UNIVERSITAT DE BARCELONA

Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques i Informàtica
Universitat de Barcelona

STOCHASTIC INTEGRALS AND WONG-ZAKAI THEOREMS

Autor: Josep-Maria Armengol Collado

Director: Dr. Carles Rovira Escofet
Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 18 de gener de 2019

Abstract

We start by characterizing Brownian motion and giving its main properties, and then we focus on studying Itô's and Stratonovich's integral. We take special interest in comparing both perspectives and proving Wong-Zakai theorems, which connect stochastic and deterministic behaviour. Finally, it is also presented a brief introduction to stochastic differential equations, demonstrating a result for the existence and uniqueness of solutions.

Resum

Comencem caracteritzant el moviment brownià i donem les seves propietats principals, per tot seguit, centrant-nos en l'estudi de les dues integrals estocàstiques més utilitzades: la d'Itô i la de Stratonovich. Posem especial interès en comparar-les, així com en provar els teoremes de Wong i Zakai que connecten el comportament estocastic i el determinista. Finalment, donem una introducció a les equacions diferencials estocàstiques i demostrem un resultat que n'assegura l'existència i unicitat de solucions.

2010 Mathematics Subject Classification. 60G05, 60H05, 60H10, 60H30, 60H35, 60H40

Acknowledgements/Agraïments

I would like to express my gratitude to my advisor Dr. Carles Rovira for his constant help, advice and because he has achieved transmitting me his patience and his way of reasoning. For having spent long hours guiding me, but has always with a smile, thank you.

També m'agradaria agrair a la meva família, no només per costejar-me els estudis sinó per estar sempre al meu costat i recolzar-me en els moments més difícils. També als meus amics, en particular, els que he conegut fent el doble grau, perquè després d'haver compartit tants moments plegats, sé que us guardo per a tota la vida.

Contents

Introduction

One of the biggest challenges for mathematicians of all times has been characterizing randomness. Although the modern formulation of probability culminated with the axiom system presented by A. N. Kolmogorov in 1933 combining the concept of sample space and measure theory, there was still some difficulties when adding uncertainty to the models that describe the evolution of a certain variable that fluctuates with time.

However, the publication of *On stochastic processes (Infinitely divisible laws of probability)* in 1942 by the Japanese mathematician Kiyosi Itô meant a paradigm shift in the field of stochastic processes. Providing a rigorous definition for stochastic integrals which generalized Lebesgue-Stieljes integration, he initiated a parallel analysis theory for the study of stochastic differential equations. This topic, one of the most active branches of mathematics now, is applicable in a wide range of scientific fields such as Physics, where it constitutes the key to insert noise to the systems, or Economics, having the most notable application in Black-Scholes model.

Although his approach is still used when modeling new phenomena, twenty years later, a Russian physicist called Ruslan Stratonovich proposed a different way of defining stochastic integrals, an alternative tool that satisfies deterministic chain rule in exchange for losing some other good properties. Then, an intuitive question arises: which is the correct interpretation? What benefits are brought by each of them? Furthermore, an article of E. Wong and M. Zakai shows how Stratonovich's description is the limit of sequences of deterministic integrals (and an analogous result for solutions of stochastic and ordinary differential equations), which will determine the election of this new representation when we are interested in approximating stochastic behaviour by well-behaved processes.

Throughout this project, we aim to lay the foundation of stochastic calculus, beginning by characterizing Brownian motion, the integrator we are going to use, but also building stochastic integrals emphasizing the benefits and differences between them. The third chapter is devoted to giving a basic introduction to stochastic differential equations, providing some illustrative examples, and rigorously proving a theorem for the existence and uniqueness of solutions. Our second goal will be studying Wong-Zakai theorems, even computing some numerical simulations to show the behavior of approximating sequences of deterministic processes predicted by theory.

As a bibliographic reference I mostly follow [4], which collects accessible but also rigorous proofs for the main results, and I also used [3, 9]. Sometimes, I needed to appeal the original articles [11, 13], becoming a challenge to adapt them to a more modern notation.

Chapter 1

Brownian Motion

With this first chapter, we want to define and give the main properties of one-dimensional Brownian motion. The name of this stochastic process comes from the botanist Robert Brown, who observing pollen in a water drop was the first in wondering the origin of its erratic movement. In 1905, Albert Einstein developed Kinetic Theory to give a satisfactory explanation to this phenomenon, but a complete mathematical construction of it did not come until the contributions of Nolbert Wiener, which explains why it is also often referred as Wiener process.

Before giving a formal definition and studying its interesting properties, we make a short reminder of notions related to stochastic processes that will appear during the whole project. Finally, we end this chapter with a brief description of white noise, a useful concept when modeling uncertainty, strongly related to Brownian motion, which is going to take center stage later in the chapter devoted to stochastic differential equations.

We fix a probability space $(\Omega, \mathscr{F}, P)$, being Ω the sample space, $\mathscr{F}$ a σ-field and P a probability measure. This will be implicit for the rest of the work.

1.1 Preliminaries

Definition 1.1. A *stochastic process* with state space S is a family $\{X_t,\ t \in T\}$ of random variables $X_t : \Omega \to S$ indexed by a *parameter set* T and defined on the same $(\Omega, \mathscr{F}, P)$.

Stochastic processes are an excellent tool to model evolution in the real world as they capture the stochasticity typical from Physics, Economics or Biology. Then T, which can be countable or not, will be usually thought as time and X_t will mean the state of the process at time t. In the sequel, we will take $T = [a, b] \subseteq [0, \infty)$ and S will be a subset of $\mathbb{R}$. Although it is possible to think of stochastic processes as random vectors, in order to study the evolution of the values of a given observable it is convenient to fix $\omega \in \Omega$.

Definition 1.2. If $\{X_t,\ t \in [a, b]\}$ is a stochastic process, for every $\omega \in \Omega$, the mapping $X(\omega) : T \to S$ defined by $X(\omega)(t) = X_t(\omega)$ is called *sample path* or *trajectory*.

We introduce now some notions that describe the main types of stochastic processes.

Definition 1.3. We say that a stochastic process $\{X_t,\ t \in [a,b]\}$ has *independent increments* if the random variables $X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}}$ are independent for any $t_1 < \dots < t_k$.

Definition 1.4. A stochastic process $\{X_t,\ t \in [a,b]\}$ is said to have *stationary increments* if for any $s < t$, $X_t - X_s$ has the same law as the random variable $X_{a+(t-s)} - X_a$.

Definition 1.5. A *filtration* is a family $\{\mathscr{F}_t,\ t \in [a,b]\}$ of sub-σ-fields of $\mathscr{F}$ such that $\mathscr{F}_s \subset \mathscr{F}_t \subset \mathscr{F}$ for any $a \leq s \leq t$. We say that a stochastic process $\{X_t,\ t \in [a,b]\}$ is adapted to $\{\mathscr{F}_t,\ t \in [a,b]\}$ if X_t is $\mathscr{F}_t$-measurable for all $t \in [a,b]$.

We should think that $\mathscr{F}_t$ contains all information available at time t so that knowing this, it is possible to decide if a random variable X_t adapted to the filtration has occurred or not at time t. There is a natural way to define a filtration for a given stochastic process $\{X_t,\ t \in [a,b]\}$ and consists of taking $\mathscr{F}_t = \sigma\{X_s,\ a \leq s \leq t\},\ t \in [a,b]$, only when it is right continuous. Fortunately, Brownian motion, the process we are going to focus on the next section satisfies this property. Next definition plays a fundamental role in stochastic theory.

Definition 1.6. A stochastic process $\{X_t,\ t \in [a,b]\}$ is called a *martingale* with respect to the filtration $\mathscr{F}_t$ if it holds:

1. For each $t \in [a,b]$, the variable X_t belongs to $L^1(\Omega)$.
2. X_t is adapted to the filtration $\mathscr{F}_t$.
3. $E(X_t|F_s) = X_s$ for every $s,t \in [a,b]$ such that $s \leq t$.

If the process satisfies $E(X_t|F_s) \geq X_s$ instead of 3, we say that it is a *submartingale*. When the inequality hold is $E(X_t|F_s) \leq X_s$, then it is a *supermartingale*.

Thus, if X_t is a martingale, knowing the value of $X_{t'}$ for $a \leq t' < b$ is equal to have the entire history of X_t up to t'. They exhibit really good properties and have been extensively studied for mathematicians because of their applications in probability theory and gambling. For this reason, it is a desired property when developing new stochastic processes. Below are some celebrated inequalities for martingales due to Joseph L. Doob, one of the initiators in the study of such processes.

Theorem 1.7. *(Doob martingale inequalities).* If $\{X_t,\ t \in [a,b]\}$ is a right continuous martingale (i.e., almost all sample paths are right continuous functions on $[a,b]$), Then for any $\epsilon > 0$ it holds

$$P\left\{\sup_{t\in[a,b]} |X_t| \geq \epsilon\right\} \leq \frac{1}{\epsilon} E\,|X_b|\,.$$

Besides, there is also a version called L^p*-inequality* that states

$$E\left(\sup_{t\in[a,b]} |X_t|^p\right) \leq \left(\frac{p}{p-1}\right)^p \sup_{t\in[a,b]} E\left(|X_t|^p\right),\ \forall p \in (1,\infty)\,.$$

1.2 Definition and features

Definition 1.8. The *Brownian motion* (also called *Wiener process*) starting in $t = 0$ is a continuous time stochastic process $\{W_t,\ t \geq 0\}$ that satisfies:

1. $W_t = 0$ a.s.
2. $\{W_t,\ t \geq 0\}$ has independent increments.
3. $W_t - W_s$ is $N(0, t-s)$, $\forall\ 0 \leq s \leq t$. Consequently, it has stationary increments too.

Unless otherwise specified, we will always assume that W_t starts in 0. As it has been described, the qualitative change in time of a Brownian particle which starts from $x = 0$ at time 0 (i.e. is a Dirac mass at zero, $\delta_{\{0\}}$) only depends on the length of the increment and the future evolution is independent of the past (*Markov property*). In addition, each W_t is normally distributed so, for $t > 0$, its density is of the form,

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

Remark 1.9. After deriving last expression it can be proved that p_t satisfies the diffusion physical equation,

$$\frac{\partial}{\partial t} p_t(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2} p_t(x).$$

The rest of this section is devoted to proving some results related to Brownian motion. The first is necessary for a future demonstration, and the second implies that this process satisfies martingale property (it is actually its most important example).

Lemma 1.10. For any $s, t \geq 0$, $E(W_s W_t) = \min(s, t)$.

Proof. It is sufficient to check that assuming $s < t$,

$$E(W_s W_t) = E(W_s (W_t - W_s) + W_s^2) = E(W_s)E(W_t - W_s) + E(W_s^2) = 0 + s = s,$$

having used conditions 2 and 3 of the definition and the basic properties of the expectation. □

Proposition 1.11. The Brownian motion $\{W_t,\ t \geq 0\}$ is a martingale.

Proof. First of all, we note that the increment $W_t - W_s$ is independent of $\mathscr{F}_s = \sigma\{W_r;\ 0 \leq r \leq s\}$ since $W_t - W_s$ and $W_r - W_0 = W_r$ are independent for $0 \leq r \leq s \leq t$ and $\mathscr{F}_s$ is generated by W_s. To prove now that $\{W_t,\ t \geq 0\}$ is a martingale, we take $\mathscr{F}_t = \sigma\{W_s;\ s \leq t\}$, and then, for any $0 \leq s < t$:

$$E(W_t|\mathscr{F}_s) = E(W_t - W_s|\mathscr{F}_s) + E(W_s|\mathscr{F}_s) = E(W_t - W_s) + W_s = W_s,$$

because we have shown that $W_t - W_s$ is independent of $\mathscr{F}_s$ and $E(W_t - W_s) = 0$, but also we use the fact that W_s is $\mathscr{F}_s$-measurable[1]. □

[1] The properties of the conditional expectation we make use here are (a) and (d) from the list we present in the appendices, subsection A.1.

1.3 Path properties

As we mentioned before, we aim to use Brownian motion to build stochastic integrals, so we should first study how its trajectories behave and which properties we can take advantage of. To do so, we need to define some previous notions.

Definition 1.12. A function $f : [a,b] \to \mathbb{R}$, $0 \leq a < b$ is called *Hölder continuous* with exponent $\gamma > 0$ at point s if there exists a real constant $K > 0$ satisfying:

$$|f(t) - f(s)| \leq K|t-s|^{\gamma}, \quad \forall t \in [a,b].$$

Remark 1.13. For any real function defined on $[a,b]$ it holds:

1. f is continuously differentiable $\to f$ is Hölder continuous with exponent 1.
2. f is γ-Hölder continuous $\to f$ is uniformly continuous.

Let us outline the following theorem, which is necessary to prove the continuity of the paths of the Brownian motion. Its proof can be read in [3].

Theorem 1.14. *(Kolmogorov's continuity criterion)* Let $\{X_t, t \in [a,b]\}$ be a stochastic process, if for some α, $\beta > 0$ and any t, s there exists a real constant $K > 0$ such that:

$$E(|X_t - X_s|^{\alpha}) \leq K|t-s|^{1+\beta},$$

then, the trajectories of the process are γ-Holder a.s. with $\gamma < \frac{\beta}{\alpha}$.

Proposition 1.15. The Brownian motion $\{W_t,\ t \geq 0\}$ is a.s γ-Hölder continuous for any $\gamma < 1/2$. In particular, their paths are continuous.

Proof. For all integers $m \geq 1$ we have

$$E(|W_t - W_s|^{2m}) = \frac{1}{\sqrt{2\pi r}} \int_{\mathbb{R}} x^{2m} e^{-\frac{x^2}{2r}}\, dx = \frac{1}{\sqrt{2\pi}} r^m \int_{\mathbb{R}} y^{2m} e^{-\frac{y^2}{2}}\, dy = Kr^m = K|t-s|^m,$$

being $r = t - s$ and $y = \frac{x}{\sqrt{r}}$ the changes of variables. More precisely, $K = \frac{(2m)!}{2^m m!}$. Thus, using Kolmogorov's theorem we know that $\{W_t,\ t \geq 0\}$ is Hölder continuous a.s. for exponents γ satisfying

$$0 < \gamma < \frac{\alpha}{\beta} = \frac{1}{2} - \frac{1}{2m}.$$

□

Theorem 1.16. For $\gamma \in (\frac{1}{2}, 1]$ the Brownian motion $\{W_t,\ t \geq 0\}$ is nowhere Hölder continuous with exponent γ. Consequently, the sample paths are nowhere differentiable and have infinite variation on each subinterval.

Proof. For simplicity, we consider only times $0 \leq t \leq 1$. Fix $\gamma \in (\frac{1}{2}, 1]$ and an integer N so big that $N(\gamma - \frac{1}{2}) > 1$. If we assume that the path $t \to W_t(\omega)$ is γ-Hölder continuous at $s \in [0,1)$, then $|W_t(\omega) - W_s(\omega)| \leq K|t-s|^\gamma$ for some constant K. Let n be large enough and set $i = [ns] + 1$. Note that for $j = i, i+1, \ldots, i+N-1$,

$$\left|W_{\frac{j}{n}}(\omega) - W_{\frac{j+1}{n}}(\omega)\right| \leq \left|W_s(\omega) - W_{\frac{j}{n}}(\omega)\right| + \left|W_s(\omega) - W_{\frac{j+1}{n}}(\omega)\right|$$
$$\leq K\left(\left|s - \frac{j}{n}\right|^\gamma + \left|s - \frac{j+1}{n}\right|^\gamma\right) \leq \frac{M}{n^\gamma},$$

given another constant M. Define now,

$$A^i_{M,n} = \left\{\left|W_{\frac{j}{n}}(\omega) - W_{\frac{j+1}{n}}\right| \leq \frac{M}{n^\gamma},\ j = i, i+1, \ldots, i+N-1\right\},$$

so $\omega \in A^i_{M,n}$ for some $1 \leq i \leq n$, $M \geq 1$ and large n. Therefore, the collection of trajectories Hölder continuous with exponent γ at s is contained in the set

$$\bigcup_{M=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\bigcup_{i=1}^{n} A^i_{M,n}.$$

Next, we show that this set has null probability. Since

$$P\left(\left|W_{\frac{1}{n}}\right| \leq \frac{M}{n^\gamma}\right) = \sqrt{\frac{n}{2\pi}}\int_{-Mn^{-\gamma}}^{Mn^{-\gamma}} e^{-\frac{nx^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\int_{-Mn^{1/2-\gamma}}^{Mn^{1/2-\gamma}} e^{-\frac{y^2}{2}}dy \leq Cn^{1/2-\gamma},$$

due to the fact that $W_{\frac{j+1}{n}} - W_{\frac{j}{n}}$ are $N\left(0, \frac{1}{n}\right)$ and independent, we have

$$P\left(\bigcap_{n=k}^{\infty}\bigcup_{i=1}^{n} A^i_{M,n}\right) \leq \liminf_{n\to\infty} P\left(\bigcup_{i=1}^{n} A^i_{M,n}\right) \leq \liminf_{n\to\infty}\sum_{i=1}^{n} P\left(A^i_{M,n}\right)$$
$$\leq \liminf_{n\to\infty} n\left(P\left(\left|W_{\frac{1}{n}}\right| \leq \frac{M}{n^\gamma}\right)\right)^N \leq \liminf_{n\to\infty} nC\left(n^{1/2-\gamma}\right)^N = 0,$$

for all k, M. Consequently, $P\left(\bigcup_{M=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\bigcup_{i=1}^{n} A^i_{M,n}\right) = 0$. On the other hand, in case that the trajectories were differentiable at s, then W_t would be Hölder continuous (with exponent 1) at this point, but is a.s not so. Moreover, if it was of finite variation at any subinterval, it would then be differentiable a.e. there[2], which would lead to a contradiction. □

Notice that the previous results do not say anything for $\gamma = 1/2$. To study what happens then, it is required a deeper knowledge in probability theory, in particular, in the analysis of the so-called *modulus of continuity*. The conclusion is that Brownian motion is not Hölder continuous of degree 1/2 either. Now, we introduce a fundamental property that provides a measure of the hardness of a function: *quadratic variation*. As it will be accurately described next chapter, this gives an explanation to why stochastic calculus behaves differently from classical deterministic one.

[2]The exact result in measure theory we make reference can be read in the appendices, subsection A.2.

Proposition 1.17. Consider a partition of our interval $[a,b]$ given by $\Delta_n = \{a = t_0^n \leq t_1^n \leq \ldots \leq t_n^n = b\}$, $n \geq 1$ such that $\|\Delta_n\| = \sup_{1\leq i\leq n}(t_i - t_{i-1})$ tends to zero if $n \to \infty$. Then, in $L^2(\Omega)$,

$$\sum_{i=1}^{n}\left(W_{t_i} - W_{t_{i-1}}\right)^2 \to b-a, \quad \text{as } n \to \infty.$$

Proof. Set from now on $\Delta_i W = W_{t_i} - W_{t_{i-1}}$ and $\Delta_i t = t_i - t_{i-1}$. We also omit the dependence on n of each t_i to avoid clutter. Our aim is proving that

$$\lim_{n\to\infty} E\left[\left(\sum_{i=1}^{n}(\Delta_i W)^2 - (b-a)\right)^2\right] = 0.$$

Note that $b - a = \sum_{i=1}^{n}\Delta_i t$, so we can reformulate the problem in terms of

$$\Phi_n = \sum_{i=1}^{n}\left[(\Delta_i W)^2 - \Delta_i t\right] = \sum_{i=1}^{n} X_i,$$

where $X_i = (\Delta_i W)^2 - \Delta_i t$. Then, $\Phi_n^2 = \sum_{i,j=1}^{n} X_i X_j$. Now, we can use the computation of moments of $W_t - W_s$ we developed indirectly in the proof of Proposition 1.15 to reason:

- If $i \neq j$, $E\left(X_i X_j\right) = 0$. This is because of the independent increments of W_t and $E\left[(W_t - W_s)^2\right] = |t-s|$.
- If $i = j$, using $E\left[(W_t - W_s)^4\right] = 3|t-s|^2$ we can compute:

$$\begin{aligned} E\left(X_i^2\right) &= E\left[(\Delta_i W)^4 - 2\Delta_i t\,(\Delta_i W)^2 + (\Delta_i t)^2\right] \\ &= 3(\Delta_i t)^2 - 2(\Delta_i t)^2 + (\Delta_i t)^2 = 2(\Delta_i t)^2. \end{aligned} \tag{1.1}$$

Therefore, we finally get

$$E\left(\Phi_n^2\right) = \sum_{i=1}^{m_n} 2(\Delta_i t)^2 \leq 2\|\Delta_n\| \sum_{i=1}^{n}\Delta_i t = 2(b-a)\|\Delta_n\| \to 0, \quad \text{as } \|\Delta_n\| \to 0.$$

□

This proposition justifies the heuristic idea that $dW_t^2 \approx dt$, so the products dt^2, $dtdW_t$ become negligible in front of dt and dW_t. These rules are collected in the called *Itô's table*, which will be an important tool for next chapter.

