Continuous martingales and stochastic calculus

February 4, 2025

Contents

These notes are based heavily on notes by Jan Obłój, Alison Etheridge and Sam Cohen from previous years' courses, and the book by Jean-François Le Gall, *Brownian motion, martingales, and stochastic calculus*, Springer 2016. The first five chapters of that book cover everything in the course (and more). Other useful references (in no particular order) include:

1. I. Karatzas and S. Shreve, Brownian motion and stochastic calculus, Springer (2nd ed.), 1991, Chapters 1-3.

2. D. Revuz and M. Yor, Continuous martingales and Brownian motion, Springer (Revised 3rd ed.), 2001, Chapters 0-4.

3. R. Durrett, Stochastic Calculus: A practical introduction, CRC Press, 1996. Sections 1.1 - 2.10.

4. F. Klebaner, Introduction to Stochastic Calculus with Applications, 3rd edition, Imperial College Press, 2012. Chapters 1, 2, 3.1–3.11, 4.1-4.5, 7.1-7.8, 8.1-8.7.

5. J. M. Steele, Stochastic Calculus and Financial Applications, Springer, 2010. Chapters 3 - 8.

6. B. Oksendal, Stochastic Differential Equations: An introduction with applications, 6th edition, Springer (Universitext), 2007. Chapters 1 - 3.

7. S. Shreve, Stochastic calculus for finance, Vol 2: Continuous-time models, Springer Finance, Springer-Verlag, New York, 2004. Chapters 3 - 4.

8. S.N. Cohen and R.J. Elliott, Stochastic Calculus and Applications, Birkhäuser, 2015, Chapters 1-5, 8-12

9. L.C.G. Rogers and D. Williams, Diffusions, Markov Processes and Martingales, Cambridge, 2000, Vol. 2, Chapters 1-6

To revise material from B8.1, you might want to look at

1. D. Williams, Probability with Martingales, Cambridge, 1991

2. M. Capińksi and E. Kopp, Measure, Integral and Probability, Springer, 1999

The appendices gather together some useful results that we take as known, or are too lazy to prove in the main text

1 Introduction

Our topic is part of the huge field devoted to the study of *stochastic processes*. Since first year, you have had the notion of a *random variable*. In this course, we want to think of random processes, that is random variables that evolve in time.

When we model deterministic quantitities that evolve with (continuous) time, such as particles moving under gravity or some other force, we often appeal to ordinary differential equations as models. In this course we develop the 'calculus' necessary to develop an analogous theory of *stochastic* (ordinary) *differential equations*. These can be used to model quantities such as the prices of assets or particles moving in turbulent fluids, where the randomness is a key part of the evolution.

An ordinary differential equation might take the form

$$\frac{dX(t)}{dt} = a(t, X(t)),$$

for a suitably nice function a. Thinking of this as an infinitesimal evolution equation we could write

$$dX(t) = a(t, X(t))dt,$$

to describe how the quantity X changes in an infinitesimal amount of time dt. A stochastic differential equation is often formally written as

$$dX(t) = a(t, X(t))dt + b(t, X(t))dB_t,$$

where the second term on the right models 'noise' or fluctuations, through a function b and a noise dB which we can think of as the infinitesimal change in a random process B, that is an 'infinitesimal normal distribution' with mean 0 and variance dt. In order to try to make sense of this we can write it equivalently in an integral form:

$$X(t) = X(0) + \int_0^t a(s, X(s))ds + \int_0^t b(s, X(s))dB_s.$$

Here the random process $(B_t)_{t\geq 0}$ is a fundamental object that we call Brownian motion. The punchline of the course is that we can give a rigorous meaning to the last term in the equation and although we will consider what appear to be more general driving noises, under rather general conditions they can all be built from Brownian motion. Indeed if we added possible (random) 'jumps' in $X(t)$, we would capture essentially the most general theory. We are not going to allow jumps, so we will be thinking of settings in which our stochastic equation has a continuous solution $t \mapsto X_t$.

2 Random processes

2.1 Measure theory

Formally, we can think of this in two parts:

- a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, describing *states of the world* $\omega \in \Omega$, *events* $A \in \mathscr{F}$ and their *probabilities* $\mathbb{P}$, and
- mappings $X : (\Omega, \mathscr{F}) \to (E, \mathscr{E})$, giving the value of the random variable in each state of the world.

We need to assume $X : \Omega \to E$ is a *measurable* mapping, so that for each $U \in \mathscr{E}$, $X^{-1}(U) \in \mathscr{F}$ and so, in particular, we can assign a probability to the event that $X \in U$. Often $(E, \mathscr{E})$ is just $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ (where $\mathscr{B}(\mathbb{R})$ is the Borel sets on $\mathbb{R}$) and this just says that for each $x \in \mathbb{R}$ we can assign a probability to the event $\{X \leq x\}$.

Definition 2.1. *A* stochastic process*, indexed by some set* $\mathscr{T}$*, is a collection of random variables* $\{X_t\}_{t\in\mathscr{T}}$*, defined on a common probability space* $(\Omega, \mathscr{F}, \mathbb{P})$ *and taking values in a common* state space $(E, \mathscr{E})$.

For us, $\mathscr{T}$ will generally be either $[0,\infty)$ or $[0,T]$ and we think of X_t as a random quantity that evolves with time.

Definition 2.2. *A collection* $\{\mathscr{F}_t, t \in [0,\infty)\}$ *of* σ*-algebras of sets in* $\mathscr{F}$ *is a* filtration *if* $\mathscr{F}_t \subseteq \mathscr{F}_{t+s}$ *for* $t,s \in [0,\infty)$*. (Intuitively,* $\mathscr{F}_t$ *corresponds to the information known to an observer at time* t*.)*

In particular, for a process X *we define* $\mathscr{F}_t^X = \sigma(\{X(s) : s \leq t\})$ *(that is* $\mathscr{F}_t^X$ *is the information obtained by observing* X *up to time* t*) to be the* natural filtration *associated with the process* X*.*

2.2 Filtrations and processes

We now wish to extend our thinking from random variables to random processes. We will work with real-valued processes, i.e. those in $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. This requires a bit of measure theoretic care to make sure everything is well defined. Recall that $\mathscr{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$, that is, the smallest σ-algebra containing all open sets/such that all continuous functions are measurable.

Definition 2.3. *The mapping* $t \mapsto X_t(\omega)$ *for a fixed* $\omega \in \Omega$*, represents a realisation of our stochastic process, called a* sample path *or* trajectory*. We shall assume that*

$$(t,\omega) \mapsto X_t(\omega) : \big([0,\infty) \times \Omega, \mathscr{B}([0,\infty)) \otimes \mathscr{F}\big) \mapsto \big(\mathbb{R}, \mathscr{B}(\mathbb{R})\big)$$

is measurable (i.e. $\forall A \in \mathscr{B}(\mathbb{R})$*,* $\{(t,\omega) : X_t \in A\}$ *is in the product* σ*-algebra* $\mathscr{B}([0,\infty)) \otimes \mathscr{F}$*). Our stochastic process is then said to be* measurable.

In discrete time, we often made statements which held 'almost surely', that is, up to a set of measure zero. In continuous time, we need to be more careful with what this means:

Definition 2.4. *Let X, Y be two stochastic processes defined on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$.*

i. We say that X is a modification *of Y if, for all $t \geq 0$, we have $X_t = Y_t$ a.s.;*

ii. We say that X and Y are indistinguishable *if*

$$\mathbb{P}\big[X_t = Y_t \text{ for all } 0 \leq t < \infty\big] = 1,$$

or equivalently, $X_t(\omega) = Y_t(\omega)$ for all t, for all $\omega \notin \mathscr{N}$ where $\mathbb{P}(\mathscr{N}) = 0$.

Example 2.5. *Let $T \sim U([0,1])$ be a uniform random variable, and take the random process $X_t = \mathbf{1}_{t=T}$. Then $Y_t := 0$ is a modification of X_t, as $Y_t = X_t$ a.s. for each t. However, Y and X are not indistinguishable, as $X \neq Y$ for* some *t with positive probability (in fact, with probability 1).*

If X and Y are modifications of one another then, in particular, they have the same finite dimensional distributions,

$$\mathbb{P}\left[(X_{t_1}, \ldots, X_{t_n}) \in A\right] = \mathbb{P}\left[(Y_{t_1}, \ldots, Y_{t_n}) \in A\right]$$

for any finite collection of times $\{t_1, t_2, \ldots, t_n\}$ for any n, and all measurable sets A, but indistinguishability is a much stronger property.

2.3 Constructing distributions on $\big(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)})\big)$

Indistinguishability takes the *sample path* as the basic object of study, so that we could think of $(X_t(\omega), t \geq 0)$ (the path) as a random variable taking values in the space $E^{[0,\infty)}$ (of all possible paths). This state space then has to be endowed with a σ-algebra of measurable sets. For definiteness, we take real-valued processes, so $E = \mathbb{R}$.

Definition 2.6. *An n-dimensional* cylinder set *in $\mathbb{R}^{[0,\infty)}$ is a set of the form*

$$C = \left\{\omega \in \mathbb{R}^{[0,\infty)} : \big(\omega(t_1), \ldots, \omega(t_n)\big) \in A\right\}$$

for some $0 \leq t_1 < t_2 \ldots < t_n$ and $A \in \mathscr{B}(\mathbb{R}^n)$.

Let $\mathscr{C}$ be the family of all finite-dimensional cylinder sets and $\mathscr{B}(\mathbb{R}^{[0,\infty)})$ the σ-algebra it generates. This is small enough to be able to build probability measures on $\mathscr{B}(\mathbb{R}^{[0,\infty)})$ using Carathéodory's Theorem (see B8.1). On the other hand $\mathscr{B}(\mathbb{R}^{[0,\infty)})$ only contains events which can be defined using at most countably many coordinates. In particular, the set

$$\left\{\omega \in \mathbb{R}^{[0,\infty)} : \omega(t) \text{ is continuous}\right\}$$

is *not* $\mathscr{B}(\mathbb{R}^{[0,\infty)})$-measurable.

We will have to do some work to show that many processes can be assumed to be continuous, or right continuous. The sample paths are then fully described by their values at times $t \in \mathbb{Q}$, which will greatly simplify the study of quantities of interest such as $\sup_{0\le s\le t}|X_s|$ or $\tau_0(\omega) = \inf\{t \ge 0 : X_t(\omega) > 0\}$.

A monotone class argument (see Appendix A.1) will tell us that a probability measure on $\mathscr{B}(\mathbb{R}^{[0,\infty)})$ is characterised by its finite-dimensional distributions – so if we can take continuous paths, then we only need to find the probabilities of cylinder sets to characterise the distribution of the process.

In this section, we are going to provide a very general result about constructing continuous time stochastic processes and a criterion due to Kolmogorov which gives conditions under which there will be a version of the process with continuous paths.

Let $\mathbb{T}$ be the set of finite increasing sequences of non-negative numbers, i.e. $\mathbf{t} \in \mathbb{T}$ if and only if $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ for some n and $0 \le t_1 < t_2 < \ldots, < t_n$.

Suppose that for each $\mathbf{t} \in \mathbb{T}$ of length n we have a probability measure $\mathbb{P}_{\mathbf{t}}$ on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. The collection $(\mathbb{P}_{\mathbf{t}} : \mathbf{t} \in \mathbb{T})$ is called a family of finite-dimensional (marginal) distributions.

Definition 2.7. *A family of finite dimensional distributions is called* consistent *if for any* $\mathbf{t} = (t_1, t_2, \ldots, t_n) \in \mathbb{T}$ *and* $1 \le j \le n$

$$\begin{aligned}\mathbb{P}_{\mathbf{t}}(A_1 \times A_2 \times \ldots \times A_{j-1} \times \mathbb{R} \times A_{j+1} \times \ldots \times A_n) \\ = \mathbb{P}_{\mathbf{s}}(A_1 \times A_2 \times \ldots \times A_{j-1} \times A_{j+1} \times \ldots \times A_n)\end{aligned}$$

where $A_i \in \mathscr{B}(\mathbb{R})$ *and* $\mathbf{s} := (t_1, t_2, \ldots, t_{j-1}, t_{j+1}, \ldots, t_n)$.

(In other words, if we integrate out over the distribution at the jth time point then we recover the corresponding marginal for the remaining lower dimensional vector.)

If we have a probability measure $\mathbb{P}$ on $\left(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)})\right)$ then it defines a consistent family of marginals via

$$\mathbb{P}_{\mathbf{t}}(A) = \mathbb{P}(\{\omega \in \mathbb{R}^{[0,\infty)} : (\omega(t_1), \ldots, \omega(t_n)) \in A\})$$

where $\mathbf{t} = (t_1, t_2, \ldots, t_n)$, $A \in \mathscr{B}(\mathbb{R}^n)$, and we note that the set in question is in $\mathscr{B}(\mathbb{R}^{[0,\infty)})$ as it depends on finitely many coordinates. But we would like to have a converse – if I give you $\mathbb{P}_{\mathbf{t}}$, does there exist a corresponding measure $\mathbb{P}$?

Theorem 2.8 (Daniell–Kolmogorov Extension Theorem)**.** *Let* $\{\mathbb{P}_{\mathbf{t}} : \mathbf{t} \in \mathbb{T}\}$ *be a consistent family of finite-dimensional distributions. Then there exists a probability measure* $\mathbb{P}$ *on* $\left(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)})\right)$ *such that for any* n, $\mathbf{t} = (t_1, \ldots, t_n) \in \mathbb{T}$ *and* $A \in \mathscr{B}(\mathbb{R}^n)$,

$$\mathbb{P}_{\mathbf{t}}(A) = \mathbb{P}[\{\omega \in \mathbb{R}^{[0,\infty)} : (\omega(t_1), \ldots, \omega(t_n)) \in A\}]. \tag{1}$$

We will not prove this here (see Appendix), but notice that (1) defines $\mathbb{P}$ on the cylinder sets and so if we can establish countable additivity then the proof reduces to an application of Carathéodory's extension theorem. Uniqueness is a consequence of the Monotone Class Lemma.

This is a remarkably general result, but it doesn't allow us to say anything meaningful about the paths of the process. For that we appeal to Kolmogorov's criterion.

Theorem 2.9 (Kolmogorov–Čentsov continuity criterion)**.** *Suppose that a stochastic process $(X_t : t \leq T)$ defined on $(\Omega, \mathscr{F}, \mathbb{P})$ satisfies*

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq C|t-s|^{1+\beta}, \quad 0 \leq s, t \leq T \tag{2}$$

for some strictly positive constants α, β and C.
Then there exists $\tilde{X}$, a modification of X, whose paths are γ-locally Hölder continuous $\forall \gamma \in (0, \beta/\alpha)$ a.s., i.e.

$$\sup_{s,t,\in[0,T]} \frac{|\tilde{X}_t - \tilde{X}_s|}{|t-s|^\gamma} < \infty \quad a.s. \tag{3}$$

In particular, the sample paths of $\tilde{X}$ are a.s. continuous (and uniformly continuous on $[0,T]$).

Proof. See appendix (not examinable) □

Remark 2.10. *Many more results and conditions in this direction are possible. See for example Cramér and Leadbetter,* Stationary and Related Stochastic Processes*, Wiley, 1967.*

3 Brownian Motion

3.1 Definition

Our fundamental building block will be Brownian motion. It is often described as an 'infinitesimal random walk', so to motivate the definition, we take a quick look at simple (discrete time) random walk.

Definition 3.1. *The discrete time stochastic process $\{S_n\}_{n\geq 0}$ is a symmetric simple random walk under the measure $\mathbb{P}$ if $S_n = \sum_{i=1}^n \xi_i$, where the ξ_i can take only the values ± 1, and are i.i.d. under $\mathbb{P}$ with $\mathbb{P}[\xi_i = -1] = 1/2 = \mathbb{P}[\xi_i = 1]$.*

Lemma 3.2. *$\{S_n\}_{n\geq 0}$ is a $\mathbb{P}$-martingale (with respect to the natural filtration) and*

$$cov(S_n, S_m) = n \wedge m.$$

To obtain a 'continuous' version of simple random walk, we appeal to the Central Limit Theorem. Since $\mathbb{E}[\xi_i] = 0$ and $\mathrm{var}(\xi_i) = 1$, we have

$$\mathbb{P}\left[\frac{S_n}{\sqrt{n}} \leq x\right] \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \text{ as } n \to \infty.$$

More generally,

$$\mathbb{P}\left[\frac{S_{[nt]}}{\sqrt{n}} \leq x\right] \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi t}} e^{-y^2/2t} dy \text{ as } n \to \infty,$$

where $[nt]$ denotes the integer part of nt.

Heuristically at least, passage to the limit from simple random walk suggests the following definition of Brownian motion.

Definition 3.3 (Brownian motion). *A real-valued stochastic process $\{B_t\}_{t\geq 0}$ is a* $\mathbb{P}$-Brownian motion *(or a* $\mathbb{P}$-Wiener process*) if for some real constant σ, under $\mathbb{P}$,*

i. for each $s \geq 0$ and $t > 0$ the random variable $B_{t+s} - B_s$ has the normal distribution with mean zero and variance $\sigma^2 t$,

ii. for each $n \geq 1$ and any times $0 \leq t_0 \leq t_1 \leq \cdots \leq t_n$, the random variables $\{B_{t_r} - B_{t_{r-1}}\}$ are independent,

iii. $B_0 = 0$,

iv. B_t is continuous in $t \geq 0$.

When $\sigma^2 = 1$, we say that we have a standard *Brownian motion.*

Notice in particular that for $s < t$,

$$\Gamma(s,t) = \mathrm{cov}(B_s, B_t) = \mathbb{E}[B_s B_t] = \mathbb{E}\left[B_s^2 + B_s(B_t - B_s)\right] = \mathbb{E}[B_s^2] = s \quad (= s \wedge t).$$

We can use this to form an equivalent definition:
Brownian motion is a real-valued, mean zero, continuous Gaussian process with covariance function $s \wedge t$.

We can write down the finite dimensional distributions using the independence of increments. They admit a density with respect to Lebesgue measure. We write $p(t,x,y)$ for the *transition density*

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right).$$

This is the density (with respect to x), of B_t given $B_0 = y$. For $0 = t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$, writing $x_0 = 0$, the joint probability density function of $B_{t_1}, \ldots, B_{t_n}$ is

$$f(x_1, \ldots, x_n) = \prod_{j=1}^{n} p(t_j - t_{j-1}, x_{j-1}, x_j).$$

We could recover the existence of Brownian motion from the general principles outlined so far (Daniell–Kolmogorov Theorem and the Kolmogorov continuity criterion), but we are now going to take a short digression to describe a beautiful construction due to Lévy. In fact, it is a little easier if we generalize our definition to more than one dimension, as follows:

Definition 3.4. *Let ν be a probability measure on $\mathbb{R}^d$. A d-dimensional stochastic process $(B_t : t \geq 0)$ on $(\Omega, \mathscr{F}, \mathbb{P})$ is called a d-dimensional Brownian motion with initial distribution ν if*

i. $\mathbb{P}[B_0 \in A] = \nu(A), \quad A \in \mathscr{B}(\mathbb{R}^d)$;

ii. $\forall 0 \leq s \leq t$ *the increment* $(B_t - B_s)$ *is independent of* $\mathscr{F}_s = \sigma(B_u : u \leq s)$ *and is normally distributed with mean* 0 *and covariance matrix* $(t-s) \times \mathbf{I}_d$;

iii. *B has continuous paths.*

Writing the d-dimensional Brownian motion as $B_t = (B_t^{(1)}, \ldots, B_t^{(d)})$, if $\mu(\{0\}) = 1$ then the coordinate processes $(B_t^{(i)})$, $1 \leq i \leq d$, are independent one-dimensional Brownian motions. If $\nu(\{x\}) = 1$ for some $x \in \mathbb{R}^d$, we say that B starts at x.

3.2 Lévy's construction

The following lemma establishes some classical and useful properties of normal distributions. Its proof is left as an exercise (see Appendix B).

Lemma 3.5. *(i) Let Z, Z' be independent random variables with $Z \sim N(\mu, \Sigma)$, $Z' \sim N(\mu', \Sigma')$. Then $Z + Z' \sim N(\mu + \mu', \Sigma + \Sigma')$. Equivalently, their densities satisfy the convolution property*

$$\int_{\mathbb{R}^d} \phi_{(\mu,\Sigma)}(y) \phi_{(\mu',\Sigma')}(x-y) dy = \phi_{(\mu+\mu',\Sigma+\Sigma')}(x).$$

(ii) If $Z_i \sim N(\mu_i, \Sigma_i)$ is a sequence of independent normal random variables such that $\mu^* = \sum_{i \in \mathbb{N}} \mu_i$ and $\Sigma^* = \sum_{i \in \mathbb{N}} \Sigma_i$ exist (i.e. the sums converge), then the sequence of partial sums $\sum_{i=1}^n Z_i$ converges in (L^2, and hence in) probability to a random variable with distribution

$$\sum_{i \in \mathbb{N}} Z_i \sim N(\mu^*, \Sigma^*).$$

(iii) If the pair (Z, Z') is a multivariate normal random variable, then Z and Z' are normal, and are independent if and only if their covariance is zero, that is, $E[(Z-\mu)(Z'-\mu')^\top] = 0$.

We begin with a countable family $\{Z_m\}$ of identically distributed random variables with $Z_m \sim N(0, I_d)$ for all m. Let $D_n = \{k2^{-n} : k, n \in \mathbb{Z}^+\}$, so that $D_n \subset D_{n+1}$, $D_0 = \mathbb{Z}^+$ and $\cup_n D_n$ is the set of Dyadic rationals. For simplicity of notation, let $\{Z_m\}$ be indexed by $m \in \cup_n D_n$ and $Z_0 := 0$.

We proceed as follows: First, we determine the value of the nth approximation X^n on the points D_n. Second, we use linear interpolation to define X_t^n for all values of t. This gives us a sequence of paths which we shall show converge.

To fix the values of X_t^n for $t \in D_n$, we define

$$X_t^0 = \sum_{\{k \in D_0 : k < t\}} Z_k.$$

Next, for every $n > 0$, define $X_t^n = X_t^{n-1}$ for all $t \in D_{n-1}$. For $t \in D_n \setminus D_{n-1}$, let

$$X_t^n = X_t^{n-1} + 2^{-(n/2+1)} Z_t. \tag{4}$$

We now linearly interpolate between these points $\{X_t^n\}_{t \in D_n}$. Formally, we can write the interpolation step as

$$X_t^n = X_{\lfloor t \rfloor_n} + \frac{t - \lfloor t \rfloor_n}{\lceil t \rceil_n - \lfloor t \rfloor_n} (X_{\lceil t \rceil_n} - X_{\lfloor t \rfloor_n}),$$

where $\lfloor t \rfloor_n = \max\{s \in D_n : s \le t\}$, $\lceil t \rceil_n = \min\{s \in D_n : s \ge t\}$. The use of linear interpolation is not vital to the construction, as we shall see (taking right-continuous step functions $X_t^n := X_{\lfloor t \rfloor_n}^n$ would work just as well for proving the existence of a limit, but would not immediately give continuity). We now seek to show that these paths converge, in a sufficiently strong sense, to a Brownian motion.

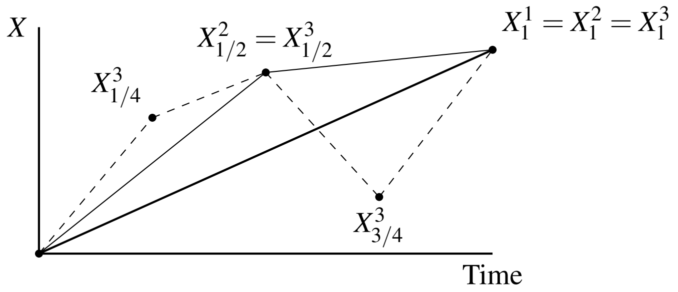

Figure 1: Three steps in Lévy's construction

Lemma 3.6. *Let $\{X^n\}_{n \in \mathbb{N}}$ be a sequence of a.s. continuous functions which converge uniformly on compacts in probability to a process X, that is, for any $\varepsilon > 0$,*

$$\lim_n P\Big(\sup_{s \in [0,t]} \|X_s^n - X_s\| < \varepsilon\Big) = 1$$

for all t. Then X is also continuous.

Proof. For fixed t, by Lemma A.7, taking a subsequence in n, we can assume that the convergence is almost sure, that is,

$$P\Big(\lim_n \big(\sup_{s\in[0,t]} \|X_s^n - X_s\|\big) = 0\Big) = 1$$

Fixing ω, this is a statement of uniform convergence of $X^{n_j} \to X$, and the continuity of the limit is classical, as for any $\varepsilon > 0$, we can find $\delta, m > 0$ such that

$$\begin{aligned}
\|X_s - X_{s+\delta}\| &\leq \|X_s^{n_m} - X_s\| + \|X_{s+\delta}^{n_m} - X_{s+\delta}\| + \|X_s^{n_m} - X_{s+\delta}^{n_m}\| \\
&\leq 2 \sup_{s\in[0,t]} \{\|X_s^{n_m} - X_s\|\} + \|X_s^{n_m} - X_{s+\delta}^{n_m}\| \\
&\leq 3\varepsilon.
\end{aligned}$$

□

Remark 3.7. *The uncountable supremum in the statement of Lemma 3.6 is measurable, as our functions are continuous (so the supremum could equally be taken over the rationals, and suprema over countable sets are always measurable).*

Theorem 3.8. *The processes X^n defined in* (4) *converge a.s. uniformly on compacts to a process X. In its natural filtration, the limit is a Brownian motion starting at zero.*

Proof. Convergence. We first show that the processes converge. We consider the case where X is a Brownian motion in two dimensions, as this implies all other cases by the triangle inequality, and is notationally simpler. From our construction, we can see that

$$\sup_{s\in[0,t]} \|X_s^n - X_s^{n+1}\| = \max_{\{s\in D_{n+1}\setminus D_n : s<t\}} \|2^{-(n/2+1)} Z_s\|.$$

The set $\{s \in D_{n+1} \setminus D_n : s < t\}$ contains at most $t2^n$ elements, and the Z_s are independent $N(0, I_d)$ random variables. It is standard that $\|Z_s\|^2$ has a χ^2-distribution with $d = 2$ degrees of freedom, so if $F(x) := P(\|Z_s\|^2 \leq x)$ is the distribution function of $\|Z_s\|^2$ we have

$$\begin{aligned}
&P\big(\sup_{s\in[0,t]} \|X_s^n - X_s^{n+1}\| > \varepsilon\big) = P\Big(\max_{\{s\in D_{n+1}\setminus D_n : s<t\}} \|Z_s\| > 2^{n/2+1}\varepsilon\Big) \\
&\leq \sum_{\substack{\{s\in D_{n+1}\setminus D_n, \\ s<t\}}} P\big(\|Z_s\| > 2^{n/2+1}\varepsilon\big) = t2^n(1 - F(2^{n+2}\varepsilon^2)).
\end{aligned}$$

By changing into polar coordinates, it is easy to show that $F(x) = 1 - e^{-x/2}$ (this simple form is the reason we chose $d = 2$). Therefore,

$$P\big(\sup_{s\in[0,t]} \|X_s^n - X_s^{n+1}\| > \varepsilon\big) \leq t2^n \exp(-2^{n+1}\varepsilon^2).$$

In particular,

$$P\big(\sup_{s\in[0,t]}\|X_s^n - X_s^{n+1}\| > n^{-3}\big) \le t2^n \exp(-2^{n+1}n^{-6}).$$

Taking N large enough that $N\log(2) - 2^{N+1}N^{-6} < -N$, for all $n > N$ we have

$$P\big(\sup_{s\in[0,t]}\|X_s^n - X_s^{n+1}\| > n^{-3}\big) \le te^{-n}.$$

By the Borel–Cantelli Lemma, as this sequence is summable we have

$$P\big(\sup_{s\in[0,t]}\|X_s^n - X_s^{n+1}\| > n^{-3} \text{ for infinitely many } n\big) = 0.$$

In particular, with probability one, taking N sufficiently large, for all $n \ge N$,

$$\sup_{s\in[0,t]}\|X_s^n - X_s^{n+1}\| \le n^{-3}$$

and by the triangle inequality, recalling that $\sum_n n^{-2} = \pi^2/6$, for $N < n < m$,

$$\sup_{s\in[0,t]}\|X_s^n - X_s^m\| \le \sum_{j=n}^{m-1}\Big(\sup_{s\in[0,t]}\|X_s^j - X_s^{j+1}\|\Big) \le \frac{\pi^2}{6n}.$$

Therefore, with probability one, the processes X^n are converging uniformly on the interval $[0,t]$. By Lemma 3.6, X is a continuous process.

X is a Brownian Motion. We now need to show that X is a Brownian motion in its natural filtration, that is, that the increment $X_t - X_s$ is normally distributed and independent of $\mathscr{F}_s = \sigma(X_u, u \le s)$. First note that for s,t with $t \in D_n \setminus D_{n+1}$ and $\lceil s\rceil_n < t$, the random variable Z_t is not involved in the construction of X_s. Hence, as X generates the filtration and the $\{Z_u\}_{u\in\cup_n D_n}$ are independent, we see that Z_t is independent of $\mathscr{F}_s$.

It is clear that if s,t are integers with $s < t$, then

$$X_t - X_s = X_t^0 - X_s^0 = \sum_{\{k\in D_0: s<k<t\}} Z_k \sim N(0,(t-s)I_d).$$

Furthermore, in this case $X_t - X_s$ is independent of $\mathscr{F}_s$, as $Z_k = Z_{\lceil k\rceil_0}$ is independent of $\mathscr{F}_s$ for all $s < k$.

Now suppose that the result holds for $s,t \in D_n$.Then we see that for any $u \in D_{n+1}\setminus D_n$,

$$X_u - X_{\lfloor u\rfloor_n} = \frac{X_{\lceil u\rceil_n} - X_{\lfloor u\rfloor_n}}{2} + 2^{-(n/2+1)}Z_u \sim N(0, 2^{-(n+1)}I_d)$$

which is independent of $\mathscr{F}_{\lfloor u\rfloor_n}$. Similarly,

$$X_{\lceil u\rceil_n} - X_u = \frac{X_{\lceil u\rceil_n} - X_{\lfloor u\rfloor_n}}{2} - 2^{-(n/2+1)}Z_u \sim N(0, 2^{-(n+1)}I_d),$$

which is independent of $\mathscr{F}_{\lfloor u \rfloor_n}$. Therefore, for any $s,t \in D_{n+1}$,

$$X_t - X_s = (X_t - X_{\lfloor t \rfloor_n}) + (X_{\lfloor t \rfloor_n} - X_{\lceil s \rceil_n}) + (X_{\lceil s \rceil_n} - X_s),$$

which is the sum of three independent normal random variables, so

$$X_t - X_s \sim N(0,(t-s)I_d).$$

The first two terms are independent of $\mathscr{F}_{\lceil s \rceil_n} \supseteq \mathscr{F}_s$. We know the last term is independent of $\mathscr{F}_{\lfloor s \rfloor_n}$, and we can compute

$$E\left[(X_{\lceil s \rceil_n} - X_s)(X_s - X_{\lfloor s \rfloor_n})^\top\right] = 0$$

so $(X_{\lceil s \rceil_n} - X_s)$ is independent of the increment $X_s - X_{\lfloor s \rfloor_n}$, as uncorrelated Gaussians are independent. As we can write

$$\mathscr{F}_s = \mathscr{F}_{\lfloor s \rfloor_n} \vee \sigma(X_s - X_{\lfloor s \rfloor_n}) \vee \sigma\big(Z_u; u \in \big]\lfloor s \rfloor_n, s\big[\big),$$

we see that $X_{\lceil s \rceil_n} - X_s$ is independent of $\mathscr{F}_s$. Therefore $X_t - X_s$ is normally distributed and independent of $\mathscr{F}_s$, as desired.

Finally, for any $s < t$ we can find sequences $s_n \downarrow s$, $t_n \uparrow t$ with $s_n, t_n \in D_n$ and $s_0 \leq t_0$. Then $X_{t_n} - X_{s_n} \sim N(0,(t_n - s_n)I_d)$, and by continuity of X we see

$$X_t - X_s = X_{t_0} - X_{s_0} + \sum_{n=1}^{\infty} \big(X_{t_n} - X_{t_{n-1}} - X_{s_n} + X_{s_{n-1}}\big) \sim N(0,(t-s)I_d).$$

All the terms in this sum are independent of $\mathscr{F}_s$, as required. As $X_0 = 0$ by construction, we see that X is a Brownian motion starting at zero, in its natural filtration. □

3.3 Wiener Measure

Let $C(\mathbb{R}_+,\mathbb{R})$ be the space of continuous functions from $[0,\infty)$ to $\mathbb{R}$. Given a Brownian motion $(B_t : t \geq 0)$ on $(\Omega, \mathscr{F}, \mathbb{P})$, consider the map

$$\Omega \to C(\mathbb{R}_+,\mathbb{R}), \quad \text{given by} \quad \omega \mapsto (B_t(\omega) : t \geq 0) \tag{5}$$

which is measurable w.r.t. $\mathscr{B}(C(\mathbb{R}_+,\mathbb{R}))$ – the smallest σ-algebra such that the coordinate mappings (i.e. $(\omega_t : t \geq 0) \mapsto \omega(t_0)$ for a fixed t_0) are measurable. (In fact $\mathscr{B}(C(\mathbb{R}_+,\mathbb{R}))$ is also the Borel σ-algebra generated by the topology of uniform convergence on compacts.)

