MAGIC089 Stochastic Processes (2023) Lecture Notes

J. Bröcker

October 30, 2023

Contents

1 Preliminaries, basic notions of probability theory, the Borel–Cantelli Lemma

The lecture notes are supposed to be as self–contained as possible, but references to other books will be unavoidable. References like [1] refer to the reading list. The notes are subdivided into Lectures numbered 1, 2, $\ldots$. Within each lecture, there will be numbered items. I will use only one counter for Definitions, Remarks, Theorems, Lemmas, etc. Exercises however will have their own counter and typically contain simple technical results which the reader is encouraged to solve herself or himself.

We start the presend section with a reminder of basic notions of probabilities and events. For a more comprehensive account, the reader is referred to [4]. Let Ω, A, B be sets. Familiarity with the notations $A \subset \Omega$, $A \cup B$, $A \cap B$, $\varnothing$ is assumed. Further

$$A \setminus B := \{x \in A; x \notin B\}, \qquad \text{read "}A\text{ without }B\text{"}$$
$$A^{\complement} := \Omega \setminus A, \qquad \text{read "Complement of }A\text{ in }\Omega\text{".}$$

The notation $A^{\complement}$ is used if Ω is clear from the context. If the elements of a set A are again sets, we call A a system or family of sets.

Let Ω be a set. A system $\mathcal{A}$ of subsets of Ω is called an *algebra* if

1. $\varnothing \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow A^{\complement} \in \mathcal{A}$.
3. $A_1, \ldots, A_n \in \mathcal{A} \Rightarrow \bigcup_{k=1}^{n} A_k \in \mathcal{A}$.

Further, $\mathcal{A}$ is a *sigma algebra* if

4. $A_1, A_2, \cdots \in \mathcal{A} \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$.

The elements of a sigma algebra are called *measurable sets*. An algebra formalises the intuition behind "events". Considering sigma algebras rather than just algebras, that is where 3 holds for countably many A_n rather than just finitely many, is important as we will see many times in this course.

Let $\mathcal{A}$ be an arbitrary family of subsets of Ω. Then there is a sigma algebra denoted by $\sigma(\mathcal{A})$ so that

1. $\sigma(\mathcal{A})$ contains $\mathcal{A}$,
2. if $\mathcal{B}$ is another sigma algebra containing $\mathcal{A}$, then $\mathcal{B}$ contains also $\sigma(\mathcal{A})$,

so roughly speaking $\sigma(\mathcal{A})$ smallest sigma algebra containing $\mathcal{A}$. The sigma algebra $\sigma(\mathcal{A})$ is uniquely defined and referred to as the sigma algebra *generated by* $\mathcal{A}$. In Exercise 1.1 you will show that this concept is well defined.

A pair $(\Omega, \mathcal{A})$ with Ω a set and $\mathcal{A}$ a sigma algebra is called a *measurable space*. If $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ are measurable spaces, a mapping

$$f : \Omega_1 \to \Omega_2$$

is called *measurable* if $f^{-1}(A) \in \mathcal{A}_1$ for every $A \in \mathcal{A}_2$. Since sigma algebras can be very large, checking measurability can be a lot of work. The situation might be easier if $\mathcal{A}_2$ is generated by some family $\mathcal{B}$ of sets, that is $\mathcal{A}_2 = \sigma(\mathcal{B})$. Then the measurability condition needs to be checked for sets in $\mathcal{B}$, only, that is, if $f^{-1}(B) \in \mathcal{A}_1$ for every $B \in \mathcal{B}$, then $f^{-1}(A) \in \mathcal{A}_1$ for every $A \in \mathcal{A}_2$.

An important class of measurable spaces arises as follows. Let E be a separable complete metric space. Such spaces are also called *polish* spaces; if you are not familiar with these concepts, you may think of E being $\mathbb{R}^d$ or an open or closed subset thereof. Sets with finite or countably many elements are also polish. The *topology* τ of E is the family of all open sets. We may thus consider the sigma algebra $\sigma(\tau)$ generated by this family. This sigma algebra is called the *Borel algebra* denoted by $\mathcal{B}_E$.

It now makes sense to consider measurable mappings

$$f : (\Omega, \mathcal{A}) \to (E, \mathcal{B}_E),$$

where $(\Omega, \mathcal{A})$ is an arbitrary measurable space and $(E, \mathcal{B}_E)$ a polish space with the Borel algebra. Such a mapping is called a *random variable*.

Let $\mathcal{A}$ be an algebra on some set Ω. A function $\mathbb{P} : \mathcal{A} \longrightarrow [0, 1]$ is a *probability* on $\mathcal{A}$ if it satisfies

1. *Normalisation:* $\mathbb{P}(\Omega) = 1$
2. *Additivity:* If $A_1, \ldots, A_n \in \mathcal{A}$, with $A_i \cap A_j = \varnothing$ for $i \neq j$, then $\sum_{k=1}^n \mathbb{P}(A_k) = \mathbb{P}(\bigcup_{k=1}^n A_k)$.
3. *Continuity at* $\varnothing$*:* If $A_1, A_2, \ldots \in \mathcal{A}$, with $A_1 \supset A_2 \supset \ldots$ and $\bigcap A_j = \varnothing$, then $\mathbb{P}(A_k) \to 0$ for $k \to \infty$.

Again, the intuition behind (1,2) is clear. The continuity at $\varnothing$ is important for technical reasons, see also Exercise 1.3 for several notions equivalent to Additivity and Continuity at $\varnothing$, in case that $\mathcal{A}$ is a sigma algebra. It is possible to construct examples of a function $\mathbb{P}$ that is additive and normalised on an algebra but not continuous at $\varnothing$.

Note that algebras are very much smaller than sigma algebras, and it is generally much easier to define a probability $\mathbb{P}$ on an algebra than on a sigma algebra. On the other hand, given an algebra $\mathcal{A}$, there is a naturally associated sigma algebra, namely $\sigma(\mathcal{A})$. The question then arises if, given probability $\mathbb{P}$ on an algebra $\mathcal{A}$, it is possible to extend $\mathbb{P}$ uniquely onto $\sigma(\mathcal{A})$ (i.e. define it on $\sigma(\mathcal{A})$ without disturbing it on $\mathcal{A}$ where it is already defined). The affirmative answer is the following celebrated theorem.

Theorem 1.1 (The Measure Extension Theorem, also known as MET or Hahn-Carathéodory theorem)**.** *Let $\mathcal{A}$ be an algebra and $\mathbb{P}$ a probability on $\mathcal{A}$. Then there exists a unique probability $\tilde{\mathbb{P}}$ on $\sigma(\mathcal{A})$ with $\tilde{\mathbb{P}}|_{\mathcal{A}} = \mathbb{P}|_{\mathcal{A}}$. Further, if $A \in \sigma(\mathcal{A})$, then for any $\epsilon > 0$ there exist disjoint sets $A_1, \dots, A_n \in \mathcal{A}$ with $\tilde{\mathbb{P}}(A \bigtriangleup \bigcup_{k=1}^n A_k) \le \epsilon$.*

Sketch of a proof, see e.g. [3]. For any $Y \subset \Omega$, put $\mathbb{P}^*(Y) = \inf \sum_{k=1}^\infty \mathbb{P}(A_k)$, inf taken over $A_1, A_2, \dots \in \mathcal{A}$, with $Y \subset \bigcup_k A_k$. Now

1. $\mathbb{P}^*|_{\mathcal{A}} = \mathbb{P}|_{\mathcal{A}}$ ("$\le$" is trivial).
2. Consider the family $\mathcal{M}$ of sets defined as follows: a set $A \subset \Omega$ is a member of $\mathcal{M}$ if $\forall E \subset \Omega$ it holds that $\mathbb{P}^*(E) \ge \mathbb{P}^*(E \cap A) + \mathbb{P}^*(E \setminus A)$. One then proves that $\mathcal{M}$ is a sigma algebra with $\mathcal{M} \supset \mathcal{A}$.
3. $\mathbb{P}^*$ is a measure on $\mathcal{M}$.
4. The approximation result is relatively straightforward from the definition of $\mathbb{P}^*$.

$\square$

We fix the uniqueness part, which is true under weaker conditions:

Theorem 1.2 (Uniqueness of probabilities)**.** *Let $\mathcal{A}$ be a family of sets which is* intersection–stable*, meaning that for any two sets $A_1 \in \mathcal{A}$, $A_2 \in \mathcal{A}$, also $A_1 \cap A_2 \in \mathcal{A}$. (This is true for instance if $\mathcal{A}$ is an algebra.) Further, let $\mathbb{P}, \mathbb{Q}$ be two probabilities on $\sigma(\mathcal{A})$, the sigma algebra generated by $\mathcal{A}$. Then if $\mathbb{P}(A) = \mathbb{Q}(A)$ for any set $A \in \mathcal{A}$, they agree on $\sigma(\mathcal{A})$.*

For a proof see [1], Proposition 2.23.

Fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$; we recall the definition of the integral for a real–valued random variable f with respect to $\mathbb{P}$, which we will write as $\int f(\omega)\, \mathrm{d}\mathbb{P}(\omega)$ or simply as $\int f\, \mathrm{d}\mathbb{P}$ (for details of the construction, see [4], Ch.9). A random variable $f : \Omega \to E$ is real–valued if $E = \mathbb{R}$ and it is non–negative if $E = \mathbb{R}_{\ge 0}$. A non–negative random variable f is *simple* if it assumes only finitely many values; it can then be written as

$$f(\omega) = \sum_{k=1}^n f_k \mathbb{1}_{A_k}(\omega), \tag{1}$$

for some measurable sets $A_1, \dots, A_n$. We define $\int f \; \mathrm{d}\mathbb{P} := \sum_{k=1}^n f_k \mathbb{P}(A_k)$ (the representation in Eq. (1) is not unique but a little algebra shows that the value of the integral is still uniquely defined). For a general non–negative random variable f, we define

$$\int f \; \mathrm{d}\mathbb{P} := \sup \int g(\omega) \; \mathrm{d}\mathbb{P}$$

where the sup is over all simple functions g with the property $0 \leq g \leq f$. (Given f, it is possible to construct a specific sequence $\{g_n, n \in \mathbb{N}\}$ of simple functions so that $g_n \uparrow f$ and $\int g_n \; \mathrm{d}\mathbb{P} \uparrow \int f \; \mathrm{d}\mathbb{P}$.)

Note that the integral of a nonnegative random variable is either a nonnegative number or infinity, but the latter possibility has to be excluded when we define the integral of general real–valued random variables. A general real–valued random variable can be written as $f = f_+ - f_-$, where $f_+ := \max\{f, 0\}$ and $f_- = \max\{-f, 0\}$ are both nonnegative random variables. If $\int |f| \; \mathrm{d}\mathbb{P} < \infty$, we say that f is *integrable* and define the integral of f as $\int f \; \mathrm{d}\mathbb{P} := \int f_+ \; \mathrm{d}\mathbb{P} - \int f_- \; \mathrm{d}\mathbb{P}$. Since $|f| = f_+ + f_-$, the condition $\int |f| \; \mathrm{d}\mathbb{P} < \infty$ implies that *both* $\int f_+ \; \mathrm{d}\mathbb{P}$ and $\int f_- \; \mathrm{d}\mathbb{P}$ are finite. Hence there can be no "$\infty - \infty$"–situation in our definition of the integral of f.

The integral we have just defined is powerful due to the following theorem, which ultimately stems from the continuity of $\mathbb{P}$ at $\varnothing$.

Theorem 1.3. *In this theorem, whenever $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of nonnegative random variables, $\sup_{n\in\mathbb{N}} f_n$ is considered pointwise.*

1. *(Monotone Convergence) If $\{f_n\}_{n\in\mathbb{N}}$ is an increasing sequence of nonnegative random variables (i.e. $f_1(\omega) \leq f_2(\omega) \leq \dots$), then $\sup_{n\in\mathbb{N}} \int f_n \; \mathrm{d}\mathbb{P} = \int \sup_{n\in\mathbb{N}} f_n \; \mathrm{d}\mathbb{P}$.*

2. *(Dominated Convergence) If $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of real–valued random variables with the property that $\int \sup_{n\in\mathbb{N}} |f_n| \; \mathrm{d}\mathbb{P} < \infty$ and which converge pointwise to a function f, then f is integrable and $\int f_n \; \mathrm{d}\mathbb{P} \to \int f \; \mathrm{d}\mathbb{P}$.*

For the rest of this lecture, fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a polish space $(E, \mathcal{B})$ where $\mathcal{B}$ =Borel algebra on E. Further, let I be either $\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{R}_{\geq 0}$ or a closed interval. We will mention when the specific choice is relevant; think of I as a time interval.

Definition 1.4. A *stochastic process* is a family of functions $\{X_t, t \in I\}$, where $X_t : \Omega \to E$ is a random variable for each $t \in I$.

This definition is preliminary only. They are two further ways of looking at a stochastic process $\{X_t, t \in I\}$. Firstly, we may consider it as a mapping $X : \Omega \to E^I, \omega \to \{X_t(\omega), t \in I\}$ that is with values being functions from E to I. Secondly, we may consider it as a mapping $X : \Omega \times I \to E, (\omega, t) \to X_t(\omega)$, that is with values in E while t and ω are arguments. We will later settle on a more precise version of the second interpretation.

