# Lecture 7: Stochastic Integration

**Readings**

Recommended:

- Pavliotis (2014) 3.1-3.2
- Oksendal (2005) 3.1-3.3, 4.1-4.2
- Grimmett and Stirzaker (2001) 13.7, 13.8

Optional:

- Koralov and Sinai (2010) 20.2-20.4, 20.7
- Karatzas and Shreve (1991) 3.1-3.3

## 7.1 Introduction

We want to start talking about phenomena that evolve stochastically. Often these models take the form of ODEs, with forcing terms that are stochastic processes. Some examples include:

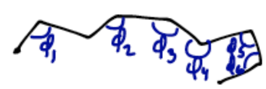
(1) A particle moving in a fluid. Its position $x(t)$ at time $t$ evolves in a velocity field $u(x,t)$ as

$$\frac{dx}{dt} = u(x,t) + \eta(t),$$

where $\eta(t)$ is a stochastic process, representing for example diffusion, or unresolved components of the velocity field.

(2) Polymer dynamics. Given the angles $\phi(t) \in \mathbb{R}^m$ of the polymer and a potential energy $U(\phi)$, the polymer might evolve as

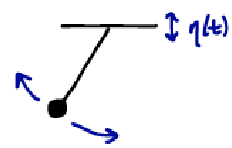
$$\frac{d\phi}{dt} = -\nabla U(\phi) + \eta(t),$$




where $\eta(t)$ is again a stochastic forcing, representing say the random collisions the solvent molecules make with the polymer.

(3) Stochastically forced harmonic oscillator. For example, let $x(t)$ be the angle of a pendulum undergoing small displacements. It could evolve as

$$\underbrace{m\frac{d^2x}{dt^2}}_{\text{acceleration}} + \underbrace{\gamma\frac{dx}{dt}}_{\text{damping}} + \underbrace{kx}_{\substack{\text{restoring}\\ \text{force, eg spring}}} = \underbrace{\eta(t)}_{\substack{\text{stochastic}\\ \text{forcing}}}.$$




(4) Population growth. Let $N(t)$ be the population size at time $t$. It could evolve with a randomly perturbed growth rate as

$$\frac{dN}{dt} = \big(\underbrace{a(t)}_{\substack{\text{growth}\\ \text{rate}}} + \underbrace{\eta(t)}_{\text{noise}}\big)N(t).$$

The general form of these equations is

$$\frac{dX_t}{dt} = b(t,X_t) + \sigma(t,X_t)\eta(t), \tag{1}$$

where $X = (X_t)_{t\geq 0}$ is a stochastic process, $b$ is the deterministic forcing, and $\eta(t)$ is a stochastic process, representing the "noise" or uncertainty in our model, which has amplitude modulated by $\sigma$.

What is a good model for the noise? We want it to be "as random as possible," because any deterministic parts could be modelled separately in $b, \sigma$, or by introducing additional variables to capture time-correlations. This suggests we should choose the noise to be stationary, with

(i) Mean $\mathbb{E}\eta(t) = 0$;
(ii) Covariance function $C_\eta(t) = \delta(t)$, so that $\eta(t_1)$, $\eta(t_2)$ are uncorrelated if $t_1 \neq t_2$.

Therefore $\eta(t)$ should be a white noise, which was introduced in the previous lecture as the derivative of a Brownian motion, $\eta(t) = \frac{dW_t}{dt}$. We also showed the derivative doesn't exist, in a classical sense. It turns out that $\eta(t)$ does exist as a "generalized" process, similar to how $\delta(t)$ is not a function, but a generalized function. We will not pursue this analogy further (though see Koralov and Sinai (2010) Ch.17 for details on generalized processes), but rather we will consider how one could make sense of (1) in an integrated sense.

Consider a discrete version of (1): at time points $t_0 < t_1 < \ldots$, let $X_k = X(t_k)$, and write

$$X_{k+1} - X_k \approx b(t_k, X_k)\Delta t_k + \sigma(t_k, X_k)\underbrace{\eta_k \Delta t_k}_{\Delta W_k}.$$

If $\eta(t) = \frac{dW_t}{dt}$, then $\eta_k \Delta t_k \approx \Delta W_k = W_{t_{k+1}} - W_{t_k}$. Therefore,

$$X_k \approx X_0 + \sum_{j=0}^{k-1} b(t_j, X_j)\Delta t_j + \sum_{j=1}^{k-1} \sigma(t_j, X_j)\Delta W_j.$$

When does the limit of the RHS exist, as $\Delta t_j \to 0$? If it exists, then we can write the solution to (1) as

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s.$$

The solution is a sum of two integrals, each involving stochastic processes. The first integral has the form

$$\int_0^t g(s,\omega)ds \qquad \text{where} \quad g(s,\omega) = b(s, X_s(\omega)),$$

and $\omega$ is an element the sample space associated with the Brownian motion. Provided $g(s,\omega)$ is integrable for each $\omega$, this integral exists as a regular Riemann integral, and produces a random variable as its output.

