M_{j2^{-n}} - M_{(j-1)2^{-n}} \big)^2 \end{aligned}$$

  is non-decreasing in $k = 1, 2, \ldots$, one can check that[19] the continuous process

$$t \;\mapsto\; M_t^2 - 2\,Y_t^{(N)} =: A_t^{(N)}$$

  is also a non-negative, non-decreasing process started at $M_0^2 - 2Y_0^{(N)} = 0$. Hence, it is an *increasing process* by Lemma 8.14 and Definition 8.15 — this shows property QV'1.

Step 3. Taking $N \to \infty$ along the integers $\mathbb{N}$ and defining

$$\langle M, M \rangle_t(\omega) \quad := \quad \begin{cases} \lim\limits_{\substack{N\to\infty \\ N \in \mathbb{N}}} A_t^{(N)}(\omega), & t \geq 0,\ \omega \in E^A, \\ 0 & t \geq 0,\ \omega \notin E^A, \end{cases}$$

where $\mathbb{P}[E^A] = 1$ for the event (*countable* union of almost sure events)

$$E^A := \big\{ \omega \in \Omega \mid t \mapsto A_t^{(N)}(\omega) \text{ is an increasing process for any } N \in \mathbb{N} \big\},$$

we see that the process $\langle M, M \rangle$ indeed satisfies the desired properties QV'1 and QV'2. □

Proof sketch of Lemma 8.40. Suppose that $n \leq n'$ and consider the mixed moment at the fixed "final" time $t = N$:

$$\mathbb{E}\big[Y_N^{(n,N)} Y_N^{(n',N)} \big] := \sum_{k=1}^{2^n N} \sum_{\ell=1}^{2^{n'} N} \mathbb{E}\Big[M_{(k-1)2^{-n}} \big(M_{k2^{-n}} - M_{(k-1)2^{-n}} \big)\, M_{(\ell-1)2^{-n'}} \big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}} \big) \Big].$$

There are two possible cases for the two dyadic intervals

$$I_k^n := [(k-1)2^{-n}, k2^{-n}) \qquad \text{and} \qquad I_\ell^{n'} := [(\ell-1)2^{-n'}, \ell 2^{-n'}).$$

[19] For more details, see [LeG16, Proof of Theorem 4.9].

▷ Either $I_k^n \cap I_\ell^{n'} = \varnothing$, in which case using the Orthogonality trick from Lemma 9.8, we obtain

$$\mathbb{E}\Big[M_{(k-1)2^{-n}}\big(M_{k2^{-n}} - M_{(k-1)2^{-n}}\big)\, M_{(\ell-1)2^{-n'}}\big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big)\Big] = 0,$$

since $0 \le (k-1)2^{-n} < k2^{-n} < (\ell-1)2^{-n'} < \ell 2^{-n'}$ and we have $M_{(k-1)2^{-n}} \in \mathrm{m}\mathcal{F}_{(k-1)2^{-n}}$ and $M_{(\ell-1)2^{-n'}} \in \mathrm{m}\mathcal{F}_{(\ell-1)2^{-n'}}$.

▷ Or $I_k^n \cap I_\ell^{n'} \neq \varnothing$, in which case $I_\ell^{n'} \subset I_k^n$. It is useful to write

$$M_{k2^{-n}} - M_{(k-1)2^{-n}} = \big(M_{k2^{-n}} - M_{\ell 2^{-n'}}\big) + \big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big) + \big(M_{(\ell-1)2^{-n'}} - M_{(k-1)2^{-n}}\big)$$

so that after another application of the Orthogonality trick from Lemma 9.8 with times $0 \le (k-1)2^{-n} \le (\ell-1)2^{-n'} < \ell 2^{-n'} \le k2^{-n}$, the terms of interest take the following form:

$$\begin{aligned}
&\mathbb{E}\Big[M_{(k-1)2^{-n}}\big(M_{k2^{-n}} - M_{(k-1)2^{-n}}\big)\, M_{(\ell-1)2^{-n'}}\big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big)\Big]\\
&= \underbrace{\mathbb{E}\Big[M_{(k-1)2^{-n}}\big(M_{k2^{-n}} - M_{\ell 2^{-n'}}\big)\, M_{(\ell-1)2^{-n'}}\big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big)\Big]}_{=0 \text{ by Lemma 9.8}}\\
&\quad + \mathbb{E}\Big[M_{(k-1)2^{-n}}\big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big)\, M_{(\ell-1)2^{-n'}}\big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big)\Big]\\
&\quad + \underbrace{\mathbb{E}\Big[M_{(k-1)2^{-n}}\big(M_{(\ell-1)2^{-n'}} - M_{(k-1)2^{-n}}\big)\, M_{(\ell-1)2^{-n'}}\big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big)\Big]}_{=0 \text{ by Lemma 9.8}}\\
&= \mathbb{E}\Big[M_{(k-1)2^{-n}}\, M_{(\ell-1)2^{-n'}}\big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big)^2\Big].
\end{aligned}$$

Lastly, note that for each fixed $\ell \in \{1, 2, \ldots, 2^{n'}N\}$, if $I_k^n \cap I_\ell^{n'} \neq \varnothing$ holds, then it holds with a *unique* $k = k(\ell) \in \{1, 2, \ldots, 2^n N\}$. Hence, we obtain

$$\begin{aligned}
&\mathbb{E}\big[(Y_N^{(n,N)} - Y_N^{(n',N)})^2\big]\\
&= \sum_{\ell=1}^{2^{n'}N} \mathbb{E}\Big[\big(M_{(k(\ell)-1)2^{-n}} - M_{(\ell-1)2^{-n'}}\big)^2 \big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big)^2\Big]\\
&\le \mathbb{E}\Bigg[\Bigg(\sup_{\ell\in\{1,\ldots,2^{n'}N\}} \big(M_{(k(\ell)-1)2^{-n}} - M_{(\ell-1)2^{-n'}}\big)^2 \Bigg(\sum_{\ell=1}^{2^{n'}N} \big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big)^2\Bigg)\Bigg]\\
&\le \Bigg(\underbrace{\mathbb{E}\Bigg[\Bigg(\sup_{\ell\in\{1,\ldots,2^{n'}N\}} \big(M_{(k(\ell)-1)2^{-n}} - M_{(\ell-1)2^{-n'}}\big)^2\Bigg)^2\Bigg]}_{\to 0 \text{ since } M \text{ continuous } \& \ [0,N] \text{ compact}}\Bigg)^{1/2} \Bigg(\underbrace{\mathbb{E}\Bigg[\Bigg(\sum_{\ell=1}^{2^{n'}N} \big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big)^2\Bigg)^2\Bigg]}_{\text{bounded uniformly in } n,n' \text{ by Exercise 8.41}}\Bigg)^{1/2}\\
&\overset{n,n'\to\infty}{\longrightarrow} \quad 0,
\end{aligned}$$

using Cauchy-Schwarz Inequality (Lemma 2.17) and Exercise 8.41. □

Exercise 8.41. Let $M \in \mathcal{M}_c^b$ and let $C \in (0,\infty)$ be a constant such that $\sup_{t\ge 0} |M_t| \le C$. Show that

$$\mathbb{E}\Bigg[\Bigg(\sum_{\ell=1}^{2^{n'}N} \big(M_{\ell 2^{-n'}} - M_{(\ell-1)2^{-n'}}\big)^2\Bigg)^2\Bigg] \le 16\, C^4.$$

8.3.5 Existence of the quadratic variation for local martingales ($\star$)

Proposition 8.42. *Let $X \in \mathcal{M}_c^{\text{loc}}$. Then, there exists a process $\langle X, X\rangle$ satisfying properties* QV1 *and* QV2. *In particular, the process $\langle X, X\rangle$ is the quadratic variation of X. Moreover, the convergence* (8.14) *holds in probability uniformly on compacts.*

Proof sketch. The idea is *localization*: up to suitable increasing stopping times, X is a bounded martingale for which Proposition 8.38 holds. Define

$$\tau_n := \inf\{t \geq 0 \mid |X_t| \geq n\}, \qquad n \in \mathbb{N},$$

which are stopping times by Lemma 5.7, as hitting times of the continuous process X to the closed set $[n, \infty)$. The stopped process $X^{\tau_n} := (X_{t\wedge\tau_n})_{t\geq 0}$ is a continuous bounded local martingale:

$$|X_t^{\tau_n}| = |X_{t\wedge\tau_n}| \leq |X_0| + n, \qquad n \in \mathbb{N},\ t \geq 0,$$

thus, it is a *continuous bounded martingale* (by Exercise 8.25). Therefore, we know by Proposition 8.38, the quadratic variation process $\langle X^{\tau_n}, X^{\tau_n}\rangle$ exists for each $n \in \mathbb{N}$.

It remains to take the limit $n \to \infty$. Since X is continuous, we have $\tau_n \uparrow \infty$ as $n \uparrow \infty$. Now,

$$\langle X, X\rangle := \lim_{n\to\infty} \langle X^{\tau_n}, X^{\tau_n}\rangle$$

is well-defined by consistency, and it has all the sought properties (Exercise: check this!). □

This concludes the sketch proof of Theorem 8.35.

8.3.6 Quadratic variation of uniformly L^2-bounded martingales ($\star$)

The next lemma will be useful when using the Hilbert space $L^2(M)$ (cf. Definition 9.1) to construct a stochastic integral w.r.t. a continuous L^2-bounded martingale $M \in \mathcal{M}_c^2$ in Section 9.3.

Lemma 8.43. *Let $M \in \mathcal{M}_c^2$. The process $M^2 - \langle M, M\rangle$ is a continuous UI martingale and*

$$\mathbb{E}[\langle M, M\rangle_\infty] = \mathbb{E}[(M_\infty - M_0)^2] < \infty. \tag{8.20}$$

Proof. By property QV1, we know that $\langle M, M\rangle \in \mathcal{V}_c^+$ is an increasing process, so its long-term limit $\langle M, M\rangle_\infty := \lim_{t\to\infty} \langle M, M\rangle_t$ exists almost surely in $[0, +\infty]$. To evaluate it, consider

$$\tau_n := \inf\{t \geq 0 \mid \langle M, M\rangle_t \geq n\}, \qquad n \in \mathbb{N},$$

(which, as before, are stopping times by Lemma 5.7, as hitting times of the continuous process $\langle M, M\rangle$ to the closed set $[n, \infty)$). Since $t \mapsto \langle M, M\rangle_t$ is increasing, we have $\tau_n \uparrow \infty$ a.s.

As before, we note that

- ▷ the stopped process $(M^2 - \langle M, M\rangle)^{\tau_n} \in \mathcal{M}_c^{\text{loc}}$ is a local martingale for each fixed $n \in \mathbb{N}$ (cf. Exercise 8.26), since $M^2 - \langle M, M\rangle \in \mathcal{M}_c^{\text{loc}}$ by property QV2;

▷ since $M \in \mathcal{M}_c^2$ (so $|||M|||_{\sup} < \infty$ by assumption), we also have

$$|(M^2 - \langle M,M\rangle)_t^{\tau_n}| = |M_{t\wedge\tau_n}^2 - \langle M,M\rangle_{t\wedge\tau_n}| \le \underbrace{\sup_{t\ge 0} M_t^2}_{\le |||M|||_{\sup}} + n < \infty, \qquad n\in\mathbb{N},\ t\ge 0,$$

so $(M^2 - \langle M,M\rangle)^{\tau_n}$ is bounded and thus a UI martingale (cf. Exercise 8.25).

Now we can use Optional Stopping Theorem for UI martingales (Theorem 7.5) to conclude that

$$\mathbb{E}\big[M_{\tau_n}^2 - \langle M,M\rangle_{\tau_n}\big] = \mathbb{E}\big[M_0^2 - \underbrace{\langle M,M\rangle_0}_{=0}\big] = \mathbb{E}\big[M_0^2\big]. \tag{8.21}$$

Taking $n\to\infty$, we obtain

$$\begin{aligned}
\mathbb{E}[M_\infty^2] = \mathbb{E}\Big[\lim_{n\to\infty} M_{\tau_n}^2\Big] &= \lim_{n\to\infty}\mathbb{E}\big[M_{\tau_n}^2\big] && \text{[by Dominated Convergence Theorem]}\\
&= \mathbb{E}[M_0^2] + \lim_{n\to\infty}\mathbb{E}[\langle M,M\rangle_{\tau_n}] && \text{[by (8.21)]}\\
&= \mathbb{E}[M_0^2] + \mathbb{E}\Big[\lim_{n\to\infty}\langle M,M\rangle_{\tau_n}\Big] && \text{[by Monotone Convergence Theorem]}\\
&= \mathbb{E}[M_0^2] + \mathbb{E}[\langle M,M\rangle_\infty].
\end{aligned}$$

Using the Sophomore's dream trick from (7.13) in Lemma 7.18, we can write this as

$$\mathbb{E}[\langle M,M\rangle_\infty] = \mathbb{E}[M_\infty^2] - \mathbb{E}[M_0^2] = \mathbb{E}[(M_\infty - M_0)^2] < \infty,$$

which proves (8.20). It remains to argue that $M^2 - \langle M,M\rangle \in \mathcal{M}_c^{\mathrm{loc}}$ is a continuous UI martingale:

$$\begin{aligned}
|M_t^2 - \langle M,M\rangle_t| &\le \underbrace{\sup_{t\ge0} M_t^2}_{\le |||M|||_{\sup}} + \underbrace{\langle M,M\rangle_t}_{\le \langle M,M\rangle_\infty} && \text{[since } t\mapsto\langle M,M\rangle_t \text{ is increasing]}\\
&\le |||M|||_{\sup} + \langle M,M\rangle_\infty \in L^1(\mathbb{P}), \qquad t\ge0, && \text{[by (8.20)]}
\end{aligned}$$

so Exercise 8.25 shows that $M^2 - \langle M,M\rangle$ is indeed a continuous UI martingale. □

8.3.7 Quadratic covariation

Proposition 8.44. (Quadratic covariation) *For $X,Y\in\mathcal{M}_c^{\mathrm{loc}}$, there exists a unique process (up to indistinguishability) denoted $\langle X,Y\rangle = (\langle X,Y\rangle_t)_{t\ge0}$ such that the following hold:*

QCV1. $\langle X,Y\rangle \in \mathcal{V}_c$ *is a (continuous) finite-variation process;*

QCV2. $t\mapsto X_tY_t - \langle X,Y\rangle_t$ *is a continuous local martingale, i.e., it belongs to $\mathcal{M}_c^{\mathrm{loc}}$.*

Moreover, the following convergence of processes holds:

$$\sum_{k=1}^{\lfloor 2^n t\rfloor}\big(X_{k2^{-n}} - X_{(k-1)2^{-n}}\big)\big(Y_{k2^{-n}} - Y_{(k-1)2^{-n}}\big) \xrightarrow[n\to\infty]{\mathbb{P}} \langle X,Y\rangle_t, \qquad t\ge0, \tag{8.22}$$

in probability uniformly on compacts (*cf. Definition 8.34*), *and in particular, in probability pointwise in time.*

Proof sketch. We know from Theorem 8.35 and linearity (Exercise 8.22) that the local martingales $X\pm Y\in\mathcal{M}_c^{\mathrm{loc}}$ have quadratic variations $\langle X\pm Y, X\pm Y\rangle\in\mathcal{V}_c^+$.

▷ The process $\langle X,Y\rangle$ can be constructed as

$$\langle X,Y\rangle := \tfrac{1}{4}\big(Q^{X+Y} - Q^{X-Y}\big),$$

where $Q^{X\pm Y} := \langle X\pm Y, X\pm Y\rangle$. Then, it follows from properties QV1 & QV2 of the quadratic variation that $\langle X,Y\rangle$ satisfies the sought properties QCV1 & QCV2 (cf. Exercise 8.45).

▷ The uniqueness of $\langle X,Y\rangle$ is the content of Exercise 8.46.

▷ The convergence (8.22) is the content of Exercise 8.47. □

Exercise 8.45. Let $X,Y \in \mathcal{M}_c^{\text{loc}}$, and define $Q_t^{X+Y} := \langle X+Y, X+Y\rangle_t$ and $Q_t^{X-Y} := \langle X-Y, X-Y\rangle_t$, and set

$$\langle X,Y\rangle_t := \tfrac{1}{4}\big(Q_t^{X+Y} - Q_t^{X-Y}\big), \qquad t \geq 0.$$

Show that $\langle X,Y\rangle$ satisfies properties QCV1 and QCV2 in Proposition 8.44.

Exercise 8.46. Let $X,Y \in \mathcal{M}_c^{\text{loc}}$. Prove that their covariation process $\langle X,Y\rangle$ is unique up to indistinguishability.

Exercise 8.47. Let $X,Y \in \mathcal{M}_c^{\text{loc}}$, and let $\langle X,Y\rangle$ be their covariation process. Prove the convergence (8.22).

Exercise 8.48. Prove that the assignment $(X,Y) \mapsto \langle X,Y\rangle$ is a $\mathbb{R}$-bilinear and symmetric map $\mathcal{M}_c^{\text{loc}} \times \mathcal{M}_c^{\text{loc}} \to \mathcal{V}_c$.