We finish this lecture with an important lemma.

Lemma 1.5 (Borel–Cantelli lemma)**.** *Let $A_1, A_2, \dots$ be a sequence of measurable sets and define*

$$A_{i.o.} := \{\omega \in \Omega; \omega \in A_k \text{ for infinitely many } k\} = \bigcap_{n\in\mathbb{N}} \bigcup_{k=n}^{\infty} A_k.$$

If $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_{i.o.}) = 0$.

Proof. $\mathbb{P}(A_{\text{i.o.}}) \le \mathbb{P}(\bigcup_{k=n}^{\infty} A_k)$ for any n, since $A_{\text{i.o.}} \subset \bigcup_{k=n}^{\infty} A_k$ for any n. It follows from the properties of probability (see Exercise 1.3, item 5) that $\mathbb{P}(\bigcup_{k=n}^{\infty} A_k) \le \sum_{k=n}^{\infty} \mathbb{P}(A_k)$. Therefore $\mathbb{P}(A_{\text{i.o.}}) \le \sum_{k=n}^{\infty} \mathbb{P}(A_k)$ for any n. Taking the limit $n \to \infty$, the right–hand–side goes to zero since $\sum_{k=1}^{\infty} \mathbb{P}(A_k) \le \infty$ by assumption. □

Exercises for Section 1

Exercise 1.1. Let Ω be a set.

1. Show that the power set 2^{Ω} is a sigma algebra.
2. Show that $\mathcal{S}_1 \cap \mathcal{S}_2$ is a sigma algebra for any two sigma algebras $\mathcal{S}_1, \mathcal{S}_2$.
3. Show the following extension of the previous item: if $\{\mathcal{S}_\lambda, \lambda \in \Lambda\}$ is an arbitrary family of sigma algebras (indexed by the arbitrary set Λ), then $\mathcal{S} := \bigcap_{\lambda\in\Lambda} \mathcal{S}_\lambda$ is a sigma algebra.
4. Let $\mathcal{A}$ be an arbitrary family of subsets of Ω. Use the previous items to show that $\sigma(\mathcal{A})$ is well defined and unique, i.e. there exist sigma algebras containing $\mathcal{A}$, and among these there exists a smallest possible one.

Exercise 1.2. In this exercise, we learn a bit more about measurable functions and the condition of measurability. Let $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2)$ be measurable spaces (i.e. $\Omega_{1,2}$ are sets with sigma algebras $\mathcal{A}_{1,2}$, resp). Consider a function $f : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$, *not necessarily measurable.*

1. Consider the family $\mathcal{B}_1$ of all sets of the form $f^{-1}(A)$ where $A \in \mathcal{A}_2$. Show that $\mathcal{B}_1$ is a sigma algebra on Ω_1. ($\mathcal{B}_1$ is referred to as the sigma algebra *generated* by f and $\mathcal{A}_2$.)

2. Consider the family $\mathcal{B}_2$ of all sets $B \subset \Omega_2$ so that $f^{-1}(B) \in \mathcal{A}_1$. Show that $\mathcal{B}_2$ is a sigma algebra on Ω_2.

3. Conclude that f is measurable with respect to the sigma algebras $\mathcal{A}_1, \mathcal{A}_2$ if $\mathcal{B}_2$ from the previous item contains $\mathcal{A}_2$.

Exercise 1.3. Let Ω be a set, $\mathcal{A}$ an algebra, $\mathbb{P} : \mathcal{A} \to [0, 1]$ a set function.

1. Show that Additivity implies $\mathbb{P}(\varnothing) = 0$.

2. Show that Additivity and Continuity at $\varnothing$ are equivalent to *sigma additivity*: If $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{A}$ with $A_i \cap A_j = \varnothing$ for any $i \neq j$, and if $\bigcup_k A_k \in \mathcal{A}$ as well, then $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \mathbb{P}(\bigcup_k A_k)$.

3. Show that Additivity and Continuity at $\varnothing$ are equivalent to *continuity from above*: If $A_1, A_2, \ldots \in \mathcal{A}$, with $A_1 \supset A_2 \supset \ldots$ and $\cap A_j = A$ with $A \in \mathcal{A}$, then $\mathbb{P}(A_k) \to \mathbb{P}(A)$ for $k \to \infty$.

4. Show that Additivity and Continuity at $\varnothing$ is equivalent to *continuity from below*: If $A_1, A_2, \ldots \in \mathcal{A}$, with $A_1 \subset A_2 \subset \ldots$ and $\cup A_j = A$ with $A \in \mathcal{A}$, then $\mathbb{P}(A_k) \to \mathbb{P}(A)$ for $k \to \infty$.

5. Show that sigma additivity implies: if $A_1, A_2, \ldots \in \mathcal{A}$, not necessarily pairwise disjoint, then $\mathbb{P}(\cup_{k\in\mathbb{N}} A_k) \leq \sum_{k\in\mathbb{N}} \mathbb{P}(A_k)$.

6. Show that Additivity implies that whenever $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{A}$, we have $\mathbb{P}(A_n) \to 0$ (in fact, $\mathbb{P}(A_n)$ must be summable).

2 Construction of Brownian Motion

We start our discussion of Brownian Motion with some motivation from signal processing. Suppose we observe a function

$$X(t) = f(t) + \xi(t), \qquad t \geq 0, \tag{2}$$

where f is a desired signal and ξ is a sum of many unwanted disturbances coming from various sources, each of them small. In reality, one never observes X at individual points in time. All measurement devices need energy, therefore any observed quantity will always be a temporal average. For instance and analog–digital converter will give you

$$x_n = \frac{1}{h}\int_{(n-1)h}^{nh} X(t)\ \mathrm{d}t = \frac{1}{h}\int_{(n-1)h}^{nh} f(t)\ \mathrm{d}t + \frac{1}{h}\int_{(n-1)h}^{nh} \xi(t)\ \mathrm{d}t = f_n + r_n, \qquad n \in \mathbb{N},$$

where h is the hold time of the AD converter. If f is not too irregular, then $f_n \cong f(nh)$. To analyse $\{r_n, n \in \mathbb{N}\}$, introduce $B(t) = \int_0^t \xi(s)\ \mathrm{d}s$. Then $r_n = \frac{1}{h}\left(B(nh) - B((n-1)h)\right)$, so instead of investigating ξ we investigate B which is a random function of time.

We will argue heuristically that B has three important properties. Firstly, $B(0) = 0$ which is evident. Secondly, the perturbations giving rise to ξ are supposed to be wildly fluctuating with a correlation time much shorter than the hold time h. Hence the values of ξ on disjoint intervals should be independent from one another and thus the same must hold for the increments of B over disjoint intervals. The third property concerns the distribution of the increments of B. If the statistical properties of ξ are time invariant, the distribution of $B(t+s) - B(t)$ should not depend on t but on s only. We assume that the increments $B(t+s) - B(t)$ have mean zero and a finite variance which must be a function of s, say $\phi(s)$. Since

$$B(t_1 + t_2) = \underbrace{B(t_1)}_{A_1} + \underbrace{B(t_1 + t_2) - B(t_1)}_{A_2},$$

with A_1, A_2 independent, taking the variance we obtain the identity

$$\phi(t_1 + t_2) = \phi(t_1) + \phi(t_2), \tag{3}$$

for any $t_1, t_2 \geq 0$. This implies that ϕ is of the form $\phi(t) = at$ (other solutions to Eq.(3) exist but are not measurable so highly irregular). We can assume $a = 1$ (or consider $\frac{1}{\sqrt{|a|}}B$ instead). We now motivate why $B(t)$ has a normal distribution for each $t \geq 0$ (a similar calculation gives that each increment $B(t+s) - B(t)$ is normal, too, with mean zero and variance s).

$$
\begin{aligned}
B(t) &= \sum_{k=1}^{N} B(k\frac{t}{N}) - B((k-1)\frac{t}{N}) \\
&= \sqrt{\frac{t}{N}} \sum_{k=1}^{N} \frac{B(k\frac{t}{N}) - B((k-1)\frac{t}{N})}{\sqrt{t/N}} \\
&\overset{\mathcal{D}}{=} \sqrt{\frac{t}{N}} \sum_{k=1}^{N} Y_k,
\end{aligned}
$$

where $\{Y_k, k \in \mathbb{N}\}$ are iid random variables with mean zero and variance equal to one (and $\overset{\mathcal{D}}{=}$ means "both sides have the same distribution"). Hence by the central limit theorem, $B(t) \overset{\mathcal{D}}{=} \mathscr{N}(0,t)$. The random function $t \to B(t)$ is called *Brownian motion*. Our discussion motivates the following definition

Definition 2.1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Brownian motion is a stochastic process $\{B(t), t \geq 0\}$ with the properties

1. $B(0) = 0$ a.s.
2. The process has independent increments, that is for $0 = t_0 < t_1 < \ldots < t_n$ arbitrary, the increments $\{B(t_k) - B(t_{k-1}), k = 1, \ldots, N\}$ are independent random variables.
3. For any $t, s \geq 0$ the increment $B(t+s) - B(t)$ is normally distributed with mean zero and variance s.
4. There is a measurable set $A \subset \Omega$ with $\mathbb{P}(A) = 1$ so that the function $t \to B(t, \omega)$ is continuous.

A word regarding the function ξ. There cannot be such a function! Indeed assume that $(t, \omega) \to \xi(t, \omega)$ is measurable (this requires a sigma algebra on $I \times \Omega$ but this can be done, permitting the use of Fubini's theorem in the following). Then $0 = \mathbb{E}(B(t+h) - B(t)) = \mathbb{E}\left(\int_t^{t+h} \xi_s \, \mathrm{d}s\right) = \int_t^{t+h} \mathbb{E}(\xi_s) \, \mathrm{d}s$ for any t, h which means that $\mathbb{E}(\xi_s) = 0$ for all $s \in I$. With a similar calculation, using the fact that nonoverlapping increments are uncorrelated, one finds that $\mathbb{E}(\xi_s \xi_t) = 0$ whenever $t \neq s$. Now assume that $\mathbb{E}(\xi_s^2) = c > 0$. We obtain

$$
\begin{aligned}
h &= \mathbb{E}((B(t+h) - B(t))^2) \\
&= \mathbb{E}\left(\int_t^{t+h} \xi_s \,\mathrm{d}s \cdot \int_t^{t+h} \xi_r \,\mathrm{d}r\right) \\
&= \mathbb{E}\left(\int_t^{t+h} \int_t^{t+h} \xi_s \xi_r \,\mathrm{d}s \,\mathrm{d}r\right) \\
&= \int_t^{t+h} \int_t^{t+h} \mathbb{E}(\xi_s \xi_r) \,\mathrm{d}s \,\mathrm{d}r \\
&= c \int_t^{t+h} \int_t^{t+h} \mathbb{1}_{s=r} \,\mathrm{d}s \,\mathrm{d}r \\
&= 0,
\end{aligned}
$$

which is a contradiction, hence the assumption that $\mathbb{E}(\xi_s^2) = c > 0$ must be false—in fact, the variance can only be infinite.

We will soon show that Brownian motion exists using a construction by Norbert Wiener. Many authors introduce Brownian motion differently though, along the following lines (for more details see the appendix). (Note: this will be added at a later stage.)

Our approach proceeds by constructing a mapping

$$X : \Omega \to C(I)$$

where (for now) $I = [0, 1]$, $C(I)$ =continuous functions from $I \to \mathbb{R}$; this space can be equipped with the topology of uniform convergence and subsequently with the Borel algebra related to that topology. Our mapping X will be measurable with respect to that Borel algebra. Furthermore, the mapping $t \to X_t(\omega)$ will be continuous for every ω by construction so property 4 of Brownian motion will be satisfied automatically.

Our construction is fixed in the following

Theorem 2.2. *Assume that there exists a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *with real–valued random variables* $\{Z_{n,k}, n, k \in \mathbb{N}\}$ *which are iid standard normal. Then there exists Brownian Motion on* $I = \mathbb{R}_{\geq 0}$.

Proof. Step 1: We start with defining for each $n \in \mathbb{N}$ a stochastic process $\{B_n(t), t \in [0, 1]\}$ which is continuous, piecewise linear, and with vertices at the dyadic points of order n, that is $D_n := \{\frac{k}{2^n}, k = 0, \ldots, 2^n\}$. Let $U_n := \{\frac{2k+1}{2^n}, k = 0, \ldots, 2^{n-1}\}$ (the "odd" points of order n), then $D_n = D_{n-1} \cup U_n$ (this union is disjoint). In the following, $\{Z_{n,k}; n, k \in \mathbb{N}_0\}$ are iid standard normal. We define B_n inductively; put $B_0(t) = tZ_{0,0}$ for $t \in [0, 1]$. Now assuming B_{n-1} has been defined, we set $B_n(t) = B_{n-1}(t) + F(t)$ for $t \in [0,1]$, where

$$F_n(t) = \begin{cases} 0 & \text{if } t \in D_{n-1}, \\ 2^{-\left(\frac{n+1}{2}\right)} Z_{n,k} & \text{if } t \in U_n, \text{ with } k \text{ so that } t = \frac{2k+1}{2^n}, \\ \text{linearly interpolated in between.} \end{cases}$$

We note that $B_{n+m} = B_n$ on D_n for any m.