The second integral has the form

$$\int_0^t f(s,\omega)dW_s \qquad \text{where} \quad f(s,\omega) = \sigma(s, X_s(\omega)).$$

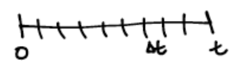
This is a new form of integral. In general it does not exist as a Riemann or Riemann-Stieltjes integral. We'll spend the rest of the lecture studying this integral: learning to define this integral properly, and studying some of its properties.

To start, let's consider an example to illustrate some of the problems that arise with defining this integral.

**Example 7.1** Let $f(s,\omega) = W_s(\omega)$. Consider the integral

$$I = \int_0^t W_s dW_s.$$

Let's try to calculate $I$ as a Riemann-Stieltjes integral. Partition the interval $[0,t]$ into equally-spaced points $\{0, \Delta t, 2\Delta t, \dots, n\Delta t\}$. Let's approximate the integrand using different points within each grid box, and compare approximations by calculating the mean of the approximation.




(1) Approximate the integrand at the left-hand endpoint:

$$I_{LH}^{(n)} = \sum_{k=0}^{n-1} W_{k\Delta t}\left(W_{(k+1)\Delta t} - W_{k\Delta t}\right), \qquad \mathbb{E} I_{LH}^{(n)} = \sum_{k=0}^{n-1} k\Delta t - k\Delta t = 0.$$

(2) Approximate the integrand at the right-hand endpoint:

$$I_{RH}^{(n)} = \sum_{k=0}^{n-1} W_{(k+1)\Delta t}\left(W_{(k+1)\Delta t} - W_{k\Delta t}\right), \qquad \mathbb{E} I_{RH}^{(n)} = \sum_{k=0}^{n-1} (k+1)\Delta t - k\Delta t = t.$$

(3) Approximate the integrand at the midpoint:

$$I_{M}^{(n)} = \sum_{k=0}^{n-1} W_{(k+\frac{1}{2})\Delta t}\left(W_{(k+1)\Delta t} - W_{k\Delta t}\right), \qquad \mathbb{E} I_{M}^{(n)} = \sum_{k=0}^{n-1} (k+\frac{1}{2})\Delta t - k\Delta t = \frac{t}{2}.$$

Depending on which point is used to approximate the integrand, the means differ by an $O(1)$ amount. Hence, as $\Delta t \to 0$, each approximation should converge to a different random variable. Therefore the Riemann-Stieltjes integral of the expression defining $I$ cannot exist.

We could have guessed the approximations would be different, since the Riemann-Stieltjes integral $\int f dg$ is only guaranteed to exist if the total variation of $g$ is finite, but we saw in the last lecture that Brownian motion has infinite variation.

How to we get around this problem? We must decide ahead of time which point to use to approximate the integrand. Each choice gives rise to a different integral. The two most common ones are defined heuristically below; we will see shortly how to construct these more rigorously.

**Definition.** The *Itô integral* is the mean-square limit of the Riemann sums using the LH endpoint to evaluate the integrand. Given a partition $0 = t_0 < t_1 < \cdots < t_n = t$, the Itô integral is

$$\int_0^t f(s,\omega) dW_s \quad = \quad \underset{\max_j |\Delta t_j| \to 0}{\text{m.s.}\lim} \sum_{j=0}^{n-1} f(t_j,\omega)\Delta W_j\,.$$

**Definition.** The *Stratonovich integral* is the mean-square limit of the Riemann sums using the trapezoidal rule evaluate the integrand. Given a partition $0 = t_0 < t_1 < \cdots < t_n = t$, the Stratonovich integral is

$$\int_0^t f(s,\omega)\circ dW_s \quad = \quad \underset{\max_j |\Delta t_j|\to 0}{\mathrm{m.s.}\lim} \sum_{j=0}^{n-1} \frac{f(t_j,\omega)+f(t_{j+1},\omega)}{2}\Delta W_j\,.$$

Equivalently, the Stratonovich integral may be defined as the mean-square limit of the Riemann sums using the midpoint rule evaluate the integrand:

$$\int_0^t f(s,\omega)\circ dW_s \quad = \quad \underset{\max_j |\Delta t_j|\to 0}{\mathrm{m.s.}\lim} \sum_{j=0}^{n} f(t_{j+\frac{1}{2}},\omega)\Delta W_j\,.$$

One can show that the trapezoidal rule and the midpoint rule give the same limit in mean-square, and hence, you can use either to compute the Stratonovich integral. In practice, calculations with the trapezoidal rule are usually easier.

Most of stochastic calculus is developed around the Itô itegral, though in physical problems the Stratonovich integral is occasionally easier to use, because we'll show later that it transforms according to the classical rules of calculus. It doesn't really matter which integral you work with, as you can always convert from one to the other.