Exercise 8.49. Let $A \in \mathcal{V}_c$ be a continuous finite-variation process and X a continuous stochastic process. Prove that the following convergence holds almost surely (yielding that $\langle A,X\rangle = 0$, as discussed below):

$$\sum_{k=1}^{\lfloor 2^n t\rfloor} (A_{k2^{-n}} - A_{(k-1)2^{-n}})(X_{k2^{-n}} - X_{(k-1)2^{-n}}) \quad \overset{n\to\infty}{\longrightarrow} \quad 0.$$

Thanks to Exercise 8.49, we can also define the quadratic variation and covariation involving any continuous finite-variation process to equal zero: for example,

$$\langle A,C\rangle := 0 \qquad \text{and} \qquad \langle A,X\rangle := 0 =: \langle X,A\rangle \qquad \text{for any } A,C \in \mathcal{V}_c \text{ and } X \in \mathcal{M}_c^{\text{loc}}.$$

This extends by linearity the definition of quadratic variation and covariation $\langle\cdot,\cdot\rangle$ to an $\mathbb{R}$- bilinear pairing for all continuous semimartingales from Definition 8.1: for a pair of semimartingales

$$\begin{cases} Y = Y_0 + X + A, \\ Z = Z_0 + N + C \end{cases}, \qquad X_0 = A_0 = 0 = N_0 = C_0, \quad Y_0, Z_0 \in \mathrm{m}\mathcal{F}_0$$

where $X,N \in \mathcal{M}_c^{\text{loc}}$ are the local martingale parts and $A,C \in \mathcal{V}_c$ the finite-variation parts, we set

$$\begin{aligned} \langle Y,Z\rangle &= \langle Y_0 + X + A, Z_0 + N + C\rangle = \langle X+A, N+C\rangle && (8.23)\\ &:= \langle X,N\rangle + \underbrace{\langle X,C\rangle}_{=0} + \underbrace{\langle A,N\rangle}_{=0} + \underbrace{\langle A,C\rangle}_{=0} = \langle X,N\rangle && (8.24) \end{aligned}$$

Exercise 8.50. Let $X, Y \in \mathcal{M}_c^{\mathrm{loc}}$, and let $\langle X, Y \rangle$ be their covariation process. Let τ be a stopping time and consider the stopped processes $X^\tau = (X_{t\wedge\tau})_{t\geq 0}$ and $Y^\tau = (Y_{t\wedge\tau})_{t\geq 0}$. Prove that we have

$$\langle X^\tau, Y \rangle_t = \langle X, Y^\tau \rangle_t = \langle X^\tau, Y^\tau \rangle_t = \langle X, Y \rangle_{t\wedge\tau}, \qquad t \geq 0.$$

Exercise 8.51. Consider Brownian motion in the Euclidean space $\mathbb{R}^n$, which is defined as the vector-valued process $\vec{B}_t = (B_t^{(1)}, \dots, B_t^{(n)})$, $t \geq 0$, where the components are independent one-dimensional Brownian motions. Calculate the quadratic covariation processes $t \mapsto \langle B^{(i)}, B^{(j)} \rangle_t$ for $i, j \in \{1, \dots, n\}$.

9 Stochastic integral

Consider still a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$ satisfying the *usual conditions*. It is possible to define a general *stochastic integral* of a progressively measurable and locally bounded process H with respect to a continuous local martingale $X \in \mathcal{M}_c^{\mathrm{loc}}$ via the following strategy.

- ▷ First, define a stochastic integral for *simple* (or *elementary*) processes H, which are bounded progressively measurable processes taking *finitely* many values (see Definition 9.4), with respect to L^2-bounded martingales $X = M \in \mathcal{M}_c^2$ (see Definition 9.5).
- ▷ Next, *approximate* a more general process H by simple processes $H^{(n)}$ in a suitable sense:

$$H^{(n)} \quad \overset{n\to\infty}{\longrightarrow} \quad H,$$

  where the convergence holds in a certain Hilbert space (see Proposition 9.11).

- ▷ Then, show that for $X = M \in \mathcal{M}_c^2$, the integrals of the simple processes,

$$\Big(\int_0^t H_s^{(n)} \, \mathrm{d}M_s \Big)_{n\in\mathbb{N}}, \qquad t \geq 0, \tag{9.1}$$

  form a *Cauchy-sequence* in the space $\mathcal{M}_c^2$ of all uniformly L^2-bounded continuous martingales. Because $\mathcal{M}_c^2$ is a Hilbert space (by Theorem 7.16), the Cauchy-sequence (9.1) must converge to some process in $\mathcal{M}_c^2$, which *by definition* is the *integral of H with respect to M*:

$$\Big(\int_0^t H_s^{(n)} \, \mathrm{d}M_s \Big)_{t\geq 0} \quad \overset{n\to\infty}{\longrightarrow} \quad \Big(\int_0^t H_s \, \mathrm{d}M_s \Big)_{t\geq 0} =: \big((H \bullet M)_t \big)_{t\geq 0}.$$

  (See Theorem 9.12.) Note that this is a weaker notion than the pathwise integral.

- ▷ Lastly, an integral process $H \bullet X$ can be defined for any progressively measurable locally bounded H and for any local martingale $X \in \mathcal{M}_c^{\mathrm{loc}}$ via a *localization* (that is, truncation by suitable stopping times and a limiting procedure) argument (see Section 9.4).

Without mentioning it explicitly, we shall consider all processes *up to indistinguishability*, that is, we identify all processes $\widetilde{Y}$ and Y that agree almost surely at all times (cf. Definition 1.20)

$$\mathbb{P}[\widetilde{Y}_t = Y_t \text{ for all } t \geq 0] = 1.$$

A fortiori, such processes are defined $\mathbb{P}$-almost everywhere.

9.1 Stochastic integral w.r.t. quadratic variation and the space $L^2(M)$

By Theorem 8.35, any martingale $M \in \mathcal{M}_c^2$ has a unique quadratic variation process $\langle M, M\rangle \in \mathcal{V}_c^+$, which importantly is an *increasing finite-variation process.* Hence, by Proposition 8.17 from Section 8.1, we know that the pathwise Lebesgue-Stieltjes integral

$$(H \bullet M)_t := \int_0^t H_s \, \mathrm{d}\langle M, M\rangle_s, \qquad t \geq 0,$$

is well-defined $\mathbb{P}$-almost everywhere for any progressively measurable process H such that[20]

$$\int_0^t |H_s| \, \mathrm{d}\langle M, M\rangle_s < \infty \qquad \text{almost surely for all } t \geq 0.$$

[20]Note that since $\langle M, M\rangle$ is an increasing process, we have $|\,\mathrm{d}\langle M, M\rangle\,| = \mathrm{d}\langle M, M\rangle$.

Moreover, in this case $H \bullet M \in \mathcal{V}_c$ is a (continuous) finite-variation process as well. It turns out that the following Hilbert space $L^2(M)$ built from the integral w.r.t. the measure $\mathrm{d}\langle M, M\rangle$ is useful for the construction of more general stochastic integrals.

Definition 9.1. (L^2-space w.r.t. L^2-martingale) Fix a uniformly L^2-bounded continuous martingale $M \in \mathcal{M}_c^2$. Define a norm

$$\| H \|_{L^2(M)} := \left(\mathbb{E}\Big[\int_0^\infty H_s^2 \,\mathrm{d}\langle M, M\rangle_s\Big]\right)^{1/2} \in [0, +\infty], \tag{9.2}$$

and define the real vector space

$$L^2(M) := \big\{ H = (H_t)_{t\geq 0} \,|\, H \text{ is progressively measurable and } \|H\|_{L^2(M)} < \infty \big\},$$

whose elements are identified up to the equivalence relation induced by identifying[a] elements with the same value of $\| \cdot \|_{L^2(M)}$:

$$\begin{aligned} H = 0 \in L^2(M) \quad &\iff \quad \| H \|_{L^2(M)} = 0 \\ &\iff \quad \int_0^\infty H_s^2 \,\mathrm{d}\langle M, M\rangle_s = 0 \quad \mathbb{P}\text{-a.s.} \\ &\iff \quad H_t \text{ is zero } \mathbb{P}\text{-a.s. almost everywhere in } [0,\infty) \ni t \\ &\qquad\quad \text{w.r.t. the quadratic variation measure of } M. \end{aligned}$$

[a]Strictly speaking $L^2(M)$ should be defined as the quotient space up to this equivalence relation.

With the above identifications, the space $L^2(M)$ also has an inner product,

$$(H, K)_{L^2(M)} := \mathbb{E}\Big[\int_0^\infty H_s K_s \,\mathrm{d}\langle M, M\rangle_s\Big], \qquad H, K \in L^2(M). \tag{9.3}$$

Importantly, the space $L^2(M)$ is a Hilbert space, see Theorem 9.3.

Exercise 9.2. Prove that (9.3) is an inner product and (9.2) a norm on $L^2(M)$ (as in Definition 2.12).

For each $M \in \mathcal{M}_c^2$, the space $L^2(M)$ is an L^2-space with respect to the *finite measure*

$$\mathrm{d}\mu_M := \mathrm{d}\mathbb{P} \otimes \mathrm{d}\langle M, M\rangle$$

on the measurable space $(\Omega \times [0,\infty), \mathscr{P})$. The measure of a set $G \in \mathscr{P}$ is

$$\mu_M[G] \;=\; \mathbb{E}\Big[\int_0^\infty \mathbb{1}_G(\omega, s) \,\mathrm{d}\langle M, M\rangle_s\Big] \;=\; \int_\Omega \int_0^\infty \mathbb{1}_G(\omega, s) \,\mathrm{d}\langle M, M\rangle_s \,\mathrm{d}\mathbb{P}(\omega).$$

The *total mass* of μ_M is

$$\mu_M[\Omega \times [0,\infty)] \;=\; \mathbb{E}\Big[\int_0^\infty \mathrm{d}\langle M, M\rangle_s\Big] \;=\; \mathbb{E}[\langle M, M\rangle_\infty],$$

which is finite by Lemma 8.43.

Theorem 9.3. *The space $L^2(M)$ is a Hilbert space with inner product* (9.3). *In particular, $(L^2(M), \| \cdot \|_{L^2(M)})$ is complete.*

Proof idea. Exercise 9.2 verifies the structures for $L^2(M)$ as an inner product space and as a normed vector space. To show that $(L^2(M), \|\cdot\|_{L^2(M)})$ is complete (cf. Definition 2.16), a very similar proof as for Theorem 2.20 works. Indeed, note that the proof of Theorem 2.20 only uses *measure theoretic* tools, which hold for any finite measure. Thus, after replacing in Theorem 2.20

- ▷ the probability measure $\mathrm{d}\mathbb{P}$ by the finite measure $\mathrm{d}\mu_M := \mathrm{d}\mathbb{P} \otimes \mathrm{d}\langle M, M\rangle$;
- ▷ the norm $\|\cdot\|_{L^2}$ by the norm $\|\cdot\|_{L^2(M)}$;
- ▷ and the norm $\|\cdot\|_{L^1} := \mathbb{E}[|\cdot|]$ by the norm

$$\|H\|_{L^1(M)} := \mathbb{E}\Big[\int_0^\infty |H_s|\,\mathrm{d}\langle M, M\rangle_s\Big],$$

and noting that both the Cauchy-Schwarz Inequality (Lemma 2.17) and Lemma 2.21 hold on these spaces as well (with very similar proofs), the same arguments go through

to show that any Cauchy-sequence in $(L^2(M), \|\cdot\|_{L^2(M)})$ converges in $(L^2(M), \|\cdot\|_{L^2(M)})$. We leave the details as an exercise for readers interested in measure theory. □

9.2 Stochastic integrals of simple processes

Similarly to the development of Lebesgue integral, the first step towards developing a general stochastic integral is to make sense of it for piecewise constant (simple) functions.

9.2.1 Integrals of simple processes

Definition 9.4. $H : (0,\infty) \times \Omega \to \mathbb{R}$ is a *simple* (*or elementary*) *process* if it has the form

$$H_t(\omega) = \sum_{k=1}^{n} \xi_k(\omega)\, \mathbb{1}_{(t_{k-1}, t_k]}(t), \tag{9.4}$$

where $0 = t_0 < t_1 < \cdots < t_n$ and each $\xi_k : \Omega \to \mathbb{R}$ is a bounded $\mathcal{F}_{t_{k-1}}$-measurable r.v.[a]

[a] Processes of the form (9.4) are also often called *predictable* — cf. Remark 8.18.

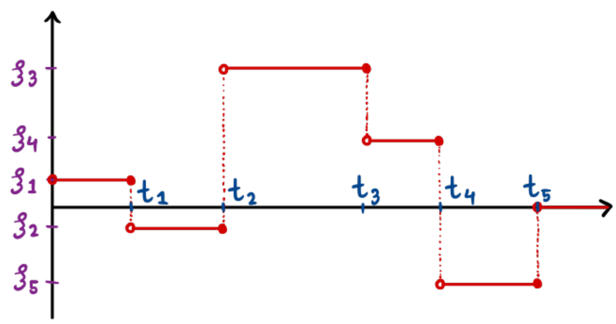


Illustration of a simple (elementary) process in Definition 9.4.

Definition 9.5. (Simple stochastic integral) Let $M \in \mathcal{M}_c^2$, and let H be a simple process as in (9.4). The *stochastic integral of H with respect to M* is defined as

$$(H \bullet M)_t(\omega) := \sum_{k=1}^{n} \xi_k(\omega)\big(M_{t\wedge t_k}(\omega) - M_{t\wedge t_{k-1}}(\omega)\big), \qquad t \geq 0,\ \omega \in \Omega. \tag{9.5}$$

One can think of (9.5) as a left Riemann sum. Note that it is important that one takes the random variables $\xi_k \in \mathcal{F}_{t_{k-1}}$ giving the values of H to be *measurable at the left endpoints* of the time intervals $(t_{k-1}, t_k]$. This will be evident in arguments using the martingale property of M.

Lemma 9.6. *Let $M \in \mathcal{M}_c^2$, and let H be a simple process as in* (9.4)*. Then, the stochastic integral* (9.5) *is a continuous uniformly L^2-bounded martingale: $H \bullet M \in \mathcal{M}_c^2$.*

Exercise 9.7. Prove Lemma 9.6. *Hint*:

- ▷ *Continuity follows from the continuity of M.*
- ▷ *The martingale property is analogous to the proof of Proposition 3.11.*
- ▷ *To prove that $H \bullet M$ is uniformly L^2-bounded, you can use its definition, expand it into a double sum, note that by Lemma 9.8 only the diagonal terms survive, and finally use the Sophomore's dream trick (Lemma 7.18) to obtain a telescoping sum. The L^2-boundedness then follows from the L^2-boundedness of M and the boundedness of the values ξ_k in* (9.4)*.*

Lemma 9.8. (Orthogonality trick) *On a filtered probability space $(\Omega, \mathscr{F}, \mathcal{F}_\bullet, \mathbb{P})$, consider a martingale M, two bounded random variables ξ, η, and times $0 \le r \le u \le s \le t$ such that $\xi \in \mathrm{m}\mathcal{F}_r$ and $\eta \in \mathrm{m}\mathcal{F}_s$. Then, we have*

$$\mathbb{E}\big[\xi\,(M_u - M_r)\,\eta\,(M_t - M_s)\big] = 0. \tag{9.6}$$

Proof. Note that $\xi\,(M_u - M_r)\,\eta \in \mathrm{m}\mathcal{F}_s$. Hence, using the Tower Property (item 4 of Lemma 2.8), we see that

$$\begin{aligned}\mathbb{E}\big[\xi\,(M_u - M_r)\,\eta\,(M_t - M_s)\big] &= \mathbb{E}\big[\mathbb{E}\big[\xi\,(M_u - M_r)\,\eta\,(M_t - M_s)\,|\,\mathcal{F}_s\big]\big] && \text{[by item 4 of Lemma 2.8]}\\ &= \mathbb{E}\big[\xi\,(M_u - M_r)\,\eta\,\mathbb{E}\big[M_t - M_s\,|\,\mathcal{F}_s\big]\big]. && \text{[by item 5 of Lemma 2.8]}\end{aligned}$$

The inner conditional expected value vanishes by item 1 of Lemma 2.8, linearity, and the martingale property MG3 of M:

$$\mathbb{E}\big[M_t - M_s\,|\,\mathcal{F}_s\big] = \mathbb{E}\big[M_t\,|\,\mathcal{F}_s\big] - M_s = 0.$$

This shows (9.6). □

9.2.2 Itô's isometry for simple integrals

Fix $M \in \mathcal{M}_c^2$. Recall the Hilbert space structure of $(\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$ from Section 7.2. The next result shows that stochastic integral of simple processes H (Definitions 9.4 and 9.5) with respect to M forms an *isometry* between $(L^2(M), \|\cdot\|_{L^2(M)})$ and $(\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$:

$$(L^2(M), \|\cdot\|_{L^2(M)}) \longrightarrow (\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2}), \qquad H \mapsto H \bullet M.$$

Proposition 9.9. (Isometry) *Let $M \in \mathcal{M}_c^2$, and let H be a simple process as in* (9.4)*. Then, the stochastic integral* (9.5) *satisfies the* isometry property

$$\|H \bullet M\|_{\mathcal{M}_c^2} := \mathbb{E}\big[(H \bullet M)_\infty^2\big] = \|H\|_{L^2(M)}^2. \tag{9.7}$$

Proof. Since $\xi_k \in \mathcal{F}_{t_{k-1}}$, we see that

$$
\begin{aligned}
\mathbb{E}\big[(H\bullet M)_\infty^2\big] &= \mathbb{E}\big[(H\bullet M)_{t_n}^2\big] && [(H\bullet M)_t \text{ constant for all } t\geq t_n]\\
&= \sum_{k,\ell=1}^{n} \mathbb{E}\big[\xi_k\big(M_{t_k}-M_{t_{k-1}}\big)\,\xi_\ell\big(M_{t_\ell}-M_{t_{\ell-1}}\big)\big]\\
&= \sum_{k=1}^{n} \delta_{k,\ell}\,\mathbb{E}\big[\xi_k\big(M_{t_k}-M_{t_{k-1}}\big)\,\xi_\ell\big(M_{t_\ell}-M_{t_{\ell-1}}\big)\big] && [\text{by Lemma 9.8}]\\
&= \sum_{k=1}^{n} \mathbb{E}\big[\xi_k^2\big(M_{t_k}-M_{t_{k-1}}\big)^2\big]\\
&= \sum_{k=1}^{n} \mathbb{E}\big[\mathbb{E}\big[\xi_k^2\big(M_{t_k}-M_{t_{k-1}}\big)^2\,|\,\mathcal{F}_{t_{k-1}}\big]\big] && [\text{by item 4 of Lemma 2.8}]\\
&= \sum_{k=1}^{n} \mathbb{E}\big[\xi_k^2\,\mathbb{E}\big[\big(M_{t_k}-M_{t_{k-1}}\big)^2\,|\,\mathcal{F}_{t_{k-1}}\big]\big] && [\text{by item 5 of Lemma 2.8}]\\
&= \sum_{k=1}^{n} \mathbb{E}\big[\xi_k^2\,\mathbb{E}\big[\big(M_{t_k}^2-M_{t_{k-1}}^2\big)\,|\,\mathcal{F}_{t_{k-1}}\big]\big] && [\text{by Sophomore's dream, Lemma 7.18}]\\
&= \sum_{k=1}^{n} \mathbb{E}\big[\xi_k^2\big(\langle M,M\rangle_{t_k}-\langle M,M\rangle_{t_{k-1}}\big)\big] && [\text{by Lemma 8.43}]\\
&= \sum_{k=1}^{n} \mathbb{E}\Big[\int_{t_{k-1}}^{t_k}\xi_k^2\,\mathrm{d}\langle M,M\rangle_s\Big] && [\xi_k^2 \text{ is constant}]\\
&= \mathbb{E}\Big[\int_0^\infty H_s^2\,\mathrm{d}\langle M,M\rangle_s\Big] = \|H\|_{L^2(M)}^2.
\end{aligned}
$$

□

Since $H\bullet M\in\mathcal{M}_c^2$ by Lemma 9.6, the left-hand side of the identity (9.7) is finite. Hence:

Corollary 9.10. *Let $M\in\mathcal{M}_c^2$. Then, any simple process H as in* (9.4) *belongs to $L^2(M)$:*

$$\mathcal{S} := \big\{H=(H_t)_{t\geq 0}\,|\,H \text{ is simple as in } (9.4)\big\} \quad\subset\quad L^2(M).$$

Moreover, any bounded[a] progressively measurable process H belongs to $L^2(M)$.

