MA50251: Applied SDEs

Alexander Cox

February 12, 2018

Preliminaries

These notes constitute the core theoretical content of the unit MA50251, which is a course on 'Applied SDEs'. Roughly speaking, the aim of this course is to introduce SDEs (Stochastic Differential Equations) as a tool for modelling random phenomena in a wide variety of contexts, to students with varied backgrounds. In particular, we do not assume a background in rigorous (measure-theoretic) probability. In one or two places, this means that we need to be a little vague about some details, however in general I hope to introduce the main issues in the field without substantial loss in rigour.

This set of notes is intended to cover the first 8 hours of lectures (4×2 hours). The remaining lectures will be given by Tony Shardlow, and be more focussed on the applications. The material in these notes will loosely follow the book of Øksendal [3], and this is recommended as a first destination when looking for a different treatment or additional examples and detail. Also useful, although generally more advanced, are the books by Karatzas and Shreve [1], Mörters and Peres [2], and Revuz and Yor [4]. Although we will try to assume only a limited background in probability, some understanding of fundamental probability is necessary. A brief summary of the core probabilistic ideas which are needed can be found at the end of these notes in Section 7, where some key notions from probability are briefly summarised.

More information about the course can be found on the webpage: `http://www.maths.bath.ac.uk/~mapamgc/AppliedSDEs.htm`. Information that will appear here includes problem sheets and solutions. Where a question on a problem sheet may be informative to the discussion in the notes, a note in the margin may be included. The lectures will also include some interactive demonstrations using an iPython notebook. This interlude is generally indicated in the notes, and a link to the relevant iPython notebook is included. The results of the notebook will also be included as a *pdf* file for reading, however it is strongly recommended that the interested reader spend time playing with the iPython examples, and modifying to better understand the ideas. Details about installing and running the notebooks can be found on the course website.

Finally, the notes also include occasional digressions into measure theory. These are not core to the material, and can safely be ignored by a reader who has not taken any courses on measure theory. They are clearly distinguished from the main text by a change of colour and the marginal designation MT.

1 Introduction

The aim of this course is to consider processes which model systems whose dynamics depend on both the current state of the system, *and* some random noise. Consider as a basic example, the proportion of a biological population which possesses some genetic mutation, $X_t \in [0,1]$ say. Then over time the proportion of the population expressing this trait may evolve in a way that depends on both the current proportion of the population which displays this trait, and also some random noise, representing for example background effects (e.g. fluctuations in the living environment at the individual level). Writing loosely, we might expect:

$$\frac{dX_t}{dt} = \mu(X_t) + \sigma(X_t) \times \text{ `noise'}. \tag{1}$$

In the absence of the noise term, this is just a classical ODE, and we can model the system using standard techniques for ODEs. The question we try to address in this course is: how to deal with the additional noise term.

Some important questions we might hope to answer are:

- How to model the noise? What (statistical) properties might be useful?
- What can we say about the *distribution* of the process X_t at a future time?
- What can we say about *path properties* of the process: for example, the probability that the process ever exceeds a certain level.

In the example of the population model above, then we might be interested in the probability that at least half of the population have a particular trait at time t_0. We could also be interested in (for example) the probability that one of the traits becomes extinct, or whether there is a long-run asymptotic distribution (that is, a probability measure μ such that $\lim_{t\to\infty} \mathbb{P}(X_t \in A) = \mu(A)$). Note that these are inherently probabilistic questions — they do not have deterministic counterparts.

An important question is then what sort of properties might we expect from the noise term? Let's start by thinking about a discretised version of the above equation, say $\Delta t = N^{-1}$, $t_k := k\Delta t$, for $k = 0, 1, \ldots, N$. Then the discretised version of the equation above becomes:

$$X_1 = X_0 + \sum_{k=0}^{N-1} \left(\mu(X_{t_k})\,\Delta t + \sigma(X_{t_k})\,Z_k\right), \tag{2}$$

where the Z_k's represent the noise terms.

The question now becomes: what properties should Z_k have in order for this equation to make sense? We can immediately make two natural assumptions: the Z_k's should be independent (we shouldn't be able to predict future noise based on past noise), and identically distributed (the noise should not depend on X_t or t).

Armed with these assumptions, we can perform some numerical experiments. One question is what should the distribution of Z_k be. This separates into two questions: first, the type of distribution (e.g. Gaussian, Bernoulli, etc.) and secondly, what parameters should we choose — in particular, as we let $N \to \infty$, how should the distribution of the Z_k's scale in N?

iPython Interlude: Go To Notebook

Of course, in the simplest case, we can use a little theory to back this up: let $\mu(x) = 0$, and $\sigma(x) = 1$. Then we have

$$X_1 = X_0 + \sum_{k=0}^{N-1} Z_k, \tag{3}$$

where our assumption is that the Z_k's are i.i.d.. But now we have a key result about convergence of sums of i.i.d. r.v.'s. Recall that the Central Limit Theorem (CLT, see Section 7.6) says that if Y_k is a sequence of i.i.d. r.v.'s with mean 0 and variance ξ^2, then

$$\frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} Y_k \xrightarrow{\mathcal{D}} N(0, \xi^2). \tag{4}$$

Setting $Y_k = \sqrt{N}\,(Z_k - \mathbb{E}[Z_k])$, we see that (3) has a sensible limit if $\mathbb{E}[Z_k] \approx \nu N^{-1}$ and $\mathbb{V}ar(Z_k) \approx \xi^2 N^{-1}$, in which case $X_1 - X_0 = N(\nu, \xi^2)$.

More generally, we can use this limiting argument to define a process X_t. Let N be large, and set

$$X_t = X_0 + \sum_{k=0}^{\lfloor Nt \rfloor - 1} Z_k.$$

Then for $t \geq s$, we have

$$X_t - X_s = \sum_{k=\lfloor Ns \rfloor}^{\lfloor Nt \rfloor - 1} Z_k$$

and for large N, by the Central Limit Theorem again, the right hand side is normally distributed with mean $\nu(t-s)$ and variance $\xi^2(t-s)$. In addition, consider times $t_1 \leq t_2 \leq \cdots \leq t_n$. Then if $i \neq j$, $X_{t_{i+1}} - X_{t_i}$ and $X_{t_{j+1}} - X_{t_j}$ have at most one Z_k term in common, the other terms contributing to the sums being independent. In the limit, we would then expect these terms to be independent. In particular that we have the following properties:

(P1): Normally distributed: $X_t - X_s \sim N(\nu(t-s), \xi^2(t-s))$,
(P2): Independent increments: if $0 = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$ then $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

Then this limiting process looks a lot like a Gaussian process: specifically, $(X_{t_0}, \ldots, X_{t_n})$ is a Gaussian process with mean function $m(t) = \nu t$ and covariance function $K(s,t) = \xi^2 \min(s,t)$. Q1.1

However, we want to introduce one further assumption. Not only do we want to consider the joint distribution of 'slices' of the process at a finite number of fixed times, but more generally, we will want to consider the *whole* process as a function of time, $t \mapsto X_t$. For example, we might want to know properties of the maximum of the process, $\sup_{s \leq T} X_t$. However in full generality, the two properties above are simply not enough: suppose we construct a process satisfying the normal and independent increments assumptions. If I take a random variable $U \sim U([0,1])$, and define a new process X'_t by:

$$X'_t = \begin{cases} X_t, & t \neq U \\ 10, & t = U \end{cases}$$

then X_t and X'_t both satisfy the normal and independent increments condition (if I check either property for X' at any finite set of times, the probability that any of these times are equal to U is zero). It follows that many of the path properties that we may be interested are not robust if we only specify the two properties above.

To rectify this, we need to make a simple additional assumption: we assume that

(P3): $t \mapsto X_t$ is a continuous process.

Note that this property would appear to be justified by inspection of the numerical output above. We will also provide a more detailed justification of this below.

MT

Note also that path continuity is also a strongly necessary property from a measure theoretic point of view. If we simply consider the (uncountable) set of random variables $(X_t; t \in \mathbb{R}_+)$, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, it is not even clear that the event $\{t \mapsto X_t \text{ is continuous}\}$ is measurable. By assuming that the process is continuous, we can check many relevant properties simply by looking at X_t at rational times, for example:

$$\left\{\sup_{s\le t} X_s \ge x\right\} = \bigcap_{\varepsilon_n \to 0} \left\{\sup_{s\le t, s\in\mathbb{Q}} X_s \ge x - \varepsilon_n\right\}$$

which is easily seen to be measurable.

To make the discussion above more rigorous, we say that two processes X_t and X'_t, defined on the same probability space, are **modifications** of each other if, for each t,

$$X_t = X'_t, \quad a.s..$$

Note that this is different to the stronger notion of indistinguishable processes: X_t and X'_t are indistinguishable if, for almost all ω:

$$X_t(\omega) = X'_t(\omega), \qquad \text{for every } t.$$

However, if we know that X_t and X'_t have continuous paths, then these are equivalent (by countability of the rationals, $\mathbb{P}(X_t = X'_t, \forall t \in \mathbb{Q}) = 1$, which is equivalent to being everywhere equal when the paths are continuous).

In particular, we have the following result:

Theorem 1.1 (Kolmogorov's Continuity Criterion)**.** *If $(X_t)_{t\ge0}$ is a real-valued process such that for all $T > 0$ there exist constants $\alpha, \beta, \gamma > 0$ with*

$$\mathbb{E}\left[|X_{t+h} - X_t|^\alpha\right] \le \gamma h^{1+\beta}$$

for every t and $0 \le h \le T$, then there exists a modification of X which has almost-surely continuous paths.

In the case where X satisfies (P1) and (P2), it is easy to check that this condition is satisfied, since[1] $\mathbb{E}\left[Z^4\right] = \nu^4h^4 + \nu^3\xi^2h^3 + 3\xi^4h^2$, when $Z \sim N(\nu h, \xi^2 h)$, and so

$$\mathbb{E}\left[|X_{t+h} - X_t|^4\right] = \nu^4h^4 + \nu^3\xi^2h^3 + 3\xi^4h^2 \le h^2(\nu^4T^2 + \nu^3\xi^2T + 3\xi^4) =: \gamma h^2.$$

where we have used the fact that $X_{t+h} - X_t \sim N(\nu h, \xi^2 h)$.

In particular, given any process X_t satisfying (P1) and (P2), there exists a *modification* X'_t which has continuous paths. It is this modification that we will work with.

[1]For example, by recalling that the moment generating function of a $N(\mu, \sigma)$ r.v. Z is $f_Z(\alpha) := \mathbb{E}\left[\exp\{\alpha Z\}\right] = e^{\mu t + \sigma^2 t^2/2}$, and so $\mathbb{E}\left[Z^4\right] = f_Z''''(0) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^2$.

2 Brownian Motion

2.1 Definition and Basic Properties

Having motivated this construction, we define:

Definition 2.1. A real-valued stochastic process $(W_t)_{t\geq 0}$ is a **(standard) Brownian motion** starting at $x \in \mathbb{R}$ if:

(P0): $W_0 = x$,
(P1): $W_t - W_s \sim N(0, t-s)$,
(P2): if $0 = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$ then $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$ are independent,
(P3): $t \mapsto W_t$ is continuous, almost surely.

Here, standard refers to the fact that we have chosen $\nu = 0, \xi = 1$. More generally, a Brownian motion with **drift** ν and **infinitesimal variance** ξ^2 is a process $X_t = \nu t + \xi W_t$, for a standard Brownian motion W_t. If the starting point is unspecified, we typically assume $W_0 = 0$. Note that we use W here as a reference to 'Wiener', who gave the first rigorous mathematical construction.

A $\mathbb{R}^d$-valued process $(\boldsymbol{W}_t)_{t\geq 0}$ is a (d-dimensional) Brownian motion if $\boldsymbol{W}_t = (W_t^1, W_t^2, \dots, W_t^d)$, where $W^1, \dots, W^d$ are i.i.d. $\mathbb{R}$-valued Brownian motions. Equivalently, $(\boldsymbol{W}_t)_{t\geq 0}$ is a stochastic process taking values in $\mathbb{R}^d$, such that $\boldsymbol{W}_t - \boldsymbol{W}_s \sim N(\mathbf{0}, (t-s)I)$, a multivariate normal with mean $\mathbf{0} \in \mathbb{R}^d$, and covariance matrix which is $(t-s)$ times the identity matrix, and such that (P2), (P3) hold.

We list here some important properties of Brownian motion[2]:

Theorem 2.2. *If $(W_t)_{t\geq 0}$ is a (one-dimensional) standard Brownian motion with $W_0 = 0$, then:*

i) $(W_t)_{t\geq 0}$ *is a Strong Markov process;*
ii) $(W_t)_{t\geq 0}$ *is a martingale;*
iii) if $\alpha \neq 0$, $(\alpha^{-1} W_{\alpha^2 t})_{t\geq 0}$ *is a standard Brownian motion (Brownian scaling);*
iv) if $h > 0$, $(W_{t+h} - W_h)_{t\geq 0}$ *is a standard Brownian motion (stationarity);*
v) $(W_1 - W_{1-t})_{t\in[0,1]}$ *is a standard Brownian motion on* $[0,1]$ *(time-reversal).*

Some of these properties generalise to e.g. d-dimensional Brownian motion, Brownian motion starting at $x \neq 0$, or Brownian motion with drift; the details can be found in many text books.

Another important theorem is:

Theorem 2.3. *Brownian motion exists and is unique. Specifically, if W and $\tilde{W}$ are both Brownian motions on (possibly different) probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, then*

$$\mathbb{P}((W_t)_{t\geq 0} \in A) = \tilde{\mathbb{P}}((\tilde{W}_t)_{t\geq 0} \in A)$$

for any[3] subset A of the continuous functions.

The existence of Brownian motion is non-trivial, but can be shown as a consequence of the construction that follows.

[2] Don't worry if you are not familiar with some of these terms — if needed, they will be introduced later in the course.
[3] (measurable)

As a simple example of how the uniqueness result is often used, consider e.g. *iii*), *iv*) and *v*) of Theorem 2.2. Suppose we can show a property is true of e.g. the rescaled version of Brownian motion, $\tilde{W}_t = (\alpha^{-1} W_{\alpha^2 t})$. Then it must also be true of W. Of course, the uniqueness does not mean that $\tilde{W}_t = W_t$!

From a modelling perspective, the key reason for introducing Brownian motion is the following result, which formalises the intuitive arguments in the introduction: let $(Z_n)_{n\in\mathbb{Z}_+}$ be a sequence of i.i.d. random variables, normalised so that $\mathbb{E}[Z_k] = 0, \mathbb{V}ar(Z_k) = 1$. Consider the **random walk**

$$Y_n = \sum_{k=1}^{n} Z_k$$

and construct a real valued, continuous function by interpolation:

$$Y(t) = Y_{\lfloor t\rfloor} + (t - \lfloor t\rfloor)(Y_{\lfloor t\rfloor+1} - Y_{\lfloor t\rfloor}).$$

This is a continuous function on $[0,\infty)$. Then we can renormalise (recall (4)) by:

$$Y^N(t) := \frac{1}{\sqrt{N}} Y(Nt).$$

We can now consider the sequence of random functions Y^n on $[0,1]$ (say) and get the following result:

Theorem 2.4 (Donsker's invariance principle)**.** *Considered as functions on the unit interval, the functions Y^N converge to a Brownian motion W_t in the sense that, for any continuous path $f : [0,1] \to \mathbb{R}$ and $\varepsilon > 0$:*

$$\mathbb{P}(\sup_{t\in[0,1]} |Y^N(t) - f(t)| < \varepsilon) \to_{N\to\infty} \mathbb{P}(\sup_{t\in[0,1]} |W_t - f(t)| < \varepsilon).$$

If we write $C([0,1])$ for the space of continuous functions equipped with the uniform norm, $||f||_\infty = \sup_{t\in[0,1]} |f(t)|$, then this is exactly convergence in distribution in this space. See [2, Theorem 5.22] for a proof of this result.

Recall that our 'physical' interpretation of the noise in (1) was that the Z_k's should be i.i.d., Donsker's Theorem provides us with the essential theoretical understanding of this characterisation: if we assume the noise is i.i.d. with second moment[4] then, in the limit, the sum of the noise looks like a Brownian motion, formally:

$$\sum_{s_i\le t} \text{noise}_{s_i} \to \int_{s\le t} \text{noise}_s \, \mathrm{d}s \approx W_t.$$

Again, completely formally, we would like to differentiate this to get:

$$\frac{dX_t}{dt} = \mu(X_t) + \sigma(X_t) \times \dot{W}_t. \tag{5}$$

but this just doesn't make sense:

Theorem 2.5 (Payley, Wiener, Zygmund, 1933)**.** *The paths of a Brownian motion are almost-surely nowhere differentiable, moreover, almost surely for all t, at least one of:*

$$\limsup_{h\downarrow 0} \frac{W_{t+h} - W_t}{h} = \infty \qquad \text{or} \qquad \liminf_{h\downarrow 0} \frac{W_{t+h} - W_t}{h} = -\infty$$

holds.

[4]That is, $\mathbb{E}[Z_k^2] < \infty$, otherwise we cannot perform the above normalisation.

Since $\dot{W}_t$ exists at t_0 if and only if

$$\limsup_{h\downarrow 0} \frac{W_{t_0+h}-W_{t_0}}{h} = \limsup_{h\downarrow 0} \frac{W_{t_0-h}-W_{t_0}}{h} = \liminf_{h\downarrow 0} \frac{W_{t_0+h}-W_{t_0}}{h} = \liminf_{h\downarrow 0} \frac{W_{t_0-h}-W_{t_0}}{h} \in \mathbb{R}$$

we see that the latter condition of the theorem is indeed stronger than the lack of differentiability.

We refer to [2, Theorem 1.39] for a proof of this result. See also Q1.5 for a proof of a similar, but slightly easier result.

2.2 Variation Properties of Brownian motion

The lack of differentiability causes problems when we want to define the integral, and we need to come up with a method around this. The essential problem is that we have a form of cancellation arising from the noise term. Think again of our discrete approximations of the Brownian motion: fix N, let $\Delta t = 1/N$, and write $t_i = i\Delta t$, and $\Delta W_i := (W_{t_{i+1}} - W_{t_i})$, so $\Delta W_0, \Delta W_1, \dots$ are i.i.d. $N(0, \Delta t)$ r.v.'s. Then for fixed $t \approx N\Delta t$, $W_t - W_0 = \sum_{i=0}^{N-1} \Delta W_i$, i.e. the sum of N i.i.d. $N(0, \Delta t)$ random variables.

Let's think about the convergence properties of this sum — the 'typical' size of a ΔW_i is $\mathbb{E}[|\Delta W_i|]$, which is easily computed (since the increments are Gaussian) to be $\sqrt{2\Delta t/\pi}$. Since $N \approx t/\Delta t$, the sum of the absolute values is approximately:

$$\sum_{i=0}^{N-1} |\Delta W_i| \approx N\sqrt{2\Delta t/\pi} = t\frac{2}{\Delta t\, \pi}.$$

As $\Delta t \to 0$, we therefore expect the sum of the increments will not be absolutely convergent. In general, what we observe is that, although the increments are not absolutely convergent, *on average* they will cancel out, since roughly as many terms are positive as negative. However, observe that this claim is a *probabilistic claim*: it is not true that for *every* sequence of Brownian increments (in the limit) this sum will be convergent, only for almost every sequence, i.e. with probability one. The tricky part is how to combine the probability with the analysis in a satisfactory way.