**Example 7.2** Evaluate $\int_0^t W_s dW_s$.

*Solution.* Start with a partition $\sigma = \{0 = t_0 < t_1 < \cdots < t_n = t\}$, let $|\sigma| = \max_j \Delta t_j$, and write the approximation using this partition as

$$\begin{aligned} I_t^\sigma &= \sum_{j=0}^{n-1} W_j\Delta W_j \\ &= \frac{1}{2}\sum_{j=0}^{n-1}(W_{j+1}^2 - W_j^2) - \frac{1}{2}\sum_{j=0}^{n-1}(W_{j+1}-W_j)^2 \qquad \text{since } 2W_j\Delta W_j = \Delta(W_j^2) - (\Delta W_j)^2 \\ &= \frac{1}{2}(W_t^2 - W_0^2) - \frac{1}{2}Q_t^\sigma(W), \end{aligned}$$

where $Q_t^\sigma(W)$ is the quadratic variation of the Brownian motion $W = (W_t)_{t\geq 0}$ with respect to the partition. We showed last lecture that $Q_t^\sigma(W) \xrightarrow{m.s.} t$ as $|\sigma|\to 0$, so

$$I_t^\sigma \quad \xrightarrow{m.s.} \quad \frac{1}{2}W_t^2 - \frac{1}{2}t \qquad \text{as } |\sigma| \to 0\,.$$

$\bowtie$

Notice that if we had treated the integral $\int_0^t W_s dW_s$ as a classical Riemann-Stieltjes integral, we would have written

$$\text{(Warning! Incorrect calculations)} \qquad d\left(\frac{1}{2}W_t^2\right) = W_t dW_t\,,$$

and obtained the incorrect result $\int_0^t d\left(\frac{1}{2}W_s^2\right) = \frac{1}{2}W_t^2$. This shows there is something fundamentally new about the Itô integral, and it also shows we don't expect the chain rule to hold in Itô calculus.

***Exercise* 7.1.** Show that $\int_0^t W_s \circ dW_s = \frac{1}{2}W_t^2$.

These exercises show that, as our calculations in Example 7.1 suggested, the Itô and Stratonovich integrals give different results. Notice that the chain rule (via the incorrect calculations above) does give the correct result for the Stratonovich integral.

## 7.2 Construction of the Itô integral

Our heuristic definition of the Itô integral is only heuristic, because we don't yet know whether this definition actually exists – is there a mean-square limit of the approximate integrals, as we refine the partition? And is this limit independent of the sequence of partitions? In this section we give an overview of how to rigorously construct the Itô integral, outlining the major steps, and filling in details when these are helpful for future calculations. The remaining details can be found in the references, e.g. Durrett (1996); Karatzas and Shreve (1991); Grimmett and Stirzaker (2001).

We start by precising the set of functions for which the Itô integral is well-defined. In the following, $W$ is a Brownian motion defined on a sample space $\Omega$ with probability measure $P$.

**Definition.** A stochastic process $f(t,\omega) : [0,\infty)\times\Omega \to \mathbb{R}$ is *adapted to W* or simply *adapted* if, $\forall t \geq 0$, $f(t,\omega)$ depends only on the values of $W_s(\omega)$ for $s \leq t$, and not on any values in the future nor on any other stochastic processes.[1] In other words, we can decide the value of $f(t,\omega)$, knowing the history of $W$ up to time $t$. A random variable $X$ is adapted to $(W_s)_{0\leq s\leq t}$ if $X 1_{t'\geq t}(t')$ is adapted, i.e. the random variable only depends on past values of $W$.

**Examples 7.3** Which of the following stochastic processes are adapted to $W$?

(i) $X_t = W_{t/2}$
(ii) $X_t = W_{2t}$
(iii) $X_t = \int_0^t W_s ds$ .
(iv) $X_t = W_t V_t$ where $V$ is a Brownian motion independent of $W$
(v) $X_t = W_t + \xi$ where $\xi$ is a Bernoulli random variable independent of $W$
(vi) $X_t = \sin(t^2+5)+e^{-t}$

*Solution.* (i), (iii), (vi) are adapted. (ii), (iv), (v) are not adapted.

Note however that when there are additional sources of randomness, as in (iv),(v), it is possible to enhance the definition of adapted to allow $X_t$ to depend on this additional randomness, as long as $W_t$ remains a Brownian motion in the enhanced probability space.

$\bowtie$

[1] The rigorous definition of adapted is that the random variable $\omega \to f(t,\omega)$ is $\mathscr{F}_t$-measurable, where $\mathscr{F}_t$ is the $\sigma$-algebra generated by the random variables $\{W_s\}_{s\leq t}$ and which includes all the null events $\mathscr{N} = \{A \in \mathscr{F} : P(A) = 0\}$, where $\mathscr{F}$ is the $\sigma$-algebra associated with $(W_t)_{t\geq 0}$. That is, $\mathscr{F}_t$ is the smallest $\sigma$-algebra containing sets of the form $\{\omega : W_{t_1}(\omega) \in F_1, W_{t_2}(\omega) \in F_2, \ldots, W_{t_k} \in F_k\}$, where $t_1,\ldots,t_k \leq t$, and $F_j \subset \mathbb{R}$ are Borel sets augmented by the null sets. See Grimmett and Stirzaker (2001), Section 13.8.

