

# Introduction to Stochastic Calculus

Math 545



# Contents

Chapter 1. Introduction	5
1. Motivations	5
2. Outline For a Course	6
Chapter 2. Probabilistic Background	9
1. Countable probability spaces	9
2. Uncountable Probability Spaces	11
3. Distributions and Convergence of Random Variables	18
Chapter 3. Wiener Process and Stochastic Processes	21
1. An Illustrative Example: A Collection of Random Walks	21
2. General Stochastic Processes	22
3. Definition of Brownian motion (Wiener Process)	23
4. Constructive Approach to Brownian motion	24
5. Brownian motion has Rough Trajectories	26
6. More Properties of Random Walks	28
7. More Properties of General Stochastic Processes	29
8. A glimpse of the connection with PDES	34
Chapter 4. Itô Integrals	37
1. Properties of the noise Suggested by Modeling	37
2. Riemann-Stieltjes Integral	38
3. A motivating example	39
4. Itô integrals for a simple class of step functions	40
5. Extension to the Closure of Elementary Processes	44
6. Properties of Itô integrals	46
7. A continuous in time version of the the Itô integral	47
8. An Extension of the Itô Integral	48
9. Itô Processes	49
Chapter 5. Stochastic Calculus	51
1. Itô's Formula for Brownian motion	51
2. Quadratic Variation and Covariation	54
3. Itô's Formula for an Itô Process	58
4. Full Multidimensional Version of Itô Formula	60
5. Collection of the Formal Rules for Itô's Formula and Quadratic Variation	64
Chapter 6. Stochastic Differential Equations	67
1. Definitions	67
2. Examples of SDEs	68
3. Existence and Uniqueness for SDEs	71
4. Weak solutions to SDEs	74

5. Markov property of Itô diffusions	76
Chapter 7. PDEs and SDEs: The connection	77
1. Infinitesimal generators	77
2. Martingales associated with diffusion processes	79
3. Connection with PDEs	81
4. Time-homogeneous Diffusions	84
5. Stochastic Characteristics	85
6. A fundamental example: Brownian motion and the Heat Equation	86
Chapter 8. Martingales and Localization	89
1. Martingales & Co.	89
2. Optional stopping	91
3. Localization	92
4. Quadratic variation for martingales	95
5. Lévy-Doob Characterization of Brownian-Motion	97
6. Random time changes	99
7. Time Change for an SDE	99
8. Martingale representation theorem	100
9. Martingale inequalities	100
Chapter 9. Girsanov's Theorem	103
1. An illustrative example	103
2. Tilted Brownian motion	104
3. Shifting Wiener Measure Simple Cameron Martin Theorem	105
4. Girsanov's Theorem for SDES	106
Bibliography	109
Appendix A. Some Results from Analysis	111

## CHAPTER 1

# Introduction

### 1. Motivations

EVOLUTIONS IN TIME WITH RANDOM INFLUENCES/RANDOM DYNAMICS. Let  $N(t)$  be the “number of rabbits in some population” or “the price of a stock”. Then one might want to make a model of the dynamics which includes “random influences”. A (very) simple example is

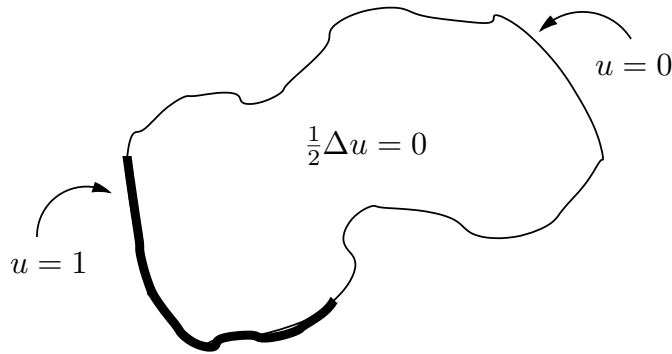
$$\frac{dN(t)}{dt} = a(t)N(t) \text{ where } a(t) = r(t) + \text{“noise”} . \quad (1.1)$$

Making sense of “noise” and learning how to make calculations with it is one of the principal objectives of this course. This will allow us predict, in a probabilistic sense, the behavior of  $N(t)$ . Examples of situations like the one introduced above are ubiquitous in nature:

- i) **The gambler’s ruin problem** We play the following game: We start with 3\$ in our pocket and we flip a coin. If the result is tail we loose one dollar, while if the result is positive we win one dollar. We stop when we have no money to bargain, or when we reach 9\$. We may ask: what is the probability that I end up broke?
- ii) **Population dynamics/Infectious diseases** As anticipated, (1.1) can be used to model the evolution in the number of rabbits in some population. Similar models are used to model the number of genetic mutations an animal species. We may also think about  $N(t)$  as the number of sick individuals in a population. Reasonable and widely applied models for the spread of infectious diseases are obtained by modifying (1.1), and observing its behavior. In all these cases, one may be interested in knowing if it is likely for the disease/mutation to take over the population, or rather to go extinct.
- iii) **Stock prices** We may think about a set of  $M$  risky investments (e.g. a stock), where the price  $N_i(t)$  for  $i \in \{1, \dots, M\}$  per unit at time  $t$  evolves according to (1.1). In this case, one would like to optimize his/her choice of stocks to maximize the total value  $\sum_{i=1}^M \alpha_i N_i(t)$  at a later time  $T$ .

CONNECTIONS WITH DIFFUSION THEORY AND PDES. There exists a deep connection between noisy processes such as the one introduced above and the deterministic theory of partial differential equations. This starling connection will be explored and expanded upon during the course, but we anticipate some examples below:

- i) **Dirichlet problem** Let  $u(x)$  be the solution to the PDE given below with the noted boundary conditions. Here  $\Delta = \frac{\partial}{\partial x} + \frac{\partial}{\partial y}$ . The amazing fact is the following: If we start a Brownian motion diffusing from a point  $(x_0, y_0)$  inside the domain then the probability that it first hits the boundary in the darker region is given by  $u(x_0, y_0)$ .
- ii) **Black Scholes Equation** Suppose that at time  $t = 0$  the person in *iii*) is offered the right (without obligation) to buy one unit of the risky asset at a specified price  $S$  and at a specified future time  $t = T$ . Such a right is called a *European call option*. How much should the person be willing to pay for such an option? This question can be answered by solving the famous Black Scholes equation, giving for any stock price  $N(t)$  the righth value  $S$  of the European option.



## 2. Outline For a Course

What follows is a rough outline of the class, giving a good indication of the topics to be covered, though there will be modifications.

- i) Week 1: Motivation and Introduction to Stochastic Process
  - (a) Motivating Examples: Random Walks, Population Model with noise, Black-Scholes, Dirichlet problems
  - (b) Themes: Direct calculation with stochastic calculus, connections with PDES
  - (c) Introduction: Probability Spaces, Expectations,  $\sigma$ -algebras, Conditional expectations, Random walks and discrete time stochastic processes. Continuous time stochastic processes and characterization of the law of a process by its finite dimensional distributions (Kolmogorov Extension Theorem). Markov Process and Martingales.
- ii) Week 2-3: Brownian motion and its Properties
  - (a) Definitions of Brownian motion (BM) as a continuous Gaussian process with independent increments. Chapman-Kolmogorov equation, forward and backward Kolmogorov equations for BM. Continuity of sample paths (Kolmogorov Continuity Theorem). BM and more Markov process and Martingales.
  - (b) First and second variation (a.k.a variation and quadratic variation) Application to BM.
- iii) Week 4: Stochastic Integrals
  - (a) The Riemann-Stieltjes integral. Why can't we use it ?
  - (b) Building the Itô and Stratonovich integrals. (Making sense of " $\int_0^t \sigma dB$ .")
  - (c) Standard properties of integrals hold: linearity, additivity
  - (d) Itô isometry:  $\mathbb{E}(\int f dB)^2 = \mathbb{E} \int f^2 ds$ .
- iv) Week 5: Itô's Formula and Applications
  - (a) Change of variable.
  - (b) Connections with PDES and the Backward Kolmogorov equation.
- v) Week 6: Stochastic Differential Equations
  - (a) What does it mean to solve an SDE ?
  - (b) Existence of solutions (Picard iteration). Uniqueness of solutions.
- vi) Week 7: Stopping Times
  - (a) Definition.  $\sigma$ -algebra associated to stopping time. Bounded stopping times. Doob's optional stopping theorem.
  - (b) Dirichlet Problems and hitting probabilities.
  - (c) Localization via stopping times
- vii) Week 8: Levy-Doob theorem and Girsanov's Theorem
  - (a) How to tell when a continuous martingale is a Brownian motion.