Definition 3.9. *The* Wiener measure $\mathbf{W}$ *is the image of* $\mathbb{P}$ *under the mapping in* (5)*; it is the probability measure on the space of continuous functions such that the canonical process, i.e.* $(B_t(\omega) = \omega(t), t \geq 0)$*, is a Brownian motion.*

In other words, $\mathbf{W}$ is the unique probability measure on $(C(\mathbb{R}_+,\mathbb{R}), \mathscr{B}(C(\mathbb{R}_+,\mathbb{R})))$ such that

i. $\mathbf{W}(\{\omega \in C(\mathbb{R}_+,\mathbb{R}), \omega(0)=0\}) = 1$;

ii. for any $n \geq 1$, $\forall\, 0 = t_0 < t_1 < \ldots < t_n$, $A \in \mathscr{B}(\mathbb{R}^n)$

$$\begin{aligned}&\mathbf{W}(\{\omega \in C(\mathbb{R}_+,\mathbb{R}) : (\omega(t_1),\ldots,\omega(t_n)) \in A\})\\ &= \int_A \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{dy_1 \cdots dy_n}{\sqrt{t_1(t_2-t_1)\ldots(t_n-t_{n-1})}} \exp\Big(-\sum_{i=1}^n \frac{(y_i-y_{i-1})^2}{2(t_i-t_{i-1})}\Big),\end{aligned}$$

where $y_0 := 0$.

(Uniqueness follows from the Monotone Class Lemma, since $\mathscr{B}(C(\mathbb{R}_+,\mathbb{R})))$ is generated by finite dimensional projections.)

3.4 Properties of Brownian motion

Although the sample paths of Brownian motion are continuous, it does not mean that they are nice in any other sense. In fact the behaviour of Brownian motion is unlike the usual functions one encounters. Here are just a few of its strange behavioural traits.

i. Although $\{B_t\}_{t\geq 0}$ is continuous everywhere, it is (with probability one) differentiable nowhere.

ii. Brownian motion will eventually hit any and every real value no matter how large, or how negative. No matter how far above the axis, it will (with probability one) be back down to zero at some later time.

iii. Once Brownian motion hits a value, it immediately hits it again (uncountably!) *infinitely* often, and then again from time to time in the future.

iv. It doesn't matter what scale you examine Brownian motion on, it looks just the same. The paths of Brownian motion are fractals almost surely.

The last property is really a consequence of the construction of the process. We will formulate the second and third more carefully later.

Proposition 3.10. *Let B be a standard real-valued Brownian motion. Then*

i. $-B_t$ *is also a Brownian motion,* *(symmetry)*

ii. $\forall c \geq 0$, cB_{t/c^2} *is a Brownian motion,* *(scaling)*

iii. $X_0 = 0$, $X_t := tB_{\frac{1}{t}}$ *is a Brownian motion,* *(time inversion)*

iv. for $t \in [0,1]$, $X_t := B_1 - B_{1-t}$ *is a Brownian motion,* *(time reversal)*

v. $\forall s \geq 0$, $\tilde{B}_t = B_{t+s} - B_s$ *is a Brownian motion independent of* $\sigma(B_u : u \leq s)$, *(simple Markov property).*

The proof is an exercise.

3.5 Fine continuity of Brownian sample paths

From now on, when we say "Brownian motion", we mean a standard real-valued Brownian motion.

We know that $t \mapsto B_t(\omega)$ is almost surely continuous.

Exercise: Use the Kolmogorov continuity criterion to show that Brownian motion admits a modification which is locally Hölder continuous of order γ for any $0 < \gamma < 1/2$.

On the other hand, as we have already remarked, the path is actually rather 'rough'. We would like to have a way to quantify this roughness.

Definition 3.11. *Let π be a partition of $[0,T]$, $N(\pi)$ the number of intervals that make up π and $\|\pi\|$ be the* mesh *of π (that is the length of the longest interval in the partition). Write $0 = t_0 < t_1 < \ldots < t_{N(\pi)} = T$ for the endpoints of the intervals of the partition. Then the* variation *of a function $f : [0,T] \to \mathbb{R}$ is*

$$\lim_{\delta \to 0} \left\{ \sup_{\pi : \|\pi\| = \delta} \sum_{j=1}^{N(\pi)} \left| f(t_j) - f(t_{j-1}) \right| \right\}.$$

If the function is 'nice', for example differentiable, then it has bounded variation. Our 'rough' paths will have *unbounded* variation. To quantify roughness we can extend the idea of variation to that of p-variation.

Definition 3.12. *In the notation of Definition 3.11, the p-variation of a function $f : [0,T] \to \mathbb{R}$ is defined as*

$$\lim_{\delta \to 0} \left\{ \sup_{\pi : \|\pi\| = \delta} \sum_{j=1}^{N(\pi)} \left| f(t_j) - f(t_{j-1}) \right|^p \right\}.$$

Notice that for $p > 1$, the p-variation will be finite for functions that are much rougher than those for which the variation is bounded. For example, roughly speaking, finite 2-variation will follow if the fluctuation of the function over an interval of order δ is order $\sqrt{\delta}$.

For a typical Brownian path, the 2-variation will be infinite. However, a slightly weaker analogue of the 2-variation *does* exist.

Theorem 3.13. *Let B_t denote Brownian motion under $\mathbb{P}$ and for a partition π of $[0,T]$ define*

$$S(\pi) = \sum_{j=1}^{N(\pi)} \left| B_{t_j} - B_{t_{j-1}} \right|^2.$$

Let π_n be a sequence of partitions with $\|\pi_n\| \to 0$. Then

$$\mathbb{E}\left[|S(\pi_n) - T|^2 \right] \to 0 \qquad \text{as } n \to \infty. \tag{6}$$

We say that the quadratic variation process of Brownian motion, which we denote by $\{\langle B\rangle_t\}_{t\geq 0}$ is $\langle B\rangle_t = t$.

Remark: In fact for *any* continuous martingale, the limit process $\langle M\rangle$ defined through a similar sequence of partitions exists and is independent of the sequence of partitions (up to integrability assumptions). We will discuss this in section 7.5.

Proof of Theorem 3.13: We expand the expression inside the expectation in (6) and make use of our knowledge of the normal distribution. Let $\{t_{n,j}\}_{j=0}^{N(\pi_n)}$ denote the endpoints of the intervals that make up the partition π_n. First observe that

$$|S(\pi_n) - T|^2 = \left|\sum_{j=1}^{N(\pi_n)} \left\{\left|B_{t_{n,j}} - B_{t_{n,j-1}}\right|^2 - (t_{n,j} - t_{n,j-1})\right\}\right|^2.$$

It is convenient to write $\delta_{n,j}$ for $\left|B_{t_{n,j}} - B_{t_{n,j-1}}\right|^2 - (t_{n,j} - t_{n,j-1})$. Then

$$|S(\pi_n) - T|^2 = \sum_{j=1}^{N(\pi_n)} \left(\delta_{n,j}^2 + 2\sum_{k>j} \delta_{n,j}\delta_{n,k}\right).$$

Note that since Brownian motion has independent increments,

$$\mathbb{E}\left[\delta_{n,j}\delta_{n,k}\right] = \mathbb{E}\left[\delta_{n,j}\right]\mathbb{E}\left[\delta_{n,k}\right] = 0 \qquad \text{if } j \neq k.$$

Also

$$\mathbb{E}\left[\delta_{n,j}^2\right] = \mathbb{E}\left[\left|B_{t_{n,j}} - B_{t_{n,j-1}}\right|^4 - 2\left|B_{t_{n,j}} - B_{t_{n,j-1}}\right|^2 (t_{n,j} - t_{n,j-1}) + (t_{n,j} - t_{n,j-1})^2\right].$$

For a normally distributed random variable, X, with mean zero and variance λ, $\mathbb{E}[|X|^4] = 3\lambda^2$, so we have

$$\begin{aligned}
\mathbb{E}\left[\delta_{n,j}^2\right] &= 3(t_{n,j} - t_{n,j-1})^2 - 2(t_{n,j} - t_{n,j-1})^2 + (t_{n,j} - t_{n,j-1})^2 \\
&= 2(t_{n,j} - t_{n,j-1})^2 \\
&\leq 2\|\pi_n\|(t_{n,j} - t_{n,j-1}).
\end{aligned}$$

Summing over j

$$\begin{aligned}
\mathbb{E}\left[\left|S(\pi_n) - T\right|^2\right] &\leq 2\sum_{j=1}^{N(\pi_n)} \|\pi_n\|(t_{n,j} - t_{n,j-1}) \\
&= 2\|\pi_n\|T \\
&\to 0 \qquad \text{as } n \to \infty.
\end{aligned}$$

□

Corollary 3.14. *Brownian sample paths are of infinite variation on any interval almost surely.*

Corollary 3.15. *Brownian sample paths are almost surely nowhere locally Hölder continuous of order $\gamma > \frac{1}{2}$.*

(The proofs are exercises.)

In fact, a very precise statement is possible.

Theorem 3.16 (Lévy's modulus of continuity (Not Examinable))**.** *For B a Brownian motion,*

$$\limsup_{\varepsilon \downarrow 0} \sup_{0 \le s < t \le 1, t-s \le \varepsilon} \frac{|B_t - B_s|}{\sqrt{2\varepsilon \log(1/\varepsilon)}} = 1 \qquad a.s.$$

Consequently, Brownian sample paths are almost surely nowhere locally Hölder continuous of order $\gamma = 1/2$, and the 2-variation is almost surely infinite.

Proof. Omitted (proof is a careful calculation with estimates of normal random variables, see, for example, Revuz & Yor, p30ff) □

3.6 Small time behaviour of Brownian motion

To study the very small time behaviour of Brownian motion, it is useful to establish the following $0-1$ law.

Theorem 3.17 (Blumenthal's 0-1 law)**.** *Fix a Brownian motion $(B_t)_{t \ge 0}$ on $(\Omega, \mathscr{F}, \mathbb{P})$. Recall $B_0 = 0$. For every $t \ge 0$ we set $\mathscr{F}_t := \sigma(B_u : u \le t)$, so that $\mathscr{F}_s \subset \mathscr{F}_t$ if $s \le t$. We also set $\mathscr{F}_{0+} := \cap_{s>0} \mathscr{F}_s$. Then the σ-field $\mathscr{F}_{0+}$ is trivial in the sense that $\mathbb{P}[A] = 0$ or 1 for every $A \in \mathscr{F}_{0+}$.*

Proof. Let $0 < t_1 < t_2 \cdots < t_k$ and let $g : \mathbb{R}^k \to \mathbb{R}$ be a bounded continuous function. Also, fix $A \in \mathscr{F}_{0+}$. Then by continuity and dominated convergence

$$\mathbb{E}[\mathbf{1}_A g(B_{t_1}, \ldots, B_{t_k})] = \lim_{\varepsilon \downarrow 0} \mathbb{E}[\mathbf{1}_A g(B_{t_1} - B_\varepsilon, \ldots, B_{t_k} - B_\varepsilon)].$$

If $0 < \varepsilon < t_1$, the variables $B_{t_1} - B_\varepsilon, \ldots, B_{t_k} - B_\varepsilon$ are independent of $\mathscr{F}_\varepsilon$ (by the Markov property) and thus also of $\mathscr{F}_{0+}$. It follows that

$$\begin{aligned} \mathbb{E}[\mathbf{1}_A g(B_{t_1}, \ldots, B_{t_k})] &= \lim_{\varepsilon \downarrow 0} \mathbb{E}[\mathbf{1}_A g(B_{t_1} - B_\varepsilon, \ldots, B_{t_k} - B_\varepsilon)] \\ &= \mathbb{P}[A] \mathbb{E}[g(B_{t_1}, \ldots, B_{t_k})]. \end{aligned}$$

We have thus obtained that $\mathscr{F}_{0+}$ is independent of $\sigma(B_{t_1}, \ldots, B_{t_k})$. Since this holds for any finite collection $\{t_1, \ldots, t_k\}$ of (strictly) positive reals, $\mathscr{F}_{0+}$ is independent of $\sigma(B_t, t > 0)$. However, $\sigma(B_t, t > 0) = \sigma(B_t, t \ge 0)$, since B_0 is the pointwise limit of B_t when $t \to 0$. Since $\mathscr{F}_{0+} \subset \sigma(B_t, t \ge 0)$, we conclude that $\mathscr{F}_{0+}$ is independent of itself and so must be trivial. □

Proposition 3.18. *Let B be a standard real-valued Brownian motion, as above.*

i. Then, a.s., for every $\varepsilon > 0$,

$$\sup_{0\le s\le\varepsilon} B_s > 0 \quad \textit{and} \quad \inf_{0\le s\le\varepsilon} B_s < 0.$$

In particular, $\inf\{t > 0 : B_t = 0\} = 0$ *a.s.*

ii. For every $a \in \mathbb{R}$, let $T_a := \inf\{t \ge 0 : B_t = a\}$ (with the convention that $\inf\emptyset = \infty$). Then a.s. for each $a \in \mathbb{R}$, $T_a < \infty$. Consequently, we have a.s.

$$\limsup_{t\to\infty} B_t = +\infty, \quad \liminf_{t\to\infty} B_t = -\infty.$$

Remark 3.19. *It is not a priori obvious that $\sup_{0\le s\le\varepsilon} B_s$ is even measurable, since this is an uncountable supremum of random variables, but since sample paths are continuous, we can restrict to* rational *values of $s \in [0,\varepsilon]$ so that we are taking the supremum over a countable set. We implicitly use this observation in what follows.*

Proof. (i) Let ε_p be a sequence of strictly positive reals decreasing to zero and set $A := \bigcap_{p\ge 0}\{\sup_{0\le s\le\varepsilon_p} B_s > 0\}$. Since this is a monotone decreasing intersection, $A \in \mathscr{F}_{0+}$. On the other hand, by monotonicity,

$$\mathbb{P}[A] = \lim_{p\to\infty}{}^{\downarrow}\Big\{\mathbb{P}[\sup_{0\le s\le\varepsilon_p} B_s > 0]\Big\},$$

where $\lim^{\downarrow}$ denotes a decreasing limit, and

$$\mathbb{P}[\sup_{0\le s\le\varepsilon_p} B_s > 0] \ge \mathbb{P}[B_{\varepsilon_p} > 0] = \frac{1}{2}.$$

So $\mathbb{P}[A] \ge 1/2$ and by Blumenthal's 0-1 law $\mathbb{P}[A] = 1$. Hence a.s. for all $\varepsilon > 0$, $\sup_{0\le s\le\varepsilon} B_s > 0$. Replacing B by $-B$ we obtain $\mathbb{P}[\inf_{0\le s\le\varepsilon} B_s < 0] = 1$.

(ii) Write

$$1 = \mathbb{P}\Big[\sup_{0\le s\le 1} B_s > 0\Big] = \lim_{\delta\downarrow 0}{}^{\uparrow}\mathbb{P}\Big[\sup_{0\le s\le 1} B_s > \delta\Big].$$

Now, writing $c = 1/\delta$ in the Brownian scaling of Proposition 3.10 ii, we have that $B_t^\delta = \delta^{-1}B_{t\delta^2}$ is a Brownian motion. Thus for any $\delta > 0$,

$$\mathbb{P}\Big[\sup_{0\le s\le 1} B_s > \delta\Big] = \mathbb{P}\Big[\sup_{0\le s\le 1/\delta^2} B_s^\delta > 1\Big] = \mathbb{P}\Big[\sup_{0\le s\le 1/\delta^2} B_s > 1\Big]. \tag{7}$$

If we let $\delta \downarrow 0$, we find

$$\mathbb{P}[\sup_{s\ge 0} B_s > 1] = \lim_{\delta\downarrow 0}\mathbb{P}\Big[\sup_{0\le s\le 1/\delta^2} B_s > 1\Big] = \lim_{\delta\downarrow 0}{}^{\uparrow}\mathbb{P}[\sup_{0\le s\le 1} B_s > \delta] = 1$$

.

Another scaling argument shows that, for every $M > 0$,

$$\mathbb{P}[\sup_{s\ge 0} B_s > M] = 1$$

and replacing B with $-B$,

$$\mathbb{P}[\inf_{s\geq 0} B_s < -M] = 1.$$

Continuity of sample paths completes the proof of (ii). □

Corollary 3.20. *The map $t \mapsto B_t$ is a.s. not monotone on any non-trivial interval.*

4 Filtrations and stopping times

4.1 Filtrations, information and adaptedness

These are concepts that you already know about in the context of discrete parameter martingales. Our definitions here mirror what you already know, but in the continuous setting one has to be slightly more careful. In the end, we will make enough assumptions to guarantee that everything goes through nicely.

Definition 4.1. *We say that $\{\mathscr{F}_t\}_{t\geq 0}$ is* right continuous *if for each $t \geq 0$,*

$$\mathscr{F}_t = \mathscr{F}_{t+} \equiv \cap_{\varepsilon>0}\mathscr{F}_{t+\varepsilon}.$$

We say that $\{\mathscr{F}_t\}_{t\geq 0}$ is complete *if $(\Omega, \mathscr{F}, \mathbb{P})$ is complete (contains all subsets of the $\mathbb{P}$-null sets) and $\{A \in \mathscr{F} : \mathbb{P}[A] = 0\} \subset \mathscr{F}_0$ (and hence $\subset \mathscr{F}_t$ for all t).*

Definition 4.2. *A filtration $\{\mathscr{F}_t\}_{t\geq 0}$ (or the filtered space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$) is said to satisfy the* usual conditions *if it is right-continuous and complete.*

Given a filtered probability space, we can always consider a natural augmentation, replacing the filtration with $\sigma(\mathscr{F}_{t+}, \mathscr{N})$, where $\mathscr{N} = \mathscr{N}(\mathbb{P}) := \{A \in \Omega : \exists B \in \mathscr{F}$ such that $A \subseteq B$ and $\mathbb{P}[B] = 0\}$. The augmented filtration satisfies the usual conditions. In Section 6.3 we will see that if we have a martingale with respect to a filtration that satisfies the usual conditions, then it has a right continuous version.

As in discrete time, adaptedness represents the idea of 'knowing X_t at time t':

Definition 4.3. *A process X is* adapted *to a filtration $\{\mathscr{F}_t\}_{t\geq 0}$ if $X(t)$ is $\mathscr{F}_t$-measurable for each $t \geq 0$ (if and only if $\mathscr{F}_t^X \subseteq \mathscr{F}_t$ for all t).*

Adaptedness tells us about measurability in ω at each time t, but nothing about regularity in time. Measurability of a process (Definition 2.3) tells us about regularity in time and space, but not about adaptedness. Putting these together we get the following:

Definition 4.4. *A process X is $\{\mathscr{F}_t\}_{t\geq 0}$-*progressive *(or* progressively measurable*) if for each $t \geq 0$, the mapping $(s, \omega) \mapsto X_s(\omega)$ is measurable on $([0,t] \times \Omega, \mathscr{B}([0,t]) \otimes \mathscr{F}_t)$.*

If X is $\{\mathscr{F}_t\}_{t\geq 0}$-progressive, then it is $\{\mathscr{F}_t\}_{t\geq 0}$-adapted, but the converse is not necessarily true. One can show, with difficulty, that any adapted and measurable process has a progressive modification. However, every right continuous $\{\mathscr{F}_t\}_{t\geq 0}$-adapted process is $\{\mathscr{F}_t\}_{t\geq 0}$-progressive and since we are interested in continuous processes, we will not need to dwell on these details.

Proposition 4.5. *An adapted process (X_t) whose paths are all right-continuous (or are all left-continuous) is* progressively measurable*.*

Proof. We present the argument for a right-continuous X. For $t > 0$, $n \geq 1$, $k = 0, 1, 2 \ldots, 2^n - 1$ let $X_0^{(n)}(\omega) = X_0(\omega)$ and $X_s^{(n)}(\omega) := X_{\frac{k+1}{2^n}t}(\omega)$ for $\frac{kt}{2^n} < s \leq \frac{k+1}{2^n}t$.

Clearly $(X_s^{(n)} : s \leq t)$ takes finitely many values and is $\mathscr{B}([0,t]) \otimes \mathscr{F}_t$–measurable. Further, by right continuity, $X_s(\omega) = \lim_{n\to\infty} X_s^{(n)}(\omega)$, and hence is also measurable (as a limit of measurable mappings). $\square$

Usually we shall consider the natural filtration associated with a process and do not specify it explicitly. On the other hand, sometimes we suppose that we are *given* a filtration $\mathscr{F}_t$. In this case;

Definition 4.6. *A process B is an $\{\mathscr{F}_t\}_{t\geq 0}$-Brownian motion if it is adapted to $\{\mathscr{F}_t\}_{t\geq 0}$, $B_0 = 0$, B has continuous paths, $B_t - B_s \sim N(0, t-s)$ for $t > s$ and $B_t - B_s$ is independent of $\mathscr{F}_s$ for all $t > s$.*

Equivalently, B is adapted, a Brownian motion in its own filtration, and $B_t - B_s$ is independent of $\mathscr{F}_s$ for all $t > s$.

Example 4.7. *Let B be a Brownian motion (in its natural filtration), and let $\mathscr{F}_t = \sigma(B_s; s \leq t) \vee \sigma(B_T)$ for some $T > 0$. Then B is* not *an $\{\mathscr{F}_t\}_{t\geq 0}$-Brownian motion.*

4.2 Stopping times

Again the definition mirrors what you know from the discrete setting.

Definition 4.8. *Let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathscr{P})$ be a filtered probability space. A random variable $\tau : \Omega \mapsto [0, +\infty]$ is called a* stopping time *(relative to $\{\mathscr{F}_t\}_{t\geq 0}$) if $\{\tau \leq t\} \in \mathscr{F}_t$, $\forall t \geq 0$.*

Stopping times are sometimes called optional times (for example, in the 'optional stopping theorem').

The 'first time a certain phenomenon occurs' will be a stopping time. Our fundamental examples will be first hitting times of sets. If X is a stochastic process and $\Gamma \in \mathscr{B}(\mathbb{R})$ we set

$$H_\Gamma(\omega) := \inf\{t \geq 0 : X_t(\omega) \in \Gamma\}. \tag{8}$$

Exercise 4.9. *Show that*

i. if X is adapted to $\{\mathscr{F}_t\}_{t\geq 0}$ and has right-continuous paths then H_Γ, for Γ an open set, is a stopping time relative to $(\mathscr{F}_{t+})$.

ii. if X has continuous paths, then H_Γ, for Γ a closed set, is a stopping time relative to $\{\mathscr{F}_t\}_{t\geq 0}$.

One can show that the hitting time of any Borel set, or of a (reasonably nice) set which changes in time, is a stopping time (assuming $\{\mathscr{F}_t\}_{t\geq 0}$ satisfies the usual conditions), but this is surprisingly difficult!

With a stopping time we can associate 'the information known at time τ':

Definition 4.10. *Given a stopping time τ relative to $\{\mathscr{F}_t\}_{t\geq 0}$ we define*

$$\mathscr{F}_\tau := \{A \in \mathscr{F} : A \cap \{\tau \leq t\} \in \mathscr{F}_t \, \forall t \geq 0\},$$
$$\mathscr{F}_{\tau-} := \sigma(\{A \cap \{\tau > t\} : t \geq 0, A \in \mathscr{F}_t\})$$

(which satisfy all the natural properties).

Proposition 4.11. *Let τ be a stopping time. Then*

(i) $\mathscr{F}_{\tau-}$ and $\mathscr{F}_\tau$ are σ-algebras and τ is $\mathscr{F}_{\tau-}$-measurable.

(ii) $\mathscr{F}_{\tau-} \subseteq \mathscr{F}_\tau$

(iii) If $\tau = t$ then $\mathscr{F}_\tau = \mathscr{F}_t$

(iv) If τ and ρ are stopping times then so are $\tau \wedge \rho$, $\tau \vee \rho$ and $\tau + \rho$ and $\{\tau \leq \rho\} \in \mathscr{F}_{\tau\wedge\rho}$. Further if $\tau \leq \rho$ then $\mathscr{F}_\tau \subseteq \mathscr{F}_\rho$.

(v) If τ is a stopping time and ρ is a $[0,\infty]$-valued random variable which is $\mathscr{F}_\tau$-measurable and $\rho \geq \tau$, then ρ is a stopping time. In particular,

$$\tau_n := \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbf{1}_{\{\frac{k}{2^n} < \tau \leq \frac{k+1}{2^n}\}} + \infty \mathbf{1}_{\{\tau=\infty\}} \tag{9}$$

is a sequence of stopping times with $\tau_n \downarrow \tau$ as $n \to \infty$.

Proof. We prove (v):

Note that $\{\rho \leq t\} = \{\rho \leq t\} \cap \{\tau \leq t\} \in \mathscr{F}_t$ since ρ is $\mathscr{F}_\tau$-measurable. Hence ρ is a stopping time. We have $\tau_n \downarrow \tau$ by definition, and clearly τ_n is $\mathscr{F}_\tau$-measurable since τ is $\mathscr{F}_\tau$-measurable. □

Lemma 4.12. *For any integrable random variable X, any stopping times ρ and τ,*

$$\mathbf{1}_{\rho\leq\tau}\mathbb{E}[X|\mathscr{F}_\rho] = \mathbf{1}_{\rho\leq\tau}\mathbb{E}[X|\mathscr{F}_{\rho\wedge\tau}].$$

Proof. As $\mathscr{F}_{\rho\wedge\tau} \subseteq \mathscr{F}_\rho$, and $\mathbf{1}_{\rho<\tau}$ is $\mathscr{F}_{\rho\wedge\tau}$-measurable,

$$\mathbf{1}_{\rho\leq\tau}\mathbb{E}[X|\mathscr{F}_{\rho\wedge\tau}] = \mathbb{E}[\mathbf{1}_{\rho\leq\tau}\mathbb{E}[X|\mathscr{F}_\rho]|\mathscr{F}_{\rho\wedge\tau}].$$

Therefore, it's enough to show that $\mathbf{1}_{\rho\leq\tau}\mathbb{E}[X|\mathscr{F}_\rho]$ is $\mathscr{F}_{\rho\wedge\tau}$-measurable. This follows from the fact that if $A \in \mathscr{F}_\rho$, then

$$A \cap \{\rho \leq \tau\} \cap \{\tau \leq t\} = (A \cap \{\rho \leq t\}) \cap \{\tau \leq t\} \cap \{\rho \wedge t \leq \tau \wedge t\},$$

so $A \cap \{\rho \leq \tau\} \in \mathscr{F}_\tau$. □

4.3 Stopped processes

It is often useful to be able to 'stop' a process at a stopping time and know that the result still has nice measurability properties.

Theorem 4.13. *Let X be a progressively measurable process and τ a stopping time. Then $X_\tau \mathbf{1}_{\tau<\infty}$ is $\mathscr{F}_\tau$-measurable. The stopped process $X^\tau = (X_{\tau\wedge t} : t \geq 0)$ is progressively measurable (where $X_\tau \mathbf{1}_{\tau<\infty}(\omega) = X_{\tau(\omega)}(\omega)\mathbf{1}_{\tau(\omega)<\infty}$).*

Proof. The first statement is 'easy' once we prove X^τ is progressively measurable. Observe $X_{\tau\wedge s}$ on $[0,t]\times\Omega$ is a composition of two maps

$$\begin{aligned}(s,\omega) &\mapsto (\tau(\omega)\wedge s, \omega),\\ ([0,t]\times\Omega, \mathscr{B}([0,t])\otimes\mathscr{F}_t) &\mapsto ([0,t]\times\Omega, \mathscr{B}([0,t])\otimes\mathscr{F}_t)\end{aligned}$$

and

$$\begin{aligned}(u,\omega) &\mapsto X_u(\omega),\\ ([0,t]\times\Omega, \mathscr{B}([0,t])\otimes\mathscr{F}_t) &\mapsto (\mathbb{R}, \mathscr{B}(\mathbb{R})),\end{aligned}$$

both of which are measurable since τ is a stopping time and X is progressively measurable. □

In other areas of analysis, we often prove results by showing that a statement holds 'locally'. This is made difficult for stochastic processes by the fact that these depend on the random sample point $\omega \in \Omega$, for which we have no topology. One useful way around this challenge is to use stopping times, which helpfully 'localize' in both time and space.

Definition 4.14. *We say that a process X* locally *has some property* C *if there exists a sequence of stopping times $\{\tau_n\}_{n\in\mathbb{N}}$ such that the stopped processes X^{τ_n} have property* C *for every n, and $\tau_n \uparrow \infty$ almost surely. The sequence τ_n is said to* localize *or* reduce *X.*

Example 4.15. *A Brownian motion is* locally bounded*, but is not bounded overall.*

When we work with processes locally, it is then useful to reconstruct the 'global' process from its local versions.

Lemma 4.16. *Given a localizing sequence $\{\tau_n\}_{n\in\mathbb{N}}$ and a family of processes $\{Y^n\}_{n\in\mathbb{N}}$ such that, for all $n \leq m$,*

$$\mathbf{1}_{t\leq\tau_n} Y^n = \mathbf{1}_{t\leq\tau_n} Y^m$$

up to indistinguishability, there exists a process X such that $\mathbf{1}_{t\leq\tau_n} Y^n = \mathbf{1}_{t\leq\tau_n} X$ for all n.

Proof. We can construct X explicitly by

$$X_t = \sum_{n\in\mathbb{N}} \mathbf{1}_{\tau_n < t \leq \tau_{n+1}} Y_t^{n+1}$$

(many constructions are possible). □

5 Strong Markov property and reflection principle

5.1 Strong Markov Property

We are going to use the sequence τ_n of stopping times in (9) to prove an important generalisation of the Markov property for Brownian motion called the *strong* Markov property. Recall that the Markov property says that Brownian motion has 'no memory' – we can start it again from B_s and $B_{t+s} - B_s$ is just a Brownian motion, independent of the path followed by B up to time s. The strong Markov property says that the same is true if we replace s by a stopping time.

Theorem 5.1. *Let $B = (B_t : t \geq 0)$ be a standard Brownian motion on the filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$ and let τ be a stopping time with respect to $\{\mathscr{F}_t\}_{t\geq 0}$. Then, conditional on $\{\tau < \infty\}$, the process*

$$B_t^{(\tau)} := B_{\tau+t} - B_\tau \tag{10}$$

is a standard Brownian motion independent of $\mathscr{F}_\tau$. This is called the strong Markov property *of Brownian motion.*

Proof. Assume that $\tau < \infty$ a.s..

We will show that $\forall\, A \in \mathscr{F}_\tau$, $0 \leq t_1 < \ldots < t_p$ and continuous and bounded functions F on $\mathbb{R}^p$ we have

$$\mathbb{E}\big[\mathbf{1}_A F(B_{t_1}^{(\tau)}, \ldots, B_{t_p}^{(\tau)})\big] = \mathbb{P}(A)\mathbb{E}\big[F(B_{t_1}, \ldots, B_{t_p})\big]. \tag{11}$$

Granted (11), taking $A = \Omega$, we find that B and $B^{(\tau)}$ have the same finite dimensional distributions, and since $B^{(\tau)}$ has continuous paths, it must be a Brownian motion. On the other hand (as usual using a monotone class argument), (11) says that $(B_{t_1}^{(\tau)}, \ldots, B_{t_p}^{(\tau)})$ is independent of $\mathscr{F}_\tau$, and so $B^{(\tau)}$ is independent of $\mathscr{F}_\tau$.

To establish (11), first observe that by continuity of B and F,

$$F(B_{t_1}^{(\tau)}, \ldots, B_{t_p}^{(\tau)}) = \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbf{1}_{\frac{k-1}{2^n} < \tau \leq \frac{k}{2^n}} F\big(B_{\frac{k}{2^n}+t_1} - B_{\frac{k}{2^n}}, \ldots, B_{\frac{k}{2^n}+t_p} - B_{\frac{k}{2^n}}\big) \quad \text{a.s.,}$$

and by the Dominated Convergence Theorem

$$\mathbb{E}\big[\mathbf{1}_A F(B_{t_1}^{(\tau)}, \ldots, B_{t_p}^{(\tau)})\big] = \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbf{1}_A \mathbf{1}_{\frac{k-1}{2^n} < \tau \leq \frac{k}{2^n}} F\big(B_{\frac{k}{2^n}+t_1} - B_{\frac{k}{2^n}}, \ldots, B_{\frac{k}{2^n}+t_p} - B_{\frac{k}{2^n}}\big)\right].$$

For $A \in \mathscr{F}_\tau$, the event $A \cap \{\frac{k-1}{2^n} < \tau \leq \frac{k}{2^n}\} \in \mathscr{F}_{\frac{k}{2^n}}$, so using the simple Markov property at $k/2^n$,

$$\begin{aligned}\mathbb{E}&\left[\mathbf{1}_{A\cap\{\frac{k-1}{2^n} < \tau \leq \frac{k}{2^n}\}} F\big(B_{\frac{k}{2^n}+t_1} - B_{\frac{k}{2^n}}, \ldots, B_{\frac{k}{2^n}+t_p} - B_{\frac{k}{2^n}}\big)\right] \\ &= \mathbb{P}\left[A \cap \left\{\frac{k-1}{2^n} < \tau \leq \frac{k}{2^n}\right\}\right] \mathbb{E}\big[F(B_{t_1}, \ldots, B_{t_p})\big].\end{aligned}$$

Sum over k and take $n \to \infty$ to recover the desired result.

If $\mathbb{P}(\tau = \infty) > 0$, the same argument gives instead

$$\mathbb{E}\left[\mathbf{1}_{A\cap\{\tau<\infty\}} F(B^{(\tau)}_{t_1},\ldots,B^{(\tau)}_{t_p})\right] = \mathbb{P}[A \cap \{\tau < \infty\}]\mathbb{E}\left[F(B_{t_1},\ldots,B_{t_p})\right].$$

□

It was not until the 1940's that Doob properly formulated the strong Markov property and it was 1956 before Hunt proved it for Brownian motion.

5.2 The reflection principle

The following result, the reflection principle, was known at the end of the 19th Century for random walk and appears in the famous 1900 thesis of Bachelier, which introduced the idea of modelling stock prices using Brownian motion (although since he had no formulation of the strong Markov property, his proof is not rigorous).

Theorem 5.2 (The reflection principle)**.** *Let B be a Brownian motion and τ a stopping time. Then the process $\tilde{B}$ defined by*

$$\tilde{B}_t = \begin{cases} B_t & t < \tau, \\ 2B_\tau - B_t & t \geq \tau. \end{cases}$$

is a standard Brownian motion.