**Definition.** Let $\mathscr{V}$ be the class of stochastic processes $f : [0,\infty)\times\Omega\to\mathbb{R}$ such that $f$ is adapted, and such that $\|f\|_{\mathscr{V}} < \infty$ where the norm $\|\cdot\|_{\mathscr{V}}$ is defined by [2]

$$\|f\|^2_{\mathscr{V}} := \mathbb{E}\left(\int_0^\infty f^2(t,\omega)dt\right). \tag{2}$$

It can be shown that $\mathscr{V}$ is a Hilbert space with norm[3] $\|\cdot\|_{\mathscr{V}}$ (Grimmett and Stirzaker (2001), p.540, and references therein).

We'll define the Itô integral for $f\in\mathscr{V}$ over an infinite interval, and write this as

$$I(f) = \int_0^\infty f(s,\omega)dW_s. \tag{3}$$

For a finite integral, simply multiply by an indicator function: $\int_0^t f dW_s = \int_0^\infty f(s,\omega)1_{s\leq t}(s)dW_s$. The Itô integral $I(f)$ is a random variable defined on the probability space $\Omega$. A useful way to compare integrals is via the $L^2(\Omega)$-norm $\|\cdot\|_2$, defined for random variables $X:\Omega\to\mathbb{R}$ as $\|X\|_2^2 = \mathbb{E}X^2$. Applying this norm to an Itô integral gives

$$\|I(f)\|_2^2 = \mathbb{E}\left(\int_0^\infty f(t,\omega)dW_t\right)^2. \tag{4}$$

Here is the strategy for constructing the Itô integral in (3):

1. Define it for a class of "simple" functions $\Phi\subset\mathscr{V}$.
2. Show that $f\in\mathscr{V}$ can be approximated by a sequence of simple functions $\{\phi_n\}_{n=1}^\infty\subset\Phi$, i.e. show there exists a sequence such that $\|\phi_n - f\|_{\mathscr{V}}\to 0$.
3. Define $I(f) = \mathrm{m.s.}\lim_{n\to\infty} I(\phi_n)$, and show this limit is well-defined.

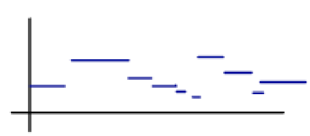
**Step 1.**

**Definition.** Let $0 = t_0 < t_1 < \cdots < t_n = T$, for some $T > 0$. A function $\phi\in\mathscr{V}$ is an *adapted step function* (called *simple function*) if

$$\phi(t,\omega) = \sum_{j=0}^{n-1} e_j(\omega)1_{(t_j,t_{j+1}]}(t),$$

where random variable $e_j$ is adapted to $(W_s)_{s\leq t}$ and has $\mathbb{E}e_j^2 < \infty$.

Such a function $\phi$ is piecewise constant, with constants that are random variables depending on values of the particular Brownian path only up to the beginning of the current time interval.



[2] We are ignoring questions of measurability here. See Durrett (1996) for more details.

[3] Technically, it is only a norm on the set of equivalence classes obtained from the equivalence relation $\psi\sim\phi$ whenever $P(\psi = \phi) = 1$.

We define the Itô integral for adapted step functions as

$$I(\phi) = \int_0^\infty \phi \, dW_t(\omega) = \sum_{j=0}^{n-1} e_j(\omega)(W_{t_{j+1}}(\omega) - W_{t_j}(\omega)). \tag{5}$$

This integral has two important properties.

**Lemma** (Non-anticipating property)**.** *If $\phi$ is a adapted step function, then*

$$\mathbb{E}\int_0^\infty \phi dW_t = 0\,. \tag{6}$$

*Proof.* Writing $\Delta W_j = W_{t_{j+1}} - W_{t_j}$, notice that $e_j, \Delta W_j$ are independent, because $e_j$ is adapted, so it only depends on values of $W_s$ for $s \le t_j$, and $\Delta W_j$ only depends on values of $W_s$ for $s \in (t_j, t_{j+1}]$. Therefore $\mathbb{E}I(\phi) = \sum_j \mathbb{E}e_j \mathbb{E}\Delta W_j = 0$. □

**Lemma** (Itô isometry for adapted step functions)**.** *If $\phi$ is a adapted step function, then*

$$\mathbb{E}\left(\int_0^\infty \phi dW_t\right)^2 = \mathbb{E}\int_0^\infty \phi^2 dt\ . \tag{7}$$

In other words, the norms (2), (4) for Itô integrals are related by $\|I(\phi)\|_2 = \|\phi\|_{\mathscr{V}}$. This relationship will be important as it gives a way to convert between the norm $\|\cdot\|_{\mathscr{V}}$ defined on adapted functions, and the norm $\|\cdot\|_2$ defined on Itô integrals – it allows us to say that if $\|\phi_1 - \phi_2\|_{\mathscr{V}}$ is small, then $\|I(\phi_1) - I(\phi_2)\|_2$ is also small.