[a] There exists $C<\infty$ such that $|H_t|\leq C$ for all $t\geq 0$ almost surely.

Proof. This follows from Proposition 9.9 with the observation that if $H_s\leq C$ almost surely for all $s\geq 0$, then using Lemma 8.43, we have

$$\int_0^\infty \underbrace{H_s^2}_{\leq C^2}\,\mathrm{d}\langle M,M\rangle_s \;\leq\; C^2\int_0^\infty \mathrm{d}\langle M,M\rangle_s \;\leq\; C^2\,\langle M,M\rangle_\infty<\infty \quad \text{almost surely.}$$

□

In Proposition 9.11 we will show that for each $M\in\mathcal{M}_c^2$, the space $\mathcal{S}\subset L^2(M)$ dense. This enables the construction of stochastic integrals as limits of simple ones (see Theorem 9.12).

9.2.3 Simple processes are dense

We shall next discuss the crucial fact that, for each $M\in\mathcal{M}_c^2$, the space $\mathcal{S}\subset L^2(M)$ of simple (elementary) processes is a dense vector subspace of $L^2(M)$. Thus, we can approximate general processes in $L^2(M)$ by simple ones.

Proposition 9.11. (Density) *Fix $M \in \mathcal{M}_c^2$. Then, the space of simple processes,*

$$(\mathcal{S}, \|\cdot\|_{L^2(M)}) \quad \subset \quad (L^2(M), \|\cdot\|_{L^2(M)}),$$

is a dense subspace of $(L^2(M), \|\cdot\|_{L^2(M)})$.

We give two sketch proofs for the density of $\mathcal{S}$ in $L^2(M)$. The first one has a functional analysis flavor — see [LeG16, Proposition 5.3] for more details. The second one is an adaptation of an analytic approximation argument that also works in the special case of Brownian motion — see [MP10, Chapter 7] for this alternative approach.

Proof sketch 1. Since $L^2(M)$ is a Hilbert space by Theorem 9.3, to prove the density

$$\overline{\mathcal{S}} = L^2(M),$$

it suffices to prove (cf. Exercise 2.14) that the orthogonal complement $\overline{\mathcal{S}}^{\perp} = \{0\}$ w.r.t. the inner product (9.3) is zero. To this end, we consider $K \in L^2(M)$ such that $(H, K)_{L^2(M)} = 0$ for all $H \in \mathcal{S}$, and we aim to prove that $K = 0$ (up to the equivalence relation in Definition 9.1):

$$K = 0 \in L^2(M) \qquad \Longleftrightarrow \qquad \begin{array}{l} K_t \text{ is zero } \mathbb{P}\text{-a.s. almost everywhere in } [0,\infty) \ni t \\ \text{w.r.t. the quadratic variation measure of } M. \end{array}$$

For this, it suffices to prove that

$$\int_0^\infty K_s \,\mathrm{d}\langle M, M\rangle_s = 0 \quad \mathbb{P}\text{-a.s.}$$

Claim. We will prove the stronger statement that the process $X = (X_t)_{t \geq 0}$,

$$X_t \;:=\; (K \bullet \langle M, M\rangle)_t \;:=\; \int_0^t K_s \,\mathrm{d}\langle M, M\rangle_s, \qquad t \geq 0,$$

is indistinguishable of the zero process: $X = 0$, $\mathbb{P}$-a.s.

Let us first record that X is well-defined. Using Cauchy-Schwarz Inequality (Lemma 2.17) for the L^2-space $L^2(M)$ and the corresponding L^1-space, we see that

$$\begin{aligned} \mathbb{E}\Big[\int_0^t |K_s| \,\mathrm{d}\langle M, M\rangle_s\Big] &\leq \bigg(\mathbb{E}\Big[\int_0^t K_s^2 \,\mathrm{d}\langle M, M\rangle_s\Big]\bigg)^{1/2} \big(\mu_M[\Omega \times [0,t]]\big)^{1/2} \\ &\leq \underbrace{\bigg(\mathbb{E}\Big[\int_0^\infty K_s^2 \,\mathrm{d}\langle M, M\rangle_s\Big]\bigg)^{1/2}}_{= \|K\|_{L^2(M)} < \infty} \big(\underbrace{\mu_M[\Omega \times [0,\infty)]}_{= \mathbb{E}[\langle M, M\rangle_\infty] < \infty}\big)^{1/2} \;<\; \infty, \qquad t \geq 0. \end{aligned}$$

(Note that the total mass of μ_M is finite by Lemma 8.43.)

We may conclude in the following manner.

- ▷ On the one hand, Proposition 8.17 shows that $X \in \mathcal{V}_c$ is a finite-variation process.
- ▷ On the other hand, if we show that $X \in \mathcal{M}_c^{\mathrm{loc}}$, then Theorem 8.28 implies that $X = 0$, $\mathbb{P}$-a.s.

In fact, $X = K \bullet \langle M, M\rangle$ is a martingale. Let us check the properties MG1–MG3.

- ▷ (MG1): For every $t \geq 0$, the random variable $X_t \;:=\; (K \bullet \langle M, M\rangle)_t$ is $\mathcal{F}_t$-measurable, since the process $K \bullet \langle M, M\rangle$ is adapted by Proposition 8.17.

▷ (MG2): For every $t \geq 0$, the random variable $X_t \in L^1(\mathbb{P})$ is integrable:

$$\mathbb{E}[|X_t|] \leq \mathbb{E}\Big[\int_0^t |K_s| \, \mathrm{d}\langle M,M\rangle_s\Big] < \infty;$$

▷ (MG3): To argue that X has the martingale property (MG3), that is, $\mathbb{E}[X_t \,|\, \mathcal{F}_s] = X_s$ almost surely for every $0 \leq s < t$, we can use property CE3 in Definition 2.1 of the conditional expected value $\mathbb{E}[X_t \,|\, \mathcal{F}_s]$. Indeed, note that for any $E \in \mathcal{F}_s$, we have

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_E (X_t - X_s)] &:= \mathbb{E}[\mathbb{1}_E \big((K \bullet \langle M,M\rangle)_t - (K \bullet \langle M,M\rangle)_s\big)] \\
&:= \mathbb{E}\Big[\mathbb{1}_E \int_s^t K_r \, \mathrm{d}\langle M,M\rangle_r\Big] \\
&= \big(\mathbb{1}_E \mathbb{1}_{(s,t]}, K\big)_{L^2(M)} = 0,
\end{aligned}$$

using the assumed orthogonality of K with the simple process $\mathbb{1}_E \mathbb{1}_{(s,t]} \in \mathcal{S}$. Then, after using the Tower Property to condition on $\mathcal{F}_s$ on both sides, we find that

$$\begin{aligned}
0 &= \mathbb{E}\big[\mathbb{E}[\mathbb{1}_E (X_t - X_s) \,|\, \mathcal{F}_s]\big] && \text{[by item 4 of Lemma 2.8]} \\
&= \mathbb{E}\big[\mathbb{E}[\mathbb{1}_E X_t \,|\, \mathcal{F}_s] - \mathbb{E}[\mathbb{1}_E X_s \,|\, \mathcal{F}_s]\big] && \text{[by linearity]} \\
&= \mathbb{E}\big[\mathbb{1}_E \big(\mathbb{E}[X_t \,|\, \mathcal{F}_s] - X_s\big)\big], && \text{[by item 5 of Lemma 2.8, as } \mathbb{1}_E, X_s \in \mathrm{m}\mathcal{F}_s]
\end{aligned}$$

which shows that $\mathbb{E}[X_t \,|\, \mathcal{F}_s] = X_s$ by property CE3 and linearity. This shows MG3.

We conclude that X is both a martingale and a finite-variation process, thus zero. □

Proof sketch 2. By Corollary 9.10, all *bounded* progressively measurable processes belong to $L^2(M)$. Moreover, any process $K \in L^2(M)$ can be approximated by such processes: *truncations*

$$(\omega, t) \longmapsto K_t(\omega) \, \mathbb{1}_{\{|K_t(\omega)| \leq n\}}(\omega, t)$$

are bounded and converge to K in $(L^2(M), \|\cdot\|_{L^2(M)})$:

$$\| K - K \mathbb{1}_{\{|K| \leq n\}} \|_{L^2(M)}^2 = \| K \|_{L^2(M)}^2 \, \mathbb{E}\Big[\int_0^\infty \mathbb{1}_{\{|K| > n\}} \, \mathrm{d}\langle M,M\rangle_s\Big] \overset{n\to\infty}{\longrightarrow} 0,$$

applying Bounded Convergence Theorem (BCT) [Kyt20, Corollary VII.21] to the finite measure $\mathrm{d}\mu_M := \mathrm{d}\mathbb{P} \otimes \mathrm{d}\langle M,M\rangle$. Hence (by the triangle inequality), to prove that the simple processes $(\mathcal{S}, \|\cdot\|_{L^2(M)})$ form a dense subspace in $(L^2(M), \|\cdot\|_{L^2(M)})$, it suffices to prove that

any *bounded* progressively measurable process can be approximated by *simple* processes in the Hilbert space $(L^2(M), \|\cdot\|_{L^2(M)})$.

▷ For a bounded *continuous* $H \in L^2(M)$, define the simple processes $H^{(n)} \in \mathcal{S}$ as

$$H_t^{(n)}(\omega) := \sum_{k=1}^{n2^n} H_{t_{k-1}^{(n)}}(\omega) \, \mathbb{1}_{(t_{k-1}^{(n)}, t_k^{(n)}]}(t), \qquad t \geq 0.$$

where $t_j^{(n)} := (j-1)2^{-j}$ are dyadic times, $j \in \{0, 1, \ldots, n2^n\}$, and $H_{t_{k-1}^{(n)}} \in \mathrm{m}\mathcal{F}_{t_{k-1}^{(n)}}$ because H is progressively measurable (hence adapted).

Then, again applying BCT [Kyt20, Corollary VII.21] gives

$$\| H^{(n)} - H \|_{L^2(M)}^2 = \mathbb{E}\Big[\int_0^\infty \Big(\sum_{k=1}^{n2^n} H_{t_{k-1}^{(n)}} \mathbb{1}_{(t_{k-1}^{(n)}, t_k^{(n)}]}(s) - H_s\Big)^2 \mathrm{d}\langle M,M\rangle_s\Big]$$

$$
\begin{aligned}
&\le \mathbb{E}\Big[\int_0^\infty \Big(\sum_{k=1}^{n2^n} \big(H_{t_{k-1}^{(n)}} - H_s\big)\,\mathbb{1}_{(t_{k-1}^{(n)},t_k^{(n)}]}(s)\Big)^2 \,\mathrm{d}\langle M,M\rangle_s\Big] \\
&\quad + 2\,\mathbb{E}\Big[\int_0^\infty \Big(|H_s| \underbrace{\sum_{k=1}^{n2^n} \mathbb{1}_{[n2^{-n},\infty)}(s)\,\mathbb{1}_{(t_{k-1}^{(n)},t_k^{(n)}]}(s)}_{=0} |H_{t_{k-1}^{(n)}} - H_s|\Big)\,\mathrm{d}\langle M,M\rangle_s\Big] \\
&\quad + \underbrace{\mathbb{E}\Big[\int_0^\infty |H_s|^2\,\mathbb{1}_{[n2^{-n},\infty)}(s)\,\mathrm{d}\langle M,M\rangle_s\Big]}_{\overset{n\to\infty}{\longrightarrow}\,0\ \text{by BCT}},
\end{aligned}
$$

where the only term that potentially survives in the limit $n \to \infty$ equals

$$
\begin{aligned}
&\mathbb{E}\Big[\int_0^\infty \Big(\sum_{k=1}^{n2^n} \big(H_{t_{k-1}^{(n)}} - H_s\big)\,\mathbb{1}_{(t_{k-1}^{(n)},t_k^{(n)}]}(s)\Big)^2 \,\mathrm{d}\langle M,M\rangle_s\Big] \\
&= \mathbb{E}\Big[\int_0^\infty \sum_{k,\ell=1}^{n2^n} \big(H_{t_{k-1}^{(n)}} - H_s\big)\big(H_{t_{\ell-1}^{(n)}} - H_s\big)\,\underbrace{\mathbb{1}_{(t_{k-1}^{(n)},t_k^{(n)}]}(s)\,\mathbb{1}_{(t_{\ell-1}^{(n)},t_\ell^{(n)}]}(s)}_{=\delta_{k,\ell}\,\mathbb{1}_{(t_{k-1}^{(n)},t_k^{(n)}]}(s)}\,\mathrm{d}\langle M,M\rangle_s\Big] \\
&= \mathbb{E}\Big[\int_0^\infty \sum_{k=1}^{n2^n} \big(H_{t_{k-1}^{(n)}} - H_s\big)^2\,\mathbb{1}_{(t_{k-1}^{(n)},t_k^{(n)}]}(s)\,\mathrm{d}\langle M,M\rangle_s\Big] \quad \overset{n\to\infty}{\longrightarrow} \quad 0,
\end{aligned}
$$

again by BCT and the continuity of $s \mapsto H_s$.

- ▷ For a *general bounded* process $N \in L^2(M)$, again by the triangle inequality (exercise) it suffices to approximate N by bounded *continuous* processes $N^{(n)} \in L^2(M)$:

$$
N_t^{(n)} := \frac{1}{\langle M,M\rangle_t - \langle M,M\rangle_{t-1/n}} \int_{t-1/n}^t N_s\,\mathrm{d}\langle M,M\rangle_s, \qquad t \ge 0,\ n \in \mathbb{N},
$$

  each of which represents an *average* of N on the time-interval $[t-1/n, t]$ w.r.t. the measure $\mathrm{d}\langle M,M\rangle$. Note that, since N is progressively measurable and $N_t^{(n)}$ only depends on its values for times up to t, the process $N^{(n)}$ is indeed progressively measurable. It is continuous similarly as in the proof of Proposition 8.17. Moreover, applying Dominated Convergence Theorem [Kyt20, Theorem VII.19] to the finite measure $\mathrm{d}\mu_M := \mathrm{d}\mathbb{P} \otimes \mathrm{d}\langle M,M\rangle$, we have

$$
\| N - N^{(n)} \|_{L^2(M)} \overset{n\to\infty}{\longrightarrow} 0,
$$

  because the following convergence holds almost surely for $\mathrm{d}\langle M,M\rangle$-almost every t:

$$
\begin{aligned}
|N_t(\omega) - N_t^{(n)}(\omega)| &\le \frac{1}{\langle M,M\rangle_t(\omega) - \langle M,M\rangle_{t-1/n}(\omega)} \int_{t-1/n}^t |N_t(\omega) - N_s(\omega)|\,\mathrm{d}\langle M,M\rangle_s(\omega) \\
&\overset{n\to\infty}{\longrightarrow} 0,
\end{aligned}
$$

  by*Lebesgue differentiation theorem* for the Radon measure $\mathrm{d}\langle M,M\rangle$ [Kin23, Theorem 4.30].