Step 2: Define $S_n := \|F_n\|_\infty$. It follows from standard analysis arguments that $\{B_n, n = 1, 2, \ldots\}$ converges uniformly if $\{S_n\}$ is summable, but since the $\{S_n\}$ are random variables, this might happen for some ω yet not for others. We will show that $\{S_n(\omega)\}$ is summable for ω in some set that has probability one. To this end, let us define the "bad sets"

$$A_k := \{\omega \in \Omega : S_k(\omega) \geq c\sqrt{k} 2^{-k/2}\};$$

the constant c will be set below. Now clearly $\{S_n(\omega)\}$ is summable if $S_k(\omega) \leq c\sqrt{k}2^{-k/2}$ except possibly for finitely many k; but that means precisely: ω is *not* in $A_{\text{i.o.}}$. Hence we have to show that $\mathbb{P}(A_{\text{i.o.}}) = 0$ or by Borel–Cantelli, $\sum_{k \in \mathbb{N}} \mathbb{P}(A_k) < \infty$.

Now note that $S_n = 2^{-(n+1)/2} \sup_{0 \leq k \leq 2^n} |Z_{n,k}|$. Therefore $S_n \geq c\sqrt{n}2^{-n/2}$ is equivalent to $\sup_{0 \leq k \leq 2^n} |Z_{n,k}|/\sqrt{2} \geq c\sqrt{n}$. Hence

$$\begin{aligned} \mathbb{P}(A_n) &= \mathbb{P}(|Z_{n,k}| \geq c\sqrt{2n} \text{ for some } k \text{ with } 0 \leq k \leq 2^n) \\ &\leq \sum_{0 \leq k \leq 2^n} \mathbb{P}(|Z_{n,k}| \geq c\sqrt{2n}) \\ &= 2^{n-1} \mathbb{P}(|Z_{n,k}| \geq c\sqrt{2n}). \end{aligned}$$

A standard estimate gives $\mathbb{P}(|Z| \geq c\sqrt{2n}) \leq \exp(-c^2 n)$ for any $c > 1$ and n large enough (see Exercise 2.1). Hence $\mathbb{P}(A_n) \leq 2^{n-1}\exp(-c^2 n)$, and therefore $\sum_{n=0}^\infty \mathbb{P}(A_n) < \infty$ if we set $c > \sqrt{\log(2)}$ so by Borel–Cantelli, $\mathbb{P}(A_{\text{i.o.}}) = 0$.

Step 3: We have constructed the process $\{B\}$ on $[0,1]$, to get a process on $\mathbb{R}_{\geq 0}$, construct independent processes $\{B^{(1)}, B^{(2)}, \ldots\}$, each on $[0,1]$ and then set

$$B(t) := B^{\lfloor t \rfloor}(t - \lfloor t \rfloor) + \sum_{k=1}^{\lfloor t \rfloor - 1} B^{(k)}(1)$$

for any $t \in \mathbb{R}_{\geq 0}$.

Step 4: B is now a randomly selected continuous function, but we still have to check that the distribution of the increments satisfies properties (2,3) of Brownian motion. We will show the equivalent statement that for any $n \in \mathbb{N}$ and any numbers $0 < t_1 < \ldots < t_n \in \mathbb{R}_{\geq 0}$, the random vector $X := (B(t_1), \ldots, B(t_n))$ is normal with mean zero and covariance $\mathrm{Cov}(X)_{i,j} = \min\{t_i, t_j\}$ (see Exercise 2.2).

B has the required properties on $D = \cup_{n\in\mathbb{N}} D_n$ (see Exercise 2.3), and it is easy to see that B has the required properties on $\tilde{D} := \cup_{n\in\mathbb{N}}\{D+n\}$ which is dense in $\mathbb{R}_{\geq 0}$. We use the following

Lemma 2.3. *Let $\{X^{(k)}, k = 1, 2, \ldots\}$ be a series of (n–dimensional) Gaussian random variables so that*

1. *$X^{(k)} \to X$ almost surely,*
2. *$\mathrm{Cov}(X^{(k)}) \to \Gamma$ and $\mathbb{E}(X^{(k)}) \to \mu$,*

then X is Gaussian with $\mathrm{Cov}(X) = \Gamma$ and $\mathbb{E}(X) = \mu$.

Let $0 < t_1 < \ldots < t_n \in \mathbb{R}_{\geq 0}$ and set $X = (B(t_1), \ldots, B(t_n))$ as discussed. There exist $0 < t_1^{(k)} < \ldots < t_n^{(k)} \in \tilde{D}$ with $t_l^{(k)} \stackrel{k\to\infty}{\longrightarrow} t_l$ for all $l = 1, \ldots, n$. We plan to apply the Lemma to $X^{(k)} := (B(t_1^{(k)}), \ldots, B(t_n^{(k)}))$. By continuity of B we obtain that $X^{(k)} \stackrel{k\to\infty}{\longrightarrow} X$. Furthermore, $X^{(k)}$ is normal with mean zero and $\mathrm{Cov}(X^{(k)})_{i,j} = \min\{t_i^{(k)}, t_j^{(k)}\}$. Applying the lemma we find that $X = (B(t_1), \ldots, B(t_n))$ is normal with mean zero and covariance $\mathrm{Cov}(X)_{i,j} = \min\{t_i, t_j\}$, completing the proof. $\square$

Exercises for Section 2

Exercise 2.1. Let X be a standard normal random variable with density $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$. Show that

$$\frac{2x}{1+x^2}\phi(x) \leq \mathbb{P}(|X| > x) \leq \frac{2}{x}\phi(x)$$

Hint: for a differentiable function h with the property $h(x)\phi(x) \to 0$ if $x \to \infty$, show that

$$\phi(x)h(x) = \int_x^\infty \phi(z)(zh(z) - h'(z))\,\mathrm{d}z.$$

Conclude that if the function $x \to xh(x) - h'(x)$ is positive and decreasing for sufficiently large x, we get

$$\frac{\phi(x)h(x)}{xh(x) - h'(x)} \geq \int_x^\infty \phi(z)\,\mathrm{d}z,$$

for sufficiently large x, while the opposite inequality occurs if $x \to xh(x) - h'(x)$ is positive and *in*creasing. Now consider $h(x) = 1/x$ and $h(x) = 1$ as examples.

Exercise 2.2. By reexamining the proof of Theorem 2.2 and the construction of $\{B\}$, show that the increments of $\{B\}$ have the required properties (2,3) on the set D. (Hint: use that any $t \in D$ is in some D_n, and that $B|_{D_n} = B_n|_{D_n}$.)

Exercise 2.3. Show that the increments of $\{B\}$ have the required properties (2,3) of Brownian Motion if and only if for any $n \in \mathbb{N}$ and any numbers $0 < t_1 < \ldots < t_n \in \mathbb{R}_{\geq 0}$, the random vector $X := (B(t_1), \ldots, B(t_n))$ is normal with mean zero and covariance $\mathrm{Cov}(X)_{i,j} = \min\{t_i, t_j\}$.

3 Scaling properties of Brownian Motion and the law of large numbers

Definition 3.1. Consider a stochastic process $\{X_t, t \in I\}$ and let $0 \leq t_1 < \ldots < t_n$ be elements of I. A *marginal* of the stochastic process (corresponding to $t_1, \ldots, t_n$) is the distribution of the random vector $(X_{t_1}, \ldots, X_{t_n})$. A stochastic process is *Gaussian* if all marginals are Gaussian distributions.

Clearly, Brownian Motion is a Gaussian process.

Lemma 3.2. *If $\{B\}$ is Brownian Motion and $a > 0$, then $t \to \frac{1}{a}B(a^2 t)$ is also Brownian Motion.*

Proof. Exercise 3.1 □

Theorem 3.3. *Let $\{X_t, t \geq 0\}$ be some stochastic process on $(\Omega, \mathcal{A}, \mathbb{P})$ so that all marginals on $D := \{\frac{k}{2^n}, k, n \in \mathbb{N}\}$ agree with that of Brownian Motion. Then there exists a measurable set $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$, and a Brownian Motion $\{B_t, t \geq 0\}$ on $(\Omega, \mathcal{A}, \mathbb{P})$ so that*

$$X_t(\omega) = B_t(\omega) \qquad \text{for all } t \in D, \omega \in A.$$

Proof (sketch). For $n \in \mathbb{N}$ put

$$B^{(n)}(t) := \begin{cases} 0 & \text{if } t = 0, \\ X(t) & \text{if } t \in \{\frac{k}{2^n}, k \in \mathbb{N}\}, \\ \text{linear in between.} \end{cases}$$

Then there is $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ so that $B^{(n)}(\omega)$ converges uniformly in t on compact intervals, except perhaps if $\omega \in A^{\complement}$. The proof is exactly as in Theorem 2.2, because the successive errors

$$F^{(n)} = B^{(n)} - B^{(n-1)}, \qquad n \in \mathbb{N} \tag{4}$$

behave exactly like the F_n in that proof. Call the limit $\{B_t, t \geq 0\}$ which is a Brownian Motion which agrees with $\{X\}$ on D by construction. □

Theorem 3.4 (Time inversion). *If $\{B_t, t \geq 0\}$ is Brownian Motion, then there exists a set $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ and a Brownian Motion $\{X_t, t \geq 0\}$ so that*

$$X_t(\omega) = \begin{cases} 0 & \textit{if } t = 0, \\ tB_{1/t}(\omega) & \textit{if } t > 0, \end{cases} \tag{5}$$

for all $\omega \in A$.

Proof. For $t > 0$ and any $\omega \in \Omega$ we define $X_t(\omega)$ as in Equation (5). In Exercise 3.2 you will check that $\{X, t > 0\}$ is a Gaussian process with $\mathbb{E}(X_t) = 0$ and independent increments. Further, for $t > 0$ and $h \geq 0$ we have

$$\begin{aligned}
\operatorname{Cov}(X(t+h), X_t) &= (t+h)t \operatorname{Cov}(B(\frac{1}{t+h}), B(\frac{1}{t})) \\
&= (t+h)t\frac{1}{t+h} \\
&= t.
\end{aligned}$$

Hence the distribution of $\{X_t, t > 0\}$ is that of Brownian Motion. It is also evident that $t \to X_t(\omega)$ is continuous for every ω. It merely remains to check continuity at $t = 0$. From Theorem 3.3, there exists a set $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ and Brownian Motion $\{\tilde{X}_t, t \geq 0\}$ so that $\tilde{X}_t(\omega) = X_t(\omega)$ for all $t \in D \setminus \{0\}$ and $\omega \in A$. Since both $t \to X_t(\omega)$ and $t \to \tilde{X}_t(\omega)$ are continuous for $t > 0$ and D is dense in $\mathbb{R}_{\geq 0}$, we obtain that $\tilde{X}_t(\omega) = X_t(\omega)$ for all $t > 0$ and $\omega \in A$. But they also agree for $t = 0$ by construction, finishing the proof. □

Corollary 3.5 (Law of large numbers)**.**

$$\lim_{t\to\infty} \frac{1}{t} X(t) = 0 \qquad \textit{almost surely.}$$

Proof. $\lim_{t\to\infty} \frac{1}{t} X(t) = \lim_{s\to 0} sX(1/s) = 0$ by Theorem 3.5. □

Exercises for Section 3

Exercise 3.1. Prove Lemma 3.2: If $\{B\}$ is Brownian Motion and $a > 0$, then $t \to \frac{1}{a} B(a^2 t)$ is also Brownian Motion.

Exercise 3.2. Let $\{X_t, t \geq 0\}$ be the process defined in Theorem 3.5. Show that $\{X_t, t > 0\}$ (note "$t > 0$" here) is a Gaussian process with $\mathbb{E}(X_t) = 0$ and independent increments.

4 Non–differentiability of Brownian Motion

The aim of this lecture is to show that for any given $\tau > 0$, the probability that Brownian Motion paths $t \to B_t(\omega)$ are differentiable at τ is zero (although the exceptional set generally depends on τ). More specifically, we aim to prove

Theorem 4.1. *Let $\{B_t, t \geq 0\}$ Brownian Motion on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then for every $t > 0$ there exists $A_t \in \mathcal{A}$ with $\mathbb{P}(A_t) = 1$ so that*

$$D_+B_t := \limsup_{h \downarrow 0} \frac{B(t+h) - B(t)}{h} = \infty, \tag{6}$$

$$D_-B_t := \liminf_{h \downarrow 0} \frac{B(t+h) - B(t)}{h} = -\infty. \tag{7}$$

The proof will be based on the following theorem, which is interesting in its own right (also compare with the Law of Large Numbers, Corollary 3.5):

Theorem 4.2. *Let $\{B_t, t \geq 0\}$ Brownian Motion on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then there exists $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ so that*

$$\limsup_{n \to \infty} \frac{B_n(\omega)}{\sqrt{n}} = \infty, \tag{8}$$

$$\liminf_{n \to \infty} \frac{B_n(\omega)}{\sqrt{n}} = -\infty, \tag{9}$$

whenever $\omega \in A$.