$\cdot$	dW_t	dt
dW_t	dt	0
dt	0	0

Table 1.1: Itô's table

1.4 White noise

In experimental sciences, the most usual way to model randomness is through small perturbations acting on the system with arbitrary directions so that the average of all of them is zero. In order to add this effect to the equations that govern the dynamics, it is common to consider a Gaussian process X_t determined by its zero mean and the covariance function $E([X_t - E(X_t)][X_s - E(X_s)]) = E(X_t X_s)$. This expectation is in fact known as the *autocorrelation function* $r(s,t)$ of X_t, and expresses how short is the effect of the pulses causing this noise. When $r(s,t) = c(t-s)$ for some function $c : \mathbb{R} \to \mathbb{R}$ and $E(X_t) = E(X_s)$, $\forall t, s$ in the parameter set we say that it is *stationary in the wide sense*. A special example of this kind of processes is white noise.

Definition 1.18. We say that a Gaussian process ξ_t is *white noise* if it is stationary in the wide sense and satisfies $E(\xi_t) = 0$ and $E(\xi_t \xi_s) = \delta_0(t-s)$, being δ_0 a Dirac delta function.

It means that ξ_t and $\xi_{t+\Delta t}$ will be uncorrelated for arbitrary small values of Δt. Moreover, it is called "white" because the spectrum of ξ_t is flat. More precisely, the *spectral density* of a process is defined as the Fourier transform of the autocorrelation function,

$$f(\lambda) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t} c(t) dt,$$

with $\lambda \in \mathbb{R}$. Then, for white noise,

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t} \delta_0(t) dt = \frac{1}{2\pi},$$

which means that all "frequencies" contribute equally to the correlation function, the same idea that makes all colors form white light.

The first attempt to describe rigorously white noise was in connection with the *theory of generalized functions* (or distributions). Here, we make a brief introduction to this field in order to justify why white noise plays the role of the "derivative" of Brownian motion, even after proving that it is nowhere differentiable. Let us begin with a general definition.

Definition 1.19. Denoting by K the space of all infinitely differentiable functions $\varphi(t)$, $t \in \mathbb{R}$ that vanish identically outside a finite interval (which in general can be different for each φ), we define a *generalized stochastic process* as the random variable $\Phi(\varphi)$, $\varphi \in K$ satisfying:

1. Linearity on K a.s.:

$$\Phi(c_1\varphi + c_2\psi) = c_1\Phi(\varphi) + c_2\Phi(\psi), \ \forall c_1, c_2 \in \mathbb{R}, \ \varphi, \psi \in K, \ a.s.$$

2. Continuity: for any sequence of functions φ_{kj} converging to φ_k in K, $\forall k = 1, 2 \ldots, n$ as $j \to \infty$, it holds that,

$$\left(\Phi(\varphi_{1j}), \ldots, \Phi(\varphi_{nj})\right) \to \left(\Phi(\varphi_1), \ldots, \Phi(\varphi_n)\right), \quad \text{as } j \to \infty \text{ in distribution.}$$

For example, we can consider the generalized stochastic process Φ created by means of taking continuous stochastic processes X and inserting them to the formula

$$\Phi_X(\varphi) = \int_{-\infty}^{\infty} \varphi(t)X(t)dt, \forall \varphi \in K.$$

In the same line, we can also define,

$$E(\Phi(\varphi)) = m(\varphi) \quad \text{and} \quad E([\Phi(\varphi) - m(\varphi)][\Phi(\psi) - m(\psi)]) = Cov(\varphi, \psi).$$

Somehow, we can imagine that our goal is measuring values of $X(t)$ but the measuring instrument affects this estimation smoothly so that $\Phi_X(\varphi)$ exhibits better properties than X, for instance, it will always exist the derivative of Φ_X (even if X does not have it) which is defined symbolically as $\dot{\Phi}(\varphi) = -\Phi(\dot{\varphi})$, or explicitly,

$$\int_{-\infty}^{\infty} \varphi(t)\dot{X}(t)dt = -\int_{-\infty}^{\infty} \dot{\varphi}(t)X(t)dt.$$

Notice that this expression is similar to the formula of integration by parts. Let us set an example, we say that a generalized process is Gaussian when $(\Phi(\varphi_1), \ldots, \Phi(\varphi_n))$ is normally distributed $\forall \varphi_1, \ldots, \varphi_n \in K$. In particular, we are interested in focusing on the Wiener process and its derivative. As we said, we can represent its generalization as

$$\Phi_W(\varphi) = \int_0^{\infty} \varphi(t)W_t dt,$$

because we have set $W_t \equiv 0$ for $t < 0$. Then, $m(\varphi) = 0$ and for the covariation, we use Lemma 1.10 to write

$$Cov(\varphi, \psi) = \int_0^{\infty}\int_0^{\infty} \min(t, s)\, \varphi(t)\psi(t)dtds.$$

Hence, integrating by parts we obtain the expression

$$Cov(\varphi, \psi) = \int_0^{\infty} [\hat{\varphi}(t) - \hat{\varphi}(\infty)]\left[\hat{\psi}(t) - \hat{\psi}(\infty)\right] dt,$$

where

$$\hat{\varphi}(t) = \int_0^t \varphi(s)ds, \quad \hat{\psi}(t) = \int_0^t \psi(s)ds.$$

On the other hand, if we compute the derivative of the Brownian motion as a generalized stochastic process, it will satisfy $\dot{m}(\varphi) = -m(\dot{\varphi})$, but also

$$\dot{Cov}(\varphi, \psi) = Cov(\dot{\varphi}, \dot{\psi}) = \int_0^{\infty} \varphi(t)\psi(t)dt = \int_0^{\infty}\int_0^{\infty} \delta(t-s)\, \varphi(t)\psi(t)dtds.$$

Consequently, white noise can be written as the derivative of the Wiener process, when both are considered as generalized stochastic processes. For this reason, we will sometimes find in the literature the short-hand notation

$$\xi_t = \dot{W}_t \quad \text{or} \quad W_t = \int_0^t \xi_t ds.$$

Chapter 2

Stochastic Integration

After explaining why a classical approach would not work when using Brownian motion as integrator, this chapter is devoted to build and compare our two main stochastic integrals: Itô's and Stratonovich's. Finally, we present and illustrate numerically the first Wong-Zakai theorem, which establishes a relation between stochastic integrals and the limits of deterministic ones.

Although this topic is usually dealt with martingales, here we decided to tackle it only for the Brownian motion, as well as we only considered one-dimensional processes, which simplifies the notation and let us go further. For brevity, we will follow using the notation $\Delta_i W = W_{t_i} - W_{t_{i-1}}$, $\Delta_i t = t_i - t_{i-1}$ and also omitting the n-dependence of the partition $\|\Delta_n\|$ to avoid clutter.

2.1 Motivation

In order to give rigorous mathematical meaning to stochastic differential equations driven by Brownian motion, it is required to introduce a new kind of integral compatible with the non-differentiability drawback of W_t. The key is that, although there exist many extensions of classical Riemann approach to integrate respect to some function g, they only allow well-behaved integrators. For instance, a bounded function f defined on $[a, b]$ is *Riemann-Stieltjes integrable*, if the following limit exists:

$$\int_a^b f(t)\, dg(t) = \lim_{\|\Delta_n\| \to 0} \sum_{i=1}^{n} f(\tau_i)\left(g(t_i) - g(t_{i-1})\right),$$

where $\Delta_n = \{t_0, t_1, \ldots, t_n\}$ is a partition of $[a, b]$, $\|\Delta_n\| = \sup_{1 \le i \le n} \Delta_i t$ and τ_i is the evaluation point in the interval $[t_{i-1}, t_i]$. A known result on measure theory asserts the ensuing.

Theorem 2.1. Let $g : [a, b] \to \mathbb{R}$ be a bounded function. If the Riemann-Stieljes integral $\int_a^b f(x)\, dg(x)$ exists for any $f \in \mathcal{C}([a, b])$, then g is of bounded variation.

As we have seen in the preceding chapter, Brownian motion has an infinite variation on each subinterval (Theorem 1.16) so it is not allowed to be used as integrator in the previous formula. This, together with its non-zero quadratic variation, make that if we try to extend the idea of the Riemann sums to the stochastic case we get defective properties, such us the fact that the election of the evaluation point τ_i will determine the result.

Example 2.2. Given a partition Δ_n of $[a,b]$, we want to compute:

$$R_n := \sum_{k=1}^{n} W_{\tau_k} \Delta_k W,$$

where $\tau_k = (1-\lambda)\, t_k + \lambda t_{k-1}$, being $0 \leq \lambda \leq 1$ a fixed parameter. We can actually rewrite last expression as:

$$R_n := \sum_{k=1}^{n} W_{\tau_k} \Delta_k W = \frac{W_{b-a}^2}{2} \overbrace{-\frac{1}{2}\sum_{k=1}^{n} (\Delta_k W)^2}^{A} + \overbrace{\sum_{k=1}^{n} \left(W_{\tau_k} - W_{t_{k-1}}\right)^2}^{B} \\ + \underbrace{\sum_{k=1}^{n} \left(W_{t_k} - W_{\tau_k}\right)\left(W_{\tau_k} - W_{t_{k-1}}\right)}_{C}.$$

In effect, if we expand

$$A = -\frac{1}{2}\sum_{k=1}^{n} \left(W_{t_k}^2 + W_{t_{k-1}}^2 - 2W_{t_k}W_{t_{k-1}}\right), \quad B = \sum_{k=1}^{n} \left(W_{\tau_k}^2 + W_{t_{k-1}}^2 - 2W_{\tau_k}W_{t_{k-1}}\right),$$
$$C = \sum_{k=1}^{n} \left(W_{t_k}W_{\tau_k} - W_{t_k}W_{t_{k-1}} - W_{\tau_k}^2 + W_{\tau_k}W_{t_{k-1}}\right),$$

we can check that

$$A + B + C = \sum_{k=1}^{n} \left(W_{\tau_k}W_{t_k} - W_{\tau_k}W_{t_{k-1}}\right) - \frac{1}{2}\sum_{k=i}^{n} \left(W_{t_k}^2 - W_{t_{k-1}}^2\right) = R_n - \frac{W_{b-a}^2}{2}.$$

Besides, according to Proposition 1.17 on quadratic variation, $A \to \frac{a-b}{2}$ as $n \to \infty$ in $L^2(\Omega)$. A similar argument reflects that $B \to \lambda(b-a)$ as $n \to \infty$. However, for C we proceed:

$$E\left(\left|\sum_{k=1}^{n} \left(W_{t_k} - W_{\tau_k}\right)\left(W_{\tau_k} - W_{t_{k-1}}\right)\right|^2\right) = \sum_{k=1}^{n} E\left(\left|W_{t_k} - W_{\tau_k}\right|^2\right) \cdot E\left(\left|W_{\tau_k} - W_{t_{k-1}}\right|^2\right)$$
$$= \sum_{k=1}^{n} (1-\lambda)\, \Delta_k t \lambda \Delta_k t \leq \lambda(1-\lambda)(a-b)\|\Delta_n\|.$$

which tends to zero as $n \to \infty$. Gathering the three last convergences together, it results in

$$\lim_{n\to\infty} R_n = \frac{W_{b-a}^2}{2} + \left(\lambda - \frac{1}{2}\right) \cdot (b-a) \quad \text{in } L^2(\Omega).$$

Remark 2.3. In particular, the limit of the Riemann sum depends on the choice of intermediate points:

- The election of $\lambda = 0$ corresponds to the definition of Itô's integral (discussed later in Section 2.2). That is:
$$\int_a^b W_t dW_t = \frac{W_{b-a}^2}{2} - \frac{b-a}{2}.$$
- Alternately, one can take $\lambda = \frac{1}{2}$ to follow the expected rules of deterministic calculus. This definition is due to Stratonovich (more deeply developed in Section 2.3) and uses the symbol "$\circ$" to be differentiated from Itô's description:
$$\int_a^b W_t \circ dW_t = \frac{W_{b-a}^2}{2}.$$

2.2 Itô's integral

To build the very first stochastic integral defined by K. Itô, we need to consider a Wiener process $\{W_t,\ t \geq 0\}$ and its associated natural filtration $\{\mathscr{F}_t,\ t \geq 0\}$. Throughout this chapter, we will work with $L^2_{ad}\left([a,b] \times \Omega\right)$, the space of stochastic processes $f_t\left(\omega\right)$ satisfying:

- f is jointly measurable in (t, ω) with respect to the σ-field $\mathcal{B}\left([a,b]\right) \otimes \mathscr{F}$.
- $\int_a^b E\left(f_t^2\right) dt < \infty$.

2.2.1 Construction

For clarity, we have divided the discussion into two parts: we start by defining the Itô's stochastic integral for the so-called step processes and after, we generalize for any function in our set $L^2_{ad}\left([a,b] \times \Omega\right)$.

Itô's integral for step processes

Definition 2.4. A stochastic process f is said to belong to the subset $\mathcal{E} \subseteq L^2_{ad}\left([a,b] \times \Omega\right)$ consisting of *step processes* if it can be written as:

$$f_t = \sum_{i=1}^{n} u_{i-1} \mathbb{1}_{[t_{i-1}, t_i)}\left(t\right),$$

with $a = t_o \leq \ldots \leq t_n = b$ and where u_{i-1} are $\mathscr{F}_{t_{i-1}}$-measurable random variables in $L^2\left(\Omega\right)$.

Note that having specified u_{i-1} for each $\mathbb{1}_{[t_{i-1}, t_i)}$ we achieve the state of the subinterval not to depend on future values (the class of the *nonanticipating processes*). Next definition corresponds to the first attempt to give a natural meaning to a stochastic integral formula.

Definition 2.5. In the above conditions, we define the *Itô's stochastic integral for a step process* $f \in \mathcal{E}$ as:

$$I(f) = \sum_{i=1}^{n} u_{i-1} \Delta_i W. \tag{2.1}$$

We will usually write $I(f)$ as $\int_a^b f_t dW_t$.

Particularly, $I(\cdot)$ is a random variable and has some good qualities summarized by next proposition.

Proposition 2.6. For any $f \in \mathcal{E}$, $I(f)$ has the following properties:

1. *Linearity*: given $g \in \mathcal{E}$ and $a, b \in \mathbb{R}$, then $af + bg \in \mathcal{E}$, but also, $I(af + bg) = aI(f) + bI(g)$.
2. *Centrality*: $E(I(f)) = 0$.
3. *Isometry*: $E\left(|I(f)|^2\right) = \int_a^b E\left(f_t^2\right) dt$.

Proof. To see the first point, we shall take a partition $a = t_o \leq \ldots \leq t_n = b$ and write

$$f_t = \sum_{i=1}^{n} u_{i-1} \mathbb{1}_{[t_{i-1}, t_i)}(t) \quad \text{and} \quad g_t = \sum_{i=1}^{n} v_{i-1} \mathbb{1}_{[t_{i-1}, t_i)}(t),$$

where u_{i-1} and v_{i-1} are $\mathscr{F}_{t_{i-1}}$-measurable random variables. Note that if the two partitions happen to be different, it will always be possible to find a common refinement of them. Then

$$af + bg = \sum_{i=1}^{n} (au_{i-1} + bv_{i-1}) \mathbb{1}_{[t_{i-1}, t_i)}(t),$$

and so

$$I(af + bg) = \sum_{i=1}^{n} (au_{i-1} + bv_{i-1}) \Delta_i W = a \sum_{i=1}^{n} u_{i-1} \Delta_i W + b \sum_{i=1}^{n} v_{i-1} \Delta_i W = aI(f) + b(g).$$

To prove that $I(f)$ is a centered random variable we just have to use the independency of u_{i-1} and $W_{t_i} - W_{t_{i-1}}$ to compute:

$$E\left(\int_a^b f_t dW_t\right) = E\left(\sum_{i=1}^{n} u_{i-1} \Delta_i W\right) = \sum_{i=1}^{n} E(u_{i-1}) E(\Delta_i W) = 0.$$

Finally, to prove the last property we proceed by splitting the sum to reason separately,

$$E\left(\int_a^b f_t dW_t\right)^2 = E\left(\sum_{i,j} u_{i-1} u_{j-1} (\Delta_i W)(\Delta_j W)\right) = \sum_{i=1}^{n} E\left(u_{i-1}^2 (\Delta_i W)^2\right) + 2 \sum_{i<j} E\left(u_{i-1} u_{j-1} (\Delta_i W)(\Delta_j W)\right).$$

The second term vanishes because for $i < j$, $u_{i-1} u_{j-1} \Delta_i W$ is independent of $\Delta_j W$, so

$$E\left(u_{i-1} u_{j-1} \Delta_i W \Delta_j W\right) = E\left(u_{i-1} u_{j-1} \Delta_i W\right) E\left(\Delta_j W\right) = 0.$$

On the other hand, the first term turns out to be:

$$\sum_{i=1}^{n} E(u_{i-1}^2)E((\Delta_i W)^2) = \sum_{i_1}^{n} E(u_{i-1}^2)\,(\Delta_i t) = \int_a^b E(f_t^2)dt,$$

again, due to the independence of u_{i-1}^2 and $(\Delta_i W)^2$. □

It is important to highlight that isometry takes equidistant points in $\mathcal{E}$ to equidistant points in L^2 so, in particular, $I(\cdot)$ maps a Cauchy sequence in $\mathcal{E}$ into a Cauchy sequence in L^2. The relevance of this observation is understood by next theorem, which states that any $f \in L^2_{ad}([a,b] \times \Omega)$ can be approximated by elements of $\mathcal{E}$.

Theorem 2.7. For all $f \in L^2_{ad}([a,b] \times \Omega)$ there exists a sequence $\{f^n\}_{n\geq 1} \subseteq \mathcal{E}$ such that:

$$\lim_{n\to\infty} \int_a^b E\left(|f_t - f_t^n|^2\right) = 0. \tag{2.2}$$

This means that $\mathcal{E}$ is dense in L^2_{ad}.

Proof. To show the result, we first consider different cases to finish with the general proof.
Case 1: $E(f_t f_s)$ is continuous of $(t,s) \in [a,b]^2$. First, we start taking a partition Δ_n of $[a,b]$ and defining a process $\{f_t^n(\omega)\}_{n\geq 1} \subseteq \mathcal{E}$ as:

$$f_t^n(\omega) = f_{t_{i-1}}(\omega), \quad t_{i-1} < t \leq t_i. \tag{2.3}$$

Due to the continuity of $E(f_t f_s)$, we have $\lim_{s\to t} E(|f_t - f_s|^2) = 0$, and consequently,

$$\lim_{n\to\infty} E\left(|f_t - f_t^n|^2\right) = 0, \quad \forall t \in [a,b]. \tag{2.4}$$

On the other hand, using the inequality $|\alpha - \beta|^2 \leq 2(|\alpha|^2 + |\beta|^2)$, it holds:

$$E\left(|f_t - f_t^n|^2\right) \leq 2\left\{E\left(|f_t|^2\right) + E\left(|f_t^n|^2\right)\right\} \leq 4 \sup_{t\in[a,b]} E\left(|f_t|^2\right).$$

Therefore, we can apply the *dominated convergence theorem*[1] to Equation (2.4), concluding that

$$\lim_{n\to\infty} \int_a^b E\left(|f_t - f_t^n|^2\right) dt = 0.$$

Case 2: f is bounded. Then, we define:

$$g_t^n(\omega) = \int_0^{n(t-a)} e^{-\tau} f_{t-n^{-1}\tau}(\omega)\, d\tau.$$

Note that $g^n \in L^2_{ad}([a,b] \times \Omega)$, $\forall n$. Next step consists of justifying

$$\int_a^b E\left(|f_t - g_t^n|^2\right) dt \to 0, \quad \text{as } n \to \infty. \tag{2.5}$$

[1]This important theorem on measure theory can be read in the appendices, subsection A.2.