- (b) Random time changes to turn a Martingale into a Brownian Motion
- (c) Hermite Polynomials and the exponential martingale
- (d) Girsanov's Theorem, Cameron-Martin formula, and changes of measure.
  - (1) The simple example of i.i.d Gaussian random variables shifted.
  - (2) Idea of Importance sampling and how to sample from tails.
  - (3) The shift of a Brownian motion
  - (4) Changing the drift in a diffusion.
- viii) Week 9: SDEs with Jumps
- ix) Week 10: Feller Theory of one dimensional diffusions
  - (a) Speed measures, natural scales, the classification of boundary point.





## CHAPTER 2

### Probabilistic Background

EXAMPLE 0.1. We begin with the following motivating example. Consider a random sequence  $\omega = \{\omega_i\}_{i=0}^N$  where

$$\omega_i = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

independent of the other  $\omega$ 's. We will also write this as

$$\mathbb{P}[(\omega_1, \omega_2, \dots, \omega_N) = (s_1, s_2, \dots, s_N)] = \frac{1}{2^N}$$

for  $s_i = \pm 1$ . We can group the possible outcomes into subsets depending on the value of  $\omega_1$ :

$$F_1 = \{\{\omega \in \Omega : \omega_1 = 1\}, \{\omega \in \Omega : \omega_1 = -1\}\}. \quad (2.1)$$

These subsets contain the possible events that we may observe knowing the value of  $\omega_1$ . In other words, this separation represents the information we have on the process at that point.

Let  $\Omega$  be the set of all such sequences of length  $N$  (i.e.  $\Omega = \{-1, 1\}^N$ ), and consider now the sequence of functions  $\{X_n : \Omega \rightarrow \mathbb{Z}\}$  where

$$\begin{aligned} X_0(\omega) &= 0 \\ X_n(\omega) &= \sum_{i=1}^n \omega_i \end{aligned} \quad (2.2)$$

for  $n \in \{1, \dots, N\}$ . Each  $X_n$  is an example of an integer-valued random variable. The set  $\Omega$  is called the sample space (or outcome space) and the measure given by  $\mathbb{P}$  on  $\Omega$  is called the probability measure.

We now recall some basic definitions from the theory of probability which will allow us to put this example on solid ground.

Intuitively, each  $\omega \in \Omega$  is a possible outcome of all of the randomness in our system. The measure  $\mathbb{P}$  gives the chance of the various outcomes. In the above setting where the outcome space  $\Omega$  consists of a finite number of elements, we are able to define everything in a straightforward way. We begin with a quick recalling of a number of definitions in the countably infinite (possibly finite) setting.

#### 1. Countable probability spaces

If  $\Omega$  is countable it is enough to define the probability of each element in  $\Omega$ . That is to say give a function  $p: \Omega \rightarrow [0, 1]$  with  $\sum_{\omega \in \Omega} p(\omega) = 1$  and define

$$\mathbb{P}[\omega] = p(\omega)$$

for each  $\omega \in \Omega$ . An event  $A$  is just a subset of  $\Omega$ . We naturally extend the definition of  $\mathbb{P}$  to an event  $A$  by

$$\mathbb{P}[A] := \sum_{\omega \in A} P[\omega].$$

Observe that this definition has a number of consequences. In particular, if  $A_i$  are disjoint events, that is to say  $A_i \subset \Omega$  and  $A_i \cap A_k = \emptyset$  if  $i \neq j$  then

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$$

and if  $A^c := \{\omega \in \Omega \text{ with } \omega \notin A\}$  is the compliment of  $A$  then  $\mathbb{P}[A] = 1 - \mathbb{P}[A^c]$ .

Given two event  $A$  and  $B$ , the *conditional probability* of  $A$  given  $B$  is defined by

$$\mathbb{P}[A|B] := \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \quad (2.3)$$

For fixed  $B$ , this is just a new probability measure  $\mathbb{P}[\cdot | B]$  on  $\Omega$  which gives probability  $\mathbb{P}[\omega|B]$  to the outcome  $\omega \in \Omega$ .

A real-valued *random variable*  $X$  is simply a real-valued function of  $\Omega$ . Similarly, a random variable taking values in some set  $\mathbf{X}$  is a function  $X: \Omega \rightarrow \mathbf{X}$ . We can then define the *expected value* of a random variable  $X$  (or simply the expectation of  $X$ ) as

$$\mathbb{E}[X] := \sum_{x \in \text{Range}(X)} x \mathbb{P}[X = x] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}[\omega]$$

Here we have used the convention that  $\{X = x\}$  is short hand for  $\{\omega \in \Omega : X(\omega) = x\}$  and the definition of  $\text{Range}(X) = \{x \in \mathbf{X} : \exists \omega, \text{ with } X(\omega) = x\} = X^{-1}(\Omega)$ .

Two events  $A$  and  $B$  are *independent* if  $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$ . Two random variable are independent if  $\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x]\mathbb{P}[Y = y]$ . Of course this implies that for any events  $A$  and  $B$  that  $\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B]$  and that  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ . A collection of events  $A_i$  is said to be *mutually independent* if

$$\mathbb{P}[A_1 \cap \dots \cap A_n] = \prod_{i=1}^n \mathbb{P}[A_i].$$

Similarly a collection of random variable  $X_i$  are mutually independent if for any collection of sets from their range  $A_i$  one has that the collection of events  $\{X_i \in A_i\}$  are mutually independent. As before, as a consequence one has that

$$\mathbb{E}[X_1 \dots X_n] = \prod_{i=1}^n \mathbb{E}[X_i].$$

Given two  $\mathbf{X}$ -valued random variables  $Y$  and  $Z$ , for any  $z \in \text{Range}(Z)$  we define the *conditional expectation* of  $Y$  given  $\{Z = z\}$  as

$$\mathbb{E}[Y|Z = z] := \sum_{y \in \text{Range}(Y)} y \mathbb{P}[Y = y|Z = z] \quad (2.4)$$

Which is to say that  $\mathbb{E}[Y|Z = z]$  is just the expected value of  $Y$  under the probability measure which is given by  $P[\cdot | Z = z]$ .

In general, for any event  $A$  we can define the conditional expectation of  $Y$  given  $A$  as

$$\mathbb{E}[Y|A] := \sum_{y \in \text{Range}(Y)} y \mathbb{P}[Y = y | A] \quad (2.5)$$

We can extend the definition  $\mathbb{E}[Y|Z = z]$  to  $\mathbb{E}[Y|Z]$  which we understand to be a function of  $Z$  which takes the value  $\mathbb{E}[Y|Z = z]$  when  $Z = z$ . More formally  $\mathbb{E}[Y|Z] := h(Z)$  where  $h: \text{Range}(Z) \rightarrow \mathbf{X}$  given by  $h(z) = \mathbb{E}[Y|Z = z]$ .