Using this theorem, we can make quick work of Theorem 4.1.

Proof of Theorem 4.1. First note that for t fixed, the process $h \to X(h) := B(t+h) - B(t)$ is also Brownian Motion so we may assume that $t = 0$. We have

$$\begin{aligned} D_+B(0) &= \limsup_{h \downarrow 0} \frac{B(h)}{h} \\ &\geq \limsup_{n \to \infty} \frac{B(1/n)}{1/n} \\ &= \limsup_{n \to \infty} nB(1/n) \\ &= \limsup_{n \to \infty} \hat{B}(n), \end{aligned}$$

where $\hat{B}(t) := tB(1/t)$ for all $t > 0$ is also Brownian Motion by Theorem 3.4. But by Theorem 4.2, we have $\limsup_{n \to \infty} \hat{B}(n) = \infty$ (actually, the theorem shows a stronger statement). This proves statement (6). The statement (7) works analogously. ☐

Some preparation will be necessary to prove Theorem 4.2.

Definition 4.3. Let $\{X_n, n \in \mathbb{N}\}$ random variables with values in $(E, \mathcal{B}_E)$ and let $B \in E^{\mathbb{N}}$. The event $A := \{\omega \in \Omega, (X_1(\omega), X_2(\omega), \ldots) \in B\}$ is *exchangeable* if it is measurable and $A = \{(X_{\sigma(1)}(\omega), X_{\sigma(2)}(\omega), \ldots) \in B\}$ for any permutation σ of finitely many indices.

Example 4.4. The event $\{X_n \geq n$ for infinitely many $n \in N\}$ is exchangeable; the event $\{\sum_{k\in\mathbb{N}} |X_n| 2^{-k} < \infty\}$ is exchangeable; the event $\{X_1 \geq 10\}$ however is *not* exchangeable.

Theorem 4.5 (Hewitt–Savage Zero–One Law)**.** *If $\{X_n, n \in \mathbb{N}\}$ are i.i.d. random variables and A is exchangeable, then $\mathbb{P}(A)$ is either zero or one.*

Proof. See Breiman [1]. □

We will use the Hewitt–Savage Zero–One Law in the

Proof of Theorem 4.2. Step I: For any $c > 0$, we define the events

$$A_c := \{\omega \in \Omega : B_n(\omega) \geq c\sqrt{n} \ \text{ inf'ly often}\}.$$

We may then write

$$\{\omega \in \Omega : \limsup_{n\to\infty} \frac{B_n(\omega)}{\sqrt{n}} = \infty\} \supset \cap_{l\in\mathbb{N}} A_l.$$

We will show (Step II) that A_c is an exchangeable event of i.i.d. random variables for every $c > 0$, hence $\mathbb{P}(A_c) = 0$ or 1 by the Hewitt–Savage Zero–One law, Theorem 4.5. The possibility $\mathbb{P}(A_c) = 0$ will then be excluded in Step III. This will conclude the proof.
Step II: We can write

$$\begin{aligned} \mathbb{P}(\{B(n) \geq c\sqrt{n} \ \text{ inf'ly often}\}) &= \mathbb{P}(\{\sum_{k=1}^{n} B(k) - B(k-1) \geq c\sqrt{n} \ \text{ i.o.}\}) \\ &= \mathbb{P}(\{\sum_{k=1}^{n} X_k \geq c\sqrt{n} \ \text{ i.o.}\}), \end{aligned} \tag{10}$$

if we define $X_k := B(k) - B(k-1)$ for $k \in N$. These are independent and identically distributed random variables, and the event $A_c := \{\sum_{k=1}^n X_k \geq c\sqrt{n}$ i.o.$\}$ is exchangeable. Hence $\mathbb{P}(A_c) = 0$ or 1 by the Hewitt–Savage Zero–One law, Theorem 4.5.

Step III: We will now show that $\mathbb{P}(A_c) > 0$ for any $c > 0$. Fix some $c > 0$. Then

$$\begin{aligned} \mathbb{P}(\{B(n) \geq c\sqrt{n} \ \text{ inf'ly often}\}) &= \mathbb{P}(\cap_{n\in\mathbb{N}} \cup_{k\geq n} \{B(k) \geq c\sqrt{k}\}) \\ &= \lim_{n\to\infty} \mathbb{P}(\cup_{k\geq n}\{B(k) \geq c\sqrt{k}\}) \\ &\geq \mathbb{P}(\cup_{k\geq n}\{B(k) \geq c\sqrt{k}\}) - \epsilon, \end{aligned} \tag{11}$$

where the latter relation holds for any $\epsilon > 0$, provided we take n "large enough", meaning that for any $\epsilon > 0$ there exists an $N_\epsilon \in \mathbb{N}$, and the estimate (11) holds provided $n \geq N_\epsilon$. Since $\mathbb{P}(\cup_{k\geq n}\{B(k) \geq c\sqrt{k}\}) \geq \sup_{k\geq n} \mathbb{P}(\{B(k) \geq c\sqrt{k}\})$ we find

$$\mathbb{P}(\{B(n) \geq c\sqrt{n} \ \text{ inf'ly often}\}) \geq \limsup_{n\to\infty} \mathbb{P}(\{B(n) \geq c\sqrt{n}\}). \tag{12}$$

By scaling (Lemma 3.2), $\mathbb{P}(B(n) > c\sqrt{n}) = \mathbb{P}(B(1) > c) > 0$ for any $c > 0$. $\square$

5 The law of the iterated logarithm

In the law of large numbers (Corollary 3.5), we have seen that $\frac{1}{t}B(t) \to 0$ almost surely if $t \to \infty$. In the previous lecture, we showed Theorem 4.1 from which we can conclude that $\frac{1}{\sqrt{t}}B(t)$ fails to converge if $t \to \infty$. In this lecture we show that there is a function ϕ in between $\sqrt{t}$ and t so that $\frac{1}{\phi(t)}B(t)$ has nonzero limsup and liminf. (But these have to be different; it is easy to see that any limit of $\frac{1}{\phi(t)}B(t)$ as $t \to \infty$ would have to be zero.) We start with a "$t \to 0$" version:

Theorem 5.1 (Law of the iterated logarithm)**.** *There is a set $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ so that for all $\omega \in A$ we have*

$$\limsup_{t \downarrow 0} \frac{B_t(\omega)}{\sqrt{2t \log(\log(1/t))}} = 1.$$

Proof. The proof will be added at a later stage. □

Corollary 5.2 (Law of the iterated logarithm for $t \to \infty$)**.** *There is a set $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ so that for all $\omega \in A$ we have*

$$\limsup_{t \to \infty} \frac{B_t(\omega)}{\sqrt{2t \log \log t}} = 1.$$

Proof. Follows from Theorem 5.1 and time inversion. □

6 The Itô integral

We will now go back to our original motivation for constructing Brownian Motion in Lecture 2, where we formally considered signals of the form (see Eq. 2)

$$x(t) = f(t) + \xi(t), \qquad t \geq 0,$$

where f is a desired signal and ξ is what we can now describe heuristically as the derivative of Brownian Motion (aka *white noise*). By formally integrating Equation (13) however we obtain the following well–defined expression

$$X_t = X_0 + \int_0^t f(s)\, \mathrm{d}s + B_t, \qquad t \geq 0, \tag{13}$$

where B is Brownian Motion, and X can be interpreted as the integral of x in Equation (13).

We would like to take this a step further and, instead of an equation like (13), consider a *stochastic differential equation*

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = f(t, X_t) + g(t, X_t)\xi(t), \qquad t \geq 0, \tag{14}$$

with some initial condition X_0, where now f and g are functions (assumed smooth in both arguments) and ξ is again white noise. Again we formally integrate Equation (14) and obtain

$$X_t = X_0 + \int_0^t f(s, X_s)\, \mathrm{d}s + \int_0^t g(s, X_s)\, \mathrm{d}B_s.$$

Yet now the precise meaning of this has to be clarified. If $\{X\}$ is continuous (something that would need to be established), then the first integral is well defined as a Riemann (or Lebesgue) integral, and it would be a continuous (even differentiable) function of t. The second integral however, which we will call *stochastic integral*, requires more work. Stochastic integrals are of the form $\int_0^t Y_s\, \mathrm{d}B_s$ where $\{Y\}$ can be a process that has, broadly speaking, the same regularity properties as Brownian Motion. (More precise definitions will of course follow.)

Stochastic integrals *cannot* be defined as Riemann–Steltjes integrals as we will now show. Take $Y = B$ as example. Then for a Riemann-Steltjes integral we have

$$\sum_{k=0}^{N-1} B_{t_k}(B_{t_{k+1}} - B_{t_k}) \to \int_0^t B_s\, \mathrm{d}B_s \tag{15}$$

for increasingly fine partitions $0 = t_0 < \ldots < t_N = t$. From the properties of Brownian Motion, we can conclude that $\mathbb{E}\big(\sum_{k=0}^{N-1} B_{t_k}(B_{t_{k+1}} - B_{t_k})\big) = 0$, so $\mathbb{E}\big(\int_0^t B_s \, \mathrm{d}B_s\big) = 0$. On the other hand, using integration by parts (which is permitted for Riemann–Steltjes integrals):

$$\int_0^t B_s \, \mathrm{d}B_s \stackrel{?}{=} B_t^2 - \int_0^t B_s \, \mathrm{d}B_s,$$

and taking expectation we find $\mathbb{E}(B_t^2) = 0$ which contradicts the fact that for Brownian Motion $\mathbb{E}(B_t^2) = t$. We note that $\mathbb{E}\big(\int_0^t B_s \, \mathrm{d}B_s\big) = 0$ depends crucially on how we approximate the integral in Equation (15); taking instead $B_{\tau_k}(B_{t_{k+1}} - B_{t_k})$ with some τ_k in between t_{k+1} and t_k, we would have $\mathbb{E}\big(\int_0^t B_s \, \mathrm{d}B_s\big) \neq 0$ in general, and even the value of the integral depends on how the τ_k are chosen! The value of the Riemann-Steltjes integral would be independent of the specific choice of the τ_k, but Brownian Motion is not regular enough for a stochastic integral to have the properties of the Riemann–Steltjes integral. We will later use (broadly speaking) the approximation (15) to define our integrals as this will give us important properties we would not get for other choices of the τ_k.

We will now explain for which integrands the Itô integral will be defined. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\{B_t, t \geq 0\}$ Brownian Motion.

Definition 6.1. Let $\mathcal{M}([0, T])$ be the class of stochastic processes $\{Y_t, t \in [0, T]\}$ with the properties

1. The mapping $Y : [0, T] \times \Omega \to \mathbb{R}, (t, \omega) \to Y_t(\omega)$ is measurable with respect to $\mathcal{B}([0, T]) \otimes \mathcal{A}$,
2. $\mathbb{E}(\int_0^T Y_t^2 \, \mathrm{d}t) < \infty$,
3. The process $\{Y\}$ is *non–anticipating*, that is for any $t \in [0, T]$, the random variable $\omega \to Y_t(\omega)$ is independent of the increments $\{B(t + h) - B(t), h \geq 0\}$.

The requirement 1 in Definition 6.1 is a technical point; the *product sigma–algebra* $\mathcal{B}([0, T]) \otimes \mathcal{A}$ is defined as the smallest sigma algebra on $[0, T] \times \Omega$ containing all rectangles of the form $B \times A$ where $B \in \mathcal{B}([0, T])$ and $A \in \mathcal{A}$. Basically, the condition ensures that Y^2 can be integrated over t and ω simultaneously (as in item 1 of Definition 6.1), and that the integrals can be interchanged (Fubini–Tonelli). Exercise 6.1 shows that we may always assume that Brownian Motion is a member of $\mathcal{M}([0, T])$ for any T; in particular it satisfies property 1 in Definition 6.1. We will from now on require all stochastic processes to have the property 1:

Convention 6.2. *A stochastic process on I is a mapping $Y : [0,T] \times \Omega \to \mathbb{R}, (t,\omega) \to Y_t(\omega)$ measurable with respect to $\mathcal{B}_I \otimes \mathcal{A}$.*

Our construction of the Itô integral proceeds in several steps.
Step I: A *simple process* is a process of the form

$$Y_t := \sum_{k=1}^{N} Y_{k-1} \mathbb{1}_{[t_{k-1},t_k)},$$

for $0 = t_0 < \ldots < t_N = T$ and random variables $Y_0, \ldots Y_{N-1}$ with the property that $\mathbb{E}(Y_k^2) < \infty$ and Y_k is independent from $\{B(t_k+h) - B(t_k), h \geq 0\}$ for each $k = 0, \ldots, N-1$. For simple processes we define

$$\int_0^T Y_s \,\mathrm{d}B_s := \sum_{k=1}^{N} Y_{k-1}(B_{t_k} - B_{t_{k-1}}).$$

Step II: Check that simple processes are in $\mathcal{M}([0,T])$ and we have

$$\mathbb{E}\Big(\int_0^T Y_s \,\mathrm{d}B_s\Big) = 0, \tag{16}$$

$$\mathbb{E}\big(\big(\int_0^T Y_s \,\mathrm{d}B_s\big)^2\big) = \mathbb{E}\Big(\int_0^T Y_s^2 \,\mathrm{d}t\Big). \tag{17}$$

Step III: On $\mathcal{M}([0,1])$ consider the norm

$$|\!|\!| Y |\!|\!| := \Big(\mathbb{E}\big(\int_0^T Y_s^2 \,\mathrm{d}s\big)\Big)^{1/2}.$$

Now Step II shows that for simple processes Y we have $\mathbb{E}\Big(\big(\int_0^T Y_s \,\mathrm{d}B_s\big)^2\Big) = |\!|\!| Y |\!|\!|^2$ so the mapping $Y \to \int_0^T Y_s \,\mathrm{d}B_s$ is a linear isometry.
Step IV: Suppose that $Y \in \mathcal{M}([0,1])$ and $Y^{(1)}, Y^{(2)}, \ldots$ is a sequence of simple processes in $\mathcal{M}([0,1])$ so that $Y^{(k)} \to Y$ in the $|\!|\!| . |\!|\!|$ norm as $k \to \infty$. Then the integrals $\int_0^T Y_s^{(k)} \mathrm{d}B_s, k = 1, 2, \ldots$ form a Cauchy sequence of random variables in $L_2(\Omega, \mathcal{A}, \mathbb{P})$; we call the limit $\int_0^T Y_s \,\mathrm{d}B_s$.