In conclusion, we have argued that any $K \in L^2(M)$ can be approximated by bounded processes in $L^2(M)$, which in turn can be approximated by simple processes in $L^2(M)$. □

9.3 Stochastic integral w.r.t. L^2-bounded martingales: Itô's isometry

We can now extend the definition of the integral $H \bullet M = \int H\,\mathrm{d}M$ to a general $H \in L^2(M)$ from the dense set of simple processes $\mathcal{S}$. The result is usually referred to as *Itô's isometry.*

Theorem 9.12. (Itô's isometry) *Fix $M \in \mathcal{M}_c^2$. Then, there exists a unique linear map*

$$I_M : (L^2(M), \|\cdot\|_{L^2(M)}) \quad \longrightarrow \quad (\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2}), \qquad H \mapsto I_M(H),$$

such that

ITO1. *for any simple $H \in \mathcal{S}$, we have $I_M(H) = H \bullet M$ (Definition 9.5)*;

ITO2. *for any $H \in L^2(M)$, the following isometry property holds*:

$$\| I_M(H) \|_{\mathcal{M}_c^2} = \| H \|_{L^2(M)}. \tag{9.8}$$

Definition 9.13. , (Stochastic integral) Using Theorem 9.12, we *define* a stochastic integral of any $H \in L^2(M)$ w.r.t. $M \in \mathcal{M}_c^2$ to be the martingale

$$\Big(\int_0^t H_s \, \mathrm{d}M_s \Big)_{t \geq 0} := \big((H \bullet M)_t \big)_{t \geq 0} := \big((I_M(H))_t \big)_{t \geq 0} \quad \in \quad \mathcal{M}_c^2.$$

Proof of Theorem 9.12. The claim already indicates that we should define $I_M(H) = H \bullet M$ for each $H \in \mathcal{S}$. Then, the isometry property (9.8) holds in this case by (9.7) in Proposition 9.9.

Step 1. For a general process $H \in L^2(M)$, by Proposition 9.11 we can pick a sequence of simple processes $H^{(n)} \in \mathcal{S}$ converging to H in $L^2(M)$:

$$\| H^{(n)} - H \|_{L^2(M)} \quad \overset{n\to\infty}{\longrightarrow} \quad 0.$$

Since all convergent sequences are Cauchy, we see that $(H^{(n)})_{n \in \mathbb{N}}$ is a Cauchy-sequence in $L^2(M)$, and the isometry property (9.8) for simple processes shows that $(I_M(H^{(n)}))_{n \in \mathbb{N}}$ is a Cauchy-sequence in $\mathcal{M}_c^2$. By completeness of $(\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$ from Theorem 7.16, it must converge to some element $N \in \mathcal{M}_c^2$. We define

$$I_M(H) := N := \lim_{n\to\infty} I_M(H^{(n)}).$$

Asserted isometry property (9.8) holds for I_M thus defined, by continuity of the norms.

It remains to be shown that the definition of $I_M(H)$ does not depend on choice of the approximating sequence for H, and that the linear functional thus defined is uniquely determined.

Step 2. To show that $I_M(H)$ does not depend on choice of the approximating sequence for H, let us consider two approximating sequences:

$$\| H^{(n)} - H \|_{L^2(M)} \quad \overset{n\to\infty}{\longrightarrow} \quad 0 \qquad \text{and} \qquad \| K^{(n)} - H \|_{L^2(M)} \quad \overset{n\to\infty}{\longrightarrow} \quad 0.$$

Indeed, the isometry property (9.8) and the triangle inequality together imply that

$$\begin{aligned} \| I_M(H^{(n)}) - I_M(K^{(n)}) \|_{\mathcal{M}_c^2} &= \| H^{(n)} - K^{(n)} \|_{L^2(M)} \\ &\leq \underbrace{\| H^{(n)} - H \|_{L^2(M)}}_{\to 0} + \underbrace{\| K^{(n)} - H \|_{L^2(M)}}_{\to 0} \quad \overset{n\to\infty}{\longrightarrow} \quad 0, \end{aligned}$$

which shows that the limits $\lim_{n\to\infty} I_M(H^{(n)})$ and $\lim_{n\to\infty} I_M(K^{(n)})$ are equal in $\mathcal{M}_c^2$.

Step 3. To show that $I_M(H)$ is uniquely determined, note that if

$$I_M, \widetilde{I}_M : (L^2(M), \|\cdot\|_{L^2(M)}) \quad \longrightarrow \quad (\mathcal{M}_c^2, \|\cdot\|_{\mathcal{M}_c^2})$$

are two linear isometries which agree on the dense set $\mathcal{S} \subset L^2(M)$, then by the triangle inequality,

$$\begin{aligned}
&\| I_M(H) - \widetilde{I}_M(H) \|_{\mathcal{M}_c^2} \\
\le\ & \underbrace{\| I_M(H) - I_M(H^{(n)}) \|_{\mathcal{M}_c^2}}_{\to 0} + \underbrace{\| I_M(H^{(n)}) - \widetilde{I}_M(H^{(n)}) \|_{\mathcal{M}_c^2}}_{=0} + \underbrace{\| \widetilde{I}_M(H^{(n)}) - \widetilde{I}_M(H) \|_{\mathcal{M}_c^2}}_{\to 0} \\
\overset{n\to\infty}{\longrightarrow}\ & 0,
\end{aligned}$$

which shows that $I_M(H) = \widetilde{I}_M(H)$ in $\mathcal{M}_c^2$. □

Using Itô's isometry, one can check that the integral thus constructed agrees with the left-Riemann approximation. Note that the right-Riemann approximation will give a different result (cf. Exercise 9.16) — unlike in the case of usual Lebesgue-Stieltjes integrals!

Proposition 9.14. (Riemann sum approximation) *Let $M \in \mathcal{M}_c^2$ and let H be an adapted continuous bounded process. The following convergence of processes holds:*

$$\sum_{k=1}^{\lfloor 2^n t \rfloor} H_{(k-1)2^{-n}} \left(M_{k2^{-n}} - M_{(k-1)2^{-n}} \right) \quad \overset{n\to\infty}{\longrightarrow} \quad \int_0^t H_s \, \mathrm{d}M_s \tag{9.9}$$

in probability uniformly on compacts (cf. Definition 8.34), and in particular, in probability pointwise in time.

Exercise 9.15. Prove Proposition 9.14 via the following steps.

1. For $t \ge 0$, define $t_n := 2^{-n} \lfloor 2^n t \rfloor$ and $H_t^{(n)} := H_{t_n}$ for $n \in \mathbb{N}$. Show that

$$(H^{(n)} \bullet M)_t \;=\; \sum_{k=1}^{\lfloor 2^n t \rfloor} H_{(k-1)2^{-n}} \left(M_{k2^{-n}} - M_{(k-1)2^{-n}} \right) \;+\; H_{t_n} \left(M_t - M_{t_n} \right).$$

2. Show that $\|H^{(n)} - H\|_{L^2(M)} \overset{n\to\infty}{\longrightarrow} 0$.

   Hint: You will need both the Bounded Convergence Theorem for an expression of type $\int_0^\infty K_s \, \mathrm{d}\langle M, M \rangle_s$ (which is a finite measure), and the Dominated Convergence Theorem for the expectation of an integral.

3. Using Itô's isometry (Theorem 9.12) and Doob's L^2-maximal inequality (Proposition 7.14), show that

$$\mathbb{E}\Big[\Big(\sup_{t\ge 0} \big| (H^{(n)} \bullet M)_t - (H \bullet M)_t \big| \Big)^2 \Big] \quad \overset{n\to\infty}{\longrightarrow} \quad 0.$$

4. Using 3 and Markov's Inequality (A.5), show that for any $\varepsilon > 0$, we have

$$\mathbb{P}\Big[\sup_{t\ge 0} \big| (H^{(n)} \bullet M)_t - (H \bullet M)_t \big| > \varepsilon \Big] \quad \overset{n\to\infty}{\longrightarrow} \quad 0.$$

5. Conclude the following convergence in probability uniformly on compacts:

$$\sum_{k=1}^{\lfloor 2^n t \rfloor} H_{(k-1)2^{-n}} \left(M_{k2^{-n}} - M_{(k-1)2^{-n}} \right) \quad \overset{n\to\infty}{\longrightarrow} \quad \int_0^t H_s \, \mathrm{d}M_s.$$

Exercise 9.16. Let $M \in \mathcal{M}_c^2$ and let H be an adapted continuous bounded process. Prove that

$$\sum_{k=1}^{\lfloor 2^n t \rfloor} H_{k2^{-n}} \left(M_{k2^{-n}} - M_{(k-1)2^{-n}} \right) \quad \xrightarrow[n\to\infty]{\mathbb{P}} \quad \int_0^t H_s \, \mathrm{d}M_s \; + \; \langle H, M \rangle_t.$$

Hint: Adapt the proof in Exercise 9.15.

Recall that $H \bullet M \in \mathcal{M}_c^2$ is a martingale for all $M \in \mathcal{M}_c^2$ and $H \in L^2(M)$, so it has a quadratic variation process $\langle H \bullet M, H \bullet M \rangle \in \mathcal{V}_c^+$ by Theorem 8.35.

Proposition 9.17. (Kunita-Watanabe identity) *Let $M \in \mathcal{M}_c^2$ and let H be a bounded*[a] *progressively measurable process. The quadratic variation of $H \bullet M$ is given by*

$$\langle H \bullet M, H \bullet M \rangle_t = \int_0^t H_s^2 \, \mathrm{d}\langle M, M \rangle_s, \qquad t \geq 0.$$

[a] There exists $C < \infty$ such that $|H_t| \leq C$ for all $t \geq 0$ almost surely.

Exercise 9.18. Prove Proposition 9.17. *Hint: Show first that the process*

$$t \quad \mapsto \quad (H \bullet M)_t^2 \; - \; \int_0^t H_s^2 \, \mathrm{d}\langle M, M \rangle_s \tag{9.10}$$

is a continuous martingale. For this, you can use Itô's isometry to verify that for any bounded stopping time τ,

$$\mathbb{E}\big[(H \bullet M)_\tau^2\big] \; = \; \mathbb{E}\Big[\int_0^\tau H_s^2 \, \mathrm{d}\langle M, M \rangle_s\Big],$$

which shows by Theorem 7.3 that (9.10) *is a continuous martingale.*

9.4 Stochastic integral w.r.t. local martingales

The extension of the definition of *Itô stochastic integral*

$$\big((H \bullet M)_t\big)_{t\geq 0} := \Big(\int_0^t H_s \, \mathrm{d}M_s \Big)_{t \geq 0} := \lim_{\substack{n\to\infty \\ \text{in } L^2(M)}} \Big(\int_0^t H_s^{(n)} \, \mathrm{d}M_s \Big)_{t\geq 0}$$

for each progressively measurable process $H \in L^2(M)$ and for each uniformly L^2-bounded martingale $M \in \mathcal{M}_c^2$ to more general integrals

$$\big((H \bullet X)_t\big)_{t\geq 0} := \Big(\int_0^t H_s \, \mathrm{d}X_s \Big)_{t\geq 0}$$

for local martingales $X \in \mathcal{M}_c^{\mathrm{loc}}$ and for nice enough (progressively measurable locally bounded) H can be done via *localization* — truncation of both X and H by suitable stopping times.

Definition 9.19. A process H is *locally bounded* if there exists a sequence of (non-random) constants $(C_n)_{n\in\mathbb{N}}$ and a sequence $(\sigma_n)_{n\in\mathbb{N}}$ of stopping times such that almost surely, $\sigma_n \uparrow \infty$ as $n \uparrow \infty$, and

$$\sup_{t\geq 0} \, |H_t| \, \mathbb{1}_{[0,\sigma_n]}(t) \; \leq \; C_n, \qquad \text{for all } n \in \mathbb{N}.$$

For example, if H is continuous, we can take

$$\sigma_n := \inf\{t \geq 0 \mid |H_t| \geq n\}, \qquad n \in \mathbb{N}.$$

Definition 9.20. (Stochastic integral) Let $X \in \mathcal{M}_c^{\mathrm{loc}}$ and let H be a locally bounded progressively measurable process. We define a stochastic integral

$$\int_0^t H_s \, \mathrm{d}X_s := (H \bullet X)_t := \lim_{n\to\infty} (H^{(n)} \bullet X^{\tau_n})_t, \qquad t \geq 0,$$

where

$$H_t^{(n)} := H_t \, \mathbb{1}_{[0,\tau_n]}(t), \qquad X_t^{\tau_n} := X_{t\wedge\tau_n}, \qquad t \geq 0,$$

and where $\tau_n := \sigma_n \wedge \nu_n$ is the minimum of the localizing stopping times $\sigma_n \uparrow \infty$ for H (Definition 9.19) and the following truncating stopping times for X:

$$\nu_n := \inf\{t \geq 0 \mid |X_t| \geq n\}, \qquad n \in \mathbb{N}.$$

Exercise 9.21. Verify the following facts.

1. $H^{(n)} \in L^2(X^{\tau_n})$ for all $n \in \mathbb{N}$.
2. For each $t \geq 0$, almost surely, the limit

$$(H^{(n)} \bullet X^{\tau_n})_t = \int_0^t H_s^{(n)} \, \mathrm{d}X_s^{\tau_n} \quad \overset{n\to\infty}{\longrightarrow} \quad (H \bullet X)_t$$

exists and is well-defined, by using Lemma 9.22 given below.

Lemma 9.22. *Let $N \in \mathcal{M}_c^2$ and $K \in L^2(N)$. Let τ be a stopping time. The following hold:*

1. $K \, \mathbb{1}_{[0,\tau]} \in L^2(N)$;
2. $K \in L^2(N^\tau)$;
3. $(K \bullet N)^\tau = (K \, \mathbb{1}_{[0,\tau]}) \bullet N = K \bullet N^\tau$.

Proof. The proof is straightforward but rather tedious. See [LeG16, Chapter 5] for details. □

Exercise 9.23. (Distribution of the integral over Brownian motion) Let $f : [0,\infty) \to [0,\infty)$ be a (non-random) locally square-integrable function. Show that, for each $t \geq 0$, the random variable

$$X_t = \int_0^t f(s) \, \mathrm{d}B_s$$

is normally distributed. Find the mean and variance of X_t.
Hint: Consider first $f(s) = \sum_{k=1}^n c_k \mathbf{1}_{[t_{k-1},t_k)}(s)$, with $0 = t_0 < t_1 < \cdots < t_n = t$ and $c_1, \ldots, c_n \in \mathbb{R}$. Approximate the general case with these. What can we say about the limits of Gaussians?

By localization, one can generalize the Kunita-Watanabe identity from Proposition 9.17.

Proposition 9.24. (Kunita-Watanabe identity) *Let $X, Y \in \mathcal{M}_c^{\text{loc}}$ and let H be a locally bounded progressively measurable process. The quadratic covariation of $H \bullet X$ and Y is given by*

$$\langle H \bullet X, Y \rangle_t = \int_0^t H_s \,\mathrm{d}\langle X, Y \rangle_s, \qquad t \geq 0.$$

Exercise 9.25. Prove Proposition 9.24. *Hint*: *Show the claim first for simple processes $H \in \mathcal{S}$ and L^2-bounded martingales $X, Y = M, N \in \mathcal{M}_c^2$. Then extend this by density of $\mathcal{S} \subset L^2(M)$ to general $H \in L^2(M)$, and finally to the general case by localization.*

Proposition 9.26. (Stochastic chain rule) *Let $X \in \mathcal{M}_c^{\text{loc}}$ and let H and K be locally bounded progressively measurable processes. Then,*

$$H \bullet (K \bullet X) = (HK) \bullet X. \tag{9.11}$$

In other words, the "stochastic differential" d is dual to the "stochastic integral" $\int$ in a certain sense. Indeed, writing (9.11) in integral notation, we have

$$\int H \,\mathrm{d} \int K \,\mathrm{d}X = \int (HK) \,\mathrm{d}X.$$

Proof. See [LeG16, Theorem 5.6]. □

10 Itô's Formula and applications

One of the greatest tools in stochastic analysis is the following generalization of the integration by parts formula to stochastic integrals, due to Kiyoshi Itô — Theorem 10.1. It applies to integrals with respect to any continuous *semimartingale* (cf. Definition 8.1):

$$Y = Y_0 + X + A, \qquad X_0 = A_0 = 0,$$

with initial value $Y_0 \in \mathrm{m}\mathcal{F}_0$, where $X \in \mathcal{M}_c^{\mathrm{loc}}$ is a continuous local martingale and $A \in \mathcal{V}_c$ a continuous finite-variation process. In particular, Brownian motion is a continuous semimartingale, since it is a continuous martingale (which has zero finite-variation part).

To set up the stage for Itô's calculus, let us recall what we have accomplished thus far.

- ▷ In Section 8.1, we have defined a stochastic integral w.r.t. any continuous finite-variation process as a pathwise Lebesgue-Stieltjes integral (in particular, in Proposition 8.17).
- ▷ In Sections 9.2–9.4, we have defined a stochastic integral w.r.t. any continuous local martingale, in terms of *approximations* by simple integral processes (see Definition 9.5 and Theorem 9.12 & Definition 9.13) — and a possible *localization* procedure using stopping times (see Definition 9.20). Such construction is a weaker notion than a pathwise integral.

Thus, in summary, for any locally bounded progressively measurable process H, we can define the *stochastic integral of H w.r.t. a semimartingale Y* by linearity:

$$\int_0^t H_s \,\mathrm{d}Y_s \;:=\; \int_0^t H_s \,\mathrm{d}X_s \;+\; \int_0^t H_s \,\mathrm{d}A_s, \qquad t \geq 0.$$

Note that here,

- ▷ the integral w.r.t. the finite-variation part A is a finite-variation process $H \bullet A = \int H \,\mathrm{d}A \in \mathcal{V}_c$;
- ▷ the integral w.r.t. the local martingale part X is a local martingale $H \bullet X = \int H \,\mathrm{d}X \in \mathcal{M}_c^{\mathrm{loc}}$.

We define integrals for multidimensional processes simply as integrals separately for each component. (For complex-valued processes, we may consider the real and imaginary parts separately.)