Indeed, since

$$f_t - g_t^n = \int_0^\infty e^{-\tau}\left(f_t - f_{t-n^{-1}\tau}\right) d\tau \Rightarrow |f_t - g_t^n|^2 \leq \int_0^\infty e^{-\tau}\left|f_t - f_{t-n^{-1}\tau}\right|^2 d\tau,$$

we can compute:

$$\begin{aligned}\int_a^b E\left(|f_t - g_t^n|^2\right) dt &\leq \int_a^b \int_0^\infty e^{-\tau} E\left(\left|f_t - f_{t-n^{-1}\tau}\right|^2\right) d\tau dt \\ &= \int_0^\infty e^{-\tau}\left(\int_a^b E\left(\left|f_t - f_{t-n^{-1}\tau}\right|^2\right) dt\right) d\tau \\ &= \int_0^\infty e^{-\tau} E\left(\int_a^b \left|f_t - f_{t-n^{-1}\tau}\right|^2 dt\right) d\tau.\end{aligned}$$

However, as f is assumed to be bounded, it yields a.s.

$$\int_a^b \left|f_t - f_{t-n^{-1}\tau}\right|^2 dt \to 0, \quad \text{as } n \to \infty, \tag{2.6}$$

which proves the desired statement in Equation (2.5). Moreover, $E\left(g_t^n g_s^n\right)$ is continuous (it is easily provable by rewriting $g_t^n = \int_a^t ne^{-n(t-u)} f_u(\omega)\, du$, with $u = t - \tau n^{-1}$, to verify that $\lim_{t\to s} E(|g_t^n - g_s^n|^2) = 0$). Therefore, we can apply *Case 1* to approximate each g^n by an adapted step process f^n such that:

$$\lim_{n\to\infty} \int_a^b E\left(|g_t^n - f_t^n|^2\right) dt \leq \frac{1}{n},$$

and using the claim in Equation (2.5), it leads to the requested formula.
General case: Having fixed $f \in L^2_{ad}([a,b] \times \Omega)$ we define for each n:

$$g_t^n(\omega) = \begin{cases} f_t(\omega) & \text{if } |f_t(\omega)| \leq n. \\ 0 & \text{if } |f_t(\omega)| > n. \end{cases}$$

Then, the *dominated converge theorem* implies that

$$\int_a^b E\left(|f_t - g_t^n|^2\right) dt \to 0, \text{ as } n \to \infty. \tag{2.7}$$

Finally, for each n we recover *Case 2* to pick an adapted step process $f_t^n(\omega)$ such that

$$\int_a^b E\left(|g_t^n - f_t^n|^2\right) dt \leq \frac{1}{n} \tag{2.8}$$

and the Equation (2.2) of the theorem follows from Equations (2.7) and (2.8). □

Once we have established the existence of such sequences of step processes, we have the necessary tools to give a more general definition by making limits in mean square.

Itô's integral for processes belonging to $L^2_{ad}([a,b]\times\Omega)$

Definition 2.8. For any process $f \in L^2_{ad}([a,b]\times\Omega)$, we define its *Itô's stochastic integral* $I(f)$ as

$$I(f) = \lim_{n\to\infty} I(f^n), \quad \text{in } L^2(\Omega). \tag{2.9}$$

In the same vein, we will denote $I(f) = \int_a^b f_t dW_t$.

Notice that to see if this definition makes sense, we should verify that the limit exists and this does not depend on the approximating sequence. The first claim is a consequence that our $f_t^n \subseteq \mathcal{E}$ is a Cauchy sequence in $L^2(\Omega)$ for all $t \in [a,b]$, which is a complete space,

$$E\left(\left|\int_a^b f_t^n dW_t - \int_a^b f_t^m dW_t\right|^2\right) = E\left(\int_a^b (f_t^n - f_t^m)^2 dt\right) \leq 2E\left(\int_a^b (f_t^n - f_t)^2 dt\right) + 2E\left(\int_a^b (f_t - f_t^m)^2 dt\right) \to 0, \quad \text{as } n,m \to 0.$$

On the other hand, to prove that the definition for a stochastic process f_t approximated by two difference sequences f_t^n and f_t^m coincide, we use the triangular inequality to compute:

$$\|I^n(f) - I^m(f)\|_{L^2} \leq \left\|I^n(f) - \int_a^b f_t^n dW_t\right\|_{L^2} + \left\|\int_a^b f_t^n dW_t - \int_a^b f_t^m dW_t\right\|_{L^2} + \left\|I^m(f) - \int_a^b f_t^m dW_t\right\|_{L^2}, \quad \forall t \in [a,b].$$

As the right-hand side of the inequality tends to zero as $n \to \infty$, the left-hand side also vanishes. This following lemma is a kind of extension of Proposition 2.6, not only considering processes in $\mathcal{E}$.

Lemma 2.9. For any $f \in L^2_{ad}([a,b]\times\Omega)$, it holds:

1. $I(af + bg) = aI(f) + bI(g), \quad \forall a,b \in \mathbb{R},\ g \in L^2_{ad}([a,b]\times\Omega)$.
2. $E(I(f)) = 0$, which means that stochastic integrals are centered random variables.
3. $E\left(|I(f)|^2\right) = \int_a^b E\left(f_t^2\right) dt$.

Proof. The fact that $I(\cdot)$ is a linear operator is a trivial consequence of the linearity for step processes. Zero mean value derives from:

$$\lim_{n\to\infty} E(I(f^n)) = E(I(f))$$

and $E(I(f^n)) = 0$, $\forall n \geq 1$. And finally, to see the isometry we use norm notation,

$$\|I(f)\|_{L^2} = \lim_{n\to\infty} \|I(f^n)\|_{L^2} = \lim_{n\to\infty} \|f^n\|_{L^2} = \|f\|_{L^2}$$

which concludes the proof. □

Remark 2.10. One can actually extend this construction to a larger set of functions consisting of all processes whose sample paths are functions in $L^2[a,b]$ (i.e. $\int_a^b |f_t|^2 dt < \infty$ a.s.). Although we omit here this tedious development, it can be found in [4].

To complete this section, we discuss a theorem that establishes a direct relation between Riemann sums and the definition of the stochastic integral given in this section.

Theorem 2.11. For any $f \in L^2_{ad}([a,b] \times \Omega)$ such that $E(f_t f_s)$ is continuous of t and s, it holds in $L^2(\Omega)$,

$$\int_a^b f_t dW_t = \lim_{\|\Delta_n\| \to 0} \sum_{i=1}^n f_{t_{i-1}} \Delta_i W,$$

where $\Delta_n = \{a = t_0 < t_1 < \ldots < t_n = b\}$ and $\|\Delta_n\| = \sup_{1 \le i \le n} \Delta_i t$.

Proof. We saw in the proof of Theorem 2.7 that defining a stochastic process as in Equation (2.3), we have

$$\lim_{n \to \infty} \int_a^b E\left(|f_t - f_t^n|^2\right) dt = 0.$$

Therefore, using Equation (2.9) and Definition 2.5 given for step processes, we get in $L^2(\Omega)$ that

$$\int_a^b f_t dW_t = \lim_{n \to \infty} I(f^n) = \lim_{n \to \infty} \sum_{i=1}^n f^n_{t_{i-1}} \Delta_i W,$$

thus we have shown that in the limit, this sequence of sums converges to Itô's integral.

□

Once we have justified this picture of the Itô's stochastic integral, we can regain the discussion on Example 2.2 to see the consistency of this approach.

2.2.2 Stochastic processes driven by Itô's integrals

Having introduced this new integral as Itô did, it involves many interesting properties we can take advantage of in stochastic calculus. Next step consists of studying indefinite Itô's stochastic integral for processes $f \in L^2_{ad}([a,b] \times \Omega)$. This is naturally defined by $X_t = \int_a^t f_s dW_s := \int_a^b f_s \mathbb{1}_{[a,t]}(s)\, dW_s$. Note first that this makes sense because

$$\int_a^t E\left(|f_s|^2\right) ds \le \int_a^b E\left(|f_s|^2\right) < \infty,$$

so $f \in L^2_{ad}([a,t] \times \Omega)$. In addition, such a stochastic process in Itô's sense is defined so that it achieves a crucial property, which is proved in Proposition 2.13. Before doing that, we first see a classical lemma of convergence of conditional random variables necessary for its demonstration.

Lemma 2.12. Let $\{X_n\}_n$ be a sequence of square integrable random variables on some probability space $(\Omega, \mathscr{F}, P)$. If $X_n \to X$ as $n \to \infty$ in $L^2(\Omega)$, then

$$E(X_n|\mathscr{G}) \to E(X|\mathscr{G}), \quad \text{as } n \to \infty,$$

for any σ-field $\mathscr{G} \subseteq \mathscr{F}$. In particular, $L^2(\Omega)$-limits of martingales are also martingales.

Proof. Consider the expression

$$|E(X|\mathscr{G}) - E(X|\mathscr{G})|^2 = |E(X_n - X|\mathscr{G})|^2 .$$

As $|\cdot|^2$ is a convex function, we can apply *Jensen's inequality*[2] to get

$$\begin{aligned} |E(X_n - X|\mathscr{G})|^2 \leq E\left(|X_n - X|^2 |\mathscr{G}\right) \Rightarrow E\left(|E(X_n|\mathscr{G}) - E(X|\mathscr{G})|^2\right) \\ \leq E\left(E\left(|X_n - X|^2 |\mathscr{G}\right)\right) = E\left(|X_n - X|^2\right) \to 0, \quad \text{as } n \to \infty, \end{aligned}$$

having used the property of conditional expectation which says that $E(E(X|\mathscr{G})) = E(X)$ for any X integrable random variable and $\mathscr{G}$ a σ-field contained in $\mathscr{F}$. □

Proposition 2.13. The process defined by $X_t = \int_a^t f_s dW_s$, with $t \in [a, b]$ is a martingale.

Proof. First of all, due to the previus lemma, proving the proposition is equal to see the result for any approximating sequence in $\mathcal{E}$,

$$X_t^n = \int_a^t f_s^n dW_s, \quad \forall t \in [a, b],$$

where f^n converges to f in $L^2(\Omega \times [a, b])$ and $f^n \in \mathcal{E}$. Fixing $a \leq s \leq t \leq b$ and assuming that $t_{k-1} < s \leq t_k < t_l < t \leq t_{l+1}$ we will have

$$X_t^n - X_s^n = u_k\left(W_{t_k} - W_s\right) + \sum_{j=k+1}^{l} u_j \Delta_j W + u_{l+1}\left(W_t - W_{t_l}\right).$$

Notice that due to Lemma 2.9

$$E\left(|X_t|^2\right) = \int_a^t E\left(|f_s|^2\right) ds < \infty \Rightarrow E(X_t) \leq \left[E\left(|X_t|^2\right)\right]^{1/2} < \infty,$$

which means that X_t is integrable and thus we can take the conditional expectation of X_t with respect to the filtration $\mathscr{F}_s$. Managing its terms[3] we get

$$\begin{aligned} E(X_t^n - X_s^n|\mathscr{F}_s) = E\left(u_k\left(W_{t_k} - W_s\right)|\mathscr{F}_s\right) + \sum_{j=k+1}^{l} E\left(E\left(u_j \Delta_j W|\mathscr{F}_{t_{j-1}}\right)|\mathscr{F}_s\right) \\ + E\left(u_{l+1} E\left(W_t - W_{t_l}|\mathscr{F}_s\right)\right) = 0, \end{aligned}$$

and so $E(X_t^n|\mathscr{F}_s) = X_s^n$ a.s., proving that X_t^n is a martingale. □

[2]To remind this result, consult the appendices, subsection A.1.

[3]More precisely, we use properties (a), (e), (f) of the conditional expectation, gathered in the appendices, subsection A.1.

One can also wonder if the stochastic process X_t defined by the indefinite Itô's integral satisfies a basic property as it is continuity. Although it could seem trivial by simple extension with respect to deterministic results, the problem here is that the stochastic integral is not defined path-wise as a Riemman(-Stieljes) integral or even Lebesgue integral. For that reason, although there exists a result confirming that X_t is continuous, namely, almost all of its paths are continuous functions on the interval $[a, b]$, its proof is not immediate at all. For more details, we recommend consulting [4].

2.2.3 Itô formula

One of the most used identities in Leibniz-Newton calculus is the well-known chain rule, which gives an easy expression for the derivative of a composite function $f(g(t))$. It states that if f and g are differentiable functions, then $f(g(t))$ is also differentiable with derivative:

$$\frac{d}{dt}f(g(t)) = f'(g(t))g'(t).$$

Or equally, using the fundamental theorem of calculus we can write

$$f(g(t)) - f(g(a)) = \int_a^t f'(g(s))g'(s)ds.$$

We can now question on the existence of such expression for Itô's stochastic calculus. Though we know that the equality $\frac{d}{dt}f(W_t) = f'(W_t)W_t'$ has no meaning since the sample paths of Brownian motion are nowhere differentiable, it could be true that $f(W_t) - f(W_a) = \int_a^t f'(W_s)dW_s$, being f a differentiable function. Nevertheless, due to the nonzero quadratic variation of Brownian motion, this formula is slightly modified with an extra term, as it shows next theorem.

Theorem 2.14. *(Itô formula, simplified version)* Let $f : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^2$-function. Then, for any $a \leq t \leq b$,

$$f(W_t) - f(W_a) = \int_a^t f'(W_s)dW_s + \frac{1}{2}\int_a^t f''(W_s)ds. \tag{2.10}$$

Observe that the first integral is an indefinite Itô's integral (as defined in subsection 2.2.2), but the second one is just a common Riemann integral for each sample path of W_s. The proof is a consequence of two technical lemmas. To discuss them, we consider as usual a partition $\Delta_n = \{a = t_0,\ t_1, \ldots, t_n = t\}$ of $[a, t]$ such that $\lim_{n\to\infty} \|\Delta_n\| = 0$.

Lemma 2.15. For any real continuous function g and $\lambda_i \in (0, 1)$, $1 \leq i \leq n$, there exists a subsequence, denoted by (n), such that

$$X_n := \sum_{i=1}^{n} \left(g\left(W_{t_{i-1}} + \lambda_i \Delta_i W\right) - g\left(W_{t_{i-1}}\right)\right)(\Delta_i W)^2$$

converges to 0 in probability when $n \to \infty$.

Proof. Define the random variable

$$Y_n = \max_{1\leq i\leq n,\, 0<\lambda<1} \left| g\left(W_{t_{i-1}} + \lambda \Delta_i W\right) - g\left(W_{t_{i-1}}\right)\right|.$$

Then, obviously it holds

$$|X_n| \leq Y_n \sum_{i=1}^{n} \left(\Delta_i W\right)^2.$$

and owing to the continuity of $g(x)$ and the Brownian motion, $\{Y_n\}_{n\geq 1}$ converges to zero a.s. On the other hand, by Proposition 1.17, $\sum_{i=1}^{n} (\Delta_i W)^2$ converges to $t-a$ in $L^2(\Omega)$. Therefore, X_n converges to zero in probability. □

Lemma 2.16. Under the same hypotheses in Lemma 2.15, the sequence

$$\sum_{i=1}^{n} g\left(W_{t_{i-1}}\right)\left(\left(\Delta_i W\right)^2 - \Delta_i t\right) \tag{2.11}$$

converges to zero in probability.

Proof. For each $L > 0$, we define the set of events

$$A_{i-1}^{(L)} \equiv \left\{ \left|W_{t_j}\right| \leq L, \quad \forall j \leq i-1 \right\}, \quad 1 \leq i \leq n.$$

Note that this is a decreasing family in i. Moreover, let S_n denote the summation in Equation (2.11) and label

$$S_{n,L} \equiv \sum_{i=1}^{n} g(W_{t_{i-1}})\mathbb{1}_{A_{i-1}^{(L)}}\left((\Delta_i W)^2 - \Delta_i t\right) = \sum_{i=1}^{n} g(W_{t_{i-1}})\mathbb{1}_{A_{i-1}^{(L)}} X_i = \sum_{i=1}^{n} Y_i, \tag{2.12}$$

where $X_i = (\Delta_i W)^2 - \Delta_i t$ and $Y_i = g(W_{t_{i-1}})\mathbb{1}_{A_{i-1}^{(L)}} X_i$. Our goal is to see that $E\left(|\sum_i Y_i|^2\right) \to 0$ as $n \to \infty$. Consider now the filtration $\mathscr{F}_t = \sigma\{W_s;\ s \leq t\}$ and fix $1 \leq i < j \leq n$. Hence, using conditional expectation[4]:

$$E\left(Y_i Y_j\right) = E\left(E\left(Y_i Y_j | \mathscr{F}_{t_{j-1}}\right)\right) = E\left(Y_i g(W_{t_{j-1}})\mathbb{1}_{A_{j-1}^{(L)}} E\left(X_j | \mathscr{F}_{t_{j-1}}\right)\right) = 0. \tag{2.13}$$

Additionally, we clearly have $Y_i^2 \leq \sup_{|x|\leq L}\left(|g(x)|^2\right) X_i^2$, and recovering Equation (1.1) in Proposition 1.17 we get

$$E\left(Y_i^2\right) \leq 2\left(\Delta_i t\right)^2 \sup_{|x|\leq L}\left(|g(x)|^2\right). \tag{2.14}$$

Consequently, from Equations (2.12), (2.13) and (2.14) we observe that

$$E\left(S_{n,L}^2\right) = \sum_{i=1}^{n} E\left(Y_i^2\right) \leq 2 \sup_{|x|\leq L}\left(|g(x)|^2\right) \sum_{i=1}^{n} (\Delta_i t)^2 \leq 2\|\Delta_n\|\,(t-a) \sup_{|x|\leq L}\left(|g(x)|^2\right) \to 0,$$

[4]The actual properties required here are (c) and (d) from the list in the appendices, subsection A.1

as $\|\Delta_n\| \to 0$. Next step consists of realizing that due to the definition of $A_{i-1}^{(L)}$ and $S_{n,L}$ it holds true that $A_{n-1}^{(L)} \subset \{S_n = S_{n,L}\}$, so

$$\{S_n \neq S_{n,L}\} \subset \left(A_{n-1}^{(L)}\right)^C \subset \left\{\sup_{s\in[a,t]} |W_s| > L\right\} \Rightarrow P\{S_n \neq S_{n,L}\} \leq P\left\{\sup_{s\in[a,t]} |W_s| > L\right\}.$$

However, Doob's inequality in Theorem 1.7 yields

$$P\left\{\sup_{s\in[a,t]} |W_s| > L\right\} \leq \frac{1}{L} E|W_s| = \frac{1}{L}\int_0^\infty \frac{2x}{\sqrt{2\pi t}} exp\left(\frac{-x^2}{2t}\right) dx = \frac{1}{L}\sqrt{\frac{2t}{\pi}}.$$

and therefore, $\forall n \geq 1$ we have $P\{S_n \neq S_{n,L}\} \leq \frac{1}{L}\sqrt{\frac{2t}{\pi}}$. Eventually, observe that $\forall \epsilon > 0$,

$$\{|S_n| > \epsilon\} \subset \{|S_{n,L}| > \epsilon\} \cup \{S_n \neq S_{n,L}\} \Rightarrow P\{|S_n| > \epsilon\} \leq P\{|S_{n,L}| > \epsilon\} + \frac{1}{L}\sqrt{\frac{2t}{\pi}},$$

so choosing L large enough so that $\frac{1}{L}\sqrt{\frac{2t}{\pi}} < \frac{\epsilon}{2}$, we have also proved that $S_{n,L} \to 0$ in probability as $\|\Delta_n\| \to 0$, thus there is $n_0 \geq 1$ satisfying $P\{|S_{n,L}| > \epsilon\} < \frac{\epsilon}{2}$, $\forall n \geq n_0$. In conclusion, $P\{|S_n| > \epsilon\} < \epsilon$, $\forall n \geq n_0$ which means that S_n converges to 0 in probability as $\|\Delta_n\| \to 0$. □

Proof. Theorem 2.14. To begin, we can write $f(W_t) - f(W_a) = \sum_{i=1}^n \left(f(W_{t_i}) - f(W_{t_{i-1}})\right)$. Since $f \in \mathcal{C}^2$, we compute its Taylor expansion up to the second order with the remainder in Lagrange form, obtaining

$$f(W_t) - f(W_a) = \sum_{i=1}^n f'(W_{t_{i-1}})\Delta_i W + \frac{1}{2}\sum_{i=1}^n f''(W_{t_{i-1}} + \lambda_i \Delta_i W)\,(\Delta_i W)^2,$$

where $\lambda_i \in (0,1)$. On the one hand, the first term in the summation converges in probability to $\int_a^t f'(W_s)dW_s$, as $\|\Delta_n\| \to 0$ through Theorem 2.11. On the the other, we rewrite the second expression as

$$\sum_{i=1}^n f''(W_{t_{i-1}} + \lambda_i \Delta_i W)\,(\Delta_i W)^2 = \sum_{i=1}^n \left(f''(W_{t_{i-1}} + \lambda_i \Delta_i W) - f''(W_{t_{i-1}})\right)(\Delta_i W)^2 \\ + \sum_{i=1}^n f''(W_{t_{i-1}})\left((\Delta_i W)^2 - \Delta_i t\right) + \sum_{i=1}^n f''(W_{t_{i-1}})\Delta_i t.$$

By Lemmas 2.15 and 2.16 (with f'' instead of g), the two first addends in the right-hand side converge to zero as $n \to \infty$, and lastly, the third term converges to $\int_a^t f''(W_s)ds$ by the classical theorem on approximation of Riemann integrals by sums. □

This formula, which captures the anomalies of stochastic calculus, can be generalized to deal with a greater set of functions. The first thought should be considering functions $f_t(x) \equiv f(t,x)$ with continuous derivatives $\partial_t f$, $\partial_x f$ and $\partial_x^2 f$. Then, a similar argument by means of Taylor expansions would prove that the corresponding formula is

$$f_t(W_t) = f_a(W_a) + \int_a^t \frac{\partial f_s}{\partial x}(W_s)dW_s + \int_a^t \left(\frac{\partial f_s}{\partial t}(W_s) + \frac{1}{2}\frac{\partial^2 f_s}{\partial x^2}(W_s)\right) ds. \tag{2.15}$$

However, there exists a more general version of Itô formula. Examining Equations (2.10) and (2.15), we realize that both expressions mix Itô's integrals together with Riemann ones, which suggests next definition.