This finishes the definition of the stochastic integral at least for all processes in $\mathcal{M}([0,T])$ which can be represented as limits of simple processes in the $|\!|\!| . |\!|\!|$ norm. We will now show that this is in fact all of $\mathcal{M}([0,T])$.

Lemma 6.3. *Every process in $\mathcal{M}([0,T])$ is the limit (in $|\!|\!| . |\!|\!|$–norm) of simple processes.*

Proof. Let $Y \in \mathcal{M}([0,T])$. Our goal is to approximate Y with simple processes in the $\|\!|.|\!\|$–norm. We will first impose more assumptions on Y which get subsequently lifted.
Step I: Assume that Y is bounded (i.e. $|Y(t,\omega)| \leq C$ for some C and all $t \in [0,T], \omega \in \Omega$), and further $t \to Y(t,\omega)$ is continuous for every $\omega \in \Omega$. For a partition $\Pi := \{(t_0, \ldots t_N); 0 = t_0 < \ldots < t_N = T\}$ of the interval $[0,T]$ with *resolution* $|\Pi| := \max\{t_k - t_{k-1}, k = 1 \ldots N\}$ we define the simple process $Y_\Pi(t,\omega) = \sum_{k=1}^{N} Y(t_{k-1},\omega)\mathbb{1}_{[t_{k-1},t_k)}$. (The reader should check that this is indeed a simple process.) If $\Pi_n, n = 1, 2, \ldots$ is a sequence of partitions so that $|\Pi_n| \to 0$, then since Y is continuous in t it follows from standard analysis results that

$$\int_0^T (Y(t,\omega) - Y_{\Pi_n}(t,\omega))^2 \, \mathrm{d}t \to 0$$

as $n \to \infty$. Since furthermore Y is bounded we obtain $\int_0^T (Y(t,\omega) - Y_{\Pi_n}(t,\omega))^2 \, \mathrm{d}t \leq 4C$ and therefore by the bounded convergence theorem we get $\|\!|Y - Y_{\Pi_n}|\!\|^2 = \mathbb{E}(\int_0^T (Y(t,\omega) - Y_{\Pi_n}(t,\omega))^2 \, \mathrm{d}t) \to 0$.
Step II: We drop the continuity assumption but still require $|Y(t,\omega)| \leq C$. For each $n \in \mathbb{N}$ consider a nonnegative and continuous function ψ_n on $\mathbb{R}$ so that $\psi_n(t) = 0$ if either $t > 0$ or $t < -1/n$ and further $\int_{\mathbb{R}} \psi_n(t) \, \mathrm{d}t = 1$. Now put $Y^{(n)}(t,\omega) := \int_0^t \psi(t-s)Y(s,\omega) \, \mathrm{d}s$. Now $Y^{(n)}$ is bounded (by C), continuous, and a member of $\mathcal{M}([0,T])$ for every $n \in \mathbb{N}$. (Use Fubini's theorem to show that $Y^{(n)}$ is non–anticipating.) Now for each $\omega \in \Omega$ we get from Lebesgue's differentiation theorem (see [2], C4) that $Y^{(n)}(t,\omega) \to Y(t,\omega)$ for almost all $t \in [0,T]$. Again by bounded convergence (over the t variable) we get $\int_0^T (Y^{(n)}(t,\omega) - Y(t,\omega))^2 \, \mathrm{d}t \to 0$ as $n \to \infty$. As in the previous step, this implies $\|\!|Y^{(n)} - Y|\!\| \to 0$ as $n \to \infty$.
Step III: Finally, let Y be an arbitrary element of $\mathcal{M}([0,1])$ and put $Y^{(n)} := (Y \wedge n) \vee (-n)$. Then $|Y^{(n)}| \leq |Y|$ for all n and also $Y^{(n)}(t,\omega) \to Y(t,\omega)$ for all t, ω. Therefore $\|\!|Y^{(n)} - Y|\!\| \to 0$ by dominated convergence theorem. □

Exercises for Section 6

Exercise 6.1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\{B_t, t \geq 0\}$ Brownian motion, and $\mathcal{M}([0,T])$ as defined in Definition 6.1. Show that there is a set $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ so that the process

$$\tilde{B}_t(\omega) = \begin{cases} B_t(\omega) & \text{if } \omega \in A, \\ 0 & \text{else.} \end{cases}$$

is an element of $\mathcal{M}([0,T])$. In particular, it is a stochastic process in the sense of Convention 6.2. (Hint: apply the same reconstruction process as in Theorem 3.3 and show that the process constructed in this way indeed adheres to Convention 6.2.)

Exercise 6.2. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\{B_t, t \in [0,T]\}$ Brownian motion. For a partition $\Pi := \{(t_0, \dots t_N); 0 = t_0 < \dots < t_N = T\}$ of the interval $[0,T]$ with resolution $|\Pi| := \max\{t_k - t_{k-1}, k = 1 \dots N\}$ consider the *quadratic variation* of Brownian motion

$$S_\Pi := \sum_{k=1}^{N} (B_{t_k} - B_{t_{k-1}})^2$$

Show that $S_\Pi \to T$ in L_2 if $|\Pi| \to 0$. Compare this with the quadratic variation of a function $t \to f(t)$ (stochastic or not) which has a continuous derivative.

7 Properties of the Itô integral; martingales

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\{B_t, t \geq 0\}$ Brownian Motion. Further, let $\{Y_t, t \in [0,T]\}$ be a stochastic process in $\mathcal{M}([0,T])$. We want to study properties of the stochastic integral $\int_0^t Y_s \, \mathrm{d}B_s$, including as a function of the upper limit t. As one would expect, the integral is linear in Y and also additive with regards to the time interval we are integrating over. Furthermore, we will see (after some work) that it is a stochastic process in the sense of definition 6.2. Here we will need to sort out an important yet subtle technical difficulty which arises because we have defined the integral $\int_0^t Y_s \, \mathrm{d}B_s$ separately for each t and these definitions now have to be spliced together to form a stochastic process. Finally, we will see that this process is a *martingale*, an important class of stochastic processes which we will learn more about.

We start with the following important lemma which we have already used and proved in the construction of the integral.

Lemma 7.1. *For $Y \in \mathcal{M}([0,T])$ we have*

$$\mathbb{E}\big(\big(\int_0^T Y_t \, \mathrm{d}B_t\big)^2\big) = \|Y\|^2.$$

In particular, if $Y^{(n)}, n \in \mathbb{N}$ are elements of $\mathcal{M}([0,T])$ with $Y^{(n)} \overset{\|\cdot\|}{\to} Y$, then $\int_0^T Y_t^{(n)} \, \mathrm{d}B_t \overset{L_2}{\to} \int_0^T Y_t \, \mathrm{d}B_t$.

Elementary properties of the Itô integral are the following.

Lemma 7.2. *Let $X, Y \in \mathcal{M}([0,T])$. Then*

1. *for $\lambda, \mu \in \mathbb{R}$ we have $\lambda X + \mu Y \in \mathcal{M}([0,T])$ and*

$$\int_0^T (\lambda X_t + \mu Y_t) \, \mathrm{d}B_t = \lambda \int_0^T X_t \, \mathrm{d}B_t + \mu \int_0^T Y_t \, \mathrm{d}B_t,$$

2. *for $0 < S < T$ we have $X|_{[0,S]} \in \mathcal{M}([0,S])$ as well as $X|_{[S,T]} \in \mathcal{M}([S,T])$, and*

$$\int_0^S X_t \, \mathrm{d}B_t + \int_S^T X_t \, \mathrm{d}B_t = \int_0^T X_t \, \mathrm{d}B_t,$$

3. *The integral $\int_0^T X_t \, \mathrm{d}B_t$ is independent from $\{B(T+h) - B(T), h \geq 0\}$*

The identities in items 1 and 2 hold for all ω in some set A with $\mathbb{P}(A) = 1$ but in general, that set depends on λ, μ (in item 1) resp on S, T (in item 2).

Proof. These identities hold for simple processes and therefore, using the isometry, they hold for all processes in $\mathcal{M}([0,T])$. Details can be found in Exercise 7.1. $\square$

For the remainder of this course, we will work with a set of stochastic processes $\mathcal{V}([0,T])$ as integrands which is smaller than $\mathcal{M}([0,T])$. This will simplify the exposition.

Definition 7.3. Let $\mathcal{V}([0,T])$ be the class of stochastic processes satisfying items 1 and 2 in the definition of $\mathcal{M}([0,T])$ (see Def. 6.1), but instead of item 3 we have

3. for any $t \in [0,T]$ the random variable $\omega \to Y_t(\omega)$ is measurable with respect to $\{B(s), 0 \leq s \leq t\}$.

We will later see that $\mathcal{V}([0,T]) \subset \mathcal{M}([0,T])$.

As said, we want to establish that the stochastic integral is a stochastic process (in fact an element of $\mathcal{V}([0,T])$) as a function of the upper integration limit. We recall that the integral $\int_0^t Y_s \, \mathrm{d}B_s$ is constructed for each $t \in [0,T]$ separately. This results in a set of random variables indexed by t but this will not in general result in a stochastic process in the sense of Definition 6.2, as will fail to have the required joint measurability in ω and t. The t–dependence of the exceptional set mentioned at the end of Lemma 7.2 is a related issue. Similarly, it will be difficult to prove any regularity property in the variable t, with ω fixed, such as continuity.

On the other hand, for every t the random variable $\int_0^t Y_s \, \mathrm{d}B_s$ is constructed by approximating $Y|_{[0,t]}$ in $\mathcal{M}([0,t])$ with $\vert\!\vert\!\vert . \vert\!\vert\!\vert$ by simple processes $Y^{(n,t)}, n \in \mathbb{N}$. Yet there is freedom in chosing the $\{Y^{(n,t)}\}$; chosing different approximating sequences will strictly speaking result in different limits for the stochastic integral, but any two of these must agree on a set of unit probability. This freedom can be exploited to achieve better regularity in t, essentially by controlling the regularity in t of the approximating sequence $\{Y^{(n,t)}\}$ of simple processes. The following theorem, which we will not prove but use extensively, explores this avenue.

Theorem 7.4. *If $Y \in \mathcal{M}([0,T])$, then there exists a stochastic process $\{Z_t, t \in [0,T]\}$ which is continuous in t for all ω and so that*

$$\int_0^t Y_s \, \mathrm{d}B_s = Z_t,$$

for all $t \in [0,T]$, and for $\omega \in A_t$, where $A_t \in \mathcal{A}$ with $\mathbb{P}(A_t) = 1$.

Note the t–dependence of the exceptional set A_t.

Proof. See [5], Theorem 3.2.5. It follows from the proof that there exists a sequence $\{Y^{(n)}, n \in \mathbb{N}\}$ of simple processes in $\mathcal{M}([0,T])$ and a set $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ so that for all $\omega \in A$,

$$\int_0^t Y_s^{(n)} \, \mathrm{d}B_s \to Z_t$$

uniformly in t. □

Theorem 7.4 may be seen as a generalisation of our construction of Brownian Motion (Thm. 2.2) where we have constructed a sequence of simple processes which (up to a set of measure zero) converge uniformly to Brownian Motion. In the same way that Theorem 2.2 proved the continuity of Brownian Motion, Theorem 7.4 proves that the stochastic integral is continuous as a function of the upper limit.

Definition 7.5. 1. A *filtration* (on $(\Omega, \mathcal{A})$) is a family $\{\mathcal{F}_t, t \geq 0\}$ of sigma algebras such that $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{A}$ for any $0 \leq s \leq t$.