EXAMPLE 1.1. *Let us compute an example of conditional probability. Using the notation from Example 0.1 we see*

$$\begin{aligned}\mathbb{E}[(X_3)^2|X_2 = 2] &= \sum_{i \in \mathbb{N}} i \mathbb{P}[(X_3)^2 = i|X_2 = 2] \\ &= (1)^2 \mathbb{P}[X_3 = 1|X_2 = 2] + (3)^2 \mathbb{P}[X_3 = 3|X_2 = 2] = 5\end{aligned}$$

Of course,  $X_2$  can also take the value  $-2$  and  $0$ . For these values of  $X_2$  we have

$$\begin{aligned}\mathbb{E}[(X_3)^2|X_2 = -2] &= (-1)^2 \mathbb{P}[X_3 = -1|X_2 = -2] + (-3)^2 \mathbb{P}[X_3 = -3|X_2 = -2] = 5 \\ \mathbb{E}[(X_3)^2|X_2 = 0] &= (-1)^2 \mathbb{P}[X_3 = -1|X_2 = 0] + (1)^2 \mathbb{P}[X_3 = 1|X_2 = 0] = 1\end{aligned}$$

Hence  $\mathbb{E}[(X_3)^2|X_2 = -2] = h(X_2)$  where

$$h(x) = \begin{cases} 5 & \text{if } x = \pm 2 \\ 1 & \text{if } x = 0 \end{cases}$$

By clever rearrangement one does not always have to calculate the function  $\mathbb{E}[Y|Z]$  so explicitly. Consider the following examples.

$$\begin{aligned}\mathbb{E}[X_7|X_6] &= \mathbb{E}\left[\sum_{i=1}^7 \omega_i|X_6\right] = \mathbb{E}\left[\sum_{i=1}^6 \omega_i + \omega_7|X_6\right] \\ &= \mathbb{E}[X_6 + \omega_7|X_6] = \mathbb{E}[X_6|\mathcal{F}_6] + \mathbb{E}[\omega_7|\mathcal{F}_6] \\ &= X_6 + \mathbb{E}[\omega_7] = X_6\end{aligned}$$

since  $\mathbb{E}[\omega_7] = 0$ . We can also do a similar calculation for the previous example.

$$\begin{aligned}\mathbb{E}[X_3^2|X_2] &= \mathbb{E}[(X_2 + \omega_3)^2|X_2] = \mathbb{E}[X_2^2 + 2\omega_3 X_2 + \omega_3^2|X_2] \\ &= \mathbb{E}[X_2^2] + 2\mathbb{E}[\omega_3]\mathbb{E}[X_2] + \mathbb{E}[\omega_3^2] = X_2^2 + 1\end{aligned}$$

since  $\mathbb{E}[\omega_3] = 0$  and  $\mathbb{E}[\omega_3^2] = 1$ . Compare this to the definition of  $h$  given above.

## 2. Uncountable Probability Spaces

We will need to consider  $\Omega$  which have uncountably many points. An example is a uniform point drawn from  $[0, 1]$ . To handle such settings completely rigorously we need ideas from basic measure theory. However if one is willing to except a few formal rules of manipulation, we can proceed with learning basic stochastic calculus without needing to distract ourselves with too much measure theory.

A  $\mathbb{R}$ -valued random variable  $X$  is called a *continuous random variable* if there exists a density function  $\rho: \mathbb{R} \rightarrow \mathbb{R}$  so that

$$\mathbb{P}[X \in [a, b]] = \int_a^b \rho(x) dx$$

for any  $[a, b] \subset \mathbb{R}$ . More generally a  $\mathbb{R}^n$ -valued random variable  $X$  is called a *continuous random variable* if there exists a density function  $\rho: \mathbb{R}^n \rightarrow \mathbb{R}$  so that

$$\mathbb{P}[X \in [a, b]] = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \rho(x_1, \dots, x_n) dx_1 \cdots dx_n = \int_{[a, b]} \rho(x) dx = \int_{[a, b]} \rho(x) \text{Leb}(dx)$$

for any  $[a, b] = \prod [a_i, b_i] \subset \mathbb{R}^n$ . The last two expressions are just different ways of writing the same thing. Here we have introduced the notation  $\text{Leb}(dx)$  for the standard Lebesgue measure on  $\mathbb{R}^n$  given by  $dx_1 \cdots dx_n$ .

Analogously to the countable case we define the expectation of a continuous random variable with density  $\rho$  by

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^n} h(x)\rho(x)dx.$$

DEFINITION 2.1. A real-valued random variable  $X$  is Gaussian with mean  $\mu$  and variance  $\sigma^2$  if

$$\mathbb{P}[X \in A] = \int_A \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

If a random variable has this distribution we will write  $X \sim \mathcal{N}(\mu, \sigma)$ .

If  $X$  and  $Y$  are  $\mathbb{R}^n$ -valued and  $\mathbb{R}^m$ -valued random variables, respectively, then the vector  $(X, Y)$  is again a continuous  $\mathbb{R}^{n+m}$ -valued random variable which has a density which is called the joint density of  $X$  and  $Y$ . If  $Y$  has density  $\rho_Y$  and  $\rho_{XY}$  is the joint density of  $X$  and  $Y$  we can define

$$\mathbb{P}[X \in A | Y = y] = \int_A \frac{\rho_{XY}(x, y)}{\rho_Y(y)} dx.$$

Hence  $X$  given  $Y = y$  is a new continuous random variable with density  $x \mapsto \frac{\rho_{XY}(x, y)}{\rho_Y(y)}$  for a fixed  $y$ . The conditional expectation is defined using this density.

While many calculations can be handled satisfactorily at this level, we will soon see that we need to consider random variables on much more complicated spaces such as the space of real-valued continuous functions on the time interval  $[0, T]$  which will be denoted  $C([0, T]; \mathbb{R})$ . To give all of the details in such a setting would require a level of technical detail which we do not wish to enter into on our first visit to the subject of stochastic calculus. If one is willing to “suspend a little disbelief” one can learn the formal rule of manipulation, much as one did when one first learned regular calculus. The technical details are important but better appreciated after one first has the big picture.

**2.1. Sigma Algebras.** To this end, we will introduce the idea of a sigma algebra (usually written  $\sigma$ -algebra or  $\sigma$ -field in [Klebaner]). In Section 1, we defined our probability measures by beginning with assigning a probability to each  $\omega \in \Omega$ . This was fine when  $\Omega$  was finite or countably infinite. When  $\Omega = [0, 1]$  as in the case of picking a uniform point from the unit interval, the probability of any given point must be zero. Otherwise the sum of all of the probabilities would be  $\infty$  since there are infinitely many points and each of them has the same probability as no point is more or less likely than another. Formally,

$$\sum_{x \in [0, 1]} \mathbb{P}[\{x\}] = \infty \quad \text{if} \quad \mathbb{P}[\{x\}] > 0.$$

This is only the tip of the iceberg. There are many more complicated issues. The solution is to fix a collection of subsets of  $\Omega$  about which we are “allowed” to ask “what is the probability of this event?”. We will be able to make this collection of subsets very large, but it will *not*, in general, contain *all* of the subsets of  $\Omega$  in situations where  $\Omega$  is uncountable. This collection of subsets is called the  $\sigma$ -algebra. The triplet  $(\Omega, \mathcal{F}, \mathbb{P})$  of an outcome space  $\Omega$ , a probability measure  $\mathbb{P}$  and a  $\sigma$ -algebra  $\mathcal{F}$  is called a *Probability Space*. For any event  $A \in \mathcal{F}$ , the “probability of this event happening” is well defined and equal to  $\mathbb{P}[A]$ . A subset of  $\Omega$  which is not in  $\mathcal{F}$  might not have a well defined probability. Essentially all of the event you will think of naturally will be in the  $\sigma$ -algebra with which we will work. In light of this, it is reasonable to ask why we bring them up

at all. It turns out that  $\sigma$ -algebras are a useful way to “encode the information” contained in a collection of events or random variables. This idea and notation is used in many different contexts. If you want to be able to read the literature, it is useful to have an operational understanding of  $\sigma$ -algebras without entering into the technical detail.

Before attempting to convey any intuition or operational knowledge about  $\sigma$ -algebras we give the formal definitions since they are short (even if unenlightening).