10.1 Itô's Formula

The simplest special case of Itô's Formula states that for any continuous semimartingale Y and a twice continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$, one can derive a formula analogous to the Fundamental Theorem of Calculus, but which possibly includes a *second order correction term* emerging from the quadratic variation of Y (as discussed in Section 8.3 before Theorem 8.35):

$$f(Y_t) - f(Y_0) \;=\; \int_0^t f'(Y_s) \,\mathrm{d}Y_s \;+\; \frac{1}{2} \int_0^t f''(Y_s) \,\mathrm{d}\langle Y, Y\rangle_s, \qquad t \geq 0, \tag{10.1}$$

where the quadratic variation process $\langle Y, Y\rangle$ of $Y = Y_0 + X + A$ is defined by using linearity and the fact that the quadratic variation of its finite-variation part A is zero (as in Equation (8.23)):

$$\langle Y, Y\rangle \;=\; \langle X, X\rangle \;+\; \underbrace{\langle A, A\rangle}_{=0} \;=\; \langle X, X\rangle.$$

In other words, the quadratic variation of Y just coincides with the quadratic variation of its local martingale part. Using this, we can also write (10.1) in the form

$$f(Y_t) - f(Y_0) \;=\; \underbrace{\int_0^t f'(Y_s) \,\mathrm{d}X_s}_{\in \mathcal{M}_c^{\mathrm{loc}}} \;+\; \underbrace{\int_0^t f'(Y_s) \,\mathrm{d}A_s \;+\; \frac{1}{2} \int_0^t f''(Y_s) \,\mathrm{d}\langle X, X\rangle_s}_{\in \mathcal{V}_c}. \tag{10.2}$$

Let us state a more general version of this formula, which holds in any finite dimension.

Theorem 10.1. (Itô's Formula) *Let $Y^{(1)}, Y^{(2)}, \dots, Y^{(d)}$ be continuous semimartingales and $f : \mathbb{R}^d \to \mathbb{R}$ twice continuously differentiable. Write $\vec{Y}_t = (Y_t^{(1)}, \dots, Y_t^{(d)})$ for brevity. Then, the following formula holds:*

$$f(\vec{Y}_t) - f(\vec{Y}_0) = \sum_{j=1}^{d} \int_0^t (\partial_j f)(\vec{Y}_s)\,\mathrm{d}Y_s^{(j)} + \frac{1}{2}\sum_{j,k=1}^{d} \int_0^t (\partial_j \partial_k f)(\vec{Y}_s)\,\mathrm{d}\langle Y^{(j)}, Y^{(k)}\rangle_s, \qquad t \geq 0.$$

Proof sketch. The key idea is Taylor expansion for f and dyadic approximation of times. The quadratic variation or covariation arises as an additional error term in the spirit of the integration by parts formula (8.13). We give a sketch here for the gist of the one-dimensional case $d = 1$ (see [LeG16, Theorem 5.10] for details; the case of general $d \geq 2$ is very similar, but has more complicated notation). We thus aim to prove Equation (10.1) for a continuous semimartingale Y and a twice continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$. For each $n \in \mathbb{N}$, we have

$$f(Y_t) = f(Y_0) + \sum_{k=1}^{\lfloor 2^n t\rfloor - 1} \big(f(Y_{k2^{-n}}) - f(Y_{(k-1)2^{-n}}) \big)$$

Using Taylor's formula, we have

$$f(Y_{k2^{-n}}) - f(Y_{(k-1)2^{-n}}) = f'(Y_{(k-1)2^{-n}})\big(Y_{k2^{-n}} - Y_{(k-1)2^{-n}}\big) + \frac{1}{2}R_k^{(n)}\big(Y_{k2^{-n}} - Y_{(k-1)2^{-n}}\big)^2,$$

where the error term includes $R_k^{(n)} := f''(\xi_k^{(n)})$ for some $\xi_k^{(n)} \in [Y_{(k-1)2^{-n}}, Y_{k2^{-n}}]$. We obtain

$$f(Y_t) - f(Y_0) = \underbrace{\sum_{k=1}^{\lfloor 2^n t\rfloor - 1} f'(Y_{(k-1)2^{-n}})\big(Y_{k2^{-n}} - Y_{(k-1)2^{-n}}\big)}_{\longrightarrow \int_0^t f'(Y_s)\,\mathrm{d}Y_s} + \frac{1}{2}\underbrace{\sum_{k=1}^{\lfloor 2^n t\rfloor - 1} R_k^{(n)}\big(Y_{k2^{-n}} - Y_{(k-1)2^{-n}}\big)^2}_{\longrightarrow \int_0^t f''(Y_s)\,\mathrm{d}\langle Y,Y\rangle_s},$$

by taking $n \to \infty$. Indeed, for the first sum, one can use similar arguments as in Exercises 8.32 and 9.15 (and localization via suitable stopping times for X if necessary) to show that

$$\begin{aligned}
&\sum_{k=1}^{\lfloor 2^n t\rfloor - 1} f'(Y_{(k-1)2^{-n}})\big(Y_{k2^{-n}} - Y_{(k-1)2^{-n}}\big) \\
= {} & \sum_{k=1}^{\lfloor 2^n t\rfloor - 1} f'(Y_{(k-1)2^{-n}})\big(X_{k2^{-n}} - X_{(k-1)2^{-n}}\big) + \sum_{k=1}^{\lfloor 2^n t\rfloor - 1} f'(Y_{(k-1)2^{-n}})\big(A_{k2^{-n}} - A_{(k-1)2^{-n}}\big) \\
\overset{n\to\infty}{\longrightarrow} {} & \int_0^t f'(Y_s)\,\mathrm{d}X_s + \int_0^t f'(Y_s)\,\mathrm{d}A_s,
\end{aligned}$$

where the convergence holds in probability uniformly on compacts. Similarly, for the second sum, one can combine similar arguments as in Exercises 8.33 and 8.49 and the proofs leading to Corollary 8.39 & Proposition 8.42 to anticipate that as $n \to \infty$, we have

$$\begin{aligned}
&\sum_{k=1}^{\lfloor 2^n t\rfloor - 1} R_k^{(n)}\big(Y_{k2^{-n}} - Y_{(k-1)2^{-n}}\big)^2 \\
= {} & \underbrace{\sum_{k=1}^{\lfloor 2^n t\rfloor - 1} R_k^{(n)}\big(X_{k2^{-n}} - X_{(k-1)2^{-n}}\big)^2}_{\longrightarrow \int_0^t f''(Y_s)\,\mathrm{d}\langle X,X\rangle_s} + \underbrace{\sum_{k=1}^{\lfloor 2^n t\rfloor - 1} R_k^{(n)}\big(A_{k2^{-n}} - A_{(k-1)2^{-n}}\big)^2}_{\longrightarrow 0}
\end{aligned}$$

$$+ \underbrace{\sum_{k=1}^{\lfloor 2^n t\rfloor -1} R_k^{(n)} \big(X_{k2^{-n}} - X_{(k-1)2^{-n}}\big)\big(A_{k2^{-n}} - A_{(k-1)2^{-n}}\big)}_{\longrightarrow 0},$$

where the convergence holds in probability uniformly on compacts. The main caveat here is that in the first term, one has to deal with the convergence of the derivative $R_k^{(n)} := f''(\xi_k^{(n)})$ as well:

$$\sum_{k=1}^{\lfloor 2^n t\rfloor -1} f''(\xi_k^{(n)}) \big(X_{k2^{-n}} - X_{(k-1)2^{-n}}\big)^2 \quad \overset{n\to\infty}{\longrightarrow} \quad \int_0^t f''(Y_s)\,\mathrm{d}\langle X, X\rangle_s$$

This relies on the uniform continuity of f and X on compact time intervals and is treated in detail in the proof of [LeG16, Theorem 5.10] (which is perhaps surprisingly elaborate). ☐

As a special case of Theorem 10.1, we obtain rigorously the stochastic integration by parts discussed earlier in Section 8.3.

Corollary 10.2. *Let X, Y be two continuous semimartingales. We have*

$$X_t Y_t \;=\; X_0 Y_0 \;+\; \int_0^t X_s\,\mathrm{d}Y_s \;+\; \int_0^t Y_s\,\mathrm{d}X_s \;+\; \langle X, Y\rangle_t, \qquad t \ge 0. \tag{10.3}$$

In particular, when $Y = X$, we have

$$X_t^2 \;=\; X_0^2 \;+\; 2\int_0^t X_s\,\mathrm{d}X_s \;+\; \langle X, X\rangle_t, \qquad t \ge 0.$$

Proof. Consider the twice continuously differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as the product $f(x,y) = xy$. Using its derivatives $(\partial_1 f)(x,y) = y$, $(\partial_2 f)(x,y) = x$, $(\partial_1^2 f)(x,y) = 0 = (\partial_2^2 f)(x,y)$, and $(\partial_1\partial_2 f)(x,y) = (\partial_2\partial_1 f)(x,y) = 1$, Itô's Formula (Theorem 10.1) gives (10.3). ☐

Exercise 10.3. Let B be a standard Brownian motion, and $f : \mathbb{R}^2 \to \mathbb{R}$ a twice continuously differentiable function. Show that

$$f(B_t, t) \;-\; f(0,0) \;=\; \int_0^t \partial_1 f(B_s, s)\,\mathrm{d}B_s \;+\; \int_0^t \Big(\partial_2 + \frac{1}{2}\partial_1^2\Big) f(B_s, s)\,\mathrm{d}s, \qquad t \ge 0.$$

Furthermore, if $(\partial_2 + \frac{1}{2}\partial_1^2)f = 0$, show that the process defined as $M_t := f(B_t, t)$ is a continuous local martingale.

Exercise 10.4. Let B be a standard Brownian motion and fix $r > 0$. Consider the exit time of B from $[-r, r]$,

$$\tau = \inf\{t \ge 0 \mid B_t \notin [-r, r]\}.$$

1. Show that $|B_\tau| = r$ almost surely.
2. Define

   $$f : \mathbb{R}^2 \to \mathbb{R}, \qquad f(x,u) = x^2 + \lambda u, \qquad \text{with } \lambda \in \mathbb{R}.$$

   Find those values of $\lambda \in \mathbb{R}$ for which the process defined as $M_t := f(B_t, t)$ is a local martingale.
3. Let M be a local martingale as found in part 2. Show that the stopped process M^τ is a martingale and calculate the expected exit time $\mathbb{E}[\tau]$.

Exercise 10.5. (Brownian motion on the circle) Consider the process $((X_t, Y_t))_{t\geq 0}$ with coordinates $X_t = \cos(B_t)$ and $Y_t = \sin(B_t)$, where B is a standard Brownian motion. Calculate $\mathbb{E}[X_t]$ for $t \geq 0$.

Hint: *Find a suitable martingale. A good guess could be $r(t)X_t$, where $r : [0,\infty) \to \mathbb{R}$ is a deterministic function. Itô's Formula leads to a differential equation for r, if you require this process to be a local martingale.*

Exercise 10.6. (Ørnstein-Uhlenbeck process) Fix $\alpha, \sigma \geq 0$. Let B be a standard Brownian motion. Prove that

$$X_t \;=\; x_0\, e^{-\alpha t} \;+\; \sigma \int_0^t e^{-\alpha(t-s)}\, \mathrm{d}B_s, \qquad t \geq 0,$$

is a solution to the stochastic differential equation

$$\mathrm{d}X_t = \sigma\, \mathrm{d}B_t - \alpha X_t\, \mathrm{d}t, \qquad X_0 = x_0 \in \mathbb{R},$$

that is, this equation holds in integrated form:

$$X_t \;=\; x_0 \;+\; \sigma \int_0^t \mathrm{d}B_s \;-\; \alpha \int_0^t X_s\, \mathrm{d}s, \qquad t \geq 0.$$

10.2 Applications — general recipe and gambler's ruin

Itô's Formula is very useful for instance in applications where one is interested in solving formulas for probabilities of given events, or expected values or distributions of given random variables, involving continuous semimartingales. The general recipe in all such applications is the same:

1. Find a suitable (local) martingale describing the quantity of interest.
2. Use Optional Stopping to solve for the desired quantity.

To find Ansätze for suitable (local) martingales, one can use Itô's Formula (Theorem 10.1) in the following manner. Recall that for a continuous semimartingale

$$Y = Y_0 + X + A, \qquad X_0 = A_0 = 0,$$

with $Y_0 \in \mathrm{m}\mathcal{F}_0$, and $X \in \mathcal{M}_c^{\mathrm{loc}}$, and $A \in \mathcal{V}_c$ (as in (8.1)), and a twice continuously differentiable function f, the process $f(Y)$ is a semimartingale, whose *Doob-Meyer decomposition* (8.1) reads[21]

$$f(Y_t) \;=\; \underbrace{f(Y_0)}_{\in\, \mathrm{m}\mathcal{F}_0} \;+\; \underbrace{\int_0^t f'(Y_s)\, \mathrm{d}X_s}_{\in\, \mathcal{M}_c^{\mathrm{loc}},\ \text{local mgle part}} \;+\; \underbrace{\int_0^t f'(Y_s)\, \mathrm{d}A_s \;+\; \frac{1}{2}\int_0^t f''(Y_s)\, \mathrm{d}\langle X, X\rangle_s}_{\in\, \mathcal{V}_c,\ \text{finite-variation part}}, \qquad t \geq 0.$$

It is common (and instructive) to write this integral equation in differential form:

$$\mathrm{d}f(Y_t) \;=\; \underbrace{f'(Y_t)\, \mathrm{d}X_t}_{\text{local mgle part}} \;+\; \underbrace{f'(Y_t)\, \mathrm{d}A_t \;+\; \frac{1}{2} f''(Y_t)\, \mathrm{d}\langle X, X\rangle_t}_{\text{finite variation part}}, \qquad t \geq 0.$$

Note that $f(Y)$ is a local martingale exactly when the finite-variation part vanishes:

$$\int_0^t f'(Y_s)\, \mathrm{d}A_s \;+\; \frac{1}{2}\int_0^t f''(Y_s)\, \mathrm{d}\langle X, X\rangle_s \;=\; 0 \qquad \text{for all } t \geq 0.$$

This gives a recipe to construct local martingales from the semimartingale Y by choosing suitable functions f. In many applications, such a choice is governed by solving a differential equation.

As an example, we consider hitting probabilities (Gambler's ruin) for Brownian motion with drift. Processes derived from Brownian motion enjoy another useful input, Markov Property.

[21] An analogous decomposition holds for the multidimensional case, with slightly more complicated notation.

Example 10.7. Let B be a standard Brownian motion. Fix $x \in \mathbb{R}$ and $\mu \in \mathbb{R}$. Consider the continuous semimartingale called[22] *Brownian motion with drift* μ*, started at* x:

$$Y_t := \underbrace{x}_{\in \mathrm{m}\mathcal{F}_0} + \underbrace{B_t}_{\in \mathcal{M}_c^{\mathrm{loc}}} + \underbrace{\mu t}_{\in \mathcal{V}_c}, \qquad t \geq 0.$$

Throughout, let us denote by $\mathbb{P}_x$ the law of Y, indicating the starting point $Y_0 = x$ in the notation; and by $\mathbb{E}_x$ the corresponding expected value. The Markov Property (Proposition 6.12) of B implies that for any fixed time $t \geq 0$, the process $(\widetilde{Y}_s)_{s\geq 0}$ defined by shifting the time by t,

$$\widetilde{Y}_s := Y_{t+s}, \qquad s \geq 0, \tag{10.4}$$

is also a Brownian motion with drift μ, which is started at the random point $\widetilde{Y}_0 = Y_t$ (and independent of the past). This property enables us to construct tautological martingales by conditioning on the time-evolution at the time instant t (see Equation (10.5)).
Fix $a \leq x \leq b$. Consider the exit time of Y from the interval $[a, b]$,

$$\tau := \inf\{t \geq 0 \mid Y_t \in \{a, b\}\}.$$

Write

$$f(x) := \mathbb{P}_x[Y_\tau = a] = 1 - \mathbb{P}_x[Y_\tau = b].$$

We can solve these probabilities using the aforementioned recipe. Let us first proceed by a useful heuristic reasoning, and then make it into a rigorous proof (by sort of reverse engineering).

Step 1. We first *find a suitable local martingale to encode the probabilities of interest.* To this end, recall that by the Tower Property (item 4 of Lemma 2.8), conditional expected values give rise to a tautological martingale (precisely, we should consider the process stopped at time τ):

$$M_t := \mathbb{E}_x[\mathbb{1}\{Y_\tau = a\} \mid \mathcal{F}_t], \qquad t \leq \tau,$$

The key point is that we can use the Markov Property to relate these conditional expected values to the probability of the event $\{Y_\tau = a\}$: indeed, writing slightly imprecisely, we find

$$\begin{aligned} M_t &:= \mathbb{E}_x[\mathbb{1}\{Y_\tau = a\} \mid \mathcal{F}_t] \\ &= \mathbb{E}_{Y_t}[\mathbb{1}\{\widetilde{Y}_{\tilde{\tau}} = a\}] \qquad \text{[by the Markov Property (Proposition 6.1)]} \\ &= \mathbb{P}_{Y_t}[\widetilde{Y}_{\tilde{\tau}} = a] = f(Y_t), \end{aligned} \tag{10.5}$$

where $\tilde{\tau}$ is the first exit time of the process $\widetilde{Y}$ defined in (10.4) from $[a, b]$. Thus, we deduce that the process (10.5) is a local martingale for times $t \in [0, \tau)$. Its starting value is the probability of interest:

$$\mathbb{P}_x[Y_\tau = a] = M_0 = f(Y_0) = f(x).$$

Step 2. Next, we derive a general Ansatz for the function $f(x)$. Anticipating that things will work out, let's *assume that* f *is twice continuously differentiable.* Then, we could use Itô's Formula (Theorem 10.1) to find a differential equation for the function f. (We would then later verify that out assumption was valid.) Indeed, Itô's Formula reads

$$M_t - M_0 = f(Y_t) - f(Y_0) = f(Y_t) - f(x)$$

[22] In such a process, the finite-variation part, which is linear in time, is called a *drift*.

$$
\begin{aligned}
&= \int_0^t f'(Y_s)\,\mathrm{d}Y_s \;+\; \frac{1}{2}\int_0^t f''(Y_s)\,\mathrm{d}\langle Y,Y\rangle_s \\
&= \int_0^t f'(Y_s)\,\mathrm{d}B_s \;+\; \mu\int_0^t f'(Y_s)\,\mathrm{d}s \;+\; \frac{1}{2}\int_0^t f''(Y_s)\,\mathrm{d}s,
\end{aligned}
$$

and since M is a local martingale, the finite-variation part of this formula must vanish identically:

$$
\mu\int_0^t f'(Y_s)\,\mathrm{d}s \;+\; \frac{1}{2}\int_0^t f''(Y_s)\,\mathrm{d}s \;=\; 0 \qquad \text{for all sufficiently small } t \ge 0.
$$

This is equivalent to the differential equation

$$
\mu f'(y) + \frac{1}{2}f''(y) = 0, \qquad y \in [a,b]. \tag{10.6}
$$

Upshot. We have found a differential equation that f *must* satisfy *if* f is twice continuously differentiable and $f(Y_t)$ is a local martingale. It is standard to find the general solution to (10.6):

$$
f(y) = C_1\, e^{-2\mu y} + C_2, \tag{10.7}
$$

where $C_1, C_2 \in \mathbb{R}$ are some constants. To fix the values of these constants, we can argue by imposing suitable *boundary conditions* for the differential equation, which we can guess from our problem at hand. (Recall that we are still doing heuristic reasoning.)