2. A stochastic process $\{Y_t, t \geq 0\}$ is *adapted* to a given filtration $\{\mathcal{F}_t, t \geq 0\}$ if Y_t is $\mathcal{F}_t$–measurable for every $t \geq 0$.

3. A stochastic process $\{M_t, t \geq 0\}$ is called a *martingale* with respect to a given filtration $\{\mathcal{F}_t, t \geq 0\}$ if

   (a) M is $\mathcal{F}_t$–adapted,

   (b) $\mathbb{E}(|M_t|) < \infty$ for all $t \geq 0$,

   (c) $\mathbb{E}(M_t | \mathcal{F}_s) = M_s$ for all $0 \leq s \leq t$.

Associated with every stochastic process $\{Y_t, t \geq 0\}$ is a filtration $\{\mathcal{F}_t, t \geq 0\}$ given by $\mathcal{F}_t := \sigma\{Y_s; s \leq t\}$ for each $t \geq 0$. It is called the *natural* filtration of Y or the filtration *generated by* Y. Of course any process is adapted to its own natural filtration. The property 3 of Definition 7.3 says that any member $Y \in \mathcal{V}([0,T])$ is adapted to the filtration generated by Brownian motion. A martingale (with respect to some filtration $\{\mathcal{F}\}$) is automatically a martingale with respect to its natural filtration.

Lemma 7.6. *Brownian motion is a martingale with respect to its natural filtration.*

Proof. Fix t, s with $0 \leq s < t$ and let $\{\mathcal{F}_t\}$ be the filtration generated by Brownian motion. We know that $\mathbb{E}(|B_t|) \leq \infty$ and furthermore Brownian motion has independent increments. This implies that $B_t - B_s$ is independent from $\mathcal{F}_s$ (this is an exercise in probability theory). Hence $\mathbb{E}(B_t - B_s|\mathcal{F}_s) = 0$ which implies the martingale property. $\square$

Theorem 7.7. *Let $Y \in \mathcal{V}([0,T])$. Then $t \to \int_0^t Y_s \, \mathrm{d}B_s$ is a continuous martingale with respect to the natural filtration of Brownian motion.*

Proof. Let $\{\mathcal{F}_t\}$ be the filtration of Brownian motion. The statement is true if $Y \in \mathcal{V}([0,T])$ is a simple process (see Exercise 7.3). Now let Y be an arbitrary element of $\mathcal{V}([0,T])$ and let $Z_t := \int_0^t Y_s \, \mathrm{d}B_s$ (where we pick a continuous version as per Theorem 7.4). It follows from the proof of Theorem 7.4 that there exists a sequence $\{Y^{(n)}, n \in \mathbb{N}\}$ of simple processes in $\mathcal{V}([0,T])$ so that the processes $t \to Z_t^{(n)} := \int_0^t Y_s \, \mathrm{d}B_s$ are martingales, and for every $t \in [0,T]$ we have $Z_t^{(n)} \to Z_t$ in L_2 as $n \to \infty$. From the Martingale property we have $\mathbb{E}(Z_t^{(n)} - Z_s^{(n)}|\mathcal{F}_s) = 0$ for any $0 \leq s \leq t$, but by the L_2 convergence, the left hand side converges to $\mathbb{E}(Z_t - Z_s|\mathcal{F}_s)$ in L_2 as well, so this quantity must be zero. This concludes the proof. $\square$

Exercises for Section 7

Exercise 7.1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\{B_t, t \geq 0\}$ Brownian motion, and $\mathcal{M}([0,T])$ as defined in Definition 6.1. If $0 < S < T$, then the definition of $\mathcal{M}([S,T])$ should also be clear from a minor modification of Definition 6.1.

1. Show that the properties listed in Lemma 7.2 hold if X, Y are simple processes in $\mathcal{M}([0,T])$.

2. Show that if the properties listed in Lemma 7.2 hold for sequences $\{X_n\}$ and $\{Y_n\}$ of elements in $\mathcal{M}([0,T])$ so that $X_n \to X$ and $Y_n \to Y$ in the $\|\!|.|\!\|$ norm, then properties listed in Lemma 7.2 also hold for X, Y.

Exercise 7.2. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\{B_t, t \geq 0\}$ Brownian motion.

1. Show that $\{B^2 - t, t \geq 0\}$ is a martingale with respect to the natural filtration of Brownian motion.

2. Show that if $\{Y_t, t \geq 0\}$ is a martingale, then $\mathbb{E}(Y_t)$ does not depend on t.

Exercise 7.3. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\{B_t, t \geq 0\}$ Brownian motion. Prove Theorem 7.7 for simple processes $Y \in \mathcal{V}([0,T])$: The process $t \to \int_0^t Y_s \, \mathrm{d}B_s$ is a martingale with respect to the natural filtration of Brownian motion.

8 Itô processes and the Itô formula

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\{B_t, t \geq 0\}$ Brownian Motion.

Definition 8.1 (Itô processes)**.** An *Itô process* is a process $\{X_t, t \in [0,T]\}$ of the form

$$X_t = X_0 + \int_0^t Y_s \,\mathrm{d}s + \int_0^t Z_s \,\mathrm{d}B_s,$$

with $Y, Z \in \mathcal{V}([0,T])$. The first (Riemann) integral is called the *absolutely continuous* part, the second (Itô) integral is called the *martingale part*.

It turns out that $X \in \mathcal{V}([0,T])$ as well:

Lemma 8.2. *If X is an Itô process on $[0,T]$, then $X \in \mathcal{V}([0,T])$ and in particular $\|X\| < \infty$.*

Proof. See Exercise 8.1. □

Definition 8.3. If $\{X_t, t \in [0,T]\}$ is an Itô process of the form

$$X_t = X_0 + \int_0^t f_s \,\mathrm{d}s + \int_0^t g_s \,\mathrm{d}B_s, \tag{18}$$

and $Y \in \mathcal{V}([0,T])$ so that also $gY \in \mathcal{V}([0,T])$, then we define the Itô integral of Y against X as

$$\int_0^t Y_s \,\mathrm{d}X_s := \int_0^t Y_s f_s \,\mathrm{d}s + \int_0^t Y_s g_s \,\mathrm{d}B_s. \tag{19}$$

We see that the Itô integral of $Y \in \mathcal{V}([0,T])$ against an Itô process X is again an Itô process. In order to remember the formula (19), we may write Equation (18) heuristically in "differential form" as

$$\mathrm{d}X_t = f_t \mathrm{d}t + g_t \mathrm{d}B_t,$$

and by multiplying with Y_t we obtain Equation (19) in differential form:

$$Y_t \mathrm{d}X_t = Y_t f_t \mathrm{d}t + Y_t g_t \mathrm{d}B_t.$$

In Exercise 8.2 we will prove an approximation result for integrals against Itô processes, similiar to integrals against Brownian Motion.

Generalising from integration against Brownian Motion to integration against Itô processes is necessary when dealing with nonlinear mappings, as we shall now explain. If $\{X_t; t \in [0,T]\}$ is an Itô process (with values in $\mathbb{R}$) and $h : \mathbb{R} \to \mathbb{R}$ a nonlinear, differentiable function, then (subject to some integrability conditions) the process $\{h(X_t); t \in [0,T]\}$ will again be an Itô process; this is the main content of Itô's theorem 8.4. It turns out that in general, $\{h(X_t)\}$ will have a nontrivial absolutely continuous part *even if* X does not have an absolutely continuous part (i.e. X is merely an integral against Brownian Motion). This might seem counterintuitive, as it is different from the "classical" situation where $\{X_t; t \in [0,T]\}$ has only an absolutely continuous part, that is $X_t = X_0 + \int_0^t f_s \, \mathrm{d}s$. Then by the ordinary chain rule $h(X_t) = h(X_0) + \int_0^t h'(X_s) f_s \, \mathrm{d}s$, so again $\{h(X_t)\}$ has only an absolutely continuous part. If $\{X_t; t \in [0,T]\}$ has only a martingale part however, that is $X_t = X_0 + \int_0^t g_s \, \mathrm{d}B_s$, then in general $h(X_t)$ is *not* given by $h(X_0) + \int_0^t h'(X_s) g_s \, \mathrm{d}B_s$. We have already seen for example that $B_t^2 \neq 2\int_0^t B_s \, \mathrm{d}B_s$; further examples can be found in the Exercises.

Theorem 8.4 (The Itô formula). *Let $\{X_t, t \in [0,T]\}$ be an Itô process of the form*

$$X_t = X_0 + \int_0^t f_s \, \mathrm{d}s + \int_0^t g_s \, \mathrm{d}B_s \qquad s \in [0,T], \tag{20}$$

and $h : [0,T] \times \mathbb{R} \to \mathbb{R}$ a function with $\partial_x h, \partial_t h, \partial_{xx}^2 h$ all continuous and bounded. Then $Y_t := h(t, X_t)$ for $t \in [0,T]$ defines again an Itô process and almost surely for each $t \in [0,T]$:

$$Y_t = Y_0 + \int_0^t \left\{ \partial_t h(s, X_s) + \frac{1}{2}\partial_{xx}^2 h(s, X_s)\, g_s^2 \right\} \mathrm{d}s + \int_0^t \partial_x h(s, X_s) \, \mathrm{d}X_s. \tag{21}$$

Here's how to remember the Itô formula (21), which heuristically can be written in differential form as

$$\mathrm{d}Y_t = \left\{ \partial_t h(s, X_s) + \frac{1}{2}\partial_{xx}^2 h(s, X_s)\, g_s^2 \right\} \mathrm{d}s + \partial_x h(s, X_s) \mathrm{d}X_s. \tag{22}$$

To obtain this formula in a formal way, use Taylor expansion to second order:

$$\begin{aligned} \mathrm{d}Y_t &= h(t + \mathrm{d}t, X_t + \mathrm{d}X_t) - h(t, X_t) \\ &= \partial_t h(t, X_t)\mathrm{d}t + \partial_x h(t, X_t)\mathrm{d}X_t \\ &\quad + \frac{1}{2}\partial_{tt}^2 h(t, X_t)(\mathrm{d}t)^2 + \frac{1}{2}\partial_{xx}^2 h(t, X_t)(\mathrm{d}X_t)^2 + \partial_{xt}^2 h(t, X_t)\mathrm{d}t\mathrm{d}X_t. \end{aligned} \tag{23}$$

It turns out that $(\mathrm{d}t)^2$ and $\mathrm{d}t\mathrm{d}X_t$ are higher–than–first–order terms which can be ignored when integrating over t, while $(\mathrm{d}X_t)^2$ is actually a term of order $\mathrm{d}t$ which cannot be ignored. To see this, and to find an expression for

$(\mathrm{d}X_t)^2$, we use the representation $\mathrm{d}X_t = f_t\mathrm{d}t + g_t\mathrm{d}B_t$ (a heuristic form of Eq.20). This gives

$$\begin{aligned}(\mathrm{d}X_t)^2 &= f_t^2(\mathrm{d}t)^2 + g_t^2(\mathrm{d}B_t)^2 + 2f_tg_t\mathrm{d}t\mathrm{d}B_t,\\ \mathrm{d}t\mathrm{d}X_t &= f_t(\mathrm{d}t)^2 + g_t\mathrm{d}t\mathrm{d}B_t.\end{aligned}$$

Note first that if we treat $(\mathrm{d}t)^2$, $\mathrm{d}t\mathrm{d}B_t$ and $(\mathrm{d}B_t)^2$ according to the following multiplication table:

$$\begin{array}{ccc} \times & \mathrm{d}t & \mathrm{d}B_t \\ \mathrm{d}t & 0 & 0 \\ \mathrm{d}B_t & 0 & \mathrm{d}t, \end{array}$$

we get $\mathrm{d}t\mathrm{d}X_t = 0$ and $(\mathrm{d}X_t)^2 = g_t^2\mathrm{d}t$. Replacing with this in Equation (23) we obtain Equation (22).