Definition 2.17. An Itô process is a stochastic process of the form

$$X_t = X_a + \int_a^t f_s dW_s + \int_a^t g_s ds, \quad a \le t \le b, \tag{2.16}$$

where X_a is $\mathscr{F}_a$-measurable, and f_t and g_t are $\mathscr{F}_t$-adapted processes such that $\int_a^b |f_t|^2 dt < \infty$ and $\int_a^b |g_t| dt < \infty$.

Alternately, we can also write it in stochastic differential form: $dX_t = f_t dW_t + g_t dt$. Nevertheless, we can not forget that it is just a convenient symbolic shorthand with no mathematical meaning by itself because Brownian sample paths are nowhere differentiable. Note that it will also be defined the Itô's integral of an Itô process. To check it, we compute

$$|X_t|^2 \le 3\left[|X_a|^2 + \left(\int_a^t f_s dW_s\right)^2 + \left(\int_a^b g_s ds\right)^2\right],$$

where we used the identity $(a+b+c)^2 \le 3(|a|^2 + |b|^2 + |c|^2)$. Then, taking the expectation and using the isometry property,

$$E|X_t|^2 \le 3\left[E|X_a|^2 + \int_a^t E\left(f_s^2\right) ds + (b-a)\left(\int_a^b E\left(g_s^2\right) ds\right)\right] \le \infty,$$

so $X_t \in L^2_{ad}([a,b] \times \Omega)$. Next theorem extends the previous Itô formula to this special class of stochastic processes.

Theorem 2.18. *(Itô formula, general form)* Let $\theta : [a,b] \times \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^{1,2}$ - function and X_t an Itô process as in Equation (2.16). Then, $\theta_t(X_t)$ is also an Itô process and it holds:

$$\theta_t(X_t) = \theta_a(X_a) + \int_a^t \frac{\partial \theta_s}{\partial x}(X_s) f_s dW_s + \int_a^t \left[\frac{\partial \theta_s}{\partial s}(X_s) + \frac{\partial \theta_s}{\partial x}(X_s) g_s + \frac{1}{2}\frac{\partial^2 \theta_s}{\partial x^2}(X_s) f_s^2\right] ds \tag{2.17}$$

Although the rigorous proof of the theorem requires long developments, we can derive symbolically Equation (2.17) using Taylor series and Table 1.1 as a hint for the differentials. Beginning by the Taylor expansion,

$$d\theta_t(X_t) = \frac{\partial \theta_t}{\partial t}(X_t) dt + \frac{\partial \theta_t}{\partial x}(X_t) dX_t + \frac{1}{2}\frac{\partial^2 \theta_t}{\partial x^2}(X_t)(dX_t)^2,$$

we can then apply the rules for the differentials in Itô's table to see that $(dX_t)^2 = f_t^2 dt$ and consequently,

$$\begin{aligned} d\theta_t(X_t) &= \frac{\partial \theta_t}{\partial t}(X_t)dt + \frac{\partial \theta_t}{\partial x}(X_t) \cdot (f_t dW_t + g_t dt) + \frac{1}{2}\frac{\partial^2 \theta_t}{\partial x^2}(X_t) f_t^2 dt \\ &= \frac{\partial \theta_t}{\partial x}(X_t) f_t dW_t + \left(\frac{\partial \theta_t}{\partial t}(X_t) + \frac{\partial \theta_t}{\partial x}(X_t) g_t + \frac{1}{2}\frac{\partial^2 \theta_t}{\partial x^2}(X_t) f_t^2\right) dt, \end{aligned}$$

which can be easily converted to the integral form in Equation (2.17).

Example 2.19. We want to apply Itô formula to the process $Y_t = e^{cW_t - \frac{c^2}{2}t}$, with $c \in \mathbb{R}$ and $t \in [a, b]$. Note that now, $X_t = W_t$ and then $f_t = 1$, $g_t = 0$ in the above expressions. Deriving properly, we obtain

$$Y_t = Y_a + c \int_a^t Y_s dW_s.$$

This means that the solution for the stochastic differential equation $dY_t = cY_t dW_t$ is $Y_t = Y_a e^{cW_t - \frac{c^2}{2}t}$ instead of the expected $Y_t = Y_a e^{cW_t}$.

2.3 Stratonovich's integral

In this section, we introduce a different way to define a stochastic integral based on the article of R. L. Stratonovich [11]. Although he developed the integral for a wider class of stochastic process (the so-called *diffusive Markov processes*), we will only particularize functions $f_t(W_t)$ in order to follow the outline from previous sections. Then, we take advantage of the conversion formula to generalize this definition for any Itô processes and after, we discuss some of the properties we can derive from this alternative description of the stochastic integral.

2.3.1 Definition and conversion formula

Here, the procedure to define Stratonovich's integral begins with its approximation by sums, as a difference with Itô's integral that we did it after all the construction. Likewise, we will use as always a partition Δ_n of the interval $[a, b]$ where is defined the time.

Definition 2.20. Let $f_t(x) \equiv f(t, x) : [a, b] \times \mathbb{R} \to \mathbb{R}$ a $\mathcal{C}^{1,1}$-function satisfying

$$\int_a^b E\left[f_t(W_t)\right]^2 dt < \infty \quad \text{and} \quad \int_a^b E\left[\partial_x f_t(W_t)\right]^2 dt < \infty,$$

Then, the *Stratonovich's integral of f_t with respect to the Brownian motion* is defined as

$$\int_a^b f_t(W_t) \circ dW_t = \lim_{\|\Delta_n\| \to 0} \sum_{i=1}^n f_{t_{i-1}} \left(\frac{W_{t_i} + W_{t_{i-1}}}{2} \right) \Delta_i W. \tag{2.18}$$

To see if the previous definition makes sense, we shall prove that the limit exists. To do it, we consider the difference between this formula and the integral in the sense of Itô (whose interpretation is discussed in Section 2.2):

$$D_\Delta = \sum_{i=1}^n \left[f_{t_{i-1}} \left(\frac{W_{t_i} + W_{t_{i-1}}}{2} \right) - f_{t_{i-1}}(W_{t_{i-1}}) \right] \Delta_i W,$$

Hence, if we see that $\lim_{\|\Delta_n\| \to 0} D_\Delta$ exists, we will have also shown the consistency of Equation (2.18). Making use of the differentiability with respect to x, we apply the Mean Value Theorem (MVT) to f_t, obtaining the expression

$$D_\Delta = \frac{1}{2} \sum_{i=1}^n \frac{\partial f_{t_{i-1}}}{\partial x} \left(\theta W_{t_i} + (1 - \theta) W_{t_{i-1}} \right) (\Delta_i W)^2,$$

being $0 \leq \theta \leq \frac{1}{2}$ a fixed parameter. Although it is not difficult to intuit that the latter tends in probability to the integral

$$\frac{1}{2}\int_a^b \frac{\partial f_t}{\partial x}(W_t)dt,$$

let us justify it rigorously. Remember that we have been picking points from the original Δ_n-partition of $[a,b]$. We now define

$$\overline{D}_{\Delta\epsilon} = \frac{1}{2}\sum_{i=1}^{n} \overline{f}^{\epsilon}_{t^{(\Delta)}_{i-1}}(W_t)\left(\Delta^{(\Delta)}_i W\right)^2 \quad \text{with} \quad \overline{f}^{\epsilon}_t(W_t) = \sup_{t\in\left[t^{(\epsilon)}_{j-1},\, t^{(\epsilon)}_j\right)} \left\{\frac{\partial f}{\partial x}(W_t)\right\},$$

$$\underline{D}_{\Delta\epsilon} = \frac{1}{2}\sum_{i=1}^{n} \underline{f}^{\epsilon}_{t^{(\Delta)}_{i-1}}(W_t)\left(\Delta^{(\Delta)}_i W\right)^2 \quad \text{with} \quad \underline{f}^{\epsilon}_t(W_t) = \inf_{t\in\left[t^{(\epsilon)}_{j-1},\, t^{(\epsilon)}_j\right)} \left\{\frac{\partial f}{\partial x}(W_t)\right\},$$

having taken an ϵ-partition of the interval satisfying $\{t^{(\epsilon)}_j,\ j = 0,1,\ldots,m\} \subset \{t^{(\Delta)}_i,\ i = 0,1,\ldots,n\}$. Note that each evaluation time is carefully indicated to belong to our ϵ or Δ partition with the superscript. By construction, it holds that $\underline{D}_{\Delta\epsilon} \leq D_\Delta \leq \overline{D}_{\Delta\epsilon}$. Moreover, we can use again the quadratic variation of the Wiener process (Proposition 1.17) to write

$$\lim_{\|\Delta_n\|\to 0} \sum_{i\in[j-1,\, j)} \left(\Delta^{(\Delta)}_i W\right)^2 = \int_{t^{(\epsilon)}_{j-1}}^{t^{(\epsilon)}_j} dt,$$

and consequently,

$$\lim_{\|\Delta_n\|\to 0} \overline{D}_{\Delta\epsilon} = \frac{1}{2}\int_a^b \overline{f}^{\epsilon}_t(W_t)dt \equiv \overline{D}_\epsilon,$$

$$\lim_{\|\Delta_n\|\to 0} \underline{D}_{\Delta\epsilon} = \frac{1}{2}\int_a^b \underline{f}^{\epsilon}_t(W_t)dt \equiv \underline{D}_\epsilon.$$

However, owing to the continuity of $\partial_x f_t\left(W_t\right)$ we can make the differences $\overline{f}^{\epsilon} - \underline{f}^{\epsilon}$ and $\overline{D}_\epsilon - \underline{D}_\epsilon$ as small as desired if we increase ϵ. Therefore, there follows the existence with probability 1 of the limit,

$$\lim_{\|\Delta_n\|\to 0} D_\Delta = \lim_{\|\epsilon_m\|\to 0} \overline{D}_\epsilon = \lim_{\|\epsilon_m\|\to 0} \underline{D}_\epsilon = \frac{1}{2}\int_a^b \frac{\partial f_t}{\partial x}(W_t)dt.$$

Then, under these assumptions, the integral in Equation (2.18) exists and is related to Itô's integral by the formula

$$\int_a^b f_t(W_t) \circ dW_t = \int_a^b f_t(W_t)dW_t + \frac{1}{2}\int_a^b \frac{\partial f_t}{\partial x}(W_t)dt. \tag{2.19}$$

Equally, we can also write last expression in stochastic differential form as $f_t(W_t) \circ dW_t = f_t dW_t + \frac{1}{2}\partial_x f_t(W_t)dt$. We now understand easily why we required some special conditions to the function f_t: we must ensure the existence of Itô's integral but also the convergence of the second addend. This formula will play a relevant role since it connects both views (Itô and Stratonovich) with a term understood as drift. As we will discuss, there is no a better approach, but each formulation is beneficial depending on situation and the properties we are requiring. For this reason, it is important to keep in mind that last expression is the key to implicate both insights and take profit of each of them.

Although last definition is the usual choice because it makes its justification computationally easier, there is an alternative way to give meaning to the Stratonovich's integral consisting of the following:

$$\int_a^b f_t(W_t) \circ dW_t = \lim_{\|\Delta_n\| \to 0} \sum_{i=1}^{n} f_{t_{i-1}} \left(W_{\frac{t_i+t_{i-1}}{2}} \right) \Delta_i W.$$

Notice that this approach recovers the interpretation given in Example 2.2 in which we stated how the main discrepancy between these two views is the evaluation point of the Brownian motion. Anyway, let us reflect that both definitions of Stratonovich's integrals coincide with a simple sketch (the rigorous steps are analogous to the previous reasoning). We also need to make the difference with Itô's integral, so inside the sum we will get

$$\left[f_{t_{i-1}}(W_{\frac{t_i+t_{i-1}}{2}}) - f_{t_{i-1}}\left(W_{t_{i-1}}\right) \right] \Delta_i W = \frac{\partial f_{t_{i-1}}}{\partial x} \left(W^*_{t_i} \right) \left(W_{\frac{t_i+t_{i-1}}{2}} - W_{t_{i-1}} \right) \Delta_i W$$

$$= \frac{\partial f_{t_{i-1}}}{\partial x} \left(W^*_{t_i} \right) \left(W_{\frac{t_i+t_{i-1}}{2}} - W_{t_{i-1}} \right)^2 + \frac{\partial f_{t_{i-1}}}{\partial x} \left(W^*_{t_i} \right) \left(W_{\frac{t_i+t_{i-1}}{2}} - W_{t_{i-1}} \right) \left(W_{t_i} - W_{\frac{t_i+t_{i-1}}{2}} \right)$$

where we used once more the MVT (taking $W^*_{t_i} = \theta W_{\frac{t_i+t_{i-1}}{2}} + (1-\theta) W_{t_{i-1}}$ for some fixed parameter $0 \le \theta \le 1$), and rewrote the expression conveniently. Nonetheless, the second term tends in probability to 0 due to the property of stationary increments of W_t and the first term has limit $\frac{1}{2}\partial_x f(W^*_{t_i})\Delta_i t$, which is coherent with the other definition.

Remark 2.21. Notice from Equation (2.19) that in general, for any f_t satisfying the above conditions,

$$\int_a^b f_t(W_t) dW_t \neq \int_a^b f_t(W_t) \circ dW_t,$$

so the properties we proved in Itô's case do not extend for Stratonovich's integral. Note also that the equality holds when f_t does not depend on the Brownian. For this kind of simpler processes, the integral carries the name of *Wiener integral*.

Although the way to understand the original idea of the Russian probabilist follows this previous reasoning, we now introduce a more general definition to deal with a greater set of stochastic processes.

Definition 2.22. For any Itô processes X_t, Y_t, we define the *Stratonovich's integral of X_t with respect to Y_t* as

$$\int_a^b X_t \circ dY_t = \int_a^b X_t dY_t + \frac{1}{2} \int_a^b (dX_t)\,(dY_t). \tag{2.20}$$

Remark 2.23. When $dY_t = dW_t$ and X_t is written as $X_t = X_a + \int_a^t f^X_s dW_s + \int_a^t g^X_s ds$ for convenient functions f^X_t, g^X_t, the expression is simplified using Itô's table,

$$\int_a^b X_t \circ dW_t = \int_a^b X_t dW_t + \frac{1}{2} \int_a^b f^X_t dt. \tag{2.21}$$

Particularly, we recover the construction developed before. More precisely, if $X_t = f_t(W_t)$, assuming regularity in the derivatives of f_t and applying Itô formula we have $dX_t = \partial_x f_t(W_t)\, dW_t + [\partial_t f_t(W_t) + \frac{1}{2}\partial_x^2 f_t(W_t)]dt$. And this, supported by Table 1.1 gives $f_t^X = \partial_x f(W_t)$, which is exactly Equation (2.19). On the other hand, it will be interesting for next chapter considering the Stratonovich's representation of processes of the form $\theta_t(X_t)$, where X_t is an Itô process written as above and θ is a real $\mathcal{C}^{1,2}$-function defined on $[a,b] \times \mathbb{R}$. By Theorem 2.18, $\theta_t(X_t)$ is another Itô process, so we can apply Definition 2.20 to reach

$$\int_a^b \theta_t(X_t) \circ dW_t = \int_a^b \theta_t(X_t)\, dW_t + \frac{1}{2}\int_a^b \frac{\partial \theta_t}{\partial x}(X_t)\, f_t dt. \tag{2.22}$$

As we will see in Section 3.4, this will be the main tool to change from Stratonovich's stochastic differential equation perspective to Itô's one and obtain a direct formula between them.

Moreover, this general definition satisfies some good properties, such as the fact that the collection of Itô processes is closed under making its Stratonovich's integral, as it asserts next theorem.

Theorem 2.24. For any X_t, Y_t Itô process, the resulting stochastic process of making its Stratonovich integral,

$$Z_t = \int_a^t X_s \circ dY_s, \quad \forall t \in [a,b]$$

is another Itô process.

Proof. As X_t and Y_t are Itô processes, there exist X_a, Y_a $\mathscr{F}_a$-measurable random variables and f_t^X, g_t^Y $\mathscr{F}_t$-measurable adapted processes with $\int_a^b \left|f_t^{X,Y}\right|^2 dt < \infty$, $\int_a^b \left|g_t^{X,Y}\right| dt < \infty$ such that

$$X_t = X_a + \int_a^t f_s^X dW_s + \int_a^t g_s^X ds, \quad \forall t \in [a,b],$$
$$Y_t = Y_a + \int_a^t f_s^Y dW_s + \int_a^t g_s^Y ds, \quad \forall t \in [a,b].$$

In stochastic differential form, we can compute with the help of Itô's table $(dX_t)(dY_t) = f_t^X f_t^Y dt$ and then, rewriting Equation (2.20) we get

$$\int_a^b X_t \circ dY_t = \int_a^b X_t f_t^Y dW_t + \int_a^b \left(X_t g_t^Y + \frac{1}{2} f_t^X f_t^Y \right) dt.$$

Now, we just have to bound each of these terms. As we mentioned, almost all of the sample paths of X_t are continuous, so we can consider its supremum to bound the expression. Then it follows,

$$\int_a^b \left| X_t f_t^Y \right|^2 dt \leq \sup_{s\in[a,b]} |X_s|^2 \int_a^b \left| f_t^Y \right|^2 dt < \infty,$$

$$\int_a^b \left| X_t g_t^Y \right| dt \leq \sup_{s\in[a,b]} |X_s| \int_a^b \left| g_t^Y \right| dt < \infty.$$

Moreover, by the *Schwarz inequality*[5] we have

$$\int_a^b \left| f_t^X f_t^Y \right| dt \leq \left(\int_a^b \left| f_t^X \right|^2 dt \right)^{1/2} \left(\int_a^b \left| f_t^Y \right|^2 dt \right)^{1/2} < \infty.$$

As the $\mathscr{F}_t$-measurability of each of the terms is fulfilled trivially, we have shown that Z_t is an Itô process too. □

In the sequel, we will always make Stratonovich's integrals with respect to the Brownian, i.e, $dY_t = dW_t$, to follow the line of previous chapters, so the Equations that will characterize these processes are (2.19) and (2.22), which carry the same information but are expressed for different contexts.

2.3.2 Advantages and disadvantages

The election of a different evaluation point for the Brownian motion makes Stratonovich's integral to exhibit different properties. For example, one can wonder whether the assertion in Lemma 2.9 is also satisfied or not. To discuss it, we will generally appeal to the conversion formula. For example, the Stratonovich's integral of an Itô processes with respect to W_t is linear. It is a trivial consequence of the same property for Itô's integrals and also linearity of the derivative and standard integral,

$$\begin{aligned}\int_a^b (aX_t + bY_t) \circ dW_t &= \int_a^b (aX_t + bY_t)\, dW_t + \frac{1}{2}\int_a^b \frac{\partial}{\partial x}\left(a f_t^X + b f_t^Y \right) dt \\ &= a\int_a^b X_t dW_t + \frac{a}{2}\int_a^b \frac{\partial f_t^X}{\partial x} dt + b\int_a^b Y_t dW_t + \frac{b}{2}\int_a^b \frac{\partial f_t^Y}{\partial x} dt \\ &= a\int_a^b X_t \circ dW_t + b\int_a^b Y_t \circ dW_t,\end{aligned}$$

when $X_t = X_a + \int_a^t f_s^X dW_s + \int_a^t g_s^X ds$ and $Y_t = Y_a + \int_a^t f_s^Y dW_s + \int_a^t g_s^Y ds$ are Itô processes defined $\forall t \in [a,b]$. However, the rest of the good conditions are not fulfilled anymore. The centrality is substituted by

$$E\left(\int_a^b X_t \circ dW_t \right) = E\left(\int_a^b X_t dW_t \right) + \frac{1}{2} E\left(\int_a^b f_t dt \right) = \frac{1}{2}\int_a^b E\left(f_t \right) dt,$$

which is not zero in general, and finally, the isometry is not true either since

$$\begin{aligned}E\left(\int_a^b X_t \circ dW_t \right)^2 &= E\left(\int_a^b X_t dW_t \right)^2 + 2E\left(\int_a^b X_t dW_t \cdot \int_a^b f_t dt \right) + E\left(\int_a^b f_t dt \right)^2 \\ &= \int_a^b E\left(X_t^2 \right) dt + 2E\left(\int_a^b X_t dW_t \cdot \int_a^b f_t dt \right) + \int_a^b \int_a^b E\left(f_t f_s \right) dt ds.\end{aligned}$$

It is also important to check whether martingale property is also lost. The following result clarifies the matter.