Step 3. Recall that $f(y) = M_0 := \mathbb{P}_y[Y_\tau = a]$. Hence, evaluating it at $y = a$ and $y = b$ gives

$$
f(a) = \mathbb{P}_a[Y_\tau = a] = 1 \qquad \text{and} \qquad f(b) = \mathbb{P}_b[Y_\tau = b] = 0,
$$

since if starting at a, we immediately hit a and never b, while if starting at b, we immediately hit b and never a. Plugging these values into the general solution (10.7), we find

$$
\mathbb{P}_x[Y_\tau = a] \;=\; f(x) \;=\; \frac{e^{-2\mu x} - e^{-2\mu b}}{e^{-2\mu a} - e^{-2\mu b}}.
$$

This is the conclusion of our heuristic considerations.

It remains to turn the heuristics into a rigorous proof, using Optional Stopping.

Claim. The probability of interest equals

$$
\mathbb{P}_x[Y_\tau = a] \;=\; \frac{e^{-2\mu x} - e^{-2\mu b}}{e^{-2\mu a} - e^{-2\mu b}}. \tag{10.8}
$$

Let us denote the left-hand side of the asserted identity (10.8) as $f(x) := \mathbb{P}_x[Y_\tau = a]$, and the right-hand side of the asserted identity (10.8) as

$$
g(x) \;:=\; \frac{e^{-2\mu x} - e^{-2\mu b}}{e^{-2\mu a} - e^{-2\mu b}}.
$$

We aim to show that $g(x) = f(x)$.

Step 4. Write $M_t := g(Y_t)$ for $t \ge 0$. Then, we can *prove* using Itô's Formula that M is a local martingale. Indeed, the right-hand side g of (10.8) is clearly twice continuously differentiable. Hence, Theorem 10.1 gives

$$
\mathrm{d}M_t \;=\; \mathrm{d}g(Y_t) \;=\; g'(Y_t)\,\mathrm{d}B_t, \qquad t \ge 0.
$$

This shows that M is a local martingale.

Step 5. We aim to apply Optional Stopping Theorem 7.5 to the local martingale $M = g(Y)$. To this end, note that since g is bounded on $[a,b] \ni x$, the stopped process $M^\tau = (M_{t\wedge\tau})_{t\geq 0}$ is a bounded local martingale, thus a "true" UI martingale (cf. Exercise 8.25). Therefore, we may indeed apply Optional Stopping Theorem 7.5 to obtain

$$g(x) \;=\; M_0^\tau \;=\; \mathbb{E}[M_\tau^\tau] \;=\; \mathbb{E}[M_\tau].$$

Step 6. On the other hand, directly from the definitions, we can write the expected value of M_τ in terms of the probabilities of interest:

$$\begin{aligned} g(x) \;=\; \mathbb{E}[M_\tau] \;&=\; g(a)\,\mathbb{P}_x[Y_\tau = a] \;+\; g(b)\,\mathbb{P}_x[Y_\tau = b] \\ &=\; \underbrace{g(a)}_{=1}\,f(x) \;+\; \underbrace{g(b)}_{=0}\,(1 - f(x)) \\ &=\; f(x). \end{aligned}$$

This is what we sought to prove. Let us summarize the discussion in the following proposition.

Proposition 10.8. *Let $Y_t := x + B_t + \mu t$, for $t \geq 0$, be a Brownian motion with drift μ started at x. Fix $a \leq x \leq b$, and consider the exit time of Y from the interval $[a,b]$,*

$$\tau := \inf\{t \geq 0 \mid Y_t \in \{a,b\}\}.$$

Then, we have

$$\mathbb{P}_x[Y_\tau = a] = \frac{e^{-2\mu x} - e^{-2\mu b}}{e^{-2\mu a} - e^{-2\mu b}}.$$

Proof. This was proven in Claim (10.8) in Example 10.7. □

Exercise 10.9. Let B be a standard Brownian motion and fix $r > 0$. Consider the exit time

$$\tau_x = \inf\{t \geq 0 \mid |B_t + x| = r\}.$$

Find the formula for the *Laplace transform* $\mathbb{E}[e^{-\theta\,\tau_x}]$, for $\theta \geq 0$, which determines the law of τ_x.

10.3 Applications — recurrence/transience of Brownian motion

Consider Brownian motion in the Euclidean space $\mathbb{R}^n$, which is defined as the process

$$\vec{B}_t = (B_t^{(1)}, \ldots, B_t^{(n)}), \qquad t \geq 0,$$

where the components are independent one-dimensional Brownian motions. Since Brownian motion is the scaling limit of simple random walk (cf. Section 1.3), one would expect that it is *recurrent* in dimensions $d = 1, 2$ and *transient* in dimensions $d \geq 3$. We will next prove that this is indeed the case — however, while in the planar case the probability that $\vec{B}$ returns to a given point is always zero, instead, $\vec{B}$ is *neighborhood recurrent*, that is, it returns almost surely to any given neighborhood of a point.

▷ For $n = 1$, we already know from Corollary 6.11 that B (started at any point) returns to zero infinitely often as $t \to \infty$, because it oscillates infinitely often. Hence, we have

$$\mathbb{P}\big[\text{for all } t_0 > 0 \text{ there exists } t > t_0 \text{ such that } B_t = 0\big] \;=\; 1.$$

This means that B is *point recurrent.*

▷ For $n = 2$, almost surely, for every non-empty open set $O \subset \mathbb{R}^2$, the set $\{t \geq 0 \mid \vec{B}_t \in O\}$ of times when $\vec{B}$ visits O is unbounded. This means that B is *neighborhood recurrent.* (See Theorem 10.14.)

▷ For $n \geq 3$, we have

$$\mathbb{P}\Big[\lim_{t\to\infty} |\vec{B}_t|\Big] \;=\; \infty.$$

This means that B is *transient.* (See Theorem 10.13.)

Example 10.10. Exercise 8.51 shows that

$$\langle B^{(j)}, B^{(k)}\rangle_t = \delta_{j,k}\, t, \qquad t \geq 0.$$

For any twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, Itô's Formula shows that

$$f(\vec{B}_t) \text{ is a continuous local martingale} \quad \text{if and only} \quad \text{if } \Delta f = 0. \tag{10.9}$$

Thus, *harmonic functions* give naturally rise to (local) martingales from Brownian motion. See [MP10] for a thorough discussion of the relation of Brownian motion with potential theory.

Let us apply the above observation to study recurrence/transience of Brownian motion in $\mathbb{R}^n$. To investigate how far Brownian motion reaches, it is useful to study a radial function

$$f(x_1,\ldots,x_n) = g_f(|\vec{x}|^2), \qquad \vec{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n,$$

where $|\vec{x}| = \sqrt{x_1^2 + \cdots + x_n^2}$ is the Euclidean norm. To construct local martingales, using the observation (10.9) from Example 10.10 we seek a function $g_f : (0,\infty) \to \mathbb{R}$ such that $\Delta f = 0$. A direct computation shows that

$$(\Delta f)(\vec{x}) \;=\; 4|\vec{x}|^2\, g_f''(|\vec{x}|^2) \;+\; 2n\, g_f'(|\vec{x}|^2).$$

Hence, we see that $\Delta f = 0$ if and only if

$$2r\, g_f''(r) \;+\; n\, g_f'(r) = 0, \qquad r > 0.$$

A solution to this differential equation is given by $g_f(r) = r^{\frac{2-n}{2}}$, which is non-trivial when $n \neq 2$. (The case of $n = 2$ is the content of Exercise 10.11.) In conclusion,

$$M_t := |\vec{B}_t|^{2-n} = \big((B_t^{(1)})^2 + \cdots + (B_t^{(n)})^2\big)^{\frac{2-n}{2}}, \qquad t \geq 0,$$

is a local martingale by Example 10.10, when the starting point of the Brownian motion is $\vec{B}_0 = \vec{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n \smallsetminus \{0\}$.

Exercise 10.11. Show that for $n = 2$, the process $M_t := \log|\vec{B}_t|$ is a local martingale.

Next, we consider how far the Brownian motion reaches. To this end, fix the starting point $\vec{B}_0 = \vec{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n \smallsetminus \{0\}$ and radii $0 < r \leq |\vec{x}| \leq R < \infty$, and define the stopping times

$$\tau_s := \inf\{t \geq 0 \mid |\vec{B}_t| = s\}, \quad s \in (0,\infty), \qquad \text{and} \qquad \tau := \tau_r \wedge \tau_R.$$

Note that τ is the exit time of $\vec{B}$ from the annulus in $\mathbb{R}^n$ centered at the origin with inner radius r and outer radius R.

Now, using the stopped martingale M^τ, which is bounded and hence a "true" UI martingale (cf. Exercise 8.25), Optional Stopping Theorem 7.5 gives the following explicit formulas.

- ▷ When $n \neq 2$, we find that

$$\begin{aligned} |\vec{x}|^{2-n} = |\vec{B}_0|^{2-n} = M_0 = \mathbb{E}_{\vec{x}}[M_\tau] &= r^{2-n}\,\mathbb{P}_{\vec{x}}[|\vec{B}_\tau| = r] + R^{2-n}\,\mathbb{P}_{\vec{x}}[|\vec{B}_\tau| = R] \\ &= r^{2-n}\,\mathbb{P}_{\vec{x}}[|\vec{B}_\tau| = r] + R^{2-n}\big(1 - \mathbb{P}_{\vec{x}}[|\vec{B}_\tau| = r]\big) \\ &= R^{2-n} + \big(r^{2-n} - R^{2-n}\big)\,\mathbb{P}_{\vec{x}}[|\vec{B}_\tau| = r], \end{aligned}$$

  from which we can solve for the probability that $\vec{B}$ hits the smaller radius r before the larger radius R:

$$\mathbb{P}_{\vec{x}}[\tau_r < \tau_R] = \mathbb{P}_{\vec{x}}[|\vec{B}_\tau| = r] = \frac{|\vec{x}|^{2-n} - R^{2-n}}{r^{2-n} - R^{2-n}}. \tag{10.10}$$

- ▷ Similarly, when $n = 2$, we find that

$$\mathbb{P}_{\vec{x}}[\tau_r < \tau_R] = \mathbb{P}_{\vec{x}}[|\vec{B}_\tau| = r] = \frac{\log R - \log|\vec{x}|}{\log R - \log r}. \tag{10.11}$$

The next result gives a "non-recurrence" property for Brownian motion.

Proposition 10.12. *Fix $n \geq 2$. Consider Brownian motion $\vec{B} = (B^{(1)}, \ldots, B^{(n)})$ in the Euclidean space $\mathbb{R}^n$ started at $\vec{B}_0 = \vec{x} \in \mathbb{R}^n \setminus \{0\}$. Then, we have*

$$\mathbb{P}_{\vec{x}}[\vec{B}_t \neq 0 \text{ for all } t \geq 0] = 1. \tag{10.12}$$

Proof. Consider shrinking radii $r_k := 1/k \downarrow 0$ as $k \uparrow \infty$. The hitting times at those radii converge a.s.: we have $\tau_{r_k} \uparrow \tau_0 := \inf\{t \geq 0 \mid \vec{B}_t = 0\}$; the first hitting time of $\vec{B}$ to the origin. (Note that $\tau_0 > 0$ almost surely, since $\vec{B}$ starts away from the origin.) The claim (10.12) is equivalent to $\mathbb{P}_{\vec{x}}[\tau_0 = +\infty] = 1$. Assuming that $n \geq 3$, we can estimate this probability using Equation (10.10):

$$\mathbb{P}_{\vec{x}}[\tau_{r_k} < \tau_N] = \mathbb{P}_{\vec{x}}[|\vec{B}_\tau| = r_k] = \frac{|\vec{x}|^{2-n} - N^{2-n}}{k^{n-2} - N^{2-n}} \overset{k\to\infty}{\longrightarrow} 0 \qquad \text{for all } N \in \mathbb{N}.$$

Therefore, we see that

$$\mathbb{P}_{\vec{x}}[\tau_0 < \tau_N] = \mathbb{P}_{\vec{x}}\Big[\lim_{k\to\infty} \tau_{r_k} < \tau_N\Big] = 0 \qquad \text{for all } N \in \mathbb{N}.$$

Using Union Bound (A.4), we thus obtain

$$\mathbb{P}_{\vec{x}}[\text{there exists } N \in \mathbb{N} \text{ such that } \tau_0 < \tau_N] \leq \sum_{N=1}^{\infty} \mathbb{P}_{\vec{x}}[\tau_0 < \tau_N] = 0.$$

Now, since $t \mapsto \vec{B}_t$ is (a.s.) continuous, we see that $\tau_N \uparrow \infty$ as $N \uparrow \infty$ (a.s.), which shows that

$$\mathbb{P}_{\vec{x}}[\tau_0 < \infty] = 0 \qquad \Longrightarrow \qquad \mathbb{P}_{\vec{x}}[\tau_0 = +\infty] = 1.$$

As we already noted, this is equivalent to the claim (10.12). One can similarly prove that the claimed property (10.12) holds also for $n = 2$, using Equation (10.11) (exercise). □

The next result gives the transience property for Brownian motion.

Theorem 10.13. *Fix $n \geq 3$. Consider Brownian motion $\vec{B}$ in the Euclidean space $\mathbb{R}^n$ started at $\vec{B}_0 = \vec{x} \in \mathbb{R}^n \smallsetminus \{0\}$. Then, almost surely, we have*

$$\lim_{t \to \infty} |\vec{B}_t| = \infty.$$

Proof. Using Monotone Convergence Theorem [Kyt20, Theorem VII.8], we see that

$$\mathbb{P}_{\vec{x}}\big[\liminf_{t \to \infty} |\vec{B}_t| < \infty\big] = \lim_{r \to \infty} \mathbb{P}_{\vec{x}}\big[\liminf_{t \to \infty} |\vec{B}_t| < r\big].$$

As before, consider the stopping times $\tau_r := \inf\{t \geq 0 \mid |\vec{B}_t| = r\}$. Using the Strong Markov Property (Proposition 6.12), for each $r > 0$ we obtain

$$\mathbb{P}_{\vec{x}}\big[\liminf_{t \to \infty} |\vec{B}_t| < r\big] \leq \mathbb{P}_{\vec{B}_{\tau_R}}\left[\tau_r < \infty\right] \overset{(10.10)}{=} \left(\frac{N}{r}\right)^{2-n} \quad N \underset{\longrightarrow}{\to} \infty \quad 0$$

using the identity (10.10) with $r > 0$ and $R \to \infty$. This implies that

$$\mathbb{P}_{\vec{x}}\big[\liminf_{t \to \infty} |\vec{B}_t| < \infty\big] = 0 \qquad \Longrightarrow \qquad \mathbb{P}_{\vec{x}}\big[\liminf_{t \to \infty} |\vec{B}_t| = \infty\big] = 1,$$

which also imples that $\lim_{t \to \infty} |\vec{B}_t| = \infty$ almost surely. □

Theorem 10.14. *Fix $n = 2$. Consider Brownian motion $\vec{B}$ in the Euclidean space $\mathbb{R}^2$ started at $\vec{B}_0 = \vec{x} \in \mathbb{R}^2 \smallsetminus \{0\}$. Then, almost surely, for every non-empty open set $O \subset \mathbb{R}^2$, the set $\{t \geq 0 \mid \vec{B}_t \in O\}$ of times when $\vec{B}$ visits O is unbounded.*

Proof. See [LeG16, Theorem 7.17]. □

We conclude this section with a few exercises applying the general recipe.

Exercise 10.15. Consider the process $((X_t, Y_t))_{t \geq 0}$ whose coordinates $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ are independent Brownian motions started respectively from $X_0 \in \mathbb{R}$ and $Y_0 \in \mathbb{R}$.

1. Consider the function $\rho : \mathbb{R}^2 \smallsetminus \{(0,0)\} \to (0, \infty)$ defined as $\rho(x,y) = x^2 + y^2$. Find a non-constant twice continuously differentiable function $g : (0, \infty) \to \mathbb{R}$ such that for $f = g \circ \rho$, the process $(f(X_t, Y_t))_{t \geq 0}$ is a local martingale. *Hint*: *You can use Itô's Formula.*

2. Given $0 < r < R$ such that $r \leq \sqrt{\rho(X_0, Y_0)} \leq R$, define a stopping time

$$\tau = \tau_{r,R} := \inf\big\{t \geq 0 \mid \sqrt{\rho(X_t, Y_t)} \notin [r, R]\big\}.$$

   Show that $(f(X_t^\tau, Y_t^\tau))_{t \geq 0}$ is a bounded martingale.