However, there is another way in which we can consider the sum above in a nice way: let's look at the sum of *squared* increments: here we have $\mathbb{E}\left[(\Delta W)^2\right] = \Delta t$, and so

$$\sum_{i=0}^{N-1} (\Delta W_i)^2 \approx N\Delta t = t.$$

iPython Interlude: Go To Notebook

We can formalise this intuitive argument as follows. We say a partition $\Pi = \{t_0, t_1, \dots\}$ of $\mathbb{R}_+$ is an increasing sequence of times, $0 = t_0 \le t_1 \le t_2 \le \dots$ such that $t_k \to \infty$ as $k \to \infty$. Then the mesh of the partition is the largest gap: $||\Pi|| := \sup_{i\in\mathbb{N}} |t_i - t_{i-1}|$. In addition, the quadratic variation of the process X along the partition Π is defined to be the process:

$$V_t^{\Pi}(X) := \sum_{i\in\mathbb{N}} (X_{t_i\wedge t} - X_{t_{i-1}\wedge t})^2,$$

where the sum is along the times $t_i \in \Pi$, and we write $a \wedge b = \min\{a, b\}$, so the final non-zero term in the sum is $(X_t - X_{t_k})$, where t_k is the largest element of Π smaller than t.

This leads us to the following definition:

Definition 2.6. Let X_t be a real-valued process. Then X_t is of **finite quadratic variation** if there exists a finite process $(\langle X\rangle_t)_{t\geq 0}$ such that for every $t \geq 0$, when $||\Pi|| \to 0$, $V_t^\Pi(X)$ converges to $\langle X\rangle_t$ in probability: i.e. for all $\varepsilon, \eta > 0$, there exists $\delta > 0$ such that

$$||\Pi|| < \delta \implies \mathbb{P}(|V_t^\Pi(X) - \langle X\rangle_t| > \eta) < \varepsilon.$$

The process $\langle X\rangle_t$ is called the **quadratic variation of** X.

Theorem 2.7. *Brownian motion has finite quadratic variation, and $\langle W\rangle_t = t$ a.s.. In addition, for Brownian motion $V_t^\Pi(W) \to t$ in $\mathcal{L}^2$ as well as in probability.*

Proof. Since convergence in $\mathcal{L}^2$ implies convergence in probability, we only need to show that for $\varepsilon > 0$, there exists $\delta > 0$ such that :

$$||\Pi|| < \delta \implies \mathbb{E}\left[|V_t^\Pi(W) - t|^2\right] < \varepsilon. \tag{6}$$

Given $\Pi = \{t_0, t_1, \dots t_N\}$, with $t_0 = 0$ and $t_N = t$, write $\Delta t_i^\Pi = t_{i+1} - t_i$ and $\Delta W_i^\Pi = W_{t_{i+1}} - W_{t_i}$. Then

$$\begin{aligned}
\mathbb{E}\left[|V_t^\Pi(W) - t|^2\right] &= \mathbb{E}\left[\left|\sum_{i=0}^{N-1}(\Delta W_i^\Pi)^2 - t\right|^2\right] \\
&= \mathbb{E}\left[\left|\sum_{i=0}^{N-1}(\Delta W_i^\Pi)^2 - \sum_{i=0}^{N-1}\Delta t_i^\Pi\right|^2\right] \\
&= \mathbb{E}\left[\left|\sum_{i=0}^{N-1}\left((\Delta W_i^\Pi)^2 - \Delta t_i^\Pi\right)\right|^2\right]
\end{aligned}$$

But by the definition of Brownian motion, $(\Delta W_i^\Pi)^2 - \Delta t_i^\Pi$ are independent random variables, each with mean zero. Hence we can expand out the terms in the square, and the cross-terms are all the product of mean-zero, independent terms, and hence are all zero. So:

$$\begin{aligned}
\mathbb{E}\left[|V_t^\Pi(W) - t|^2\right] &= \mathbb{E}\left[\sum_{i=0}^{N-1}\left((\Delta W_i^\Pi)^2 - \Delta t_i^\Pi\right)^2\right] \\
&= \sum_{i=0}^{N-1}\mathbb{E}\left[\left(\Delta W_i^\Pi\right)^4 - 2(\Delta W_i^\Pi)^2(\Delta t_i^\Pi) + (\Delta t_i^\Pi)^2\right]
\end{aligned}$$

Since Z is a centred normal random variable, $\mathbb{E}\left[Z^4\right] = 3\mathbb{E}\left[Z^2\right]^2$ and also $\mathbb{E}\left[(\Delta W_i^\Pi)^2\right] = \Delta t_i^\Pi$, and we finally get

$$\begin{aligned}
\mathbb{E}\left[|V_t^\Pi(W) - t|^2\right] &= 2\sum_{i=0}^{N-1}(\Delta t_i^\Pi)^2 \\
&\leq 2\left(\max_i \Delta t_i^\Pi\right)\sum_{i=0}^{N-1}\Delta t_i^\Pi = 2||\Pi||t.
\end{aligned}$$

Taking $\delta = \varepsilon/(2t)$, we conclude that (6) holds. □

Remark 2.8. There are some subtleties here that we need to be careful about: while the theorem above tells us that

$$\sup_{||\Pi||<\delta} \mathbb{P}(|V_t^\Pi(W) - t| > \eta) < \varepsilon,$$

if we can choose the partition based on the path we observe, then the quadratic variation can be made arbitrarily large:

$$\mathbb{P}\left(\sup_{||\Pi||<\delta} V_t^\Pi(W) = \infty\right) = 1$$

for all $t, \delta > 0$.

3 Stochastic Integration

In this section we consider further the problem of making sense of an equation such as (5). We know from the discussion above that $\dot{W}$ does not make sense, so we cannot 'define' a classical integral, but we can try at least to make sense of this term in some cases. Writing (5) in its integrated form, we have:

$$X_t - X_0 = \int_0^t \mu(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s$$

where $\mathrm{d}W_s$ is a nicer way of writing the (meaningless) $\dot{W}_s\,\mathrm{d}s$. The first term on the right-hand side is not (usually) difficult to interpret: it is simply the average of a function over time, and we can handle this using classical integration. Our aim is to try and understand the $\mathrm{d}W_s$ integral.

In this section, we therefore concentrate on how to interpret integrals of this form, and (slightly more generally) how to understand

$$\int_0^t \varphi_s\,\mathrm{d}W_s$$

where W_t is a standard Brownian motion, and $(\varphi_t)_{t\geq 0}$ is a stochastic process — in particular, it may depend on the processes W_t and X_t say.

We begin by considering a simple case: $\varphi_t = W_t$. Following the numerical examples in Section 1, and recalling that we used (2), where we now take $Z_k = \Delta W_k$, we might guess:

$$\int_0^1 W_s\,\mathrm{d}W_s \approx \sum_{k=0}^{N-1} W_{t_k} \times \Delta W_k,$$

where we recall that, for fixed N, $\Delta t = 1/N$, $t_k = k\Delta t$ and $\Delta W_k = W_{t_{k+1}} - W_{t_k}$. Then the right-hand side of the equation is:

$$\sum_{k=0}^{N-1} W_{t_k}(W_{t_{k+1}} - W_{t_k}).$$

But we can use the simple algebraic identity: $a(b-a) = \frac{1}{2}b^2 - \frac{1}{2}a^2 - \frac{1}{2}(b-a)^2$ to see:

$$\begin{aligned}
\sum_{k=0}^{N-1} W_{t_k}(W_{t_{k+1}} - W_{t_k}) &= \sum_{k=0}^{N-1}\left[\frac{1}{2}W_{t_{k+1}}^2 - \frac{1}{2}W_{t_k}^2 - \frac{1}{2}(W_{t_{k+1}} - W_{t_k})^2\right] \\
&= \frac{1}{2}W_{t_N}^2 - \frac{1}{2}W_{t_0}^2 - \frac{1}{2}\sum_{k=0}^{N-1}(W_{t_{k+1}} - W_{t_k})^2 \\
&= \frac{1}{2}\left(W_1^2 - W_0^2\right) - \frac{1}{2}\sum_{k=0}^{N-1}(W_{t_{k+1}} - W_{t_k})^2. \qquad (7)
\end{aligned}$$

If W_t were 'classically' differentiable, at this point we would say that the final term is zero and the first term is what we would expect by the chain rule ('$(\frac{1}{2}W_t^2)' = W_t\,\mathrm{d}W_t$'). However, using Theorem 2.7, since $\langle W\rangle_1 = 1$, we have

$$\sum_{k=0}^{N-1} W_{t_k}(W_{t_{k+1}} - W_{t_k}) \to \frac{1}{2}\left(W_1^2 - W_0^2\right) - \frac{1}{2}$$

as $N \to \infty$, and where the convergence is in both probability and $\mathcal{L}^2$.

So this suggests a general procedure: we approximate the process φ which we wish to integrate against the Brownian motion, and look at the limit of a sequence of sums. However, there is one additional aspect about which we really need to be careful. Why did we choose to approximate $\varphi_t = W_t$ over the interval $[t_i, t_{i+1})$ by φ_{t_i}? By the continuity of Brownian paths, it really should not matter if we choose $\varphi_{(t_i+t_{i+1})/2}$ or $\varphi_{t_{i+1}}$, since in the limit as $N \to \infty$, these processes will also converge to the same process. *But* a simple calculation should give us concern: using $b(b-a) = \frac{1}{2}b^2 - \frac{1}{2}a^2 + \frac{1}{2}(b-a)^2$, we get

$$\begin{aligned}\sum_{k=0}^{N-1} W_{t_{k+1}}(W_{t_{k+1}} - W_{t_k}) &= \sum_{k=0}^{N-1} \left[\frac{1}{2}W_{t_{k+1}}^2 - \frac{1}{2}W_{t_k}^2 + \frac{1}{2}(W_{t_{k+1}} - W_{t_k})^2\right] \\ &= \frac{1}{2}W_{t_N}^2 - \frac{1}{2}W_{t_0}^2 + \frac{1}{2}\sum_{k=0}^{N-1}(W_{t_{k+1}} - W_{t_k})^2 \\ &= \frac{1}{2}\left(W_1^2 - W_0^2\right) + \frac{1}{2}\sum_{k=0}^{N-1}(W_{t_{k+1}} - W_{t_k})^2.\end{aligned}$$

In particular, in the limit, our 'integral' converges to $\frac{1}{2}\left(W_1^2 - W_0^2\right) + \frac{1}{2}$, not $\frac{1}{2}\left(W_1^2 - W_0^2\right) - \frac{1}{2}$.

iPython Interlude: Go To Notebook

In the second case, the integrand has been allowed to peek slightly into the future, and this has led to a different result. For the moment, we will concentrate on the case where our integrand is the value of the left-hand endpoint of the time interval (this is usually called the **Itô integral**), but we will later discuss briefly another case, where we choose the average value of the integrand at the times t_i and t_{i+1}, which is also known as the **Stratonovich integral**. In practice, which integral is more relevant should depend on the application.

To avoid issues such as this, we will from now on assume that the process φ_t is **adapted**: that is, for each time t, φ_t is a random variable which depends only on the information available at time t — the process φ_t is not allowed to 'look ahead' and use information from the future. Using the notation from Section 7.4, so that $\mathcal{F}_t$ represents the information up to time t, note that this means in particular that $\mathbb{E}[\varphi_t|\mathcal{F}_s] = \varphi_t$ for all $s \geq t$ — i.e. φ_t is the best guess at the value of φ_t given all the information known at time $s \geq t$. Note also that we (implicitly) assume that our Brownian motion W_t is known at time t.

MT

Strictly speaking, at this point we should introduce a filtered probability space, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, such that the Brownian motion W_t is $\mathcal{F}_t$-adapted, and moreover, such that $W_s - W_t$ is independent of $\mathcal{F}_t$ for all $s \geq t$. Alternatively, one could take a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$, and a Brownian motion, and define a filtration $\mathcal{F}_t^W$ by setting $\mathcal{F}_t^W$ to be the smallest σ-algebra such that W_s is $\mathcal{F}_t^W$-measurable, for all $s \leq t$. Such a filtration is called the **natural** filtration.

In fact, when constructing the stochastic integral, it is usual to take a filtration satisfying the usual conditions: that is, we assume that is, we assume that in addition $\mathcal{F}_t$ is right-continuous (i.e. $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$), and $\mathcal{F}_0$ contains all the $\mathbb{P}$-negligible events (i.e. $\mathbb{P}(F) = 0, F \in \mathcal{F} \implies F \in \mathcal{F}_0$). It is easy to see how such a filtration can be constructed from the natural filtration by including all the negligible events in $\mathcal{F}_0$ and taking the right-limit of the resulting filtration. It can be shown that in this filtration, W_t remains a Brownian motion.

3.1 Construction of the integral for simple processes

The strategy we will follow is to consider a class $\mathcal{V}$ of processes satisfying some natural conditions. Within this class, we can identify a natural subclass of *simple processes* for which the Itô integral can be easily defined. Then our goal will be to show that there is a natural manner in which a general element of $\mathcal{V}$ can be approximated by simple processes, and moreover, that a limit of the resulting integrals also exists in a natural sense, in which case we can define the limit of the random variables to be the Itô integral of the limiting $\varphi \in \mathcal{V}$.

Definition 3.1. We define the set $\mathcal{V}$ to be the set of processes — that is functions $\varphi : \mathbb{R}_+ \times \Omega \to \mathbb{R}$, usually written[5] $\varphi_t(\omega) := \varphi(t, \omega)$, such that

i) $\varphi_t(\omega)$ is adapted,

ii) $\mathbb{E}\left[\int_0^\infty \varphi_t^2 \,\mathrm{d}t\right] < \infty$.

MT

And really we need an additional measurability criteria, which is that $\varphi : (t, \omega) \to \varphi_t(\omega)$ is $\mathcal{B}([0,t]) \times \mathcal{F}$-measurable, where $\mathcal{B}([0,t])$ denotes the Borel σ-algebra. Processes satisfying this criteria are called **progressively measurable**.

Note that this is strictly stronger than the class of adapted processes, in particular, the measurability criteria above implies adaptedness. However, any adapted, measurable process (i.e. $\varphi : \mathcal{B}(\mathbb{R}_+) \times \mathcal{F} \to \mathbb{R}$ is measurable) has a progressively measurable modification, and moreover if the paths of the process φ are either left-continuous, or right-continuous almost surely, then φ is also progressively measurable. See [1, Propositions 1.13, 1.14]. To be rigorous, a formal treatment should replace 'adapted' by 'progressively measurable' in what follows.

If $\varphi \in \mathcal{V}$ has the form:

$$\varphi_t(\omega) = \sum_{j=0}^{N-1} e_j(\omega) \mathbf{1}_{[t_j, t_{j+1})}(t),$$

for some sequence of (fixed) times $t_0 \le t_1 \le \cdots \le t_N$, then we say that φ_t is **simple** and write $\varphi \in \mathcal{S}$. Note that the assumption that φ_t is adapted implies e_j is known at time t_j.

Now we are in a position to define the stochastic integral for simple processes:

Definition 3.2. Let φ_t be a simple process. Then the **stochastic integral** of φ_t with respect to a Brownian motion W_t is defined to be:

$$\int_0^\infty \varphi_t \,\mathrm{d}W_t = \sum_{j=0}^{N-1} e_j (W_{t_{j+1}} - W_{t_j}). \tag{8}$$

More generally, if $u \le v$ are fixed times, we can define the stochastic integral:

$$\int_u^v \varphi_t \,\mathrm{d}W_t = \int_0^\infty \varphi_t \mathbf{1}_{[u,v)}(t) \,\mathrm{d}W_t, \tag{9}$$

where it is easily verified that φ_t is simple implies $\varphi_t \mathbf{1}_{[u,v)}(t)$ is also simple.

[5] Or also commonly as just φ_t; as is common in probability, we often omit the ω unless we want to emphasise the fact that this is a random variable.

3.2 Properties of the integral of simple processes

In this section we discuss some useful properties of *simple* processes. One aim will be to show that these properties actually extend to the limit, and so they are also true of the general stochastic integral. However, we will also need some of these properties to show that the limit exists. The main result is the following:

Lemma 3.3. *Suppose $\varphi, \psi \in \mathcal{S}$ and $\alpha, \beta \in \mathbb{R}$, $0 \leq s \leq u \leq t, t \in \mathbb{R}_+ \cup \{\infty\}$. Then*

i) $\int_s^t \varphi_r \,\mathrm{d}W_r = \int_s^u \varphi_r \,\mathrm{d}W_r + \int_u^t \varphi_r \,\mathrm{d}W_r$.

ii) $\int_s^t (\alpha\varphi_r + \beta\psi_r) \,\mathrm{d}W_r = \alpha \int_s^t \varphi_r \,\mathrm{d}W_r + \beta \int_s^t \psi_r \,\mathrm{d}W_r$.

iii) $\mathbb{E}\left[\int_s^t \varphi_r \,\mathrm{d}W_r\right] = \mathbb{E}\left[\int_s^t \varphi_r \,\mathrm{d}W_r | \mathcal{F}_s\right] = 0$.

iv) $t \mapsto \int_0^t \varphi_r \,\mathrm{d}W_r$ *is a continuous, adapted process.*

v) (Itô Isometry:) $\mathbb{E}\left[\left(\int_0^t \varphi_u \,\mathrm{d}W_u\right)^2\right] = \mathbb{E}\left[\int_0^t \varphi_u^2 \,\mathrm{d}u\right]$.

We leave the proofs of many of these to a question sheet. Here we show v). Q2.3

Proof of v). Since φ_t is a simple function, we can write $\varphi_u = \sum_{i=0}^{N-1} \varphi_{t_i} \mathbf{1}_{[t_i, t_{i+1})}(u)$ for some sequence $0 = t_0 \leq t_1 \leq \cdots \leq t_N = t$. Then we have:

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^t \varphi_u \,\mathrm{d}W_u\right)^2\right] &= \mathbb{E}\left[\left(\sum_{i=0}^{N-1} \varphi_{t_i}\left(W_{t_{i+1}} - W_{t_i}\right)\right)^2\right] \\ &= \mathbb{E}\left[\sum_{i,j=0}^{N-1} \varphi_{t_i}\varphi_{t_j}\Delta W_i \Delta W_j\right]\end{aligned}$$

where $\Delta W_i = W_{t_{i+1}} - W_{t_i}$. Now consider $i \neq j$ with $i < j$ say. Then

$$\begin{aligned}\mathbb{E}\left[\varphi_{t_i}\varphi_{t_j}\Delta W_i \Delta W_j\right] &= \mathbb{E}\left[\mathbb{E}\left[\varphi_{t_i}\varphi_{t_j}\Delta W_i \Delta W_j | \mathcal{F}_{t_j}\right]\right] \\ &= \mathbb{E}\left[\varphi_{t_i}\varphi_{t_j}\Delta W_i \mathbb{E}\left[\Delta W_j | \mathcal{F}_{t_j}\right]\right] \\ &= \mathbb{E}\left[\varphi_{t_i}\varphi_{t_j}\Delta W_i \mathbb{E}\left[\Delta W_j\right]\right] \\ &= 0\end{aligned}$$

where we have used the Tower property in the first line, the fact that $\varphi_{t_i}, \varphi_{t_j}$ and ΔW_{t_i} are all known at time t_j, and finally the fact that ΔW_j is independent of $\mathcal{F}_{t_j}$, and has mean zero.