DEFINITION 2.2. *Given a set  $\Omega$ , a  $\sigma$ -algebra  $\mathcal{F}$  is a collection of subsets of  $\Omega$  such that*

- i)  $\Omega \in \mathcal{F}$
- ii)  $A \in \mathcal{F} \implies A^c = \Omega \setminus A \in \mathcal{F}$
- iii) *Given  $\{A_n\}$  a countable collection of subsets of  $\mathcal{F}$ , we have  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ .*

*In this case the pair  $(\Omega, \mathcal{F})$  are referred to as a measurable space*

*For us a  $\sigma$ -algebra is the embodiment of information.*

EXAMPLE 2.3. *If  $\Omega = \mathbb{R}^n$  or any subset of it, we talk about the Borel  $\sigma$ -algebra as the  $\sigma$ -algebra generated by all of the intervals  $[a, b]$  with  $a, b \in \Omega$ . This  $\sigma$ -algebra contains essentially any event you would think about in most all reasonable problems. Using  $(a, b)$ , or  $[a, b)$  or  $(a, b]$  or some mixture of them makes no difference.*

Given any collection of subsets  $\mathcal{G}$  of  $\Omega$  we can talk about the “ $\sigma$ -algebra generated by  $\mathcal{G}$ ” as simply what we get by taking all of the elements of  $\mathcal{G}$  and exhaustively applying all of the operations listed above in the definition of a  $\sigma$ -algebra. More formally,

DEFINITION 2.4. *Given  $\Omega$  and  $F$  a collection of subsets of  $\Omega$ ,  $\sigma(F)$  is the  $\sigma$ -algebra generated by  $F$ . This is defined as the smallest (in terms of numbers of sets)  $\sigma$ -algebra which contains  $F$ . Intuitively  $\sigma(F)$  represents all of the probability data contained in  $F$ .*

EXAMPLE 2.5 (Example 0.1 continued). *In (2.1) we defined the set  $F_1$  as the collection of sets of possible outcomes fixing  $\omega_1$ . This collection of sets generates a  $\sigma$ -algebra on  $\Omega$ , given by*

$$\mathcal{F}_1 := \{\emptyset, \Omega, \{\omega \in \Omega : \omega_1 = 1\}, \{\omega \in \Omega : \omega_1 = -1\}\}, \quad (2.6)$$

*representing the information we have on the process knowing  $\omega_1$ .*

To complete our measurable space  $(\Omega, \mathcal{F})$  into a probability space we need to add a probability measure. Since we will not build our measure from its definition on individual  $\omega \in \Omega$  as we did in Section 1 we will instead assume that it satisfies certain reasonable properties which follow from this construction in the countable or finite case. The fact that the following assumptions is all that is needed would be covered in a measure theoretical probability or analysis class.

DEFINITION 2.6. *A measure  $\mathbb{P}$  on a measurable space  $(\Omega, \mathcal{F})$  is a probability measure if*

- i)  $\mathbb{P}[\Omega] = 1$ ,
- ii)  $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$  for all  $A \in \mathcal{F}$ .
- iii) *Given  $\{A_i\}$  a finite collection of pairwise disjoint sets in  $\mathcal{F}$ ,  $\mathbb{P}[\bigcup_{i=1}^n A_i] = \sum_{i=1}^n \mathbb{P}[A_i]$ ,*

*In this case the triplet  $(\Omega, \mathcal{F}, \mathbb{P})$  is referred to as a probability space*

DEFINITION 2.7. *If  $(\Omega, \mathcal{F})$  and  $(\mathbf{X}, \mathcal{B})$  are measurable spaces, then  $\xi : \Omega \rightarrow \mathbf{X}$  is a  $\mathbf{X}$ -valued random variable if for all  $B \in \mathcal{B}$  we have  $\xi^{-1}(B) \in \mathcal{F}$ . In other words, admissible events in  $\mathbf{X}$  get mapped to admissible events in  $\Omega$ .*

Given any events  $A$  and  $B$  in  $\mathcal{F}$ , we define the conditional probability just as before, namely

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Given real-valued random variable  $X$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , we define the expected value of  $X$  in a way analogous to before:

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega).$$

We will take for granted that this integral makes sense. However, it follows from the general theory of measure spaces.

DEFINITION 2.8. *Given a random variable on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  taking values in a measurable space  $(\mathbf{X}, \mathcal{B})$ , we define the  $\sigma$ -algebra generated by the random variable  $X$  as*

$$\sigma(X) = \sigma(\{X^{-1}(B) | B \in \mathcal{B}\}).$$

The idea is that  $\sigma(X)$  contains all of the information contained in  $X$ . If an event is in  $\sigma(X)$  then whether this event happens or not is completely determined by knowing the value of the random variable  $X$ .

EXAMPLE 2.9 (Example 0.1 continued). *By definition (2.2), since  $X_1 = \omega_1$  the  $\sigma$ -algebra generated by the random variable  $X_1$  is  $\sigma(X_1) = \mathcal{F}_1$  from (2.6). However, the  $\sigma$ -algebra generated by  $X_2 = \omega_1 + \omega_2$  is given by*

$$\sigma(X_2) = \{\emptyset, \Omega, \{\omega \in \Omega : (\omega_1, \omega_2) = (1, 1)\}, \{\omega \in \Omega : (\omega_1, \omega_2) = (-1, -1)\}, \{\omega \in \Omega : \omega_1 + \omega_2 = 0\}\}.$$

*Note that this  $\sigma$ -algebra is different than  $\mathcal{F}_2 = \sigma(\{\{\omega \in \Omega : (\omega_1, \omega_2) = (s_1, s_2)\} : s_1, s_2 \in \{-1, 1\}\})$ . Indeed, knowing the value of  $X_2$  is not always sufficient to know the value of  $\omega_1 = X_1$ . Contrarily, knowing the value of  $(\omega_1, \omega_2)$  (contained in  $\mathcal{F}_2$ ) definitely implies that you know the value of  $X_2$ . In other words, (the information of)  $\sigma(X_2)$  is contained in  $\mathcal{F}_2$ , concisely  $\sigma(X_2) \subset \mathcal{F}_2$ .*

*Now compare  $\sigma(X_2)$  and  $\sigma(Y)$  where  $Y = X_2^2$ . Lets consider three events  $A = \{X_2 = 2\}$ ,  $B = \{X_2 = 0\}$ ,  $C = \{X_2 \text{ is even}\}$ . Clearly all three events are in the  $\sigma$ -algebra generated by  $X_2$  (i.e.  $\sigma(X_2)$ ) since if you know that value of  $X_2$  then you **always** know whether the events happen or not. Next notice that  $B \in \sigma(Y)$  since if you know that  $Y = 0$  then  $X_2 = 0$  and if  $Y \neq 0$  then  $X_2 \neq 0$ . Hence no matter what the value of  $Y$  is knowing it you can decide if  $X_2 = 0$  or not. However, knowing the value of  $Y$  does not **always** tell you if  $X_2 = 2$ . It does sometimes, but not always. If  $Y = 0$  then you know that  $X_2 \neq 0$ . However if  $Y = 4$  then  $X_2$  could be equal to either 2 or  $-2$ . We conclude that  $A \notin \sigma(Y)$  but  $B \in \sigma(Y)$ . Since  $X_2$  is always even, we do not need to know any information to decide  $C$  and it is in fact in both  $\sigma(X_2)$  and  $\sigma(Y)$ . In fact,  $C = \Omega$  and  $\Omega$  is in any  $\sigma$ -algebra since by definition  $\Omega$  and the empty set  $\emptyset$  are always included. Lastly, since whenever we know  $X_2$  we know  $Y$ , it is clear that  $\sigma(X_2)$  contains all of the information contained in  $\sigma(Y)$ . In fact it follows from the definition and the fact that  $\sigma(Y) \subset \sigma(X_2)$ . To say that one  $\sigma$ -algebra is contained in another is to say that the second contains all of the information of the first and possibly more.*

DEFINITION 2.10. *We say that a real-valued random variable  $X$  is measurable with respect to  $\sigma$ -algebra  $\mathcal{G}$  if every set of the form  $X^{-1}([a, b])$  is in  $\mathcal{G}$ .<sup>1</sup>*

---

<sup>1</sup>If  $X$  is a random variable taking values in a measurable space  $(\mathbf{X}, \mathcal{B})$  (recall that  $\mathcal{B}$  is a  $\sigma$ -algebra over  $\mathbf{X}$ ) then we require that  $X^{-1}(B) \in \mathcal{G}$  for all  $B \in \mathcal{B}$ .