*Proof.*

$$\mathbb{E}\left(\int_0^\infty \phi dW_t\right)^2 = \mathbb{E}\Big(\sum_i e_i \Delta W_i\Big)^2 = \mathbb{E}\left(\sum_j e_j^2 (\Delta W_j)^2 + 2\sum_{j<k} e_j e_k \Delta W_j \Delta W_k\right).$$

As before $e_j, \Delta W_j$ are independent. We also have that if $j < k$, then $\Delta W_k$ is independent of $e_j, e_k, \Delta W_j$. Therefore the above equals

$$\sum_j \mathbb{E}e_j^2 \Delta t_j + 2\sum_{j<k} \mathbb{E}(e_j e_k \Delta W_j)\underbrace{\mathbb{E}\Delta W_k}_{=0} \;=\; \mathbb{E}\sum_j e_j^2 \Delta t_j \;=\; \mathbb{E}\int_0^\infty \phi^2 dt.$$

□

**Step 2.**

**Proposition.** *Given $f \in \mathscr{V}$, there exists a sequence of adapted step functions $\phi = \{\phi_n\}_{n\in\mathbb{N}}$ such that $\|\phi_n - f\|_{\mathscr{V}} \to 0$ as $n \to \infty$.*

For a proof, see Grimmett and Stirzaker (2001), Section 13.8, or Oksendal (2005), Section 3.1.

**Step 3.** Given $f \in \mathscr{V}$ and a sequence of adapted step functions $\phi = \{\phi_n\}_{n\in\mathbb{N}}$ such that $\|\phi_n - f\|_{\mathscr{V}} \to 0$ as $n \to \infty$, the Itô integral is defined as

$$I(f) = \underset{n\to\infty}{\text{m.s.}\lim} I(\phi_n)\,. \tag{8}$$

To show that $I(f)$ is well-defined, we need to show (i) that the limit exists, and (ii) that it is independent of the particular sequence of adapted step functions. (These calculations are from Grimmett and Stirzaker (2001), p.542.) For (i), notice that $\phi_m - \phi_n$ is an adapted step function, so

$$\begin{aligned}
\|I(\phi_m) - I(\phi_n)\|_2 &= \|I(\phi_m - \phi_n)\|_2 && \text{Itô integral is linear for step functions}\\
&= \|\phi_m - \phi_m\|_{\mathscr{V}} && \text{Itô isometry for step functions, (7)}\\
&\le \|\phi_m - f\|_{\mathscr{V}} + \|\phi_n - f\|_{\mathscr{V}} && \text{triangle inequality on norm } \|\cdot\|_{\mathscr{V}}\\
&\to 0 \quad \text{as } m,n\to\infty && \text{by construction of } \phi_n.
\end{aligned}$$

Therefore the sequence $I(\phi_n)$ is a Cauchy sequence in $\|\cdot\|_2$, so from a theorem in probability theory,[4] there is a random variable $I(\phi)$ such that $I(\phi_n) \xrightarrow{m.s.} I(\phi)$ as $n \to \infty$.

To show (ii), suppose there is another sequence of adapted step functions $\rho = \{\rho_n\}_{n\in\mathbb{N}}$ sch that $\|\rho_n - f\|_{\mathscr{V}} \to 0$ as $n \to \infty$, and let $I(\rho)$ be the limit of the corresponding integrals. By the triangle inequality,

$$\|I(\phi) - I(\rho)\|_2 \;\le\; \|I(\phi) - I(\phi_n)\|_2 + \|I(\phi_n) - I(\rho_n)\|_2 + \|I(\rho_n) - I(\rho)\|_2\,.$$

The first and third terms on the right-hand side go to zero as $n \to \infty$. The second term can be written using the Itô isometry as as

$$\|I(\phi_n - \rho_n)\|_2 \;=\; \|\phi_n - \rho_n\| \;\le\; \|\phi_n - f\| + \|\rho_n - f\|\,,$$

which also goes to zero as $n \to \infty$. Therefore $\|I(\phi) - I(\rho)\|_2 = 0$, so $I(\phi) = I(\rho)$ with probability 1.

We obtain a construction of the Itô integral, from (8).

## 7.3 Properties of the Itô integral

Given $f, g \in \mathscr{V}$, the Itô integral has the following properties.

(i) (Linearity) for $a, b \in \mathbb{R}$, $\int_0^\infty (af + bg)dW_t) = a\int_0^\infty f dW_t + b\int_0^\infty g dW_t$.

(ii) (Non-anticipating property)

$$\mathbb{E}\int_0^\infty f dW_t = 0\,. \tag{9}$$

(iii) (Itô isometry)

$$\mathbb{E}\left(\int_0^\infty f(t,\omega)dW_t\right)^2 = \mathbb{E}\int_0^\infty f^2(t,\omega)dt. \tag{10}$$

***Exercise* 7.2.** Show these properties, by showing they are true for adapted step functions, and then show they must be true in the limit (8).

[4] The theorem says that if a sequence of random variables $\{X_n\}$ is a Cauchy sequence, meaning that $X_n - X_m \xrightarrow{\gamma} 0$ as $n, m \to \infty$, where $\gamma$ is any of the stochastic modes of convergence, then $\exists\, X$ such that $X_n \xrightarrow{\gamma} X$ (see e.g. Breiman (1992), Section 2.8).

Another property is:

(iv) The integral $I_t = \int_0^t f(s,\omega)dW_s$ can be chosen to depend continuously on $t$ almost surely.