3. For $R > 0$, define

$$\tau_R := \inf\big\{t \geq 0 \mid \sqrt{\rho(X_t, Y_t)} = R\big\}.$$

   Use Optional Stopping Theorem for the martingale in part 2 to compute the probability

$$\mathbb{P}\left[\tau_R < \tau_r\right].$$

   What is the limit of this probability as $r \downarrow 0$? How about the limit as $R \uparrow \infty$?

Exercise 10.16. Let $n \geq 3$. Consider the process $\vec{X} = (\vec{X}_t)_{t\geq 0}$, $\vec{X}_t = (X_t^{(1)}, \ldots, X_t^{(n)})$ whose coordinates are independent Brownian motions started from $X_0^{(1)}, \ldots, X_0^{(n)} \in \mathbb{R}$ such that the starting point $\vec{X}_0 = (X_0^{(1)}, \ldots, X_0^{(n)}) \in \mathbb{R}^n \smallsetminus \{(0, \ldots, 0)\}$ is not the origin.

1. Show that the following process $M = (M_t)_{t\geq 0}$ is a local martingale:

$$M_t = R_t^{2-n} \qquad \text{where} \qquad R_t = \sqrt{(X_t^{(1)})^2 + \cdots + (X_t^{(n)})^2}.$$

   Hint: *You can use Itô's Formula.*

2. Prove that M is a supermartingale.

   Hint: *You can use Exercise 8.23.*

3. Use the transience of the n-dimensional Brownian motion to show that $M_t \to 0$ almost surely as $t \to \infty$.

4. Show that $\mathbb{E}[M_t] \to 0$ as $t \to \infty$ and conclude that M is not a martingale.

Exercise 10.17. Let B be a standard Brownian motion and fix $r > 0$. Consider the exit time of B from $[-r, r]$,

$$\tau_x = \inf\{t \geq 0 \mid |B_t + x| = r\}.$$

Find the formula for the *Laplace transform* $\mathbb{E}[e^{\theta\,\tau_x}]$, for $\theta \in \mathbb{R}$, which determines the law of τ_x.

A Basic concepts from probability theory & stochastic processes

A.1 Basic definitions

Power set $\mathscr{P}(\Omega)$ of a given set Ω is defined as $\mathscr{P}(\Omega) := \{E \mid E \subset \Omega\}$.

Exercise A.1. Find the power set $\mathscr{P}(\Omega)$ for the sets

- ▷ $\Omega = \{1\}$
- ▷ $\Omega = \{a, b, c\}$
- ▷ $\Omega = \mathbb{N} := \{1, 2, 3, \dots\}$
- ▷ $\Omega = \mathbb{R}$.

What is the cardinality of $\mathscr{P}(\Omega)$ in these cases?

Probability space $(\Omega, \mathscr{F}, \mathbb{P})$ consists of:

- ▷ *sample space* Ω of possible *outcomes* $\omega \in \Omega$;
- ▷ *sigma-algebra* $\mathscr{F} \subset \mathscr{P}(\Omega)$ of *events* $E \in \mathscr{F}$, satisfying by definition the properties
  1. (contains the whole space): $\Omega \in \mathscr{F}$,
  2. (is closed under complement): $E \in \mathscr{F} \;\Rightarrow\; \Omega \setminus E \in \mathscr{F}$,
  3. (is closed under countable union): $E_1, E_2, \dots \in \mathscr{F} \;\Rightarrow\; \bigcup_{n=1}^{\infty} E_n \in \mathscr{F}$;

  $(\Omega, \mathscr{F})$ is called a *measurable space*, and any $E \subset \Omega$ such that $E \in \mathscr{F}$ is called *measurable.*
- ▷ *probability measure* $\mathbb{P}$ on the measurable space $(\Omega, \mathscr{F})$, satisfying by definition the properties
  1. (is a non-negative function): $\mathbb{P} : \mathscr{F} \to [0, +\infty]$,
  2. (is additive for countable unions): $E_1, E_2, \dots \in \mathscr{F}$ disjoint $\Rightarrow\; \mathbb{P}\big[\bigcup_{n=1}^{\infty} E_n\big] = \sum_{n=1}^{\infty} \mathbb{P}[E_n]$,
  3. (gives zero mass for empty set): $\mathbb{P}[\varnothing] = 0$,
  4. (has total mass equal to one): $\mathbb{P}[\Omega] = 1$.

  Recall that $\mathbb{P}$ satisfying properties 1–3 is called a *measure*, and the additional property 4 is a normalization condition such that the total probability equals one. For an event $E \in \mathscr{F}$, the value[23] $\mathbb{P}[E]$ is called the *probability of* E, or sometimes the *mass of* E.

  Remark. Alternatively, a probability measure can be defined via the properties

  1'. $\mathbb{P} : \mathscr{F} \to [0, 1]$,
  2'. $E_1, E_2, \dots \in \mathscr{F}$ disjoint $\Rightarrow\; \mathbb{P}\big[\bigcup_{n=1}^{\infty} E_n\big] = \sum_{n=1}^{\infty} \mathbb{P}[E_n]$,
  3'. $\mathbb{P}[\Omega] = 1$,

  since properties 2' & 3' imply 3:

$$1 = \mathbb{P}[\Omega] = \mathbb{P}[\varnothing \cup \Omega] = \mathbb{P}[\varnothing] + \mathbb{P}[\Omega] = \mathbb{P}[\varnothing] + 1 \qquad \Longrightarrow \qquad \mathbb{P}[\varnothing] = 0.$$

[23] Notation: it is conventional to write $\mathbb{P}[E]$ instead of $\mathbb{P}(E)$, but this is of course a matter of choice.

Exercise A.2. Consider a finite or countably infinite set Ω with $\mathscr{F} = \mathscr{P}(\Omega)$. Let $p : \Omega \to [0,1]$ be a *probability mass function* satisfying $\sum_{\omega\in\Omega} p(\omega) = 1$. Show that $(\Omega, \mathscr{P}(\Omega))$ is a measurable space and

$$\mathbb{P} : \mathscr{F} \to [0,1], \qquad \mathbb{P}[E] := \sum_{\omega\in E} p(\omega) \qquad \text{for all } E \in \mathscr{F},$$

is a probability measure on it. Conversely, given $(\Omega, \mathscr{F}, \mathbb{P})$, construct a suitable p.

Exercise A.3. Write down $(\Omega, \mathscr{F}, \mathbb{P})$ for k repeated independent coin tosses. How about taking $k \to \infty$?

Random variable ξ on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with values in a measurable space $(S, \mathscr{S})$ is a function $\xi : \Omega \to S$ which is *measurable*, i.e., it satisfies the property

$$\xi^{-1}(A) := \{\omega \in \Omega \mid \xi(\omega) \in A\} \in \mathscr{F} \qquad \text{for all (measurable) } A \in \mathscr{S}.$$

Commonly we abbreviate the left-hand side as $\xi^{-1}(A) =: \{\xi \in A\}$, which represents the values of the random quantity ξ more intuitively — $\{\xi \in A\}$ asks for those values of ξ that belong to A.

Exercise A.4. (Coin tosses) Take $\Omega = \{\mathrm{H}, \mathrm{T}\}$ with $\mathscr{F} = \mathscr{P}(\Omega)$ and $\mathbb{P}$ determined[a] by $\mathbb{P}[\mathrm{H}] = \frac{1}{2} = \mathbb{P}[\mathrm{T}]$. Consider a *Bernoulli random variable* $\xi : \Omega \to \{-1, 1\}$ defined as

$$\xi(\mathrm{H}) := -1 \qquad \text{and} \qquad \xi(\mathrm{T}) := 1.$$

Find the probabilities of the events $\{\xi = -1\}$ and $\{\xi = 1\}$ and verify that

$$\mathbb{E}[\xi] = 0 \qquad \text{and} \qquad \mathrm{Var}(\xi) = 1.$$

[a] For notational simplicity, we write $\mathbb{P}[\mathrm{H}]$ and $\mathbb{P}[\xi = 1]$ etc. instead of $\mathbb{P}[\{\mathrm{H}\}]$ and $\mathbb{P}[\{\xi = 1\}]$ etc.

Example A.5. Different types of random variables:

- ▷ Real-valued random variable (i.e., random real number): $S = \mathbb{R}$ with $\mathscr{S} = \mathscr{B}(\mathbb{R})$ Borel sets[24].
- ▷ Random continuous function on the unit interval:

$$S = \mathcal{C}([0,1], \mathbb{R}) := \{f : [0,1] \to \mathbb{R} \mid f \text{ is continuous}\}$$

with $\mathscr{S} = \mathscr{B}(\mathcal{C}([0,1], \mathbb{R}))$ the Borel sigma-algebra with respect to the topology induced by the *uniform norm* (i.e., *sup-norm*) $\|f\|_\infty := \sup_{x\in[0,1]} |f(x)|$.

Distribution (law) of random variable $\xi : \Omega \to S$ is the probability measure

$$\mathsf{P}_\xi : \mathscr{S} \to [0,1], \qquad \mathsf{P}_\xi[A] := \mathbb{P}[\xi \in A] \quad \text{for all } A \in \mathscr{S}$$

on $(S, \mathscr{S})$ induced by $\mathbb{P}$ and ξ. (Exercise: check for yourself that this makes sense.)

[24] Recall that the Borel sigma-algebra $\mathscr{B}(S)$ of a topological space S is generated by all open subsets of S.

Probability density f_ξ (with respect to the Lebesgue measure on $\mathbb{R}$) of a real-valued random variable $\xi : \Omega \to \mathbb{R}$ on $(\Omega, \mathscr{F}, \mathbb{P})$ is a function $f_\xi : \mathbb{R} \to [0, +\infty)$ such that

$$\mathsf{P}_\xi[A] = \int_A f_\xi(x)\,\mathrm{d}x \qquad \text{for all (Borel sets) } A \in \mathscr{B}(\mathbb{R}).$$

Note that not all real-valued random variables have a density. We say that ξ has a *continuous distribution* if it has a density. In this case, if f_ξ satisfies the additional property

$$\int_{\mathbb{R}} |x|\, f_\xi(x)\,\mathrm{d}x < \infty,$$

then ξ is $L^1(\mathbb{P})$-*integrable*, i.e., $\int_\Omega |\xi(\omega)|\,\mathrm{d}\mathbb{P}(\omega) < \infty$. Then we also have

$$\mathbb{E}[\xi] := \int_\Omega \xi(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}} x\, f_\xi(x)\,\mathrm{d}x = \mathfrak{m} < \infty, \qquad \text{(expected value)} \qquad \text{(A.1)}$$

$$\mathrm{Var}(\xi) := \mathbb{E}[(\xi - \mathfrak{m})^2] = \int_{\mathbb{R}} (x - \mathfrak{m})^2 f_\xi(x)\,\mathrm{d}x < \infty \qquad \text{(variance).} \qquad \text{(A.2)}$$

(These properties have been proven in the Probability Theory course, see [Kyt20, Chapter VIII.1].)

Remark A.6. The notion of density naturally generalizes to random variables $\xi : \Omega \to \mathbb{R}^n$ taking values in a higher dimensional Euclidean space: $f_\xi : \mathbb{R}^n \to [0, +\infty)$ such that

$$\mathsf{P}_\xi[A] = \int_A f_\xi(x_1, \ldots, x_n)\,\mathrm{d}x_n \cdots \mathrm{d}x_1 \qquad \text{for all (Borel sets) } A \in \mathscr{B}(\mathbb{R}^n).$$

Exercise A.7. Prove that if ξ has a continuous distribution with density f_ξ, then

$$\int_{\mathbb{R}} |x|\, f_\xi(x)\,\mathrm{d}x < \infty \qquad \Longleftrightarrow \qquad \int_\Omega |\xi(\omega)|\,\mathrm{d}\mathbb{P}(\omega) < \infty$$

Verify also the above formulas (A.1, A.2) for the expected value and variance of ξ.

Example A.8. Different types of densities:

▷ Gaussian: for $\mathfrak{m} \in \mathbb{R}$ and $\mathfrak{s} > 0$,

$$f(x) = \frac{1}{\sqrt{2\pi\mathfrak{s}^2}} \exp\Big(- \frac{(x - \mathfrak{m})^2}{2\mathfrak{s}^2} \Big). \qquad \text{(A.3)}$$

If $f_\xi = f$, then we denote $\xi \sim N(\mathfrak{m}, \mathfrak{s}^2)$ and say that ξ is a *Gaussian random variable.*

▷ Exponential: for $\lambda > 0$,

$$f(x) = \begin{cases} \lambda\, e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

If $f_\xi = f$, then we denote $\xi \sim \mathrm{Exp}(\lambda)$ and say that ξ is an *Exponential random variable.*

Independent random variables $\xi, \eta : \Omega \to \mathbb{R}$ on the same probability space satisfy by definition the following equivalent properties:

1. (factorization of joint cumulative distribution function): $\mathbb{P}[\xi \leq x,\ \eta \leq y] = \mathbb{P}[\xi \leq x]\,\mathbb{P}[\eta \leq y]$,
2. (factorization of joint law[25]): $\mathsf{P}_{\xi,\eta} = \mathsf{P}_\xi \otimes \mathsf{P}_\eta$

[25] Recall that $\mathsf{P}_{\xi,\eta}[A] := \mathbb{P}[(\xi, \eta) \in A]$ for all $A \in \mathscr{B}(\mathbb{R}^2)$ defines the *joint law* of ξ and η.

If they have a joint density $f_{\xi,\eta} : \mathbb{R}^2 \to [0,+\infty)$, then properties 1 and 2 are also equivalent to

3. (factorization of joint density): $f_{\xi,\eta}(x,y) = f_\xi(x)\, f_\eta(y)$ for Lebesgue-almost all $(x,y) \in \mathbb{R}^2$.

(These properties have been proven in the Probability Theory course, see [Kyt20, Chapter X.3].)

Exercise A.9. Suppose that ξ and η have a joint density $f_{\xi,\eta} : \mathbb{R}^2 \to [0,+\infty)$. Show that then both ξ and η have individual densities (called *marginals*). Does the converse hold?

Exercise A.10. Suppose $\xi \sim N(\mathfrak{m}_1, \mathfrak{s}_1^2)$ and $\eta \sim N(\mathfrak{m}_2, \mathfrak{s}_2^2)$ are independent. Calculate the characteristic function

$$\varphi_{\xi+\eta}(\theta) = \mathbb{E}\left[e^{\mathrm{i}\,\theta(\xi+\eta)}\right]$$

and conclude that $\xi + \eta \sim N(\mathfrak{m}_1 + \mathfrak{m}_2, \mathfrak{s}_1^2 + \mathfrak{s}_2^2)$. Find also the joint distribution of (ξ, η).

Exercise A.11. Let $C \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix[a], and let $\mathfrak{m} \in \mathbb{R}^n$ be a vector. Define

$$p : \mathbb{R}^n \to \mathbb{R}, \qquad p(\boldsymbol{x}) := \frac{1}{\mathcal{Z}} \exp\Big(-\frac{1}{2}(\boldsymbol{x}-\mathfrak{m})^T\, C^{-1}\, (\boldsymbol{x}-\mathfrak{m})\Big), \qquad \boldsymbol{x} \in \mathbb{R}^n,$$

where $\mathcal{Z}$ is a constant.

(a) Calculate $\int_{\mathbb{R}^n} p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$, and show that p is a (correctly normalized) probability density on $\mathbb{R}^n$ if

$$\mathcal{Z} = (2\pi)^{n/2}\sqrt{\det(C)}.$$

Hint: *First, make a change of variables (translation) to reduce to the case* $\mathfrak{m} = 0$. *Then, make an orthogonal change of variables to a basis in which* C *is diagonal.*

(b) Choose $\mathcal{Z}$ as in part (a). Let $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$ be a random vector in $\mathbb{R}^n$ which has probability density $p : \mathbb{R}^n \to \mathbb{R}$ as above. Let $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$. Calculate the characteristic function of the linear combination

$$\boldsymbol{a}\cdot\boldsymbol{\xi} = \sum_{j=1}^n a_j \xi_j .$$

[a] *Symmetric* means $C_{ij} = C_{ji}$ for all i,j. *Positive definite* means $v^T C v > 0$ for any $v \in \mathbb{R}^n \setminus \{0\}$.

Stochastic process X on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ indexed by a set I with values in a measurable space $(S, \mathscr{S})$ is a collection $X = (X_i)_{i\in I}$ of random variables $X_i : \Omega \to S$. The index set I is usually either a discrete set, such as $\mathbb{N}$, or a subset of the reals, such as $[0,1]$ or $[0,\infty)$. In these cases, we speak of a discrete-time or a continuous-time stochastic process, respectively. Each value $X_i(\omega)$ describes a random position of the process. When $I = [0,\infty)$, it is also natural to think of $t \mapsto X_t(\omega)$ as a random function $[0,\infty) \to S$, or alternatively, a random path on S.

Two stochastic processes X and Y are said to be *independent* if, for any sets $t_1, \ldots, t_n \geq 0$ and $s_1, \ldots, s_n \geq 0$ of times, the random vectors $(X_{t_1}, \ldots, X_{t_n})$ and $(Y_{s_1}, \ldots, Y_{s_n})$ are independent.

A.2 Useful tools

We gather here without proof useful results from probability theory.

The countable sub-additivity property gives a very useful bound for probability measures.