Proof. By definition $\tilde{B}$ is a Brownian motion up to the stopping time τ. For $t > \tau$ we write $t = \tau + t'$, and let $\bar{B}_{t'} = B_{\tau+t'} - B_\tau$, which is a Brownian motion independent of (τ, B_τ) by the strong Markov property. Using this and the symmetry of Brownian motion, so that $\bar{B} = -\bar{B}$ in distribution, for $t > \tau$ we have

$$\begin{aligned} B_t &= B_{\tau+t'} - B_\tau + B_\tau \\ &= \bar{B}_{t'} + B_\tau \\ &= -\bar{B}_{t'} + B_\tau \text{ (in distribution by symmetry)} \\ &= 2B_\tau - B_t = \tilde{B}_t. \end{aligned}$$

Thus $\tilde{B}$ has the law of Brownian motion as required. □

Corollary 5.3. *Let $S_t := \sup_{u\leq t} B_u$. For $a \geq 0$ and $b \leq a$ we have*

$$\mathbb{P}[S_t \geq a, B_t \leq b] = \mathbb{P}[B_t \geq 2a - b] \quad \forall t \geq 0.$$

In particular S_t and $|B_t|$ have the same distribution.

Proof. We apply the reflection principle with the stopping time T_a (so $B_{T_a} = a$)

$$\begin{aligned}\mathbb{P}[S_t \geq a, B_t \leq b] &= \mathbb{P}[T_a \leq t, B_t \leq b] \\ &= \mathbb{P}[T_a \leq t, \tilde{B}_t \leq b] \\ &= \mathbb{P}[T_a \leq t, 2a - B_t \leq b] \\ &= \mathbb{P}[B_t \geq 2a - b],\end{aligned}$$

as $2a - b > a$, so $\{B_t \geq 2a - b\} \subseteq \{T_a \leq t\}$.

We have proved that $\mathbb{P}[S_t \geq a, B_t \leq b] = \mathbb{P}[B_t \geq 2a - b]$. For the last assertion of the theorem, taking $b = a$, observe that

$$\begin{aligned}\mathbb{P}[S_t \geq a] &= \mathbb{P}[S_t \geq a, B_t \geq a] + \mathbb{P}[S_t \geq a, B_t \leq a] \\ &= 2\mathbb{P}[B_t \geq a] = \mathbb{P}[B_t \geq a] + \mathbb{P}[B_t \leq -a] \quad \text{(symmetry)} \\ &= \mathbb{P}[|B_t| \geq a].\end{aligned}$$

□

6 (Sub/super-)Martingales in continuous time

The results in this section will to a large extent mirror what you proved last term for discrete parameter martingales (and we use those results repeatedly in our proofs). We assume throughout that a filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$ is given.

6.1 Definitions

Definition 6.1. *An adapted stochastic process* $(X_t)_{t\geq 0}$ *such that* $X_t \in \mathrm{L}^1(\mathbb{P})$ *(i.e.* $\mathbb{E}[|X_t|] < \infty$*) for any* $t \geq 0$*, is called*

i. a martingale *if* $\mathbb{E}[X_t | \mathscr{F}_s] = X_s$ *for all* $0 \leq s \leq t$*,*

ii. a super-martingale *if* $\mathbb{E}[X_t | \mathscr{F}_s] \leq X_s$ *for all* $0 \leq s \leq t$*,*

iii. a sub-martingale *if* $\mathbb{E}[X_t | \mathscr{F}_s] \geq X_s$ *for all* $0 \leq s \leq t$*.*

Exercises: Suppose $(Z_t : t \geq 0)$ is an adapted process with independent increments, i.e. for all $0 \leq s < t$, $Z_t - Z_s$ is independent of $\mathscr{F}_s$. The following give us examples of martingales:

i. if $\forall t \geq 0$, $Z_t \in \mathrm{L}^1$, then $\tilde{Z}_t := Z_t - \mathbb{E}[Z_t]$ is a martingale,

ii. if $\forall t \geq 0$, $Z_t \in \mathrm{L}^2$, then $\tilde{Z}_t^2 - \mathbb{E}[\tilde{Z}_t^2]$ is a martingale,

iii. if for some $\theta \in \mathbb{R}$, and $\forall t \geq 0$, $\mathbb{E}[e^{\theta Z_t}] < \infty$, then $\frac{e^{\theta Z_t}}{\mathbb{E}[e^{\theta Z_t}]}$ is a martingale.

In particular, B_t, $B_t^2 - t$ and $e^{\theta B_t - \theta^2 t/2}$ are all martingales with respect to a filtration $\{\mathscr{F}_t\}_{t\geq 0}$ for which $(B_t)_{t\geq 0}$ is a Brownian motion.

Warning: It is important to remember that a process is a martingale *with respect to a filtration* – giving yourself more information (enlarging the filtration) may destroy the martingale property. For us, even when we don't explicitly mention it, there is a filtration implicitly assumed (usually the natural filtration associated with the process, augmented to satisfy the usual conditions).

Given a martingale (or submartingale) it is easy to generate many more.

Proposition 6.2. *Let $(X_t)_{t\geq 0}$ be a martingale (respectively sub-martingale) and $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex (respectively convex and increasing) such that $\mathbb{E}[|\varphi(X_t)|] < \infty$ for any $t \geq 0$. Then $(\varphi(X_t))_{t\geq 0}$ is a sub-martingale.*

Proof. Apply the conditional Jensen inequality (see appendix, Lemma A.28). □

In particular, if $(X_t)_{t\geq 0}$ is martingale with $\mathbb{E}[|X_t|^p] < \infty$, for some $p \geq 1$ and all $t \geq 0$, then $|X_t|^p$ is a sub-martingale (and consequently, $t \mapsto \mathbb{E}[|X_t|^p]$ is non-decreasing).

6.2 Doob's maximal inequalities

Doob was the person who placed martingales on a firm mathematical foundation (beginning in the 1940's). He initially called them 'processes with the property E', but reverted to the term martingale in his monumental book.

Doob's inequalities are fundamental to proving convergence theorems for martingales. You already encountered them in the discrete setting and we shall recall those results that underpin our proofs in the continuous world here. They allow us to control the running maximum of a martingale. The following results (Corollary A.24 and Theorem A.26) are standard in discrete time.

Theorem 6.3. *If $(X_n)_{n\geq 0}$ is a discrete martingale (or a nonnegative submartingale) w.r.t. some filtration $(\mathscr{F}_n)$, then for any $N \in \mathbb{N}$, $p \geq 1$ and $\lambda > 0$,*

$$\lambda^p \mathbb{P}\Big[\sup_{n\leq N} |X_n| \geq \lambda\Big] \leq \mathbb{E}\big[|X_N|^p\big]$$

and for any $p > 1$

$$E\big[|X_N|^p\big] \leq E\left[\sup_{n\leq N} |X_n|^p\right] \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[|X_N|^p\big].$$

We would now like to extend this to continuous time.

Suppose that X is indexed by $t \in [0,\infty)$. Take a countable dense set D in $[0,T]$, e.g. $D = \mathbb{Q} \cap [0,T]$, and an increasing sequence of finite subsets $D_n \subseteq D_{n+1} \subseteq D$ such that $\cup_{n=1}^{\infty} D_n = D$.

The above inequalities hold for X indexed by $t \in D_n \cup \{T\}$. Monotone convergence then yields the result for $t \in D$. If X has regular sample paths (e.g. right continuous) then the supremum over a countable dense set in $[0,T]$ is the same as over the *whole* of $[0,T]$ and so:

Theorem 6.4 (Doob's maximal and L^p inequalities)**.**
If $(X_t)_{t\geq 0}$ is a right continuous martingale or nonnegative sub-martingale, then for any $T \geq 0$, $\lambda > 0$,

$$\begin{aligned} \mathbb{P}\Big[\sup_{t\leq T}|X_t| \geq \lambda\Big] &\leq \frac{1}{\lambda^p}\mathbb{E}\big[|X_T|^p\big], \quad p \geq 1 \\ E\left[\sup_{t\leq T}|X_t|^p\right] &\leq \Big(\frac{p}{p-1}\Big)^p\mathbb{E}\big[|X_T|^p\big], \quad p > 1. \end{aligned} \tag{12}$$

As an application of Doob's maximal inequality, we derive a useful bound for Brownian motion.

Proposition 6.5. *Let $(B_t)_{t\geq 0}$ be Brownian motion and $S_t = \sup_{u\leq t} B_u$. For any $\lambda > 0$ we have*

$$\mathbb{P}[S_t \geq \lambda t] \leq e^{-\frac{\lambda^2 t}{2}}.$$

Proof. Recall that $e^{\alpha B_t - \alpha^2 t/2}$, $t \geq 0$, is a non-negative martingale. It follows that, for $\alpha \geq 0$, using Doob,

$$\begin{aligned} \mathbb{P}[S_t \geq \lambda t] &\leq \mathbb{P}\Big[\sup_{u\leq t}\big(e^{\alpha B_u - \alpha^2 t/2}\big) \geq e^{\alpha\lambda t - \alpha^2 t/2}\Big] \\ &\leq \mathbb{P}\Big[\sup_{u\leq t}\big(e^{\alpha B_u - \alpha^2 u/2}\big) \geq e^{\alpha\lambda t - \alpha^2 t/2}\Big] \\ &\leq e^{-\alpha\lambda t + \alpha^2 t/2}\underbrace{\mathbb{E}\left[e^{\alpha B_t - \alpha^2 t/2}\right]}_{=1}. \end{aligned}$$

The bound now follows since $\min_{\alpha\geq 0} e^{-\alpha\lambda t + \alpha^2 t/2} = e^{-\lambda^2 t/2}$ (with the minimum achieved when $\alpha = \lambda$). □

In the next subsection, we are going to show that even if a supermartingale is not right continuous, it has a right continuous version (this is Doob's Regularisation Theorem). To prove this, we need a slight variant of the maximal inequality – this time for a supermartingale – which in turn relies on Doob's Optional Stopping (or Sampling) Theorem for discrete supermartingales.

Theorem 6.6 (Doob's Optional Stopping Theorem for discrete supermartinagles)**. (bounded case)**

If $(Y_n)_{n\geq 1}$ is a supermartingale, then for any choice of bounded stopping times S and T such that $S \leq T$, we have

$$Y_S \geq \mathbb{E}[Y_T | \mathscr{F}_S].$$

Here's the version of the maximal inequality that we shall need.

Proposition 6.7. *Let* $(X_t : t \geq 0)$ *be a supermartingale. Then*

$$\mathbb{P}\left[\sup_{t\in[0,T]\cap\mathbb{Q}} |X_t| \geq \lambda\right] \leq \frac{1}{\lambda}\left(2\mathbb{E}[|X_T|]+\mathbb{E}[|X_0|]\right), \quad \forall \lambda, T > 0. \tag{13}$$

In particular, $\sup_{t\in[0,T]\cap\mathbb{Q}} |X_t| < \infty$ *a.s.*

Proof. Take a sequence of rational numbers $0 = t_0 < t_1 < \ldots < t_n = T$. Applying Theorem 6.6 with $S = \min\{t_i : X_{t_i} \geq \lambda\} \wedge T$, we obtain

$$\mathbb{E}[X_0] \geq \mathbb{E}[X_S] \geq \lambda \mathbb{P}[\sup_{1\leq i\leq n} X_{t_i} \geq \lambda] + \mathbb{E}[X_T \mathbf{1}_{\sup_{1\leq i\leq n} X_{t_i} < \lambda}].$$

Rearranging,

$$\lambda \mathbb{P}(\sup_{1\leq i\leq n} X_{t_i} \geq \lambda) \leq \mathbb{E}[X_0] + \mathbb{E}[X_T^-]$$

(where $X_T^- = -\min(X_T, 0)$). Now X_T^- is a non-negative submartingale and so we can apply Doob's inequality directly to it, from which

$$\lambda \mathbb{P}(\sup_{1\leq i\leq n} X_{t_i}^- \geq \lambda) \leq \mathbb{E}[X_T^-],$$

and, since $\mathbb{E}[X_T^-] \leq \mathbb{E}[|X_T|]$, taking the (monotone) limit in nested sequences in $[0,T]\cap\mathbb{Q}$, gives the result. □

6.3 Convergence and regularisation theorems

As advertised, our aim in this section is to prove that, provided the filtration satisfies 'the usual conditions', any martingale has a version with right continuous paths.

First we recall the notion of upcrossing numbers.

Definition 6.8. *Let* $f : I \to \mathbb{R}$ *be a function defined on a subset* I *of* $[0,\infty)$. *If* $a < b$, *the upcrossing number of* f *along* $[a,b]$, *which we shall denote* $U([a,b],(f_t)_{t\in I})$ *is the maximal integer* $k \geq 1$ *such that there exists a sequence* $s_1 < t_1 < \cdots < s_k < t_k$ *of elements of* I *such that* $f(s_i) < a$ *and* $f(t_i) > b$ *for every* $i = 1,\ldots,k$.

If even for $k = 1$ *there is no such sequence, we take* $U([a,b],(f_t)_{t\in I}) = 0$. *If such a sequence exists for* every $k \geq 1$, *we set* $U([a,b],(f_t)_{t\in I}) = \infty$.

Upcrossing numbers are a convenient tool for studying the regularity of functions. We omit the proof of the following analytic lemma.

Lemma 6.9. *Let* D *be a countable dense set in* $[0,\infty)$ *and let* f *be a real function defined on* D. *Assume that for every* $T \in D$

i. f *is bounded on* $D \cap [0,T]$;

ii. for all rationals a *and* b *such that* $a < b$

$$U([a,b],(f_t)_{t\in D\cap[0,T]}) < \infty.$$

Then the right limit

$$f(t+) = \lim_{s \downarrow t, s \in D} f(s)$$

exists for every real $t \geq 0$, and similarly the left limit

$$f(t-) = \lim_{s \uparrow t, s \in D} f(s)$$

exists for any real $t > 0$.

Furthermore, the function $g : \mathbb{R}_+ \to \mathbb{R}$ defined by $g(t) = f(t+)$ is càdlàg *('continue à droite avec des limites à gauche'; i.e. right continuous with left limits) at every $t > 0$.*

Lemma 6.10 (Doob's upcrossing lemma in discrete time)**.** *Let $(X_t)_{t \geq 0}$ be a supermartingale and F a finite subset of $[0,T]$. If $a < b$ then*

$$\mathbb{E}\Big[U\big([a,b],(X_n : n \in F)\big)\Big] \leq \sup_{n \in F} \frac{\mathbb{E}\big[(X_n - a)^-\big]}{b-a} \leq \frac{\mathbb{E}\big[(X_T - a)^-\big]}{b-a}.$$

The last inequality follows since $(X_t - a)^-$ is a submartingale. By monotone convergence

$$\lim_{k \to \infty} \mathbb{E}\big[U\big([a, b-1/k],(X_n : n \in F)\big)\big] = \mathbb{E}\big[U\big([a,b),(X_n : n \in F)\big)\big]$$

satisfies the same bound (and similarly for other intervals)

Taking an increasing sequence F_n and setting $\cup_n F_n = F$, this immediately extends to a countable $F \subset [0,T]$. From this we deduce:

Theorem 6.11. *If (X_t) is a right-continuous super-martingale and $\sup_t \mathbb{E}[X_t^-] < \infty$ then $X_\infty = \lim_{t \to \infty} X_t$ exists (convergence a.s.) and X_∞ is in L^1. In particular, a non-negative right-continuous supermartingale converges a.s. as $t \to \infty$.*

Proof. By right continuity, for any $\varepsilon > 0$,

$$U([a,b],(X_t)_{t \in [0,T]}) \leq U([a, b-\varepsilon],(X_t)_{t \in [0,T] \cap \mathbb{Q}}).$$

Also, by Lemma 6.9, a bounded sequence $(x_n)_{n \geq 1}$ converges if and only if the number of upcrossings is finite, that is $U([a,b],(x_n)_{n \geq 1}) < \infty$ for all $a < b$ with $a, b \in \mathbb{Q}$. By the above calculations and Lemma 6.10, these statements can be taken to hold almost surely for the paths of our supermartingale X. Hence $\{X_{t_n}\}$ converges a.s. for any sequence $t_n \uparrow \infty$, but this implies X_t converges a.s. as $t \to \infty$.

As X is a supermartingale

$$\mathbb{E}[|X_t|] = \mathbb{E}[X_t] + 2\mathbb{E}[X_t^-] \leq \mathbb{E}[X_0] + 2E[X_t^-]$$

so by Fatou's inequality

$$\mathbb{E}[|X_\infty|] = \mathbb{E}[\lim_t |X_t|] \leq \liminf_t \mathbb{E}[|X_t|] < \infty,$$

that is, $X_\infty \in L^1$. □

Remark 6.12. *Note the convergence here is almost sure, not in L^1 (that is, we usually don't have $\mathbb{E}[|X_t - X_\infty|] \to 0$ or $\mathbb{E}[X_t] \to \mathbb{E}[X_\infty]$)!*

Example 6.13. *By direct calculation, we know $X_t = \exp(\theta B_t - \theta^2 t/2)$ defines a martingale, and clearly $X \geq 0$, so X_t converges almost surely as $t \to \infty$. Restricting to $t \in \mathbb{N}$, from the strong law of large numbers, we know that*

$$\frac{B_t}{t} = \frac{1}{t}\sum_{s=1}^{t}(B_s - B_{s-1}) \to 0$$

and hence as $t \to \infty$

$$\theta B_t - \frac{\theta^2 t}{2} = t\Big(\theta\frac{B_t}{t} - \frac{\theta^2}{2}\Big) \to -\infty.$$

It follows that $X_t \to X_\infty = 0$ a.s., but

$$\mathbb{E}[|X_t - X_\infty|] = \mathbb{E}[X_t] = 1 \not\to 0 \qquad \textit{and} \qquad X_t \neq \mathbb{E}[X_\infty|\mathscr{F}_t].$$

6.4 (Super)martingale continuity

The next result says that we can talk about left and right limits for a general supermartingale and then our analytic lemma will tell us how to find a càdlàg version.

Theorem 6.14. *If $(X_t : t \geq 0)$ is a supermartingale, then for $\mathbb{P}$-almost every $\omega \in \Omega$,*

$$\forall t \in (0,\infty) \quad \lim_{r\uparrow t,\, r\in\mathbb{Q}} X_r(\omega) \text{ and } \lim_{r\downarrow t,\, r\in\mathbb{Q}} X_r(\omega) \text{ exist and are finite.} \tag{14}$$

Proof. Fix $T > 0$. From Lemma 6.10, as $\mathbb{E}[(X_T - a)^-] \leq \mathbb{E}[|X_T|] + a < \infty$, there exists $\Omega^T \subseteq \Omega$, with $\mathbb{P}(\Omega^T) = 1$, such that for any $\omega \in \Omega^T$

$$\forall a,b \in \mathbb{Q} \text{ with } a < b, \qquad U\big([a,b],(X_t(\omega) : t \in [0,T]\cap\mathbb{Q})\big) < \infty.$$

Also, by Proposition 6.7,

$$\sup_{t\in[0,T]\cap\mathbb{Q}} |X_t(\omega)| < \infty.$$

It follows by Lemma 6.9 that the limits in (14) are well defined and finite for all $t \leq T$ and $\omega \in \Omega^T$. To complete the proof, take $\Omega := \Omega^1 \cap \Omega^2 \cap \Omega^3 \cap \ldots$. □

Using this, even if X is not right-continuous, its right-continuous version is a.s. well defined. The following fundamental regularisation result is again due to Doob. We begin by recalling Vitali's convergence theorem:

Theorem 6.15 (Vitali convergence theorem)**.** *Let $\{Y_k\}$ be family of random variables, and suppose $Y_k \to Y_\infty$ in probability (or a.s.). Then $Y_k \to Y_\infty$ in L^1 if and only if $\{Y_k\}$ is a uniformly integrable family.*

Proof. See appendix (Theorem A.18) □

Theorem 6.16. *Let X be a supermartingale with respect to a right-continuous and complete filtration $\{\mathscr{F}_t\}_{t\geq 0}$. If $t \mapsto \mathbb{E}[X_t]$ is right continuous (e.g. if X is a martingale) then X admits a modification with càdlàg paths, which is also an $\{\mathscr{F}_t\}_{t\geq 0}$-supermartingale.*

Corollary 6.17. *If X is a martingale then its càdlàg modification is also a martingale.*

Proof. By Theorem 6.14, there exists $\Omega_0 \subseteq \Omega$, with $\mathbb{P}[\Omega_0] = 1$, such that the process

$$X_{t+}(\omega) = \begin{cases} \lim_{r\downarrow t,\, r\in\mathbb{Q}} X_r(\omega) & \omega \in \Omega_0 \\ 0 & \omega \notin \Omega_0 \end{cases}$$

is well defined and adapted to $\{\mathscr{F}_t\}_{t\geq 0}$. By Lemma 6.9, it has càdlàg paths.

To check that we really have only produced a modification of X_t, that is $X_t = X_{t+}$ almost surely, let $t_n \downarrow t$ be a sequence of rationals. Then $\{X_{t_k}\}$ is uniformly integrable (see appendix[1], Lemma A.19) and converges a.s. to X_{t+}. By Vitali's convergence theorem, $X_{t_k} \to X_{t+}$ in L^1, so we can pass to the limit $n \to \infty$ in the inequality $X_t \geq \mathbb{E}[X_{t_n}|\mathscr{F}_t]$ to obtain $X_t \geq \mathbb{E}[X_{t+}|\mathscr{F}_t] = X_{t+}$ a.s. as $\mathscr{F}_t$ is right continuous.

Right continuity of $t \mapsto \mathbb{E}[X_t]$ implies $\mathbb{E}[X_{t+} - X_t] = 0$, so that $X_t = X_{t+}$ almost surely. It follows that

$$X_{s+} = X_s \geq \mathbb{E}[X_t|\mathscr{F}_s] = \mathbb{E}[X_{t+}|\mathscr{F}_s] \qquad a.s.$$

which confirms that the right-continuous modification is a supermartingale. Applying this to X and $-X$ gives the corollary. □

Given this result, we will now often *assume* that our (sub/super)-martingales are càdlàg.

Remark 6.18. *Let's make some comments on the assumptions of the theorem.*

i. *The assumption that the filtration is right continuous is necessary. For example, let $\Omega = \{-1, +1\}$ and $\mathbb{P}[\{1\}] = \mathbb{P}[\{-1\}] = 1/2$. We set*

$$X_t(\omega) = \begin{cases} 0, & 0 \leq t \leq 1, \\ \omega, & t > 1. \end{cases}$$

Then X is a martingale with respect to its natural filtration (which is complete since there are no nonempty negligible sets), but no modification of X can be right continuous at $t = 1$.

ii. *Similarly, take $X_t = f(t)$, where $f(t)$ is deterministic, non-increasing and not right continuous. Then no modification can have right continuous sample paths.*

[1] An alternative is to show that $\{X_{t_k}\}$ is a *backwards supermartingale* and then use the backwards supermartingale convergence theorem, which effectively amounts to proving uniform integrability

6.5 Martingale convergence and optional stopping

In Theorem 6.11 we showed that:

Theorem. *Let X be a supermartingale with right continuous sample paths. Assume that $(X_t)_{t\geq 0}$ is bounded in L^1, i.e.* $\sup_t \mathbb{E}[|X_t|] < \infty$ *(or more generally* $\sup_t \mathbb{E}[X_t^-] < \infty$*). Then there exists $X_\infty \in L^1$ such that* $\lim_{t\to\infty} X_t = X_\infty$ *almost surely.*

Under the assumptions of this theorem, X_t may not converge to X_∞ in L^1. The next result gives, for martingales, necessary and sufficient conditions for L^1-convergence.

Definition 6.19. *A martingale is said to be* closed *if there exists a random variable $Z \in L^1$ such that for every $t \geq 0$, $X_t = \mathbb{E}[Z|\mathscr{F}_t]$.*

Theorem 6.20 (Martingale Convergence Theorem)**.** *Let $(X_t : t \geq 0)$ be a martingale with right continuous sample paths. Then TFAE:*

i. X is closed;

ii. the collection $(X_t)_{t\geq 0}$ is uniformly integrable;

iii. X_t converges almost surely and in L^1 as $t \to \infty$.

Moreover, if these properties hold, $X_t = \mathbb{E}[X_\infty|\mathscr{F}_t]$ for every $t \geq 0$, where $X_\infty \in L^1$ is the almost sure limit of X_t as $t \to \infty$.

Proof. That the first condition implies the second is easy. If $Z \in L^1$, then $\mathbb{E}[Z|\mathscr{G}]$, where $\mathscr{G}$ varies over sub σ-fields of $\mathscr{F}$ is uniformly integrable.

As ii implies Theorem 6.11, under both ii and iii we have almost sure convergence. Vitali's theorem then states that ii and iii are equivalent.

Finally, if the third condition holds, for every $s \geq 0$, pass to the limit as $t \to \infty$ in the equality $X_s = \mathbb{E}[X_t|\mathscr{F}_s]$ (using the fact that conditional expectation is continuous for the L^1-norm, see appendix, Lemma A.29) and obtain $X_s = \mathbb{E}[X_\infty|\mathscr{F}_s]$. □

We would now like to establish conditions under which we have an optional stopping theorem for continuous martingales. As usual, our starting point will be the corresponding discrete time result and we shall pass to a suitable limit.

Theorem 6.21 (Optional stopping for uniformly integrable discrete time martingales)**.** *Let $(Y_n)_{n\in\mathbb{N}}$ be a uniformly integrable martingale with respect to the filtration $(\mathscr{G}_n)_{n\in\mathbb{N}}$, and let Y_∞ be the a.s. limit of Y_n when $n \to \infty$. Then, for every choice of the stopping times S and T such that $S \leq T$, we have $Y_T \in L^1$ and*

$$Y_S = \mathbb{E}[Y_T|\mathscr{G}_S],$$

where

$$\mathscr{G}_S = \{A \in \mathscr{G}_\infty : A \cap \{S = n\} \in \mathscr{G}_n \text{ for every } n \in \mathbb{N}\},$$

with the convention that $Y_T = Y_\infty$ on the event $\{T = \infty\}$, and similarly for Y_S.

Let $(X_t)_{t\geq 0}$ be a right continuous martingale or supermartingale such that X_t converges almost surely as $t \to \infty$ to a limit X_∞. Then for every stopping time T, we define

$$X_T(\omega) = \mathbf{1}_{\{T(\omega)<\infty\}} X_{T(\omega)}(\omega) + \mathbf{1}_{\{T(\omega)=\infty\}} X_\infty(\omega).$$

Theorem 6.22 (Optional Stopping Theorem for UI continuous time martingales)**.** *Let $(X_t)_{t\geq 0}$ be a uniformly integrable martingale with right continuous sample paths. Let S and T be two stopping times with $S \leq T$. Then X_S and X_T are in L^1 and $X_S = \mathbb{E}[X_T|\mathscr{F}_S]$.*

In particular, for every stopping time S we have $X_S = \mathbb{E}[X_\infty|\mathscr{F}_S]$ and $\mathbb{E}[X_S] = \mathbb{E}[X_\infty] = \mathbb{E}[X_0]$.

Proof. For any integer $n \geq 0$ set

$$T_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbf{1}_{\{k2^{-n}<T\leq(k+1)2^{-n}\}} + \infty \mathbf{1}_{\{T=\infty\}},$$

$$S_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbf{1}_{\{k2^{-n}<S\leq(k+1)2^{-n}\}} + \infty \mathbf{1}_{\{S=\infty\}}.$$

Then T_n and S_n are sequences of stopping times that decrease respectively to T and S. Moreover, $S_n \leq T_n$ for every $n \geq 0$.

For each fixed n, $2^n S_n$ and $2^n T_n$ are stopping times of the discrete filtration $\mathscr{G}_k^{(n)} = \mathscr{F}_{k/2^n}$ and $Y_k^{(n)} = X_{k/2^n}$ is a discrete martingale with respect to this filtration.

From Theorem 6.21, $Y^{(n)}_{2^n S_n}$ and $Y^{(n)}_{2^n T_n}$ are in L^1 and

$$X_{S_n} = Y^{(n)}_{2^n S_n} = \mathbb{E}[Y^{(n)}_{2^n T_n}|\mathscr{G}^{(n)}_{2^n S_n}] = \mathbb{E}[X_{T_n}|\mathscr{F}_{S_n}].$$

Let $A \in \mathscr{F}_S$. Since $\mathscr{F}_S \subseteq \mathscr{F}_{S_n}$ we have $A \in \mathscr{F}_{S_n}$ and so $\mathbb{E}[\mathbf{1}_A X_{S_n}] = \mathbb{E}[\mathbf{1}_A X_{T_n}]$. By right continuity, $X_S = \lim_{n\to\infty} X_{S_n}$ and $X_T = \lim_{n\to\infty} X_{T_n}$. The limits also hold in L^1 (in fact, by Theorem 6.21, $X_{S_n} = \mathbb{E}[X_\infty|\mathscr{F}_{S_n}]$ for every n and so $(X_{S_n})_{n\geq 1}$ and $(X_{T_n})_{n\geq 1}$ are uniformly integrable). L^1 convergence implies that the limits X_S and X_T are in L^1 and allows us to pass to a limit, $\mathbb{E}[\mathbf{1}_A X_S] = \mathbb{E}[\mathbf{1}_A X_T]$. This holds for all $A \in \mathscr{F}_S$ and so since X_S is $\mathscr{F}_S$-measurable we conclude that $X_S = \mathbb{E}[X_T|\mathscr{F}_S]$, as required. □

Corollary 6.23. *In particular, for any martingale with right continuous paths and two* bounded *stopping times, $S \leq T$, we have X_S, $X_T \in L^1$ and $X_S = \mathbb{E}[X_T|\mathscr{F}_S]$.*

Proof. Let a be such that $S \leq T \leq a$. The martingale $(X_{t\wedge a})_{t\geq 0}$ is closed by X_a and so we may apply our previous results. □

Corollary 6.24. *Suppose that $(X_t)_{t\geq 0}$ is a martingale with right continuous paths and T is a stopping time.*

i. $X^T = (X_{t\wedge T})_{t\geq 0}$ is a martingale;

ii. if, in addition, $(X_t)_{t\geq 0}$ is uniformly integrable, then $X^T = (X_{t\wedge T})_{t\geq 0}$ is uniformly integrable and for every $t \geq 0$, $X_{t\wedge T} = \mathbb{E}[X_T|\mathscr{F}_t]$.

Proof. We know $X^T_t = X_{t\wedge T} = X^t_T$, and that X_t is integrable. Hence, by the optional stopping theorem applied to the stopped process X^t, we see that X^T_t is integrable for every t. Furthermore, for any $s < t$, as $T\wedge s$ and $T\wedge t$ are bounded stopping times, by the optional stopping theorem and Lemma 4.12,

$$\begin{aligned} X^T_s = X_{T\wedge s} = \mathbb{E}[X_{T\wedge t}|\mathscr{F}_{T\wedge s}] &= \mathbf{1}_{T<s}X_T + \mathbf{1}_{T\geq s}\mathbb{E}[X_{T\wedge t}|\mathscr{F}_{T\wedge s}] \\ &= \mathbf{1}_{T<s}X_{T\wedge t} + \mathbf{1}_{T\geq s}\mathbb{E}[X_{T\wedge t}|\mathscr{F}_s] = \mathbb{E}[X^T_t|\mathscr{F}_s]. \end{aligned}$$

Therefore X^T is a martingale. □

A converse result is also possible:

Theorem 6.25. *Suppose M is a right-continuous process defined for $t < \infty$, and adapted to a right-continuous filtration $\{\mathscr{F}_t\}_{t<\infty}$. Then M is a martingale if, and only if, for every bounded stopping time T we know $\mathbb{E}\big[|M_T|\big] < \infty$ and $\mathbb{E}[M_T] = \mathbb{E}[M_0]$.*

Proof. By considering the process $\{M_t - M_0\}_{t\geq 0}$, we can assume without loss of generality that $E[M_T] = E[M_0] = 0$. If M is a martingale, then $M_t = E[M_T|\mathscr{F}_t]$, and the result follows by optional stopping and Jensen's inequality.

Conversely, consider any times $s < t \in [0,\infty)$ and any $A \in \mathscr{F}_s$. Define a random time T by putting $T(\omega) = s$ if $\omega \in A$ and $T(\omega) = t$ if $\omega \notin A$. Then T is a stopping time. By hypothesis

$$\mathbb{E}[\mathbf{1}_A M_s] + \mathbb{E}[\mathbf{1}_{A^c} M_t] = \mathbb{E}[M_T] = 0 = \mathbb{E}[M_t] = \mathbb{E}[\mathbf{1}_A M_t] + \mathbb{E}[\mathbf{1}_{A^c} M_t].$$

Therefore

$$\mathbb{E}[\mathbf{1}_A M_s] = \mathbb{E}[\mathbf{1}_A M_t]$$

for all $A \in \mathscr{F}_s$, so $M_s = E[M_t|\mathscr{F}_s]$ almost surely. □

Above all, optional stopping is a powerful tool for explicit calculations.

Example 6.26. *Fix $a > 0$ and let T_a be the first hitting time of a by standard Brownian motion. Then for each $\lambda > 0$,*

$$\mathbb{E}[e^{-\lambda T_a}] = e^{-a\sqrt{2\lambda}}.$$

Recall that $N^\lambda_t = \exp(\lambda B_t - \frac{\lambda^2}{2}t)$ is a martingale. So $N^\lambda_{t\wedge T_a}$ is still a martingale and it is in the bounded interval $[0, e^{\lambda a}]$ and hence is uniformly integrable, so $\mathbb{E}[N^\lambda_{T_a}] = \mathbb{E}[N^\lambda_0]$. That is,

$$e^{a\lambda}\mathbb{E}[e^{-\lambda^2 T_a/2}] = \mathbb{E}[N^\lambda_0] = 1.$$

Replace λ by $\sqrt{2\lambda}$ and rearrange.