For a proof, see Oksendal (2005), Section 3.2.

Finally, a generalisation of the Itô isometry is

(v) For $g,h \in \mathscr{V}$,

$$\mathbb{E}\left(\int_0^t g(s,\omega)dW_s \int_0^t h(s,\omega)dW_s\right) = \int_0^t \mathbb{E}[g(s,\omega)h(s,\omega)]ds\,. \tag{11}$$

To prove this property, apply Itô's isometry with $f = h + g$.

Formally, the Itô isometry (10) and its generalisation (11) can be derived from the substitutions $\mathbb{E}dW_u = \mathbb{E}dW_v = 0$, $\mathbb{E}dW_u dW_v = \delta(u-v)dudv$, and the fact that $g(u), h(v)$ are adapted, so they are each independent of $dW_u$, $dW_v$ respectively. That is, write

$$\mathbb{E}\left[\left(\int_0^t g(u)dW_u\right)\left(\int_0^t h(v)dW_v\right)\right] = \int_0^t\int_0^t \mathbb{E}[g(u)h(v)dW_u dW_v]\,.$$

Now decompose the integrand into different pieces, depending on the relationship between $u,v$. Since $1_{v<u}(u,v) + 1_{u<v}(u,v) + 1_{u=v}(u,v) = 1$, where $1_A(u,v)$ is the indicator function for set $A$, we can write

$$\begin{aligned}\mathbb{E}[g(u)h(v)dW_u dW_v] = \mathbb{E}[g(u)h(v)1_{v<u}(u,v)dW_u dW_v] + \mathbb{E}[g(u)h(v)1_{u<v}(u,v)dW_u dW_v]\\ + \mathbb{E}[g(u)h(v)1_{u=v}(u,v)dW_u dW_v]\,.\end{aligned}$$

Therefore,

$$\begin{aligned}\int_0^t\int_0^t \mathbb{E}[g(u)h(v)dW_u dW_v] &= \int_0^t\int_0^t (\mathbb{E}g(u)h(v))\,\delta(u-v)dudv + \int_0^t\int_0^t (\mathbb{E}g(u)h(v)1_{u<v}dW_u)(\mathbb{E}dW_v)\\ &\quad + \int_0^t\int_0^t (\mathbb{E}g(u)h(v)1_{v<u}dW_v)(\mathbb{E}dW_u)\\ &= \int_0^t (\mathbb{E}g(u)h(u))du\,.\end{aligned}$$

## 7.4 Itô formula

We can finally begin to make sense of the "stochastic" ODE (1). Rewrite the equation as

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t, \tag{12}$$

to see that $X_t$ should be the solution to the following integral equation:

$$X_t = X_0 + \int_0^t b(s,X_s)ds + \int_0^t \sigma(s,X_s)dW_s. \tag{13}$$

In practice, one usually writes (12), to mean (13). A solution to (13) is called a *diffusion process*. We will study equations of this form much more in the coming lectures. We may also consider the more general equation

$$X_t = X_0 + \int_0^t b(s,\omega)ds + \int_0^t \sigma(s,\omega)dW_s \tag{14}$$

where $b, \sigma \in \mathscr{V}$. A solution to (14) is called an *Itô process*. Clearly, a diffusion process is a particular kind of Itô process.

Suppose we have an Itô process $X_t$, and suppose $Y_t = g(t, X_t)$. What equation does $Y_t$ satisfy?

If the rules of classical calculus were satisfied, we would write $\frac{dY_t}{dt} = \frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}\frac{dX_t}{dt}$ and so

$$\text{(Incorrect!)} \qquad dY_t = \frac{\partial g}{\partial t}(t,X_t)dt + \frac{\partial g}{\partial x}(t,X_t)b(t,\omega)dt + \frac{\partial g}{\partial x}(t,X_t)\sigma(t,\omega)dW_t. \tag{15}$$

We saw in Example 7.2 that this chain rule doesn't hold for Itô integrals. Therefore, we need to develop a new version of the chain rule that works for the Itô integral.

**Itô formula.** *Let $X_t$ be the solution to*

$$dX_t = b(t,\omega)dt + \sigma(t,\omega)dW_t,$$

*where $b, \sigma$ are adapted functions. Given $g \in C^2([0,\infty)\times\mathbb{R})$, the process $Y_t = g(t,X_t)$ solves the equation*

$$dY_t = \frac{\partial g}{\partial t}(t,X_t)dt + \frac{\partial g}{\partial x}(t,X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)(dX_t)^2 , \tag{16}$$

*where $(dX_t)^2$ is computed according to the rules*

$$dt\cdot dt = dt \cdot dW_t = dW_t \cdot dt = 0, \qquad dW_t \cdot dW_t = dt.$$

*Specifically,*

$$dY_t = \left(\frac{\partial g}{\partial t}(t,X_t) + \frac{\partial g}{\partial x}(t,X_t)b(t,\omega) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)\sigma^2(t,\omega)\right)dt + \frac{\partial g}{\partial x}(t,X_t)\sigma(t,\omega)dW_t . \tag{17}$$

Compared to the chain rule from classical calculus (15), the Itô formula includes an extra drift term $\frac{1}{2}\sigma^2\frac{\partial g}{\partial x}dt$.