Lemma A.12. (Union Bound) *For any events* $E_1, E_2, \ldots \in \mathscr{F}$ *in a probability space* $(\Omega, \mathscr{F}, \mathbb{P})$, *we have*

$$\mathbb{P}\Big[\bigcup_{n=1}^{\infty} E_n\Big] \leq \sum_{n=1}^{\infty} \mathbb{P}[E_n]. \tag{A.4}$$

Proof. See, e.g., [Kyt20, Theorem II.20]. □

Lemma A.13. (First Borel-Cantelli lemma) *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space. For any events* $E_1, E_2, \ldots \in \mathscr{F}$ *such that*

$$\sum_{n=1}^{\infty} \mathbb{P}[E_n] < +\infty,$$

we have

$$\mathbb{P}[E_n \text{ occurs infinitely often}] = 0.$$

Proof. See, e.g., [Kyt20, Lemma V.7]. □

Lemma A.14. (Second Borel-Cantelli lemma) *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space. For any* independent *events* $E_1, E_2, \ldots \in \mathscr{F}$ *such that*

$$\sum_{n=1}^{\infty} \mathbb{P}[E_n] = +\infty,$$

we have

$$\mathbb{P}[E_n \text{ occurs infinitely often}] = 1.$$

Proof. See, e.g., [Kyt20, Lemma V.8]. □

This event can be alternatively written in the form

$$\{E_n \text{ occurs infinitely often}\} = \Big\{\limsup_{n\to\infty} E_n\Big\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k.$$

If $\mathbb{P}[E_n \text{ occurs infinitely often}] = 0$, then almost surely, there are only finitely (but randomly) many indices $n \in \mathbb{N}$ for which the event E_n occurs.

Lemma A.15. (Markov's Inequality, a.k.a. first Chebyshev's inequality) *Let* $\xi : \Omega \to \mathbb{R}$ *be a random variable on a probability space* $(\Omega, \mathscr{F}, \mathbb{P})$. *For any* $\lambda > 0$, *we have*

$$\mathbb{P}[|\xi| \geq \lambda] \leq \frac{\mathbb{E}[|\xi|]}{\lambda}. \tag{A.5}$$

Proof. See, e.g., [Kyt20, Lemma XI.6]. □

Recall that a random variable $\xi : \Omega \to \mathbb{R}$ is called square-integrable if $\mathbb{E}[\xi^2] < \infty$. The space of all square-integrable random variables is denoted as $L^2(\mathbb{P})$.

Lemma A.16. (Second Chebyshev's inequality) *Let $\xi : \Omega \to \mathbb{R}$ be a square-integrable random variable (i.e., $\xi \in L^2(\mathbb{P})$) on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Denote $\mathbb{E}[\xi] = \mathfrak{m}$ and* $\mathrm{Var}(\xi) = \mathfrak{s}^2$. *For any $\lambda > 0$, we have*

$$\mathbb{P}[|\xi - \mathfrak{m}| \geq \lambda] \leq \frac{\mathfrak{s}^2}{\lambda^2}. \tag{A.6}$$

Proof. See, e.g., [Kyt20, Corollary XI.7]. □

Lemma A.17. (Fatou's lemma) *For any sequence $(f_n)_{n\in\mathbb{N}}$ of non-negative measurable functions on a measure space $(S, \mathscr{S}, \mu)$, we have*[a]

$$\int_S \liminf_{n\to\infty} f_n \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_S f_n \, \mathrm{d}\mu.$$

Proof. See, e.g., [Kyt20, Lemma VII.20]. □

[a]Note that the integrals may be infinite.

Lemma A.18. (Reverse Fatou's lemma) *For any sequence $(f_n)_{n\in\mathbb{N}}$ of non-negative measurable functions on a measure space $(S, \mathscr{S}, \mu)$, we have*[a]

$$\limsup_{n\to\infty} \int_S f_n \, \mathrm{d}\mu \leq \int_S \limsup_{n\to\infty} f_n \, \mathrm{d}\mu.$$

Proof. See, e.g., [Kyt20, Lemma VII.21]. □

[a]Note that the integrals may be infinite.

Recall that for two measure spaces $(S_1, \mathscr{S}_1, \mu_1)$ and $(S_2, \mathscr{S}_2, \mu_2)$, where μ_1 and μ_2 are finite measures, the product space $(S_1 \times S_2, \mu_1 \otimes \mu_2)$ is naturally a measure space:

▷ the *product sigma-algebra* $\mathscr{S}_1 \otimes \mathscr{S}_2$ is defined as the smallest sigma-algebra on $S_1 \times S_2$ with respect to which the projections

$$\mathrm{pr}_1 : S_1 \times S_2 \to S_1 \qquad \text{and} \qquad \mathrm{pr}_2 : S_1 \times S_2 \to S_2$$

are measurable (see, e.g., [Kyt20, Definition IX.2]);

▷ the *product measure* $\mu_1 \otimes \mu_2$ is defined by

$$\begin{aligned}(\mu_1 \otimes \mu_2)[A] &:= \int_{S_1} \Big(\int_{S_2} \mathbb{1}_A \, \mathrm{d}\mu_2(s_2) \Big) \mathrm{d}\mu_1(s_1) \\ &:= \int_{S_2} \Big(\int_{S_1} \mathbb{1}_A \, \mathrm{d}\mu_1(s_1) \Big) \mathrm{d}\mu_2(s_2), \qquad A \in \mathscr{S}_1 \otimes \mathscr{S}_2\end{aligned}$$

(see, e.g., [Kyt20, Definition IX.6 & Lemma IX.7]).

Theorem A.19. (Fubini's theorem) *For a function* $f : S_1 \times S_2 \to [-\infty, +\infty]$*, consider*

$$\int_{S_1 \times S_2} f(s_1, s_2)\, \mathrm{d}(\mu_1 \otimes \mu_2)(s_1, s_2), \tag{A.7}$$

$$\int_{S_1} \Big(\int_{S_2} f(s_1, s_2)\, \mathrm{d}\mu_2(s_2) \Big)\, \mathrm{d}\mu_1(s_1), \tag{A.8}$$

$$\int_{S_2} \Big(\int_{S_1} f(s_1, s_2)\, \mathrm{d}\mu_1(s_1) \Big)\, \mathrm{d}\mu_2(s_2). \tag{A.9}$$

Then, the following hold.

1. *If f is non-negative and measurable, then the integrals* (A.7, A.8, A.9) *are all in* $[0, +\infty]$*, and they are all equal.*
2. *If f is integrable, then* (A.7, A.8, A.9) *are all in* $\mathbb{R} = (-\infty, \infty)$*, and they are all equal.*

Proof. See, e.g., [Kyt20, Theorem IX.9]. □

A.3 Various notions of convergence

Convergence in probability. A sequence $(\xi_n)_{n \in \mathbb{N}}$ of real-valued random variables is said to *converge in probability* to a real-valued random variable ξ if

$$\lim_{n \to \infty} \mathbb{P}[|\xi_n - \xi| < \varepsilon] = 1 \qquad \text{for all } \varepsilon > 0. \tag{A.10}$$

In this case, we write $\xi_n \xrightarrow{\mathbb{P}} \xi$ as $n \to \infty$.

Exercise A.20. Show that (A.10) is equivalent to

$$\lim_{n \to \infty} \mathbb{P}[|\xi_n - \xi| \geq \varepsilon] = 0 \qquad \text{for all } \varepsilon > 0.$$

Convergence almost surely. A sequence $(\xi_n)_{n \in \mathbb{N}}$ of real-valued random variables is said to *converge almost surely* (a.s.) to a real-valued random variable ξ if

$$\mathbb{P}[\lim_{n \to \infty} \xi_n = \xi] = 1. \tag{A.11}$$

In this case, we write $\xi_n \xrightarrow{\text{a.s.}} \xi$ as $n \to \infty$. The a.s. convergence can also be phrased analytically by saying that the measurable functions $\xi_n : \Omega \to \mathbb{R}$ converge pointwise for *almost every* $\omega \in \Omega$.

Convergence almost surely is a strong notion of convergence for random variables. Indeed, it implies their convergence in probability, while the converse does not necessarily hold.

Exercise A.21. Show that convergence almost surely (A.11) implies convergence in probability (A.10).

Exercise A.22. Find an example of a sequence of random variables $(\xi_n)_{n \in \mathbb{N}}$ that converges in probability, but does not converge almost surely.

Convergence in L^1. A sequence $(\xi_n)_{n\in\mathbb{N}}$ in $L^1(\mathbb{P})$ is said to *converge in L^1* to $\xi \in L^1(\mathbb{P})$ if

$$\lim_{n\to\infty} \mathbb{E}[|\xi_n - \xi|] = 0. \tag{A.12}$$

In this case, we write $\xi_n \xrightarrow{L^1} \xi$.

Exercise A.23. Show that convergence in L^1 (A.12) implies convergence in probability (A.10).

Exercise A.24. Find an example of a sequence of random variables $(\xi_n)_{n\in\mathbb{N}}$ such that

1. $(\xi_n)_{n\in\mathbb{N}}$ converges in L^1, but it does not converge almost surely.
2. $(\xi_n)_{n\in\mathbb{N}}$ converges almost surely, but it does not converge in L^1.

Even though in general, almost sure convergence does not imply convergence in L^1, it does under a suitable condition on the growth of the sequence (termed uniform integrability).

Uniformly integrable (UI) collection $(\xi_i)_{i\in I}$ of real-valued random variables satisfies by definition the following tail bound:

$$\lim_{R\to\infty} \sup_{i\in I} \mathbb{E}\Big[|\xi_i| \, \mathbb{1}\{|\xi_i| \geq R\}\Big] = 0.$$

Proposition A.25. *Consider a sequence $\xi_1, \xi_2, \ldots \in L^1(\mathbb{P})$ of random variables. Also, let $\xi \in L^1(\mathbb{P})$. The following are equivalent:*

1. *The sequence $(\xi_n)_{n\in\mathbb{N}}$ converges to ξ in L^1.*
2. *The sequence $(\xi_n)_{n\in\mathbb{N}}$ is UI and converges to ξ in probability.*

Proof. See, for instance, [Wil91, Theorem 13.7]. □

A.4 Laws of large numbers

Let us consider a sequence $\xi_1, \xi_2, \ldots$ of independent real-valued random variables. We dDefine a random walk with these steps as

$$S_0 = 0 \qquad \text{and} \qquad S_n := \sum_{j=1}^{n} \xi_j \qquad \text{for } n = 1, 2, \ldots.$$

When the second moments of the steps are bounded, we have the following asymptotic result. It follows easily from Markov (A.5) and second Chebyshev (A.6) inequalities.

Theorem A.26. (Weak Law of Large Numbers) *Suppose that $\mathbb{E}[\xi_j] = \mathfrak{m} < \infty$ and there exists $K < \infty$ such that $\mathbb{E}[\xi_j^2] \leq K$ for all j. Then, we have*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathfrak{m} \qquad \text{as } n \to \infty,$$

where the convergence takes place in probability.

Proof. See, e.g., [Kyt20, Theorem XI.4]. □

Theorem A.27. (Strong Law of Large Numbers) *Suppose that* $\mathbb{E}[\xi_j] = \mathfrak{m} < \infty$ *and there exists* $K < \infty$ *such that* $\mathbb{E}[\xi_j^4] \leq K$ *for all* j*. Then, we have*

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mathfrak{m} \qquad \text{as } n \to \infty,$$

where the convergence takes place $\mathbb{P}$*-almost surely.*

Proof. See [Kyt20, Theorem XI.5]. □

In fact, if we assume that $\xi_1, \xi_2, \ldots$ are i.i.d. real-valued random variables with $\mathbb{E}[\xi_1] = \mathfrak{m} < \infty$ (so $\mathbb{E}[\xi_j] = \mathfrak{m} < \infty$ for all j), then the statement in Theorem A.27 also holds, both $\mathbb{P}$-almost surely and in L^1. However, while the proof of Theorem A.27 is elementary assuming $\mathbb{E}[\tau^4] < \infty$, the case without this assumption is quite hard to prove — see, e.g., [Wil91, Chapters 12 & 14].

A.5 Dynkin's Identification Theorem

Sigma-algebra $\sigma(\Pi)$ generated by $\Pi \subset \mathcal{P}(\Omega)$ is the *smallest* sigma-algebra containing Π.

Pi-system is a collection $\Pi \subset \mathcal{P}(\Omega)$ closed under non-empty intersection:

$$A_1, A_2 \in \Pi \quad \text{and} \quad A_1 \cap A_2 \neq \varnothing \qquad \Longrightarrow \qquad A_1 \cap A_2 \in \Pi.$$

Exercise A.28. Consider the collection $\Pi(\mathbb{R}) := \{(-\infty, a] \mid a \in \mathbb{R}\}$ of semi-infinite intervals on $\mathbb{R}$.

1. Show that $\Pi(\mathbb{R})$ is a pi-system.
2. Show that $\Pi(\mathbb{R})$ generates the Borel sigma-algebra $\mathscr{B}(\mathbb{R})$ on $\mathbb{R}$.

Theorem A.29. (Dynkin's Identification Theorem) *Let* $\mathbb{P}_1$ *and* $\mathbb{P}_2$ *be probability measures on a common measurable space* $(\Omega, \mathscr{F})$*. Let* $\Pi \subset \mathcal{P}(\Omega)$ *be a pi-system such that the sigma-algebra generated by is* full: $\sigma(\Pi) = \mathscr{F}$*. Then, the following are equivalent.*

DYN1. *The two probability measures are equal*: $\mathbb{P}_1 = \mathbb{P}_2$.

DYN2. $\mathbb{P}_1[E] = \mathbb{P}_2[E]$ *for all* $E \in \Pi$.

Proof. See, e.g., [Kyt20, Theorem II.26 and Appendix C.3]. □

A.6 Monotone Class Theorem

Consider a set S, for example, $S = \Omega$ or $S = \Omega \times [0, \infty)$.

Monotone class $\mathcal{H}$ is a collection of bounded functions $h : S \to \mathbb{R}$ such that

- ▷ the constant function $1 \in \mathcal{H}$;
- ▷ $\mathcal{H}$ is an $\mathbb{R}$-vector space;
- ▷ if $(h_n)_{n\in\mathbb{N}}$ is a sequence of elements of $\mathcal{H}$ such that $0 \le h_n(x) \uparrow h(x)$ as $n \uparrow \infty$ pointwise for $x \in S$, where the limit $h : S \to \mathbb{R}$ is bounded, then $h \in \mathcal{H}$.

Theorem A.30. (Monotone Class Theorem) *Consider a set S. Let*

- ▷ *$\mathcal{H}$ be a monotone class of bounded functions from S to $\mathbb{R}$, and*
- ▷ *let $\Pi \subset \mathcal{P}(S)$ be a pi-system.*

If $\mathcal{H}$ contains the indicator function $\mathbb{1}_G$ of every set $G \in \Pi$ in the pi-system, then $\mathcal{H}$ contains all bounded measurable functions $h : (S, \sigma(\Pi)) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Proof. See, e.g., [Kyt20, Theorem C.2 and Appendix C.4]. □

References

[Bac00] L. Bachelier. Théorie de la spéculation. *Ann. Sci. Éc. Norm. Supér.*, 17(3): 21–86, 1900.

[BS73] F. Black and M. Scholes. The Pricing of Options and Corporate Liabilities. *J. Political Econ.*, 81(3): 637–654, 1973.

[Bro28] R. Brown. A Brief Account of Microscopical Observations Made in the Months of June, July, and August, 1827, on the Particles Contained in the Pollen of Plants; and on the General Existence of Active Molecules in Organic and Inorganic Bodies. *Edinburgh New Philos. J.*, 5: 358–371, 1828.

[Dup05] B. Duplantier. Brownian Motion, Diverse and Undulating. In *Einstein, 1905–2005*: 201–293, Progress in Mathematical Physics, vol. 47. Birkhäuser Basel, 2005.
Available online: `http://arxiv.org/pdf/0705.1951.pdf`

[Ein06] A. Einstein. Eine neue Bestimmung der Moleküldimensionen. *Ann. Phys.*, 19: 289–306, 1906.

[Fey48] R. P. Feynman. Space-Time Approach to Non-Relativistic Quantum Mechanics. *Rev. Mod. Phys*, 20(2): 367–387, 1948.

[Gou88] L.-G. Gouy. Note sur le Mouvement Brownien. *J. Phys.*, 7(2): 561–564, 1888.

[Jan08] S. Janson. Gaussian Hilbert spaces. Cambridge University Press, 2008.

[Kac47] M. Kac. Random Walk and the Theory of Brownian Motion. *Am. Math. Mon.*, 54(7): 369–391, 1947.

[Kal21] Olav Kallenberg. *Foundations of Modern Probability.* Springer, second edition, 2002.

[KK19] A. Kemppainen and K. Kytölä. *Large random systems.* Lecture notes at Aalto University, 2019. Available online:
`http://math.aalto.fi/~kkytola/files_KK/LRS2019/Large_random_systems-2019.pdf`

[Kin23] J. Kinnunen. *Real analysis.* Lecture notes at Aalto University, 2023. Available online: `http://math.aalto.fi/~jkkinnun/files/real_analysis.pdf`

[Kyt20] K. Kytölä. *Probability theory.* Lecture notes at Aalto University, 2020. Available online:
`http://math.aalto.fi/~kkytola/files_KK/ProbaTh2019/ProbaTh-2019.pdf`

[LeG16] J.-F. Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus.* Springer-Verlag, 2016.

[MP10] P. Mörters and Y. Peres. *Brownian Motion.* Cambridge University Press, 2010.

[OS13] M. Ondrejat and J. Seidler. On existence of progressively measurable modifications. *Electron. Commun. Probab.*, 18: 1–6, 2013.

[Per13] J. B. Perrin. *Les Atomes.* Félix Alcan, Paris, 1913. (Réédition Champs Flammarion, 1991.)

[RY05] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, 2005. Reprint of the 3rd ed. Berlin Heidelberg New York 1999.

[Sut05] W. Sutherland. A Dynamical Theory for Non-Electrolytes and the Molecular Mass of Albumin. *Phil. Mag. S.6*, 9(54): 781–785, 1905.

[Wie23] N. Wiener. Differential-Space. *J. Math. Phys.*, 2(1-4): 131–174, 1923.

[Wil91] David Williams. *Probability with Martingales.* Cambridge University Press, 1991.