Warning: This argument fails if $\lambda < 0$ – the reason being that we lose the uniform integrability.

7 Continuous semimartingales

Recall that our original goal was to make sense of differential equations driven by 'rough' inputs. In fact, we will recast our differential equations as integral equations and so we must develop a theory that allows us to integrate with respect to 'rough' driving processes. The class of processes with which we work are called *semimartingales*, and we shall specialise to the continuous ones.

We are going to start with functions for which the integration theory that we already know is adequate – these are called functions of finite variation.

Throughout, we assume that a filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions is given.

7.1 Functions of finite variation

Throughout this section we only consider real-valued right-continuous functions on $[0,\infty)$. Our arguments will be shift invariant so, without loss of generality, we assume that any such function a satisfies $a(0) = 0$. Recall the following definition.

Definition 7.1. *The (total) variation of a function a over* $[0,T]$ *is defined as*

$$V(a)_T = \sup_{\pi} \sum_{i=0}^{m(\pi)-1} |a_{t_{i+1}} - a_{t_i}|,$$

where the supremum is over partitions $\pi = \{0 = t_0 < t_1 < \ldots < t_{N(\pi)} = T\}$ *of* $[0,T]$. *We say that a is of* finite variation *on* $[0,T]$ *if* $V(a)_T < \infty$. *The function a is of* finite variation *if* $V(a)_T < \infty$ *for all* $T \geq 0$ *and of bounded variation if* $\lim_{T\to\infty} V(a)_T < \infty$.

Remark 7.2. *Note that* $t \to V(a)_t$ *is non-negative, right-continuous (whenever finite) and non-decreasing in t. This follows since any partition of* $[0,s]$ *may be included in a partition of* $[0,t]$, $t \geq s$.

Proposition 7.3. *The function a is of finite variation if and only if it is equal to the difference of two non-decreasing functions,* a_1 *and* a_2.

Moreover, if a is of finite variation, then a_1 *and* a_2 *can be chosen so that* $V(a)_t = a_1(t) + a_2(t)$. *If a is càdlàg then* $V(a)_t$ *is also càdlàg.*

Proof. Let

$$a_1(t) = \frac{1}{2}(V(a)_t - a(t)) = \frac{1}{2}\sup_{\pi} \sum_{i=0}^{N(\pi)-1} \big(|a(t_{i+1}) - a(t_i)| - \big(a(t_{i+1}) - a(t_i)\big)\big).$$

This is a non-decreasing function of t, as is $a_2(t) = (V(a)_t + a(t))/2$, and these functions have the required properties. □

If we define measures μ_+, μ_- by

$$\mu_+((0,t]) = \frac{V(a)_t + a(t)}{2}, \qquad \mu_-((0,t]) = \frac{V(a)_t - a(t)}{2},$$

then we can develop a theory of integration with respect to a by declaring that

$$\int_0^t f(s)da(s) = \int_0^t f(s)\mu_+(ds) - \int_0^t f(s)\mu_-(ds),$$

provided that

$$\int_0^t |f(s)||\mu|(ds) = \int_0^t |f(s)|\left(\mu_+(ds) + \mu_-(ds)\right) < \infty,$$

when we say that f is a-integrable. We say that $\mu = \mu_+ - \mu_-$ is the *signed measure* associated with a, μ_+, μ_- is its *Jordan decomposition* and $\int_0^t f(s)da(s)$ is the *Lebesgue-Stieltjes* integral of f with respect to a.

We sometimes use the notation

$$(f \cdot a)(t) = \int_0^t f(s)da(s).$$

The function $(f \cdot a)$ will be right continuous and of finite variation whenever a is of finite variation and f is integrable with respect to a (exercise).

Example 7.4. *For some $\lambda \in \mathbb{R}$, let $a(t) = 1 - e^{-\lambda t}$. Then $\mu([a,b)) = e^{-\lambda a} - e^{-\lambda b} = \int_a^b \lambda e^{-\lambda t} dt$, and we find*

$$(f \cdot a)(t) = \int_0^t f(s)\lambda e^{-\lambda s} ds.$$

The same approach works whenever a is any distribution function.

Remark 7.5. *A function $a \in C^1$ is of finite variation and $\int f(s)\mathrm{d}a(s) = \int f(s)a'(s)\mathrm{d}s$.*

7.2 Finite variation integrals

It is worth recording that our integral can be obtained through the limiting procedure that one might expect. Let $f : [0,T] \to \mathbb{R}$ be left-continuous and $0 = t_o^n < t_1^n < \cdots < t_{p_n}^n = T$ be a sequence of partitions of $[0,T]$ with mesh tending to zero. Then

$$\int_0^T f(s)da(s) = \lim_{n\to\infty} \sum_{i=1}^{p_n} f(t_{i-1}^n)\big(a(t_i^n) - a(t_{i-1}^n)\big).$$

The proof is easy: let $f_n : [0,T] \to \mathbb{R}$ be defined by $f_n(s) = f(t_{i-1}^n)$ if $s \in (t_{i-1}^n, t_i^n]$, $1 \le i \le p_n$, and $f_n(0) = 0$. Then

$$\sum_{i=1}^{p_n} f(t_{i-1}^n)\big(a(t_i^n) - a(t_{i-1}^n)\big) = \int_{[0,T]} f_n(s)\mu(ds),$$

where μ is the signed measure associated with a. The desired result now follows by the Dominated Convergence Theorem.

In the argument above, f_n took the value of f at the *left* endpoint of each interval. In the finite variation case, we could equally have approximated by f_n taking the value of f at the midpoint of the interval, or the right hand endpoint, or any other point in between, but the limits could differ if a were not continuous.

Proposition 7.6 (Associativity)**.** *Let a be of finite variation as above and f,g measurable functions, f is a-integrable and g is $(f \cdot a)$-integrable. Then gf is a-integrable and*

$$\int_0^t g(s)\mathrm{d}(f\cdot a)(s) = \int_0^t g(s)f(s)\mathrm{d}a(s).$$

In our 'dot'-notation:

$$g\cdot(f\cdot a) = (gf)\cdot a. \tag{15}$$

Proposition 7.7 (Stopping)**.** *Let a be of finite variation as above and fix $t \geq 0$. Set $a^t(s) = a(t\wedge s)$. Then a^t is of finite variation and for any measurable a-integrable function f*

$$\int_0^{u\wedge t} f(s)\mathrm{d}a(s) = \int_0^u f(s)\mathrm{d}a^t(s) = \int_0^u f(s)\mathbf{1}_{[0,t]}(s)\mathrm{d}a(s), \quad u\in[0,\infty].$$

Proposition 7.8 (Integration by parts)**.** *Let a and b be two right-continuous functions of finite variation with $a(0) = b(0) = 0$. Then for any t*

$$a(t)b(t) = \int_0^t a(s-)\mathrm{d}b(s) + \int_0^t b(s-)\mathrm{d}a(s) + \sum_{s\in[0,t]}\Delta a(s)\Delta b(s)$$

where $\Delta a(t) = a(t) - a(t-)$ and $a(t-) = \lim_{s\uparrow t} a(s)$.

Remark 7.9. *As a and b are right-continuous they have at most countably many discontinuities, and as they are of finite variation, the left-limits exist.*

Sketch. For a partition π_n, take a telescoping sum

$$\begin{aligned} a(t)b(t) &= \sum_{t_i\in\pi_n}(a(t_i)b(t_i) - a(t_{i-1})b(t_{i-1})) \\ &= \sum_{t_i\in\pi_n} a(t_{i-1})\big(b(t_i) - b(t_{i-1})\big) + \sum_{t_i\in\pi_n} b(t_{i-1})\big(a(t_i) - a(t_{i-1})\big) \\ &\quad + \sum_{t_i\in\pi_n}\big(a(t_i) - a(t_{i-1})\big)\big(b(t_i) - b(t_{i-1})\big). \end{aligned}$$

By dominated convergence, these converge to the stated integrals. □

Proposition 7.10 (Chain-rule)**.** *If F is a C^1 function and a is continuous of finite variation, then $F(a(t))$ is also of finite variation and*

$$F(a(t)) = F(a(0)) + \int_0^t F'(a(s))\mathrm{d}a(s).$$

Proof. The statement is trivially true for $F(x) = x$. Now by Proposition 7.8, it is straightforward to check that if the statement is true for F, then it is also true for $xF(x)$. Hence, by induction, the statement holds for all polynomials. To complete the proof, approximate $F\in C^1$ by a sequence of polynomials. □

Proposition 7.11 (Change of variables)**.** *If a is non-decreasing and right-continuous then so is its 'right inverse'*

$$c(s) := \inf\{t \geq 0 : a(t) > s\},$$

where $\inf\emptyset = +\infty$*. Let* $a(0) = 0$*. Then, for any Borel measurable function* $f \geq 0$ *on* $\mathbb{R}_+$*, we have*

$$\int_0^\infty f(u)\mathrm{d}a(u) = \int_0^{a(\infty)} f(c(s))\mathrm{d}s.$$

Proof. If $f(u) = \mathbf{1}_{[0,v]}(u)$, then the claim becomes

$$a(v) = \int_0^\infty \mathbf{1}_{\{c(s)\leq v\}} ds = \inf\{s : c(s) > v\},$$

and equality holds by definition of c. Take differences to get indicators of sets $(u, v]$. The Monotone Class Theorem allows us to extend to functions of compact support and then take increasing limits to obtain the formula in general. □

7.3 Processes of finite variation

Recall that a filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions is given.

Definition 7.12. *An adapted right-continuous process* $A = (A_t : t \geq 0)$ *is called a* finite variation process *(or a* process of finite variation*) if* $A_0 = 0$ *and* $t \mapsto A_t$ *is (a function) of finite variation a.s..*

Proposition 7.13. *Let A be a finite variation process and K a progressively measurable process s.t.*

$$\forall t \geq 0,\ \forall \omega \in \Omega,\ \int_0^t |K_s(\omega)||\mathrm{d}A_s(\omega)| < \infty.$$

Then $((K \cdot A)_t : t \geq 0)$*, defined as* $(K \cdot A)_t(\omega) := \int_0^t K_s(\omega)\mathrm{d}A_s(\omega)$*, is a finite variation process.*

Proof. The right continuity is immediate from the deterministic theory, but we need to check that $(K \cdot A)_t$ is *adapted* (and hence progressive, by Proposition 4.5). For this we check that if $t > 0$ is fixed and $h : [0,t] \times \Omega \to \mathbb{R}$ is measurable with respect to $\mathscr{B}([0,t]) \otimes \mathscr{F}_t$, and if

$$\int_0^t |h(s,\omega)||dA_s(\omega)| < \infty$$

for every $\omega \in \Omega$, then

$$\int_0^t h(s,\omega)dA_s(\omega)$$

is $\mathscr{F}_t$-measurable.

Fix $t > 0$. Consider first h defined by $h(s,\omega) = \mathbf{1}_{(u,v]}(s)\mathbf{1}_\Gamma(\omega)$ for $(u,v] \subseteq [0,t]$ and $\Gamma \in \mathscr{F}_t$. Then

$$(h \cdot A)_t = \mathbf{1}_\Gamma(A_v - A_u)$$

is $\mathscr{F}_t$-measurable. By the Monotone Class Theorem, $(h \cdot A)_t$ is $\mathscr{F}_t$-measurable for any $h = \mathbf{1}_G$ with $G \in \mathscr{B}([0,t]) \otimes \mathscr{F}_t$, or, more generally, any bounded $\mathscr{B}([0,t]) \otimes \mathscr{F}_t$-measurable function h. If h is a general $\mathscr{B}([0,t]) \otimes \mathscr{F}_t$-measurable function satisfying

$$\int_0^t |h(s,\omega)||\mathrm{d}A_s(\omega)| < \infty \quad \forall \omega \in \Omega,$$

then h is a pointwise limit, $h = \lim_{n\to\infty} h_n$, of simple functions with $|h| \geq |h_n|$. The integrals $\int h_n(s,\omega)dA_s(\omega)$ converge by the Dominated Convergence Theorem, and hence $\int_0^t h(s,\omega)\mathrm{d}A_s(\omega)$ is also $\mathscr{F}_t$-measurable (as a limit of $\mathscr{F}_t$-measurable functions). In particular, $(K \cdot A)_t(\omega)$ is $\mathscr{F}_t$-measurable since by progressive measurability, $(s,\omega) \mapsto K_s(\omega)$ on $[0,t]$ is $\mathscr{B}([0,t]) \otimes \mathscr{F}_t$-measurable. □

7.4 Continuous local martingales

We now want to extend our integration theory to processes which are *not* of finite variation. The processes that make our theory work are slight generalisations of martingales.

Definition 7.14. *An adapted process $(M_t : t \geq 0)$ is called a continuous* local *martingale if it has continuous trajectories a.s. and if there exists a non-decreasing sequence of stopping times $(\tau_n)_{n\geq 1}$ such that $\tau_n \uparrow \infty$ a.s. and for each n, $M^{\tau_n} = (M_{t\wedge\tau_n} : t \geq 0)$ is a (wlog uniformly integrable) martingale. We say (τ_n)* reduces *or* localizes *M.*

Any martingale is a local martingale, but the converse is false.

Example 7.15. *Let ξ be a random variable not in L^1, and Z be an independent Bernoulli random variable with $p = 1/2$. Define a filtration*

$$\mathscr{F}_t = \begin{cases} \{\emptyset, \Omega\} & t < 1 \\ \sigma(\xi) & t \in [1,2) \\ \sigma(\xi, Z) & t \geq 2 \end{cases}$$

and a process

$$X_t = \begin{cases} 0 & t < 2 \\ \xi Z & t \geq 2 \end{cases}$$

By taking the stopping times $\tau_n = n\mathbf{1}_{\{|\xi|<n\}}$, we see that X is a local martingale, but cannot be a martingale as $E[|X_2|] \not< \infty$.

Example 7.16. *Let B be a Brownian motion, and ξ an independent nonnegative random variable not in L^1. Then define $X_t = B_{\xi^2 t}$, in the filtration $\{\mathscr{F}^X_{t+}\}_{t\geq 0}$. Then*

$$\mathbb{E}[|X_t|] = \mathbb{E}[|B_{\xi^2 t}|] = \mathbb{E}\Big[\mathbb{E}\Big[|B_{\xi^2 t}|\Big|\xi\Big]\Big] = \sqrt{2t/\pi}\mathbb{E}[\xi] = \infty$$

so X is not a martingale. However, ξ is $\mathscr{F}^X_0$-measurable (we will see this from the fact $\langle X\rangle_t = \xi^2 t$ and right-continuity), so we can use the stopping times $\tau_n = n\mathbf{1}_{\{\xi<n\}}$ to localize and hence verify X^{τ_n} is a martingale. As X is continuous, we can also localize with $\tau_n = \inf\{t : |X_t| \geq n\}$.

There are other more subtle examples of local martingales, including cases where $\{X_t\}_{t\in\mathbb{R}}$ is uniformly integrable, which are not martingales.

Proposition 7.17. *i. A non-negative continuous local martingale such that $M_0 \in L^1$ is a supermartingale.*

ii. A continuous local martingale M such that there exists a random variable $Z \in L^1$ with $|M_t| \leq Z$ for every $t \geq 0$ is a uniformly integrable martingale.

iii. If M is a continuous local martingale and $M_0 = 0$ (or more generally $M_0 \in L^1$), the sequence of stopping times

$$T_n = \inf\{t \geq 0 : |M_t| \geq n\}$$

reduces M.

iv. If M is a continuous local martingale, then for any stopping time ρ, the stopped process M^ρ is also a continuous local martingale.

Proof. (i) Write $M_t = M_0 + N_t$. By definition, there exists a sequence T_n of stopping times that reduces N. Thus, if $s \leq t$, for every n,

$$N_{s\wedge T_n} = \mathbb{E}[N_{t\wedge T_n}|\mathscr{F}_s].$$

We can add M_0 to both sides (M_0 is $\mathscr{F}_s$-measurable and in L^1) and we find

$$M_{s\wedge T_n} = \mathbb{E}[M_{t\wedge T_n}|\mathscr{F}_s].$$

Since M takes non-negative values, let $n \to \infty$ and apply Fatou's lemma for conditional expectations to find

$$M_s \geq \mathbb{E}[M_t|\mathscr{F}_s]. \tag{16}$$

Taking $s = 0$, $\mathbb{E}[M_t] \leq \mathbb{E}[M_0] < \infty$. So $M_t \in L^1$ for every $t \geq 0$, and (16) says that M is a supermartingale.

(ii) By the same argument,

$$M_{s\wedge T_n} = \mathbb{E}[M_{t\wedge T_n}|\mathscr{F}_s].$$

Since $|M_{t\wedge T_n}| \leq Z$, this time apply the Dominated Convergence Theorem to see that $M_{t\wedge T_n}$ converges in L^1 (to M_t) and $M_s = \mathbb{E}[M_t|\mathscr{F}_s]$.

The other two statements are immediate. □

Theorem 7.18. *A continuous local martingale M with $M_0 = 0$ a.s., is a process of finite variation if and only if M is indistinguishable from zero.*

Remark 7.19. *Continuity* is *critical here.*

Proof. Assume M is a continuous local martingale and of finite variation. Let

$$\tau_n = \inf\{t \geq 0 : \int_0^t |dM_s| \geq n\} = \inf\{t \geq 0 : V(M)_t \geq n\},$$

which are stopping times since $V(M)_t = \int_0^t |dM_s|$ is continuous and adapted.
Let $N = M^{\tau_n}$, which is bounded since

$$|N_t| = |M_{t\wedge\tau_n}| \leq |\int_0^{t\wedge\tau_n} dM_u| \leq \int_0^{t\wedge\tau_n} |dM_u| \leq n,$$

and hence (N_t) is a martingale.
Let $t > 0$ and $\pi = \{0 = t_0 < t_1 < t_2 < \ldots < t_{N(\pi)} = t\}$ be a partition of $[0,t]$. Then

$$\begin{aligned}
\mathbb{E}[N_t^2] &= \sum_{i=1}^{N(\pi)} \mathbb{E}[N_{t_i}^2 - N_{t_{i-1}}^2] = \sum_{i=1}^{N(\pi)} \mathbb{E}\big[(N_{t_i} - N_{t_{i-1}})^2\big] \\
&\leq \mathbb{E}\Big[\big(\sup_{1\leq i\leq N(\pi)} |N_{t_i} - N_{t_{i-1}}|\big) \cdot \underbrace{\sum |N_{t_i} - N_{t_{i-1}}|}_{\leq V(N)_t = V(M)_{t\wedge\tau_n} \leq n}\Big] \\
&\leq n\mathbb{E}\big[\sup_{1\leq i\leq N(\pi)} |N_{t_i} - N_{t_{i-1}}|\big] \to 0 \quad \text{as } \|\pi\| \to 0
\end{aligned}$$

(where $\|\pi\|$ is the mesh of π), by the Dominated Convergence Theorem (since $|N_{t_i} - N_{t_{i-1}}| \leq V(N)_t \leq n$ and so n is a dominating function).
It then follows by Fatou's Lemma that

$$\mathbb{E}[M_t^2] = \mathbb{E}[\lim_{n\to\infty} M_{t\wedge\tau_n}^2] \leq \lim_{n\to\infty} \mathbb{E}[M_{t\wedge\tau_n}^2] = 0$$

which implies that $M_t = 0$ a.s., and so by continuity of paths, $\mathbb{P}[M_t = 0 \, \forall t \geq 0] = 1$. □

7.5 Quadratic variation of a continuous local martingale

If our martingales are going to be interesting, then they're going to have unbounded variation. But remember that we said that we would use Brownian motion as a basic building block, and that while Brownian motion has infinite variation, it has bounded *quadratic variation*, defined over $[0,T]$ by the limit (in probability)

$$\lim_{\|\pi_n\|\to 0} \sum_{j=1}^{N(\pi_n)} \left(B_{t_j} - B_{t_{j-1}}\right)^2 = T.$$

where $\{\pi_n\}$ is a sequence of partitions.

We are now going to see that the analogue of this process exists for any continuous local martingale. Ultimately, we shall see that the quadratic variation is in some sense a 'clock' for a local martingale, but that will be made more precise in the very last result of the course.

Theorem 7.20. *Let M be a continuous local martingale. There exists a unique (up to indistinguishability) non-decreasing, continuous adapted finite variation process $(\langle M,M\rangle_t : t \geq 0)$, starting in zero, such that $(M_t^2 - \langle M,M\rangle_t : t \geq 0)$ is a continuous local martingale.*

Furthermore, for any $T > 0$ and any sequence of partitions $\pi_n = \{0 = t_0^n < t_1^n < \ldots < t_{N(\pi_n)}^n = T\}$ with $\|\pi_n\| = \sup_{1\leq i\leq N(\pi_n)}(t_i^n - t_{i-1}^n) \to 0$ as $n \to \infty$

$$\langle M,M\rangle_T = \lim_{n\to\infty} \sum_{i=1}^{N(\pi_n)} (M_{t_i^n} - M_{t_{i-1}^n})^2, \tag{17}$$

where the limit is in probability.

The process $\langle M,M\rangle$ is called the quadratic variation of *M, or simply* the increasing process of *M, and is often denoted $\langle M,M\rangle_t = \langle M\rangle_t$.*

Sketch of a Proof (NOT EXAMINABLE). Uniqueness is a direct consequence of Theorem 7.18 since if A, A' are two such processes then $\big(M^2 - A - (M^2 - A')\big) = A - A'$ is a local martingale starting in zero and of finite variation, which implies $A = A'$ by Theorem 7.18.

The idea of existence is as follows. First suppose that M is *bounded*. Take a sequence of partitions $\pi_n = \{0 = t_0^n < \cdots < t_{N(\pi_n)}^n = T\}$ with mesh tending to zero. Then check that

$$X_t^n := \sum_{i=1}^{N(\pi_n)} M_{t_{i-1}^n}(M_{t_i^n\wedge t} - M_{t_{i-1}^n\wedge t})$$

is a (bounded) martingale. Now observe that

$$M_{t_j^n}^2 - 2X_{t_j^n}^n = \sum_{i=1}^{j}(M_{t_i^n} - M_{t_{i-1}^n})^2.$$

A direct computation gives

$$\lim_{n,m\to\infty} \mathbb{E}[(X_t^n - X_t^m)^2] = 0,$$

and by Doob's L^2-inequality

$$\lim_{n,m\to\infty} \mathbb{E}[\sup_{t\leq T}(X_t^n - X_t^m)^2] = 0.$$

By passing to a subsequence, $X^{n_k} \to Y$ almost surely on $[0,T]$ where $(Y_t)_{t\leq T}$ is a continuous process which inherits the martingale property from X.

$$M_{t_j^n}^2 - 2X_{t_j^n}^n = \sum_{i=1}^{j}(M_{t_i^n} - M_{t_{i-1}^n})^2 =: QV_{t_j^n}^{\pi_n}(M)$$

is non-decreasing along $t_j^n : j \leq N(\pi_n)$. Letting $n \to \infty$, $M_t^2 - 2Y_t$ is almost surely non-decreasing and we set $\langle M, M \rangle_t = M_t^2 - 2Y_t$.

To move to a general continuous local martingale, we consider a sequence of stopped processes.

Details are in, for example, Le Gall's book. □

7.6 $\mathscr{H}^2$ space

Our theory of integration is going to be an 'L^2-theory'. Let us introduce the martingales with which we are going to work. We are going to think of them as being defined up to indistinguishability – nothing changes if we change the process on a null set. Think of this as analogous to considering Lebesgue integrable functions as being defined 'almost everywhere'.

Definition 7.21. *Let $\mathscr{H}^2$ be the space of* L^2*-bounded càdlàg martingales, i.e.*

$$(\{\mathscr{F}_t\}_{t \geq 0}, \mathbb{P})\text{–martingales } M \text{ s.t. } \sup_{t \geq 0} \mathbb{E}[M_t^2] < \infty,$$

and $\mathscr{H}^{2,c}$ the subspace consisting of continuous L^2*-bounded martingales. Finally, let $\mathscr{H}_0^{2,c} = \{M \in \mathscr{H}^{2,c} : M_0 = 0 \text{ a.s.}\}$.*

We note that the space $\mathscr{H}^2$ is also sometimes denoted $\mathscr{M}^2$.

It follows from Doob's L^2-inequality that

$$\mathbb{E}\left[\sup_{t \geq 0} M_t^2\right] \leq 4 \sup_{t \geq 0} \mathbb{E}[M_t^2] < +\infty, \quad M \in \mathscr{H}^2.$$

Consequently, $\{M_t : t \geq 0\}$ is bounded by a square integrable random variable $(\sup_{t \geq 0} |M_t|)$ and in particular is uniformly integrable. It follows from the martingale convergence theorem that $M_t = \mathbb{E}[M_\infty | \mathscr{F}_t]$ for some square integrable random variable M_∞.

Conversely, we can start with a random variable $Y \in \mathrm{L}^2(\Omega, \mathscr{F}_\infty, \mathbb{P})$ and define a martingale $M_t := \mathbb{E}[Y | \mathscr{F}_t] \in \mathscr{H}^2$ (and $M_\infty = Y$).

Two L^2-bounded martingales M, M' are indistinguishable if and only if $M_\infty = M'_\infty$ a.s. and so if we endow $\mathscr{H}^2$ with the norm

$$\|M\|_{\mathscr{H}^2} := \sqrt{\mathbb{E}[M_\infty^2]} = \|M_\infty\|_{\mathrm{L}^2(\Omega, \mathscr{F}_\infty, \mathbb{P})}, \quad M \in \mathscr{H}^2, \tag{18}$$

then $\mathscr{H}^2$ can be identified with the familiar $\mathrm{L}^2(\Omega, \mathscr{F}_\infty, \mathbb{P})$ space.

Remark 7.22. *It's worth noticing that $\|M\|_{\mathscr{H}^2}^2 = \mathbb{E}[M_\infty^2] = \mathrm{var}(M_\infty)$ (recalling that $\mathbb{E}[M_\infty] = 0$).*

Theorem 7.23. *$\mathscr{H}^{2,c}$ is a closed subspace of $\mathscr{H}^2$.*

Proof. This is almost a matter of writing down definitions. Suppose that the sequence $M^n \in \mathscr{H}^{2,c}$ converges in $\|\cdot\|_{\mathscr{H}^2}$ to some $M \in \mathscr{H}^2$. By Doob's L^2-inequality

$$\mathbb{E}\left[\sup_{t\geq 0} |M_t^n - M_t|^2\right] \leq 4\|M^n - M\|^2_{\mathscr{H}^2} \longrightarrow 0, \quad \text{as } n \to \infty.$$

Passing to a subsequence, we have $\sup_{t\geq 0} |M_t^{n_k} - M_t| \to 0$ a.s. and hence M has continuous paths a.s.,which completes the proof. $\square$

For continuous local martingales, the norm in (18) can be re-expressed in terms of the quadratic variation:

Theorem 7.24. *Let M be a continuous local martingale with $M_0 \in L^2$.*

i. TFAE

(a) M is a martingale, bounded in L^2;

(b) $\mathbb{E}[\langle M,M\rangle_\infty] < \infty$.

Furthermore, if these properties hold, $M_t^2 - \langle M,M\rangle_t$ is a uniformly integrable martingale and, in particular, $\mathbb{E}[M_\infty^2] = \mathbb{E}[M_0^2] + \mathbb{E}[\langle M,M\rangle_\infty]$.

ii. TFAE

(a) M is a martingale and $M_t \in L^2$ for every $t \geq 0$;

(b) $\mathbb{E}[\langle M,M\rangle_t] < \infty$ for every $t \geq 0$.

Furthermore, if these properties hold, $M_t^2 - \langle M,M\rangle_t$ is a martingale.

Proof. The second statement will follow from the first on applying it to $M_{t\wedge a}$ for every choice of $a \geq 0$.

To prove the first set of equivalences, without loss of generality, suppose that $M_0 = 0$ (or replace M by $M - M_0$).

Suppose that M is a martingale, bounded in L^2. Doob's L^2-inequality implies that for every $T > 0$,

$$\mathbb{E}[\sup_{0\leq t\leq T} M_t^2] \leq 4\mathbb{E}[M_T^2],$$

and so, letting $T \to \infty$,

$$\mathbb{E}[\sup_{t\geq 0} M_t^2] \leq 4 \sup_{t\geq 0} \mathbb{E}[M_t^2] = C < \infty.$$

Let $S_n = \inf\{t \geq 0 : \langle M,M\rangle_t \geq n\}$. Then the continuous local martingale $M_{t\wedge S_n}^2 - \langle M,M\rangle_{t\wedge S_n}$ is dominated by $\sup_{s\geq 0} M_s^2 + n$, which is integrable. By Proposition 7.17 this continuous local martingale is a uniformly integrable martingale, so $\mathbb{E}[M_{t\wedge S_n}^2 - \langle M\rangle_{t\wedge S_n}] = 0$, and hence

$$\mathbb{E}[\langle M,M\rangle_{t\wedge S_n}] = \mathbb{E}[M_{t\wedge S_n}^2] \leq \mathbb{E}[\sup_{s\geq 0} M_s^2] \leq C < \infty.$$

Let n and then t tend to infinity and use the Monotone Convergence Theorem to obtain $\mathbb{E}[\langle M,M\rangle_\infty] < \infty$.

Conversely, assume that $\mathbb{E}[\langle M,M\rangle_\infty] < \infty$. Set $T_n = \inf\{t \geq 0 : |M_t| \geq n\}$. Then the continuous local martingale $M^2_{t\wedge T_n} - \langle M,M\rangle_{t\wedge T_n}$ is dominated by $n^2 + \langle M,M\rangle_\infty$ which is integrable. From Proposition 7.17 again, this continuous local martingale is a uniformly integrable martingale and hence for every $t \geq 0$,

$$\mathbb{E}[M^2_{t\wedge T_n}] = \mathbb{E}[\langle M,M\rangle_{t\wedge T_n}] \leq \mathbb{E}[\langle M,M\rangle_\infty] = C' < \infty. \tag{19}$$

Let $n \to \infty$ and use Fatou's lemma to see that $\mathbb{E}[M_t^2] \leq C' < \infty$, so $(M_t)_{t\geq 0}$ is bounded in L^2.

We still have to check that $(M_t)_{t\geq 0}$ is a martingale. However, (19) shows that $(M_{t\wedge T_n})_{n\geq 1}$ is uniformly integrable and so converges both almost surely and in L^1 to M_t for every $t \geq 0$. Recalling that M^{T_n} is a martingale, L^1 convergence implies, for $s > t$,

$$M_t = \lim_n M_t^{T_n} = \lim_n \mathbb{E}[M_s^{T_n}|\mathscr{F}_t] = \mathbb{E}[\lim_n M_s^{T_n}|\mathscr{F}_t] = \mathbb{E}[M_s|\mathscr{F}_t]$$

so M is a martingale.

Finally, if the two properties hold, then $M^2 - \langle M,M\rangle$ is dominated by $\sup_{t\geq 0} M_t^2 + \langle M,M\rangle_\infty$, which is integrable, and so Proposition 7.17 again says that $M^2 - \langle M,M\rangle$ is a uniformly integrable martingale. □

Our previous theorem immediately yields that for a local martingale M with $M_0 = 0$, if $\mathbb{E}[\langle M\rangle_\infty] < \infty$ then

$$\|M\|^2_{\mathscr{H}^2} = \mathbb{E}[M^2_\infty] = \mathbb{E}[\langle M\rangle_\infty].$$

We can also deduce a complement to Theorem 7.18.

Corollary 7.25. *Let M be a continuous local martingale with $M_0 = 0$. Then the following are equivalent:*

i. M is indistinguishable from zero,

ii. $\langle M\rangle_t = 0$ for all $t \geq 0$ a.s.,

iii. M is a process of finite variation.

Proof. We already know that the first and third statements are equivalent. That the first implies the second is trivial, so we must just show that the second implies the first. We have $\langle M\rangle_\infty = \lim_{t\to\infty}\langle M\rangle_t = 0$. From Theorem 7.24, $M \in \mathscr{H}^2$ and $\mathbb{E}[M^2_\infty] = \mathbb{E}[\langle M\rangle_\infty] = 0$ and so $M_t = \mathbb{E}[M_\infty|\mathscr{F}_t] = 0$ almost surely. □

7.7 Quadratic covariation

We can see that the quadratic variation of a martingale is telling us something about how its variance increases with time. We also need an analogous quantity for the 'covariance' between two martingales. This is most easily defined through *polarisation*.

Definition 7.26. *The quadratic* co-variation *between two continuous local martingales M,N is defined by*

$$\langle M,N\rangle := \frac{1}{2}\left(\langle M+N,M+N\rangle - \langle M,M\rangle - \langle N,N\rangle\right). \tag{20}$$

It is often called the (angle) bracket process of M and N.

Proposition 7.27. *For two continuous local martingales M,N*

i. *the process $\langle M,N\rangle$ is the unique finite variation process, zero at zero, such that $(M_tN_t - \langle M,N\rangle_t : t \geq 0)$ is a continuous local martingale;*

ii. *the mapping $M,N \mapsto \langle M,N\rangle$ is bilinear and symmetric;*

iii. *for any stopping time τ,*

$$\langle M^\tau,N^\tau\rangle_t = \langle M^\tau,N\rangle_t = \langle M,N^\tau\rangle_t = \langle M,N\rangle_{\tau\wedge t}, \quad t \geq 0,\ a.s.; \tag{21}$$

iv. *for any $t > 0$ and a sequence of partitions π_n of $[0,t]$ with mesh converging to zero*

$$\sum_{t_i\in\pi_n}(M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}) \to \langle M,N\rangle_t, \tag{22}$$

the convergence being in probability.