For $i = j$, the terms in the sum simplify to:

$$\begin{aligned}\mathbb{E}\left[\varphi_{t_i}^2 (\Delta W_i)^2\right] &= \mathbb{E}\left[\varphi_{t_i}^2 \mathbb{E}\left[(\Delta W_i)^2 | \mathcal{F}_{t_i}\right]\right] \\ &= \mathbb{E}\left[\varphi_{t_i}^2 \mathbb{E}\left[(\Delta W_i)^2\right]\right] \\ &= \mathbb{E}\left[\varphi_{t_i}^2 (t_{i+1} - t_i)\right] \\ &= \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \varphi_u^2 \,\mathrm{d}u\right].\end{aligned}$$

So

$$\mathbb{E}\left[\left(\int_0^t \varphi_u \, \mathrm{d}W_u\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=j} + \sum_{i\neq j}\right) \varphi_{t_i}\varphi_{t_j}\Delta W_i \Delta W_j\right]$$
$$= \sum_i \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \varphi_u^2 \, \mathrm{d}u\right] = \mathbb{E}\left[\int_0^t \varphi_u^2 \, \mathrm{d}u\right].$$

□

3.3 Itô Isometry and the Stochastic Integral

Throughout this section, we assume that $T \in \mathbb{R}_+ \cup \{\infty\}$ is fixed and with slight abuse of notation, write $\mathcal{V}$ for the set of processes defined on $[0,T]$ such that $\mathbb{E}\left[\int_0^T \varphi_t^2 \, \mathrm{d}t\right] < \infty$. We assume that W is a standard Brownian motion.

The key idea is now to use the Itô isometry to extend the integral from $\mathcal{S}$ to $\mathcal{V}$. An important point is that, for simple integrands, we have a map from $\mathcal{S}$ to the integral $\int_0^t \varphi_u \, \mathrm{d}W_u$, which is an element of the metric space $\mathcal{L}^2(\mathbb{P})$, where

$$\mathcal{L}^2(\mathbb{P}) := \{X : \Omega \to \mathbb{R} | \mathbb{E}\left[X^2\right] < \infty\},$$

which is complete, that is, every Cauchy sequence converges, or equivalently $\mathbb{E}\left[|X_n - X_m|^2\right] \to 0$ as $m, n \to \infty$ implies $X_n \to X$, some $X \in \mathcal{L}^2(\mathbb{P})$.

To exploit this fact, we look for a sequence $\varphi^n \in \mathcal{S}$ approximating $\varphi \in \mathcal{V}$, such that both

$$\mathbb{E}\left[\int_0^T (\varphi_u^n - \varphi_u)^2 \, \mathrm{d}u\right] \to 0, \text{ and } \mathbb{E}\left[\int_0^T (\varphi_u^n - \varphi_u^m)^2 \, \mathrm{d}u\right] \to 0$$

as $n, m \to \infty$. Then, by the completeness of $\mathcal{L}^2(\mathbb{P})$ and the Itô Isometry, there exists a limit of the sequence $\int_0^T \varphi_u^n \, \mathrm{d}W_u$, and we can *define* the integral of φ to be this limit. With this aim in mind, we have:

Theorem 3.4. *Let $\varphi \in \mathcal{V}$. Then there exists a sequence $\varphi^n \in \mathcal{S}$ such that*

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T (\varphi_u - \varphi_u^n)^2 \, \mathrm{d}u\right] = 0. \tag{10}$$

Moreover $I := \lim_{n\to\infty} \int_0^T \varphi_u^n \, \mathrm{d}W_u$ exists (in $\mathcal{L}^2(\mathbb{P})$) and is independent of the choice of the sequence φ^n.

A full proof of this result is technical and difficult. We refer the reader to [1, Lemma 3.2.4] for complete details. We sketch the proof here:

Sketch Proof. We first prove the existence of a suitable approximating sequence in three steps:

i) Suppose $\varphi \in \mathcal{V}$ is bounded (i.e. $\varphi_t(\omega) \le M$ for all t and almost all ω) and continuous in t, for each ω. Then taking $t_i^n := i2^{-n}$, we have

$$\varphi_t^n := \sum_i \varphi_{t_i^n} \mathbf{1}_{[t_i^n, t_{i+1}^n)}(t)$$

defines a process in $\mathcal{S}$, and (10) holds by bounded convergence.

ii) Suppose $\varphi \in \mathcal{V}$ is bounded. Define

$$\varphi_t^n := n \int_{(t-1/n)\vee 0}^t \varphi_u \, \mathrm{d}u.$$

Then (it can be shown — this is not trivial) $\varphi^n \in \mathcal{V}$, φ^n is continuous and (10) holds.

iii) Finally, consider a general $\varphi \in \mathcal{V}$. Then this can be approximated by bounded $\varphi^n \in \mathcal{V}$ by taking

$$\varphi_t^n = (-n) \vee (n \wedge \varphi_t),$$

and we have convergence in the sense of (10) by dominated convergence.

By combining each of the three steps, we are able to provide an approximating sequence of φ by elements of $\mathcal{S}$.

Moreover, for $\varphi, \psi, \xi \in \mathcal{V}$, we have:

$$\mathbb{E}\left[\int_0^T (\varphi_u - \psi_u)^2 \, \mathrm{d}u\right] \leq 2\mathbb{E}\left[\int_0^T (\varphi_u - \xi_u)^2 \, \mathrm{d}u\right] + 2\mathbb{E}\left[\int_0^T (\xi_u - \psi_u)^2 \, \mathrm{d}u\right]$$

where we have used the identity $(x-y)^2 \leq 2(x-z)^2 + 2(y-z)^2$. It follows immediately that, whenever (10) holds, then

$$\mathbb{E}\left[\int_0^T (\varphi_u^m - \varphi_u^n)^2 \, \mathrm{d}u\right] \to 0 \text{ as } m, n \to \infty.$$

In particular, if we write $I^n := \int_0^T \varphi_u^n \, \mathrm{d}W_u$, then, by the Itô Isometry for simple processes, and the fact that $\varphi^n - \varphi^m$ is also simple:

$$\mathbb{E}\left[(I^n - I^m)^2\right] = \mathbb{E}\left[\int_0^T (\varphi_u^n - \varphi_u^m)^2 \, \mathrm{d}u\right] \to 0 \text{ as } m, n \to \infty.$$

Hence I^n forms a Cauchy sequence in $\mathcal{L}^2(\mathbb{P})$, which is a complete space, and so I^n converges to a limit, $I \in \mathcal{L}^2$, say.

Finally, suppose $\varphi^n, \tilde{\varphi}^n$ are both sequences of simple functions, approximating $\varphi \in \mathcal{V}$. By the argument just given, these sequences both converge to limits I and $\tilde{I}$, say. Consider also the sequence ψ^n obtained by interleaving these sequences. This sequence also converges to φ in the sense of (10), and so the corresponding integrals must converge in $\mathcal{L}^2(\mathbb{P})$ to some limit I'. It follows immediately that $I' = I = \tilde{I}$, and so the limit is independent of the approximating sequence chosen. □

Definition 3.5. Let $\varphi \in \mathcal{V}$, fix $T > 0$. Then the **Itô integral** of φ from 0 to T is defined to be

$$\int_0^T \varphi_u \, \mathrm{d}W_u = \lim_{n\to\infty} \int_0^T \varphi_u^n \, \mathrm{d}W_u,$$

where the limit is in $\mathcal{L}^2(\mathbb{P})$, and $\varphi_u^n \in \mathcal{S}$ is a sequence of simple processes satisfying (10).

Corollary 3.6 (Itô Isometry)**.** *If $\varphi \in \mathcal{V}$, then*

$$\mathbb{E}\left[\left(\int_0^T \varphi_u \, \mathrm{d}W_u\right)^2\right] = \mathbb{E}\left[\int_0^T \varphi_u^2 \, \mathrm{d}u\right].$$

Proof. This now follows immediately from the definition of the stochastic integral and the fact that if $X_n \to X$ in $\mathcal{L}^2(\mathbb{P})$, then $\mathbb{E}\left[X_n^2\right] \to \mathbb{E}\left[X^2\right]$. □

Example 3.7. Recall the example at the start of this chapter, where we tried to compute $\int_0^1 W_s \,\mathrm{d}W_s$. Let $t_j^N := jN^{-1}$, and set $\varphi_t^N := \sum_j W_{t_j^N} \mathbf{1}_{[t_j^N, t_{j+1}^N)}(t)$. Then $\varphi^N \in \mathcal{S}$, and it is not hard to show:

$$\mathbb{E}\left[\int_0^1 \left(\varphi_u^N - W_u\right)^2 \,\mathrm{d}u\right] \to 0 \text{ as } N \to \infty.$$

But in (7) we showed

$$\int_0^1 \varphi_u^N \,\mathrm{d}W_u = \sum_{j=0}^{N-1} W_{t_j^N}(W_{t_{j+1}^N} - W_{t_j^N}) = \frac{1}{2}(W_1^2 - W_0^2) - \frac{1}{2}\sum_{k=0}^{N-1}(W_{t_{k+1}^N} - W_{t_k^N})^2,$$

and we know from Theorem 2.7 that the right hand side converges in $\mathcal{L}^2(\mathbb{P})$ to $\frac{1}{2}(W_1^2 - W_0^2) - \frac{1}{2}$. It follows therefore that

$$\int_0^1 W_u \,\mathrm{d}W_u = \frac{1}{2}(W_1^2 - W_0^2) - \frac{1}{2},$$

and more generally, it is easy to check that

$$\int_0^t W_u \,\mathrm{d}W_u = \frac{1}{2}(W_t^2 - W_0^2) - \frac{1}{2}t.$$

iPython Interlude: Go To Notebook

Finally, we observe that:

Corollary 3.8. *All the properties of Lemma 3.3 hold for $\varphi \in \mathcal{V}$.*

In general, these results hold in a very similar manner to the proof of Corollary 3.6, see the example sheet. The one non-trivial statement is the claim that the stochastic integral is con- Q2.3
tinuous as a function of t — in fact, this statement should be more carefully stated as: there exists a modification of the stochastic integral which is continuous (so far, we have only defined the stochastic integral for fixed $T > 0$, and as a limit). In future, we will always *assume* that the continuous modification of the stochastic integral has been chosen. See the example sheet for more details. Q2.8

3.4 Stratonovich Integral

Recall the discussion at the start of this chapter. In particular, we showed that there were different integrals depending on how we chose to approximate the integrand in the limiting sum. The Itô integral arose when we chose the left-hand endpoint of the interval $[t_i, t_{i+1}]$. A more appropriate choice might have been to consider the limit of the sums:

$$\sum_i \left(\frac{\varphi_{t_i} + \varphi_{t_{i+1}}}{2}\right)\left(W_{t_{i+1}} - W_{t_i}\right) \tag{11}$$

rather than

$$\sum_i \varphi_{t_i}\left(W_{t_{i+1}} - W_{t_i}\right).$$

It is possible to show[6], under certain conditions, that the term in (11) converges to a limit in probability, which we denote $\int_0^t \varphi_u \circ \mathrm{d}W_u$. We call this limit the **Stratonovich Integral** of φ_u.

[6]Indeed, we can also consider the approximation $\sum_i \varphi_{t_i^*}\left(W_{t_{i+1}} - W_{t_i}\right)$, where $t_i^* = (t_i + t_{i+1})$, and this gives the same result, as does any similar choice which 'averages' out to give t_i^* over $[t_i, t_{i+1}]$.

Theorem 3.9. *Suppose $\varphi \in \mathcal{V}$, and the Stratonovich integral exists. Then*

$$\int_0^T \varphi_u \circ \mathrm{d}W_u = \int_0^T \varphi_u \, \mathrm{d}W_u + \frac{1}{2}\langle \varphi, W\rangle_T,$$

where

$$\langle \varphi, W\rangle_T := \lim \sum \left(\varphi_{t_{i+1}} - \varphi_{t_i}\right)\left(W_{t_{i+1}} - W_{t_i}\right)$$

is the **(quadratic) covariation** *of φ, W, and the limit is in the sense of convergence in probability, along partitions with mesh going to zero.*

Remark 3.10. Observe that the quadratic variation $\langle X\rangle_T$ is equal to the covariation of X with itself, $\langle X, X\rangle_T$. Moreover, the quadratic covariation, $\langle \varphi, W\rangle$ can be defined in terms of the quadratic variation, using the polarisation identity:

$$\langle \phi, W\rangle_T = \frac{1}{4}\left[\langle \varphi + W\rangle_T - \langle \varphi - W\rangle_T\right].$$

In a specific modelling situation, it now becomes important to understand which of the integrals, the Itô integral, or the Stratonovich integral, is the relevant integral. In practice, this needs a careful understanding of the application, and exactly how the integral arises. For example, in financial applications, one often uses the stochastic integral to model the value of a portfolio, where one integrates against the value of some asset, and ϕ_s becomes the number of units held. Since one cannot know the price in the future, the Itô integral becomes more relevant. On the other hand, for some physical processes, or for applications in filtering, the Stratonovich integral turns out to be more useful. In addition, there are some technical advantages to each integral:

- The Itô integral has certain advantages from a technical perspective (in particular, it can be defined in more general situations).
- The Itô integral of an integrand against a martingale preserves the martingale property. This can be extremely useful for computing expectations.
- The Stratonovich has a nice chain rule: see the problem sheet, and Section 4 for further discussion on 'chain-rules' for stochastic integrals. Q2.6

3.5 Stochastic integration in higher dimensions

Let $\boldsymbol{W} = \left(W^1, \ldots, W^d\right)$ be a d-dimensional Brownian motion, and let $\mathcal{V}^{n\times d}$ denote the set of $n \times d$ matrices, $\boldsymbol{\varphi}_t := \left(\varphi_t^{ij}\right)$, where $\varphi_t^{ij} \in \mathcal{V}$ for each i, j.

Then we define the n-dimensional stochastic integral

$$\boldsymbol{I}_t := \int_0^t \boldsymbol{\varphi}_u \, \mathrm{d}\boldsymbol{W}_u = \begin{pmatrix} I_t^1 \\ I_t^2 \\ \vdots \\ I_t^n \end{pmatrix} \in \mathbb{R}^n,$$

where $I_t^i = \sum_{j=1}^d \int_0^t \varphi_u^{ij} \, \mathrm{d}W_u^j$.

4 Stochastic Calculus

In this section we introduce some rules to manipulate the stochastic integral. As usual, we consider a standard Brownian motion, W_t, and consider processes of the form:

$$X_t = X_0 + \int_0^t \alpha_u \, \mathrm{d}u + \int_0^t \varphi_u \, \mathrm{d}W_u.$$

Our aim is to understand the process $f(t, X_t)$ for such an X_t.

Our first observation is that we can weaken slightly the assumption on φ, so that this is no longer in $\mathcal{V}$. Specifically, consider a process φ which is adapted, and such that

$$\mathbb{P}\left(\int_0^t \varphi_u^2 \, \mathrm{d}u < \infty, \forall t \geq 0\right) = 1, \tag{12}$$

so that this class contains $\mathcal{V}$ as a subset, since $\varphi \in \mathcal{V}$ implies that $\int_0^\infty \varphi_u^2 \, \mathrm{d}u < \infty$ almost surely. On the other hand, given φ_u satisfying (12), we can define $\varphi_u^N \in \mathcal{V}$ by:

$$\varphi_t^N := \begin{cases} \varphi_t & \int_0^t \varphi_u^2 \, \mathrm{d}u \leq N \\ 0 & \int_0^t \varphi_u^2 \, \mathrm{d}u > N \end{cases}$$

it is easily checked that $\varphi_t^N \in \mathcal{V}$. Moreover, it is easy to show that $\int_0^t \varphi_u^N \, \mathrm{d}W_u = \int_0^t \varphi_u^{N'} \, \mathrm{d}W_u$ whenever $\int_0^t \varphi_u^2 \, \mathrm{d}u \leq \max\{N, N'\}$, since we may take approximating sequences of simple functions for φ^N and $\varphi^{N'}$ which agree on this event. Hence, $\lim_{N\to\infty} \int_0^t \varphi_u^N \, \mathrm{d}W_u$ is well defined for all $t \geq 0$, whenever $\int_0^t \varphi_u^2 \, \mathrm{d}u \leq N$ for some N. By (12), this is a set of probability 1, so we can define $\int_0^t \varphi_u \, \mathrm{d}W_u$ to be the limit *in probability* of the sequence $\int_0^t \varphi_u^N \, \mathrm{d}W_u$.

However, we note that this sequence may *not* converge in $\mathcal{L}^2(\mathbb{P})$. Q3.8

Denote the set of adapted processes satisfying (12) by $\mathcal{H}$. Moreover, write $\hat{\mathcal{H}}$ for the class of adapted processes α_u for which the integral of the absolute value is almost surely finite for all t, i.e. $\alpha_t \in \hat{\mathcal{H}}$ if and only if:

$$\mathbb{P}\left(\int_0^t |\alpha_u| \, \mathrm{d}u < \infty, \forall t \geq 0\right) = 1. \tag{13}$$

We note that it can be shown that any $\varphi \in \mathcal{H}$ and $\alpha \in \hat{\mathcal{H}}$ can be approximated by simple functions such that $\int_0^t |\varphi_u^N - \varphi_u|^2 \, \mathrm{d}u$ converges to zero in probability, and then the limit of the integrals of the simple functions converge in probability to the integral of φ_u as above (and hence, uniquely).

The reason for doing this is that we can now define a class of processes X_t which are represented as integrals in time and a stochastic integral:

Definition 4.1. A (one-dimensional) **Itô process** is a stochastic process on $\mathbb{R}$ of the form:

$$X_t = X_0 + \int_0^t \alpha_u \, \mathrm{d}u + \int_0^t \varphi_u \, \mathrm{d}W_u, \tag{14}$$

for some process $\varphi \in \mathcal{H}$ and $\alpha \in \hat{\mathcal{H}}$.

We will sometimes also refer to such a process as a stochastic integral.

Then, for example, $W_t \in \mathcal{H}$, and $1/2 \in \hat{\mathcal{H}}$, so

$$\frac{W_t^2}{2} = \frac{W_0^2}{2} + \int_0^t \frac{1}{2}\,\mathrm{d}u + \int_0^t W_u\,\mathrm{d}W_u$$

and hence $W_t^2/2$ is an Itô process. We often write this in the abbreviated form:

$$\mathrm{d}\left(\frac{W_t^2}{2}\right) = \frac{1}{2}\,\mathrm{d}t + W_t\,\mathrm{d}W_t,$$

or more generally, (14) becomes:

$$\mathrm{d}X_t = \alpha_t\,\mathrm{d}t + \varphi_t\,\mathrm{d}W_t.$$

Note also that the choice of X_0 in (14) can be treated as an 'initial condition', so we may often look at an equation of the form above with the initial condition $X_0 = x_0$, for some constant x_0. In some situations, we may also be interested in a random starting point, so X_0 may be given as some random variable (in which case, we need the randomness to be included in the information available at time 0).