Roughly, the Itô formula comes from Taylor-expanding $g(t,X_t)$ near some point $(t,x)$:

$$\Delta Y = \frac{\partial g}{\partial t}\Delta t + \frac{\partial g}{\partial x}\Delta X + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(\Delta X)^2 + \frac{1}{3!}\frac{\partial^3 g}{\partial x^3}(\Delta X)^3 + \ldots$$

To approximate $\Delta Y$ to $O(\Delta t)$, one must account for the fact that $\Delta W \sim O(\Delta t^{1/2})$, and therefore we must go to second-order in the Taylor expansion to retain all the terms that are first-order in $\Delta t$. We keep terms involving $(\Delta W)^2 \sim O(\Delta t)$, but throw out terms containing $\Delta t^2$ or $\Delta t \Delta W \sim O(\Delta t^{3/2})$.

**Example 7.4** Let $Y_t = \frac{1}{2}W_t^2$. What is $dY_t$?

*Solution.* Let $X_t = W_t$, $g(t,x) = \frac{1}{2}x^2$, $Y_t = g(t,X_t) = \frac{1}{2}W_t^2$. We know that $dX_t = dW_t$. Then

$$dY_t = \frac{\partial g}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(dX_t)^2 = X_t dX_t + \frac{1}{2}(dX_t)^2 = W_t dW_t + \frac{1}{2}dt .$$

Let's solve this equation to check its consistency.

$$Y_t = \int_0^t W_s dW_s + \int_0^t \frac{1}{2}ds = \frac{1}{2}W_t^2 - \frac{1}{2}t + \frac{1}{2}t = \frac{1}{2}W_t^2 .$$

We obtain the correct formula for $Y_t$. ⋈

**Example 7.5** Let $Y_t = e^{W_t}$. What is $dY_t$?

*Solution.* Here $g(t,x) = e^x$, so Itô's formula yields

$$dY_t = \frac{\partial g}{\partial x} dW_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(dW_t)^2 = \frac{1}{2}e^{W_t}dt + e^{W_t}dW_t = \frac{1}{2}Y_t dt + Y_t dW_t \,.$$

Notice the extra drift term $\frac{1}{2}Y_t dt$ that arises in Itô calculus. $\bowtie$

*Sketch proof of Itô's formula.* (from E et al. (2014)) Assume that $g, g_t, g_x, g_{xx}, g_{xxx}$ are bounded, and that $b, \sigma$ are bounded adapted step functions. When they aren't, take limits of adapted step functions.[5] Let's show Itô's formula for $g = g(x)$; the proof for $g = g(t,x)$ is very similar. We show Itô's formula on a bounded time interval $[0,T]$.

Let $\{t_i\}_{i=1}^N$ be a partition of $[0,T]$. Without loss of generality we can assume the discontinuities of $b, \sigma$ lie on the grid points of the partition.We have

$$Y_t - Y_0 = \sum_j g(X_{t_{j+1}}) - g(X_{t_j}) \,,$$

which holds exactly when $b, \sigma$ are step functions, so by Taylor-expanding $g(X_{t_{j+1}})$ about point $X_{t_j}$, we obtain

$$Y_t - Y_0 = \sum_j g(X_{t_{j+1}}) - g(X_{t_j}) = \sum_j \left( \underbrace{g'(X_{t_j})\Delta X_{t_j}}_{\text{Term 1}} + \frac{1}{2}\underbrace{g''(X_{t_j})(\Delta X_{t_j})^2}_{\text{Term 2}} + R_j \right) .$$

Here $\Delta X_{t_j} = X_{t_{j+1}} - X_{t_j}$, and $R_j = o(|\Delta X_{t_j}|^2)$ is a remainder term in the Taylor expansion of $g$. It is bounded by $\max(\partial_x g, \partial_{xx} g)\left((\Delta t_j)^3 + (\Delta t_j)^2 \Delta W_j + \Delta t_j (\Delta W_j)^2 + (\Delta W_j)^3\right)$, hence $\sum_j R_j$ can be shown to go to zero as $\sup_j \Delta t_j \to 0$, using the same techniques as used in what follows.

Let's estimate each of the remaining terms. Since $b, \sigma$ are step functions, we have that $\Delta X_{t_j} = b(t_j)\Delta t_j + \sigma(t_j)\Delta W_{t_j}$, without making any approximation. (When they aren't step functions, then Taylor-expand them, and bound the remainder terms.) Therefore

$$\begin{aligned} \text{Term 1} &= \sum g'(X_{t_j}) b(t_j) \Delta t_j + \sum_j g'(t_j)\sigma(t_j)\Delta W_{t_j} \\ &\xrightarrow{m.s.} \int_0^t b(s) g'(X_s) ds + \int_0^t \sigma(s) g'(X_s) dW_s \qquad\qquad \text{as } \sup_j |\Delta t_j| \to 0, \end{aligned}$$

by the definition of the Itô integral. This gives us part of (16). For the second term:

$$\text{Term 2} = \sum_j g''(X_{t_j})\left(b^2(t_j)(\Delta t_j)^2 + 2b(t_j)\sigma(t_j)\Delta t_j \Delta W_{t_j} + \sigma^2(t_j)(\Delta W_{t_j})^2\right).$$

[5] For a full proof, see Ikeda and Watanabe (1981) or Karatzas and Shreve (1991).