Proof. (i) $(M+N)_t^2 - \langle M+N,M+N\rangle_t$ is a continuous local martingale and by adding and subtracting terms it is equal to

$$\underbrace{M_t^2 - \langle M,M\rangle_t}_{\text{l.mat}} + \underbrace{N_t^2 - \langle N,N\rangle_t}_{\text{l.mat}} + 2\underbrace{\left(M_tN_t - \frac{1}{2}(\langle M+N,M+N\rangle_t - \langle M,M\rangle_t - \langle N,N\rangle_t)\right)}_{\text{hence also a l.mat}}$$

Uniqueness follows from Theorem 7.18.

(iv) Note that

$$(M_t + N_t - M_s - N_s)^2 - (M_t - M_s)^2 - (N_t - N_s)^2 = 2(M_t - M_s)(N_t - N_s).$$

The asserted convergence then follows from Theorem 7.20

(ii) Both properties follow from (iv). Symmetry is obvious from the definition in (20).

(iii) Follows from (iv). □

Definition 7.28. *Two continuous local martingales M, N, are said to be* (very strongly) orthogonal *if* $\langle M,N\rangle = 0$.

For example, if B and B' are independent Brownian motions, then $\langle B,B'\rangle = 0$.

Remark 7.29. *It follows that if M and N are two martingales bounded in L^2 and with $M_0N_0 = 0$ a.s., then $(M_tN_t - \langle M,N\rangle_t, t \geq 0)$ is a uniformly integrable martingale. In particular, for* every *stopping time τ,*

$$\mathbb{E}[M_\tau N_\tau] = \mathbb{E}[\langle M,N\rangle_\tau]. \tag{23}$$

Remark 7.30. *Note that $\langle M,N\rangle = 0$ is a stronger statement than $\mathbb{E}[M_\infty N_\infty] = \mathbb{E}[\langle M,N\rangle_\infty] = 0$. For example, consider W a Brownian motion and $N = \xi W$, for ξ independent of W with mean zero ($\mathscr{F}_0$-measurable). Then $\langle W,N\rangle = \xi t \neq 0$, but $\mathbb{E}[\langle W,N\rangle_\infty] = 0$.*

Take $M,N \in \mathscr{H}_0^{2,c}$, which we recall had the norm $\|M\|^2_{\mathscr{H}^{2,c}} = \mathbb{E}[\langle M,M\rangle_\infty] = \mathbb{E}[M_\infty^2]$. Then we see that this norm is consistent with the inner product on $\mathscr{H}^{2,c} \times \mathscr{H}^{2,c}$ given by $\mathbb{E}[\langle M,N\rangle_\infty] = \mathbb{E}[M_\infty N_\infty]$ and, by the usual Cauchy–Schwarz inequality,

$$\mathbb{E}[|M_\infty N_\infty|] \leq \sqrt{\mathbb{E}[\langle M\rangle_\infty]\mathbb{E}[\langle N\rangle_\infty]}.$$

It is easy to obtain an almost sure result also, using that

$$\left|\sum \left(M_{t_{i+1}} - M_{t_i}\right)\left(N_{t_{i+1}} - N_{t_i}\right)\right| \leq \sqrt{\sum \left(M_{t_{i+1}} - M_{t_i}\right)^2}\sqrt{\sum \left(N_{t_{i+1}} - N_{t_i}\right)^2}$$

and taking limits to deduce that

$$|\langle M,N\rangle_t| \leq \sqrt{\langle M\rangle_t}\sqrt{\langle N\rangle_t}.$$

It's often convenient to have a more general version of this inequality.

Theorem 7.31 (Kunita–Watanabe inequality)**.** *Let M,N be continuous local martingales and K,H two measurable processes. Then for all $0 \leq t \leq \infty$,*

$$\int_0^t |H_s||K_s||\mathrm{d}\langle M,N\rangle_s| \leq \left(\int_0^t H_s^2 \mathrm{d}\langle M\rangle_s\right)^{1/2} \left(\int_0^t K_s^2 \mathrm{d}\langle N\rangle_s\right)^{1/2} \ a.s.. \tag{24}$$

We omit the proof which approximates H, K by simple functions and then essentially uses the Cauchy–Schwarz inequality for sums noted above.

7.8 Continuous semimartingales

Definition 7.32. *A stochastic process $X = (X_t : t \geq 0)$ is called a continuous semimartingale if it can be written as*

$$X_t = X_0 + M_t + A_t, \quad t \geq 0 \tag{25}$$

where M is a continuous local martingale, A is a continuous process of finite variation, and $M_0 = A_0 = 0$ a.s..

The decomposition is unique (up to indistinguishability). It should be remembered that there is a filtration $\{\mathscr{F}_t\}_{t\geq 0}$ and a probability measure $\mathbb{P}$ implicit in our definition.

Proposition 7.33. *A continuous semimartingale is of finite quadratic variation and in the notation above* $\langle X,X\rangle = \langle M,M\rangle$.

Proof. Fix $t \geq 0$ and consider a sequence of partitions of $[0,t]$, $\pi_m = \{0 = t_0 < t_1 < \ldots < t_{n_m} = t\}$ with $\|\pi_m\| \to 0$ as $m \to \infty$. Then

$$\sum_{i=1}^{n_m}(X_{t_i} - X_{t_{i-1}})^2 = \underbrace{\sum_{i=1}^{n_m}(M_{t_i} - M_{t_{i-1}})^2}_{(i)} + \underbrace{\sum_{i=1}^{n_m}(A_{t_i} - A_{t_{i-1}})^2}_{(ii)} + 2\underbrace{\sum_{i=1}^{n_m}(M_{t_i} - M_{t_{i-1}})(A_{t_i} - A_{t_{i-1}})}_{(iii)}.$$

It follows from the properties of M and A that, as $m \to \infty$,

$$\begin{aligned}
&(i) \to \langle M,M\rangle_t, \\
&(ii) \leq \sup_{1\leq i\leq n_m} |A_{t_i} - A_{t_{i-1}}| \cdot V_t(A) \to 0 \quad \text{a.s. ,} \\
&(iii) \leq \sup_{1\leq i\leq n_m} |M_{t_i} - M_{t_{i-1}}| \cdot V_t(A) \to 0 \quad \text{a.s. .}
\end{aligned}$$

□

If X,Y are two continuous semimartingales, we can define their co-variation $\langle X,Y\rangle$ via the polarisation formula that we used for martingales. If $X_t = X_0 + M_t + A_t$ and $Y_t = Y_0 + N_t + A'_t$, then $\langle X,Y\rangle_t = \langle M,N\rangle_t$.

8 Stochastic Integration

At the beginning of the course we argued that whereas classically differential equations take the form

$$dX(t) = a(t,X(t))dt,$$

in many settings, the dynamics of the physical quantity in which we are interested may also have a random component and so perhaps take the form

$$dX_t = a(t,X_t)dt + b(t,X_t)dB_t.$$

We actually understand equations like this in the integral form:

$$X_t - X_0 = \int_0^t a(s,X_s)ds + \int_0^t b(s,X_s)dB_s.$$

If a is nice enough, then the first term has a classical interpretation. It is the second term, or rather a generalisation of it, that we want to make sense of now.

The first approach will be to mimic what we usually do for construction of the Lebesgue integral, namely work out how to integrate simple functions and then extend to general functions through passage to the limit. We will then provide a very slick, but not at all intuitive, approach that nonetheless gives us some 'quick wins' in proving properties of the integral.

8.1 Stochastic integral w.r.t. L^2-bounded martingales

Remark on Notation: We are going to use the notation $\varphi \bullet M$ for the (Itô) stochastic integral of φ with respect to M. This is not universally accepted notation; many authors would write $\int_0^t \varphi_s dM_s$ for $(\varphi \bullet M)_t$. Moreover, for emphasis, when the integrator is stochastic, we have used '$\bullet$' in place of the '$\cdot$' that we used for the Stieltjes integral.

8.1.1 Riemann sums and simple integrals

We are going to develop a theory of integration w.r.t. martingales in $\mathscr{H}^{2,c}$. Recall that $\mathscr{H}_0^{2,c}$ is the space of continuous martingales M, zero at zero, which are bounded in L^2. It is a Hilbert space with the inner product $\langle M,N\rangle_{\mathscr{H}^{2,c}} = \mathbb{E}[M_\infty N_\infty]$ and induced norm

$$\|M\|_{\mathscr{H}^{2,c}} = \sqrt{\mathbb{E}[M_\infty^2]} = \sqrt{\mathbb{E}[\langle M\rangle_\infty]}.$$

(In a very real sense we are identifying $\mathscr{H}^{2,c}$ with L^2.)

Define $\mathscr{E}$ to be the space of simple bounded processes of the form

$$\varphi_t = \sum_{i=0}^{m} \varphi^{(i)} \mathbf{1}_{(t_i,t_{i+1}]}(t), \quad t \geq 0, \tag{26}$$

for some $m \in \mathbb{N}$, $0 \leq t_0 < t_1 < \ldots < t_{m+1}$ and where $\varphi^{(i)}$ are bounded $\mathscr{F}_{t_i}$-measurable random variables. Define the stochastic integral $\varphi \bullet M$ of φ in (26) with respect to $M \in \mathscr{H}^{2,c}$ via

$$(\varphi \bullet M)_t := \sum_{i=0}^{m} \varphi^{(i)} (M_{t\wedge t_{i+1}} - M_{t\wedge t_i}), \quad t \geq 0. \tag{27}$$

If we write $M_t^i := \varphi^{(i)}(M_{t\wedge t_{i+1}} - M_{t\wedge t_i})$ then clearly $M^i \in \mathscr{H}^{2,c}$ and so $\varphi \bullet M$ is a martingale. Moreover, since for $i \neq j$ the intervals $(t_i,t_{i+1}]$ and $(t_j,t_{j+1}]$ are disjoint, $M_t^i M_t^j$ is a martingale and hence $\langle M^i, M^j\rangle_t = 0$. Using the bilinearity of the bracket process then yields

$$\langle \varphi \bullet M\rangle_t = \sum_{i=0}^{m} \langle M^i\rangle_t = \sum_{i=0}^{m} \left(\varphi^{(i)}\right)^2 \left(\langle M\rangle_{t_{i+1}\wedge t} - \langle M\rangle_{t_i\wedge t}\right) = \int_0^t \varphi_s^2 \mathrm{d}\langle M\rangle_s, \quad t \geq 0. \tag{28}$$

We already used the notation that if K is progressively measurable and A is of finite variation, then

$$(K\cdot A)_t = \int_0^t K_s(\omega) dA_s(\omega), \qquad t \geq 0.$$

In that notation

$$\langle \varphi \bullet M\rangle = \varphi^2 \cdot \langle M\rangle.$$

More generally, for $N \in \mathscr{H}^{2,c}$,

$$\begin{aligned}\langle \varphi \bullet M, N\rangle_t &= \sum_{i=0}^{m} \langle M^i, N\rangle_t = \sum_{i=0}^{m} \varphi^{(i)} \left(\langle M, N\rangle_{t_{i+1}\wedge t} - \langle M, N\rangle_{t_i \wedge t}\right) \\ &= \int_0^t \varphi_s \mathrm{d}\langle M, N\rangle_s = (\varphi \cdot \langle M, N\rangle)_t .\end{aligned} \tag{29}$$

Proposition 8.1. *Let $M \in \mathscr{H}^{2,c}$. The mapping $\varphi \mapsto \varphi \bullet M$ is a linear map from $\mathscr{E}$ to $\mathscr{H}_0^{2,c}$. Moreover,*

$$\|\varphi \bullet M\|^2_{\mathscr{H}^{2,c}} = \mathbb{E}\left[\int_0^\infty \varphi_t^2 \mathrm{d}\langle M\rangle_t\right]. \tag{30}$$

The proof is easy – we just need to show linearity. But given φ, $\psi \in \mathscr{E}$, we use a refinement of the partitions on which they are constant to write them as simple functions with respect to the same partition and the result is trivial.

Remark 8.2. *If we were considering martingales with jumps, then it would be important that the processes in $\mathscr{E}$ are left continuous.*

8.1.2 Simple integrals are dense

We are expecting an L^2-theory – we have already found an expression for the 'L^2-norm' of $\varphi \bullet M$. Let us define the appropriate spaces more carefully.

Definition 8.3. *Given $M \in \mathscr{H}^{2,c}$ we denote by $\mathrm{L}^2(M)$ the space of progressively measurable processes K such that*

$$\|K\|^2_{\mathrm{L}^2(M)} := \mathbb{E}\left[\int_0^\infty K_t^2 \mathrm{d}\langle M\rangle_t\right] < +\infty. \tag{31}$$

$L^2(M)$ is a Hilbert space, with inner product

$$H, K \mapsto \mathbb{E}\left[\int_0^\infty H_t K_t d\langle M\rangle_t\right] = \mathbb{E}\left[(HK \cdot \langle M\rangle)_\infty\right].$$

We have $\mathscr{E} \subseteq L^2(M)$ and (30) tells us that the map $\mathscr{E} \to \mathscr{H}_0^{2,c}$ given by $\varphi \mapsto \varphi \bullet M$ is a linear isometry. If we can show that the elementary functions are dense in $L^2(M)$, this observation will allow us to define integrals of functions from $L^2(M)$ with respect to M via a limiting procedure.

Proposition 8.4. *Let $M \in \mathscr{H}^{2,c}$. Then $\mathscr{E}$ is a dense vector subspace of* $\mathrm{L}^2(M)$.

Proof. It is enough to show that if $K \in \mathrm{L}^2(M)$ is orthogonal to φ for all $\varphi \in \mathscr{E}$, then $K = 0$ (as an element of $\mathrm{L}^2(M)$). So suppose that $\langle K, \varphi\rangle_{\mathrm{L}^2(M)} = 0$ for all $\varphi \in \mathscr{E}$. Let $X = K \cdot \langle M\rangle$, i.e. $X_t = \int_0^t K_u \mathrm{d}\langle M\rangle_u$. This is well defined and, by Cauchy–Schwarz

$$\mathbb{E}[|X_t|] \leq \mathbb{E}\left[\int_0^t |K_u| \mathrm{d}\langle M\rangle_u\right] \leq \sqrt{\mathbb{E}\left[\int_0^t K_u^2 \mathrm{d}\langle M\rangle_u\right]} \sqrt{\mathbb{E}\langle M\rangle_t} < +\infty$$

since $M \in \mathscr{H}^{2,c}$ and $K \in \mathrm{L}^2(M)$ (we took one of the functions to be identically one in Cauchy–Schwarz).

Taking $\varphi = \xi \mathbf{1}_{(s,t]} \in \mathscr{E}$, with $0 \le s < t$ and ξ a bounded $\mathscr{F}_s$-measurable r.v., we have

$$0 = \langle K, \varphi \rangle_{\mathrm{L}^2(M)} = \mathbb{E}\left[\xi \int_s^t K_u \mathrm{d}\langle M \rangle_u\right] = \mathbb{E}\left[\xi (X_t - X_s)\right].$$

Since this holds for any $\mathscr{F}_s$-measurable bounded ξ, we conclude that $\mathbb{E}[(X_t - X_s)|\mathscr{F}_s] = 0$. In other words, X is a martingale. But X is also continuous and of finite variation and hence $X \equiv 0$ a.s. Thus $K = 0$ $\mathrm{d}\langle M \rangle - a.e.$ $a.s.$ and hence $K = 0$ in $\mathrm{L}^2(M)$. □

8.1.3 General integrals in $L^2(M)$

We now know that any $K \in L^2(M)$ is a limit of simple processes $\varphi^n \to K$. For each φ^n we can define the stochastic integral $\varphi^n \bullet M$. The isometry property then shows that $\{\varphi^n \bullet M\}_{n \in \mathbb{N}}$ converges in $\mathscr{H}^{2,c}$ to some element that we denote $K \bullet M$ and which does not depend on the choice of approximating sequence φ^n.

Theorem 8.5. *Let $M \in \mathscr{H}^{2,c}$. The mapping $\varphi \mapsto \varphi \bullet M$ from $\mathscr{E}$ to $\mathscr{H}_0^{2,c}$ defined in* (27) *has a unique extension to a linear isometry from* $\mathrm{L}^2(M)$ *to $\mathscr{H}_0^{2,c}$ which we denote $K \mapsto K \bullet M$.*

Remark 8.6. *For $K \in \mathrm{L}^2(M)$, the martingale $K \bullet M$ is called the* Itô stochastic integral of K with respect to M *and is often written as $(K \bullet M)_t = \int_0^t K_u \mathrm{d}M_u$. The isometry property, called the* Itô isometry, *may then be written as*

$$\|K \bullet M\|^2_{\mathscr{H}^{2,c}} = \mathbb{E}\left[\left(\int_0^\infty K_t \mathrm{d}M_t\right)^2\right] = \mathbb{E}\left[\int_0^\infty K_t^2 \mathrm{d}\langle M \rangle_t\right] = \|K\|^2_{\mathrm{L}^2(M)}. \tag{32}$$

Example 8.7. *Let $M = B^T$ be a standard Brownian motion stopped at a time $T > 0$. Consider $K_t = t^3$. To define $K \bullet B$, we take a sequence of simple functions K^n which converge to K in the sense*

$$\mathbb{E}\Big[\int_0^\infty (K_t^n - K_t)^2 d\langle M \rangle_t\Big] = \mathbb{E}\Big[\int_0^T (K_t^n - t^3)^2 dt\Big] \to 0.$$

Then the approximate integrals $K^n \bullet M$ (which are defined by Riemann sums) converge in $\mathscr{H}^2$ to a process which we call $K \bullet M$, or

$$(K \bullet M)_t = \int_0^t s^3 dB_s, \qquad \text{for } t \le T.$$

Notice that if B is standard Brownian motion and we calculate $(B \bullet B)_t$, then

$$(B \bullet B)_t = \lim_{\|\pi\| \to 0} \sum_{j=0}^{N(\pi)-1} B_{t_j}\left(B_{t_{j+1}} - B_{t_j}\right). \tag{33}$$

We also know already that the quadratic variation is

$$t = \lim_{\|\pi\|\to 0} \sum_{j=0}^{N(\pi)-1} \Big(B_{t_{j+1}} - B_{t_j}\Big)^2 = B_t^2 - B_0^2 - 2 \sum_{j=0}^{N(\pi)-1} B_{t_j}\Big(B_{t_{j+1}} - B_{t_j}\Big),$$

and so rearranging we find

$$\int_0^t B_s dB_s = \frac{1}{2}\Big(B_t^2 - B_0^2 - t\Big) = \frac{1}{2}\Big(B_t^2 - t\Big).$$

This is *not* what one would have predicted from classical integration theory (the extra term here comes from the quadratic variation).

Even more strangely, it *matters* that in (33) we took the *left* endpoint of the interval for evaluating the integrand. On the problem sheet, you are asked to evaluate

$$\lim_{\|\pi\|\to 0} \sum B_{t_{j+1}}\Big(B_{t_{j+1}} - B_{t_j}\Big), \quad \text{and} \lim_{\|\pi\|\to 0} \sum \frac{B_{t_j} + B_{t_{j+1}}}{2}\Big(B_{t_{j+1}} - B_{t_j}\Big).$$

Each gives a different answer.

We can more generally define

$$\int_0^T f(B_s) \circ dB_s = \lim_{\|\pi\|\to 0} \sum \Big(\frac{f(B_{t_j}) + f(B_{t_{j+1}})}{2}\Big)\Big(B_{t_{j+1}} - B_{t_j}\Big).$$

This is the so-called *Stratonovich integral*, and has the advantage that from the point of view of calculations, the rules of Newtonian calculus hold true. From a modelling perspective however, it can be the wrong choice. For example, suppose that we are modelling the change in a population size over time and we use $[t_i, t_{i+1})$ to represent the $(i+1)$st generation. The change over (t_i, t_{i+1}) will be driven by the number of *adults*, so the population size at the *beginning* of the interval.

8.2 Intrinsic characterisation of stochastic integrals using the quadratic co-variation

We can also characterise the Itô integral in a slightly different way.

Theorem 8.8. *Let $M \in \mathscr{H}^{2,c}$. For any $K \in \mathrm{L}^2(M)$ there exists a unique element in $\mathscr{H}_0^{2,c}$, denoted $K \bullet M$, such that*

$$\langle K \bullet M, N\rangle = K \cdot \langle M, N\rangle, \quad \forall N \in \mathscr{H}^{2,c}. \tag{34}$$

Furthermore, $\|K \bullet M\|_{\mathscr{H}^{2,c}} = \|K\|_{\mathrm{L}^2(M)}$ and the map

$$\begin{gathered} K \mapsto K \bullet M \\ L^2(M) \to \mathscr{H}_0^{2,c} \end{gathered}$$

is a linear isometry.

Proof. We first check uniqueness. Suppose that there are two such elements, X and X'. Then

$$\langle X,N\rangle - \langle X',N\rangle = \langle X-X',N\rangle \equiv 0, \quad \forall N \in \mathscr{H}^{2,c}.$$

Taking $N = X - X'$ we conclude, by Corollary 7.25, that $X = X'$.

Now let us verify (34) for the Itô integral.

Fix $N \in \mathscr{H}^{2,c}$. First note that for $K \in L^2(M)$ the Kunita–Watanabe inequality shows that

$$\mathbb{E}\left[\int_0^\infty |K_s||d\langle M,N\rangle_s|\right] \le \|K\|_{L^2(M)}\|N\|_{\mathscr{H}^{2,c}} < \infty$$

and thus the variable

$$\int_0^\infty K_s d\langle M,N\rangle_s = \Big(K\cdot\langle M,N\rangle\Big)_\infty$$

is well defined and in L^1.

If K is an elementary process, in the notation of (27) and (28),

$$\langle K\bullet M,N\rangle = \sum_{i=0}^m \langle M^i,N\rangle$$

and

$$\langle M^i,N\rangle_t = K^{(i)}\Big(\langle M,N\rangle_{t_{i+1}\wedge t} - \langle M,N\rangle_{t_i\wedge t}\Big),$$

so

$$\langle K\bullet M,N\rangle_t = \sum K^{(i)}\Big(\langle M,N\rangle_{t_{i+1}\wedge t} - \langle M,N\rangle_{t_i\wedge t}\Big) = \int_0^t K_s d\langle M,N\rangle_s.$$

Now observe that the mapping $X \mapsto \langle X,N\rangle_\infty$ is continuous from $\mathscr{H}^{2,c}$ into L^1. Indeed, by Kunita–Watanabe

$$\mathbb{E}\Big[|\langle X,N\rangle|\Big] \le \mathbb{E}\Big[\langle X,X\rangle_\infty\Big]^{1/2}\mathbb{E}\Big[\langle N,N\rangle_\infty\Big]^{1/2} = \|N\|_{\mathscr{H}^{2,c}}\|X\|_{\mathscr{H}^{2,c}}.$$

So if K^n is a sequence in $\mathscr{E}$ such that $K^n \to K$ in $L^2(M)$,

$$\langle K\bullet M,N\rangle_\infty = \lim_{n\to\infty}\langle K^n\bullet M,N\rangle_\infty = \lim_{n\to\infty}\Big(K^n\cdot\langle M,N\rangle\Big)_\infty = \Big(K\cdot\langle M,N\rangle\Big)_\infty,$$

where the convergence is in L^1 and the last equality is again a consequence of Kunita–Watanabe by writing

$$\mathbb{E}\left[\left|\int_0^\infty \Big(K_s^n - K_s\Big)d\langle M,N\rangle_s\right|\right] \le \mathbb{E}\Big[\langle N,N\rangle_\infty\Big]^{1/2}\|K^n - K\|_{L^2(M)}.$$

We have thus obtained

$$\langle K\bullet M,N\rangle_\infty = \Big(K\cdot\langle M,N\rangle\Big)_\infty,$$

but replacing N by the stopped martingale N^t in this identity also gives

$$\langle K\bullet M,N\rangle_t = \Big(K\cdot\langle M,N\rangle\Big)_t$$

which completes the proof of (34). □

8.3 Properties of the stochastic integral

We could write the relationship (34) as

$$\left\langle \int_0^{\cdot} K_s dM_s, N \right\rangle_t = \int_0^t K_s d\langle M,N\rangle_s ;$$

that is, the stochastic integral 'commutes' with the bracket. One important consequence is that if $M \in \mathscr{H}^{2,c}$ and $K \in L^2(M)$, then applying (34) twice gives

$$\langle K \bullet M, K \bullet M \rangle = K \cdot \Big(K \cdot \langle M,M\rangle \Big) = K^2 \cdot \langle M,M\rangle.$$

In other words, the bracket process of $\int K_s dM_s$ is $\int K_s^2 d\langle M,M\rangle_s$. More generally, for N another martingale and $H \in L^2(N)$,

$$\left\langle \int_0^{\cdot} H_s dN_s, \int_0^{\cdot} K_s dM_s \right\rangle_t = \int_0^t H_s K_s d\langle M,N\rangle_s .$$

Proposition 8.9 (Associativity of stochastic integration)**.** *Let $H \in L^2(M)$. If K is progressive, then $KH \in L^2(M)$ if and only if $K \in L^2(H \bullet M)$. In that case,*

$$(KH) \bullet M = K \bullet (H \bullet M).$$

(This is the analogue of what we already know for finite variation processes, where $K \cdot (H \cdot A) = (KH) \cdot A$.)

Proof.

$$\mathbb{E}\left[\int_0^\infty K_s^2 H_s^2 d\langle M,M\rangle_s \right] = \mathbb{E}\left[\int_0^\infty K_s^2 d\langle H \bullet M, H \bullet M\rangle_s \right],$$

which gives the first assertion.

For the second, for $N \in \mathscr{H}^{2,c}$ we write

$$\begin{aligned} \langle (KH) \bullet M, N\rangle &= KH \cdot \langle M,N\rangle = K \cdot (H \cdot \langle M,N\rangle) \\ &= K \cdot \langle H \bullet M, N\rangle = \langle K \bullet (H \bullet M), N\rangle, \end{aligned}$$

and by uniqueness in (34) this implies

$$(KH) \bullet M = K \bullet (H \bullet M).$$

□

Recall that if $M \in \mathscr{H}^{2,c}$ and τ is a stopping time, then $M^\tau = (M_{t\wedge\tau}, t \geq 0)$ denotes the stopped process, which is itself a martingale and clearly $M^\tau \in \mathscr{H}^{2,c}$. For any $N \in \mathscr{H}^{2,c}$ we have

$$\langle M^\tau, N\rangle = \langle M,N\rangle^\tau = \mathbf{1}_{[0,\tau]} \cdot \langle M,N\rangle = \langle \mathbf{1}_{[0,\tau]} \bullet M, N\rangle,$$

so by uniqueness in Theorem 8.8, $\mathbf{1}_{[0,\tau]} \bullet M = M^\tau$.

In fact a much more general property holds true.

Proposition 8.10 (Stopped stochastic integrals)**.** *Let $M \in \mathscr{H}^{2,c}$, $K \in \mathrm{L}^2(M)$ and τ be a stopping time. Then*

$$(K \bullet M)^{\tau} = K \bullet M^{\tau} = K\mathbf{1}_{[0,\tau]} \bullet M.$$

Proof. We already argued above that the result holds for $K \equiv 1$.
Associativity says

$$K \bullet M^{\tau} = K \bullet \left(\mathbf{1}_{[0,\tau]} \bullet M\right) = K\mathbf{1}_{[0,\tau]} \bullet M.$$

Applying the same result to the martingale $K \bullet M$ we obtain

$$(K \bullet M)^{\tau} = \mathbf{1}_{[0,\tau]} \bullet (K \bullet M) = \mathbf{1}_{[0,\tau]} K \bullet M,$$

which gives the desired equalities. □

8.4 Stochastic integration with respect to continuous local martingales

Definition 8.11. *For a continuous local martingale M, denote by $\mathrm{L}^2_{loc}(M)$ the space of progressively measurable processes K such that*

$$\forall t \geq 0 \quad \int_0^t K_s^2 \mathrm{d}\langle M\rangle_s < \infty \ a.s.$$

Theorem 8.12. *Let M be a continuous local martingale. For any $K \in \mathrm{L}^2_{loc}(M)$ there exists a unique continuous local martingale, zero in zero, denoted by $K \bullet M$ and called the Itô integral of K with respect to M, such that for any continuous local martingale N*

$$\langle K \bullet M, N\rangle = K \cdot \langle M, N\rangle. \tag{35}$$

If $M \in \mathscr{H}^{2,c}$ and $K \in \mathrm{L}^2(M)$ then this definition coincides with the previous one.

Proof. We only sketch the proof. Not surprisingly, we use a stopping argument.
For every $n \geq 1$, set

$$\tau_n = \inf\left\{t \geq 0 : \int_0^t (1+K_s^2)\mathrm{d}\langle M\rangle_s \geq n\right\},$$

so that τ_n is a sequence of stopping times that increases to infinity. Since $\langle M^{\tau_n}\rangle_\infty = \langle M\rangle_{\tau_n} \leq n$, the stopped martingale M^{τ_n} is in $\mathscr{H}^{2,c}$. Also

$$\int_0^\infty K_s^2 d\langle M^{\tau_n}, M^{\tau_n}\rangle_s = \int_0^{\tau_n} K_s^2 d\langle M, M\rangle_s \leq n,$$

so that $K \in \mathrm{L}^2(M^{\tau_n})$ and the definition of $K \bullet M^{\tau_n}$ makes sense. If $m > n$,

$$K \bullet M^{\tau_n} = (K \bullet M^{\tau_m})^{\tau_n}$$

so there is a unique process, that we denote $K \bullet M$ such that

$$(K \bullet M)^{\tau_n} = K \bullet M^{\tau_n}$$

and $(K \bullet M)_t = \lim_{n\to\infty}(K \bullet M^{\tau_n})_t$ and so, since $(K \bullet M^{\tau_n})$ is a martingale, the process $K \bullet M$ is a continuous local martingale with reducing sequence τ_n.

If N is a continuous local martingale (and without loss of generality $N_0 = 0$), we consider a reducing sequence

$$\tilde{\tau}_n = \inf\{t \geq 0 : |N_t| \geq n\} \quad \text{and set} \quad \rho_n := \tau_n \wedge \tilde{\tau}_n.$$

Then $N^{\rho_n} \in \mathscr{H}_0^{2,c}$ and hence

$$\begin{aligned}\langle K \bullet M, N\rangle^{\rho_n} &= \langle (K \bullet M)^{\rho_n}, N^{\rho_n}\rangle \overset{\tau_n \geq \rho_n}{=} \langle (K \bullet M^{\tau_n})^{\rho_n}, N^{\rho_n}\rangle \overset{(21)}{=} \langle K \bullet M^{\tau_n}, N^{\rho_n}\rangle \\ &\overset{\text{Thm 8.8}}{=} K \cdot \langle M^{\tau_n}, N^{\rho_n}\rangle \overset{(21)}{=} K \cdot \langle M, N\rangle^{\rho_n} = (K \cdot \langle M, N\rangle)^{\rho_n},\end{aligned}$$

so that $\langle K \bullet M, N\rangle = K \cdot \langle M, N\rangle$ as required. Uniqueness of $K \bullet M$ follows as in Theorem 8.8. □

8.5 Stochastic integration with respect to continuous semimartingales

Naturally, we are going to define an integral with respect to a continuous semimartingale $X = X_0 + M + A$ as a sum of integrals w.r.t. M and w.r.t. A.

Definition 8.13. *Let $X = X_0 + M + A$ be a continuous semimartingale. The space of X-*stochastically integrable processes *is given by*

$$L(X) := \mathrm{L}^2_{loc}(M) \cap L^1_{loc}(|dA|),$$

that is $K \in L(X)$ if there are stopping times $\tau_n \to \infty$ such that

$$\mathbb{E}\Big[\int_0^{\tau_n} K_t^2 d\langle M\rangle_t\Big] < \infty \qquad \text{and} \qquad \int_0^{\tau_n} |K_t||dA_t| < \infty \quad \text{a.s.}$$

A subset of such integrands, particularly convenient since it does not depend on X, is given by

Definition 8.14. *We say that a progressively measurable process K is locally bounded if*

$$\sup_{u \leq t} |K_u| < \infty \quad \forall t \geq 0, \quad a.s.$$

In particular, any adapted process with continuous sample paths is locally bounded.

Proposition 8.15. *If K is progressively measurable and locally bounded, then it is in $L(X)$ for every continuous semimartingale X.*

Proof. Take $\tau_n = \inf\{T : \int_0^T K_t^2 d\langle M\rangle_t + \int_0^T |K_t||dA_t| \geq n\}$. □

Definition 8.16. *Let $X = X_0 + M + A$ be a continuous semimartingale and $K \in L(X)$. The* Itô stochastic integral *of K with respect to X is the continuous semimartingale $K \bullet X$ defined by*

$$K \bullet X := K \bullet M + K \cdot A$$

often written

$$(K \bullet X)_t = \int_0^t K_s \mathrm{d}X_s = \int_0^t K_s \mathrm{d}M_s + \int_0^t K_s \mathrm{d}A_s.$$

This integral inherits all the nice properties of the Stieltjes integral and the Itô integral that we have already derived (linearity, associativity, stopping etc.).

And of course, it is still the case for an elementary function $\varphi \in \mathscr{E}$ that

$$(\varphi \bullet X)_t = \sum_{i=1}^m \varphi^{(i)} \left(X_{t_{i+1}\wedge t} - X_{t_i \wedge t} \right).$$

8.6 Stochastic dominated convergence

We should also like to know how our integral behaves under limits.