With this in mind, we are able to extend the notion of a stochastic integral against Brownian motion to the case where our integrand is an Itô process:

Definition 4.2. Let X_t be an Itô process with representation

$$X_t = X_0 + \int_0^t \alpha_u\,\mathrm{d}u + \int_0^t \varphi_u\,\mathrm{d}W_u,$$

and suppose ψ_u is an adapted process in the set

$$\mathcal{H}[X] := \left\{\psi : \int_0^t |\alpha_u\psi_u|\,\mathrm{d}u + \int_0^t (\varphi_u\psi_u)^2\,\mathrm{d}u < \infty, \mathbb{P} - as \text{ for all } t \geq 0\right\}.$$

Then we define $Y_t := \int_0^t \psi_u\,\mathrm{d}X_u$ by

$$Y_t = \int_0^t \psi_u\alpha_u\,\mathrm{d}u + \int_0^t \psi_u\varphi_u\,\mathrm{d}W_u.$$

Remark 4.3. Taking this definition, we can extend the stochastic integral to the set of Itô processes. Of course, a natural question is whether this still corresponds to our original definition for Brownian motion, and this can indeed be shown: if $\psi \in \mathcal{H}[X]$, and ψ^n is a sequence of simple processes in $\mathcal{H}[X]$ such that $\int_0^t(\psi_u^n - \psi_u)^2\varphi_u^2\,\mathrm{d}u \to 0$ and $\int_0^t |\psi_u - \psi_u^n|\alpha_u\,\mathrm{d}u \to 0$ in probability, for all $t \geq 0$, then the limit in probability of the discrete sums:

$$\sum_i \psi_{t_i}^n (X_{t_{i+1}} - X_{t_i})$$

exists in probability, and agrees with the integral defined above.[7]

Note also the following corollary:

Corollary 4.4. *Let X be an Itô process, and for $\psi \in \mathcal{H}[X]$ define $Y := \int \psi_u\,\mathrm{d}X_u$. Then Y_t is an Itô process, and if $\xi \in \mathcal{H}[Y]$, then $\xi_u\psi_u \in \mathcal{H}[X]$ and*

$$\int \xi_u\,\mathrm{d}Y_u = \int \xi_u\psi_u\,\mathrm{d}X_u.$$

[7] The approach we take here is slightly non-standard. It is more common to define the integral in general in this manner, and then to show that as a consequence, the integral against an Itô process has this defining property.

4.1 Ito's Lemma

The main reason for considering Itô processes is that this class of processes are closed when we apply a (suitably nice) function to the process, and moreover, we can explicitly compute their form:

Theorem 4.5 (Itô's Lemma). *Let X_t be an Itô process given by* (14)*, and suppose $g \in C^2([0,\infty)\times\mathbb{R})$ (i.e. $g(t,x)$ is twice continuously differentiable in both t and x). Then*

$$Y_t := g(t, X_t)$$

is an Itô process, and

$$\mathrm{d}Y_t = \left(\frac{\partial g}{\partial t}(t,X_t) + \alpha_t \frac{\partial g}{\partial x}(t,X_t) + \frac{1}{2}\varphi_t^2 \frac{\partial^2 g}{\partial x^2}(t,X_t)\right)\mathrm{d}t + \varphi_t \frac{\partial g}{\partial x}(t,X_t)\,\mathrm{d}W_t. \tag{15}$$

An important interpretation of (15) is the following: we know from Theorem 2.7, that (roughly) $\mathrm{d}W_t \sim \sqrt{\mathrm{d}t}$ and $(\mathrm{d}W_t)^2 \sim \mathrm{d}t$, so e.g. $\mathrm{d}W_t\,\mathrm{d}t \sim (\mathrm{d}W_t)^3 \sim (\mathrm{d}t)^{3/2}$, and hence we might expect these latter terms to be zero, so:

$$\begin{aligned}(\mathrm{d}X_t)^2 &= (\alpha_t\,\mathrm{d}t + \varphi_t\,\mathrm{d}W_t)\\ &= \alpha_t^2\,\mathrm{d}t^2 + 2\alpha_t\varphi_t\,\mathrm{d}W_t\,\mathrm{d}t + \varphi_t^2\,(\mathrm{d}W_t)^2\\ &= \varphi_t^2\,\mathrm{d}t.\end{aligned}$$

It is easy then to check that (15) is exactly what we get if we expand

$$\mathrm{d}Y_t = \frac{\partial g}{\partial t}(t,X_t)\,\mathrm{d}t + \frac{\partial g}{\partial x}(t,X_t)\,\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)(\mathrm{d}X_t)^2$$

using these rules. The interpretation of the formula above is that this is the second order Taylor expansion of $g(t,X_t)$ over a small time step, and using the additional approximations that $(\mathrm{d}t)^2 = (\mathrm{d}t\,\mathrm{d}X_t) = 0$ (which are also consequences of similar expansions).

We prove the theorem under the slightly stronger assumption that $g \in C^3$, but this is not necessary.[8]

Proof of Theorem 4.5. We first observe that Y_t is indeed an Itô process, and specifically, that $\beta_t \in \hat{\mathcal{H}}$, $\psi \in \mathcal{H}$, where:

$$\begin{aligned}\beta_t &:= \frac{\partial g}{\partial t}(t,X_t) + \alpha_t \frac{\partial g}{\partial x}(t,X_t) + \frac{1}{2}\varphi_t^2 \frac{\partial^2 g}{\partial x^2}(t,X_t)\\ \psi_t &:= \varphi_t \frac{\partial g}{\partial x}(t,X_t).\end{aligned}$$

Note first that to show (12), (13), it is sufficient to show e.g. that $\int_0^t \varphi_u^2\,\mathrm{d}u < \infty$ with probability 1 for all $t > 0$. So fix $t > 0$, and then our first observation is that, since X_t has continuous paths, the sets

$$A_n := \{\omega : |X_u(\omega)| \le n, \forall u \le t\}$$

satisfy $\mathbb{P}(A_n) \to 1$. It suffices therefore to verify (12), (13) for finite t, on each A_n. But now by the continuity of the derivatives, each of the partial derivatives are bounded on $[0,t]\times[-n,n]$,

[8] See e.g. [1, Theorem 3.3]; in fact, we only require $g \in C^{1,2}([0,\infty)\times\mathbb{R})$ for the theorem to hold.

and so it is sufficient to observe that, with probability 1, each of $\int_0^t \left(1+|\alpha_u|+|\varphi_u^2|\right) \mathrm{d}u$ and $\int_0^t \varphi_u^2 \mathrm{d}u$ are all finite almost surely.

We first make some simplifying comments: first, by multiplying by a smooth function $\eta(t,x)$ such that $\eta(t,x) = 1$ for $\max\{|x|,t\} < N$, and $\eta(t,x) = 0$ for $\max\{|x|,t\} > N+1$, we can assume that g is bounded, and has bounded derivatives up to order 3.

Second, recalling the discussion at the start of the chapter, it is sufficient to show the result when α_t, φ_t are simple functions such that α_u, φ_u are bounded. In particular, if we can take a sequence of α, φ converging to some limits in probability, it follows from the assumption of bounded derivatives, that the corresponding terms β_t, ψ_t converge in probability, and hence the integrals are indeed the same.

So we consider the case where α_u, φ_u are simple, and consider a sequence of partitions Π^n of $[0,t]$, $0 = t_0^n \le t_1^n \le \cdots \le t_{m_n}^n = t$ such that α, φ are constant for $t \in [t_i^n, t_{i+1}^n)$ (so in particular, these are all refinements of the usual partitions associated with α, φ). We then consider a sequence with $||\Pi^n|| \to 0$ as $n \to \infty$.

Write $\alpha_{j,n} = \alpha_{t_j^n}$, $\varphi_{j,n} = \varphi_{t_j^n}$, $\Delta t_{j,n} = t_{j+1}^n - t_j^n$, $\Delta X_{j,n} = X_{t_{j+1}^n} - X_{t_j^n}, \Delta W_{j,n} = W_{t_{j+1}^n} - W_{t_j^n}$, and then by Taylor's Theorem

$$
\begin{aligned}
g(t,X_t) &= g(0,X_0) + \sum_{j=0}^{m_n-1} \left(g(t_{j+1}^n, X_{t_{j+1}^n}) - g(t_j^n, X_{t_j^n}) \right) \\
&= g(0,X_0) + \sum_{j=0}^{m_n} \frac{\partial g}{\partial t}(t_j^n, X_{t_j^n}) \Delta t_{j,n} + \sum_{j=0}^{m_n} \frac{\partial g}{\partial x}(t_j^n, X_{t_j^n}) \Delta X_{j,n} \\
&\quad + \frac{1}{2} \sum_{j=0}^{m_n} \frac{\partial^2 g}{\partial t^2}(t_j^n, X_{t_j^n})(\Delta t_{j,n})^2 + \sum_{j=0}^{m_n} \frac{\partial^2 g}{\partial t \partial x}(t_j^n, X_{t_j^n})(\Delta t_{j,n})(\Delta X_{j,n}) \\
&\quad + \frac{1}{2} \sum_{j=0}^{m_n} \frac{\partial^2 g}{\partial x^2}(t_j^n, X_{t_j^n})(\Delta X_{j,n})^2 + \sum_{j=0}^{m_n} R_j
\end{aligned}
$$

where R_n are remainder terms, in particular, $R_j = o(|\Delta t_{j,n}|^2 + |\Delta X_{j,n}|^2)$, uniformly over j, n by the assumption that $g \in C^3$.

As we let $n \to \infty$, so $\Delta t_{j,n} \to 0$, we get the convergence in probabilities of:

$$
\sum_{j=0}^{m_n} \frac{\partial g}{\partial t}(t_j^n, X_{t_j^n}) \Delta t_{j,n} \to \int_0^t \frac{\partial g}{\partial t}(u, X_u)\, \mathrm{d}u \quad \text{and} \quad \sum_{j=0}^{m_n} \frac{\partial g}{\partial x}(t_j^n, X_{t_j^n}) \Delta X_{j,n} \to \int_0^t \frac{\partial g}{\partial x}(u, X_u)\, \mathrm{d}W_u
$$

by the fact that $\frac{\partial g}{\partial t}$ is bounded, and Remark 4.3. Also $\sum_{j=0}^{m_n} R_j \to 0$ in probability.

Since $\frac{\partial^2 g}{\partial t^2}$ is bounded (by g^* say), we also have $\sum_{j=0}^{m_n} \left| \frac{\partial^2 g}{\partial t^2}(t_j^n, X_{t_j^n})(\Delta t_{j,n})^2 \right| \le g^* ||\Pi^n|| t \to 0$, and also

$$
\sum_{j=0}^{m_n} \frac{\partial^2 g}{\partial t \partial x}(t_j^n, X_{t_j^n})(\Delta t_{j,n}) \mathbf{1}_{[t_j^n, t_{j+1}^n)}(t) \to 0
$$

in the sense of Remark 4.3, so that also this term converges to zero.

Finally, we need to consider the term $\sum_{j=0}^{m_n} \frac{\partial^2 g}{\partial x^2}(t_j^n, X_{t_j^n})(\Delta X_{j,n})^2$. By assumption, since α, φ are simple functions, we have:

$$\begin{aligned}\sum_{j=0}^{m_n} \frac{\partial^2 g}{\partial x^2}(t_j^n, X_{t_j^n})(\Delta X_{j,n})^2 &= \sum_{j=0}^{m_n} \frac{\partial^2 g}{\partial x^2}(t_j^n, X_{t_j^n})(\alpha_{j,n}\Delta t_{j,n} + \varphi_{j,n}\Delta W_{j,n})^2 \\ &= \sum_{j=0}^{m_n} \frac{\partial^2 g}{\partial x^2}\alpha_{j,n}^2(\Delta t_{j,n})^2 + 2\sum_{j=0}^{m_n} \frac{\partial^2 g}{\partial x^2}\alpha_{j,n}(\Delta t_{j,n})\varphi_{j,n}(\Delta W_{j,n}) \\ &\quad + \sum_{j=0}^{m_n} \frac{\partial^2 g}{\partial x^2}\varphi_{j,n}^2(\Delta W_{j,n})^2.\end{aligned}$$

Again, we observe that for the first term on the right hand side, $\sum_{j=0}^{m_n} \frac{\partial^2 g}{\partial x^2}\alpha_{j,n}^2(\Delta t_{j,n})^2$, by boundedness of the g and α terms, this converges to zero as $\Delta t \to 0$. Moreover, by a similar reasoning to above, using Remark 4.3, the second term goes to zero. It only remains to consider the final term. Write $\gamma_t := \frac{\partial^2 g}{\partial x^2}(t, X_t)\varphi_t^2$, and $\gamma_{j,n} := \gamma_{t_j^n}$. We show the convergence in $\mathcal{L}^2(\mathbb{P})$:

$$\sum_{j=0}^{m_n} \gamma_{j,n}(\Delta W_{j,n})^2 \to \int_0^t \gamma_u \,\mathrm{d}u.$$

The argument follows closely that of Theorem 2.7. Consider:

$$\begin{aligned}&\mathbb{E}\left[\left(\sum_{j=0}^{m_n} \gamma_{j,n}(\Delta W_{j,n})^2 - \sum_{j=0}^{m_n} \gamma_{j,n}(\Delta t_{j,n})\right)^2\right] \\ &\quad = \sum_{i,j} \mathbb{E}\left[\gamma_{j,n}\gamma_{i,n}\left((\Delta W_{j,n})^2 - \Delta t_{j,n}\right)\left((\Delta W_{i,n})^2 - \Delta t_{i,n}\right)\right] \\ &\quad = \sum_{j} \mathbb{E}\left[\gamma_{j,n}^2\left((\Delta W_{j,n})^2 - \Delta t_{j,n}\right)^2\right] \\ &\quad = \sum_{j} \mathbb{E}\left[\gamma_{j,n}^2\right]\mathbb{E}\left[\left((\Delta W_{j,n})^2 - \Delta t_{j,n}\right)^2\right] \\ &\quad = 2\sum_{j} \mathbb{E}\left[\gamma_{j,n}^2\right](\Delta t_{j,n})^2 \to 0.\end{aligned}$$

Here, we have used in the second line the tower property, and the fact that

$$\mathbb{E}\left[\left((\Delta W_{j,n})^2 - \Delta t_{j,n}\right)\right] = 0$$

(see also Theorem 2.7 for a similar calculation with extra details). Then the subsequent line follows by independence of the Brownian increment, and adaptedness of γ. Next we use the fact that $\mathbb{E}\left[\left((\Delta W_{j,n})^2 - \Delta t_{j,n}\right)^2\right] = 2(\Delta t_{j,n})^2$, which was proved in Theorem 2.7. Finally, $\sum_{j=0}^{m_n} \gamma_{j,n}(\Delta t_{j,n}) \to \int_0^t \gamma_s \,\mathrm{d}s$ in $\mathcal{L}^2(\mathbb{P})$, since γ_s is bounded, and $\Delta t_{j,n} \to 0$ as $n \to \infty$. □

Example 4.6. As a very simple example, we consider the case where $Y_t = W_t^2$, so $f(t,x) = x^2$ and $X_t = W_t$. Then we get:

$$\mathrm{d}Y_t = \frac{\partial f}{\partial x}(t, W_t)\,\mathrm{d}W_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(\mathrm{d}W_t)^2 = 2W_t\,\mathrm{d}W_t + \mathrm{d}t$$

which we can integrate to get the (by now) familiar:

$$W_t^2 = W_0^2 + 2\int_0^t W_s\,\mathrm{d}W_s + t.$$

4.2 Multi-dimensional Itô formula

Let $\boldsymbol{W}$ be a d-dimensional Brownian motion, and consider the n-dimensional process $\boldsymbol{\alpha}_t = (\alpha^1_t, \dots, \alpha^n_t)$ and the $n \times d$-dimensional process $\boldsymbol{\varphi}_t = (\varphi^{i,j}_t)_{1 \le i \le n, 1 \le j \le d}$, such that $\alpha^i \in \hat{\mathcal{H}}$, $\varphi^{i,j} \in \mathcal{H}$. Then we define the n-dimensional Itô process $\boldsymbol{X}_t = (X^1_t, X^2_t, \dots, X^n_t)$ by

$$\mathrm{d}X^i_t = \alpha^i_t \,\mathrm{d}t + \varphi^{i,1}_t \,\mathrm{d}W^1_t + \cdots + \varphi^{i,d}_t \,\mathrm{d}W^d_t, \tag{16}$$

or in matrix notation,

$$\mathrm{d}\boldsymbol{X}_t = \boldsymbol{\alpha}_t \,\mathrm{d}t + \boldsymbol{\varphi}_t \,\mathrm{d}\boldsymbol{W}_t. \tag{17}$$

Then there is a multi-dimensional version of Itô's Lemma, which states:

Theorem 4.7. *Suppose $\boldsymbol{X}_t$ is a n-dimensional Itô process, given by* (17). *Let $g : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^m$ be a C^2 function. Then $\boldsymbol{Y}_t := g(t, \boldsymbol{X}_t)$ is an m-dimensional Itô process, given by $(Y^1_t, \dots, Y^m_t)$, where*

$$\mathrm{d}Y^i_t = \frac{\partial g_i}{\partial t}(t, \boldsymbol{X}_t) \,\mathrm{d}t + \sum_{j=1}^n \frac{\partial g_i}{\partial x_j}(t, \boldsymbol{X}_t) \mathrm{d}X^j_t + \frac{1}{2} \sum_{j,k=1}^n \frac{\partial^2 g_i}{\partial x_j \partial x_k}(t, \boldsymbol{X}_t) \,\mathrm{d}X^j_t \mathrm{d}X^k_t, \tag{18}$$

where $g(t, x) = (g_1(t, x), \dots, g_m(t, x))$, and where $\mathrm{d}X^j_t \,\mathrm{d}X^k_t$ are computed by expanding (16), *and using the rules:*

$$\mathrm{d}W^j_t \,\mathrm{d}W^k_t = \delta_{j,k} \,\mathrm{d}t, \quad \mathrm{d}W^j_t \,\mathrm{d}t = \mathrm{d}t \,\mathrm{d}W^j_t = (\mathrm{d}t)^2 = 0. \tag{19}$$

In essence, this result is proved in the same manner as the one-dimensional version of the result. The main additional concern comes from the 'cross'-terms, $\Delta W^j \Delta W^k$. In particular, we have an additional term (as in the proof of Theorem 4.5) of the form $\sum_{i=0}^{m_n} \gamma_{i,n} (\Delta W^j_{i,n})(\Delta W^k_{i,n})$. In the case where $j = k$, this converges as above to the time-integral of γ, and hence the claim $\mathrm{d}W^j_t \,\mathrm{d}W^j_t = \mathrm{d}t$ is justified.

In the case where $j \neq k$, we consider

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=0}^{m_n} \gamma_{i,n} (\Delta W^j_{i,n})(\Delta W^k_{i,n})\right)^2\right] &= \mathbb{E}\left[\sum_{i,l=0}^{m_n} \gamma_{i,n}\gamma_{l,n} (\Delta W^j_{i,n})(\Delta W^k_{l,n})(\Delta W^j_{l,n})(\Delta W^k_{i,n})\right] \\
&= \sum_{i=0}^{m_n} \mathbb{E}\left[\gamma^2_{i,n} (\Delta W^j_{i,n})^2 (\Delta W^k_{i,n})^2\right] \\
&= \sum_{i=0}^{m_n} \mathbb{E}\left[\gamma^2_{i,n}\right] (\Delta t_{i,n})^2
\end{aligned}$$

where in the second line, we use the independence of the increments of W^j and W^k (recall, these are assumed independent processes), and the Tower property, and in the final line, we use independence, and conditioning again (since $\mathbb{E}\left[(\Delta W^j_{i,n})^2\right] = \mathbb{E}\left[(\Delta W^k_{i,n})^2\right] = \Delta t_{i,n}$). It follows from the convergence of the mesh of the partition to zero that this last term goes to zero.