Speaking intuitively, a random variable is measurable with respect to a given  $\sigma$ -algebra if the information in the  $\sigma$ -algebra is **always** sufficient to fix the value of the random variable. Of course the random variable  $X$  is always measurable with respect to  $\sigma(X)$ . More specifically,  $\sigma(X)$  is the smallest  $\sigma$ -algebra  $\mathcal{G}$  on  $\Omega$  such that  $X$  is  $\mathcal{G}$ -measurable. In the previous example,  $Y$  is measurable with respect to  $\sigma(X_2)$  since knowing the value of  $X_2$  fixes the value of  $Y$ .

DEFINITION 2.11. *If a random variable  $X$  is measurable with respect to a  $\sigma$ -algebra  $\mathcal{F}$  then we will write  $X \in \mathcal{F}$ . While this is a slight abuse of notation, it will be very convenient.*

EXAMPLE 2.12. *Let  $X$  be a random variable taking values in  $[-1, 1]$ . Let  $g$  be the function from  $[-1, 1] \rightarrow \{-1, 1\}$  such that  $g(x) = -1$  if  $x \leq 0$  and  $g(x) = 1$  if  $x > 0$ . Define the random variable  $Y$  by  $Y(\omega) = g(X(\omega))$ . Hence  $Y$  is a random variable taking values in  $\{-1, 1\}$ . Let  $\mathcal{F}_Y$  be the  $\sigma$ -algebra generated by the random variable  $Y$ . That is  $\mathcal{F}_Y = \sigma(Y) := \{Y^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$ . In this case, we can figure out exactly what  $\mathcal{F}_Y$  looks like. Since  $Y$  takes on only two values, we see that for any subset  $B$  in  $\mathcal{B}(\mathbb{R})$  (the Borel  $\sigma$ -algebra of  $\mathbb{R}$ )*

$$Y^{-1}(B) = \begin{cases} Y^{-1}(-1) := \{\omega : Y(\omega) = -1\} & \text{if } -1 \in B, 1 \notin B \\ Y^{-1}(1) := \{\omega : Y(\omega) = 1\} & \text{if } 1 \in B, -1 \notin B \\ \emptyset & \text{if } -1 \notin B, 1 \notin B \\ \Omega & \text{if } -1 \in B, 1 \in B \end{cases}$$

Thus  $\mathcal{F}_Y$  consists of exactly four sets, namely  $\{\emptyset, \Omega, Y^{-1}(-1), Y^{-1}(1)\}$ . For a function  $f : \Omega \rightarrow \mathbb{R}$  to be measurable with respect to the  $\sigma$ -algebra  $\mathcal{F}_Y$ , the inverse image of any set  $B \in \mathcal{B}(\mathbb{R})$  must be one of the four sets in  $\mathcal{F}_Y$ . This is another way of saying that  $f$  must be constant on both  $Y^{-1}(-1)$  and  $Y^{-1}(1)$ . Note that together  $Y^{-1}(-1) \cup Y^{-1}(1) = \Omega$ .

We now re-examine the idea of a conditional expectation of a random variable with respect to a  $\sigma$ -algebra. To do so, we introduce the indicator function.

DEFINITION 2.13. *Given  $(\Omega, \mathcal{F}, \mathbb{P})$  a probability space and  $A \in \mathcal{F}$ , the indicator function of  $A$  is*

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & \text{otherwise} \end{cases} \quad (2.7)$$

and is a measurable function. Fixing a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , we have

PROPOSITION 2.14. *If  $X$  is a random variable on  $(\Omega, \mathcal{F}, \mathbb{P})$  with  $\mathbb{E}[|X|] < \infty$ , and  $\mathcal{G} \subset \mathcal{F}$  is a  $\sigma$ -algebra, then there is a unique random variable  $Y$  on  $(\Omega, \mathcal{G}, \mathbb{P})$  such that*

- i)  $\mathbb{E}[|Y|] < \infty$ ,
- ii)  $\mathbb{E}[\mathbb{1}_A Y] = \mathbb{E}[\mathbb{1}_A X]$  for all  $A \in \mathcal{G}$ .

DEFINITION 2.15. *We define the conditional expectation with respect to a  $\sigma$ -algebra  $\mathcal{G}$  as the unique random variable  $Y$  from Proposition 2.14, i.e.,  $\mathbb{E}[X|\mathcal{G}] := Y$ .*

The intuition behind Proposition 2.15 is that the conditional expectation WRT a  $\sigma$ -algebra  $\mathcal{G} \subset \mathcal{F}$  of a random variable  $X \in \mathcal{F}$  is that random variable  $Y \in \mathcal{G}$  that is equivalent or identical (in terms of expected value, or predictive power) to  $X$  given the information contained in  $\mathcal{G}$ . In other words,  $Y = \mathbb{E}[X|\mathcal{G}]$  is **the best approximation** of the value of  $X$  given the information in  $\mathcal{G}$ . The previous definition of conditional expectation WRT a fixed set of events is obtained by *evaluating*

the random variable  $\mathbb{E}[X|\mathcal{G}]$  on the events of interest, *i.e.*, by *fixing* the events in  $\mathcal{G}$  that may have occurred.

When we condition on a random variable we are really conditioning on the *information* that random variable is giving to us. In other words, we are conditioning on the  $\sigma$ -algebra generated by that random variable:

$$\mathbb{E}[X|Z] := \mathbb{E}[X|\sigma(Z)].$$

As in the discrete case, one can show that there exists a function  $h : \text{Range}(Z) \rightarrow \mathbf{X}$  such that

$$\mathbb{E}[Y|Z(\omega)] := h(Z(\omega)),$$

and hence we can think about the conditional expectation as a function of  $Z(\omega)$ . In particular, this allows to define

$$\mathbb{E}[Y|Z = z] := h(z).$$

EXAMPLE 2.16. When the set  $\Omega$  is countable, we can write every random variable  $Y$  on  $\Omega$  as

$$Y(\omega) = \sum_{y \in \text{Range}(Y)} y \mathbb{1}_{Y=y}(\omega).$$

Consequently, writing

$$\mathbb{E}[X|Z](\omega) = \sum_{z \in \text{Range}(Z)} \mathbb{E}[X|Z = z] \mathbb{1}_{Z=z}(\omega),$$

we obtain

$$\mathbb{E}[\mathbb{1}_{Z=z}(\omega)X(\omega)] = \mathbb{E}[\mathbb{1}_{Z=z}(\omega)\mathbb{E}[X|Z = z]] = \mathbb{E}[X|Z = z] \cdot \mathbb{E}[\mathbb{1}_{Z=z}(\omega)]$$

Now, recognizing that  $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}[A]$ , if  $\mathbb{P}[Z = z] \neq 0$  we finally obtain

$$\mathbb{E}[X|Z = z] = \frac{\mathbb{E}[\mathbb{1}_{Z=z}(\omega)X(\omega)]}{\mathbb{P}[Z = z]} = \sum_{x \in \text{Range}(X)} \frac{x \mathbb{P}[Z = z, X = x]}{\mathbb{P}[Z = z]},$$

and recover (2.4).

We now list some properties of the conditional expectation:

- **Linearity:** for all  $\alpha, \beta \in \mathbb{R}$  we have

$$\mathbb{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha \mathbb{E}[X|\mathcal{G}] + \beta \mathbb{E}[Y|\mathcal{G}],$$

- if  $X$  is  $\mathcal{G}$ -measurable then

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

Intuitively, since  $X \in \mathcal{G}$  ( $X$  is measurable wrt the  $\sigma$ -algebra  $\mathcal{G}$ ), the best approximation of  $X$  on the sets contained in  $\mathcal{G}$  is  $X$  itself, so we do not need to approximate it.