Why is it justified to treat the higher–order terms in this way upon integration? Let $\Pi := \{0 = t_0 < \ldots < t_N = T\}$ be a partition of the interval $[0,T]$ and write $|\Pi| := \max_{k=1,\ldots,N}|t_k - t_{k-1}|$. Now heuristically (for some continuous function r)

$$\begin{aligned}\text{“}\int_0^T r_t\,\mathrm{d}t\mathrm{d}B_t\text{”} &\cong \sum_{k=1}^{N} r_{t_k}(t_{k+1}-t_k)(B_{t_{k+1}} - B_{t_k})\\ &\leq \max_m |B_{t_{m+1}} - B_{t_m}| \cdot \sum_{k=1}^{N} |r_{t_k}|(t_{k+1}-t_k).\end{aligned}$$

If $|\Pi| \to 0$, the second term converges to $\int_0^T |r_t|\,\mathrm{d}t$ while the first converges to zero since the paths of B are uniformly continuous on $[0,T]$. Therefore we may set $\mathrm{d}t\mathrm{d}B_t = 0$ as claimed. A similar argument holds for terms of order $(\mathrm{d}t)^2$. Terms of order $(\mathrm{d}B_t)^2$ however behave differently; to analyse those terms it is essential to assume that $r \in \mathcal{V}([0,T])$. Then

$$\begin{aligned}\text{“}\int_0^T r_t\,(\mathrm{d}B_t)^2\text{”} &\cong \sum_{k=1}^{N} r_{t_k}(B_{t_{k+1}} - B_{t_k})^2\\ &= \underbrace{\sum_{k=1}^{N} r_{t_k}(t_{k+1}-t_k)}_{=:S_I}\\ &\quad + \underbrace{\sum_{k=1}^{N} r_{t_k}\left((t_{k+1}-t_k) - (B_{t_{k+1}} - B_{t_k})^2\right)}_{=:S_{II}}.\end{aligned}$$

The sum S_I on the right hand side converges to $\int_0^T r_t \, \mathrm{d}t$ if $|\Pi| \to 0$. The second sum S_{II} converges to zero in L_2 if $|\Pi| \to 0$, as we will now demonstrate. We use the abbreviations $\Delta_k t := t_{k+1} - t_k$ and $\Delta_k B := B_{t_{k+1}} - B_{t_k}$; the terms in S_{II} can now be written as $r_{t_k}(\Delta_k t - (\Delta_k B)^2)$. Using that r_{t_k} is independent from $\Delta_k t - (\Delta_k B)^2$ for $k = 1, \ldots, N$ (since $r \in \mathcal{V}([0,T])$) as well as the properties of the increments $\Delta_k B$ we obtain

$$\begin{aligned} \mathbb{E}\big[r_{t_k}(\Delta_k t - (\Delta_k B)^2)\big] &= 0, \\ \mathbb{E}\big[r_{t_k} r_{t_l}(\Delta_k t - (\Delta_k B)^2)(\Delta_l t - (\Delta_l B)^2)\big] &= \begin{cases} 2\mathbb{E}\big[r_{t_k}^2\big](\Delta_k t)^2 & \text{if } k = l, \\ 0 \quad \text{else.} \end{cases} \end{aligned} \tag{24}$$

Therefore $\mathbb{E}\big[S_{II}^2\big] = 2\sum_{k=1}^N \mathbb{E}\big[r_{t_k}^2\big](\Delta_k t)^2 \to 0$ as $|\Pi| \to 0$. This demonstrates, on a heuristic level, that $(\mathrm{d}B_t)^2 = \mathrm{d}t$.

Proof of Itô's theorem. The proof of Itô's theorem basically proceeds by making the discussed heuristics rigorous. For the sake of simplicity, we will assume that h does not depend on t. We can also take t in Equation (21) to be equal to T (the proof for general $t \in [0,T]$ is the same). Furthermore, there are several steps that will be relegated to the exercises. Let $\Pi = \{0 = t_0 < \ldots < t_n = T\}$ a partition of the interval $[0,T]$; as before we write $|\Pi| = \max_{k=1,\ldots,n} t_k - t_{k-1}$.

Step I: Using Taylor's theorem, we obtain

$$\begin{aligned} Y_T &= Y_0 + \sum_{k=1}^n h(X_{t_k}) - h(X_{t_{k-1}}) \\ &= Y_0 + \sum_{k=1}^n \partial_x h(X_{t_k})(X_{t_k} - X_{t_{k-1}}) && (25) \\ &\quad + \frac{1}{2}\sum_{k=1}^n (\partial_{xx}^2 h(X_{t_k}) + R_k)(X_{t_k} - X_{t_{k-1}})^2, && (26) \end{aligned}$$

where for Taylor's remainder terms we have $S_n := \max_k |R_k| \to 0$ almost surely if $|\Pi| \to 0$.

Step II: The sum in (25) converges in $L_2(\Omega, \mathcal{A}, \mathbb{P})$ to $\int_0^T \partial_x h(X_s) \, \mathrm{d}X_s$ if $|\Pi| \to 0$. To show this, according to Exercise 8.2 we need to prove that the approximating process $t \to \sum_{k=1}^n \partial_x h(X_{t_k}) \mathbb{1}_{[t_k, t_{k+1})}(t)$ converges in the $|||.|||$ norm to the process $t \to \partial_x h(X_t)$. This holds because X is continuous in t and $\partial_x h$ is continuous and bounded (see Exercise 8.5 for details).

Step III: We now consider the sum in (26):

$$\frac{1}{2}\sum_{k=1}^{n}(\partial^2_{xx}h(X_{t_k}) + R_k)(X_{t_k} - X_{t_{k-1}})^2$$
$$= \frac{1}{2}\sum_{k=1}^{n}\partial^2_{xx}h(X_{t_k})(X_{t_k} - X_{t_{k-1}})^2 \tag{27}$$
$$+ \frac{1}{2}\sum_{k=1}^{n}R_k(X_{t_k} - X_{t_{k-1}})^2 \tag{28}$$

To continue we need the following Lemma which we will use several times in the following:

Lemma 8.5. *Suppose $Y_n, Z_n, n = 1, 2, \dots$ are sequences of nonnegative random variables so that $\mathbb{E}(Z_n) \leq \zeta$ for all n and $Y_n \to 0$ almost surely as $n \to \infty$. Then $Y_n Z_n, n = 1, 2, \dots$ converges to zero in probability.*

Proof. For any $x, \epsilon > 0$ we have

$$\begin{aligned}\mathbb{P}(Y_n Z_n \geq \epsilon) &= \mathbb{P}(\{Y_n Z_n \geq \epsilon\} \cap \{Z_n < x\}) + \mathbb{P}(\{Y_n Z_n \geq \epsilon\} \cap \{Z_n \geq x\})\\ &\leq \mathbb{P}(Y_n x \geq \epsilon) + \mathbb{P}(Z_n \geq x)\\ &\leq \mathbb{P}(Y_n x \geq \epsilon) + \frac{\zeta}{x},\end{aligned}$$

the last inequality following from Markov's inequality. We may now choose first x to make the second term small. Since $Y_n x$ converges to zero in probability, we can now render the first term arbitrarily small by chosing n large enough. □

We will use Lemma 8.5 to show that the sum in (28) converges to zero in probability. Since $|\sum_{k=1}^{n} R_k(X_{t_k} - X_{t_{k-1}})^2| \leq S_n \sum_{k=1}^{n}(X_{t_k} - X_{t_{k-1}})^2$ and $\{S_n, n \in \mathbb{N}\}$ converges to zero almost surely if $|\Pi| \to 0$, we can apply Lemma 8.5 with $Y_n := S_n$ and $Z_n := \sum_{k=1}^{n}(X_{t_k} - X_{t_{k-1}})^2$, if we can show that $\mathbb{E}(\sum_{k=1}^{n}(X_{k+1} - X_{t_k})^2)$ can be bounded independent of n. This is done in Exercise 8.6 by direct use of the representation (20) of X.
Step IV: We now consider the term in (25), which we need to split up, using

the representation of X

$$
\begin{aligned}
&= \frac{1}{2}\sum_{k=1}^{n} \partial^2_{xx} h(X_{t_k})(X_{t_k} - X_{t_{k-1}})^2 \\
&= \frac{1}{2}\sum_{k=1}^{n} \partial^2_{xx} h(X_{t_k}) \left(\int_{t_k}^{t_{k+1}} g_s \, \mathrm{d}B_s \right)^2 && (29) \\
&\quad + \frac{1}{2}\sum_{k=1}^{n} \partial^2_{xx} h(X_{t_k}) \left(\int_{t_k}^{t_{k+1}} f_s \, \mathrm{d}s \right)^2 && (30) \\
&\quad + \sum_{k=1}^{n} \partial^2_{xx} h(X_{t_k}) \int_{t_k}^{t_{k+1}} f_s \, \mathrm{d}s \int_{t_k}^{t_{k+1}} g_s \, \mathrm{d}B_s. && (31)
\end{aligned}
$$

We first show that the terms (30,31) go to zero in probability if $|\Pi| \to 0$. To this end, we note that $|\partial^2_{xx} h(.)| \leq C$ by assumption. Further, $|\int_{t_k}^{t_{k+1}} f_s \, \mathrm{d}s| \leq \int_{t_k}^{t_{k+1}} |f_s| \, \mathrm{d}s$. Finally, since the function $t \to \int_0^t g_s \, \mathrm{d}B_s$ is uniformly continuous on $[0, T]$, we can conclude that $S_n := \max_{k \leq n} |\int_{t_k}^{t_{k+1}} g_s \, \mathrm{d}B_s| \to 0$ a.s. if $|\Pi| \to 0$. We therefore find for the third term (31) that

$$
\begin{aligned}
&|\sum_{k=1}^{n} \partial^2_{xx} h(X_{t_k}) \int_{t_k}^{t_{k+1}} f_s \, \mathrm{d}s \int_{t_k}^{t_{k+1}} g_s \, \mathrm{d}B_s| \\
&\leq C S_n \sum_{k=1}^{n} \int_{t_k}^{t_{k+1}} |f_s| \, \mathrm{d}s \leq C S_n \int_0^T |f_s| \, \mathrm{d}s,
\end{aligned}
$$

and we can now apply Lemma 8.5 with $Y_n := S_n$ and $Z_n := \int_0^T |f_s| \, \mathrm{d}s$. The same argument works for the term (30) since the function $t \to \int_0^t f_s \, \mathrm{d}s$, too, is uniformly continuous on $[0, T]$.
Step V: In this step, which is the most difficult part, we show that (roughly speaking) for $|\Pi|$ very small we have $\left(\int_{t_k}^{t_{k+1}} g_s \, \mathrm{d}B_s\right)^2 \cong \int_{t_k}^{t_{k+1}} g_s^2 \, \mathrm{d}s$ in the sum (29). Thus we need to compare the sum (29) with

$$
\frac{1}{2}\sum_{k=1}^{n} \partial^2_{xx} h(X_{t_k}) \int_{t_k}^{t_{k+1}} g_s^2 \, \mathrm{d}s. \tag{32}
$$

But first note that the sum in display (32) converges to $\frac{1}{2}\int_0^T \partial^2_{xx} h(X_s) g_s^2 \, \mathrm{d}s$ almost surely if $|\Pi| \to 0$, which is the desired term. To show that the difference between the sums in displays (32) and (29) converges to zero in probability, we introduce the abbreviations

$$
\begin{aligned}
m_t &:= \int_0^t g_s^2 \,\mathrm{d}s, \quad M_t := \int_0^t g_s \,\mathrm{d}B_s, \\
\delta_k &:= \int_{t_k}^{t_{k+1}} g_s^2 \,\mathrm{d}s, \quad \Delta_k := \int_{t_k}^{t_{k+1}} g_s \,\mathrm{d}B_s.
\end{aligned}
\tag{33}
$$

With this, we square the difference between the sums in (32) and (29), take the expectation, and find

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{k=1}^{n} \partial_{xx}^2 h(X_{t_k}) \left(\int_{t_k}^{t_{k+1}} g_s \,\mathrm{d}B_s\right)^2 - \sum_{k=1}^{n} \partial_{xx}^2 h(X_{t_k}) \int_{t_k}^{t_{k+1}} g_s^2 \,\mathrm{d}s\right)^2 \\
&= \mathbb{E}\left(\sum_{k=1}^{n} \partial_{xx}^2 h(X_{t_k})(\Delta_k^2 - \delta_k)\right)^2 \\
&= \sum_{k=1}^{n} \mathbb{E}\left((\partial_{xx}^2 h(X_{t_k}))^2 (\Delta_k^2 - \delta_k)^2\right) \\
&\leq 2C\left(\sum_{k=1}^{n} \mathbb{E}(\Delta_k^4) + \sum_{k=1}^{n} \mathbb{E}(\delta_k^2)\right).
\end{aligned}
\tag{34}
$$

The second equality follows as in Equation (24) because the terms under the sum are uncorrelated (see also Exercise 8.7; the important point here is that the discrete–time stochastic process $\{\sum_{k=1}^{n} \partial_{xx}^2 h(X_{t_k})(\Delta_k^2 - \delta_k), n \in \mathbb{N}\}$ is a Martingale). For the remainder of step V, we assume that $|m_t| \leq L, |M_t| \leq L$ for all $t \in [0, T]$ and all $\omega \in \Omega$. (This restriction will be removed in step VI.) We may now estimate $\sum_{k=1}^{n} \delta_k^2 \leq \max_k \delta_k \cdot m_T$ which goes to zero almost surely because the process $t \to m_t$ is uniformly continuous. By bounded convergence, it also goes to zero in expectation. For the first term on the last line of display (34) we have (using the Cauchy–Schwartz inequality)

$$
\begin{aligned}
\sum_{k=1}^{n} \mathbb{E}\Delta_k^4 &\leq \mathbb{E}(\max_k \Delta_k^2 \cdot \sum_{k=1}^{n} \Delta_k^2) \\
&\leq \sqrt{\mathbb{E}(\max_k \Delta_k^4)} \sqrt{\mathbb{E}(\sum_{k=1}^{n} \Delta_k^2)^2}
\end{aligned}
\tag{35}
$$