4.3 Integration by Parts

We can deduce from the multi-dimensional Itô formula a particularly useful one-dimensional result: let X_t, Y_t be Itô processes, say:

$$\begin{aligned}
\mathrm{d}X_t &= \alpha_t \,\mathrm{d}t + \varphi_t \,\mathrm{d}W_t \\
\mathrm{d}Y_t &= \beta_t \,\mathrm{d}t + \psi_t \,\mathrm{d}W_t.
\end{aligned}$$

Then we can apply Theorem 4.7 in the case where $n = 2$ and $d = 1$, with $g(t, x, y) = xy$ to get the **Integration by Parts formula**:

$$\mathrm{d}(X_t Y_t) = X_t\,\mathrm{d}Y_t + Y_t\,\mathrm{d}X_t + \mathrm{d}X_t\,\mathrm{d}Y_t$$

where $\mathrm{d}X_t\,\mathrm{d}Y_t$ is computed using the usual rules:

$$\mathrm{d}W_t\,\mathrm{d}W_t = \mathrm{d}t, \quad \mathrm{d}W_t\,\mathrm{d}t = \mathrm{d}t\,\mathrm{d}W_t = (\mathrm{d}t)^2 = 0. \tag{20}$$

In particular, we have:

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s\,\mathrm{d}Y_s + \int_0^t Y_s\,\mathrm{d}X_s + \int_0^t \varphi_s \psi_s\,\mathrm{d}s.$$

4.4 Rules of Stochastic Calculus

We summarise here the previous results, and how they allow us to manipulate Itô processes:

Suppose X_t and Y_t are Itô processes with representation

$$\begin{aligned} \mathrm{d}X_t &= \alpha_t\,\mathrm{d}t + \varphi_t\,\mathrm{d}W_t \\ \mathrm{d}Y_t &= \beta_t\,\mathrm{d}t + \psi_t\,\mathrm{d}W_t \end{aligned}$$

where all processes are assumed to be sufficiently well-behaved. Then:

- $Z_t = \int_0^t \gamma_s\,\mathrm{d}X_s$ is given by

$$\mathrm{d}Z_t = \gamma_t\,\mathrm{d}X_t = \gamma_t(\alpha_t\,\mathrm{d}t + \varphi_t\,\mathrm{d}W_t) = \gamma_t\alpha_t\,\mathrm{d}t + \gamma_t\varphi_t\,\mathrm{d}W_t.$$

- *Itô's Lemma*:

$$\mathrm{d}g(t, X_t) = \frac{\partial g}{\partial t}(t, X_t)\,\mathrm{d}t + \frac{\partial g}{\partial x}(t, X_t)\,\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)(\mathrm{d}X_t)^2$$

  where $(\mathrm{d}X_t)^2$ can be calculated using the rules:

$$\mathrm{d}W_t\,\mathrm{d}W_t = \mathrm{d}t, \quad \mathrm{d}W_t\,\mathrm{d}t = \mathrm{d}t\,\mathrm{d}W_t = (\mathrm{d}t)^2 = 0.$$

  Note in particular that this is essentially a formal Taylor expansion of g, combined with these rules. The general (multi-dimensional) version can be derived in an identical manner.

- *Integration by parts*:

$$\mathrm{d}(X_t Y_t) = X_t\,\mathrm{d}Y_t + Y_t\,\mathrm{d}X_t + \mathrm{d}X_t\,\mathrm{d}Y_t$$

  where, again, $\mathrm{d}X_t\,\mathrm{d}Y_t$ can be calculated using the same rules as above.

5 Stochastic Differential Equations

We now return to the study of equations with an unknown X appearing on both sides. Specifically, we say that X_t solves the **Stochastic Differential Equation** (SDE):

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \qquad X_0 = x_0 \tag{21}$$

if $X_t = x_0 + \int_0^t \mu(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s$ for the given functions μ, σ. Note that we can have x_0 to be a random variable, usually assumed independent of the Brownian motion W_t.

5.1 Simple Examples

Example 5.1 (Geometric Brownian Motion)**.** Fix $\mu \in \mathbb{R}$, $x_0, \sigma > 0$. Let X_t be given by

$$X_t = x_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right\}$$

so $X_t = f(t, W_t)$ where $f(t, x) = x_0 \exp\{(\mu - \frac{1}{2}\sigma^2)t + \sigma x\}$. Then

$$\frac{\partial f}{\partial x} = \sigma f(t, x), \quad \frac{\partial^2 f}{\partial x^2} = \sigma^2 f(t, x) \quad \frac{\partial f}{\partial t} = \left(\mu - \frac{1}{2}\sigma^2\right) f(t, x).$$

So, by Itô's Lemma, Theorem 4.5, we have:

$$\begin{aligned}
\mathrm{d}X_t &= \mathrm{d}f(t, W_t) \\
&= \left(\mu - \frac{1}{2}\sigma^2\right) f(t, W_t)\,\mathrm{d}t + \sigma f(t, W_t)\,\mathrm{d}W_t + \frac{1}{2}\sigma^2 f(t, W_t)\,(\mathrm{d}W_t)^2 \\
&= \mu X_t\,\mathrm{d}t + \sigma X_t \mathrm{d}W_t.
\end{aligned}$$

In particular, this process satisfies an SDE with $\mu(t, x) = \mu x, \sigma(t, x) = \sigma x$, and we call such a process **Geometric Brownian motion**.

Example 5.2 (Ornstein Uhlenbeck process)**.** Let $\lambda, \sigma \in \mathbb{R}$. Consider the process

$$Y_t = \sigma \int_0^t \mathrm{e}^{\lambda s}\,\mathrm{d}W_s.$$

Then $\mathrm{d}Y_t = \sigma \mathrm{e}^{\lambda t}\,\mathrm{d}W_t$. Now consider $X_t := \mathrm{e}^{-\lambda t} Y_t$. By Itô's Lemma (applied to $f(t, y) = \mathrm{e}^{-\lambda t} y$) we have

$$\begin{aligned}
\mathrm{d}X_t &= \mathrm{e}^{-\lambda t}\,\mathrm{d}Y_t - \lambda \mathrm{e}^{-\lambda t} Y_t\,\mathrm{d}t \\
&= -\lambda X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.
\end{aligned}$$

In particular, we have $X_t = \mathrm{e}^{-\lambda t} x_0 + \sigma \int_0^t \mathrm{e}^{(s-t)\lambda}\,\mathrm{d}W_s$ solves the SDE:

$$\mathrm{d}X_t = -\lambda X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_0 = x_0.$$

Example 5.3 (Exponential Martingales)**.** Let

$$X_t = x_0 \exp\left\{\int_0^t \alpha_s\,\mathrm{d}W_s - \frac{1}{2}\int_0^t \alpha_s^2\,\mathrm{d}s\right\}, \quad x_0 \in \mathbb{R}.$$

Then, writing $Y_t = \int_0^t \alpha_s \,\mathrm{d}W_s - \frac{1}{2}\int_0^t \alpha_s^2 \,\mathrm{d}s$, we get $X_t = x_0 \exp(Y_t)$, and so by Itô's Lemma,

$$\begin{aligned}
\mathrm{d}X_t &= X_t \,\mathrm{d}Y_t + \frac{1}{2}X_t \,(\mathrm{d}Y_t)^2 \\
&= X_t(\alpha_t \,\mathrm{d}W_t - \frac{1}{2}\alpha_t^2 \,\mathrm{d}t) + \frac{1}{2}X_t \left(\alpha_t \,\mathrm{d}W_t - \frac{1}{2}\alpha_t^2 \,\mathrm{d}t\right)^2 \\
&= X_t\alpha_t \,\mathrm{d}W_t + \left(\frac{1}{2} - \frac{1}{2}\right)\alpha_t^2 X_t \,\mathrm{d}t \\
&= X_t\alpha_t \,\mathrm{d}W_t.
\end{aligned}$$

Note that, provided $\mathbb{E}\left[\int_0^t \alpha_t^2 X_t^2 \,\mathrm{d}t\right] < \infty$, we have:

$$\mathbb{E}\left[\int_0^t \alpha_s X_s \,\mathrm{d}W_s\right] = 0$$

by Corollary 3.6, so:

$$\begin{aligned}
\mathbb{E}\left[X_t\right] &= \mathbb{E}\left[x_0 \exp\left\{\int_0^t \alpha_s \,\mathrm{d}W_s - \frac{1}{2}\int_0^t \alpha_s^2 \,\mathrm{d}s\right\}\right] \\
&= \mathbb{E}\left[X_0 + \int_0^t \alpha_s X_s \,\mathrm{d}W_s\right] \\
&= \mathbb{E}\left[X_0\right] = x_0.
\end{aligned}$$

Example 5.4 (Integrals of deterministic processes)**.** Let $X_t = x_0 + \int_0^t \alpha(s)\,\mathrm{d}s + \int_0^t \varphi(s)\,\mathrm{d}W_s$, where $\alpha(s)$ and $\varphi(s)$ are *deterministic* functions. Then we will show

$$X_t \sim N(x_0 + \int_0^t \alpha(s)\,\mathrm{d}s, \int_0^t \varphi(s)^2 \,\mathrm{d}s).$$

Proof. We use the following fact about the moment generating function of a random variable Z:

$$Z \sim N(\mu, \sigma^2) \iff \mathbb{E}\left[\mathrm{e}^{\eta Z}\right] = \mathrm{e}^{\eta\mu + \frac{1}{2}\eta^2\sigma^2} \quad \forall \eta \in \mathbb{R}.$$

Consider $Y_t = \exp\{\eta X_t\}$. Then

$$\begin{aligned}
\mathrm{d}Y_t &= \eta Y_t \,\mathrm{d}X_t + \frac{1}{2}\eta^2 Y_t \,(\mathrm{d}X_t)^2 \\
&= \eta Y_t \alpha(t)\,\mathrm{d}t + \eta Y_t \varphi(t)\,\mathrm{d}W_t + \frac{1}{2}\eta^2 Y_t \varphi(t)^2 \,\mathrm{d}t
\end{aligned}$$

that is:

$$Y_t = Y_0 + \int_0^t \left(\eta\alpha(s) + \frac{1}{2}\eta^2\varphi(s)^2\right) Y_s \,\mathrm{d}s + \frac{1}{2}\int_0^t \eta Y_s \varphi(s)\,\mathrm{d}W_s.$$

Assuming that $\mathbb{E}\left[\int_0^t \eta^2\varphi(s)^2 Y_s^2 \,\mathrm{d}s\right] < \infty$ (it is...) then the stochastic integral term has zero expectation. So

$$\mathbb{E}\left[Y_t\right] = \exp\{x_0\eta\} + \int_0^t \mathbb{E}\left[Y_s\right]\left(\eta\alpha(s) + \frac{1}{2}\eta^2\varphi(s)^2\right)\,\mathrm{d}s$$

or

$$\frac{d}{dt}\left\{\mathbb{E}\left[e^{\eta X_t}\right]\right\} = \mathbb{E}\left[e^{\eta X_t}\right]\left(\eta\alpha(t) + \frac{1}{2}\eta^2\varphi(t)^2\right)$$

which has solution

$$\mathbb{E}\left[e^{\eta X_t}\right] = \exp\left\{\left(x_0 + \int_0^t \alpha(s)\,\mathrm{d}s\right)\eta + \frac{1}{2}\eta^2\int_0^t \varphi(s)^2\,\mathrm{d}s\right\}$$

□

5.2 Existence and Uniqueness of Solutions to SDEs

The class of solutions to SDEs have important and useful properties, as demonstrated by the examples above. To work with these objects, it will be useful to first establish existence and uniqueness of SDEs of the form (21).

First note that, even if we consider the simplification $\sigma \equiv 0$, where we get a 'classical' ODE, things can still go wrong with both existence and uniqueness:

i) Consider the ODE: $\frac{\mathrm{d}X_t}{\mathrm{d}t} = X_t^2$, with $X_0 = 1$. Then we get the 'solution' $X_t = \frac{1}{1-t}$ for $0 \le t < 1$, which *explodes* at time 1. Hence there is no global solution on $[0,\infty)$.

ii) Consider the ODE: $\frac{\mathrm{d}X_t}{\mathrm{d}t} = 3X_t^{2/3}$, $X_t = 0$. Then the functions

$$X_t = \begin{cases} 0 & t \le a \\ (t-a)^3 & t > a \end{cases}$$

solve the ODE for any $a \ge 0$: hence there is no unique solution.

In the classical theory, it is usual to put a Lischitz condition on the function in the ODE to get well-behaved solutions, and this is essentially the approach we need to take here:

Theorem 5.5. *Let $T > 0$ and suppose $\mu, \sigma : [0,T] \times \mathbb{R} \to \mathbb{R}$ are functions such that*

i) there exists a constant $C \ge 0$ for which

$$|\mu(t,x)| + |\sigma(t,x)| \le C(1+|x|), \quad \forall x \in \mathbb{R}, t \in [0,T]; \tag{22}$$

ii) and there exists a constant $D \ge 0$ for which

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \le D|x-y|, \quad \forall x,y \in \mathbb{R}, t \in [0,T]. \tag{23}$$

In addition, let Z be a r.v. such that $\mathbb{E}\left[Z^2\right] < \infty$, and such that Z is independent of $(W_t)_{t\ge 0}$.

Then there exists a unique t-continuous solution to (21) with $X_0 = Z$ and such that X_t depends only on $Z, (W_s)_{s\le t}$ and

$$\mathbb{E}\left[\int_0^T X_s^2\,\mathrm{d}s\right] < \infty.$$

The version of the theorem given above is stated for the case of a one-dimensional process, but the result easily extends to higher dimensional processes.

Proof. We first show uniqueness. For slight additional generality, we consider initially the case where we have two solutions X_t, X_t' to (21) with possibly different starting values, $X_0 = Z, X_0' = Z'$. Then, writing $\mu_s = \mu(s, X_s)$ and $\mu_s' = \mu(s, X_s')$, etc.:

$$\begin{aligned}
\mathbb{E}\left[(X_t - X_t')^2\right] &= \mathbb{E}\left[\left(Z - Z' + \int_0^t (\mu_s - \mu_s')\,\mathrm{d}s + \int_0^t (\sigma_s - \sigma_s')\,\mathrm{d}W_s\right)^2\right] \\
&\leq 3\mathbb{E}\left[(Z-Z')^2\right] + 3\mathbb{E}\left[\left(\int_0^t (\mu_s - \mu_s')\,\mathrm{d}s\right)^2\right] + 3\mathbb{E}\left[\left(\int_0^t (\sigma_s - \sigma_s')\,\mathrm{d}W_s\right)^2\right] \\
&\leq 3\mathbb{E}\left[(Z-Z')^2\right] + 3t\mathbb{E}\left[\int_0^t (\mu_s - \mu_s')^2\,\mathrm{d}s\right] + 3\mathbb{E}\left[\int_0^t (\sigma_s - \sigma_s')^2\,\mathrm{d}s\right] \\
&\leq 3\mathbb{E}\left[(Z-Z')^2\right] + 3(1+t)D^2 \int_0^t \mathbb{E}\left[(X_s - X_s')^2\right]\,\mathrm{d}s.
\end{aligned}$$

Writing $m(t) = \mathbb{E}\left[(X_t - X_t')^2\right]$ we have

$$m(t) \leq \underbrace{3\mathbb{E}\left[(Z-Z')^2\right]}_{C_1} + \underbrace{3(1+T)D^2}_{C_2} \int_0^t m(s)\,\mathrm{d}s$$

for some constants C_1, C_2. By Gronwall's Inequality[9] we get: $m(t) \leq C_1 \mathrm{e}^{C_2 t}$, and in the case where $C_1 = 0$, we in addition have $m(t) = 0$ for all $t \in [0, T]$.

It follows that $X_t = X_t'$ almost surely at every rational, and by the continuity of paths, we therefore have almost sure equality of the paths (i.e. X, X' are indistinguishable).

We now consider the existence of solutions. Our proof will follow a similar structure to the classical proof of existence of solutions to ODEs.

We iteratively define a sequence of stochastic processes by: $Y_t^{(0)} = X_0$, and

$$Y_t^{(k+1)} = X_0 + \int_0^t \mu(s, Y_s^{(k)})\,\mathrm{d}s + \int_0^t \sigma(s, Y_s^{(k)})\,\mathrm{d}W_s.$$

By the same argument as above, we get:

$$\mathbb{E}\left[(Y_t^{(k+1)} - Y_t^{(k)})^2\right] \leq 3(1+T)D^2 \int_0^t \mathbb{E}\left[(Y_t^{(k)} - Y_t^{(k-1)})^2\right]\,\mathrm{d}s$$

for $k \geq 1$, and

$$\mathbb{E}\left[(Y_t^{(1)} - Y_t^{(0)})^2\right] \leq 2C^2t^2(1 + \mathbb{E}\left[X_0^2\right]) + 2C^2t(1 + \mathbb{E}\left[X_0^2\right]) \leq A_1 t$$

where A_1 is a constant depending only on $T, C, \mathbb{E}\left[X_0^2\right]$.

By induction on k, we get

$$\mathbb{E}\left[(Y_t^{(k+1)} - Y_t^{(k)})^2\right] \leq \frac{A_2^{k+1}t^{k+1}}{(k+1)!}$$

[9]Gronwall's Inequality: If $m(t) \leq u(t) + \int_0^t m(s)v(s)\,\mathrm{d}s$ for continuous functions m, v, and non-decreasing function u, and all $t \geq 0$, then $m(t) \leq u(t)\mathrm{e}^{\int_0^t v(s)\,\mathrm{d}s}$ for all $t \geq 0$.

for an additional constant A_2. But then:

$$\begin{aligned}\mathbb{E}\left[\int_0^T (Y_s^{(m)} - Y_s^{(n)})^2 \,\mathrm{d}s\right]^{1/2} &= \mathbb{E}\left[\int_0^T \sum_{k=n}^{m-1} (Y_s^{(k+1)} - Y_s^{(k)})^2 \,\mathrm{d}s\right]^{1/2} \\ &\leq \sum_{k=n}^{m-1} \mathbb{E}\left[\int_0^T (Y_s^{(k+1)} - Y_s^{(k)})^2 \,\mathrm{d}s\right]^{1/2} \\ &\leq \sum_{k=n}^{m-1} \left(\int_0^T \frac{A_2^{k+1} s^{k+1}}{(k+1)!} \,\mathrm{d}s\right)^{1/2} \\ &\leq \sum_{k=n}^{m-1} \left(\frac{A_2^{k+1} T^{k+2}}{(k+2)!}\right)^{1/2} \\ &\to 0 \qquad \text{as } m, n \to \infty.\end{aligned}$$

It follows that $(Y_t^n)_{t\in[0,T]}$ is a Cauchy sequence, and therefore that the limit exists. Moreover, by Hölder's inequality and the Itô isometry, we get convergence of e.g. $\int_0^t b(s, Y_s^n) \,\mathrm{d}s \to \int_0^t b(s, Y_s) \,\mathrm{d}s$, and so the desired process exists, and has the required properties. □

5.3 Weak and Strong Solutions to SDEs

Observe that, in the section above, we have the following interpretation of a solution to an SDE: essentially, we assume that we are *given* a probability space on which there is a Brownian motion, and based on this, we look to construct a solution which depends on this Brownian motion, and which solves equation (21).