Proposition 8.17 (Stochastic Dominated Convergence Theorem)**.** *Let X be a continuous semimartingale and K^n a sequence in $L(X)$ with $K_t^n \to 0$ as $n \to \infty$ for all t almost surely. Further suppose that $|K_t^n| \leq K_t$ for all n where $K \in L(X)$. Then $K^n \bullet X$ converges to zero in probability and, more precisely,*

$$\forall t \geq 0 \quad \sup_{s \leq t} \left| \int_0^s K_u^n \mathrm{d}X_u \right| \longrightarrow 0 \text{ in probability as } n \to \infty.$$

Proof. We can treat the finite variation part, $X_0 + A$, and the local martingale part, M, separately. For the first, note that

$$\begin{aligned} \left| \int_0^t K_u^n \mathrm{d}A_u \right| &= \left| \int_0^t K_u^n \mathrm{d}A_u^+ - \int_0^t K_u^n \mathrm{d}A_u^- \right| \\ &\leq \int_0^t |K_u^n| \mathrm{d}A_u^+ + \int_0^t |K_u^n| \mathrm{d}A_u^- = \int_0^t |K_u^n||\mathrm{d}A_u|. \end{aligned}$$

The a.s. pointwise convergence of K^n to 0, together with the bound $|K^n| \leq K$, allow us to apply the (usual) Dominated Convergence Theorem to conclude that, for any $t > 0$, $\int_0^t |K_u^n||\mathrm{d}A_u|$ converges to 0 a.s. (in fact, as $\int_0^t |K_u^n||\mathrm{d}A_u|$ is non-decreasing in t, the convergence is uniform on any compact interval).

For the continuous local martingale part M, let (τ_m) be a reducing sequence such that $M^{\tau_m} \in \mathscr{H}_0^{2,c}$ and $K \in \mathrm{L}^2(M^{\tau_m})$. Then, by the Itô isometry,

$$\|K^n \bullet M^{\tau_m}\|_{\mathscr{H}^{2,c}}^2 = \mathbb{E}\left[\left(\int_0^{\tau_m} K_t^n \mathrm{d}M_t \right)^2 \right] = \mathbb{E}\left[\int_0^\infty (K_t^n)^2 \mathbf{1}_{[0,\tau_m]}(t) \mathrm{d}\langle M\rangle_t \right] = \|K^n\|_{\mathrm{L}^2(M^{\tau_m})}^2.$$

The right hand side tends to zero by the usual Dominated Convergence Theorem. For a fixed $t \geq 0$, and any given $\varepsilon > 0$, we may take m large enough that $\mathbb{P}[\tau_m \leq t] \leq \varepsilon/2$. We then have

$$\begin{aligned}\mathbb{P}\left[\sup_{s\leq t}|(K^n \bullet M)_s| > \varepsilon\right] &\leq \mathbb{P}\left[\sup_{s\leq t\wedge\tau_m}|(K^n \bullet M)_s| > \varepsilon\right] + \varepsilon/2 \\ &\leq \frac{1}{\varepsilon^2}\|K^n \bullet M^{\tau_m}\|^2_{\mathscr{H}^{2,c}} + \varepsilon/2 \leq \varepsilon,\end{aligned}$$

for n large enough. □

From this we can also confirm that even in their most general form our stochastic integrals can be thought of as limits of integrals of simple functions.

Proposition 8.18. *Let X be a continuous semimartingale and K a left-continuous process in $L(X)$. If π^n is a sequence of partitions of $[0,t]$ with mesh converging to zero then*

$$\sum_{t_i\in\pi^n} K_{t_i}(X_{t_{i+1}} - X_{t_i}) \longrightarrow \int_0^t K_s \mathrm{d}X_s \text{ in probability as } n \to \infty.$$

9 Itô's formula and its applications

9.1 Itô's formula and integration by parts

We already saw that the stochastic integral of Brownian motion with respect to itself did not behave as we would expect from Newtonian calculus. So what are the analogues of integration by parts and the chain rule for stochastic integrals?

Proposition 9.1 (Integration by parts)**.** *If X and Y are two continuous semimartingales then*

$$\begin{aligned} X_tY_t &= X_0Y_0 + \int_0^t X_s \mathrm{d}Y_s + \int_0^t Y_s \mathrm{d}X_s + \langle X,Y\rangle_t, \quad t \geq 0 \quad a.s. \qquad (36)\\ &= X_0Y_0 + (X\bullet Y)_t + (Y\bullet X)_t + \langle X,Y\rangle_t. \end{aligned}$$

Proof. Fix t and let π^n be a sequence of partitions of $[0,t]$ with mesh converging to zero. Note that

$$X_tY_t - X_sY_s = X_s(Y_t - Y_s) + Y_s(X_t - X_s) + (X_t - X_s)(Y_t - Y_s)$$

so for any n

$$\begin{aligned} X_tY_t - X_0Y_0 &= \sum_{t_i\in\pi^n}\Big(X_{t_i}(Y_{t_{i+1}} - Y_{t_i}) + Y_{t_i}(X_{t_{i+1}} - X_{t_i}) + (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i})\Big) \\ &\longrightarrow (X\bullet Y)_t + (Y\bullet X)_t + \langle X,Y\rangle_t \quad \text{as } n\to\infty. \end{aligned}$$

□

Remark 9.2. *Comparing with the finite variation case with jumps (Proposition 7.8), we see that the 'product of jumps' term has become the 'quadratic variation' term.*

Theorem 9.3 (Itô's formula)**.** *Let $X^1,\ldots,X^d$ be continuous semimartingales and $F:\mathbb{R}^d\to\mathbb{R}$ a C^2 function. Then $(F(X_t^1,\ldots,X_t^d):t\geq 0)$ is a continuous semimartingale and up to indistinguishability*

$$\begin{aligned}F(X_t^1,\ldots,X_t^d)=&F(X_0^1,\ldots,X_0^d)+\sum_{i=1}^d\int_0^t\frac{\partial F}{\partial x^i}(X_s^1,\ldots,X_s^d)\mathrm{d}X_s^i\\&+\frac{1}{2}\sum_{1\leq i,j\leq d}\int_0^t\frac{\partial^2 F}{\partial x^i\partial x^j}(X_s^1,\ldots,X_s^d)\mathrm{d}\langle X^i,X^j\rangle_s.\end{aligned}\tag{37}$$

In particular, for $d=1$, we have

$$F(X_t)=F(X_0)+\int_0^t F'(X_s)\mathrm{d}X_s+\frac{1}{2}\int_0^t F''(X_s)\mathrm{d}\langle X\rangle_s.$$

Proof. Let $X^i=X_0^i+M^i+A^i$ be the semimartingale decomposition of X^i and denote by V^i the total variation process of A^i. Let

$$\tau_r^i=\inf\{t\geq 0:|X_t^i|+V_t^i+\langle M^i\rangle_t>r\},$$

and $\tau_r=\min\{\tau_r^i,i=1,\ldots,d\}$. Then $(\tau_r)_{r\geq 0}$ is a family of stopping times with $\tau_r\uparrow\infty$. It is sufficient to prove (37) up to time τ_r. We will prove that the result holds for polynomials and then the full result follows by approximating C^2 functions by polynomials.

First note that it is obvious that the set of functions for which the formula holds is a vector space containing the functions $F\equiv 1$ and $F(x_1,\ldots,x_d)=x_i$ for $i\leq d$.

We now check that if (37) holds for two functions F and G, then it holds for the product FG. Integration by parts yields

$$F_tG_t-F_0G_0=\int_0^t F_s\mathrm{d}G_s+\int_0^t G_s\mathrm{d}F_s+\langle F,G\rangle_t.\tag{38}$$

By associativity of stochastic integration, and because (37) holds for G,

$$\int_0^t F_s\mathrm{d}G_s=\sum_{i=1}^d\int_0^t F(X_s)\frac{\partial G_s}{\partial x^i}\mathrm{d}X_s^i+\frac{1}{2}\sum_{1\leq i,j\leq d}\int_0^t F(X_s)\frac{\partial^2 G_s}{\partial x^i\partial x^j}\mathrm{d}\langle X^i,X^j\rangle_s,$$

with a similar expression for $\int_0^t G_s\mathrm{d}F_s$. Using the fact that (37) holds for F and G, we also have

$$\langle F,G\rangle_t=\sum_{i=1}^d\sum_{j=1}^d\int_0^t\frac{\partial F_s}{\partial x^i}\frac{\partial G_s}{\partial x^j}\mathrm{d}\langle X^i,X^j\rangle_s.$$

Substituting these into (38), we obtain Itô's formula for FG.

To pass to a general C^2 function F, the Stone–Weierstrass theorem (see appendix, Theorem D.12) allows us to approximate the second derivative of F uniformly on compacts by a polynomial (and hence F' and F are also uniformly approximated on compacts). Using the dominated convergence theorem (and the fact that everything is nicely bounded up to time τ_r), we have the result up to time τ_r, and then we send $r \to \infty$. □

9.2 Applications of Itô's formula

As a first application of this, suppose that M is a continuous local martingale and A is a process of finite variation. Then $\langle M,A\rangle \equiv 0$ and applying Itô's formula with $X^1 = M$ and $X^2 = A$ yields

$$F(M_t,A_t) = F(M_0,A_0) + \int_0^t \frac{\partial F}{\partial m}(M_s,A_s)\mathrm{d}M_s + \int_0^t \frac{\partial F}{\partial a}(M_s,A_s)\mathrm{d}A_s + \frac{1}{2}\int_0^t \frac{\partial^2 F}{\partial m^2}(M_s,A_s)\mathrm{d}\langle M\rangle_s.$$

Note that this gives us the semimartingale decomposition of $F(M_t,A_t)$ and we can, for example, read off the conditions on F under which we recover a local martingale. In particular, taking $F(x,y) = \exp(\lambda x - \frac{\lambda^2}{2}y)$ with $X^1 = M$ and $X^2 = \langle M,M\rangle$, we obtain:

Proposition 9.4. *Let M be a continuous local martingale and $\lambda \in \mathbb{R}$. Then*

$$\mathscr{E}^\lambda(M)_t := \exp\left(\lambda M_t - \frac{\lambda^2}{2}\langle M\rangle_t\right), \quad t \geq 0, \tag{39}$$

is a continuous local martingale. In fact the same holds true for any $\lambda \in \mathbb{C}$ with the real and imaginary parts being local martingales.

Proof. Let $F(x,y) = \exp\left(\lambda x - \frac{\lambda^2}{2}y\right)$. $F \in C^2(\mathbb{R}^2,\mathbb{C})$ so we may apply Itô's formula to $\mathscr{E}^\lambda(M)_t = F(M_t,\langle M\rangle_t)$. Computing the partial derivatives and simplifying gives:

$$\mathscr{E}^\lambda(M)_t = \mathscr{E}^\lambda(M)_0 + \int_0^t \frac{\partial}{\partial x}F^\lambda(M_s,\langle M\rangle_s)\mathrm{d}M_s.$$

□

Note that we have $\frac{\partial}{\partial x}F(x,y) = \lambda F(x,y)$ so that we could have written this as

$$\mathscr{E}^\lambda(M)_t = \mathscr{E}^\lambda(M)_0 + \lambda\int_0^t \mathscr{E}^\lambda(M)_s \mathrm{d}M_s$$

or in 'differential form' as

$$\mathrm{d}\mathscr{E}^\lambda(M)_t = \lambda\mathscr{E}^\lambda(M)_t\mathrm{d}M_t$$

which shows $\mathscr{E}^\lambda(M)$ solves the stochastic exponential differential equation driven by M: $\mathrm{d}Y_t = \lambda Y_t \mathrm{d}M_t$.

9.2.1 Lévy's characterization

Here is a beautiful application of exponential martingales:

Theorem 9.5 (Lévy's characterisation of Brownian motion)**.** *Let M be a continuous local martingale starting at zero. Then M is a standard Brownian motion if and only if $\langle M\rangle_t = t$ a.s. for all $t \geq 0$.*

Proof. We know that the quadratic variation of a Brownian motion B is given by $\langle B\rangle_t = t$.

Suppose M is a continuous local martingale starting in zero with $\langle M\rangle_t = t$ a.s. for all $t \geq 0$. Then, by Proposition 9.4,

$$\exp\left(\mathrm{i}\xi M_t + \frac{\xi^2}{2}t\right), \quad t \geq 0$$

is a local martingale for any $\xi \in \mathbb{R}$ and, since it is bounded, it is a martingale. Let $0 \leq s < t$. We have

$$\mathbb{E}\left[\exp\left(\mathrm{i}\xi M_t + \frac{\xi^2}{2}t\right)\Big|\mathscr{F}_s\right] = \exp\left(\mathrm{i}\xi M_s + \frac{\xi^2}{2}s\right)$$

which we can rewrite as

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi(M_t - M_s)}\Big|\mathscr{F}_s\right] = \mathrm{e}^{-\frac{\xi^2}{2}(t-s)}. \tag{40}$$

In other words, $M_t - M_s$ is centred Gaussian with (conditional) variance $t-s$.

It follows also from (40) that for $A \in \mathscr{F}_s$,

$$\mathbb{E}\left[\mathbf{1}_A e^{i\xi(M_t - M_s)}\right] = \mathbb{P}[A]\mathbb{E}\left[e^{i\xi(M_t - M_s)}\right],$$

so fixing $A \in \mathscr{F}_s$ with $\mathbb{P}[A] > 0$ and writing $\mathbb{P}_A = \mathbb{P}[\cdot \cap A]/\mathbb{P}[A]$ (which is a probability measure on $\mathscr{F}_s$) for the conditional probability given A, we have that $M_t - M_s$ has the same distribution under $\mathbb{P}$ as under $\mathbb{P}_A$ and so $M_t - M_s$ is independent of $\mathscr{F}_s$ and we have that M is an $\{\mathscr{F}_t\}_{t\geq 0}$-Brownian motion. □

So the quadratic variation is capturing all the information about M. This is surprising – recall that it is a special property of Gaussians that they are characterised by their means and the variance-covariance matrix, but in general we need to know much more. It also shows we didn't really need the Gaussian assumption in our definition of Brownian motion, it's guaranteed by the independence and variance assumptions.

9.2.2 Dambis–Dubins–Schwarz Theorem

It turns out that what we just saw for Brownian motion has a powerful consequence for all continuous local martingales – they are characterised by their quadratic variation and, in fact, they are all time changes of Brownian motion.

Theorem 9.6 (Dambis–Dubins–Schwarz Theorem)**.** *Let M be an $(\{\mathscr{F}_t\}_{t\geq 0},\mathbb{P}))$-continuous local martingale with $M_0 = 0$ and $\langle M\rangle_\infty = \infty$ a.s. Let $\tau_s := \inf\{t \geq 0 : \langle M\rangle_t > s\}$. Then the process B defined by $B_s := M_{\tau_s}$, is an $(\{\mathscr{F}_{\tau_s}\}_{s\geq 0},\mathbb{P})$-Brownian motion and $M_t = B_{\langle M\rangle_t}$, $\forall t \geq 0$ a.s.*

Proof. Note that τ_s is the first hitting time of an open set (s,∞) for an adapted process $\langle M\rangle$ with continuous sample paths, and hence τ_s is a stopping time (recall that $\{\mathscr{F}_t\}_{t\geq 0}$ is right-continuous). Further, $\langle M\rangle_\infty = \infty$ a.s. implies that $\tau_s < \infty$ a.s. The process $(\tau_s : s \geq 0)$ is non-decreasing and right-continuous (in fact $s \to \tau_s$ is the right-continuous inverse of $t \to \langle M\rangle_t$). Let $\mathscr{G}_s := \mathscr{F}_{\tau_s}$. Note that it satisfies the usual conditions. The process B is right continuous by continuity of M and right-continuity of τ. We have

$$\lim_{u\uparrow s} B_u = \lim_{u\uparrow s} M_{\tau_u} = M_{\tau_{s-}}.$$

But$[\tau_{s-},\tau_s]$ is either a point or an interval of constancy of $\langle M\rangle$. The latter are known (exercise) to coincide a.s. with the intervals of constancy of M and hence $M_{\tau_{s-}} = M_{\tau_s} = B_s$ so that B has a.s. continuous paths. To conclude that B is a $(\mathscr{G}_s)$-Brownian motion, by Lévy's theorem, it remains to show that (B_s) and $(B_s^2 - s)$ are $(\mathscr{G}_s)$-local martingales.

Note that M^{τ_n} and $(M^{\tau_n})^2 - \langle M\rangle^{\tau_n}$ are uniformly integrable martingales. Taking $0 \leq u < s < n$ and applying the Optional Stopping Theorem we obtain

$$\mathbb{E}[B_s|\mathscr{G}_u] = \mathbb{E}[M_{\tau_s}^{\tau_n}|\mathscr{F}_{\tau_u}] = M_{\tau_u}^{\tau_n} = M_{\tau_u} = B_u$$

and

$$\mathbb{E}[B_s^2 - s|\mathscr{G}_u] = \mathbb{E}\left[(M_{\tau_s}^{\tau_n})^2 - \langle M\rangle_{\tau_s}^{\tau_n}|\mathscr{F}_{\tau_u}\right] = (M_{\tau_u}^{\tau_n})^2 - \langle M\rangle_{\tau_u}^{\tau_n} = (M_{\tau_u})^2 - \langle M\rangle_{\tau_u} = B_u^2 - u,$$

where we used continuity of $\langle M\rangle$ to write $\langle M\rangle_{\tau_u} = u$. It follows that B is indeed a $(\mathscr{G}_s)$-Brownian motion.

Finally, $B_{\langle M\rangle_t} = M_{\tau_{\langle M\rangle_t}} = M_t$, again since the intervals of constancy of M and of $\langle M\rangle$ coincide a.s. so that $s \to \tau_s$ is constant on $[t,\tau_{\langle M\rangle_t}]$. $\square$

A Review of some basic measure theoretic probability

A.1 The Monotone Class Lemma/ Dynkin's $\pi - \lambda$ Theorem

There are multiple names used for this result (often with slightly different formulations).

Let E be an arbitrary set and let $\mathscr{P}(E)$ be the set of all subsets of E.

Definition A.1. *A subset $\mathscr{M}$ of $\mathscr{P}(E)$ is called a* monotone class*, or a Dynkin system/ λ-system, if*

i. $E \in \mathscr{M}$;

ii. if $A, B \in \mathscr{M}$ and $A \subset B$, then $B \backslash A \in \mathscr{M}$;

iii. if $(A_n)_{n \geq 0}$ is an increasing sequence of subsets of E such that $A_n \in \mathscr{M}$, then $\cup_{n \geq 0} A_n \in \mathscr{M}$.

The monotone class generated by an arbitrary subset $\mathscr{C}$ of $\mathscr{P}(E)$ is

$$\mathscr{M}(\mathscr{C}) = \bigcap \{\mathscr{D} : \mathscr{D} \text{ monontone class}, \mathscr{C} \subset \mathscr{D}\}.$$

Equivalently, $\mathscr{M}$ is a montone class if

i. $E \in \mathscr{M}$;

ii. if $A, B \in \mathscr{M}$ and $A \subset B$, then $B \backslash A \in \mathscr{M}$;

iii. if $(A_n)_{n \geq 0}$ is a sequence of subsets of E such that $A_i \cap A_j = \emptyset$ for $i \neq j$, then $\cup_{n \geq 0} A_n \in \mathscr{M}$.

Definition A.2. *A collection $\mathscr{I}$ of subsets of E such that $\emptyset \in \mathscr{I}$ and for all A, $B \in \mathscr{I}$, $A \cap B \in \mathscr{I}$ is called a π-system.*

You may have seen the result expresed as:

Theorem A.3 (Dynkin's $\pi - \lambda$ Theorem)**.** *If P is a π-system and D is a λ-system such that $P \subseteq D$, then $\sigma(P) \subseteq \sigma(D)$.*

Le Gall's (equivalent) formulation is:

Lemma A.4 (Monotone class lemma)**.** *If $\mathscr{C} \subset \mathscr{P}(E)$ is stable under finite intersections, then $\mathscr{M} = \sigma(\mathscr{C})$.*

In other words, a Dynkin system which is also a π-system is a σ-algebra.

Here are some useful consequences:

i. Let $\mathscr{A}$ be a σ-field of E and let μ, ν be two probability measures on $(E, \mathscr{A})$. Assume that there exists $\mathscr{C} \subset \mathscr{A}$ which is stable under finite intersections and such that $\sigma(\mathscr{C}) = \mathscr{A}$ and $\mu(A) = \nu(A)$ for every $A \in \mathscr{C}$, then $\mu = \nu$. (Use that $\mathscr{G} = \{A \in \mathscr{A} : \mu(A) = \nu(A)\}$ is a montone class.)

ii. Let $(X_i)_{i\in I}$ be an arbitrary collection of random variables, and let $\mathscr{G}$ be a σ-field on some probability space. In order to show that the σ-fields $\sigma(X_i : i \in I)$ and $\mathscr{G}$ are independent, it is enough to verify that $(X_{i_1}, \ldots, X_{i_p})$ is independent of $\mathscr{G}$ for any choice of the finite set $\{i_1, \ldots, i_p\} \subset I$. (Observe that the class of all events that depend on a finite number of the X_i is stable under finite intersections and generates $\sigma(X_i, i \in I)$.)

iii. Let $(X_i)_{i\in I}$ be an arbitrary collection of random variables and let Z be a bounded real variable. Let $i_0 \in I$. In order to verify that $\mathbb{E}[Z|X_i, i \in I] = \mathbb{E}[Z|X_{i_0}]$, it is enough to show that $\mathbb{E}[Z|X_{i_0}, X_{i_1} \ldots X_{i_p}] = \mathbb{E}[Z|X_{i_0}]$ for any choice of the finite collection $\{i_1, \ldots i_p\} \subset I$. (Observe that the class of all events A such that $\mathbb{E}[\mathbf{1}_A Z] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[Z|X_{i_0}]]$ is a monotone class.)

A.2 Convergence of random variables.

We only consider real-valued random variables here.

a) $X_n \to X$ a.s. iff $\mathbb{P}[\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)] = 1$.

b) $X_n \to X$ in probability iff $\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0$.

c) X_n converges to X in distribution (denoted $X_n \Rightarrow X$) iff $\lim_{n\to\infty} \mathbb{P}\{X_n \leq x\} = \mathbb{P}\{X \leq x\} \equiv F_X(x)$ for all x at which F_X is continuous.

Theorem A.5. *a) implies b) implies c).*

Proof. ($b \Rightarrow c$) Let $\varepsilon > 0$. Then

$$\begin{aligned} \mathbb{P}\{X_n \leq x\} - \mathbb{P}\{X \leq x + \varepsilon\} &= \mathbb{P}\{X_n \leq x, X > x + \varepsilon\} - \mathbb{P}\{X \leq x + \varepsilon, X_n > x\} \\ &\leq \mathbb{P}\{|X_n - X| > \varepsilon\} \end{aligned}$$

and hence $\limsup \mathbb{P}\{X_n \leq x\} \leq \mathbb{P}\{X \leq x + \varepsilon\}$. Similarly, $\liminf \mathbb{P}\{X_n \leq x\} \geq \mathbb{P}\{X \leq x - \varepsilon\}$. Since ε is arbitrary, the implication follows. □

A.3 Convergence in probability.

a) If $X_n \to X$ in probability and $Y_n \to Y$ in probability, then $aX_n + bY_n \to aX + bY$ in probability.

b) If $f : \mathbb{R} \to \mathbb{R}$ is continuous and $X_n \to X$ in probability, then $f(X_n) \to f(X)$ in probability.

c) If $X_n \to X$ in probability and $X_n - Y_n \to 0$ in probability, then $Y_n \to X$ in probability.

Remark A.6. *(b) and (c) hold with convergence in probability replaced by convergence in distribution; however (a) is not in general true for convergence in distribution.*

Lemma A.7. *Suppose* $\{X_n\}_{n\in\mathbb{N}}$ *is a sequence of random variables converging in probability to* X. *Then there exists a* sub-*sequence* $\{X_{n_k}\}_{k\in\mathbb{N}}$ *which converges almost surely to* X.

Proof. For any $k \in \mathbb{N}$, there exists N_k such that for $n \geq N_k$ we know

$$\mathbb{P}\big(\{\omega : |X(\omega - X_n(\omega)| > 2^{-k}\}\big) < 2^{-k}$$

Let $B_k = \{\omega : |X(\omega) - X_{N_k}(\omega)| > 2^{-k}\}$ and $A = \cap_{k=1}^{\infty} \cup_{i=k}^{\infty} B_k$. Then if $\omega \notin A$, it follows that

$$|X(\omega) - X_{N_i}(\omega)| < 2^{-i} \quad \text{for all } i \text{ sufficiently large,}$$

so $X_n \to X$ on $\Omega \setminus A$. Finally, we note that for any k,

$$\mathbb{P}(A) \leq \mathbb{P}(\cup_{i=k}^{\infty} E_i) \leq \sum_{i=k}^{\infty} 2^{-i} = 2^{-k+1}$$

so $\mathbb{P}(A) = 0$. □

Note that the limit of a sequence converging in measure is uniquely defined, up to equality almost everywhere.

A.4 The basic convergence theorems

Theorem A.8. *(Bounded Convergence Theorem) Suppose that* $X_n \Rightarrow X$ *and that there exists a constant b such that* $\mathbb{P}(|X_n| \leq b) = 1$. *Then* $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Proof. Let $\{x_i\}$ be a partition of $\mathbb{R}$ such that F_X is continuous at each x_i. Then

$$\sum_i x_i \mathbb{P}\{x_i < X_n \leq x_{i+1}\} \leq \mathbb{E}[X_n] \leq \sum_i x_{i+1} \mathbb{P}\{x_i < X_n \leq x_{i+1}\}$$

and taking limits we have

$$\begin{aligned} \sum_i x_i \mathbb{P}\{x_i < X \leq x_{i+1}\} &\leq \underline{\lim}_{n\to\infty} E[X_n] \\ &\leq \overline{\lim}_{n\to\infty} E[X_n] \leq \sum_i x_{i+1} \mathbb{P}\{x_i < X \leq x_{i+1}\} \end{aligned}$$

As $\max |x_{i+1} - x_i| \to 0$, the left and right sides converge to $\mathbb{E}[X]$, as required. □

Lemma A.9. *Let* $X \geq 0$ *a.s. Then* $\lim_{M\to\infty} \mathbb{E}[X \wedge M] = \mathbb{E}[X]$.

Proof. Check the result first for X having a discrete distribution and then extend to general X by approximation. □

Theorem A.10. *(Monotone Convergence Theorem.) Suppose* $0 \leq X_n \leq X$ *and* $X_n \to X$ *in probability. Then* $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.

Proof. For $M > 0$

$$\mathbb{E}[X] \geq \mathbb{E}[X_n] \geq \mathbb{E}[X_n \wedge M] \to \mathbb{E}[X \wedge M]$$

where the convergence on the right follows from the bounded convergence theorem. It follows that

$$\mathbb{E}[X \wedge M] \leq \liminf_{n\to\infty} \mathbb{E}[X_n] \leq \limsup_{n\to\infty} \mathbb{E}[X_n] \leq \mathbb{E}[X]$$

and the result follows by Lemma A.9. □

Lemma A.11. *(Fatou's lemma.) If $X_n \geq 0$ and $X_n \Rightarrow X$, then* $\liminf \mathbb{E}[X_n] \geq \mathbb{E}[X]$.

Proof. Since $\mathbb{E}[X_n] \geq \mathbb{E}[X_n \wedge M]$ we have

$$\liminf \mathbb{E}[X_n] \geq \liminf \mathbb{E}[X_n \wedge M] = \mathbb{E}[X \wedge M].$$

By the Monotone Convergence Theorem $\mathbb{E}[X \wedge M] \to \mathbb{E}[X]$ and the result follows. □

Theorem A.12. *(Dominated Convergence Theorem) Assume $X_n \Rightarrow X$, $Y_n \Rightarrow Y$, $|X_n| \leq Y_n$, and $\mathbb{E}[Y_n] \to \mathbb{E}[Y] < \infty$. Then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.*

Proof. For simplicity, assume in addition that $X_n + Y_n \Rightarrow X + Y$ and $Y_n - X_n \Rightarrow Y - X$ (otherwise consider subsequences along which $(X_n, Y_n) \Rightarrow (X, Y)$). Then by Fatou's lemma $\liminf \mathbb{E}[X_n + Y_n] \geq \mathbb{E}[X + Y]$ and $\liminf \mathbb{E}[Y_n - X_n] \geq \mathbb{E}[Y - X]$. From these observations $\liminf \mathbb{E}[X_n] + \lim \mathbb{E}[Y_n] \geq \mathbb{E}[X] + \mathbb{E}[Y]$, and hence $\liminf \mathbb{E}[X_n] \geq \mathbb{E}[X]$. Similarly $\liminf \mathbb{E}[-X_n] \geq \mathbb{E}[-X]$ and $\limsup \mathbb{E}[X_n] \leq \mathbb{E}[X]$ □

Lemma A.13. *(Markov's inequality)*

$$\mathbb{P}\{|X| > a\} \leq \mathbb{E}[|X|]/a, \quad a \geq 0.$$

Proof. Note that $|X| \geq a\mathbf{1}_{\{|X|>a\}}$. Taking expectations proves the desired inequality. □

A.5 Uniform Integrability

If X is an integrable random variable (that is $\mathbb{E}[|X|] < \infty$) and Λ_n is a sequence of sets with $\mathbb{P}[\Lambda_n] \to 0$, then $\mathbb{E}[|X\mathbf{1}_{\Lambda_n}|] \to 0$ as $n \to \infty$. (This is a consequence of the DCT since $|X|$ dominates $|X\mathbf{1}_{\Lambda_n}|$ and $|X\mathbf{1}_{\Lambda_n}| \to 0\, a.s.$) Uniform integrability demands that this type of property holds *uniformly* for random variables from some class.

Definition A.14 (Uniform Integrability). *A class $\mathscr{C}$ of random variables is called* uniformly integrable *if given $\varepsilon > 0$ there exists $K \in (0, \infty)$ such that*

$$\mathbb{E}[|X|\mathbf{1}_{\{|X|>K\}}] < \varepsilon \qquad \text{for all } X \in \mathscr{C}.$$

Proposition A.15. *Suppose that* $\{X_\alpha, \alpha \in I\}$ *is a uniformly integrable family of random variables on some probability space* $(\Omega, \mathscr{F}, \mathbb{P})$. *Then*

i.
$$\sup_\alpha \mathbb{E}[|X_\alpha|] < \infty,$$

ii.
$$\mathbb{P}[|X_\alpha| > N] \to 0 \quad \textit{as } N \to \infty, \text{ uniformly } \textit{in } \alpha.$$

iii.
$$\mathbb{E}[|X_\alpha| \mathbf{1}_\Lambda] \to 0 \quad \textit{as } \mathbb{P}[\Lambda] \to 0, \text{ uniformly } \textit{in } \alpha.$$

Conversely, either i and *iii or ii* and *iii implies uniform integrability.*

Theorem A.16 (de la Vallée-Poussin criterion)**.** *Let* K *be a subset of* L^1. *Suppose there is a positive function* ϕ *defined on* $[0,\infty[$ *such that* $\lim_{t\to\infty} t^{-1}\phi(t) = +\infty$ *and* $\sup_{X\in K} \mathbb{E}\left[\phi(|X|)\right] < \infty$. *Then* K *is uniformly integrable.*

Proof. Write $\lambda = \sup_{X\in K} \mathbb{E}[\phi \circ |X|]$ and fix $\varepsilon > 0$. Put $a = \varepsilon^{-1}\lambda$ and choose c large enough that $t^{-1}\phi(t) \geq a$ if $t \geq c$. Then, on the set $\{|X| \geq c\}$, we have

$$|X| \leq a^{-1}(\phi \circ |X|),$$

so

$$\int_{\{|X|\geq c\}} |X(\omega)|\, dP(\omega) \leq a^{-1} \int_{\{|X|\geq c\}} (\phi \circ |X|)\, dP \leq a^{-1}\mathbb{E}\left[\phi \circ |X|\right] \leq \varepsilon.$$

Therefore, K is uniformly integrable. □

Remark A.17. *A common application example of the above result is when* $\phi(x) = x^p$, *for* $p > 1$. *Then if* K *is a subset of* L^p *with* $\sup_{X\in K} E[X^p] < \infty$, *we know* K *is uniformly integrable.*

The power of uniform integrability is given by the following generalization of the dominated convergence theorem

Theorem A.18 (Vitali convergence theorem)**.** *Suppose* $\{X_n\}_{n\in\mathbb{N}}$ *is a sequence of integrable random variables which converge in probability to a random variable* X. *Then the following are equivalent:*

(i) X_n *converges to* X *in* L^1 *(that is,* $\|X_n - X\|_1 = \mathbb{E}[|X_n - X_\infty|] \to 0$*),*

(ii) the collection $K = \{X_n\}_{n\in\mathbb{N}}$ *is uniformly integrable.*

In either case, the limit X *is also integrable.*

Proof. (i $\Rightarrow$ ii) Suppose $X_n \to X$ in the norm of L^1, so that X itself is in L^1. For any n, $\mathbb{E}[|X_n|] \le \|X_n - X\|_1 + \|X\|_1$, and we see that the expectations $\mathbb{E}\big[|X_n|\big] = \mathbb{E}[|X_n|]$ are uniformly bounded.

For any $\varepsilon > 0$, let N be such that

$$\|X_n - X\|_1 < \varepsilon/3$$

for all $n \ge N$. For any $n \ge N$ and any set $A \in \mathbb{F}$, this implies

$$\mathbb{E}[|X_n|\mathbf{1}_A] < \mathbb{E}[|X|I_A] + \|X_n - X\|_1 < \mathbb{E}[|X|\mathbf{1}_A] + \varepsilon/3.$$

For any $n < N$ and any set $A \in \mathbb{F}$,

$$\begin{aligned}\mathbb{E}[|X_n|\mathbf{1}_A] &\le \mathbb{E}[|X|\mathbf{1}_A] + \mathbb{E}[|X_n - X_N|\mathbf{1}_A] + \|X_N - X\|_1 \\ &< \mathbb{E}[|X|\mathbf{1}_A] + \mathbb{E}[|X_n - X_N|\mathbf{1}_A] + \varepsilon/3\end{aligned}$$

As X is integrable, we can find a $\delta_\infty > 0$ such that $\mathbb{E}[|X|\mathbf{1}_A] < \varepsilon/3$ whenever $\mathbb{P}(A) \le \delta_\infty$. Similarly, for each $n \le N$ we can find a $\delta_n > 0$ such that $\mathbb{E}[|X_n - X_N|\mathbf{1}_A] < \varepsilon/3$ whenever $\mathbb{P}(A) \le \delta_n$. Let $\delta = \delta_\infty \wedge \min_{n \le N} \delta_n$. Then, whenever $P(A) < \delta$, we have $\mathbb{E}[|X_n|\mathbf{1}_A] < \varepsilon$. By our proposition, this shows that $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable.