- **Tower property:** if  $\mathcal{G}$  and  $\mathcal{H}$  are both  $\sigma$ -algebras with  $\mathcal{G} \subset \mathcal{H}$ , then

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

Since  $\mathcal{G}$  is a smaller  $\sigma$ -algebra, the functions which are measurable with respect to it are contained in the space of functions measurable with respect to  $\mathcal{H}$ . More intuitively, being measurable with respect to  $\mathcal{G}$  means that only the information contained in  $\mathcal{G}$  is left free to vary.  $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}]$  means first give me your best guess given only the information contained in  $\mathcal{H}$  as input and then reevaluate this guess making use of only the information in  $\mathcal{G}$  which is a subset of the information in  $\mathcal{H}$ . Limiting oneself to the information in  $\mathcal{G}$  is the bottleneck so in the end it is the only effect one sees. In other words, once one takes the conditional expectation with respect to a smaller  $\sigma$  algebra one is losing information. Therefore, by doing  $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$  one is losing information (in the innermost expectation) that cannot be recovered by the second one.



- **Jensen's inequality** If  $g : I \rightarrow \mathbb{R}$  is convex<sup>2</sup> on  $I \subseteq \mathbb{R}$  for a random variable  $X \in \mathcal{G}$  with  $\text{range}(X) \subseteq I$  we have

$$g(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[g(X)|\mathcal{G}],$$

- **Chebysheff inequality** For a random variable  $X \in \mathcal{G}$  we have that for any  $\lambda > 0$

$$\mathbb{P}[X > \lambda|\mathcal{G}] \leq \frac{\mathbb{E}[X|\mathcal{G}]}{\lambda},$$

- **Optimal approximation** The conditional expectation with respect to a  $\sigma$ -algebra  $\mathcal{G} \subset \mathcal{F}$  by

$$\mathbb{E}[Y|\mathcal{G}] = \underset{Z \text{ meas w.r.t. } \mathcal{G}}{\text{argmin}} \mathbb{E}[Y - Z]^2 \quad (2.8)$$

This should be thought of as the best guess of the value of  $Y$  given the information in  $\mathcal{G}$ .

EXAMPLE 2.17 (Example 2.12 continued). *In the previous example,  $\mathbb{E}\{X|\mathcal{F}_Y\}$  is the best approximation to  $X$  which is measurable with respect to  $\mathcal{F}_Y$ , that is constant on  $Y^{-1}(-1)$  and  $Y^{-1}(1)$ . In other words,  $\mathbb{E}\{X|\mathcal{F}_Y\}$  is the random variable built from a function  $h_{min}$  composed with the random variable  $Y$  such that the expression*

$$\mathbb{E} \left\{ (X - h_{min}(Y))^2 \right\}$$

*is minimized. Since  $Y(\omega)$  takes only two values in our example, the only details of  $h_{min}$  which matter are its values at 1 and -1. Furthermore, since  $h_{min}(Y)$  only depends on the information in  $Y$ , it is measurable with respect to  $\mathcal{F}_Y$ . If by chance  $X$  is measurable with respect to  $\mathcal{F}_Y$ , then the best approximation to  $X$  is  $X$  itself. So in that case  $\mathbb{E}\{X|\mathcal{F}_Y\}(\omega) = X(\omega)$ .*

In light of (2.8), we see that

$$\mathbb{E}[X|Y_1, \dots, Y_k] = \mathbb{E}[X|\sigma(Y_1, \dots, Y_k)]$$

This fits with our intuitive idea that  $\sigma(Y_1, \dots, Y_k)$  embodies the information contained in the random variables  $Y_1, Y_2, \dots, Y_k$  and that  $\mathbb{E}[X|\sigma(Y_1, \dots, Y_k)]$  is our best guess at  $X$  if we only know the information in  $\sigma(Y_1, \dots, Y_k)$ .

DEFINITION 2.18. *Given  $(\Omega, \mathcal{F}, \mathbb{P})$  a probability space and  $A, B \in \mathcal{F}$ , we say that  $A$  and  $B$  are independent if*

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B] \quad (2.9)$$

*Similarly, random variables  $\{X_i\}$  are jointly independent if for all  $C_i$ ,*

$$\mathbb{P}[X_1 \in C_1 \text{ and } \dots \text{ and } X_n \in C_n] = \prod_{i=1}^n \mathbb{P}[X_i \in C_i] \quad (2.10)$$

It is important to note that given two independent random variables  $X$  and  $Y$  one has

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \quad (2.11)$$

---

<sup>2</sup>A function  $g$  is convex on  $I \subseteq \mathbb{R}$  if for all  $x, y \in I$  with  $[x, y] \subseteq I$  and for all  $\lambda \in [0, 1]$  one has  $g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y)$

### 3. Distributions and Convergence of Random Variables

DEFINITION 3.1. We say that two  $\mathbf{X}$ -valued random variables  $X$  and  $Y$  have the same distribution or have the same law if for all bounded (measurable) functions  $f : \mathbf{X} \rightarrow \mathbb{R}$  we have  $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$ . This equivalence is sometimes written

$$\text{Law}(X) = \text{Law}(Y) \quad (2.12)$$

REMARK 3.2. Either of the following are equivalent to two random variable  $X$  and  $Y$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  having the same distribution.

- i)  $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$  for all continuous  $f$  with compact support.
- ii)  $\mathbb{P}[X \in A] = \mathbb{P}[Y \in A]$  for all  $A \in \mathcal{F}$ .

There are many ways a sequence of random variables  $\{X_n\}_{n \in \mathbb{N}}$  can converge to another random variable  $X$ :

DEFINITION 3.3. Let  $\{X_n\}_{n \in \mathbb{N}}$  be a sequence of random variables on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , and let  $X$  be a random variable on the same space. Then

- **almost sure convergence**  $\{X_n\}$  converges to  $X$  almost surely if

$$\mathbb{P}[\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}] = 1,$$

- **convergence in probability**  $\{X_n\}$  converges to  $X$  in probability if, for all  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}[\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}] = 0,$$

- **weak convergence**  $\{X_n\}$  converges weakly (or in distribution) to  $X$  if, for all  $A \in \mathcal{B}$

$$\lim_{n \rightarrow \infty} \mathbb{P}[X_n(\omega) \in A] = \mathbb{P}[X(\omega) \in A].$$

REMARK 3.4. The above definitions can be ordered by strength: we have the following implications  
almost sure convergence  $\Rightarrow$  convergence in probability  $\Rightarrow$  weak convergence.

Moreover, we note that in order to have convergence in distributions the two random variables do not need to live on the same probability space.

A useful method of showing that the distribution of a sequence of random variables converges to another is to consider the associated sequence of Fourier transforms, or the characteristic function of a random variable as it is called in probability theory.

DEFINITION 3.5. The characteristic function (or Fourier Transform) of a random variable  $X$  is defined as

$$\psi(t) = \mathbb{E}[\exp(itX)]$$

for all  $t \in \mathbb{R}$ .

It is a basic fact that the characteristic function of a random variable uniquely determines the distribution of a random variable. Furthermore, the following convergence theorem is a classical theorem from probability theory.