It turns out that this is a rather strong notion of solution. A weaker notion of solution might allow us to *construct* a probability space on which there exists both a Brownian motion W_t, and a process X_t satisfying (21). We call such a notion of solution a *weak* solution. As we shall see below, there are examples of functions σ and μ such that it is not possible to construct a strong solution, but it is possible to construct a weak solution. Observe that, from Theorem 5.5, if there is no strong solution, then necessarily conditions (22) and (23) must not be satisfied.

From a modelling perspective, the latter form of solution may be more natural. Often we do not think of the noise as a given external signal, which can be measured directly, but something which may only be observable through its influence on the observed process X_t. In this case, there is no reason to assume that W_t is observed independently of the process X_t, and so it may be no loss of generality to assume that we only have a weak solution to (21).

Example 5.6 (The Tanaka SDE)**.** The idea here is to consider the SDE:

$$\mathrm{d}X_t = \mathrm{sign}(X_t) \,\mathrm{d}W_t, \tag{24}$$

where $\mathrm{sign}(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$.

If X_t is a solution to this SDE, then we should be able to recover the process W_t, since:

$$\mathrm{d}W_t = \mathrm{sign}(X_t)^2 \mathrm{d}W_t = \mathrm{sign}(X_t)\mathrm{d}X_t.$$

Moreover, if X_t solves (24), then X_t is also a Brownian motion[10].

[10] This is Levy's Theorem: if X_t is a continuous martingale with $\langle X\rangle_t = t$, then X_t is a standard Brownian motion. It is easy to check that these conditions hold for X_t given by (24).

Now consider a Brownian motion, or more specifically, the approximation we get for N large through Donsker's Theorem and i.i.d. noise,

$$\tilde{X}_n = \tilde{X}_{n-1} + \begin{cases} N^{-1/2} & \text{with probability } \frac{1}{2} \\ -N^{-1/2} & \text{with probability } \frac{1}{2} \end{cases}$$

so $X_t = \tilde{X}_{\lfloor tN \rfloor} + (tN - \lfloor tN \rfloor)(\tilde{X}_{\lfloor tN \rfloor+1} - \tilde{X}_{\lfloor tN \rfloor})$. Then we have

$$\int_0^t \operatorname{sign}(X_s)\,\mathrm{d}X_s \approx \sum_{i=0}^{\lfloor Nt \rfloor - 1} \operatorname{sign}(\tilde{X}_t)(\tilde{X}_{i+1} - \tilde{X}).$$

Now observe that this sum can be rewritten as:

$$\sum_{i=0}^{\lfloor Nt \rfloor - 1} \operatorname{sign}(\tilde{X}_i)(\tilde{X}_{i+1} - \tilde{X}_i) = |\tilde{X}_{\lfloor Nt \rfloor}| - |\tilde{X}_0| - 2N^{-1/2} \sum_{i=0}^{\lfloor Nt \rfloor - 1} \mathbf{1}\{\tilde{X}_i = 0, \tilde{X}_{i+1} = -N^{-1/2}\}.$$

However, as $N \to \infty$, by the strong law of large numbers, we have

$$2 \sum_{i=0}^{\lfloor Nt \rfloor - 1} \mathbf{1}\{\tilde{X}_i = 0, \tilde{X}_{i+1} = -N^{-1/2}\} \to \sum_{i=0}^{\lfloor Nt \rfloor - 1} \mathbf{1}\{\tilde{X}_i = 0\} = \sum_{i=0}^{\lfloor Nt \rfloor - 1} \mathbf{1}\{|\tilde{X}_i| = 0\}.$$

In particular,

$$\sum_{i=0}^{\lfloor Nt \rfloor - 1} \operatorname{sign}(\tilde{X}_i)(\tilde{X}_{i+1} - \tilde{X}_i) \to |\tilde{X}_{\lfloor Nt \rfloor}| - |\tilde{X}_0| - N^{-1/2} \sum_{i=0}^{\lfloor Nt \rfloor - 1} \mathbf{1}\{|\tilde{X}_i| = 0\}.$$

So $\int_0^t \operatorname{sign}(X_s)\,\mathrm{d}X_s$ should only depend on $|X_s|$, not X_s. Writing[11]

$$L_t := \lim_{N \to \infty} N^{-1/2} \sum_{i=0}^{\lfloor Nt \rfloor - 1} \mathbf{1}\{|\tilde{X}_i| = 0\}$$

we have

$$\int_0^t \operatorname{sign}(X_s)\,\mathrm{d}X_s = |X_t| - |X_0| - L_t,$$

where the right-hand side only depends on $(|X_s|)_{s \le t}$, and not $(X_s)_{s \le t}$.

Therefore, if there were a strong solution to (24), we would be able to follow the sequence of constructive steps:

$$(W_s)_{s \le t} \xmapsto[\int \operatorname{sign}(X_t)\mathrm{d}W_t]{} (X_s)_{s \le t} \xmapsto[\int \operatorname{sign}(X_t)\mathrm{d}X_t]{} (W_s)_{s \le t}$$

i.e. this is an invertible set of maps.

However, by the argument above, again assuming that there is a strong solution, we also can map constructively by:

$$(W_s)_{s \le t} \xmapsto[\int \operatorname{sign}(X_t)\mathrm{d}W_t]{} (X_s)_{s \le t} \xmapsto[|\cdot|]{} (|X_s|)_{s \le t} \xmapsto[|X_t| - |X_0| - L_t]{} (W_s)_{s \le t}$$

Again, we have an invertible sequence of mappings, however, by applying these in the correct order, this implies that we can construct $(X_s)_{s \le t}$ from its absolute value $(|X_s|)_{s \le t}$. However,

[11] This is a process called the **Local time**. It is essentially a measure of the time spent by the process at 0.

there is no way that the sign of X_t can be inferred simply from its absolute value, and hence this is not possible. In particular, the assumption that there exists a strong solution to (24) must be false.

Finally, it is easy to see that we *can* construct a weak solution. Simply take a probability space on which we have a Brownian motion, X_t, and define W_t by $X_t = |X_t| - |X_0| - L_t$. Then by the arguments above, X_t solves the SDE (24) in the weak sense, but not in the strong sense.

6 Diffusion Processes

In this section, we analyse further the properties of solutions to SDEs of the form:

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \qquad X_0 = x_0. \tag{25}$$

We call solutions to these equations **diffusions**, and we call μ the **drift coefficient** and σ (or sometimes, confusingly, $\frac{1}{2}\sigma^2$) the **diffusion coefficient**.

More specifically, we often drop the time-dependence in the coefficients, and then:

Definition 6.1. A (time-homogenous) (Itô) **diffusion** is a process X_t such that

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad t \geq s, X_s = x \tag{26}$$

where $|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq D|x - y|$, for some constant D.

We denote the unique (strong) solution to (25) by $(X_t^{s,x})_{t \geq s}$, and $X^{0,x} = X^x$. Then in particular, we expect the process $(X_t^{s,x})$ to be time-homogenous, that is, $(X_{s+t}^{s,x})_{s \geq 0}$ and $(X_t^{0,x})_{t \geq 0}$ should have the same laws. We also write $\mathbb{E}^x[\cdot]$ for the expectation of $X^{0,x}$, i.e. given $X_0 = x$.

Theorem 6.2 (Markov Property)**.** *We have*

$$\mathbb{E}^x\left[f(X_{t+h})|\mathcal{F}_t\right] = \mathbb{E}^{X_t}\left[f(X_h)\right] \tag{27}$$

We will not prove this result, but rather motivate this result: we have

$$\begin{aligned} X_t &= X_0 + \int_0^t \sigma(X_u)\,\mathrm{d}W_u + \int_0^t \mu(X_u)\,\mathrm{d}u \\ &= \underbrace{\left(X_0 + \int_0^s \sigma(X_u)\,\mathrm{d}W_u + \int_0^s \mu(X_u)\,\mathrm{d}u\right)}_{=X_s} + \int_s^t \sigma(X_u)\,\mathrm{d}W_u + \int_s^t \mu(X_u)\,\mathrm{d}u, \end{aligned}$$

so conditioning on $X_s = x$ is essentially the same as starting at x at time s.

A stronger version of the Markov Property involves starting at *random times*.

Definition 6.3. A random time $\tau : \Omega \to [0, \infty]$ is a **stopping time** if $\{\tau \leq t\} \in \mathcal{F}_t$, i.e. we know at time t if it has already happened.

For example:

$$\begin{aligned} \tau_1(\omega) &= \inf\{t \geq 0 \big| |W_t| \geq x\} \\ \tau_2(\omega) &= \inf\{t \geq 0 \big| 2W_t \geq x + \sup_{s \leq t} W_s\} \end{aligned}$$

are both stopping times, while

$$\tau_3(\omega) = \sup\{t \geq 0 \big| W_t \geq 2t\}$$

is not a stopping time — we need to 'see' the future in order to know whether or not it has happened at the current time.

Since stopping times are not able to see the future, we do not expect the behaviour up to a stopping time τ to influence the future. In particular, we have:

Theorem 6.4 (Strong Markov Property)**.** *It X_t is a (time-homogenous, Itô) diffusion, and τ is a stopping time such that $\tau < \infty$ almost surely, then*

$$\mathbb{E}^x \left[f(X_{\tau+h}) | \mathcal{F}_t \right] = \mathbb{E}^{X_\tau} \left[f(X_h) \right]$$

where $\mathcal{F}_\tau$ denotes the information available at the random time τ

MT

Strictly speaking, $A \in \mathcal{F}_\tau$ if and only if $A \in \mathcal{F}$ and $A \cap \{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

The Strong Markov property can be described in words as:

'The behaviour of the diffusion after a stopping time is independent of the past, given the current position.'

6.1 Generators of diffusions and the Dynkin formula

There is an important connection between second order partial differential operators and diffusions, which we now investigate.

Recall Questions Q1.2 and Q1.3 from Example Sheet 1. Here, we examined $h^{-1} \left[\mathbb{E}^x \left[f(X_h) \right] - f(x) \right]$ for Brownian motion and the Brownian Bridge, and we were able to show that for example

$$\lim_{h \searrow 0} \frac{\mathbb{E}^x \left[f(W_h) \right] - f(x)}{h} = \frac{1}{2} f''(x),$$

for 'nice' functions f. In the case of the Brownian Bridge, we obtained a (time-inhomogeneous) limit: $\frac{y-x}{1-t} f'(x) + \frac{1}{2} f''(x)$.

Definition 6.5. Let X_t be a (time-homogeneous, Itô) diffusion on $\mathbb{R}^n$. The **(infinitesimal) generator** $\mathcal{A}$ of X_t is defined by

$$\mathcal{A}f(x) = \lim_{h \searrow 0} \frac{\mathbb{E}^x \left[f(X_t) \right] - f(x)}{t}, x \in \mathbb{R}^n. \tag{28}$$

The set of functions $f : \mathbb{R}^n \to \mathbb{R}$ for which the limit exists at x is denoted $D_{\mathcal{A}}(x)$, and $D_{\mathcal{A}}$ is the set of functions for which the limit exists at all $x \in \mathbb{R}^n$.

Remark 6.6. It can be show that if $f \in D_{\mathcal{A}}$, and U_n is a sequence of open sets decreasing to $\{x\}$ (that is, $U_n \supset U_{n+1}$ and $\cap_{n=1}^\infty U_n = \{x\}$), then

$$\mathcal{A}f(x) = \lim_{U_n \searrow \{x\}} \frac{\mathbb{E}^x \left[f(X_{\tau_{U_n}}) - f(x) \right]}{\mathbb{E}^x \left[\tau_{U_n} \right]},$$

where $\tau_U := \inf\{t \geq 0 : X_t \notin U\}$.

Theorem 6.7. *Let X_t be the solution to* (26)*. Then*

$$\mathrm{d}X_t = \mu(X_t) \, \mathrm{d}t + \sigma(X_t) \, \mathrm{d}W_t.$$

If $f \in C_0^2(\mathbb{R})$ (the set of compactly supported, twice differentiable functions), then $f \in D_{\mathcal{A}}$ and

$$\mathcal{A}f(x) = \mu(x) \frac{\partial f}{\partial x} + \frac{1}{2} \sigma(x)^2 \frac{\partial^2 f}{\partial x^2}.$$

More generally, if X_t is a diffusion[12] *in $\mathbb{R}^n$, and $f \in C_0^2(\mathbb{R}^n)$ then $f \in D_{\mathcal{A}}$ and*

$$\mathcal{A}f(\boldsymbol{x}) = \sum_{i=1}^{n} \mu^i(\boldsymbol{x}) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{n} (\sigma(\boldsymbol{x})\sigma^T(\boldsymbol{x}))^{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Proof for $n = d = 1$. By Itô's Lemma,

$$f(X_t) = f(X_0) + \int_0^t \left(\frac{\partial f}{\partial x}(X_s)\mu(X_s) + \frac{1}{2}\sigma(X_s)^2 \frac{\partial^2 f}{\partial x^2} \right) \mathrm{d}s + \int_0^t \frac{\partial f}{\partial x}\sigma(X_s)\, \mathrm{d}W_s.$$

Taking expectations, we get

$$\begin{aligned}\mathbb{E}\left[f(X_t) - f(X_0)\right] = \mathbb{E}&\left[\int_0^t \left(\frac{\partial f}{\partial x}(X_s)\mu(X_s) + \frac{1}{2}\sigma(X_s)^2 \frac{\partial^2 f}{\partial x^2} \right) \mathrm{d}s\right] \\ &+ \mathbb{E}\left[\int_0^t \frac{\partial f}{\partial x}\sigma(X_s)\, \mathrm{d}W_s\right]\end{aligned}$$

Now, since f is compactly supported, and $\sigma(x)$ is continuous (by the assumption that $|\sigma(x) - \sigma(y)| \leq D|x-y|$ in Definition 6.1), we have $\frac{\partial f}{\partial x}\sigma(x)$ is bounded on $\mathbb{R}$, and so $\mathbf{1}_{[0,t]}\frac{\partial f}{\partial x}(x)\sigma(x) \in \mathcal{V}$. Hence (Corollary 3.8) $\mathbb{E}\left[\int_0^t \frac{\partial f}{\partial x}\sigma(X_s)\,\mathrm{d}W_s\right] = 0$.

Taking limits as $t \searrow 0$, we get

$$\begin{aligned}\mathcal{A}f(x) &= \lim_{t\searrow 0} \mathbb{E}\left[\frac{1}{t}\int_0^t \left(\frac{\partial f}{\partial x}(X_s)\mu(X_s) + \frac{1}{2}\sigma(X_s)^2 \frac{\partial^2 f}{\partial x^2} \right) \mathrm{d}s\right] \\ &= \mu(x)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma(x)^2\frac{\partial^2 f}{\partial x^2}.\end{aligned}$$

□

A closely related result, which is very useful is:

Theorem 6.8 (Dynkin's Lemma)**.** *Let $f \in C_0^2$, and suppose that τ is a stopping time with $\mathbb{E}^x[\tau] < \infty$. Then:*

$$\mathbb{E}^x\left[f(X_\tau)\right] = f(x) + \mathbb{E}^x\left[\int_0^\tau \mathcal{A}f(X_s)\,\mathrm{d}s\right].$$

Proof. As $f \in C_0^2$ and σ, μ are continuous, $|\mathcal{A}f|$ is bounded, by M say. Then:

$$\int_0^\tau |\mathcal{A}f(X_s)|\,\mathrm{d}s \leq M\tau,$$

and by bounded convergence, we have

$$\mathbb{E}\left[\int_0^{\tau\wedge t} \mathcal{A}f(X_s)\,\mathrm{d}s\right] \to \mathbb{E}\left[\int_0^\tau \mathcal{A}f(X_s)\,\mathrm{d}s\right]$$

as $t \to \infty$.

[12] So $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times d}$, W a d-dimensional Brownian motion, and $\mu : \mathbb{R}^n \to \mathbb{R}^n$, and we denote individual elements of σ and μ by e.g. $\sigma^{i,j}$ or μ^j. Recall the discussion at the start of Section 4.2.

But applying the same argument as above to the Itô process $\mathrm{d}\hat{X}_t = \alpha_t \,\mathrm{d}t + \varphi_t \,\mathrm{d}W_t$, where $\alpha_t = \mu(\hat{X}_t)\mathbf{1}\{t \le \tau\}$ and $\varphi_t = \sigma(\hat{X}_t)\mathbf{1}\{t \le \tau\}$, so we have $\hat{X}_t = X_t$ on $\{t \le \tau\}$ and $\hat{X}_t = X_\tau$ on $\{t \ge \tau\}$.

Then applying the same argument as in the previous proof, we get:

$$\mathbb{E}^x[f(X_\tau)] \leftarrow \mathbb{E}^x[f(X_{t\wedge\tau})] = f(x) + \mathbb{E}^x\left[\int_0^{t\wedge\tau} \mathcal{A}f(X_s)\,\mathrm{d}s\right] \to f(x) + \mathbb{E}^x\left[\int_0^{\tau} \mathcal{A}f(X_s)\,\mathrm{d}s\right].$$

□

Example 6.9. The generator of d-dimensional Brownian motion is:

$$\mathcal{A}f(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{d} \frac{\partial^2 f}{\partial {x_i}^2}(\boldsymbol{x}) =: \frac{1}{2}\Delta f(\boldsymbol{x}), \qquad \boldsymbol{x} = (x^1, \dots, x^d) \in \mathbb{R}^d.$$

Here Δ is commonly known as the **Laplacian**.

Let D be an open subset of $\mathbb{R}^d$, and write $\tau_D := \inf\{t \ge 0 : X_t \notin D\}$. Suppose X_t is a diffusion on $\mathbb{R}^d$. Then we set

$$u(x) := \mathbb{E}^x[g(X_{\tau_D})]$$

for some (nice) function g, and observe that, for any open set $U \subset D$, we have:

$$u(x) = \mathbb{E}^x\left[\mathbb{E}^{X_{\tau_U}}[g(X_{\tau_D})]\right] = \mathbb{E}^x[u(X_{\tau_U})]$$

by the Strong Markov property (Theorem 6.4). Hence, by Remark 6.6 we have

$$\mathcal{A}u(x) = \lim_{U \searrow \{x\}} \frac{\mathbb{E}^x[u(X_{\tau_U})] - u(x)}{\mathbb{E}^x[\tau_U]} = 0.$$

In particular, if $u \in D_{\mathcal{A}}$, then we must have $\mathcal{A}u(x) = 0$.

6.2 Kolmogorov's Backward Equation, Boundary Value Problems

The close connection between the generator and second order partial differential operators allows us to provide some probabilistic insight into classical second order problems. We give some basic examples here, but note that there are many deep connections, and much more can be said than we do here.