[5] In fact, it is a particular case of Hölder's inequality, an important result in mathematical analysis which is reminded in the appendices, subsection A.2.

Lemma 2.25. Given an Itô process $X_t = X_a + \int_a^t f_s dW_s + \int_a^t g_s ds$, then the indefinite Stratonovich's integral $Y_t = \int_a^t X_t \circ dW_t$,

- Is a submartingale when $E(f_t) > 0,\ \forall t \in [a,b]$.
- Is a supermartingle when $E(f_t) < 0,\ \forall t \in [a,b]$.
- Is a martingale when $E(f_t) = 0,\ \forall t \in [a,b]$.

In general, we can only say that Y_t can be expressed as a martingale plus another process of finite variation.

Proof. We first fix $a \leq s \leq t \leq b$. Turning again to the conversion formula for our two different integrals in Equation (2.21), we see that Y_t is integrable since it is $X_t = \int_a^t f_t(W_s)dt$ too and the hypothesis on the coefficient $\int_a^b |f_t|^2 < \infty$ implies

$$E\left(\int_a^t f_t dt\right) \leq E\left(\int_a^b f_t^2 dt\right)^{1/2} = \left(\int_a^b E\left(f_t^2\right) dt\right)^{1/2} < \infty.$$

Hence, we can compute the conditional expectation of Y_t with respect to the filtration $\mathscr{F}_s$, and using its linearity we get

$$E\left(Y_t|\mathscr{F}_s\right) = E\left(X_t|\mathscr{F}_s\right) + E\left(\int_a^t f_s ds|\mathscr{F}_s\right).$$

As we shew in Proposition 2.13, Itô's integral is a martingale so the first term is equal to X_s. Therefore, it holds that

$$E\left(Y_t|\mathscr{F}_s\right) = X_s + E\left(\int_a^t f_t dt|\mathscr{F}_s\right) \begin{cases} < X_s + E\left(\int_a^s f_t dt|\mathscr{F}_s\right) = Y_s & \text{if } \mathrm{E}(f_t) > 0,\ \forall t \in [a,b] \\ > X_s + E\left(\int_a^s f_t dt|\mathscr{F}_s\right) = Y_s & \text{if } \mathrm{E}(f_t) < 0,\ \forall t \in [a,b] \\ = X_s = Y_s & \text{if } \mathrm{E}(f_t) = 0,\ \forall t \in [a,b] \end{cases}$$

proving the different cases stated in the lemma. □

Until here, it seems as if we had lost properties with respect to Itô's integral, even though, having chosen the intermediate of the interval to evaluate W_t makes the process to satisfy deterministic techniques of integration.

Theorem 2.26. *Let $F_t(x)$ be an antiderivative in x of a function $f_t(x)$. Suppose that $\partial_t F$, $\partial_t f$ and $\partial_x f$ are continuous. Then,*

$$\int_a^b f_t\left(W_t\right) \circ dW_t = F_t\left(W_t\right)\Big]_a^b - \int_a^b \frac{\partial F_t}{\partial t}\left(W_t\right) dt. \tag{2.23}$$

Particularly, when f has no dependecy on time, we get

$$\int_a^b f_t\left(W_t\right) \circ dW_t = F_t\left(W_t\right)\Big]_a^b.$$

Proof. Consider a function $F_t(x)$ with continuous derivatives $\partial_t F$, $\partial^2_{xx}F$ and $\partial^2_{tx}F$. Taking the stochastic differential form of Equation (2.19) applied to $\partial_x F$ gives

$$\frac{\partial F_t}{\partial x}(W_t) \circ dW_t = \frac{\partial F_t}{\partial x}(W_t)\, dW_t + \frac{1}{2}\frac{\partial^2 F_t}{\partial x^2}(W_t)\, dt.$$

Otherwise, by Itô formula, we know

$$dF_t(W_t) = \frac{\partial F_t}{\partial t}(W_t)\, dt + \frac{\partial F_t}{\partial x}(W_t)\, dW_t + \frac{1}{2}\frac{\partial^2 F_t}{\partial x^2}(W_t)\, dt,$$

so we obtain the equivalence

$$\frac{\partial F_t}{\partial x}(W_t) \circ dW_t = dF_t(W_t) - \frac{\partial F_t}{\partial t}(W_t)\, dt,$$

which corresponds to the differential form of Equation (2.23). □

This crucial result means that Stratonovich's integral behaves like standard integral in Leibniz-Newton calculus. In addition, we can use it to evaluate Stratonovich's integrals and then use the conversion formula to obtain the respective Itô's solution faster.

Example 2.27. We want to compute $\int_a^b \cos(W_t) dW_t$. To do it, it is easier to consider first its Stratonovich's representation since the integral will be direct,

$$\int_a^b \cos(W_t) \circ dW_t = \sin(W_t)]_b^a = \sin(W_a) - \sin(W_b).$$

Thus, with the conversion formula we obtain

$$\begin{aligned}\int_a^b \cos(W_t) dW_t &= \int_a^b \cos(W_t) \circ dW_t - \frac{1}{2}\int_a^b [-\sin(W_t)]\, dt \\ &= \sin(W_a) - \sin(W_b) + \frac{1}{2}\int_a^b \sin(W_t) dt.\end{aligned}$$

Consequently, an important question arises over what is the correct interpretation for the stochastic integral. The answer is not clear and will depend on the situation. For example, due to the martingale property of Itô's integral, it is easier to compute the conditional expectation of a process so this perspective would be more appropriate in mathematical theory. Moreover, imagine that a stochastic integral for given observable f_t comes from a noisy term as

$$\int_a^b f_t dW_t = \int_a^b f_t \xi_t dt.$$

Then, guessing that f_t is a $\mathscr{F}_t$-adapted process seems to be the suitable in gambling, since the decision must be taken before you are provided with new information, and in Itô's integral the current knowledge is independent of the noise increment produced at this time. In other words, choosing Itô's approach makes the stochastic process f_t to be uncorrelated with the white noise acting on the system at this same time. That is why it is usually called *nonanticipating*.

Hence, while Itô's view is more used in Mathematics and Economics, Stratonovich's is the common choice in experimental sciences. Next section tries to explain the because of this advisable election.

2.4 Approximation of stochastic integrals

As we commented previously, many physical systems are influenced by white noise (for example, if they are driven by the *Langevin equation*, described next chapter). However, Brownian motion is an idealization as it cannot be achievable in real world, but approximated by a sequence $\{\Phi_t^n\}_{n\geq 1}$ of well-behaved stochastic processes. For that reason, it becomes interesting to study the convergence of integrals $\int_a^b h_t(\Phi_t^n)d\Phi_t^n$, where each Φ_t^n is a reasonable smooth function satisfying ordinary calculus. Before giving the main result in this area due to E. Wong and M. Zakai, we distinguish the following types of approximations to the Brownian motion.

(C1) For almost all ω, $\Phi_t^n \to W_t$ a.s. $\forall t \in [a,b]$ when $n \to \infty$ and $\Phi_t^n(\omega)$ are continuous functions of bounded variation $\forall n \geq 1$.

(C2) Condition **(C1)** and also that for almost all ω, $\exists\ n_0(\omega), k(\omega) < \infty$ such that $\Phi_t^n(\omega) \leq k(\omega)$, $\forall n > n_0$, $\forall t \in [a,b]$.

Observe that **(C1)** ensures the existence of the Riemann-Stieljes integral because of Theorem 2.1. In the sequel, we examine a result which gives a direct relation between these sequences of ordinary integrals and the Stratonovich's stochastic integral defined before. In its proof, we require some tools from a course on measure theory that are collected in the appendices, subsection A.2.

Theorem 2.28. *(Wong and Zakai, I part)* Let $h_t(x)$ a continuous function with continuous partial derivative $\partial_x h_t(x)$, and $\partial_t h_t(x)$ Moreover, $\{\Phi_t^n\}_{n\geq 1}$ will be a sequence of stochastic processes satisfying **(C2)**. Then, a.s.,

$$\lim_{n\to\infty} \int_a^b h_t(\Phi_t^n)\, d\Phi_t^n = \int_a^b h_t(W_t) \circ dW_t.$$

In addition, if $h_t(x)$ is independent of t, the same assertion is correct under conditon **(C1)**.

Proof. We define a function

$$F_t(x) = \int_0^x h_t(y)dy \quad t \in [a,b]\,,\ x \in \mathbb{R},$$

whose differential evaluated on Φ_t^n is, due to the chain rule in Newton-Leibniz deterministic calculus

$$dF_t(\Phi_t^n) = \frac{\partial F_t}{\partial t}(\Phi_t^n)\, dt + h_t(\Phi_t^n)\, d\Phi_t^n.$$

Consequently,

$$\int_a^b h_t\left(\Phi_t^n\right) d\Phi_t^n = F_b\left(\Phi_b^n\right) - F_a\left(\Phi_a^n\right) - \int_a^b \frac{\partial F_t}{\partial t}\left(\Phi_t^n\right) dt.$$

When making the limit of the previous expression, the first two terms in the right-side do not show problem as the sequence is on the evaluation point of the integral, however, for the last one, we proceed more subtly. As $\partial_x h_t(x)$ is a continuous function defined on the interval $[a, b]$, it will bounded and then, applying the *theorem of derivation under the integral sign* we will get

$$\frac{\partial F_t}{\partial t}(x) = \frac{\partial}{\partial t}\left(\int_a^x h_t(y)dy\right) = \int_a^x \frac{\partial h_t}{\partial t}(y)dy.$$

Therefore, using **(C2)**,

$$\frac{\partial F_t}{\partial t}(\Phi_t^n) = \int_a^{\Phi_t^n} \frac{h_t}{\partial t}(y)dy \Rightarrow \left|\frac{\partial F_t}{\partial t}(\Phi_t^n)\right| \leq \int_a^k \left|\frac{\partial h_t}{\partial t}(y)\right| dy \leq \max_{y\in[a,k]} \frac{\partial_t h}{\partial t}(y)\,(k-a)\,.$$

Notice that the existence of such a maximum is due to Weierstrass theorem ($\partial_t h_t$ is a continuous function defined on a closed interval). Consequently, we can use the *dominated convergence theorem* to obtain

$$\lim_{n\to\infty}\int_a^b \frac{\partial F_t}{\partial t}(\Phi_t^n)dt = \int_a^b \left(\lim_{n\to\infty}\frac{\partial F_t}{\partial t}(\Phi_t^n)\right)dt = \int_a^b \frac{\partial F_t}{\partial t}(W_t)dt,$$

and so,

$$\lim_{n\to\infty}\int_a^b h_t\left(\Phi_t^n\right)d\Phi_t^n = F_b\left(W_b\right) - F_a\left(W_a\right) - \int_a^b \frac{\partial F_t}{\partial t}\left(W_t\right)dt. \tag{2.24}$$

In case that $\partial_t h_t = 0$, the last integral would vanish and the statement of the theorem is satisfied if we replace **(C1)** by **(C2)**. On the other hand, we can take Itô formula in Equation (2.15) to write F as

$$F_b\left(W_b\right) = F_a\left(W_a\right) + \int_a^b \frac{\partial F_t}{\partial t}\left(W_t\right)dt + \int_a^b h_t\left(W_t\right)dW_t + \frac{1}{2}\int_a^b \frac{\partial h_t}{\partial x}\left(W_t\right)dt \tag{2.25}$$

Again, we used the continuity of $\partial_t h_t$ and $\partial_x h_t$ to derive the integral under the integral sign. Then, at the final point we only have to connect Equations (2.24) and (2.25) to reach the expression

$$\lim_{n\to\infty}\int_a^b h_t\left(\Phi_t^n\right)d\Phi_t^n = \int_a^b h_t\left(W_t\right)dW_t + \frac{1}{2}\int_a^b \frac{\partial h_t}{\partial x}\left(W_t\right)dt,$$

where the left-hand side of the equation corresponds to the definition of the Stratonovich's integral in terms of Itô's one. □

As a consequence, if we come back to Itô-Stratonovich dilemma, this tells that the second will be the right choice when the noise is approximated by smoothly well-defined processes. This is the case of a particle suspended in a fluid, where the collisions happen in a smaller scale of time than the average motion of the object. Then, as the speed is finite its correlation time τ will be tiny, but we ideally expect recovering white noise when $\tau \to 0$ and the frequency spectrum tends to be uniform. It is important to emphasize that this is not applicable to Finance, where the processes are usually discontinuous and thus Itô's calculus is chosen.

Furthermore, we can generalize this result in case that ξ_t does not satisfy $E\left(\xi_t\xi_{t+\Delta t}\right) = 0$, the so-called *coloured noise*. Then, the Stratonovich's integral is also the limit of continuous time coloured noise.

2.4.1 Numerical simulation I

The aim of this part is applying the theory developed in this section to study a case numerically. In particular, we focus on solving Example 2.2, showing the main differences between Itô's and Stratonovich's integration and give evidence of Wong-Zakai theorem for approximating integrals. For the sake of simplicity, we take $a = 0, b = 1$. In other words, we want to compute $\int_0^t W_s dW_s$ and $\int_0^t W_s \circ dW_s$, $0 \leq t \leq 1$, so as to exemplify the discrepancy of both points of views.

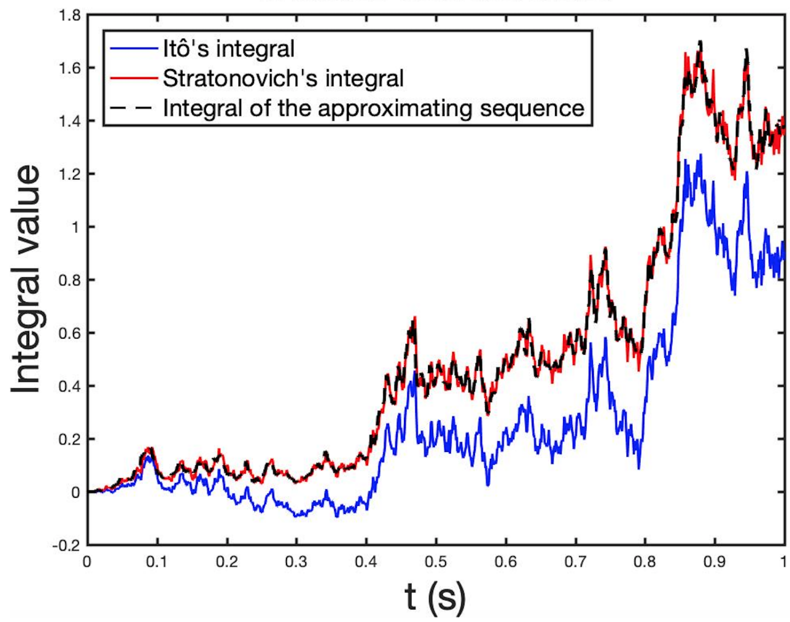

Figure 2.1: Value of the indefinite integral of W_t for Itô's and Stratonovich's description and also using an approximating sequence of well-behaved functions. The simulation of the Brownian was taken for 1000 steps in time and only 300 to approximate it with splines.

To do so, we fix a time scale and a partition of it with a considerable number of time steps of length $\Delta_j t$ where the functions will be evaluated. Then, we approximately build $\Delta_j W \approx \gamma_j \sqrt{\Delta_j t}$, being γ_j a normal distributed random number, and W is the cumulative sum of the vector dW. After that, a simple way of computing stochastic integrals is recovering Theorem 2.11 and Definition 2.18 to obtain values for each indefinite integral in the "Riemann style". Finally, if we also want to compute its approximating deterministic integrals, we proceed as follows:

- We first set n the number of points of intermediate points at which we will take the exact values of W_t. The resulting number of steps in time must be lower than the taken to build the original Brownian vector.
- We interpolate the rest of the values of the time vector with *splines*. These are well-behaved functions and satisfy the assumptions in Theorem 2.28.

- We make the numerical derivative of this, for example, with the *five-point method*.
- To estimate the Riemann-Stieljes integral, we solve with *Simpson 1/3 rule* the following integral,
$$\int_a^b \Phi_t^n \, d\Phi_t^n = \int_a^b \Phi_t^n \, \frac{d\Phi_t^n}{dt} dt.$$
- Then, we can repeat the same argument for a higher number of intermediate points.

The results are shown in Figure 2.1. Note that we can distinguish different results for Itô's and Stratonovich's approaches, and the approximating sequence of deterministic integrals clearly tends to the Stratonovich's integral, which is coherent with Theorem 2.28. Moreover, we note that in this case, Stratonovich's integral always takes greater values of Itô's one (even for other simulations of the same problem). This is direct consequence of the conversion formula, because here $f_t(W_t) = W_t$ and so $\frac{1}{2}\int_0^t \frac{\partial f_t}{\partial x}(W_t) = \frac{t}{2} > 0, \forall t \in [0,1]$.

Chapter 3

Stochastic Differential Equations

In this section, we introduce the study of Stochastic Differential Equations (SDEs). On the one hand, using both Itô's and Stratonovich's differantials we give some examples to explore the behaviour of the solutions and then we prove a crucial result which states existence and uniqueness of these processes under some hypothesis we will mention. On the other, the chapter finishes with a discussion referred to the approximation of solutions of SDEs by the analogous solutions of Ordinary Differential Equations (ODEs).

As a historical note, the first person proposing a "SDE" was Louis Bachelier, who in 1900 used W_t to describe the market fluctuations affecting the price of the Paris Stock Exchange. Of course he did not mention the term Brownian motion, but he used a process with the same properties.

3.1 Introductory examples

Our goal is solving equations of the form:

$$dX_t = \sigma_t(X_t)dW_t + f_t(X_t)dt, \; X_a = x_a,$$

where $\sigma_t(x)$ and $f_t(x)$ are measurable functions of $t \in [a, b]$, $x \in \mathbb{R}$, and also knowing an initial condition x_a, which is a random variable independent of the Brownian motion W_t. Remember, as usual, that the accurate sense of this kind of expressions comes from its integral representation, called Stochastic Integral Equation (SIE):

$$X_t = x_a + \int_a^t \sigma_s(X_s)dW_s + \int_a^t f_s(X_s)ds, \quad t \in [a, b], \tag{3.1}$$

because dW_t does not have meaning by itself. Commonly, the coefficients $f_t(X_t)$ and $\sigma_t(X_t)$ are known as *drift* and *diffusion* coeffient respectively. Note that when the latter vanishes, we recover the ordinary differential approach $\frac{dX_t}{dt} = b_t(X_t)$.

To give some heuristic interpretation to a SDE, we can think that the increment $\Delta X_t = X_{t+\Delta t} - X_t$ is the sum of two terms: one comes from a deterministic sight, $b_t(X_t)$, and the other is a random impulse $\sigma_t(X_t)$, so that the distribution of the increment will be normal with mean $b_t(X_t)\Delta t$ and variance $\sigma_t^2(X_t)\Delta t$. This view is connected with the following example.

Example 3.1. (*Langevin equation*) To model the velocity of a Brownian particle under frictional forces, we use the equation

$$\frac{dX_t}{dt} = -\beta X_t + \alpha\xi, \quad X_a = x_a,$$

where $\beta > 0$ and $t \geq a$. For the sake of simplicity, we will also take $\alpha \in \mathbb{R}$, which defines the noise as *additive*. If we had taken $\alpha = \alpha\,(X_t)$, it is said *multiplicative noise* and requires a more sophisticated treatment. Notice that Langevin equation is not more than the 2nd Newton's Law with a diffusion extra term. Written in stochastic differential notation it is

$$dX_t = -\beta X_t dt + \alpha dW_t, \quad X_a = x_a.$$

To solve this, we use the technique of *variation of constants*. We first consider the solution of the same equation without the diffusion term,

$$dX_t = -\beta X_t dt \Rightarrow X_t = x_a(t)e^{-\beta t},\ t \geq a.$$

Since $x_a(t)$ is a random variable, in order take the differential form of X_t we shall use Itô formula, obtaining

$$dX_t = dx_a(t)e^{-\beta t} - \beta x_a(t)e^{-\beta t}dt = dx_a(t)e^{-\beta t} - \beta X_t dt,$$

which must be equal to the original SDE,

$$dx_a(t)e^{-\beta t} - \beta X_t dt = -\beta X_t dt + \alpha dW_t \Rightarrow dx_a(t) = \alpha e^{\beta t}dW_t \Rightarrow x_a(t) = x_a + \alpha\int_a^t e^{\beta s}dW_s,$$

so we finally get

$$X_t = x_a e^{-\beta t} + \int_a^t e^{-\beta(t-s)}dW_s.$$

One further point: this kind of solutions has its own name: *Ornstein-Uhlenbech processes*, and they play an important role in Physics and Mathematical Finance.