**THEOREM 3.6.** *Let  $X_n$  be a sequence of real-valued random variables and let  $\psi_n$  be the associated characteristic functions. Assume that there exists a function  $\psi$  so that for each  $t \in \mathbb{R}$*

$$\lim_{n \rightarrow \infty} \psi_n(t) = \psi(t).$$

*If  $\psi$  is continuous at zero then there exists a random variable  $X$  so that the distribution of  $X_n$  converges to the distribution of  $X$ . Furthermore the characteristic function of  $X$  is  $\psi$ .*

**EXAMPLE 3.7.** *Note that using a Fourier transform,*

$$\mathbb{E}[e^{i\lambda x}] = e^{i\lambda m - \frac{\sigma^2 \lambda^2}{2}},$$

*for all  $\lambda$ . Using this, we say that  $X = (X_1, \dots, X_k)$  is a  $k$ -dimensional Gaussian if there exists  $m \in \mathbb{R}^k$  and  $R$  a positive definite symmetric  $k \times k$  matrix so that for all  $\lambda \in \mathbb{R}^k$  we have*

$$\mathbb{E}[e^{i\lambda \cdot x}] = e^{i\lambda \cdot m - \frac{(R\lambda) \cdot \lambda}{2}}.$$



## Wiener Process and Stochastic Processes

### 1. An Illustrative Example: A Collection of Random Walks

Fixing an  $n \geq 0$ , let  $\{\xi_k^{(n)} : k = 1, \dots, 2^n\}$  be a collection of independent random variables each distributed as normal with mean zero and variance  $2^{-n}$ . For  $t = k2^{-n}$  for some  $k \in \{1, \dots, 2^n\}$ , we define

$$B^{(n)}(t) = \sum_{j=1}^k \xi_j^{(n)}. \quad (3.1)$$

for intermediate times  $t \in [0, 1]$  not of the form  $k2^{-n}$  for some  $k$  we define the function as the linear function connecting the two nearest points of the form  $k2^{-n}$ . In other words, if  $t \in [s, r]$  were  $s = k2^{-n}$  and  $r = (k+1)2^{-n}$  then

$$B^{(n)}(t) = \frac{t-s}{2^{-n}}B^{(n)}(s) + \frac{r-t}{2^{-n}}B^{(n)}(r)$$

We will see momentarily that  $B^{(n)}$  has the following properties independent of  $n$ :

- i)  $B^{(n)}(0) = 0$ .
- ii)  $\mathbb{E}B^{(n)}(t) = 0$  for all  $t \in [0, 1]$ .
- iii)  $\mathbb{E}|B^{(n)}(t) - B^{(n)}(s)|^2 = t - s$  for  $0 \leq s < t \leq 1$  of the form  $k2^{-n}$ .
- iv) The distribution of  $B^{(n)}(t) - B^{(n)}(s)$  is Gaussian for  $0 \leq s < t \leq 1$  of the form  $k2^{-n}$ .
- v) The collection of random variable

$$B^{(n)}(t_i) - B^{(n)}(t_{i-1})$$

are mutually independent as long as  $0 \leq t_0 < t_1 < \dots < t_m \leq 1$  for some  $m$  and the  $t_i$  are of the form  $k2^{-n}$ .

The first property is clear since the sum in (3.1) is empty. The second property for  $t = k2^{-n}$  follows from

$$\mathbb{E}B^{(n)}(t) = \sum_{j=1}^k \mathbb{E}\xi_j^{(n)} = 0$$

since  $\mathbb{E}\xi_j^{(n)} = 0$ . For general  $t$ , we have  $\mathbb{E}B^{(n)}(t) = \frac{t-s}{2^{-n}}\mathbb{E}B^{(n)}(s) + \frac{r-t}{2^{-n}}\mathbb{E}B^{(n)}(r)$  with  $s, r$  of the form  $k2^{-n}$  for different  $k$ .

To see the second moment calculation take  $s = m2^{-n}$  and  $t = k2^{-n}$  and observe that

$$\begin{aligned} \mathbb{E}\left[|B^{(n)}(t) - B^{(n)}(s)|^2\right] &= \mathbb{E}\left[\left(\sum_{j=m}^k \xi_j\right)\left(\sum_{\ell=m}^k \xi_\ell\right)\right] = \sum_{j=m}^k \sum_{\ell=m}^k \mathbb{E}[\xi_j \xi_\ell] \\ &= \sum_{j=m}^k \mathbb{E}[\xi_j^2] + \sum_{j=m}^k \sum_{\substack{\ell=m \\ \ell \neq j}}^k \mathbb{E}[\xi_j] \mathbb{E}[\xi_\ell] = \sum_{j=m}^k 2^{-n} = k2^{-n} - m2^{-n} = t - s \end{aligned}$$

since  $\xi_j$  and  $\xi_\ell$  are independent if  $j \neq \ell$ , and  $\mathbb{E}[\xi_j] = 0$ , and  $\mathbb{E}[\xi_j^2] = 2^{-n}$ .

Since  $B^n(t)$  is just the sum of independent Gaussians, it is also distributed Gaussian with a mean and a variance which is just the sum of the individual means and variances respectively. Because for disjoint time intervals the differences of the  $B^n(t_i) - B^n(t_{i-1})$  are sums over disjoint collections of  $\xi'_i$ s, they are mutually independent.

Since all of these properties are independent of  $n$  it is tempting to think about the limit as  $n \rightarrow \infty$  and the mesh becoming increasingly fine. It is not clear that such a limit would exist as the curves  $B^n$  become increasingly “wiggly.” We will see in fact that it does exist. We begin by taking an abstract perspective in the next sections though we will return to a more concrete perspective at the end.

## 2. General Stochastic Processes

Motivated by the example of the previous section, we pause to discuss the idea of a stochastic process more generally.

**DEFINITION 2.1.** *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and let  $(\mathbf{X}, \mathcal{B})$  be a measurable space. Also let  $\mathcal{T}$  be an indexing set which for our purposes will typically be  $\mathbb{R}, \mathbb{R}_+, \mathbb{N}$ , or  $\mathbb{Z}$ . Suppose that for each  $t \in \mathcal{T}$  we have  $X_t : \Omega \rightarrow \mathbf{X}$  a measurable function. Then the set  $\{X_t\}$  is a stochastic process on  $T$  with values in  $\mathbf{X}$ . Also, given  $\omega \in \Omega$ ,  $\xi_t(\omega) : \mathcal{T} \rightarrow \mathbf{X}$  is called a path or trajectory of  $\xi_t$ .*

**REMARK 2.2.** *Commonly used notations for stochastic processes include  $\{X_t\}, X_t(\omega), X_t(\omega), \dots$*

We want to define a type of equivalence between stochastic processes. But first of recall the following notion of equivalence of random variables.

**DEFINITION 2.3.** *We say that two stochastic processes have the same distribution or have the same law if for all  $t_1 < \dots < t_n \in \mathcal{T}$  we have*

$$\text{Law}(X_{t_1}, \dots, X_{t_n}) = \text{Law}(Y_{t_1}, \dots, Y_{t_n})$$

where we think of the vector  $(X_{t_1}, \dots, X_{t_n})$  as a random variable taking values in the product space  $\mathbf{X}^n$ .

We would like to state a nice extension theorem for constructing stochastic processes, but first we need a definition.

**DEFINITION 2.4.** *Given a set of finite dimensional distributions  $\{\mu\}$  over an indexing set  $\mathcal{T}$  on  $\mathbf{X}$  we say that the set is compatible if*

i) *For all  $t_1 < \dots < t_{m+1} \in \mathcal{T}$  and  $A_1, \dots, A_m \in \mathcal{B}$  we have*

$$\mu_{t_1 \dots t_m}(A_1, \dots, A_m) = \mu_{t_1 \dots t_{m+1}}(A_1, \dots, A_m, \mathbf{X})$$

ii) *For all  $t_1 < \dots < t_m, A_1, \dots, A_m \in \mathcal{B}$ , and  $\sigma$  a permutation on  $m$  letters, we have*

$$\mu_{t_1 \dots t_m}(A_1, \dots, A_m) = \mu_{t_{\sigma(1)} \dots t_{\sigma(m)}}(A_{\sigma(1)}, \dots, A_{\sigma(m)})$$

The first condition is roughly saying that if one considers a null condition (the total space) in a higher-dimensional measure, one gets the same result without the null condition in the lower dimensional measure. The second condition is saying that the order of the indexing of the  $\mu$  doesn't matter.