(ii $\Rightarrow$ i) Conversely, suppose the set $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable. Then the set of expectations $\mathbb{E}[|X_n|]$ is bounded and so, by Fatou's inequality applied to an almost surely converging subsequence,

$$\mathbb{E}[|X|] = \mathbb{E}[\lim_n |X_n|] \le \liminf_n \mathbb{E}[|X_n|] < \infty.$$

Now, writing

$$X_c = \mathbf{1}_{\{|X| \ge c\}} X, \qquad X^c = X - X_c = \mathbf{1}_{\{|X| < c\}} X$$

we have

$$\|X_n - X\|_1 \le \|(X_n)^c - X^c\|_1 + \|(X_n)_c\|_1 + \|X_c\|_1.$$

Fix $\varepsilon > 0$. Because the collection $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable, there exists a number $c > 0$ such that $\|X_c\|_1 < \varepsilon/3$ and $\|(X_n)_c\|_1 < \varepsilon/3$ for all n. We know that X_n^c converges to X^c in probability and $|X_n^c - X^c| \le 2c$ so by Lebesgue's dominated convergence theorem, $\lim_n \|X_n^c - X^c\|_1 = 0$.

There is, therefore, an integer N such that $\|X_n^c - X^c\|_1 \le \varepsilon/3$ if $n > N$. Consequently, if $n > N$, we have $\|X_n - X\| < \varepsilon$, and $X_n \to X$ in L^1. Because

$$\big|\,\mathbb{E}[|X_n|] - \|X\|_1\,\big| \le \|X_n - X\|_1,$$

$\mathbb{E}\big[|X_n|\big]$ converges to $\mathbb{E}\big[|X|\big]$. □

Uniform integrability is particularly useful in martingale theory. An example of this is the following result:

Lemma A.19. *Let X be a supermartingale with respect to the filtration $\{\mathscr{F}_t\}_{t\geq 0}$. Let s_n be a nonincreasing sequence in $[0,T]$. Then $\{X_{s_n}\}_{n\in\mathbb{N}}$ is uniformly integrable.*

Proof. As $\mathbb{E}[X_{s_n}]$ is an increasing function of n, the limit $\alpha = \lim_n \mathbb{E}[X_{s_n}]$ exists. As $\mathbb{E}[X_{s_n}] \leq \mathbb{E}[X_0]$ for all n, $\alpha < \infty$. For any $\varepsilon > 0$, there exists an integer k such that $\alpha - \mathbb{E}[X_{s_k}] < \varepsilon/2$ and so, for all $n \geq k$,

$$0 \leq \mathbb{E}[X_{s_n}] - \mathbb{E}[X_{s_k}] \leq \varepsilon/2.$$

Consider any $\lambda > 0$ and suppose $n \geq k$. Then

$$\begin{aligned}
I(n,\lambda) &:= \mathbb{E}[\mathbf{1}_{\{|X_{s_n}|>\lambda\}}|X_{S_n}|] \\
&= \mathbb{E}[\mathbf{1}_{\{X_{s_n}<-\lambda\}}(-X_{S_n})] + \mathbb{E}[\mathbf{1}_{\{X_{s_n}>\lambda\}}X_{S_n}] \\
&= -\mathbb{E}[\mathbf{1}_{\{X_{s_n}<-\lambda\}}X_{S_n}] + \mathbb{E}[X_{s_n}] - \mathbb{E}[\mathbf{1}_{\{X_{s_n}\leq\lambda\}}X_{s_n}] \\
&\leq -\mathbb{E}[\mathbf{1}_{\{X_{s_n}<-\lambda\}}X_{s_n}] + \mathbb{E}[X_{S_k}] - \mathbb{E}[\mathbf{1}_{\{X_{s_n}\leq\lambda\}}X_{s_n}] + \varepsilon/2.
\end{aligned}$$

As $X_{s_n} \geq \mathbb{E}[X_{s_k}|\mathscr{F}_{s_n}]$, we have that

$$\begin{aligned}
&\mathbb{E}[\mathbf{1}_{\{X_{s_n}\leq\lambda\}}X_{s_n}] + \mathbb{E}[\mathbf{1}_{\{X_{s_n}<-\lambda\}}X_{s_n}] \\
&\geq \mathbb{E}[\mathbf{1}_{\{X_{s_n}\leq\lambda\}}X_{S_k}] + \mathbb{E}[\mathbf{1}_{\{X_{s_n}<-\lambda\}}X_{s_k}]
\end{aligned}$$

and so, by rearrangement,

$$I(n,\lambda) \leq \mathbb{E}[\mathbf{1}_{\{|X_{s_n}|>\lambda\}}|X_{s_k}|] + \varepsilon/2.$$

By Jensen's inequality, we also have that

$$\mathbb{E}\big[|X_{s_n}|\big] = \mathbb{E}[X_{s_n}] + 2\mathbb{E}[X^-_{s_n}] \leq \alpha + 2\mathbb{E}[X^-_{s_0}] =: \beta$$

and by Markov's inequality,

$$P(|X_{s_n}| > \lambda) \leq \frac{\mathbb{E}\big[|X_{s_n}|\big]}{\lambda} \leq \frac{\beta}{\lambda}.$$

By absolute continuity of the measure $\nu(A) = \mathbb{E}[\mathbf{1}_A|X_{s_k}|]$ with respect to $\mathbb{P}$, we know that there exists a λ_0 such that

$$\mathbb{E}[\mathbf{1}_{\{|X_{s_n}|>\lambda\}}|X_{s_k}|] \leq \varepsilon/2$$

for all $\lambda \geq \lambda_0$, $n \geq k$. That is, $I(n,k) \leq \varepsilon$ for all $\lambda \geq \lambda_0$, $n \geq k$. For $n < k$, there exists a λ_1 such that $I(n,k) \leq \varepsilon$ whenever $n < k$, $\lambda \geq \lambda_1$, and so if $\lambda \geq \lambda_0 \vee \lambda_1$, we observe the requirement for uniform integrability. □

Remark A.20. *Given the optional stopping theorem, you can easily check the above works for a nonincreasing sequence of stopping times also.*

A.6 Doob's inequalities

In discrete time, Doob showed that martingales had strong boundedness properties. You may have seen a version of these in an earlier course – the proofs of the versions in these notes are presented here for completeness.

Lemma A.21. *Suppose $\{X_n\}_{n\in\mathbb{Z}^+}$, is a supermartingale. For every $\alpha \geq 0$*

$$\alpha\mathbb{P}\big(\sup_n X_n \geq \alpha\big) \leq \mathbb{E}[X_0] + \sup_n \mathbb{E}[X_n^-] \leq 2\sup_n \mathbb{E}[|X_n|].$$

Proof. Put $T(\omega) = \min\{n : X_n \geq \alpha\}$ and define a sequence of stopping times $\{T_k = T\wedge k\}_{k\in\mathbb{Z}^+}$. By the optional stopping theorem, for each k, $\mathbb{E}[X_{T_k}] \leq \mathbb{E}[X_0]$. Either

$$X_{T_k}(\omega) \geq \alpha \quad \text{or} \quad X_{T_k}(\omega) = X_k(\omega),$$

therefore,

$$\alpha\mathbb{P}\big(\sup_{n\leq k} X_n \geq \alpha\big) + \int_{\{\sup_{n\leq k} X_n < \alpha\}} X_k\, dP \leq \mathbb{E}[X_{T_k}] \leq \mathbb{E}[X_0].$$

As $\alpha \geq 0$, we know $\{X_k < 0\} \subseteq \{X_k < \alpha\} \subseteq \{\sup_{n\leq k} X_n < \alpha\}$. Hence

$$-\mathbb{E}[X_k^-] = \int_{\{X_k \leq 0\}} X_k\, dP \leq \int_{\{\sup_{n\leq k} X_n < \alpha\}} X_k\, dP,$$

and so

$$\alpha\mathbb{P}\big(\sup_{n\leq k} X_n \geq \alpha\big) \leq E[X_0] + E[X_k^-].$$

Letting $k \to \infty$ we have

$$\alpha\mathbb{P}\big(\sup_n X_n \geq \alpha\big) \leq \mathbb{E}[X_0] + \sup_n \mathbb{E}[X_n^-] \leq 2\sup_n \mathbb{E}[|X_n|].$$

□

Lemma A.22. *Suppose $\{X_n\}_{n\in\mathbb{Z}^+}$, is a supermartingale. For every $\alpha \geq 0$*

$$\alpha\mathbb{P}\big(\inf_n X_n \leq -\alpha\big) \leq \sup_n \mathbb{E}[X_n^-].$$

Proof. Put $S(\omega) = \min\{n : X_n(\omega) \leq -\alpha\}$ and define a sequence of stopping times $\{S_k := S\wedge k\}_{k\in\mathbb{Z}^+}$. By optional stopping, similarly as in the previous lemma $\mathbb{E}[X_{S_k}] \geq \mathbb{E}[X_k]$ for every $k \in \mathbb{Z}^+$. Therefore,

$$\mathbb{E}[X_k] \leq -\alpha\mathbb{P}\big(\inf_{n\leq k} X_n \leq -\alpha\big) + \int_{\{\inf_{n\leq k} X_n > -\alpha\}} X_k\, d\mathbb{P},$$

so

$$\begin{aligned}\alpha\mathbb{P}\big(\inf_{n\leq k} X_n \leq -\alpha\big) &\leq \mathbb{E}[-X_k] + \int_{\{\inf_{n\leq k} X_n > -\alpha\}} X_k\, d\mathbb{P} \\ &= \int_{\{\inf_{n\leq k} X_n \leq -\alpha\}} (-X_k)\, d\mathbb{P} \qquad (41) \\ &\leq \mathbb{E}[X_k^-].\end{aligned}$$

Letting $k \to \infty$, the result follows. □ □

Corollary A.23. *Suppose $\{X_n\}_{n\in\mathbb{Z}^+}$ is a supermartingale. For every $\alpha \geq 0$,*

$$\alpha\mathbb{P}\big(\sup_n |X_n| \geq \alpha\big) \leq 3\sup_n \mathbb{E}[|X_n|].$$

Corollary A.24 (Doob's Maximal Inequality)**.** *Suppose $\{X_n\}_{n\in\mathbb{Z}^+}$ is a martingale and $p \geq 1$. For every $\alpha \leq 0$,*

$$\alpha^p P\big(\sup_n |X_n| \geq \alpha\big) \leq \sup_n \mathbb{E}[|X_n|^p].$$

Proof. From Jensen's inequality, if $Y_n = -|X_n|^p$, then Y is a (negative) supermartingale (provided X^p is integrable) and

$$\mathbb{E}[|Y_n|] = \mathbb{E}[|X_n|^p] = E[Y_n^-].$$

Also

$$\big\{\inf_n Y_n \leq -\alpha^p\big\} = \big\{\sup_n |X_n| \geq \alpha\big\},$$

so the result follows from Lemma A.22. □ □

Lemma A.25. *Suppose X and Y are two nonnegative random variables such that $X \in L^p$ for some $p > 1$, and for every $\alpha > 0$, $\alpha\mathbb{P}(Y \geq \alpha) \leq \int_{\{Y\geq\alpha\}} X\,d\mathbb{P}$. Then $\mathbb{E}[|Y|^p] \leq q^p\mathbb{E}[|X|^p]$, where $p^{-1}+q^{-1} = 1$.*

Proof. Let $\tilde{F}(\lambda) = 1 - F_Y(\lambda) = P(Y > \lambda)$ where F_Y is the distribution function of Y. As λ^p is continuous, integration by parts yields,

$$\begin{aligned}
E[Y^p] &= -\int_{[0,\infty]} \lambda^p\,dF(\lambda) \\
&= \int_{[0,\infty]} \tilde{F}(\lambda)\,d(\lambda^p) - \lim_{h\to\infty}[\lambda^p\tilde{F}(\lambda)]_0^h \\
&\leq \int_{[0,\infty]} \tilde{F}(\lambda)\,d(\lambda^p) \\
&\leq \int_{[0,\infty]} \lambda^{-1}\Big(\int_{\{Y\geq\lambda\}} X\,dP\Big)\,d(\lambda^p) && \text{by hypothesis} \\
&= E\Big[X\int_{[0,Y]} \lambda^{-1}\,d(\lambda^p)\Big] && \text{by Fubini's theorem} \\
&= \Big(\frac{p}{p-1}\Big)\mathbb{E}[XY^{p-1}] \\
&\leq q\mathbb{E}[|X|^p]^{1/p}\mathbb{E}[|Y^{p-1}|^q]^{1/q} && \text{by Hölder's inequality.}
\end{aligned}$$

We have, therefore, proved that

$$\mathbb{E}[Y^p] \leq q\mathbb{E}[|X|^p]^{1/p}(E[|Y|^{pq-q}])^{\frac{1}{q}} = q\mathbb{E}[|X|^p]^{1/p}(E[|Y|^p])^{\frac{1}{q}}.$$

If $\mathbb{E}[|Y|^p]$ is finite, as $1-q^{-1} = p^{-1}$ the inequality follows immediately. Otherwise, the random variable $Y_n := Y \wedge n$ satisfies the hypotheses and is in L^p for every n. Therefore

$$\mathbb{E}[|Y_n|^p] \leq q^p\mathbb{E}[|X|^p],$$

and the result follows by letting $n \to \infty$ and monotone convergence. □

Theorem A.26 (Doob's L^p inequality)**.** *Suppose X is a martingale or nonnegative submartingale. Then for $1 < p \leq \infty$, we have*

$$\sup_n |X_n| \in L^p \text{ if and only if } \sup_n \mathbb{E}[|X_n|^p] < \infty.$$

Furthermore, for $p > 1$ and $p^{-1} + q^{-1} = 1$ we have

$$E[\sup_n |X_n|^p] \leq q^p \sup_n \mathbb{E}[|X_n|^p].$$

Proof. When $p = \infty$ the first part of the theorem is immediate. Clearly, for $1 < p \leq \infty$ if $\sup_n |X_n| \in L^p$ then $\sup_n \mathbb{E}[|X_n|^p] \leq \mathbb{E}[\sup_n |X_n|^p] < \infty$.

To show the converse inequality, we know by assumption that

$$\sup_n \mathbb{E}[(-X_n)^-] = \sup_n \mathbb{E}[X_n^+] \leq \sup_n \mathbb{E}\big[|X|^p\big] < \infty$$

and from the supermartingale convergence theorem applied to the supermartingale $-X$ we know $\lim_{n\to\infty} X_n(\omega) = X_\infty(\omega)$ exists and is integrable. By Fatou's lemma

$$\begin{aligned}\mathbb{E}\big[|X_\infty|^p\big] &= \mathbb{E}\big[|\lim_n X_n|^p\big] = \mathbb{E}\big[\lim_n |X_n|^p\big] \\ &\leq \liminf_n \mathbb{E}\big[|X_n|^p\big] \leq \sup_n \mathbb{E}\big[|X_n|^p\big] < \infty,\end{aligned}$$

so $X_\infty \in L^p$ and $E[|X_\infty|^p] \leq \sup_n E[|X_n|^p]$.

Using the calculation in (41), as $-|X|$ is a supermartingale, for any $\alpha > 0$ we have

$$\begin{aligned}\alpha\mathbb{P}\big(\sup_{n\leq k} |X_n| \geq \alpha\big) &= \alpha\mathbb{P}\big(\inf_{n\leq k}(-|X_n|) \leq -\alpha\big) \\ &\leq \int_{\{\sup_{n\leq k} |X_n| \geq \alpha\}} X_k \, d\mathbb{P} \\ &\leq \int_{\{\sup_n |X_n| \geq \alpha\}} X_k^+ \, d\mathbb{P}.\end{aligned}$$

Letting $k \to \infty$, as $X_k^+ \leq \sup_n |X_n|$ which is integrable, by dominated convergence we have that for any $\alpha > 0$

$$\alpha\mathbb{P}(\sup_n |X_n| \geq \alpha) \leq \lim_k \int_{\{\sup_n |X_n| \geq \alpha\}} X_k^+ \, d\mathbb{P} = \int_{\{\sup_n |X_n| \geq \alpha\}} X_\infty^+ \, d\mathbb{P}.$$

Consequently, we can apply Lemma A.25 with $Y = \sup_n |X_n|$ and $X = X_\infty^+$ to deduce that

$$\mathbb{E}[\sup_n |X_n|^p] \leq q^p \mathbb{E}[(X_\infty^+)^p] \leq q^p \mathbb{E}[|X_\infty|^p].$$

□

A.7 Information and independence.

Information obtained by observations of the outcome of a random experiment is represented by a sub-σ-algebra $\mathscr{D}$ of the collection of events $\mathscr{F}$. If $D \in \mathscr{D}$, then the oberver "knows" whether or not the outcome is in D.

An S-valued random variable Y is *independent* of a σ-algebra $\mathscr{D}$ if

$$P(\{Y \in B\} \cap D) = P(\{Y \in B\})P(D), \quad \forall B \in \mathscr{B}(S), D \in \mathscr{D}.$$

Two σ-algebras $\mathscr{D}_1, \mathscr{D}_2$ are independent if

$$P(D_1 \cap D_2) = P(D_1)P(D_2), \quad \forall D_1 \in \mathscr{D}_1, D_2 \in \mathscr{D}_2.$$

Two S-valued random variables X and Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent, that is, if

$$P(\{X \in B_1\} \cap \{Y \in B_2\}) = P(\{X \in B_1\})P(\{Y \in B_2\}), \quad \forall B_1, B_2 \in \mathscr{B}(S).$$

A.8 Conditional expectation.

Interpretation of conditional expectation in L_2.

Problem: Approximate $X \in L_2$ using information represented by $\mathscr{D}$ such that the mean square error is minimized, i.e., find the $\mathscr{D}$-measurable random variable Y that minimizes $E[(X-Y)^2]$.

Solution: Suppose Y is a minimizer. For any $\varepsilon \neq 0$ and any $\mathscr{D}$-measurable random variable $Z \in L_2$

$$E[|X-Y|^2] \le E[|X-Y-\varepsilon Z|^2] = E[|X-Y|^2] - 2\varepsilon E[Z(X-Y)] + \varepsilon^2 E[Z^2].$$

Hence $2\varepsilon E[Z(X-Y)] \le \varepsilon^2 E[Z^2]$. Since ε is arbitrary, $E[Z(X-Y)] = 0$ and hence

$$E[ZX] = E[ZY] \tag{42}$$

for every $\mathscr{D}$-measurable Z with $E[Z^2] < \infty$. □

With (42) in mind:

Definition A.27. *For an integrable random variable X, the* conditional expectation *of X, denoted $E[X|\mathscr{D}]$, is the unique (up to changes on events of probability zero) random variable Y satisfying*

A) Y is $\mathscr{D}$-measurable.

B) $\int_D X dP = \int_D Y dP$ for all $D \in \mathscr{D}$.

Note that Condition B is a special case of (42) with $Z = I_D$ (where I_D denotes the indicator function for the event D) and that Condition B implies that (42) holds for all bounded $\mathscr{D}$-measurable random variables. Existence of conditional expectations is a consequence of the Radon–Nikodym theorem.

Lemma A.28 (Conditional Jensen Inequality)**.** *Suppose ϕ is a convex map of $\mathbb{R}$ into $\mathbb{R}$ and suppose X is an integrable random variable such that $\phi(X)$ is integrable. Then*

$$\phi\left(\mathbb{E}[X|\mathscr{D}]\right) \leq \mathbb{E}[\phi(X)|\mathscr{D}] \quad a.s.$$

Proof. As ϕ is convex, it is the upper envelope of a countable family of affine functions

$$\lambda_n(x) = \alpha_n x + \beta_n, \quad x \in \mathbb{R}, \quad n \in \mathbb{N},$$

that is, $\phi(x) = \sup_n\{\lambda_n(x)\}$. Note that this implies ϕ is Borel measurable. The random variables $\lambda_n(X)$ are integrable and

$$\lambda_n(\mathbb{E}[X|\mathscr{D}]) = \mathbb{E}[\lambda_n(X)|\mathscr{D}] \leq \mathbb{E}[\phi(X)|\mathscr{D}] \quad \text{a.s.}$$

Taking the supremum with respect to n, the result follows. □

Lemma A.29 (Continuity of conditional expectation)**.** *For any σ-algebra $\mathscr{D}$, the conditional expectation $\mathbb{E}[\cdot|\mathscr{D}]$ is continuous in L^1, that is, if $X_n \to X$ in L^1, then $\mathbb{E}[X_n|\mathscr{D}] \to \mathbb{E}[X|\mathscr{D}]$ in L^1.*

Proof. By Jensen,

$$\begin{aligned}\mathbb{E}\Big[\Big|\mathbb{E}[X_n|\mathscr{D}] - \mathbb{E}[X|\mathscr{D}]\Big|\Big] &= \mathbb{E}\Big[\Big|\mathbb{E}[X_n - X|\mathscr{D}]\Big|\Big] \\ \leq \mathbb{E}\Big[\mathbb{E}\big[|X_n - X|\big|\mathscr{D}\big]\Big] &= \mathbb{E}[|X_n - X|] \to 0.\end{aligned}$$

□

B An overview of Gaussian variables

There are a variety of ways for us to look at Brownian motion. One good way to start[2] is as a extension to 'processes' of Gaussian random variables.

Brownian motion is a special example of a Gaussian process – or at least a version of one that is assumed to have continuous sample paths. In this section we give an overview of Gaussian variables.

B.1 Gaussian variables in one dimension

Definition B.1. *A random variable X is called a* centred standard Gaussian*, or* standard normal*, if its distribution has density*

$$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad x \in \mathbb{R}$$

with respect to Lebesgue measure. We write $X \sim \mathcal{N}(0,1)$.

[2]This is effectively how Einstein thought of it in his original 1905 paper.

It is elementary to calculate its moment generating function (Laplace transform):

$$\mathbb{E}[e^{\lambda X}] = e^{\frac{\lambda^2}{2}},\ \lambda \in \mathbb{R},$$

or extending to complex values the characteristic function (Fourier transform)

$$\mathbb{E}[e^{i\xi X}] = e^{\frac{-\xi^2}{2}},\ \xi \in \mathbb{R}.$$

We say Y has Gaussian (or normal) distribution with mean m and variance σ^2, written $Y \sim \mathscr{N}(m, \sigma^2)$, if $Y = \sigma X + m$ where $X \sim \mathscr{N}(0,1)$. Then

$$\mathbb{E}[e^{i\xi Y}] = \exp\left(im\xi - \frac{\sigma^2\xi^2}{2}\right), \quad \xi \in \mathbb{R},$$

and if $\sigma > 0$, the density on $\mathbb{R}$ is

$$p_Y(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

We think of a *constant* 'random' variable as being a *degenerate Gaussian*. Then the space of Gaussian variables (resp. distributions) is closed under convergence in probability (resp. distribution).

Proposition B.2. *Let (X_n) be a sequence of Gaussian random variables with $X_n \sim \mathscr{N}(m_n, \sigma_n^2)$, which converges in distribution to a random variable X. Then*

(i) X is also Gaussian, $X \sim \mathscr{N}(m, \sigma^2)$ with $m = \lim_{n\to\infty} m_n$, $\sigma^2 = \lim_{n\to\infty} \sigma_n^2$; and

(ii) if $(X_n)_{n\geq 1}$ converges to X in probability, then the convergence is also in L^p for all $1 \leq p < \infty$.

Proof. Convergence in distribution is equivalent to saying that the characteristic functions converge:

$$\mathbb{E}\left[e^{\mathrm{i}\xi X_n}\right] = \exp(\mathrm{i}m_n\xi - \sigma_n^2\xi^2/2) \longrightarrow \mathbb{E}\left[e^{\mathrm{i}\xi X}\right], \quad \xi \in \mathbb{R}. \tag{43}$$

Taking the modulus we see that the sequence $\exp(-\sigma_n^2\xi^2/2)$ converges, which in turn implies that $\sigma_n^2 \to \sigma^2 \in [0,\infty)$ (where we ruled out the case $\sigma_n \to \infty$ since the limit has to be the modulus of a characteristic function and so, in particular, has to be continuous). We deduce that

$$e^{\mathrm{i}m_n\xi} \longrightarrow e^{\frac{1}{2}\sigma^2\xi^2}\mathbb{E}\left[e^{\mathrm{i}\xi X}\right], \quad \xi \in \mathbb{R}.$$

We now argue that this implies that the sequence m_n converges to some finite m. Suppose first that the sequence $\{m_n\}_{n\geq 1}$ is bounded, and consider any two convergent subsequences, converging to m and m' say. Then rearranging (43) yields $m = m'$ and so the sequence converges.

Now suppose that $\limsup_{n\to\infty} m_n = \infty$ (the case $\liminf_{n\to\infty} m_n = -\infty$ is similar). There exists a subsequence $\{m_{n_k}\}_{k\geq 1}$ which tends to infinity. Given M, for k large enough that $m_{n_k} > M$,

$$\mathbb{P}[X_{n_k} \geq M] \geq \mathbb{P}[X_{n_k} \geq m_{n_k}] = \frac{1}{2},$$

and so using convergence in distribution

$$\mathbb{P}[X \geq M] \geq \limsup_{k\to\infty} \mathbb{P}[X_{n_k} \geq M] \geq \frac{1}{2}, \quad \text{for any } M > 0.$$

This is clearly impossible for any fixed (real valued) random variable X and gives us the desired contradiction. This completes the proof of (i).

To show (ii), observe that the convergence of σ_n and m_n implies, in particular, that

$$\sup_n \mathbb{E}\left[e^{\theta X_n}\right] = \sup_n e^{\theta m_n + \theta^2 \sigma_n^2/2} < \infty, \quad \text{for any } \theta \in \mathbb{R}.$$

Since $\exp(|x|) \leq \exp(x) + \exp(-x)$, this remains finite if we take $|X_n|$ instead of X_n. This implies that $\sup_n \mathbb{E}[|X_n|^p] < \infty$ for any $p \geq 1$ and hence also

$$\sup_n \mathbb{E}[|X_n - X|^p] < \infty, \quad \forall p \geq 1. \tag{44}$$

Fix $p \geq 1$. Then the sequence $|X_n - X|^p$ converges to zero in probability (by assumption) and is uniformly integrable, since $\sup_n \mathbb{E}[(|X_n - X|^p)^q] < \infty$ for some (in fact all) $q > 1$. It follows that we also have convergence of X_n to X in L^p. □

B.2 Gaussian vectors

So far we have considered only real-valued Gaussian variables.

Definition B.3. *A random vector taking values in $\mathbb{R}^d$ is called* Gaussian *if and only if*

$$\langle u, X\rangle := u^{\mathrm{T}} X = \sum_{i=1}^{d} u_i X_i \quad \textit{is Gaussian for all } u \in \mathbb{R}^d.$$

It follows immediately that the image of a Gaussian vector under a linear transformation is also Gaussian: if $X \in \mathbb{R}^d$ is Gaussian and A is an $m \times d$ matrix, then AX is Gaussian in $\mathbb{R}^m$.

Lemma B.4. *Let X be a Gaussian vector and define $m_X := (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])$ and $\Gamma_X := \left(\mathrm{cov}(X_i, X_j)_{1\leq i,j\leq d}\right)$, the mean vector and the covariance matrix respectively. Then $q_X(u) := u^{\mathrm{T}} \Gamma_X u$ is a non-negative quadratic form and*

$$u^{\mathrm{T}} X \sim \mathcal{N}\left(u^{\mathrm{T}} m_X, q_X(u)\right), \quad u \in \mathbb{R}^d. \tag{45}$$

Proof. Clearly $\mathbb{E}[u^{\mathrm{T}}X] = u^{\mathrm{T}}m_X$ and

$$\mathrm{var}(u^{\mathrm{T}}X) = \mathbb{E}\left[\left(\sum_{i=1}^{d} u_i(X_i - \mathbb{E}[X_i])\right)^2\right] = \sum_{1\le i,j\le d} u_i u_j \mathrm{cov}(X_i, X_j) = u^{\mathrm{T}}\Gamma_X u = q_X(u),$$

which also shows that $q_X(u) \ge 0$. □

The identification in (45) is equivalent to

$$\mathbb{E}\left[e^{\mathrm{i}\langle u,X\rangle}\right] = e^{\mathrm{i}\langle u,m_X\rangle - \frac{1}{2}q_X(u)}, \quad u \in \mathbb{R}^d.$$

From this we derive easily the following important fact:

Proposition B.5. *Let X be a Gaussian vector and Γ_X its covariance matrix. Then $X_1,\dots,X_d$ are independent if and only if Γ_X is a diagonal matrix (i.e. the variables are pairwise uncorrelated).*

Proof. If (and only if) Γ_X is diagonal, we can write

$$\mathbb{E}\left[e^{\mathrm{i}\langle u,X\rangle}\right] = \prod_k e^{\mathrm{i}u_k m_k - \frac{1}{2}\xi_k^2\Gamma_{kk}} = \prod_k \mathbb{E}\left[e^{\mathrm{i}u_k X_k}\right].$$

As the characteristic function of a vector can be written as a product of the characteristics of its components if and only if the components are independent, we are done. □

Warning: It is crucial to assume that the vector X is Gaussian and not just that $X_1,\dots,X_d$ are Gaussian. For example, consider $X_1 \sim \mathcal{N}(0,1)$ and ε an independent random variable with $\mathbb{P}[\varepsilon = 1] = 1/2 = \mathbb{P}[\varepsilon = -1]$. Let $X_2 := \varepsilon X_1$. Then $X_2 \sim \mathcal{N}(0,1)$ and $\mathrm{cov}(X_1,X_2) = 0$, while clearly X_1, X_2 are *not* independent.

By definition, a Gaussian vector X remains Gaussian if we add to it a deterministic vector $m \in \mathbb{R}^d$. Hence, without loss of generality, by considering $X - m_X$, it suffices to consider centred Gaussian vectors. The variance-covariance matrix Γ_X is symmetric and non-negative definite (as observed above). Conversely, for any such matrix Γ, there exists a Gaussian vector X with $\Gamma_X = \Gamma$, and, indeed, we can construct it as a linear transformation of a Gaussian vector with i.i.d. coordinates.

Theorem B.6. *Let Γ be a symmetric non-negative definite $d \times d$ matrix. Let $(\varepsilon_1,\dots,\varepsilon_d)$ be an orthonormal basis in $\mathbb{R}^d$ which diagonalises Γ, i.e. $\Gamma\varepsilon_i = \lambda_i\varepsilon_i$ for some $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_r > 0 = \lambda_{r+1} = \ldots = \lambda_d$, where $1 \le r \le d$ is the rank of Γ.*

(i) A centred Gaussian vector X with covariance matrix $\Gamma_X = \Gamma$ exists.

(ii) Further, any such vector can be represented as

$$X = \sum_{i=1}^{r} Y_i \varepsilon_i, \tag{46}$$

where $Y_1,\dots,Y_r$ are independent Gaussian random variables with $Y_i \sim \mathcal{N}(0,\lambda_i)$.

(iii) If $r = d$, then X admits a density given by

$$p_X(x) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det(\Gamma)}} \exp\Big(-\frac{1}{2} x^T \Gamma^{-1} x \Big), \quad x \in \mathbb{R}^d.$$

Proof. Let A be a matrix whose columns are ε_i so that $\Gamma_X = A\Lambda A^T$ where Λ is the diagonal matrix with entries λ_i on the diagonal. Let $Z_1, \ldots, Z_n$ be i.i.d. standard centred Gaussian variables and $Y_i = \sqrt{\lambda_i} Z_i$. Let X be given by (46), i.e. $X = AY$. Then

$$\langle u, X \rangle = \langle u, AY \rangle = \langle A^T u, Y \rangle = \sum_{i=1}^{d} \sqrt{\lambda_i} \langle u, \varepsilon_i \rangle Z_i$$

is centred Gaussian. Its variance is given by

$$\mathrm{var}(\langle u, X \rangle) = \sum_{i=1}^{d} \lambda_i (u^{\mathrm{T}} \varepsilon_i)^2 = \sum_{i=1}^{d} u^{\mathrm{T}} \varepsilon_i \lambda_i \varepsilon_i^{\mathrm{T}} u = (A^T u)^T \Lambda A^T u = u^T A \Lambda A^T u = u^T \Gamma u,$$

and (i) is proved.
Conversely, suppose X is a centred Gaussian vector with covariance matrix Γ and let $Y = A^T X$. For $u \in \mathbb{R}^d$, $\langle u, Y \rangle = \langle Au, X \rangle$ is centred Gaussian with variance

$$(Au)^T \Gamma A u = u^T A^T \Gamma A u = u^T \Lambda u$$

and we conclude that Y is also a centred Gaussian vector with covariance matrix Λ. Independence between $Y_1, \ldots, Y_d$ then follows from Proposition B.5. It follows that when $r = d$, Y admits a density on $\mathbb{R}^d$ given by

$$p_Y(y) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det(\Lambda)}} \exp\Big(-\frac{1}{2} y^T \Lambda^{-1} y \Big), \quad y \in \mathbb{R}^d.$$

Change of variables, together with $\det(\Lambda) = \det(\Gamma)$ and $|\det(A)| = 1$ gives the desired density for x. □

Once we know that we can write $X = \sum_{i=1}^{r} Y_i \varepsilon_i$ in this way, we have an easy way to compute conditional expectations within the family of random variables which are linear transformations of a Gaussian vector X. To see how it works, suppose that X is a Gaussian vector in $\mathbb{R}^d$ and define $Z := (X_1 - \sum_{i=2}^{d} a_i X_i)$ with the coefficients a_i chosen in such a way that Z and X_i are uncorrelated for $i = 2, \ldots, d$; that is

$$\mathrm{cov}(X_1, X_i) - \sum_{j=2}^{d} a_j \mathrm{cov}(X_j, X_i) = 0, \; 2 \le i \le d.$$

Evidently Z is Gaussian (it is a linear combination of Gaussians) and since it is uncorrelated with $X_2, \ldots, X_d$, by Proposition B.5, it is independent of them. Then

$$\begin{aligned}\mathbb{E}[X_1 | \sigma(X_2, \ldots, X_d)] &= \mathbb{E}\Big[Z + \sum_{i=2}^{d} a_i X_i \Big| \sigma(X_2, \ldots, X_d) \Big] \\ &= \mathbb{E}[Z | \sigma(X_2, \ldots, X_d)] + \mathbb{E}\Big[\sum_{i=2}^{d} a_i X_i \Big| \sigma(X_2, \ldots, X_d) \Big] = \sum_{i=2}^{d} a_i X_i,\end{aligned}$$

where we have used independence to see that $\mathbb{E}[Z|\sigma(X_2,\ldots,X_d)] = \mathbb{E}[Z] = 0$.