Before moving forward, we give two more examples that show how stochastic differential equations can sometimes exhibit defective properties.

Example 3.2. Let us analyze now the following SDE:

$$dX_t = X_t^2 dW_t + X_t^3 dt,\ X_a = 1.$$

To find the solution for $t \in [a,b]$, it is convenient to apply Itô formula to compute

$$d\left(\frac{1}{X_t}\right) = -\frac{1}{X_t^2}dX_t + \frac{1}{2}\frac{2}{X_t^3}\left(dX_t\right)^2 = -\frac{1}{X_t^2}\left(X_t^2 dW_t + X_t^3 dt\right) + X_t dt = -dW_t.$$

Hence, we have that $\frac{1}{X_t} = -W_t + C$, being C a constant. This, together with the initial condition result in the solution

$$X_t = \frac{1}{1-(W_t - W_a)}.$$

Consequently, the process X_t exists only up to $\tau = \inf_{t\in[a,b]}\left\{W_t = 1 + W_a\right\}$, which is called the *explosion time* of the SDE.

Example 3.3. Consider also the following equation to solve

$$dX_t = 3X_t^{2/3}dW_t + 3X_t^{1/3}dt, \quad X_0 = 0.$$

It is easily verifiable that $X_t = (W_t - c)^3 \mathbb{1}_{\{W_t \geq c\}}$ is solution of the equation $\forall c > 0$ (just using anew Itô formula), which means that there exist infinite processes satisfying such SDE.

These two last examples, which actually have equivalent versions in ordinary differential calculus, show the necessity to impose further conditions to the functions $\sigma_t(x)$ and $f_t(x)$ in order to achieve existence and uniqueness of solutions.

3.2 Preparatory lemmas and hypothesis

We aim to prove the analogous Peano and Picard's theorem of ODEs for our SDEs coming Equation (3.1). Before doing that, we should specify what kind of solutions we are searching for and which assumptions we need to reach the desired result.

Definition 3.4. We will say that a $\mathscr{F}_t$-adapted stochastic process X_t is a *solution of our SIE* in Equation (3.1) $\forall t \in [a,b]$ if it holds:

1. $\sigma \in L^2_{ad}([a,b] \times \Omega)$, so that $\int_a^t \sigma_s(X_s)dW_s$ is an Itô's integral, $\forall t \in [a,b]$.
2. $f \in L^1([a,b] \times \Omega)$, $\forall t \in [a,b]$.
3. For any $t \in [a,b]$, Equation (3.1) holds true a.s.

Definition 3.5. We say that the Equation (3.1) has a *path-wise unique solution* if given any $X_t^{(1)}$ and $X_t^{(2)}$ satisfying Definition (3.4), then

$$P\left\{X_t^{(1)} = X_t^{(2)},\ t \in [a,b]\right\} = 1.$$

What is to say, they are *indistinguishable*.

Let us now talk about the requirements for drift and diffusion coefficients, discussing a bit why they are necessary.

Definition 3.6. Given a measurable function $f_t(x)$ on $[a,b] \times \mathbb{R}$, we say it is *Lipschitz in the x variable* when it is Hölder continuous of degree 1. What is to say, if there exists a constant $K > 0$ such that:

$$|f_t(x) - f_t(y)| \leq K|x-y|, \quad \forall t \in [a,b], \quad \forall x, y \in \mathbb{R}.$$

This is some kind of bound for the derivative of the function (in case that f had one) and was also present in the deterministic case. Notice that Example 3.3 does not satisfy Lipschitz condition on x so it seems to be required to ensure uniqueness of solutions.

Definition 3.7. We say that a measurable function $f_t(x)$ on $[a,b] \times \mathbb{R}$ follows *linear growth condition* if there exists a constant $K > 0$ such that:

$$|f_t(x)| \leq K(1+|x|), \quad \forall t \in [a,b], \quad \forall x \in \mathbb{R}.$$

In fact, this condition is equivalent to the existence of a constant $K > 0$ satisfying $|f_t(x)|^2 \leq K\left(1+x^2\right)$. This is decisive in order to guarantee the existence avoiding explosion. Notice that, for example, the process in Example 3.2 does not satisfy it. Moreover, it is important to remark that this condition is weaker than continuity in time. Indeed, for any continuous time process $f_t(x)$, also Lipschitz in x, we can bound

$$|f_t(x)| \leq |f_t(x) - f_t(0)| + |f_t(0)| \leq K(1+|x|),$$

where K is now the maximum between the Lipschitz constant and the bound for the continuous function $f_t(0)$. To finish this section, we give a pair of lemmas that will appear in the proof of the theorems.

Lemma 3.8. *(a Gronwall lemma)* Let $u, v : [a,b] \to \mathbb{R}_+$ be continuous functions. If

$$v(t) \leq \alpha + \int_a^t u(s)v(s)ds, \ \forall t \in [a,b],$$

for some constant $\alpha > 0$, then $v(t) \leq \alpha e^{\int_a^t u(s)ds}$.

Proof. Define a function ψ as $\psi(t) := \alpha + \int_a^t u(s)v(s)ds$ whose derivative is $\psi'(t) = u(t)v(t)$. Hence,

$$\left(e^{-\int_a^t u(s)ds}\psi(t)\right)' = \left[\psi'(t) - u(t)\psi(t)\right] e^{-\int_a^t u(s)ds} \leq \left[u(t)\psi(t) - u(t)\psi(t)\right] e^{-\int_a^t u(s)ds} = 0,$$

and so

$$\psi(t)e^{-\int_a^t u(s)ds} \leq \psi(a)e^{-\int_a^a u(s)ds} = \alpha.$$

Finally, the result follows from $v(t) \leq \psi(t) \leq \alpha e^{\int_a^t u(s)ds}$. □

Lemma 3.9. For any process $f \in L^2_{ad}([a,b] \times \Omega)$, and $\epsilon, N > 0$, it holds

$$P\left\{\left|\int_a^b f_t dW_t\right| > \epsilon\right\} \leq P\left\{\int_a^b f_t^2 dt > N\right\} + \frac{N}{\epsilon^2}.$$

We omit the demonstration of this lemma, but we want to emphasize that it is one of the crucial tools we need to extend the definition of Itô's integral for a larger set of processes, as we mentioned in Remark 2.10.

3.3 A result on existence and uniqueness of solutions

In the following, we prove these two notable results always under the considerations we made in the previous section. As we will see, there are involved many notions studied during the degree, so we highly recommend the reader consulting the appendix as a reminder.

Theorem 3.10. Let $\sigma_t(x)$ and $f_t(x)$ be measurable functions on $[a,b]\times\mathbb{R}$ satisfying Lipschitz and linear growth conditions on x. Assume also that the initial state X_a is a $\mathscr{F}_a$-measurable random variable with $E\left(X_a^2\right)<\infty$. Under these hypothesis, the SDE

$$dX_t=\sigma_t(X_t)dW_t+f_t(X_t)dt,\ X_a=x_a \tag{3.2}$$

has a continuous solution X_t.

Proof. We are going to proceed similarly as with the proof for ODEs. Given a sequence of continuous stochastic processes $\{X_t^{(n)}\}_{n\geq 1}$, we set Picard's iteration scheme in the following way:

$$X_t^{(0)}=X_a,\ X_t^{(n+1)}=X_0+\int_a^t\sigma_s\left(X_s^{(n)}\right)dW_s+\int_a^t f_s\left(X_s^{(n)}\right)ds,\ n\geq 0, \tag{3.3}$$

with $t\in[a,b]$. To begin, we should see that this sequence defined recursively belongs to $L^2_{ad}\left([a,b]\times\Omega\right)$. Let us do it by induction. Trivially, $X_t^{(0)}$ fulfills this assumption, and suppose now that it is also true for a process $X_t^{(n)}$. Then, using linear growth condition (being K_1, K_2 the respective constants),

$$E\int_a^b\sigma\left(t,X_t^{(n)}\right)^2dt\leq K_1\left(b-a\right)+K_1E\int_a^b\left|X_t^{(n)}\right|^2dt<\infty, \tag{3.4}$$

$$\int_a^t\left|f_s\left(X_s^{(n)}\right)\right|ds\leq K_2\int_a^b\left(1+\left|X_t^{(n)}\right|\right)dt\leq K_2\left(b-a\right)+\int_a^b\left|X_t^{(n)}\right|<\infty, \tag{3.5}$$

and we used the inductive hypothesis to prove that such expressions are finite. With this, we prove that $\sigma_s(X_s^{(n)})\in L^2_{ad}\left([a,b]\times\Omega\right)$ and $f_t(X_s^{(n)})\in L^1\left([a,b]\times\Omega\right)$. On the other hand, applying the identity $|a+b+c|^2\leq 3\left(a^2+b^2+c^2\right)$ to the iterative Equation (3.3) we get

$$\left|X_t^{(n+1)}\right|^2\leq 3\left[X_0^2+\left(\int_a^t\sigma_s\left(X_s^{(n)}\right)dW_s\right)^2+\left(\int_a^t f_s\left(X_s^{(n)}\right)ds\right)^2\right],$$

which together with Equations (3.4) and (3.5) lead us to conclude

$$E\int_a^b\left|X_t^{(n+1)}\right|^2dt<\infty\Rightarrow X_t^{(n+1)}\in L^2_{ad}\left([a,b]\times\Omega\right).$$

Next, we compute $D_t^n:=E(|X_t^{(n+1)}-X_t^{(n)}|^2)$. As in the previous step, we will prove by induction on n that

$$D_t^n\leq\frac{[M(t-a)]^{n+1}}{(n+1)!}, \tag{3.6}$$

for some constant M, depending on $(b-a)$, X_a and L which is defined to be the maximum of the two linear growth constants. The case for $n=0$ follows from

$$\begin{aligned}D_t^0&=E\left(\left|X_t^{(1)}-X_t^{(0)}\right|^2\right)=E\left(\left|\int_a^t f_s(X_a)ds+\int_a^t\sigma_s(X_a)dW_s\right|^2\right)\\&\leq 2E\left(\left|\int_a^t L\left(1+|X_a|\right)ds\right|^2\right)+2E\left(\int_a^t L^2\left(1+|X_a|\right)^2ds\right)\\&\leq 2L^2\left(b-a\right)E\left(\int_a^t\left(1+|X_a|\right)^2ds\right)+2L^2E\left(\int_a^t\left(1+|X_a|\right)^2ds\right)\leq M\left(t-a\right),\end{aligned}$$

being $M = 2L^2(1+b-a)E(1+|X_a|^2)$. After this, we make the hypothesis that the inequality is valid for any natural number $m \leq n-1$. Then,

$$D_t^{(n)} = E\left(\left|\int_a^t f_s\left(X_s^{(n)}\right) - f_s\left(X_s^{(n-1)}\right) ds + \int_a^t \sigma_s\left(X_s^{(n)}\right) - \sigma_s\left(X_s^{(n-1)}\right) dW_s\right|^2\right)$$
$$\leq 2(b-a)L^2 E\left(\int_a^t \left|X_s^{(n)} - X_s^{(n-1)}\right|^2 ds\right) + 2L^2 E\left(\int_a^t \left|X_s^{(n)} - X_s^{(n-1)}\right|^2 ds\right),$$

which by the induction hypothesis results in

$$D_t^{(n)} \leq 2L^2(1+b-a)\int_a^t \frac{M^n s^n}{n!} ds \leq \frac{M^{n+1}(t-a)^{n+1}}{(n+1)!},$$

if we carefully choose the same value of M as above. Hence, the desired claim is proven. After this, we consider

$$\sup_{t\in[a,b]}\left|X_t^{(n+1)} - X_t^{(n)}\right|^2 \leq 2(b-a)L^2\int_a^b \left|X_t^{(n)} - X_t^{(n-1)}\right|^2 ds$$
$$+ 2\sup_{t\in[a,b]}\left|\int_a^t \left[\sigma_s\left(X_t^{(n)}\right) - \sigma_s\left(X^{(n-1)}\right)\right] dW_s\right|^2.$$

Now, we can apply Dood's L^2 inequality (stated in Theorem 1.7) to the second term so that we reach to

$$E\left(\sup_{t\in[a,b]}\left|X_t^{(n+1)} - X_t^{(n)}\right|^2\right) \leq 2(b-a)L^2 E\left(\int_a^b \sup_{t\in[a,b]}\left|X_t^{(n+1)} - X_t^{(n)}\right|^2 ds\right)$$
$$+ 8L^2 E\left(\int_a^b \sup_{t\in[a,b]}\left|X_t^{(n+1)} - X_t^{(n)}\right|^2 ds\right) \leq C\frac{M^n(t-a)^n}{n!},$$

where we used the Inequality (3.6) proved before to bound the whole result, and $C > 0$ collects some of the other constants. Moreover, using *Chevychev's inequality* we can write

$$P\left(\sup_{t\in[a,b]}\left|X_t^{(n+1)} - X_t^{(n)}\right| > \frac{1}{2^n}\right) \leq 2^{2n} E\left(\sup_{t\in[a,b]}\left|X_t^{(n+1)} - X_t^{(n)}\right|^2\right) \leq 2^{2n} C\frac{M^n(t-a)^n}{n!}.$$

Therefore, as $\sum_{n=1}^{\infty} 2^{2n}\frac{M^n(t-a)^n}{n!} < \infty$ (easily demonstrable using *Ratio Test*), *first Borel Cantelli lemma* implies that

$$P\left(\liminf_n\left\{\sup_{t\in[a,b]}\left|X_t^{(n+1)} - X_t^{(n)}\right| \leq \frac{1}{2^n}\right\}\right) = 1.$$

In other words, $\forall\omega$ a.s., $\exists m_0(\omega)$ such that $\forall n \geq m_0(\omega)$

$$\sup_{t\in[a,b]}\left|X_t^{(n+1)} - X_t^{(n)}\right| \leq \frac{1}{2^n}.$$

And consequently, making use of *Weierstrass M-test* we obtain that the following series

$$X_t^{(m)} = X_a + \sum_{k=0}^{m-1} \left[X_t^{(k+1)} + X_t^{(k)} \right],$$

converges uniformly on $[a, b]$ to a process $X_t = \lim_{n\to\infty} X_t^{(n)}$. Anyway, we still have to check some conditions to see that this solution meets the requirements in Definition 3.4. To prove that X_t satisfies Equation (3.1), we distinguish:

- The a.s. convergence of the path-wise integrals is a consequence of

$$\left| \int_a^t \left[f_s(X_s^{(n)}) - f_s(X_s) \right] ds \right| \leq L \int_a^t \left| X_s^{(n)} - X_s \right| ds \leq L \sup_{s\in[a,t]} \left| X_s^{(n)} - X_s \right| \to 0.$$

- However, for the stochastic integral, we need the Lemma 3.9, and thus

$$P\left(\left| \int_a^t \left[\sigma_s(X_s^{(n)}) - \sigma_s(X_s) \right] dW_s \right| > \epsilon \right) \leq P\left(\int_a^t \left| \sigma_s(X_s^{(n)}) - \sigma_s(X_s) \right|^2 ds > N \right) + \frac{N}{\epsilon^2}.$$

  The first term in the right-hand side tends to zero due to the uniform convergence of $\sigma_t(X_t^{(n)}) \to \sigma_t(X_t)$ (true because σ is continuous), so choosing appropriate ϵ and N the assertion is proven.

Finally, the conditions on the coefficients σ and f are direct from Equations (3.4) and (3.5) and we have also seen

$$\left\| X_t^{(n+1)} - X_t^{(n)} \right\|_{L^2} \leq \frac{CM^n (t-a)^n}{n!} \Rightarrow \|X_t\|_{L^2} \leq \|X_0\|_{L^2} + \sum_{n=1}^{\infty} \frac{CM^n (t-a)^n}{n!}.$$

Therefore, $E \int_a^b |X_t|^2 dt < \infty$ and so $X_t \in L^2_{ad}([a,b] \times \Omega)$, which completes the proof.

□

Theorem 3.11. Under the same hypothesis in Theorem 3.10, the solution to the stochastic differential Equation (3.2) is unique.

Proof. In the following, K will denote a big enough constant which can be different in each expression. Let us suppose that X_t and Y_t are different continuous solutions of our SIE. We can write $Z_t = X_t - Y_t$ as

$$Z_t = \int_a^t [\sigma_s(X_s) - \sigma_s(Y_s)] dW_s + \int_a^t [f_s(X_s) - f_s(Y_s)] ds.$$

Then, using again the identity $(a+b)^2 \leq 2(a^2 + b^2)$ we get

$$Z_t^2 \leq 2 \left\{ \left(\int_a^t [\sigma_s(X_s) - \sigma_s(Y_s)] dW_s \right)^2 + \left(\int_a^t [f_s(X_s) - f_s(Y_s)] ds \right)^2 \right\}. \tag{3.7}$$

Now, applying the expectation to the each of its terms, and using Lipschitz property for both $\sigma_t(x)$ and $f_t(x)$ gives us

$$E\left(\int_a^t [\sigma_s(X_s) - \sigma_s(Y_s)]\, dW_s\right)^2 = E\int_a^t [\sigma_s(X_s) - \sigma_s(Y_s)]^2\, ds \leq K^2 \int_a^t E(Z_s^2)ds,$$

$$E\left(\int_a^t [f_s(X_s) - f_s(Y_s)]\, ds\right)^2 = (t-a)\, E\int_a^t [f_s(X_s) - f_s(Y_s)]^2\, ds \leq (t-a)\, K^2 \int_a^t E(Z_s^2)ds.$$

All in all, inserting these two last expressions in Equation (3.7), we reach to

$$E\left(Z_t^2\right) \leq 2K^2\,(1+b-a)\int_a^t E\left(Z_s^2\right) ds.$$

Therefore, Gronwall's Lemma (3.8) with $\alpha = 0$ implies that $E\left(Z_t^2\right) = 0$, $\forall t \in [a,b]$. Hence, $X_t^{(1)} = X_t^{(2)}$, $\forall t \in [a,b]$ a.s. and in particular $X_r^{(1)} = X_r^{(2)}$ for all rational numbers $r \in [a,b]$ except for a set of probability zero. However, as $X_t^{(1)}$ and $X_t^{(2)}$ are assumed to have continuous sample paths a.s., and consequently

$$P\left(\sup_{t\in[a,b]} \left|X_t^{(1)} - X_t^{(2)}\right| > 0\right) = 0.$$

□

3.4 Stratonovich's view of a SDE

Once we have solved some examples and proven the existence and uniqueness of solutions for SDEs of the form as Equation (3.2), one can wonder whether it is possible to change Itô's perspective of the differentials to the other kind of representation we studied in the chapter devoted to stochastic integration. To analyse this, we first consider

$$dX_t = \sigma_t(X_t) \circ dW_t + f_t(X_t)dt,\ X_a = x_a, \tag{3.8}$$

which is just the same expression, replacing Itô's differential by Stratonovich's one denoted with "$\circ$". The coefficients $f_t\,(X_t)$ and $\sigma_t\,(X_t)$ satisfy conditions in Definition 3.4, and we also assume that $\sigma_t\,(X_t) \in \mathcal{C}^{1,2}$ in $[a,b] \times \mathbb{R}$, with $|\partial_x \sigma_t\,(X_t)| < \infty$. Then, we can apply the formula in Equation (2.22) with $\theta_t\,(X_t) = \sigma_t\,(X_t)$ and also $f_t\,(X_t) = \sigma_t\,(X_t)$ so that

$$\sigma_t(X_t) \circ dW_t = \sigma_t(X_t)dW_t + \frac{1}{2}\frac{\partial \sigma_t}{\partial x}\,(X_t)\,\sigma_t\,(X_t)\,dt.$$

Thus, we finally get that the equivalent Equation (3.8) by substituting this expression there, is

$$dX_t = \sigma_t(X_t)dW_t + \left[f_t(X_t) + \frac{1}{2}\frac{\partial \sigma_t}{\partial x}\,(X_t)\,\sigma_t\,(X_t)\right] dt,\ X_a = x_a. \tag{3.9}$$

Note that the existence and uniqueness of solution of this equation is guaranteed by Theorems 3.10 and 3.11, always if we appropriately suppose Lipschitz and linear growth conditions for the new coefficients. We only have to check that the second term in bounded.