Theorem 6.10 (Kolmogorov's Backward Equation)**.** *Let X_t be an Itô diffusion in $\mathbb{R}^n$ with generator $\mathcal{A}$. Suppose $f \in C_0^2(\mathbb{R}^n)$, and set $u(t,x) := \mathbb{E}^x[f(X_t)]$. Then:*

i) $u(t,\cdot) \in D_{\mathcal{A}}$, and

$$\begin{cases} \frac{\partial u}{\partial t} &= \mathcal{A}u \qquad t > 0, x \in \mathbb{R}^n \\ u(0,x) &= f(x), \qquad x \in \mathbb{R}^n \end{cases} \tag{29}$$

ii) Moreover, if $w \in C^{1,2}$ solves (29), then $w = u$.

Sketch Proof. We have

$$\begin{aligned}
\mathcal{A}u(t,x) &= \frac{\mathbb{E}^x\left[u(t,X_r)\right] - u(t,x)}{r} \\
&= \frac{1}{r}\mathbb{E}^x\left[\mathbb{E}^{X_r}\left[f(X_t)\right] - \mathbb{E}^x\left[f(X_t)\right]\right] \\
&= \frac{1}{r}\mathbb{E}^x\left[\mathbb{E}^x\left[f(X_{t+r})|\mathcal{F}_r\right] - \mathbb{E}^x\left[f(X_t)|\mathcal{F}_t\right]\right] \\
&= \frac{1}{r}\mathbb{E}^x\left[f(X_{t+r}) - f(X_t)\right] \\
&= \frac{u(t+r,x)-u(t,x)}{r} \to \frac{\partial u}{\partial t}(t,x)
\end{aligned}$$

Conversely, given w solving (29), and applying Itô's Lemma (as in the proof of Dynkin's Lemma), we get:

$$\mathbb{E}^x\left[w(s-t,X_t)\right] = w(s,X_t) + \mathbb{E}^x\left[\int_0^t \underbrace{\left(-\frac{\partial w}{\partial t} + \mathcal{A}w\right)}_{=0}(s-r,X_r)\,\mathrm{d}r\right]$$

taking $s = t$, we see that $w(t,x) = \mathbb{E}^x\left[w(0,X_t)\right] = \mathbb{E}^x\left[f(X_t)\right] = u(t,x)$. □

Theorem 6.11 (Dirichlet-Poisson Problem)**.** *Let D be a bounded, open set, such that $\mathbb{E}^x\left[\tau_D\right] < \infty$, for all $x \in D$, and suppose that $g \in C(\partial D), h \in C(D)$ are bounded, and $w \in C^2(D)\cap C(\bar{D})$[13], where*

$$w(x) = \mathbb{E}^x\left[g(X_{\tau_D}) + \int_0^{\tau_D} h(X_s)\,\mathrm{d}s\right].$$

Then w is the unique solution to the Dirichlet-Poisson problem:

$$\begin{cases} \mathcal{A}w(x) = -h(x) & x \in D \\ w(x) = g(x) & x \in \partial D \end{cases}. \tag{30}$$

Sketch Proof. To see that w solves (30), we apply Dynkin's formula to w (or strictly speaking, to compact subsets of D, so that we can ensure sufficient differentiability on the whole of $\mathbb{R}^d$). Then:

$$\begin{aligned}
\mathbb{E}^x\left[w(X_{\tau_D})\right] &= w(x) + \mathbb{E}^x\left[\int_0^{\tau_D}\mathcal{A}w(X_s)\,\mathrm{d}s\right] \\
\implies \qquad \mathbb{E}^x\left[g(X_{\tau_D})\right] &= w(x) - \mathbb{E}^x\left[\int_0^{\tau_D} h(X_s)\,\mathrm{d}s\right]
\end{aligned}$$

For uniqueness, suppose $\tilde{w}$ also solves (30). Then we can also apply Dynkin to $w - \tilde{w}$ to get:

$$\mathbb{E}^x\left[(w-\tilde{w})(X_{\tau_D})\right] = (w-\tilde{w})(x) - \mathbb{E}^x\left[\int_0^{\tau_D}\mathcal{A}(w-\tilde{w})\,\mathrm{d}s\right]$$

which implies $(w - \tilde{w})(x) = 0$. □

The arguments here are really only the start of many related results connecting second order PDEs to diffusions. In particular, diffusions prove to be very powerful tools for proving existence results of solutions to PDEs. See e.g. [3] for more on this topic.

[13] That is, w is twice continuously differentiable in D, and has a continuous extension to the boundary of D.

We end with one last example. In this case, we use Theorem 6.11 to compute a numerical solution to a PDE. Specifically, we simulate a diffusion up to the exit time of a domain, and calculate the expectation given in the theorem. By simulating this diffusion many times, we are able to compute an estimate of the true value of the expectation. These methods are called *Monte Carlo* methods, and prove to be very important in certain situations.

iPython Interlude: Go To Notebook

7 Probability Background

7.1 Probability spaces, probability measures and events

Let's recall the key mathematical ingredients that we need to discuss probability.

A **sample space**, Ω, which is a set of points $\omega \in \Omega$, which we think of as the possible outcomes of an experiment. When we run our experiment, one of these ω gets picked at random according to some probability law $\mathbb{P}$.

The second ingredient is an **event**. An event, F is a subset of Ω, $F \subset \Omega$. The event F occurs if and only if (iff) the chosen ω is an element of F.

Example 7.1. Infinite coin flips: G_p is event that proportion of heads converges to p:

$$G_p = \left\{ \omega = (\omega_1, \omega_2, \omega_3, \ldots) \in \Omega_{C_\infty} \middle| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_H(\omega_k) \to p \text{ as } n \to \infty \right\}.$$

Where we have introduced the **indicator function**: for an event $Y \subset \Omega$, defined by:

$$\mathbf{1}_Y(\omega) = \begin{cases} 1 & \text{if } \omega \in Y \\ 0 & \text{if } \omega \notin Y \end{cases}.$$

Collection of events $\mathcal{F}$: this all the events of interest. If Ω is a finite or countable set, usually $\mathcal{F} = \mathcal{P}(\Omega)$, the set of all subsets of Ω. If $|\Omega| = N$, $|\mathcal{P}(\Omega)| = 2^N$.

Remember that, if Ω is countable, we can take $\mathcal{F}$ to be the power set of Ω, however this leads to problems if Ω is uncountable. Fortunately, even for uncountable Ω we can still choose $\mathcal{F}$ to be big enough that it contains all events of interest, and so that $\mathcal{F}$ is a **σ-algebra**: $\mathcal{F} \subset \mathcal{P}(\Omega)$ is a σ-algebra if:

i) $\emptyset \in \mathcal{F}$;
ii) $F \in \mathcal{F} \implies F^{\complement} \in \mathcal{F}$;
iii) $F_n \in \mathcal{F}$, for $n \in \mathbb{N}$, then $\cup_{n\in\mathbb{N}} F_n \in \mathcal{F}$

NB: $\mathcal{F}$ closed under complements and countable unions. Since $\emptyset \in \mathcal{F}$ and $\Omega = \emptyset^{\complement}$, then $\Omega \in \mathcal{F}$. Also $\cap_{n\in\mathbb{N}} F_n = \left(\cup_{n\in\mathbb{N}} F_n^{\complement}\right)^{\complement}$, so countable intersection of $F_n \in \mathcal{F}$ also in $\mathcal{F}$.

If $A \subset \mathcal{P}(\Omega)$ is a collection of subsets of Ω, then we define $\sigma(A)$**, the σ-algebra generated by** A, to be the smallest σ-algebra on Ω that contains A. Note that since the intersection of two σ-algebras is a (smaller) σ-algebra, this can be alternatively defined as the intersection of all σ-algebras containing A. It is also easy to check that $\mathcal{P}(\Omega)$ is a σ-algebra, so there is at least one σ-algebra containing A.

If $E \subset \Omega$ is a single event, then $\sigma(E) = \{\emptyset, E, E^{\complement}, \Omega\}$. (Check that this is a σ-algebra, and that every σ-algebra containing E must contain all these elements.)

Now we can discuss probability! Given sample space Ω and σ-algebra $\mathcal{F}$ on Ω, the map $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a **probability measure** on $(\Omega, \mathcal{F})$ if:

i) $\mathbb{P}(\emptyset) = 0$;

ii) $\mathbb{P}(\Omega) = 1$;

iii) (Countable additivity) If $(F_n : n \in \mathbb{N})$ is a sequence of **disjoint** (i.e. $F_i \cap F_j = \emptyset$ for all $i \neq j$) events in $\mathcal{F}$, then $\mathbb{P}(\cup_{n\in\mathbb{N}} F_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(F_n)$.

Then we call $(\Omega, \mathcal{F}, \mathbb{P})$ a **probability triple**, or a **probability space**, and we interpret $\mathbb{P}(F)$ as the probability of the event F happening — that is, the probability that the selected ω is in F.

Lemma 7.2 (Properties of $\mathbb{P}$). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple. Then:*

i) If $A \in \mathcal{F}$ then $\mathbb{P}(A^{\complement}) = 1 - \mathbb{P}(A)$;

ii) If $A, B \in \mathcal{F}$ and $A \subset B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$;

iii) (Finite Additivity) If $F_1, F_2, \ldots, F_n \in \mathcal{F}$ are disjoint then $\mathbb{P}(\cup_{j=1}^n F_j) = \sum_{j=1}^n \mathbb{P}(F_j)$;

iv) (Boole's Inequality) For any $F_1, F_2, \ldots, F_n \in \mathcal{F}$, $\mathbb{P}(\cup_{j=1}^n F_j) \leq \sum_{j=1}^n \mathbb{P}(F_j)$;

v) (Continuity of $\mathbb{P}$)

   a) Suppose $F_1 \subset F_2 \subset F_3 \subset \cdots$ and $F = \cup_{n\in\mathbb{N}} F_n$, then $\mathbb{P}(F_n) \uparrow \mathbb{P}(F)$ (i.e. $\mathbb{P}(F_n)$ is an increasing sequence, with limit $\mathbb{P}(F)$).

   b) Suppose $F_1 \supset F_2 \supset F_3 \supset \cdots$ and $F = \cap_{n\in\mathbb{N}} F_n$, then $\mathbb{P}(F_n) \downarrow \mathbb{P}(F)$.

When Ω is countable, then we can define a probability measure simply by assigning a probability to each $\omega \in \Omega$, and provided $\sum_{\omega\in\Omega} \mathbb{P}(\omega) = 1$, and $\mathbb{P}(\omega) \in [0,1]$ then we can define $\mathbb{P}(F) = \sum_{\omega\in F} \mathbb{P}(\omega)$, and this satisfies the definition of a probability. It's not so easy when Ω is uncountable, but consider $\Omega = [0,1]$ and define the **Borel σ-algebra** on $[0,1]$ to be $\sigma(G)$ where $G = \{[0,x] : x \in [0,1]\}$. Then:

Theorem 7.3 (Lebesgue and Borel). *There exists a unique probability measure $\mathbb{P}$ on the space $([0,1], \mathcal{B}([0,1]))$ such that $\mathbb{P}([0,x]) = x$ for all $x \in [0,1]$.*

Some terminology that will be used frequently: we say that an event A happens **almost surely** if $\mathbb{P}(A) = 1$. Note that this may be different to $A = \Omega$: there may be elements $\omega \notin A$, however this set of events is assigned weight 0. Sometimes, these properties can seem counter-intuitive: a simple example is the probability measure given above, which has $\mathbb{P}(\{x\}) = 0$ for all $x \in [0,1]$, but $\mathbb{P}(\cup_{x\in[0,1]}\{x\}) = 1$. This does not contradict the countably additivity property since $[0,1]$ has uncountably elements. However, we can make statements such as: almost surely, ω is irrational under $\mathbb{P}$.

Two useful results about probability measures are the Borel-Cantelli Lemmas. The first lemma says:

Lemma 7.4 (First Borel-Cantelli Lemma). *Let $F_1, F_2, F_3 \ldots$ be a sequence of events. If*

$$\sum_{k=1}^{\infty} \mathbb{P}(F_k) < \infty$$

then almost surely, only finitely many of the events F_n occur (or equivalently $\mathbb{P}(\sum_k \mathbf{1}_{F_k} < \infty) = 1$).

7.2 Random Variables, Expectation

A **random variable (r.v.)** X is a function $X : \Omega \to \mathbb{R}$ such that $\{X \leq x\} := \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$.

NB: Since $\mathcal{B}(\mathbb{R}) = \sigma((-\infty, x], x \in \mathbb{R})$ and $\mathcal{F}$ is a σ-algebra, $X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$ for any $A \in \mathcal{B}(\mathbb{R})$. Hence for any $A \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}(X \in A)$ makes sense!

When this happens, we say that X is a $\mathcal{F}$-measurable function from Ω to $\mathbb{R}$. All the functions we meet in this course will be suitably measurable.

The **distribution function** F_X of a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is the map $F_X : \mathbb{R} \to [0,1]$ where $F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}(\{\omega : X(\omega) \le x\})$.

If X takes values in a countable set $X(\Omega) := \{X(\omega) : \omega \in \Omega\}$, we say that X is a **discrete** r.v., and the **probability mass function** (p.m.f.) is the function $p_X(x) := \mathbb{P}(X = x) = \mathbb{P}(\{\omega : X(\omega) = x\})$.

A r.v. is called **continuous** if there exists a non-negative function $f_X(x)$ on $\mathbb{R}$ such that $\mathbb{P}(a < X \le b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\,\mathrm{d}x$, in which case $f_X(x)$ is the **probability density function (p.d.f.)** of X.

Let X be a r.v. with $0 \le X \le \infty$. If X is discrete, the **expectation** of X is:

$$\mathbb{E}X = \sum_{x \in X(\Omega)} x\,\mathbb{P}(X = x) = \sum_{x \in X(\Omega)} x\,p_X(x) \le \infty.$$

Note: $\mathbb{E}\mathbf{1}_A = 0 \times \mathbb{P}(\mathbf{1}_A = 0) + 1 \times \mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(A)$, since $\{\mathbf{1}_A(\omega) = 1\} = \{\omega \in A\} = A$.

If X is a continuous random variable, then we define

$$\mathbb{E}X = \int_{x \in \mathbb{R}} x f_X(x)\,\mathrm{d}x.$$

More generally, we can define the expectation of a r.v. X with $0 \le X \le \infty$ by:

$$\mathbb{E}X = \sup\{\mathbb{E}Y | Y \text{ is discrete and } Y \le X\}$$

where $Y \le X$ means $Y(\omega) \le X(\omega)$ for all $\omega \in \Omega$.

For a r.v. X on $\mathbb{R}$, write $X^+(\omega) = \max\{0, X(\omega)\}$ and $X^-(\omega) = \max\{0, -X(\omega)\}$. Then X^+ and X^- are both non-negative, and $X = X^+ - X^-$ and $|X| = X^+ + X^-$. If $\mathbb{E}X^+ < \infty$ and $\mathbb{E}X^- < \infty$, so $\mathbb{E}|X| = \mathbb{E}X^+ + \mathbb{E}X^- < \infty$, we say X is **integrable**, write $X \in \mathcal{L}^1$, and define $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$.

Lemma 7.5 (Properties of Expectation). *i)* $|\mathbb{E}[X]| \le \mathbb{E}|X|$. *(e.g. if X is 1 or -1 with equal probability, $\mathbb{E}|X| = 1 \ne 0 = |\mathbb{E}[X]|$.)*

ii) If $X, Y \in \mathcal{L}^1$ and $\lambda, \mu \in \mathbb{R}$ then $\lambda X + \mu Y \in \mathcal{L}^1$ and $\mathbb{E}[\lambda X + \mu Y] = \lambda\mathbb{E}X + \mu\mathbb{E}Y$.

iii) If X is a discrete r.v. and $h : X(\Omega) \to \mathbb{R}$ then $\mathbb{E}h(X) = \sum_{x \in X(\Omega)} h(x)\mathbb{P}(X = x)$, and $h(X)$ is integrable iff $\sum_{x \in X(\Omega)} |h(x)|\mathbb{P}(X = x) < \infty$.

iv) If X is a continuous r.v. with p.d.f. f_X and h is a piecewise continuous function, $\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x) f_X(x)\,\mathrm{d}x$ where $h(X)$ is integrable iff $\int_{\mathbb{R}} |h(x)| f_X(x)\,\mathrm{d}x < \infty$.

v) (Monotone Convergence Theorem) If (X_n) is a sequence of non-negative r.v.'s such that $0 \le X_n \uparrow X$ then $\mathbb{E}X_n \uparrow \mathbb{E}X \le \infty$.

vi) (Fatou's Lemma) If (X_n) is a sequence of non-negative r.v.'s then

$$\mathbb{E}\left[\liminf_n X_n\right] \le \liminf_n \mathbb{E}[X_n].$$

vii) (Cauchy-Schwartz Inequality) If X, Y are integrable r.v.'s, then

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right].$$

In *v*), convergence is to be understood pointwise, so $X_n \uparrow X$ means $X_n(\omega) \uparrow X(\omega)$ for (almost all) $\omega \in \Omega$. Similarly, $\liminf_n X_n$ is the random variable X defined by $X(\omega) = \liminf_n X_n(\omega)$.

Given the notion of expectation, we are able to define two further key properties of random variables: the **variance** of a random variable X is defined to be

$$\mathbb{V}ar(X) := \mathbb{E}\left[X^2 - \mathbb{E}[X]^2\right],$$

and the **covariance** of two random variables X and Y is given by

$$\mathbb{C}ov(X, Y) := \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right].$$

Note that $\mathbb{C}ov(X, X) = \mathbb{V}ar(X)$.

7.3 Conditional Probability and Independence

Given $(\Omega, \mathcal{F}, \mathbb{P})$ and $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, the **conditional probability of A given B** is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

In addition, $\mathbb{P}(\cdot|B) : \mathcal{F} \to [0, 1]$ is a probability measure on $(\Omega, \mathcal{F})$.

Events $A, B \in \mathcal{F}$ are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Then $\mathbb{P}(A|B) = \mathbb{P}(A|B^{\complement}) = \mathbb{P}(A)$. So whether or not B happens does not affect the probability of A happening.

More generally, a finite collection of events, $F_1, \dots, F_n$ are **independent** if, for every $k = 2, \dots, n$ and $i_1 < i_2 < \dots < i_k$ we have:

$$\mathbb{P}(F_{i_1} \cap F_{i_2} \cap \dots \cap F_{i_k}) = \mathbb{P}(F_{i_1})\mathbb{P}(F_{i_2}) \cdots \mathbb{P}(F_{i_k}).$$

Once we have independence, we can state the second Borel-Cantelli Lemma:

Lemma 7.6 (Second Borel-Cantelli Lemma)**.** *Let $F_1, F_2, \dots$ be a sequence of independent events such that $\sum_{n=1}^{\infty} \mathbb{P}(F_n) = \infty$. Then*

$$\mathbb{P}(\textit{infinitely many } F_n \textit{ occur}) = \mathbb{P}\left(\sum_{n \in \mathbb{N}} \mathbf{1}_{F_n} = \infty\right) = 1.$$

Random variables, $X_1, X_2, X_3, \dots$ in a finite or infinite sequence are **independent** if, whenever $i_1, i_2, \dots, i_r$ are distinct integers, and $x_{i_1}, x_{i_2}, \dots, x_{i_r} \in \mathbb{R}$, then

$$\mathbb{P}(X_{i_1} \leq x_{i_1}, X_{i_2} \leq i_2, \dots, X_{i_r} \leq x_{i_r}) = \mathbb{P}(X_{i_1} \leq x_{i_1})\mathbb{P}(X_{i_2} \leq i_2) \cdots \mathbb{P}(X_{i_r} \leq x_{i_r})$$

where we write $\mathbb{P}(A, B)$ as shorthand for $\mathbb{P}(A \cap B)$. We sometimes use the notation $X \perp Y$ to mean X and Y are independent.