The most striking feature is that $\mathbb{E}[X|\sigma(K)]$ is an element of the vector space K itself and not a general $\sigma(K)$-measurable random variable. In particular, if (X_1,X_2,X_3) is a Gaussian vector then the best (in the L^2 sense, see Appendix A.8) approximation of X_1 in terms of X_2,X_3 is in fact a *linear* function of X_1 and X_2. This extends to the more general setting of Gaussian spaces to which we now turn.

B.3 Gaussian spaces

Note that to a Gaussian vector X in $\mathbb{R}^d$, we can associate the vector space spanned by its coordinates:

$$\left\{\sum_{i=1}^{d} u_i X_i : u_i \in \mathbb{R}\right\},$$

and by definition all elements of this space are Gaussian random variables. This is a simple example of a *Gaussian space* and it is useful to think of such spaces in much greater generality.

Definition B.7. *A closed linear subspace $H \subset \mathrm{L}^2(\Omega,\mathscr{F},\mathbb{P})$ is called a* Gaussian space *if all of its elements are centred Gaussian random variables.*

In analogy to Proposition B.5, two elements of a Gaussian space are independent if and only if they are uncorrelated, which in turn is equivalent to being orthogonal in L^2. More generally we have the following result.

Theorem B.8. *Let H_1,H_2 be two Gaussian subspaces of a Gaussian space H. Then H_1,H_2 are orthogonal if and only if $\sigma(H_1)$ and $\sigma(H_2)$ are independent*[3].

The theorem follows from monotone class arguments, which (see Appendix A.1) reduce it to checking that it holds true for any finite subcollection of random variables – which is Proposition B.5.

Corollary B.9. *Let H be a Gaussian space and K a closed subspace. Let p_K denote the orthogonal projection onto K. Then for $X \in H$*

$$\mathbb{E}[X|\sigma(K)] = p_K(X). \tag{47}$$

Proof. Let $Y = X - p_K(X)$ which, by Theorem B.8 is independent of $\sigma(K)$. Hence

$$\mathbb{E}[X|\sigma(K)] = \mathbb{E}[p_K(X)|\sigma(K)] + \mathbb{E}[Y|\sigma(K)] = p_K(X) + \mathbb{E}[Y] = p_K(X),$$

where we have used that Y is a centred Gaussian and so has zero mean. □

Warning: For an arbitrary $X \in \mathrm{L}^2$ we would have

$$\mathbb{E}[X|\sigma(K)] = p_{\mathrm{L}^2(\Omega,\sigma(K),\mathbb{P})}(X).$$

It is a special property of Gaussian random variables X that it is enough to consider the projection onto the much smaller space K.

[3]Orthogonality here is in the L^2 sense, that is, $E[X_1X_2] = 0$ for all $X_1 \in H_1, X_2 \in H_2$. In other words, the spaces are made up of uncorrelated (and hence independent) Gaussian random variables.

B.4 Gaussian processes

Definition B.10. *A stochastic process $(X_t : t \geq 0)$ is called a (centred) Gaussian process if any finite linear combination of its coordinates is a (centred) Gaussian variable.*

Equivalently, X is a centred Gaussian process if for any $n \in \mathbb{N}$ and $0 \leq t_1 < t_2 < \ldots < t_n$, $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ is a (centred) Gaussian vector. It follows that the distribution of a centred Gaussian process (as a measure on $\mathscr{B}(\mathbb{R}^{[0,\infty)})$), is characterised by the covariance function, or *Kernel*, $\Gamma : [0,\infty)^2 \to \mathbb{R}$, i.e.

$$\Gamma(s,t) := \mathrm{cov}(X_t, X_s).$$

For any fixed n-tuple $(X_{t_1}, \ldots, X_{t_n})$ the covariance matrix $(\Gamma(t_i,t_j))$ has to be symmetric and positive semi-definite. As the following result shows, the converse also holds – for any such function, Γ, we may construct an associated Gaussian process.

Theorem B.11. *Let $\Gamma : [0,\infty)^2 \to \mathbb{R}$ be symmetric and such that for any $n \in \mathbb{N}$ and $0 \leq t_1 < t_2 < \ldots < t_n$,*

$$\sum_{1 \leq i,j, \leq n} u_i u_j \Gamma(t_i,t_j) \geq 0, \quad u \in \mathbb{R}^n.$$

Then there exists a centred Gaussian process with covariance function Γ.

This result will follow from the (more general) Daniell–Kolmogorov Theorem 2.8 below.

Recalling from Proposition B.2 that an L^2-limit of Gaussian variables is also Gaussian, we observe that the closed linear subspace of L^2 spanned by the variables $(X_t : t \geq 0)$ is a Gaussian space.

C Some useful measure theoretic results

C.1 Daniell–Kolmogorov Extension Theorem

NOT EXAMINABLE

Definition C.1. *Let $\mathbb{T} = [0,\infty]$ or $[0,\infty[$. Then $(\mathbb{R}^d)^{\mathbb{T}}$ denotes the space of real functions $x : \mathbb{T} \to \mathbb{R}^d$. The 'cylinder topology' on $(\mathbb{R}^d)^{\mathbb{T}}$ is given by finite intersections of sets of the form $\{x_t \in B\}$, where $t \in \mathbb{T}$ and B is an open set in $\mathbb{R}^d$. This in turn defines the Borel cylinder σ-algebra, denoted $\mathscr{B}((\mathbb{R}^d)^{\mathbb{T}})$.*

We write $\mathscr{B}((\mathbb{R}^d)^\infty)$ for the Borel σ-algebra on sequences in $\mathbb{R}^d$, which is given by the product $\otimes_{n\in\mathbb{N}}\mathscr{B}(\mathbb{R}^d)$.

Lemma C.2. *A set $A \in \mathscr{B}((\mathbb{R}^d)^{\mathbb{T}})$ if and only if there are a countable collection of points $t_1, t_2, \ldots$ and a Borel set $B \in \mathscr{B}((\mathbb{R}^d)^\infty)$ with*

$$A = \{x : (x_{t_1}, x_{t_2}, \ldots) \in B\}.$$

A set of this form where the collection of points $t_1, t_2 \ldots$ is finite is called a 'cylinder set'.

Proof. Exercise (monotone class argument). $\square$

Lemma C.3. *A set $A \in \mathscr{B}(\mathbb{R}^d)$ is* regular*, that is, for every finite measure μ on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ and every $\varepsilon > 0$ we can find a compact set C and an open set B such that $C \subseteq A \subseteq B$ and $\mu(B \setminus C) < \varepsilon$.*

Proof. We use a monotone class argument. The result is easy for a set of the form $A = \big(]a_1, b_1] \times \ldots \times]a_n, b_n]\big) \cap \mathbb{R}^d$ (where $a_i, b_i = \infty$ is permitted). These sets form an algebra generating $\mathscr{B}(\mathbb{R}^d)$ Now suppose $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of regular sets with associated $\{B_n\}_{n\in\mathbb{N}}, \{C_n\}_{n\in\mathbb{N}}$ such that $\mu(B_n \setminus C_n) \leq \varepsilon 2^{-n}$. If $\{A_n\}_{n\in\mathbb{N}}$ is increasing, then $B = \cup_n B_n$ is an open set containing $\cup_n A_n$. As the measure is finite, we know $\mu\big(B \setminus (\cup_{m\leq n} B_m)\big) \downarrow 0$, and so for any n

$$\begin{aligned}\mu\big(B \setminus (\cup_{m\leq n} C_m)\big) &\leq \mu\big(B \setminus (\cup_{m\leq n} B_m)\big) + \sum_{m\leq n} \mu(B_m \setminus C_m) \\ &\leq \mu\big(B \setminus (\cup_{m\leq n} B_m)\big) + \varepsilon\end{aligned}$$

which can be made arbitrarily small. As $\cup_{m\leq n} C_m$ is compact for any finite m, we see that $\cup_n A_n$ is regular. Conversely, if A_n is a decreasing sequence, then we see that $C = \cap_n C_n$ is a compact contained within $\cap_n A_n$. Again, by finiteness of μ we have $\mu(C_n \setminus C) \downarrow 0$. For any n

$$\mu\big((\cap_{m\leq n} B_m) \setminus C\big) \leq \mu(B_n \setminus C_n) + \mu(C_n \setminus C) \leq \varepsilon 2^{-n} + \mu(C_n \setminus C).$$

As this can also be made arbitrarily small and $\cap_{m\leq n} B_m$ is open for finite m, we see that $\cap_n A_n$ is regular. Hence the regular sets form a monotone class, and the result holds for all Borel measurable sets. $\square$

Theorem C.4 (Daniell–Kolmogorov Extension Theorem)**.** *Let $\{\mathbb{P}_{\mathbf{t}}\}$ be a consistent family of probability measures, where $\mathbb{P}_T$ is defined for all finite sets $\mathbf{t} \subseteq \mathbb{T}$. Then there is a unique probability measure $\mathbb{P}$ on $((\mathbb{R}^d)^{\mathbb{T}}, \mathscr{B}((\mathbb{R}^d)^{\mathbb{T}}))$ such that $\mathbb{P}(\pi_{\mathbf{t}}(B)) = P_{\mathbf{t}}(B)$ for all $B \in \mathscr{B}((\mathbb{R}^d)^{\mathbf{t}})$, where $\pi_{\mathbf{t}}$ is the projection $\pi_{\mathbf{t}}(B) = \{x \in (\mathbb{R}^d)^{\mathbb{T}} : (x_{t_1}, x_{t_2}, \ldots, x_{t_N}) \in B\}$.*

Proof. Let $\mathscr{A}$ denote the subalgebra of $\mathscr{B}(\mathbb{R}^{\mathbb{T}})$ given by sets of the form

$$\{x \in (\mathbb{R}^d)^{\mathbb{T}} : (x_{t_1}, x_{t_2}, \ldots, x_{t_n}) \in B, B \in \mathscr{B}((\mathbb{R}^d)^n)\}$$

for finite sequences $\{t_1, t_2, \ldots, t_n\}$. Note that, as we restrict to finite sequences, $\mathscr{A}$ is an algebra, but not a σ-algebra. We can define a measure $\mathbb{P}$ on the algebra $\mathscr{A}$ by $\mathbb{P}(\pi_T(B)) = \mathbb{P}_T(B)$ for all $B \in \mathscr{B}((\mathbb{R}^d)^T)$.

We need to show that $\mathbb{P}$ is countably additive. It is enough to show that $\lim_n \mathbb{P}(A_n) = 0$ for $\{A_n\}_{n\in\mathbb{N}}$ any nonincreasing sequence of sets in $\mathscr{A}$ with $\cap_n A_n = \emptyset$. As the sequence $\{A_n\}_{n\in\mathbb{N}}$ is nonincreasing, we know that $\mathbb{P}(A_n) = \mathbb{P}(A_n \setminus$

$A_{n+1}) + \mathbb{P}(A_{n+1}) \geq \mathbb{P}(A_{n+1})$, so the limit $\lim_n \mathbb{P}(A_n)$ exists, and is within $[0,1]$ by construction.

Suppose that $\lim_n \mathbb{P}(A_n) = \varepsilon > 0$. As $A_n \in \mathscr{A}$, we know that there exists a sequence $t_1, t_2, \ldots$ such that we can write

$$A_n = \{x \in (\mathbb{R}^d)^{\mathbb{T}} : (x_{t_1}, x_{t_2}, \ldots, x_{t_{k(n)}}) \in B_n, B_n \in \mathscr{B}((\mathbb{R}^d)^{k(n)})\}$$

for some function $k : \mathbb{N} \to \mathbb{N}$. By Lemma C.3, we can find compact sets $C_n \subseteq B_n$ such that the corresponding events

$$D_n = \big\{x \in (\mathbb{R}^d)^{\mathbb{T}} : (x_{t_1}, x_{t_2}, \ldots, x_{t_{k(n)}}) \in C_n\big\} \subseteq A_n$$

satisfy $\mathbb{P}(A_n \setminus D_n) \geq \varepsilon 2^{-n}$. Taking an intersection, we see that

$$\cap_{i\leq n} D_i = \big\{x \in (\mathbb{R}^d)^{\mathbb{T}} : (x_{t_1}, \ldots, x_{t_{k(n)}}) \in \tilde{C}_n := \cap_{i\leq n}(C_i \times (\mathbb{R}^d)^{(t_{k(n)} - t_{k(i)})})\big\}.$$

Therefore,

$$\begin{aligned}\mathbb{P}(\cap_{i\leq m} D_i) &= \mathbb{P}(A_m) - \mathbb{P}\big(A_m \setminus (\cap_{n\leq m} D_n)\big) = \mathbb{P}(A_m) - \mathbb{P}\big(\cup_{n\leq m}(A_m \setminus D_n)\big) \\ &\geq \mathbb{P}(A_m) - \mathbb{P}\big(\cup_{n\leq m}(A_n \setminus D_n)\big) \geq \varepsilon - \varepsilon \sum_{n\leq m} 2^{-m} > 0.\end{aligned}$$

Therefore, for each m, $\cap_{i\leq m} D_i$ is nonempty, and so $\tilde{C}_n$ is also nonempty. Take an arbitrary $c_m = (c_m^1, c_m^2, \ldots, c_m^{k(m)}) \in \tilde{C}_m$. As the sets $\tilde{C}_m$ are nonincreasing, the sequence $\{(c_m^1, \ldots, c_m^{k(1)})\}_{m\in\mathbb{N}} \subset (\mathbb{R}^d)^{k(1)}$ is within the compact set $\tilde{C}_1$, and hence has a convergent subsequence with limit $(c^1, \ldots, c^{k(1)})$. Similarly, taking subsequences, the sequence $\{(c_m^1, \ldots, c_m^{k(2)})\}_{m\in\mathbb{N}} \subset (\mathbb{R}^d)^{k(2)}$ is within the compact set $\tilde{C}_2$, and so we have a limit $(c^1, \ldots, c^{k(2)})$.

We therefore obtain a sequence $(c^1, c^2, \ldots)$ with $(c^1, c^2, \ldots, c^{k(n)}) \in \cap_{i\leq n} \tilde{C}_i$ for every n. It follows that the event $(\cap_{i=1}^{\infty}\{x_{t_i} = c^i\}) \in D_n$ for every n, which implies that $\cap_n D_n$ is nonempty. This is a contradiction with the fact that $\cap_n D_n \subseteq \cap_n A_n = \emptyset$. Therefore, we must have $\lim_n \mathbb{P}(A_n) = 0$, and so P is countably additive.

By Carathéodory's extension theorem we can now extend $\mathbb{P}$ uniquely to a measure on $\sigma(\mathscr{A})$, which is equal to $\mathscr{B}((\mathbb{R}^d)^{\mathbb{T}})$, by Lemma C.2. □

C.2 Kolmogorov–Čentsov continuity criterion

NOT EXAMINABLE

Theorem C.5 (Kolmogorov–Čentsov continuity criterion)**.** *Let X be a measurable process (valued in any Banach space) such that, for some positive α, β, c, for all $s < t$,*

$$E\big[\|X_t - X_s\|^{\alpha}\big] \leq c|t-s|^{1+\beta}.$$

Then there exists a modification $\tilde{X}$ of X which is almost surely locally Hölder γ-continuous for all $\gamma \in]0, \beta/\alpha[$. In particular, for each T, there exists a constant $k > 0$ such that for all $\delta > 0$,

$$P\Big(\sup_{\{s<t<T\}}\Big\{\frac{\|\tilde{X}_t - \tilde{X}_s\|}{|t-s|^{\gamma}}\Big\} > \delta\Big) \leq k\delta^{-\alpha}.$$

Proof. Fix $T = 1$ for simplicity, the general result will hold by induction. Let D_n be the dyadic rationals of the form $D_n = \{k2^{-n}\}_{n,k\in\mathbb{Z}^+} \subset [0,1[$, and let $\Delta_n = 2^{-n}$ and $\lfloor t \rfloor_n = \max\{s \in D_n : s \leq t\}$, $\lceil t \rceil_n = \min\{s \in D_n : s \geq t\}$, as in Lévy's construction of Brownian motion. Then for $t \in D_n$, the assumption states that

$$E\big[\sup_{t\in D_n} |X_{t+\Delta_n} - X_t|^\alpha\big] \leq \sum_{t\in D_n} E[|X_{t+\Delta_n} - X_t|^\alpha] \leq c2^{n+1}\Delta_n^{1+\beta} = c2^{1-n\beta}.$$

Define $\mu_n = \sup_{t\in D_n} |X_{t+\Delta_n} - X_t|$.

For $s,t \in \cup_n D_n$, with $s < t$, the sequences $\{\lceil s \rceil_n\}_{n\in\mathbb{N}}$, $\{\lfloor t \rfloor_n\}_{n\in\mathbb{N}}$ will equal s,t respectively after at most finitely many steps. Then we can write

$$X_t - X_s = \Big(X_{\lfloor t \rfloor_m} + \sum_{i=m}^{\infty}(X_{\lfloor t \rfloor_{i+1}} - X_{\lfloor t \rfloor_i})\Big) - \Big(X_{\lceil s \rceil_m} + \sum_{i=m}^{\infty}(X_{\lceil s \rceil_{i+1}} - X_{\lceil s \rceil_i})\Big)$$

where all the sums are in fact finite. Therefore, as $\lfloor t \rfloor_{i+1} = \lfloor t \rfloor_i + \Delta_{i+1}$ unless $\lfloor t \rfloor_i = t$,

$$\begin{aligned}\|X_t - X_s\| &\leq \|X_{\lfloor t \rfloor_m} - X_{\lceil s \rceil_m}\| + \sum_{i=m}^{\infty}\|X_{\lfloor t \rfloor_{i+1}} - X_{\lfloor t \rfloor_i}\| + \sum_{i=m}^{\infty}\|X_{\lceil s \rceil_{i+1}} - X_{\lceil s \rceil_i}\| \\ &\leq \|X_{\lfloor t \rfloor_m} - X_{\lceil s \rceil_m}\| + 2\sum_{i=m+1}^{\infty}\mu_i.\end{aligned}$$

In particular, if $|t-s| < 2^{-m}$, then $\lfloor t \rfloor_m = \lceil s \rceil_m$, and so $\|X_t - X_s\| \leq 2\sum_{i=m+1}^{\infty}\mu_i$. From this, we see

$$\begin{aligned}M_\gamma := \sup_{\{s,t\in\cup_n D_n\}} \frac{\|X_t - X_s\|}{|t-s|^\gamma} &\leq \sup_{\substack{m,n\in\mathbb{Z}^+\\ m<n}} \sup_{\substack{s,t\in D_n\\ |t-s|<2^{-m}}} \frac{\|\tilde{X}_t - \tilde{X}_s\|}{|t-s|^\gamma} \\ &\leq \sup_{\substack{m,n\in\mathbb{Z}^+\\ m<n}} \frac{2\sum_{i=m+1}^{\infty}\mu_i}{2^{-\gamma n}} = \sup_{m\in\mathbb{Z}^+}\Big\{2^{1+\gamma(m+1)}\sum_{i=m+1}^{\infty}\mu_i\Big\} \\ &\leq 2^{1+\gamma}\sum_{i=1}^{\infty}2^{\gamma i}\mu_i.\end{aligned}$$

If $\alpha \geq 1$, by Minkowski's inequality

$$\begin{aligned}E\big[|M_\gamma|^\alpha\big]^{1/\alpha} &\leq 2^{1+\gamma}\sum_{i=1}^{\infty}2^{\gamma i}E\big[|\mu_i|^\alpha\big]^{1/\alpha} \leq 2^{1+\gamma}\sum_{i=1}^{\infty}2^{\gamma i}c^{1/\alpha}2^{(1-i\beta)/\alpha} \\ &= c^{1/\alpha}2^{1+\gamma+1/\alpha}\sum_{i=1}^{\infty}2^{(\gamma-\beta/\alpha)i} = \frac{c^{1/\alpha}2^{1+\gamma+1/\alpha}}{2^{(\gamma-\beta/\alpha)}-1} < \infty.\end{aligned}$$

For $\alpha < 1$, the same bound holds for $E\big[|M_\gamma|^\alpha\big]$.

Therefore, as M_γ is almost surely finite, for almost all ω, X is uniformly continuous on $\cup_n D_n$. Defining $\tilde{X}_t := \lim_n X_{\lfloor t \rfloor_n}$, by Fatou's inequality we see $X_t = \tilde{X}_t$ almost surely, and

$$\sup_{\{s<t<T\}} \frac{\|\tilde{X}_t - \tilde{X}_s\|}{|t-s|^\gamma} = M_\gamma,$$

so $\tilde{X}$ is a uniformly continuous modification of X. Finally, by Markov's inequality, $P(M_\gamma > \delta) \le E\left[|M_\gamma|^\alpha\right]\delta^{-\alpha}$. A Borel–Cantelli argument (with $\delta = 2^{k/\alpha}$) gives the almost-sure Hölder continuity. □

D A very short primer in functional analysis

We start with a brief recall of basic notions of functional analysis leading to Hilbert spaces and identification of their dual.

D.1 Normed vector spaces

We start with basics. A vector space V over $\mathbb{R}$ is a set endowed with two binary operations: addition and multiplication by a scalar, which satisfy the natural axioms. We focus the discussion on real scalars as this is relevant for us but most of what follows applies to spaces over complex numbers (or more general fields).

Definition D.1. *A norm $\|\cdot\|$ on a vector space V is a mapping from V to $[0,\infty)$ such that*

(i) for any $a \in \mathbb{R}$, $v \in V$, $\|av\| = |a|\|v\|$ (absolute homogeneity);

(ii) for any $x,y \in V$, $\|x+y\| \le \|x\| + \|y\|$ (triangle inequality);

(iii) $\|v\| = 0$ if and only if v is the zero vector in V (separates points).

Note that a norm induces a metric on V through $d(x,y) = \|x-y\|$ and hence a topology on V. A norm is then a continuous function from V to $\mathbb{R}$. The space of continuous linear functionals plays a special role:

Definition D.2. *Given a normed vector space over reals $(V, \|\cdot\|_V)$, its dual V' is the space of all continuous linear maps (functionals) from V to $\mathbb{R}$. V' itself is a vector space over $\mathbb{R}$ equipped with a norm*

$$\|\phi\|_{V'} := \sup_{v\in V, \|v\|_V \le 1} |\phi(v)|.$$

The classical examples of spaces to consider are spaces of sequences or of functions. Let $(\mathscr{S}, \mathbb{F}, \mu)$ be a measurable space endowed with a σ-finite measure. Then for a real valued measurable function f on $\mathscr{S}$ we can consider

$$\|f\|_p := \left(\int_{\mathscr{S}} |f(x)|^p \mu(\mathrm{d}x)\right)^{1/p}$$

and let $\mathscr{L}^p(\mathscr{S}, \mathbb{F}, \mu)$ be the space of such functions for which $\|f\|_p < \infty$. Observe that $\|\cdot\|_p$ is not yet a norm on $\mathscr{L}^p$ – indeed it fails to satisfy (iii) in Definition D.1 since if $f = 0$ μ-a.e. but is not zero, e.g. $f = \mathbf{1}_A$ for a measurable $A \in \mathbb{F}$ with $\mu(A) = 0$, then still $\|f\|_p = 0$. We then say that $\|\cdot\|_p$ is a semi-norm on $\mathscr{L}^p$.

To remedy this, we consider the space $\mathrm{L}^p(\mathscr{S},\mathbb{F},\mu)$ which is the quotient of $\mathscr{L}^p$ with respect to the equivalence relation $f \sim g$ iff $f = g$ μ-a.e. Put differently, L^p is the space of equivalence classes of functions equal μ-a.e. and which are integrable with p^{th} power. Then $(\mathrm{L}^p(\mathscr{S},\mathbb{F},\mu), \|\cdot\|_p)$ is a normed vector space for $p \geq 1$. The triangle inequality for $\|\cdot\|_p$ is simply the *Minkowski inequality*.

A more geometric notion of measuring the relation between vectors is given by an inner product.

Definition D.3. *Given a vector space V over $\mathbb{R}$, a mapping $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ is called an* inner product *if*

(i) it is bilinear and symmetric: $\langle ax+bz,y\rangle = a\langle x,y\rangle + b\langle z,y\rangle$ and $\langle x,y\rangle = \langle y,x\rangle$ for $a,b \in \mathbb{R}$, $x,y,z \in V$;

(ii) for any $x \in V$, $\langle x,x\rangle \geq 0$;

(iii) $\langle x,x\rangle = 0$ if and only if x is the zero vector in V.

This notion is very familiar on $V = \mathbb{R}^n$ where an inner product is given by

$$\langle x,y\rangle = x^{\mathrm{T}}y = \sum_{i=1}^{n} x_i y_i.$$

An inner product satisfies the Cauchy–Schwartz inequality

Proposition D.4. *An inner product on a vector space satisfies*

$$|\langle x,y\rangle| \leq \sqrt{\langle x,x\rangle}\sqrt{\langle y,y\rangle}, \quad x,y \in V. \tag{48}$$

Proof. Let $\|x\| := \sqrt{\langle x,x\rangle}$, $x \in V$. Fix $x,y \in V$ and define a quadratic function $Q : \mathbb{R} \to \mathbb{R}$ by

$$Q(r) = \|x+ry\|^2 = \|y\|^2 r^2 + 2\langle x,y\rangle r + \|x\|^2, \quad r \in \mathbb{R}$$

which clearly is non-negative and hence its discriminant has to be non-positive i.e.

$$4|\langle x,y\rangle|^2 - 4\|x\|^2\|y\|^2 \leq 0 \text{ that is } |\langle x,y\rangle| \leq \|x\|\|y\|,$$

as required. We note also that equality holds if and only if the vectors x,y are linearly dependent i.e. $x = ry$ for some $r \in \mathbb{R}$. □

The above implies that $\|x\| := \sqrt{\langle x,x\rangle}$ is a norm on V. We say that the norm is induced by an inner product. Among spaces L^p defined above only L^2 has norm which is induced by an inner product, namely by

$$\langle f,g\rangle = \int_{\mathscr{S}} f(x)g(x)\mu(\mathrm{d}x). \tag{49}$$

D.2 Banach spaces

We first define Cauchy sequences which embody the idea of a converging sequence when we do not know the limiting element.

Definition D.5. *A sequence (x_n) of elements in a normed vector space $(X, \|\cdot\|)$ is called a Cauchy sequence if for any $\varepsilon > 0$ there exists $N \geq 1$ such that for all $n, m \geq N$ we have $\|x_n - x_m\| \leq \varepsilon$.*

Definition D.6. *A normed vector space $(X, \|\cdot\|)$ is complete if every Cauchy sequence converges to an element $x \in X$. It is then called a* Banach space. *Further, if the norm is induced by an inner product, then $(X, \|\cdot\|)$ is called a* Hilbert space.

Naturally, the Euclidean space $\mathbb{R}^d$ is a Banach space (and in fact a Hilbert space) with the norm $\|x\| = \sqrt{\sum_{i=1}^d x_i^2}$. This implies (reasoning for $d = 1$) that

Proposition D.7. *If $(X, \|\cdot\|_X)$ is a normed vector space over $\mathbb{R}$ then its dual $(X', \|\cdot\|_{X'})$ in Definition D.2 is a Banach space.*

In many cases it is interesting to build linear functionals satisfying certain additional properties. This is often done using the Hahn–Banach theorem. It states in particular that a continuous linear functional defined on a linear subspace Y of X can be continuously extended to the whole of X without increasing its norm. A version of this is also known as the separating hyperplane theorem since it allows to separate two convex sets (one open) using an affine hyperplane.

An important step in studying continuous linear functionals on X is achieved by describing the structure of X'. We have

Proposition D.8. *Let $(\mathscr{S}, \mathbb{F}, \mu)$ be a measurable space with a σ-finite measure. Then for any $p \geq 1$, $\mathrm{L}^p(\mathscr{S}, \mathbb{F}, \mu)$ is a Banach space and for $p > 1$ its dual is equivalent to (isometric to) the space $\mathrm{L}^q(\mathscr{S}, \mathbb{F}, \mu)$, where $1/p + 1/q = 1$.*

In particular we see that L^2 is its own dual. This means that any continuous linear functional on L^2 can be identified with an element in L^2. This property remains true for any Hilbert space:

Proposition D.9. *Let $(X, \|\cdot\|)$ be a Hilbert space with the norm induced by an inner product, $\|x\| = \sqrt{\langle x, x\rangle}$. If $\phi : X \to \mathbb{R}$ is a continuous linear map, then there exists an element $x_\phi \in X$ such that*

$$\phi(y) = \langle x, y\rangle, \quad \forall y \in X.$$

In particular, if $\phi : \mathrm{L}^2(\mathscr{S}, \mathbb{F}, \mu) \to \mathbb{R}$ is a continuous linear map, then there exists an element $f_\phi \in \mathrm{L}^2(\mathscr{S}, \mathbb{F}, \mu)$ such that

$$\phi(g) = \int_{\mathscr{S}} g(x) f_\phi(x) \mu(\mathrm{d}x), \quad \forall g \in \mathrm{L}^2(\mathscr{S}, \mathbb{F}, \mu).$$

Note that the inner product $\langle x,y\rangle$, or the integral $\int_{\mathscr{S}} g(x) f_\phi(x)\mu(\mathrm{d}x)$, in the above statement is well defined by (48)–(49).

On a (separable) Hilbert space, we can also state an infinite dimensional analogue of the Pythagorean theorem. Recall that $\mathbb{R}^n$ is a Hilbert space with inner product $\langle x,y\rangle = \sum_{i=1}^n x_i y_i$. This uses the canonical basis in $\mathbb{R}^n$ but if we take any orthonormal basis in $\mathbb{R}^n$, say $(\varepsilon_1,\dots,\varepsilon_n)$, then

$$x = \sum_{i=1}^n \langle x,\varepsilon_i\rangle \varepsilon_i \quad \text{and hence } \|x\|^2 = \sum_{i=1}^n \langle x,\varepsilon_i\rangle^2,$$

which is the Pythagorean theorem. The same reasoning gives $\langle x,y\rangle = \sum_{i=1}^n \langle x,\varepsilon_i\rangle\langle y,\varepsilon_i\rangle$. The infinite dimensional version is known as the

Proposition D.10 (Parseval's identity). *Let $(X,\|\cdot\|)$ be a separable Hilbert space with the norm induced by an inner product $(x,y)\to\langle x,y\rangle$ and let $(\varepsilon_n : n\geq 1)$ be an orthonormal basis of X. Then for any $x,y\in X$*

$$\langle x,y\rangle = \sum_{n\geq 1}\langle x,\varepsilon_n\rangle\langle y,\varepsilon_n\rangle, \quad \textit{and in particular } \|x\|^2 = \sum_{n\geq 1}\langle x,\varepsilon_n\rangle^2.$$

Finally we state one more result, which is crucial for the construction of the stochastic integral.

Proposition D.11. *Suppose $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ are two Banach spaces, $\mathscr{E}\subset X$ is a dense vector subspace in X and $I:\mathscr{E}\to Y$ is a linear isometry, i.e. a linear map which preserves the norm, $\|I(x)\|_Y = \|x\|_X$ for all $x\in X$. Then I may be extended in an unique way to a linear isometry from X to Y.*

Proof. Take $x\in X$ and $x_n\to x$ with $x_n\in\mathscr{E}$. Then, by the isometry property, $(I(x_n))$ is Cauchy in Y since (x_n) is Cauchy in X. It follows that it converges to some element which we denote $I(x)$. Further, if we have two sequences in $\mathscr{E}$, (x_n) and (y_n), both converging to $x\in X$ and giving raise to potentially two elements $I(x)$ and $I(x)'$ then we can build a third sequence $z_{2n} = x_n$, $z_{2n+1} = y_n$ which also converges to x and we see that $I(z_n)$ has to converge and the limit has to agree with both $I(x)$ and $I(x)'$, so that $I(x) = I(x)'$ is unique. It follows that we defined $I(x)\in Y$ uniquely for all $x\in X$. Further,

$$\|I(x)\|_Y = \lim_n \|I(x_n)\|_Y = \lim_n \|x_n\|_X = \|x\|_X$$

so that I is norm-preserving. Finally, if $x,y\in X$ then we can write them as limits of sequences of elements in $\mathscr{E}$, say (x_n) and (y_n) respectively. For any $a,b\in\mathbb{R}$, $ax_n+by_n\in\mathscr{E}$ since $\mathscr{E}$ is a vector space, and then by the above and linearity of I on $\mathscr{E}$ we have

$$I(ax+by) = \lim_n I(ax_n+by_n) = \lim_n \Big(aI(x_n)+bI(y_n)\Big) = aI(x)+bI(y),$$

so that I is linear on X as required. ☐

The following theorem is fundamental to many approximations. A proof can be found in the Lecture notes of B4.1 (Functional Analysis I)

Theorem D.12 (Stone–Weierstrass theorem)**.** *Let $K \subset \mathbb{R}^n$ be compact. Then the space of polynomials is dense in $C(K)$, i.e. for every $f \in C(K)$ there exists a sequence of polynomials $\{p_n\}$ so that $p_n \to f$ uniformly on K.*