Starting with the triangular inequality, we can separate

$$\int_a^b \left| f_t(X_t) + \frac{1}{2}\frac{\partial \sigma_t}{\partial x}(X_t)\,\sigma_t(X_t) \right| dt \leq \int_a^b |f_t(X_t)|\, dt + \frac{1}{2}\int_a^b \left| \frac{\partial \sigma_t}{\partial x}(X_t)\,\sigma_t(X_t) \right| dt.$$

The first term is bounded by hypothesis. For the second, as the integrator is a continuous function, it is easily bounded by its maximum in the interval $[a, b]$.

Remark 3.12. Notice that from the conversion formula (3.9), we see that both approaches coincide when σ_t does not depend on X_t, for example, Langevin equation with additive white noise has the same solution no matter which differential you choose. Anyway, it is important to highlight that in general, the solution of a SDE depends on which is the view we take, and its properties can considerably differ as it shows next example.

Example 3.13. We want to analyze a population growth model in presence of white noise given by the following SDE,

$$\frac{dN_t}{dt} = N_t\left(\alpha + \beta\xi_t\right), \quad \forall t \geq 0,$$

where N_t denotes the number of individuals at time t and $\alpha, \beta \in \mathbb{R}$ are the average growth constant and the noise intensity respectively. Remember also that ξ_t is white noise. Let us begin solving it with Itô's description assuming an initial population N_0. We can rewrite the previous equation as

$$dN_t = \alpha N_t dt + \beta N_t dW_t \Rightarrow \frac{dN_t}{N_t} = \alpha dt + \beta dW_t. \tag{3.10}$$

Now, it is convenient to use Itô formula to compute

$$d\left(\log N_t\right) = \frac{1}{N_t}\beta N_t dW_t + \left(\frac{1}{N_t}N_t\alpha - \frac{1}{2}\beta^2\right) dt = \frac{dN_t}{N_t} - \frac{1}{2}\beta^2 dt,$$

because we can replace this expression in Equation (3.10) and integrate from 0 to t to get

$$\log\left(\frac{N_t}{N_0}\right) = \left(\alpha - \frac{1}{2}\beta^2\right) t + \beta W_t \Rightarrow N_t = N_0 \exp\left[\left(\alpha - \frac{1}{2}\beta^2\right) dt + \beta W_t\right].$$

On the contrary, if we had taken Stratonovich differential, the equation whould have been

$$dN_t = \alpha N_t dt + \beta N_t \circ dW_t \quad \text{with} \quad \beta N_t \circ dW_t = \beta N_t dW_t + \frac{1}{2}\beta^2 N_t dt,$$

so instead of Equation (3.10) we would have

$$\frac{dN_t}{N_t} = \left(\alpha + \frac{1}{2}\beta\right) dt + \beta dW_t,$$

and the finale solution would have been

$$N_t^{Strat} = N_0 \exp\left(\alpha t + \beta W_t\right).$$

Notice that the solutions disagree in the accompanying factor of time, and this gives different predictions for this process. For example, we can find the expectation of such processes. For Itô's solution,

$$E\left(N_t\right) = N_0 \exp\left[\left(\alpha - \frac{1}{2}\beta^2\right)t\right] E\left[\exp\left(\beta W_t\right)\right], \tag{3.11}$$

and in order to compute that, we define $Y_t = \exp\left(\beta W_t\right)$ so

$$dY_t = \beta \exp\left(\beta W_t\right) dW_t + \frac{1}{2}\beta^2 \exp\left(\beta W_t\right) dt \Rightarrow Y_t = Y_0 + \beta \int_0^t Y_s dW_s + \frac{1}{2}\beta^2 \int_0^t Y_s ds.$$

Then, using that Itô's integrals are centered random variables (Lemma 2.9), we apply the expectation to both sides of the equality to obtain

$$E\left(Y_t\right) = E\left(Y_0\right) + \frac{1}{2}\beta^2 E\left[\int_0^t Y_s ds\right] \Rightarrow \frac{d}{dt}E\left(Y_t\right) = \frac{1}{2}\beta^2 E\left(Y_t\right),$$

so we finally have $E\left(Y_t\right) = \exp\left(\frac{1}{2}\beta^2 t\right)$ which substituted in Equation (3.11) gives

$$E\left(N_t\right) = N_0 \exp\left(\alpha t\right).$$

Note that it coincides with the solution of a non-noisy model. On the other hand, we can repeat the same argument for the Stratonovich's solution reaching to the following expression

$$E\left(N_t^{Strat}\right) = N_0 \exp\left[\left(\alpha + \frac{1}{2}\beta^2\right)t\right],$$

which is greater than the other result.

All in all, although is could seem that the choice of the vision would determine the behaviour of our solution X_t, this is not the reality. We should think that depending on the formalism, the coefficients f_t and σ_t take different forms but the mathematics behind them is the same.

3.5 Approximation of solutions of SDEs

In the line with the discussion of Section 2.4, we are now interested in studying which is the limit of solutions of well-behaved stochastic processes in ODEs and how it is related to its analogous SDE associated. Before tackling the problem presenting the solution proposed by Wong and Zakai, we need to prove the following lemma.

Lemma 3.14. *(Another Gronwall inequality)* Let $u : [a,b] \to \mathbb{R}_+$ be a continuous function and suppose also that $\int_a^b v(s)ds < \left(\rho\mu e^{\,\mu\rho(b-a)}\right)^{-1}$. Then, if

$$\log\left(1 + \frac{u(t)}{\mu}\right) \leq \log\left(1 + v(t)\right) + \rho \int_a^t u(s)ds, \tag{3.12}$$

for some fixed constants $0 < \mu < \infty$, $\rho > 0$, it holds the following inequality

$$u(t) \leq \mu \frac{v(t) + \rho\mu e^{\,\rho\mu(a-b)} \int_a^b v(t)dt}{1 - \rho\mu e^{\,\rho\mu(a-b)} \int_a^b v(t)dt}. \tag{3.13}$$

Proof. By making the exponential of the whole expression in Equation (3.12) we have

$$\frac{1 + \frac{u(t)}{\mu}}{e^{\,\rho \int_a^t u(s)ds}} \leq 1 + v(t). \tag{3.14}$$

Now, we arrange the equation by multiplying both sides of the equality by $\mu\rho e^{-\rho\mu t}$ to realize that the left-hand side can be written as a simple derivative,

$$\frac{\rho\mu + \rho u(t)}{e^{\left(\rho \int_a^t u(s)ds + \rho\mu t\right)}} = -\frac{d}{dt} e^{\left(-\rho \int_a^t u(s)ds - \rho\mu t\right)} \leq \rho\mu \left(1 + v(t)\right) e^{-\rho\mu t}.$$

Hence, integrating from a to t, we will get

$$e^{-\rho\mu a} - e^{-\rho \int_a^t u(s)ds - \rho\mu t} \leq e^{-\rho\mu a} - e^{-\rho\mu t} + \rho\mu e^{-\rho\mu a} \int_a^b v(t)dt,$$

and thus, simplifying a bit we obtain

$$e^{-\rho \int_a^t u(s)ds} \geq 1 - \rho\mu e^{\,\rho\mu(t-a)} \int_a^b v(t)dt.$$

Finally, since $\rho\mu e^{\,\rho\mu(b-a)} \int_a^b v(t)dt < 1$ we can invert last expression to reach

$$e^{\,\rho \int_a^t f(s)ds} \leq \frac{1}{1 - \rho\mu e^{\,\rho\mu(t-a)} \int_a^b v(t)} \leq \frac{1}{1 - \rho\mu e^{\,\rho\mu(b-a)} \int_a^b v(t)},$$

and the desired Inequality (3.13) is a consequence of inserting this to the expression (3.14) □

Now, to the conditions we gave in the previous chapter, we add these two more,

(C3) Condition **(C2)** and also that $\Phi_t^n(\omega)$ has a piecewise continuous derivative $\forall n \geq 1$.

(C4) Condition **(C3)** and also that $\Phi_t^n(\omega) \to W_t(\omega)$ as $n \to \infty$ uniformly $\forall t \in [a, b]$.

which are the hypothesis of the following notable result.

Theorem 3.15. *(Wong and Zakai, II part)* Assume that:

1. The functions $f_t(x)$, $\sigma_t(x)$, $\partial_x \sigma_t(x)$ and $\partial_t \sigma_t(x)$ exist and are continuous $\forall x \in \mathbb{R}$ and $\forall t \in [a, b]$.
2. Lipschitz condition in x is satisfied by $f_t(x)$, $\sigma_t(x)$ and $\partial_x \sigma_t^2(x)$.
3. $\exists C_1 > 0$ such that $|\sigma_t(x)| \geq C_1$ and also $|\partial_t \sigma_t(x)| \leq C_2 \sigma_t^2(x)$ for some other positive constant C_2.

Then, if $\{\Phi_t^n\}_{n\geq 1}$ is a sequence of random variables satisfying **(C3)** and $X_t^{(n)}$ the solution of the ODE

$$dX_t^{(n)} = \sigma_t(X_t^{(n)})d\Phi_n(t) + f_t(X_t^{(n)})dt, \quad X_a^{(n)} = x_a,$$

it holds that $X_t^{(n)} \to X_t$ a.s. as $n \to \infty$, being X_t the solution of the SDE

$$dX_t = \sigma_t(X_t) \circ dW_t + f_t(X_t)\,dt, \quad X_a = x_a.$$

Furthermore, if $\{\Phi_t^n\}_{n\geq 1}$ satisfies **(C4)**, the convergence is uniform in $[a,b]$.

Proof. Let us define now an auxiliary function $G_t(x)$ by

$$G_t(x) = \int_0^x \frac{dy}{\sigma_t(y)}, \quad t \in [a,b], \ \forall x \in \mathbb{R}.$$

Then, the differential of $G_t(X_t^{(n)})$, due to Newton-Leibniz chain rule, is

$$dG_t(X_t^{(n)}) = \frac{\partial G_t}{\partial t}(X_t^{(n)})dt + \frac{dX_t^{(n)}}{\sigma_t(X_t^{(n)})} = \frac{\partial G_t}{\partial t}(X_t^{(n)})dt + d\Phi_t^{(n)} + \frac{f_t(X_t^{(n)})}{\sigma_t(X_t^{(n)})}dt,$$

and in integral notation takes the form

$$G_t(X_t^{(n)}) = G_a(X_a^{(n)}) + \Phi_t^{(n)} - \Phi_a^{(n)} + \int_a^t \left[\frac{\partial G_s}{\partial s}(X_s^{(n)}) + \frac{f_s(X_s^{(n)})}{\sigma_s(X_s^{(n)})}\right] ds. \tag{3.15}$$

On the other hand, applying Itô formula to our solution of the SDE, developed as an Itô's integral we get the expression

$$\begin{aligned} G_t(X_t) &= G_a(X_a) + \int_a^t \left[\frac{\partial G_s}{\partial x}(X_s) \cdot \sigma_s(X_s)\right] dW_s + \int_a^t \frac{\partial G_s}{\partial s}(X_s)\,ds \\ &+ \int_a^t \frac{\partial G_s}{\partial x}(X_s) \cdot \left[f_s(X_s) + \frac{1}{2}\sigma_s(X_s)\frac{\partial \sigma_s}{\partial x}(X_s)\right] ds + \frac{1}{2}\int_a^t \frac{\partial^2 G_s}{\partial x}(X_s) \cdot \sigma_s^2(X_s)\,ds. \end{aligned}$$

Now, using the fundamental theorem of calculus, we will have $\partial_x G_s(X_s) = \sigma^{-1}(X_s)$, and therefore,

$$\begin{aligned} G_t(X_t) &= G_a(X_a) + \int_a^t dW_s + \int_a^t \frac{G_s}{\partial s}(X_s)\,ds + \int_a^t \left[\frac{f_t(X_s)}{\sigma_s(X_s)} + \frac{1}{2}\frac{\partial \sigma_s}{\partial x}(X_s)\right] ds \\ &+ \frac{1}{2}\int_a^t \frac{\partial}{\partial x}\left(\frac{1}{\sigma_s(X_s)}\right) \cdot \sigma_s^2(X_s)\,ds = G_a(X_a) + W_t - W_a \\ &+ \int_a^t \frac{\partial G_s}{\partial s}(X_s)\,ds + \int_a^t \frac{f_t(X_s)}{\sigma_s(X_s)}ds. \end{aligned} \tag{3.16}$$

Our objective is bounding the difference between Equation (3.15) and (3.16):

$$\begin{aligned} G_t(X_t^{(n)}) - G_t(X_t) &= \int_a^t \left\{\left[\frac{\partial G_s}{\partial s}(X_s^{(n)}) - \frac{\partial G_s}{\partial s}(X_s)\right] - \left[\frac{f_t(X_s^{(n)})}{\sigma_s(X_s^{(n)})} - \frac{f_t(X_s)}{\sigma_s(X_s)}\right]\right\} ds \\ &+ \left(\Phi_t^{(n)} - W_t\right) + \left(\Phi_a^{(n)} - W_a\right), \end{aligned} \tag{3.17}$$

but before, we need some considerations. In the sequel, K will denote a big enough constant, not necessary equal at all inequalities, that we use to simplify the notation.

(a) Because of the continuity and Lipschitz condition on $f_t(x)$, it will also satisfy linear growth, so we can develop

$$\begin{aligned}\left|\frac{f_t(x)}{\sigma_t(x)}-\frac{f_t(y)}{\sigma_t(y)}\right| &\le \left|\frac{f_t(x)}{\sigma_t(x)}-\frac{f_t(y)}{\sigma_t(x)}\right|+\left|\frac{f_t(y)}{\sigma_t(x)}-\frac{f_t(y)}{\sigma_t(y)}\right| \\ &= \left|\frac{1}{\sigma_t(x)}\right| |f_t(x)-f_t(y)| + |f_t(y)| \left|\frac{1}{\sigma_t(x)}-\frac{1}{\sigma_t(y)}\right| \\ &\le \left|\frac{1}{\sigma_t(x)}\right| K|x-y| + \left|\frac{1}{\sigma_t(x)}-\frac{1}{\sigma_t(y)}\right| K(1+|y|) \\ &\le K|x-y|(1+|y|).\end{aligned}$$

(b) As $\sigma_t^{-2}(x)\cdot\partial_t\sigma_t(x)$ is uniformly bounded by hypothesis,

$$\left|\frac{\partial G_t}{\partial t}(x)-\frac{\partial G_t}{\partial t}(y)\right| = \left|\int_y^x \frac{-1}{\sigma_t^2(z)}\frac{\partial\sigma_t}{\partial t}(z)\,dz\right| \le K|x-y|.$$

(c) We also need to prove that

$$|G_t(x)-G_t(y)| \ge K\log\left(1+\frac{|x-y|}{1+|y|}\right),$$

and to do so, we require linear growth condition on σ, which is consequence of Lipschitz on x and time continuity, so $|\sigma_t(x)|\le K(1+|x|)$. Then, we make a case-based reasoning depending on the sign of x and y.

- If they have the same sign, we can assume without loss of generality that both are non-negative. Then, by defining $u=\max(x,y)$ and $v=\min(x,y)$, we get

$$\begin{aligned}|G_t(x)-G_t(y)| &\ge K\int_v^u \frac{dz}{1+z} = K[\log(1+u)-\log(1+v)] \\ &= K\log\left(1+\frac{|u-v|}{1+v}\right) \ge K\log\left(1+\frac{|x-y|}{1+|y|}\right).\end{aligned}$$

- Otherwise, if their signs are opposite and supposing that $u\le|v|$,

$$\begin{aligned}|G_t(x)-G_t(y)| &\ge K[\log(1+u)+\log(1+|v|)] \ge K\log(1+u+|v|) \\ &= K\log(1+|x-y|) \ge K\log\left(1+\frac{|x-y|}{1+|y|}\right).\end{aligned}$$

All together, if we apply these inequalities with $x=X_t^{(n)}$ and $y=X_t$ in Equation (3.17) and define a random variable $\mu := 1+\max_{t\in[a,b]} X_t$ we have that

$$\log\left(1+\frac{\left|X_t^{(n)}-X_t\right|}{\mu}\right) \le K\left(\Phi_t^{(n)}-W_t\right)+K\left(\Phi_a^{(n)}-W_a\right)+\mu K\int_a^t \left|X_s^{(n)}-X_s\right| ds.$$

Finally, we identify $\epsilon_t^{(n)}=\epsilon(t)=\exp\{K[(\Phi_t^{(n)}-W_t)+(\Phi_a^{(n)}-W_a)]\}$, in order to apply Lemma 3.14 and proving that $X_t^{(n)}\to X_t$ as $n\to\infty$ a.s. To do so, it is required to make use of *dominated convergence theorem* to see that $\int_a^b \epsilon_t^{(n)}$ and $\epsilon_t^{(n)} \to 0$ as $n \to \infty$, supported by the bounding of $\Phi_t^{(n)}$ (remember condition (**C2**)) and W_t (it is a continuous function defined on the interval $[a,b]$). In addition, under (**C4**), for almost all samples, the convergence $X_t^{(n)} \to X_t$ is uniform in $[a,b]$.

□

3.5.1 Numerical simulation II

With the same spirit we tried to capture the differences between Itô's and Stratonovich's stochastic integrals numerically, we now want to plot the distinct behaviour of solutions of SDEs when we take one approach or the other. To do that, we focus on Example 3.13 about the noisy population model expecting to recover the analytic results but also estimating the solution of this SDE by means of Wong-Zakai theorem.

Before showing the plots, we describe briefly the algorithm used to solve the SDE numerically: *Euler-Maruyama method*. This basically consists of approximating a general SDE given as

$$dX_t = \sigma_t(X_t)dW_t + f_t(X_t)dt,$$

by

$$X_{t_j} = X_{t_{j-1}} + g_{t_{j-1}}(X_{t_{j-1}})\Delta_j W + f_{t_{j-1}}(X_{t_{j-1}})\Delta_j t,$$

with $j = 1, \ldots, n$ a partition of the interval of work. To justify it, one has to appeal the corresponding SIE and estimate

$$\int_{t_{j-1}}^{t_j} g_t\left(X_t\right) dW_t \approx g_{t_{j-1}}(X_{t_{j-1}})\Delta_j W \quad \text{and} \quad \int_{t_{j-1}}^{t_j} f_t\left(X_t\right) dt \approx f_{t_{j-1}}(X_{t_{j-1}})\Delta_j t,$$

where $\Delta_j W = \gamma_j \sqrt{\Delta_j t}$, and γ_j is chosen from $N\left(0,1\right)$. This is a modification of the classical Euler method to solve ordinary differential equations to deal with an extra term that does not increase $\Delta_j t$ but $\sqrt{\Delta_j t}$. Notice that this lets us solve both Itô's and Stratonovich's differential equations because we can rewrite the latter in form of Itô due to the conversion formula in Equation (3.9). Furthermore, if we are interested in observing the convergence of solutions of ordinary to stochastic differential equations, we recover the picture in Subsection 2.4.1 to interpolate points using spline functions, then make its numerical derivation to end up with the standard Euler method to solve the differential equation numerically.

We can even get more juice to this example trying to reproduce the expectation of each description by computing the mean of several simulations. This two quantities should be the same a.s. due to the Strong Law of Large Numbers[1].

[1] It is collected in the appendix, in the part concerning probability notions, subsection A.1.

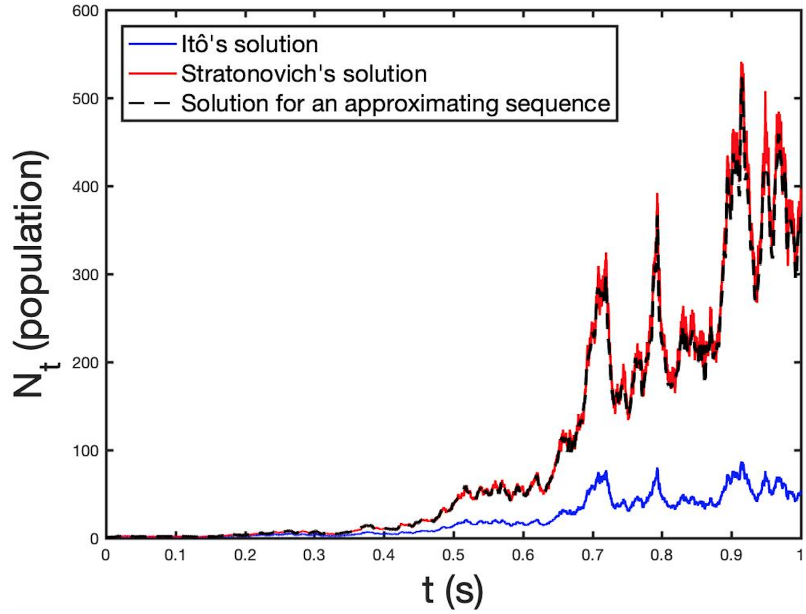

Figure 3.1: Illustrative plot that shows how the approximating successions of solutions of ODEs tend to the corresponding solution of SDE under Stratonovich's view. This is particularized for the example of the growth of a population N_t under uncertainty with parameters $\alpha = 6$, $\beta = 2$, $N_0 = 1$. The number of intermediate steps in time is 10000.