If $X_1, X_2, \ldots$ are independent r.v.'s, and $f_1, f_2, \ldots$ are 'nice' functions on $\mathbb{R}$ then $f(X_1), f(X_2), \ldots$ are also independent, and in fact, if f is 'nice' then

$$f(X_1, X_2, \ldots, X_n), X_{n+1}, X_{n+2}, \ldots$$

are also independent. (Here, 'nice' relates to the Borel σ-algebra, and every function we consider will be 'nice').

Lemma 7.7 (Independence means multiply)**.** *Suppose $X \perp Y$. Then*

i) If $X, Y \geq 0$ then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
ii) If $X, Y \in \mathcal{L}^1$, then $XY \in \mathcal{L}^1$ and $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

But: $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \not\Rightarrow X \perp Y$.

7.4 Conditional Expectation

Definition 7.8. For an integrable r.v. X and an event $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, we define the **conditional expectation of X given A** to be:

$$\mathbb{E}[X|A] = \frac{\mathbb{E}[X;A]}{\mathbb{P}(A)} = \frac{\mathbb{E}[X\mathbf{1}_A]}{\mathbb{P}A}.$$

In particular, if X takes values in a countable set, so $X(\Omega) = \{x_1, x_2, \ldots\}$, then

$$\mathbb{E}[X|A] = \sum x_j \mathbb{P}(X = x_j|A).$$

Definition 7.9. Suppose X and Y are random variables.

i) **Conditional Expectation (Discrete)** If Y is a discrete r.v., so $Y(\Omega) = \{y_1, y_2, \ldots\} = \{y_i : i \in I\}$, where either $I = \{1, 2, \ldots, n\}$ or $I = \mathbb{N}$, depending on whether $|Y(\Omega)| < \infty$ or $|Y(\Omega)| = \infty$, and $\mathbb{P}(Y = y_i) > 0$ for all $i \in I$, and $\sum_{i \in I} \mathbb{P}(Y = y_i) = 1$, then we define the **Conditional Expectation of X given Y** to be the random variable:

$$\mathbb{E}[X|Y] = \sum_{i \in I} \mathbf{1}_{Y = y_i} \mathbb{E}[X|Y = y_i].$$

Recall that $\mathbf{1}_A = \mathbf{1}_A(\omega)$ is a r.v. which takes the value 1 if $\omega \in A$, and 0 otherwise. Since the events $\{Y = y_i\}$ are disjoint, only one may happen, and therefore at most one term in the sum is not zero.
Note that if we define the function

$$\psi(y) = \mathbb{E}[X|Y = y]$$

then $\mathbb{E}[X|Y] = \psi(Y)$.
More generally, if $Y_0, Y_1, \ldots, Y_n$ are discrete random variables, with Y_k taking values in a set I_k, we can define the conditional expectation

$$\begin{aligned}&\mathbb{E}[X|Y_0, Y_1, \ldots, Y_n] \\ &\quad = \sum_{i_0 \in I_0, i_1 \in I_1, \ldots, i_n \in I_n} \mathbf{1}_{Y_0 = y_{i_0}, Y_1 = y_{i_1}, \ldots, Y_n = y_{i_n}} \mathbb{E}[X|Y_0 = y_{i_0}, Y_1 = y_{i_1}, \ldots, Y_n = y_{i_n}],\end{aligned}$$

and it follows that $\mathbb{E}[X|Y_0, Y_1, \ldots, Y_n] = \psi(Y_0, Y_1, \ldots, Y_n)$ if we define:

$$\psi(y_0, y_1, \ldots, y_n) = \mathbb{E}[X|Y_0 = y_0, Y_1 = y_1, \ldots, Y_n = y_n].$$

Alternatively, we can define $(\mathcal{F}_n)_{n\geq 0}$ to be the **filtration** of the sequence of random variables $(Y_i)_{i\geq 0}$. That is, $\mathcal{F}_n$ is the collection of events depending only on $Y_0, Y_1, \dots, Y_n$. We usually think of Y_k as the value of something observed at time k, and then $\mathcal{F}_n$ is the 'state of knowledge' or 'history' up to time n.
Given a filtration, we can then define the conditional expectation of X 'given the history up to time n' as

$$\begin{aligned}\mathbb{E}\left[X|\mathcal{F}_n\right] &= \mathbb{E}\left[X|Y_0, Y_1, \dots, Y_n\right] \\ &= \sum_{i_0\in I_0, i_1\in I_1,\dots,i_n\in I_n} \mathbf{1}_{Y_0=y_{i_0},Y_1=y_{i_1},\dots,Y_n=y_{i_n}} \mathbb{E}\left[X|Y_0 = y_{i_0}, Y_1 = y_{i_1}, \dots, Y_n = y_{i_n}\right],\end{aligned}$$

We will also frequently use the shorthand: $\mathbb{E}\left[X|\mathcal{F}_n\right] = \mathbb{E}_n\left[X\right]$.

ii) If X and Y are both continuous random variables with joint density $f_{X,Y}(x, y)$, and marginal densities $f_X(x)$ and $f_Y(y)$ respectively, then the **conditional p.d.f.** of X given $Y = y$ is:

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

If we define the function

$$\psi(y) = \mathbb{E}\left[X|Y = y\right] = \int_{x\in\mathbb{R}} x\, f_{X|Y}(x|y)\,\mathrm{d}x,$$

then the **conditional expectation of** X **given** Y is $\psi(Y)$.
Similarly, if $Y_0, Y_1, \dots, Y_n$ are continuous random variables, with $X, Y_0, \dots, Y_n$ having joint density $f_{X,Y_0,\dots,Y_n}(x, y_0, \dots, y_n)$ and

$$f_{Y_0,\dots,Y_n}(y_0, \dots, y_n) = \int_{x\in\mathbb{R}} f_{X,Y_0,\dots,Y_n}(x, y_0, \dots, y_n),$$

then we can define

$$f_{X|Y_0,\dots,Y_n}(x|y_0, \dots, y_n) = \frac{f_{X,Y_0,\dots,Y_n}(x, y_0, \dots, y_n)}{f_{Y_0,\dots,Y_n}(y_0, \dots, y_n)},$$

and

$$\psi(y_0, y_1, \dots, y_n) = \int_{x\in\mathbb{R}} x\, f_{X|Y_0,Y_1,\dots,Y_n}(x|y_0, y_1, \dots, y_n)\,\mathrm{d}x,$$

and we define the conditional expectation by:

$$\mathbb{E}\left[X|Y_0, Y_1, \dots, Y_n\right] = \psi(Y_0, \dots, Y_n).$$

The notion of a filtration can be generalised. In particular, we can define the σ-algebra $\mathcal{F}_n := \sigma(Y_0, Y_1, \dots, Y_n)$ to be the smallest σ-algebra $\mathcal{G}$ such that $\{Y_j \leq y\} \in \mathcal{G}$ for all $j \in \{0, 1, \dots, n\}$ and all $y \in \mathbb{R}$. Such a σ-algebra exists (since the Y_j's are random variables, this is true for $\mathcal{F}$; then take the intersection over all sub-σ-algebras of $\mathcal{F}$ (that is, all σ-algebras $\mathcal{G}$ such that $A \in \mathcal{G} \implies A \in \mathcal{F}$) such that the Y_j's are measurable — it is easily verified that the intersection of arbitrary σ-algebras is a σ-algebra). Then we can define the conditional expectation of a random variable X with respect to this filtration by defining $\mathbb{E}\left[X|\mathcal{F}_n\right]$ to be the (unique) random variable Z which is measurable with respect to $\mathcal{F}_n$, and such that

$$\mathbb{E}\left[Z; A\right] = \mathbb{E}\left[X; A\right]$$

for all $A \in \mathcal{F}_n$. To further understand such a definition, one really needs a course in measure theory! However, for our purposes, it is enough to understand that this definition coincides with the previous definitions:

$$\mathbb{E}\left[X|A\right] = \mathbb{E}\left[X|\sigma(A)\right], \quad \mathbb{E}\left[X|Y\right] = \mathbb{E}\left[X|\sigma(Y)\right], \quad \mathbb{E}\left[X|Y_0, Y_1, \dots, Y_n\right] = \mathbb{E}\left[X|\sigma(Y_0, Y_1, \dots, Y_n)\right].$$

However, this definition also turns out to be more flexible: in particular, we will want to think of the information as being generated by a continuous time process: $(Y_t)_{t\geq 0}$, and often we want to think about the information we receive by observing the path $(Y_s)_{s\in[0,t]}$ — that is, the information we have observed up to time t. In this case, we can define the filtration $\mathcal{F}_t$ to be the smallest σ-algebra such that Y_s is $\mathcal{F}_t$-measurable for all $s \leq t$. In this way, we are able to condition on complicated properties of the path Y_s (for example, we could condition on the events $\left\{\sup_{s\leq t} Y_s \geq y\right\}$ or $\left\{\int_{s\leq t} Y_s \, \mathrm{d}s \geq y\right\}$) which may not be easily expressible in terms of only finitely many of the values of the Y_i's.

Finally, in terms of σ-algebras, we can discuss independence: in particular, we say that Y is independent of $\mathcal{G}$ if Y is independent of A for every $A \in \mathcal{G}$.

Lemma 7.10 (Properties of Conditional Expectation)**.** *The following properties hold for all the forms of conditional expectation above, if $X, Z, Y_0, \ldots, Y_n$ are r.v.'s, and $\mathcal{G}$ a σ-algebra:*

i) (Taking out what is known): for a 'nice' function h

$$\mathbb{E}\left[h(Y_0, Y_1, \ldots, Y_n)X | Y_0, \ldots, Y_n\right] = h(Y_0, Y_1, \ldots, Y_n)\mathbb{E}\left[X | Y_0, \ldots, Y_n\right];$$

or equivalently: if Y is $\mathcal{G}$-measurable,

$$\mathbb{E}\left[XY | \mathcal{G}\right] = Y\mathbb{E}\left[X | \mathcal{G}\right].$$

ii) (Tower Property): if $m \leq n$,

$$\mathbb{E}\left[\mathbb{E}\left[X | Y_0, \ldots, Y_n\right] | Y_0, \ldots, Y_m\right] = \mathbb{E}\left[X | Y_0, \ldots, Y_m\right],$$

and in particular, if $m = 0$, then

$$\mathbb{E}\left[\mathbb{E}\left[X | Y_0, \ldots, Y_n\right]\right] = \mathbb{E}\left[X\right];$$

or equivalently: if $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then

$$\mathbb{E}\left[\mathbb{E}\left[X | \mathcal{G}\right] | \mathcal{H}\right] = \mathbb{E}\left[X | \mathcal{H}\right].$$

iii) (Combining (i) and (ii)): for a 'nice' function h

$$\mathbb{E}\left[h(Y_0, Y_1, \ldots, Y_n)X\right] = \mathbb{E}\left[h(Y_0, Y_1, \ldots, Y_n)\mathbb{E}\left[X | Y_0, \ldots, Y_n\right]\right];$$

or equivalently: if Y is $\mathcal{G}$-measurable:

$$\mathbb{E}\left[XY\right] = \mathbb{E}\left[Y\mathbb{E}\left[X | \mathcal{G}\right]\right].$$

iv) (Independence): if X is independent of $Y_0, \ldots, Y_n$

$$\mathbb{E}\left[X | Y_0, \ldots, Y_n\right] = \mathbb{E}\left[X\right];$$

or equivalently: if X is independent of $\mathcal{G}$:

$$\mathbb{E}\left[X | \mathcal{G}\right] = \mathbb{E}\left[X\right].$$

v) (Linearity): for $\alpha, \beta \in \mathbb{R}$

$$\mathbb{E}\left[\alpha X + \beta Z | Y_0, \ldots, Y_n\right] = \alpha\mathbb{E}\left[X | Y_0, \ldots, Y_n\right] + \beta\mathbb{E}\left[Z | Y_0, \ldots, Y_n\right];$$

or equivalently:

$$\mathbb{E}\left[\alpha X + \beta Z | \mathcal{G}\right] = \alpha\mathbb{E}\left[X | \mathcal{G}\right] + \beta\mathbb{E}\left[Z | \mathcal{G}\right].$$

vi) *(Conditional Monotone Convergence) Let $(X_n)_{n\in\mathbb{Z}_+}, X$ be non-negative random variables such that $X_n \nearrow X$ (i.e. for almost every $\omega \in \Omega$, $X_n(\omega) \nearrow X(\omega)$), then*

$$\mathbb{E}\left[X_n|\mathcal{G}\right] \nearrow \mathbb{E}\left[X_n|\mathcal{G}\right],$$

and in the special case where $\mathcal{G}$ is the trivial σ-algebra, $\{\emptyset, \Omega\}$, we therefore have $\mathbb{E}\left[X_n\right] \nearrow \mathbb{E}\left[X\right]$.

vii) *(Conditional Jensen's Inequality) Let φ be a convex function. Then:*

$$\mathbb{E}\left[\varphi(X)|\mathcal{G}\right] \geq \varphi(\mathbb{E}\left[X|\mathcal{G}\right]).$$

7.5 Convergence of Random Variables

Let X, X_n be real-valued Random Variables. We are interested in what it means for the sequence X_n to converge to X. We write $\mathcal{L}^k(\mathbb{P})$ (or just $\mathcal{L}^k$ if the choice of $\mathbb{P}$ is clear) for the set of random variables with finite k^{th}moment, $\mathbb{E}\left[|X|^k\right] < \infty$.

Definition 7.11. *i)* We say that X_n **converges to X in distribution** if:

$$\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x) \qquad \text{as } n \to \infty;$$

ii) We say that X_n **converges to X in probability** if, for all $\varepsilon > 0$:

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0 \qquad \text{as } n \to \infty;$$

iii) We say that X_n **converges to X in $\mathcal{L}^k$** if

$$X_n, X \in \mathcal{L}^k \qquad \text{and} \qquad \mathbb{E}\left[|X_n - X|^k\right] \to 0 \qquad \text{as } n \to \infty;$$

iv) We say that X_n **converges almost surely** to X if

$$\mathbb{P}\left(\{\omega : X_n(\omega) \to X(\omega)\}\right) = 1.$$

In general, we get the following implications: for $p \geq 1$, convergence in $\mathcal{L}^{p'}$ implies convergence in $\mathcal{L}^p$, for $p' > p$, convergence in $\mathcal{L}^p$ (for $p \geq 1$) implies convergence in probability, which in turn implies convergence in distribution. Note that for convergence in distribution, the random variables do not need to be defined on the same probability space, wheras for the stronger versons of convergence, the random variables need to be defined on the same probability space. Almost sure convergence implies convergence in probability. In general, almost sure convergence does not imply convergence in $\mathcal{L}^p$, or vice-versa.

It is also often useful to be able to deduce convergence results in expectation from almost sure convergence results. In general, the implication does not hold[14], however there are some special cases where we can show that convergence almost surely implies convergence in expectation, and we summarise these below:

Lemma 7.12. *i) (Monotone Convergence Theorem) If (X_n) is a sequence of non-negative r.v.'s such that $0 \leq X_n \uparrow X$ (i.e. for almost all $\omega \in \Omega$, $X_n(\omega)$ is an increasing sequence, converging to $X(\omega)$) then $\mathbb{E}X_n \uparrow \mathbb{E}X \leq \infty$.*

[14]Consider the following example, where we have $\Omega = [0,1]$ with the uniform measure (so $\mathbb{P}([a,b]) = b - a$ for $0 \leq a \leq b \leq 1$). If we take the sequence of random variables $X_n(\omega) = n\mathbf{1}_{(0,\frac{1}{n})}(\omega)$, then $X_n(\omega) \to 0$ almost surely, since for any ω, $X_n(\omega)$ is zero for n sufficiently large. However, $\mathbb{E}\left[X_n\right] = n\mathbb{P}((0,\frac{1}{n})) = 1$, so $\lim_n \mathbb{E}\left[X_n\right] = 1 \neq 0 = 0$. Observe that this sequence only satisfies the conditions for Fatou's Lemma in Lemma 7.12, and also therefore shows that the inequality in *ii)* can be strict.

ii) (Fatou's Lemma) If (X_n) is a sequence of non-negative r.v.'s then

$$\mathbb{E}\left[\liminf_n X_n\right] \leq \liminf_n \mathbb{E}\left[X_n\right].$$

iii) (Bounded Convergence) If X_n is a sequence of random variables bounded by a constant ($X_n(\omega) \leq M$ for all ω, n, M independent of n, ω), and $X_n \to X$ almost surely, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

iv) (Dominated Convergence) If X_n is a sequence of random variables bounded in absolute value by an integrable random variable Z (i.e. $|X_n|(\omega) \leq Z(\omega)$ for all ω, n and $\mathbb{E}[Z] < \infty$), and $X_n \to X$ almost surely, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Note that all of the integration results above have corresponding conditional versions.

7.6 Central Limit Theorem, Weak Law and Strong Law

One important probabilistic result concerns what happens when we look at the average of many repeated, independent experiments. There are two types of results here. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.'s. Laws of Large numbers essentially state that the mean of the X_i's, $m^N := N^{-1}\sum_{i=1}^N X_i$ converges as $N \to \infty$ to $\mathbb{E}[X_1]$, for different notions of convergence. The Strong law looks at the limit of $z^N := N^{-1/2}\sum_{i=1}^N (X_i - \mathbb{E}[X_1])$, and shows that this converges in distribution to a normally distributed random variable:

Theorem 7.13 (Laws of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be i.i.d., integrable random variables with mean $\mu = \mathbb{E}[X_1]$. Then*

- **Weak Law of Large Numbers:** *m^N converges to μ in probability;*
- **Strong Law of Large Numbers:** *m^N converges to μ almost surely.*

Since almost sure convergence implies convergence in probability, in some sense the weak law above is redundant, however it is useful to separate the two since there are cases where one may only be able to prove a weak law if some of the assumptions are weakened (for example, under milder independence assumptions, or where the r.v.'s are not i.i.d.).

We also have:

Theorem 7.14 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be i.i.d., random variables with mean $\mu = \mathbb{E}[X_1]$ and $\sigma^2 := \mathbb{V}ar(X_i) < \infty$. Then*

$$z^N := \frac{\sum_{i=1}^N (X_i - \mu)}{\sigma\sqrt{N}} \to Z$$

where $Z \sim N(0,1)$ and the convergence is in distribution, i.e. $\mathbb{P}(z^N \leq x) \to \mathbb{P}(Z \leq x)$ for all $x \in \mathbb{R}$, as $N \to \infty$.

References

[1] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus.* Vol. 113. Springer New York, 1998.

[2] P. Mörters and Y. Peres. *Brownian motion*. Cambridge University Press, 2010.

[3] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 2010.

[4] D. Revuz and M. Yor. *Continuous Martingales and Brownian motion*. Springer-Verlag, 1991.