Let's bound some of these terms. Let $K$ be a bound for $b, \sigma, g''$ on $[0,T]$.

$$\left|\sum_j g''(X_{t_j})b^2(t_j)(\Delta t_j)^2\right| \le K\sum_j |\Delta t_j|^2 \le KT\sup_j \Delta t_j \to 0 \quad \text{a.s.}.$$

$$\left|\sum_j g''(X_{t_j})b(t_j)\sigma(t_j)\Delta t_j \Delta W_{t_{j+1}}\right| \le K\sum_j |\Delta t_j \Delta W_{t_j}| \le KT\sup_j |\Delta W_{t_j}| \to 0 \quad \text{a.s.}.$$

The fact that $\sup_j |\Delta W_{t_j}| \to 0$ as $\sup_j |\Delta t_j| \to 0$ follows from the Law of the Iterated Logarithm. Finally, we have that

$$\sum_j g''(X_{t_j})\sigma^2(t_j)(\Delta W_{t_j})^2 \xrightarrow{m.s.} \int_0^t \sigma^2(s) g''(X_s)ds \qquad \text{as } \sup_j |\Delta t_j| \to 0.$$

Put this all together to get the result. □

## 7.5 Itô calculus in higher dimensions

We may wish to calculate integrals that depend on several independent Brownian motions, such as $\int W_t^{(2)} dW_t^{(1)}$, $\int e^{W_t^{(1)}+W_t^{(2)}} dW_t^{(3)}$, etc. The Itô integral can be extended to this case.

**Definition.** Let $W_t = (W_t^{(1)}, \ldots, W_t^{(n)})^T$ be an $n$-dimensional Brownian motion. Let $M = M(t,\omega)$ be a matrix-valued process that is adapted to $W_t$. The *multi-dimensional Itô integral* of $M$ is

$$\int_S^T M dW_t = \int_S^T \begin{pmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ \vdots & & & \\ M_{m1}\cdots & & M_{mn} & \end{pmatrix} \begin{pmatrix} dW_t^{(1)} \\ \vdots \\ dW_t^{(n)} \end{pmatrix} \tag{18}$$

That is, the integral is a vector whose $k$th component is $\sum_{j=1}^n \int_S^T M_{kj}(s,\omega) dW^{(j)}(s,\omega)$.

Itô's formula can be extended to higher dimensions.

**Itô formula** (Multidimensional)**.** *Let $X_t$ solve*

$$dX_t = b(t,\omega)dt + \sigma(t,\omega)dW_t\,,$$

*where $X_t, b \in \mathbb{R}^n$, $\sigma \in \mathbb{R}^{n\times m}$, $W_t \in \mathbb{R}^m$, and $b, \sigma$ are adapted to $W_t$. Let $Y_t = f(X_t)$, where $f \in C^2(\mathbb{R}^n)$. Then*

$$dY_t = \nabla f(X_t)\cdot dX_t + \frac{1}{2}(dX_t)^T \nabla^2 f(X_t)\, dX_t, \tag{19}$$

*where $\nabla^2 f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{i,j}$ is the Hessian matrix of $f$, and products of increments are evaluated using the rules following* (16) *plus the additional rule*

$$dW_t^{(i)}\cdot dW_t^{(i)} = dt\,, \qquad dW_t^{(i)}\cdot dW_t^{(j)} = 0 \qquad \text{for } i \ne j\,.$$

*Therefore $Y_t$ solves the equation*

$$dY_t = \left(b\cdot\nabla f + \frac{1}{2}\sigma\sigma^T : \nabla^2 f\right)dt + (\nabla f)^T \sigma\, dW_t,$$

*where $A : B = Tr(A^T B) = \sum_{i,j} A_{ij}B_{ij}$.*

*Proof.* Very similar to the 1d case. □

***Exercise* 7.3.** Give a heuristic justification for (19), based on Taylor-expansion as in the example in the previous section.

# References

Breiman, L. (1992). *Probability*. SIAM.

Durrett, R. (1996). *Stochastic Calculus: A practical introduction*. CRC Press, Taylo & Francis Group.

E, W., Li, T., and Vanden-Eijnden, E. (2014). *Applied Stochastic Analysis*. In preparation.

Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*. Oxford University Press.

Ikeda, N. and Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes*. Elsevier.

Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*. Springer.

Koralov, L. B. and Sinai, Y. G. (2010). *Theory of Probability and Random Processes*. Springer.

Oksendal, B. (2005). *Stochastic Differential Equations*. Springer, 6 edition.

Pavliotis, G. A. (2014). *Stochastic Processes and Applications*. Springer.

Varadhan, S. R. S. (2007). *Stochastic Processes*, volume 16 of *Courant Lecture Notes*. American Mathematical Society.