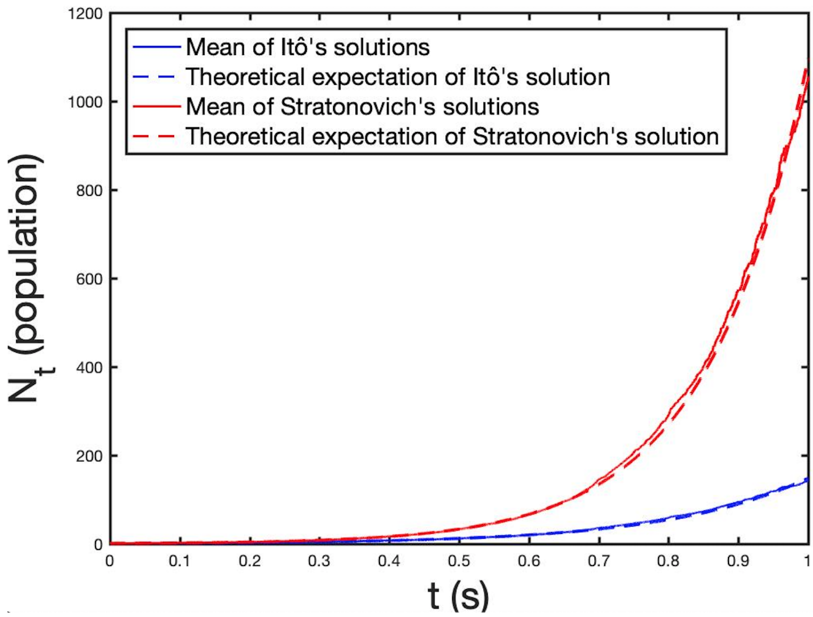

Figure 3.2: The solid curves are the mean of 5000 different curves, and the theoretical dashed lines are the corresponding to the expressions we deduced in Example 3.13. For both descriptions they fit well, which is coherent with the Law of Large Numbers.

Conclusions

With this project I had the opportunity of learning the fundamentals of stochastic calculus up to point of applying them to model different phenomena related to Physics or Ecology. Although I just had minute knowledge of stochastic processes before starting, thanks to my tutor and the books I read, I really enjoyed initiating myself in such an uprising field.

I began by providing the basic definitions related to stochastic processes, and we quickly immersed in the first main concept we would deal with during all the work: Brownian motion. Here, we focused on proving some path properties that make it special, basically the non-differentiability a.e. and the value of its quadratic variation. We also defined white noise, connecting it with the derivative of the Brownian motion considered as a generalized stochastic process.

Next step has been giving a rigorous construction of stochastic integrals for the two different descriptions we studied. We carefully built Itô's integral, first for step processes and later extending it for a larger set of functions, and besides, we defined Stratonovich's integral, establishing a conversion formula for both approaches. It has been taken particular interest in how choosing a different evaluating point when defining stochastic integrals can cause so different properties: Statonovich's follows deterministic calculus rules but loses good attributes such us centrality or the fact of being a martingale, as Itô's is. At that point, we discussed about the dilemma of electing the best integral, coming to the conclusion that it would strongly depend on the situation. We also realized how Itô formula, the principal tool when computing stochastic integrals, captures stochasticity with an extra term which does not appear in Newton-Leibniz chain rule.

Then, we made a brief introduction to the study of stochastic differential equations, taking illustrative examples to justify the assumptions needed to ensure existence and uniqueness of solutions. When proving this theorem, we made use of several tools studied during this project and also recovering results viewed in different subjects of the degree, which has been compelling. Again, we proposed a different way of interpreting SDEs using Stratonovich's integral, highlighting that the resulting process might have different properties.

Finally, we have also discussed and simulated Wong-Zakai theorems, a powerful result that establishes a direct connection between stochastic integrals/solutions of a SDE, and the deterministic integrals/limits of solutions of ODEs, constituting a vital instrument when modelling white noise in experimental sciences.

Moreover, while studying all these mathematics, I realized in future outlook which could be done. Beginning by rewriting all this theory for martingales and making extensions of Itô's and Stratonovich's integrals for even more kind of functions, to exhaustly examining the properties of the solutions of SDEs or studying in more detail their applications to Finance or the field of Partial Differential Equations. Analyzing the convergence of the used numerical method as well as developing other (for instance, *Milstein algorithm*) would have been interesting too. Particularly, I would have liked to deeply understand issues related with multiplicative noise and filtering problems, as they have to be with the choice of the integral we have discussed here.

Appendix A

Basic Notions

Here we present some propositions and theorems that are assumed to be known during the work, usually used to prove the main results. We do not show the demonstrations but they can be easily found in any classical book of its respective field.

A.1 Tools of probability theory

Properties of the conditional expectation

If X, Y are random variables in a probability space (Ω, F, P), and $\mathscr{G}$ is a sub-σ-field of $\mathscr{F}$, then it holds,

(a) Linearity: $\forall c_1, c_2 \in \mathbb{R}$, $E\left(c_1 X + c_2 Y|\mathscr{G}\right) = c_1 E\left(X|\mathscr{G}\right) + c_2 E\left(Y|\mathscr{G}\right)$.

(b) Monotony: $E\left(X|\mathscr{G}\right) \leq E\left(X|\mathscr{G}\right)$ when $X \leq Y$.

(c) $E\left(E\left(X|\mathscr{G}\right)\right) = E\left(X\right)$.

(d) If X is $\mathscr{G}$-measurable, $E\left(X|\mathscr{G}\right) = X$.

(e) Assuming (d) and also that X is bounded, it is true $E\left(XY|\mathscr{G}\right)$=$XE\left(Y|\mathscr{G}\right)$.

(f) If $\mathscr{G}_1, \mathscr{G}_2$ are σ-fields and $\mathscr{G}_1 \subset \mathscr{G}_2$, $E\left(E\left(X|\mathscr{G}_1\right)|\mathscr{G}_2\right) = E\left(E\left(X|\mathscr{G}_2\right)|\mathscr{G}_1\right) = E\left(X|\mathscr{G}_1\right)$.

First lemma of Borel-Cantelli

Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events in some probability space. Then,

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P\left(\limsup_{n\to\infty} A_n\right) = 0.$$

Conditional Jensen's inequality

Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a convex function and X an integrable random variable on $(\Omega, \mathscr{F}, P)$ with $f(X)$ also integrable. Then it holds a.s.,

$$f\left(E\left(X|\mathscr{G}\right)\right) \leq E\left(f(X)|\mathscr{G}\right),$$

for any σ-field $\mathscr{G}$ on Ω contained in $\mathscr{F}$.

Chebyshev's inequality

For any random variable X with finite moment of order p, we have

$$P\left(|X| \geq \lambda\right) \leq \frac{1}{\lambda^p} E\left(|X|^p\right),$$

with $p \in [1, \infty)$ and $\lambda > 0$.

Strong Law of Large Numbers

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed random variables. If $E|X_i| < \infty, \forall i$, then a.s.,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to E\left(X_i\right), \text{ as } n \to \infty.$$

A.2 Tools of analysis and measure theory

Ratio Test

Given a series of the form $\sum_{n\geq 1} a_n$ where each a_n is a real number, we consider the limit $A = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|$. Ratio test states the following:

- If $A > 1$, then the series is divergent.
- If $A < 1$, then the series converges uniformly.
- If $A = 1$, this test is inconclusive.

Real Weierstrass M-test

Consider a sequence of real functions $\{f_n\}_{n\geq 1}$ defined on $\mathbb{R}$. If there exists another sequence of positive numbers $\{M_n\}_{n\geq 1}$ with

$$|f_n(x)| \leq M_n, \ \forall n \geq 1, \forall x \in \mathbb{R}, \quad \text{with} \quad \sum_{n\geq 1} M_n < \infty,$$

then $\sum_{n\geq 1} f_n$ converges absolutely and uniformly on $\mathbb{R}$.

Results on a.e. differentiabilty of functions of bounded variation

Given a function $f : [a, b] \to \mathbb{R}$, it holds,

- If f is of bounded variation, then it can be written as the difference of two monotone functions real-valued functions defined on $[a, b]$ (the reciprocal implication is also true).
- If f is an increasing function, then f is differentiable a.e.

As a consequence, any function of bounded variation on $[a, b]$ is a.e. differentiable there.

Hölder's inequality

Given a measure space $(X, \mathscr{A}, \mu)$, if $1 \leq p, q \leq \infty$ and $p^{-1} + q^{-1} = 1$, it holds,

$$\int_X |f(x)g(x)|\, d\mu(x) \leq \left(\int_X |f(x)|^p\, d\mu(x) \right)^{1/p} \left(\int_X |g(x)|^q\, d\mu(x) \right)^{1/q}$$

for any f, g measurable functions taking real values. When $p = q = 2$, this is called *Cauchy-Schwarz inequality*.

Dominated Convergence Theorem

Given a measure space $(X, \mathscr{A}, \mu)$, we consider a sequence of $\mathscr{A}$-measurable functions $(f_n)_n$ defined on X and taking real values. If $\lim_n f_n(x) = f(x)$ a.e. x and exists $g \in L^1(\mu)$ satisfying $|f_n(x)| \leq g(x)$ a.e. x for all n, then:

$$\lim_n \int_X f_n(x) d\mu(x) = \int_X f(x) d\mu(x). \tag{A.1}$$

In particular, if we consider a succession of random variables $(X_n)_n \to X$ with $|X_n| \leq Y$ *a.s.* $\forall n$, for some Y integrable random variable, we will have,

$$E(X_n) \to E(X), \quad \text{as } n \to \infty. \tag{A.2}$$

Derivation under the integral sign

Given a measure space $(X, \mathscr{A}, \mu)$, we take $0 \leq g \in L^1(\mu)$ and $f(\cdot, t)$ an integrable function $\forall t$ in an interval of $\mathbb{R}$. Consider also

$$\Phi(t) = \int f(x, t)\, d\mu(x).$$

Then,

1. If $f(x, \cdot)$ is contunous in t_0 and $|f(x, t)| \leq g(x)$ for all $x \in X$ and for all $t \in I$, then Φ is continuous in t_0.
2. If for all $x \in X$, $f(x, \cdot)$ is derivable for all $t \in I$ and $\left|\frac{\partial f(x,t)}{\partial x}\right| \leq g(x)$ for all $x \in X$, it holds
$$\Phi'(t) = \int \frac{\partial f(x, t)}{\partial x} d\mu(x).$$

Appendix B

Matlab Script of Simulations

```
clear all
close all

% Parameters
T = 1;                          % Final time
n = 10000;                      % Steps in time
dt = T/n;                       % Time increment
time = [0:dt:T];                % Time vector

% Initialize arrays to zero to improve eficiency
dW = zeros(1,n);
W = zeros(1,n);
ito = zeros(1,n);
strat = zeros(1,n);
I = zeros(1,n);
ito_num = zeros(1,numel(time));
strat_num = zeros(1,numel(time));

% Simulation of a brownian sample path
dW = sqrt(dt)*randn(1,n);       % Increments
W = cumsum(dW);                 % Cumulative sum

%{
% Plotting the trajectory of Brownian Motion
figure(1)
plot(time,[0,W]);
xlabel('t', 'FontSize', 16)
ylabel('W(t)', 'FontSize', 16,'Rotation', 0)
%}

% SIMULATION OF THE STOCHASTIC INTEGRAL INT(W */o dW)

ito = cumsum([0,W(1:end-1)].*dW);            % Ito's integral
strat = cumsum((0.5*([0,W(1:end-1)]+...      % Stratonovich's integral
        [0,W(2:end)])).*dW);

%{
% Plotting Ito and Stratonovich integrals
```

```
figure(2)
plot(time,[0,ito]);
hold on
plot(time,[0,strat]);
xlabel('t', 'FontSize', 16)
ylabel('W_t', 'FontSize', 16,'Rotation', 0)

% Proving the conversion formula between both integrals
figure(3)
plot(time,[0,strat]);
hold on
plot(time,[0,ito]+0.5.*time)
%}

% Simulating the approximating sequence of deterministic integrals
for num = 100:200:500            % num is the number of intermediate points

    t = linspace(0,T,num);

    jump = n/num;
    for i=1:num
        y(i)=W(floor(i*jump));
    end

    % Spline interpolation
    pp = interp1(t, y, [0:dt:T], 'spline');

    % Numerical derivation
    pp_der = fstderivative(pp, dt, n+1);
    f = pp.*pp_der;

    % Comtuping the integral using Simpson 1/3
    for i=2:n+1
        I(i-1)= dt/3*(f(1)+2*sum(f(2:2:i-2))+4*sum(f(1:2:i-1))+f(i));
    end

    %{
    % Plotting the appromiting sequences of deterministic intragral
    figure(4)
    plot(time,[0,I]);
    hold on
    %}

end

% Plotting Ito, Stratonovich and the integral of the approximating sequence
figure(5)
plot(time,[0,ito], 'b', 'Linewidth', 1.5);
hold on
plot(time,[0,strat],'r', 'Linewidth', 1.5);
hold on
plot(time,[0,I],'k--', 'Linewidth', 1.5);
hold on
xlabel('t␣(s)', 'Fontsize', 24)
ylabel('Integral␣value', 'Fontsize', 24)
set (gca, 'LineWidth', 1.7)
```

```
title ('Initial␣condition', 'Fontsize', 30)
legend({'Ito''s␣integral', 'Stratonovich''s␣integral', ...
    'Integral␣of␣the␣approximating␣sequence'}, 'FontSize', ...
    15, 'Location', 'northwest')


% SIMULATION FOR THE SOLUTIONS OF SDE AND ODE, EXAMPLE OF GROWTH POPULATION

c_alpha = 6;      % Average growth constant
c_beta = 2;       % Noise intensity
init_pop = 1;     % Initial population

%{
% Plotting the analytical solution deduced in the example
figure(9)
ito_theo = init_pop*exp((c_alpha-0.5*c_beta*c_beta)*time+c_beta*[0,W]);
strat_theo = init_pop*exp(c_alpha*time+c_beta*[0,W]);
plot(time,ito_theo);
hold on
plot(time,strat_theo);
%}

% Solving Ito's SDE with Euler-Maruyama method
ito_num(1)= init_pop;
for i=2:n+1
    ito_num(i) = ito_num(i-1)+ito_num(i-1)*...
                 c_alpha*dt+c_beta*ito_num(i-1)*dW(i-1);
end

%{
% Comparing theorical and numerical solution of Ito's SDE
figure(10)
plot(time,ito_num)
hold on
plot(time,ito_theo)
%}

% Solving Stratonovich's SDE with Euler-Maruyama method
strat_num(1)= init_pop;
for i=2:n+1
    strat_num(i) = strat_num(i-1)+strat_num(i-1)*...
                   (c_alpha+0.5*c_beta*c_beta)*dt+c_beta*strat_num(i-1)*dW(i-1);
end

%{
% Comparing theorical and numerical solution of Ito's SDE
figure(11)
plot(time,strat_num)
hold on
plot(time,strat_theo)
%}

for num = 100:200:500          % num is the number of intermediate points

    s=[];
    y=[];
```

```
    t = linspace(0,T,num);

    jump = n/num;
    for i=1:num
        y(i)=W(floor(i*jump));
    end

    % Spline interpolation
    pp = interp1(t, y, time, 'spline');
    % Numerical derivation
    pp_der = fstderivative(pp, dt, n+1);

    s(1)= init_pop;
    for i=2:n+1
        s(i) = s(i-1)+(s(i-1)*(c_alpha+(c_beta*pp_der(i-1)))*dt);
    end

    %{
    % Plotting the solution of the ODE for the approximating sequences
    figure(12)
    plot(time,s);
    hold on
    %}
end


% Plotting Ito's and Stratonovich's numerical solution solution together
% with the solution for the approximating sequence
figure(13)
plot(time, ito_num, 'b', 'LineWidth', 1.5);
hold on
plot(time, strat_num, 'r', 'LineWidth', 1.5);
hold on
plot(time, s, 'k--', 'LineWidth', 1.5);
xlabel('t (s)', 'Fontsize', 24)
ylabel('N_t (population)', 'Fontsize', 24)
set (gca, 'LineWidth', 1.7)
title ('Initial condition', 'Fontsize', 30)
legend({'Ito''s solution', 'Stratonovich''s solution', ...
    'Solution for an approximating sequence'}, 'FontSize', ...
    15, 'Location', 'northwest')

% SIMULATING STRONG LAW OF LARGE NUMBERS

nc = 50;                    % Number of curves to compute the average

% Renitializing arrays
dW = zeros(nc,n);
W = zeros(nc,n);

for i=1:nc

    dW(i,:)= sqrt(dt)*randn(1,n);
    W(i,:) = cumsum(dW(i,:));

    ito_theo(i,:) = init_pop*exp((c_alpha-0.5*c_beta*c_beta)*...
```

```
                    time+c_beta*[0,W(i,:)]);
    strat_theo(i,:) = init_pop*exp(c_alpha*time+c_beta*[0,W(i,:)]);

end

% Computing the mean for nc theoretical solutions of the SDE
s_ito = mean(ito_theo);
s_strat = mean(strat_theo);

% Plotting the mean together with the expectation deduced by the example
figure (16)
plot(time, s_ito, 'b', 'LineWidth', 1.5);
hold on
plot(time, init_pop*exp(time*c_alpha), 'b--', 'LineWidth', 1.5);
hold on
plot(time, s_strat, 'r', 'LineWidth', 1.5);
hold on
plot(time, init_pop*exp(time*(c_alpha+0.5*c_beta^2)), 'r--', 'LineWidth', 1.5);
xlabel('t (s)', 'Fontsize', 24)
ylabel('N_t (population)', 'Fontsize', 24)
set (gca, 'LineWidth', 1.7)
title ('Initial condition', 'Fontsize', 30)
legend({'Mean of Ito''s solutions', ...
    'Theoretical expectation of Ito''s solution', ...
    'Mean of Stratonovich''s solutions', ...
    'Theoretical expectation of Stratonovich''s solution'}, ...
    'FontSize', 15, 'Location', 'northwest')

% FUNCTION TO COMPUTE THE FIRST DERIVATIVE (of time) USING CENTERED FINITE
% DIFERENCES (error Oh^4)

function fsdx = fstderivative (f, h, t)

    fsdx = zeros(1,t);

    fsdx(1) = (-3*f(1)+4*f(2)-f(3))/(2*h);
    fsdx(2) = (-3*f(2)+4*f(3)-f(4))/(2*h);

    for i=3:(t-2)
        fsdx(i) = (f(i-2)-8*f(i-1)+8*f(i+1)-f(i+2))/(12*h);
    end

    fsdx(t-1) = (3*f(t-1)-4*f(t-2)+f(t-3))/(2*h);
    fsdx(t) = (3*f(t)-4*f(t-1)+f(t-2))/(2*h);

end
```

Bibliography

[1] L. Arnold, *Stochastic Differential Equations. Theory and Applications*, Willey & Sons, 1974.

[2] Z. Brzezniak, T. Zastawniak, *Basic Stochastic Processes*, Springer, 1999.

[3] L.C. Evans, *An Introduction to Stochastic Differential Equations*, American Mathematical Society, 2013.

[4] H.H. Kuo, *Introduction to Stochastic Integration*, Springer, 2006.

[5] D. Nualart, *Stochastic Processes*, Lecture notes of a course in The University of Kansas, [http://nualart.faculty.ku.edu/]

[6] B. Øksendal, *Stochastic Differential Equations. An Introduction with Applications*, Springer Verlag, 2007.

[7] H.L. Royden, P.M. Fitzpatrick, *Real Analysis*, Pearson, 2010.

[8] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill Inc, 1964.

[9] M. Sanz-Solé, *An Introduction to Stochastic Calculus*, Lecture notes in the Master Degree in Advanced Mathematics, Universitat de Barcelona, 2017. [http://www.ub.edu/plie/Sanz-Sole/]

[10] L. Smith, *Itô and Stratonovich, a guide for the perplexed*, 2018. [http://www.robots.ox.ac.uk/ lsgs/posts/2018-09-30-ito-strat.html]

[11] R. L. Stratonovich, *A new representation for stochastic integrals and equations*, J. Siam Control, **4** (1966), 362-371.

[12] R. Toral, P. Colet, *Stochastic Numerical Methods. An Introduction for Students and Scientists*, Wiley-VCH, 2014.

[13] E. Wong, M. Zakai, *On the convergence of ordinary integrals to stochastic integrals*, Ann. Math. Stat., **36** (1965) 1560-1564.