Since $\max_k \Delta_k^4$ goes to zero almost surely (because the process $t \to M_t$ is uniformly continuous) it also goes to zero in expectation by bounded convergence, and the remaining task is to demonstrate that $\mathbb{E}(\sum_{k=1}^{n} \Delta_k^2)^2$ remains bounded as $|\Pi| \to 0$.

$$\begin{aligned}\mathbb{E}(\sum_{k=1}^{n} \Delta_k^2)^2 &= \sum_{k=1}^{n} \sum_{l=1}^{n} \mathbb{E}\Delta_k^2 \Delta_l^2 \\ &= \sum_{k=1}^{n} \mathbb{E}\Delta_k^4 + 2\sum_{k=1}^{n} \sum_{l=k+1}^{n} \mathbb{E}\Delta_k^2 \Delta_l^2.\end{aligned}$$

Note that we have encountered the first term on the right hand side already in Equation (35) but now we merely need to show that this term is bounded. But this follows from the first inequality in (35) as $\max_k \Delta_k^2 \leq L^2$ and $\mathbb{E}(\sum_{k=1}^{n} \Delta_k^2) = \int_0^T \mathbb{E}(g_s^2)\, \mathrm{d}s$ (this follows from the definition of the Itô integral, see Lemma 7.1). For the mixed term, we note that if $l > k$ then $\mathbb{E}(\Delta_k^2 M_{t_{l+1}} M_{t_l}) = \mathbb{E}(\Delta_k^2 M_{t_l}^2)$ by the martingale property. Hence

$$\begin{aligned}\mathbb{E}(\Delta_k^2 \Delta_l^2) &= \mathbb{E}(\Delta_k^2 (M_{t_{l+1}}^2 + M_{t_l}^2 - 2M_{t_{l+1}} M_{t_l})) \\ &= \mathbb{E}(\Delta_k^2 (M_{t_{l+1}}^2 - M_{t_l}^2)),\end{aligned}$$

and therefore

$$\begin{aligned}\sum_{k=1}^{n} \sum_{l=k+1}^{n} \mathbb{E}(\Delta_k^2 \Delta_l^2) &= \sum_{k=1}^{n} \mathbb{E}(\Delta_k^2 (M_n^2 - M_k^2)) \\ &\leq 2L^2 \sum_{k=1}^{n} \mathbb{E}(\Delta_k^2) \\ &= 2L^2 \int_0^T \mathbb{E}(g_s^2)\, \mathrm{d}s.\end{aligned}$$

Step VI: We now remove the condition that m and M are bounded and define

$$K_n := \sum_{k=1}^{n} \partial_{xx}^2 h(X_{t_k})(\Delta_k^2 - \delta_k), \tag{36}$$

This is the quantity investigated in Step V, see Equation (33), but for M, m bounded. In order to apply the results from Step V, we introduce *stopping times*

$$T_L := \inf\{t \in [0,T], |M_t| \geq L \text{ or } |m_t| \geq L\},$$

and let $K_n^{(L)}$ be as in Equation (36), but with M, m replaced with $\{M_{t\wedge T_L}, t \in [0,T]\}$ and $\{m_{t\wedge T_L}, t \in [0,T]\}$, respectively. Then Step V can be applied to $K_n^{(L)}$ since $M_{.\wedge T_L}$ is now a bounded Martingale, with $m_{.\wedge T_L}$ also bounded (see Exercises 8.8,8.9 for more details on this step).

Now

$$\begin{aligned}\mathbb{P}(|K_n| \geq c) &= \mathbb{P}(\{|K_n| \geq c\} \cap \{T_L \geq T\}) + \mathbb{P}(\{|K_n| \geq c\} \cap \{T_L < T\}) \\ &= \mathbb{P}(\{|K_n^{(L)}| \geq c\} \cap \{T_L \geq T\}) + \mathbb{P}(\{|K_n| \geq c\} \cap \{T_L < T\}) \\ &\leq \mathbb{P}(|K_n^{(L)}| \geq c) + \mathbb{P}(T_L < T),\end{aligned}$$

where we use that $K_n^{(L)} = K_n$ if m, M do not reach their bounds and the stopping never happens. But applying Step V to $K_n^{(L)}$, the first term goes to zero with $|\Pi| \to 0$ while the second term goes to zero if $L \to \infty$ since M, m have bounded paths with probability one. □

Exercises for Section 8

Exercise 8.1. Prove Lemma 8.2. It may help to prove the lemma for the absolutely continuous part and the martingale part separately, and by taking Y, Z to be simple processes first. Also note the general fact that if a sequence $\{f_n\}$ of random variables is measurable with respect to some sigma algebra $\mathcal{G}$ and converges in L_2 to some f, then also f is measurable with respect to $\mathcal{G}$.

Exercise 8.2. Show the following **Lemma:** If $\{X\}$ is an Itô process of the form (18) and $\{Y^{(n)}, n \in \mathbb{N}\}$ is a sequence of processes in $\mathcal{V}([0,T])$ so that $Y^{(n)}$ converges to Y and also $Y^{(n)}g$ converges to Yg in the $\|\!|.\|\!|$ norm, then $\int_0^T Y_t^{(n)} \,\mathrm{d}X_t$ converges to $\int_0^T Y_t \,\mathrm{d}X_t$ in L_2.

Note:

Exercises 8.3,8.4 discuss some applications of the Itô formula, while Exercises 8.5–8.9 fill in bits of the proof of the Itô formula.

Exercise 8.3 (A stochastic differential equation)**.** Let $\{B_t, t \geq 0\}$ be Brownian motion on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Find the Itô process $\{X_t, t \geq 0\}$ that solves the stochastic differential equation

$$\mathrm{d}X_t = (-B_t^2 + 2t^2 B_t^2 - t)X_t \mathrm{d}t - 2tB_t X_t \mathrm{d}B_t$$

with initial condition $X_0 = 1$. (Hint: Apply Ito's theorem to $X_t = \exp(\phi(t, B_t))$ and try to find ϕ.)

Exercise 8.4. Let $\{B_t, t \geq 0\}$ be Brownian motion on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

1. Find a function ϕ so that the process $t \to X_t := \exp(B_t - \phi(t))$ can be written as an Itô–process *without* absolutely continuous part (i.e. up to a constant, $\{X\}$ is a stochastic integral).

2. Use the first item to prove that $\mathbb{E}(X_t) = \exp(-\phi(0))$ for all $t \geq 0$.

Exercise 8.5. To complete Step II of the proof of the Itô formula, show that the approximating process $t \to \sum_{k=1}^{n} \partial_x h(X_k) \mathbb{1}_{[t_k, t_{k+1})}(t)$ converges in the $\|.\|$ norm to the process $t \to \partial_x h(X_t)$. (Hint: because X is continuous in t and $\partial_x h$ is continuous, this holds pointwise for all t, ω. Now use that $\partial_x h$ is bounded.)

Exercise 8.6. In Step III, show that $\mathbb{E}(\sum_{k=1}^{n}(X_{k+1} - X_k)^2)$ can be bounded independent of n. Hint: use the representation (20) of X.

Exercise 8.7 (More on martingales)**.** In this exercise, we find the reason why the cross terms in Step V of the proof of Itô's theorem cancel.

1. Let $\{M_t, t \geq 0\}$ be a martingale with respect to some filtration $\{\mathcal{F}_t, t \geq 0\}$, and suppose that $\mathbb{E}(M_t^2) < \infty$ for all $t \geq 0$. Show that

$$\mathbb{E}\left(\left[\sum_{k=0}^{n-1} M_{t_{k+1}} - M_{t_k}\right]^2\right) = \sum_{k=0}^{n-1} \mathbb{E}\left[M_{t_{k+1}} - M_{t_k}\right]^2$$

for any partition $0 = t_0 \leq \ldots \leq t_n \in \mathbb{R}$.

2. Let $\{B_t, t \geq 0\}$ Brownian motion, $\{\mathcal{F}_t, t \geq 0\}$ the filtration generated by Brownian Motion, and $g \in \mathcal{V}([0,T])$. Put $M_t = \int_0^t g_s \mathrm{d}B_s$ and show that $V_t = M_t^2 - \int_0^t g_s^2 \mathrm{d}s$ is a martingale w.r.t. $\{\mathcal{F}_t, t \geq 0\}$. (Hint: Prove this first for $g \in \mathcal{V}([0,T])$ simple.)

Exercise 8.8 (Stopped martingales)**.** In this exercise, we learn how a martingale can be stopped to give a bounded martingale. We will consider the discrete time case first. You might want to read a little bit on "Optional Stopping" of martingales (e.g. Breiman).

Let $\{M_n, n \in \mathbb{N}\}$ be a martingale in discrete time, with respect to some filtration $\{\mathcal{F}_n, n \in \mathbb{N}\}$. That is $\mathbb{E}(|M_n|) < \infty$, M_n is $\mathcal{F}_n$ measurable for all $n \in \mathbb{N}$, and $\mathbb{E}(M_n|\mathcal{F}_k) = M_k$ if $k \leq n$. Further, a *stopping time* is a random variable $T : \Omega \to \mathbb{N}$ so that the event $\{T \leq k\} \in \mathcal{F}_k$ for all $k \in \mathbb{N}$.

1. Show that the event $\{T > k\} \in \mathcal{F}_k$ for all $k \in \mathbb{N}$.

2. Prove that $M_{T \wedge n}$ is $\mathcal{F}_n$-measurable for all $k \in \mathbb{N}$. (Hint: Show that every random variable Y can be written as $\sum_{k=1}^{n} Y \mathbb{1}_{\{T=k\}} + Y \mathbb{1}_{\{T>n\}}$, apply this with $Y = M_{T \wedge n}$ and use the previous item).

3. Prove that $\mathbb{E}(M_{T\wedge n}|\mathcal{F}_k) = M_{T\wedge k}$ for all $k \leq n$. (Hint: Use the representation in item 2 with $Y = M_{T\wedge n} - M_{T\wedge k}$).

4. Prove that $\mathbb{E}(|M_n| \cdot \mathbb{1}_{\{|M_{T\wedge n}|>x\}}) \geq \mathbb{E}(|M_{T\wedge n}| \cdot \mathbb{1}_{\{|M_{T\wedge n}|>x\}})$ for all $x \geq 0$. (Hint: Use the representation in item 2 with $Y = (|M_n| - |M_{T\wedge n}|) \cdot \mathbb{1}_{\{|M_{T\wedge n}|>x\}}$).

Exercise 8.9 (Stopped martingales in continuous time)**.** In this exercise, we show that the results of the previous exercise hold in continuous time. Let $\{M_t, t \geq 0\}$ be a *continuous* martingale w.r.t. some filtration $\{\mathcal{F}_t, t \geq 0\}$. Now, a *stopping time* is a random variable $T : \Omega \to \mathbb{R}_{\geq 0}$ so that the event $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. You can use without proof that the mapping $\omega \to M_{T(\omega)\wedge t}$ is measurable with respect to $\mathcal{A}$.

1. If T is a stopping time that assumes only countably many values, prove for all $t \geq s \geq 0$ that
$$\mathbb{E}(M_{T\wedge t}|\mathcal{F}_s) = M_{T\wedge s}$$
and for all $t \geq 0$ and $x \geq 0$ that
$$\mathbb{E}(|M_t| \cdot \mathbb{1}_{\{|M_{T\wedge t}|>x\}}) \geq \mathbb{E}(|M_{T\wedge t}| \cdot \mathbb{1}_{\{|M_{T\wedge t}|>x\}})$$
(Hint: Use the results of the previous exercise).

2. If T is a stopping time, show that for each $n \in \mathbb{N}$,
$$S^{(n)} = \frac{k}{2^n} \qquad \text{if} \qquad \frac{k-1}{2^n} < T \leq \frac{k}{2^n}$$
is a stopping time assuming only countably many values.

3. If T is a stopping time and $S^{(n)}$ defined as in the previous item, show that $M_{S^{(n)}\wedge t} \to M_{T\wedge t}$ almost surely if $n \to \infty$. (Hint: Use continuity.)

4. Conclude from the other items that if T is a stopping time, then $\{M_{T\wedge t}, t \geq 0\}$ is a martingale w.r.t. $\{\mathcal{F}_t, t \geq 0\}$, that is for all $0 \leq s \leq t$
$$\mathbb{E}(M_{T\wedge t}|\mathcal{F}_s) = M_{T\wedge s}$$
(Hint: The idea is to apply the first relation in item 1 to the stopping times $S^{(n)}$ and take the limit. But item 3 only gives you a.s. convergence. You need the second relation in item 1 to conclude that the random variables $M_{S^{(n)}\wedge t}$ are uniformly integrable; check the references on probability theory, e.g. Breiman, for this concept.)

References

[1] Leo Breiman. *Probability*. Addison-Wesley, Reading, Mass, 1973.

[2] L.C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 1998.

[3] P.R. Halmos. *Measure theory*. Number 18 in Graduate Texts in Mathematics. Springer, 1974.

[4] J. Jacod and P.E. Protter. *Probability Essentials*. Hochschultext / Universitext. Springer, 2000.

[5] Bernt Öksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, fifth edition, 1998. An introduction with applications.