**REMARK 2.5.** *The first of the above two requirements is sometimes called the Chapman-Kolmogorov equation.*

**THEOREM 2.6** (Kolmogorov Extension Theorem). *Given a set of compatible finite dimensional distributions  $\{\mu_{t_1 \dots t_m}\}$  with indexing set  $\mathcal{T}$ , there exists a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and a stochastic process  $\{X_t\}$  so that  $X_t$  has the required finite dimensional distributions, i.e. for all  $t_1 < \dots < t_m \in \mathcal{T}$  and  $A_1, \dots, A_m \in \mathcal{B}$  we have*

$$\mathbb{P}[X_{t_1} \in A_1 \text{ and } \dots \text{ and } X_{t_m} \in A_m] = \mu_{t_1 \dots t_m}(A_1, \dots, A_m) \quad (3.2)$$

### 3. Definition of Brownian motion (Wiener Process)

Looking back at Section 1, the list of properties of  $B^{(n)}$  suggest a reasonable collection of compatible finite distributions. Namely independent increments with each increment distributed normally with mean zero and variance proportional to the time interval. and we make the following definition.

**DEFINITION 3.1.** *Standard Brownian motion  $\{B_t\}$  is a stochastic process on  $\mathbb{R}$  such that*

- i)  $B_0 = 0$  almost surely (i.e.  $\mathbb{P}[\{\omega \in \Omega : B_0 \neq 0\}] = 0$ ),
- ii)  $B_t$  has independent increments: for any  $t_1 < t_2 < \dots < t_n$ ,  $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$  are independent,
- iii) The increments  $B_t - B_s$  are Gaussian random variables with mean 0 and variance given by the length of the interval:

$$\text{Var}(B_t - B_s) = |t - s|.$$

Since this is a compatible collection of finite dimensional distributions, Theorem 2.6 guarantees the existence of the process we have described.

Looking back at  $B^{(n)}$  from (3.1) it might be reasonable to hope that the Brownian motion  $\{B_t\}$  defined above might be a continuous function of time. Notice that the above definition makes no mention of the continuity. The following definition makes it clear what we mean by continuous.

**DEFINITION 3.2.** *A stochastic process is continuous if all the trajectories  $t \rightarrow X_t(\omega)$  are continuous.*

It turns out that the finite dimensional distributions can not *guarantee* that a stochastic process is almost surely continuous. However they can imply that it is possible for a given process to be continuous.

**DEFINITION 3.3.** *A stochastic process  $\{X_t\}$  is a version (or modification) of a second stochastic process  $\{Y_t\}$  if for all  $t$ ,  $\mathbb{P}[X_t = Y_t] = 1$ . Notice that this is a symmetric relation.*

**THEOREM 3.4** (Kolmogorov Continuity Theorem (a version)). *Suppose that a stochastic process  $\{X_t\}$ ,  $t \geq 0$  satisfies the estimate:*

*for all  $T > 0$  there exist positive constants  $\alpha, \beta, D$  so that*

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq D|t - s|^{1+\beta} \quad \forall t, s \in [0, T], \quad (3.3)$$

*then there exist a version of  $X_t$  which is continuous.*

**REMARK 3.5.** *The estimate in (3.3) holds for a Brownian motion. We give the details in one-dimension. First recall that if  $X$  is a gaussian random variable with mean 0 and variance  $\sigma^2$  then  $\mathbb{E}[X^4] = 3\sigma^4$ . Applying this to Brownian motion we have  $\mathbb{E}|B_t - B_s|^4 = 3|t - s|^2$  and conclude that (3.3) holds with  $\alpha = 4, \beta = 1, D = 3$ . Hence it is not incompatible with the all ready assumed properties of Brownian motion to assume that  $B_t$  is continuous almost surely.*

Remark 3.5 shows that continuity is a fundamental attribute of Brownian motion. In fact we have the following second (and equivalent) definition of Brownian motion which assumes a form of continuity as a basic assumption, replacing other assumptions.

**THEOREM 3.6.** *Let  $B_t$  be a stochastic process such that the following conditions hold:*

- i)  $\mathbb{E}(B_1^2) = \text{constant}$ ,
- ii)  $B_0 = 0$  almost surely,
- iii)  $B_{t+h} - B_t$  is independent of  $\{B_s : s \leq t\}$ .
- iv) The distribution of  $B_{t+h} - B_t$  is independent of  $t \geq 0$  (stationary increments),
- v) (Continuity in probability.) For all  $\delta > 0$ ,

$$\lim_{h \rightarrow 0} \mathbb{P}[|B_{t+h} - B_t| > \delta] = 0$$

then  $B_t$  is Brownian motion. When  $\mathbb{E}B(1)^2 = 1$  we call it standard Brownian motion.

The process introduced above can be straightforwardly generalized to  $n$  dimensions:

**DEFINITION 3.7.**  *$n$ -dimensional Standard Brownian motion  $\{B_t\}$  is a stochastic process on  $\mathbb{R}^n$  such that*

- i)  $B_0 = 0$  almost surely (i.e.  $\mathbb{P}[\{\omega : B_0 \neq 0\}] = 0$ ),
- ii)  $B_t$  has independent increments: for any  $t_1 < t_2 < \dots < t_n$ ,  $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$  are independent,
- iii) The increments  $B_t - B_s$  are Gaussian random variables with mean 0 and variance given by the length of the interval: denoting by  $(B_t)_i$  the  $i$ -th component of  $B_t$

$$\text{Var}((B_t)_i - (B_s)_j) = \begin{cases} |t - s| & \text{if } i = j \\ 0 & \text{else} \end{cases}.$$

#### 4. Constructive Approach to Brownian motion

Returning to the construction of Section 1, one might be tempted to hope that the random walks  $B^{(n)}$  converge to a Brownian motion  $B_t$  as  $n \rightarrow \infty$ . While this is true that the distribution of  $B^{(n)}$  converges weakly to that of  $B_t$  as  $n \rightarrow \infty$ , a moments reflection shows that there is not hope that the sequence converges almost surely since  $B^{(n)}$  and  $B^{(n+1)}$  have no relation for a given realization of the underlying random variable  $\{\xi_k^{(n)}\}$ .

We will now show that by cleverly rearranging the randomness we can construct a new sequence of random walks  $W^{(n)}(t)$  so that the stochastic process  $W^{(n)}$  has the same distribution as  $B^{(n)}$  yet  $W^{(n)}$  will converge almost surely to a realization of Brownian motion.

We begin by defining a new collection of random variables  $\{\eta_k^{(n)}\}$  from the  $\{\xi_k^{(n)}\}$ . We define  $\eta_1^{(0)}$  to be a normal random variable with mean 0 and variance 1 which is independent of all of the  $\xi$ 's. Then for  $n \geq 0$  and  $k \in \{1, \dots, 2^k\}$  we define

$$\eta_{2k}^{(n+1)} = \frac{1}{2}\eta_k^{(n)} + \frac{1}{2}\xi_k^{(n)} \quad \text{and} \quad \eta_{2k-1}^{(n+1)} = \frac{1}{2}\eta_k^{(n)} - \frac{1}{2}\xi_k^{(n)}$$

Since each  $\eta_k^{(n)}$  is the sum of independent Gaussian random variables they are themselves Gaussian random variables. It is easy to see that  $\eta_k^{(n)}$  is mean zero and has variance  $2^{-n}$ . Since for any  $n \geq 0$  and  $j, k \in \{1, \dots, 2^k\}$  with  $j \neq k$ , we see that  $\mathbb{E}\eta_k^{(n)}\eta_j^{(n)} = 0$  and we conclude that because the



variables are Gaussian that the collection of random variables  $\{\eta_k^{(n)} : k \in \{1, \dots, 2^k\}\}$  are mutually independent. Hence if we define

$$W^{(n)}(k2^{-n}) = \sum_{j=1}^k \eta_j^{(n)}$$

and at intermediate times as the value of the line connecting the two nearest points, then  $W^{(n)}$  has the same distribution as  $B^{(n)}$  from Section 1.

**THEOREM 4.1.** *With probability one, the sequence of functions  $(W^{(n)}(t))(\omega)$  on converges uniformly to a continuous function  $B_t(\omega)$  as  $n \rightarrow \infty$ , and the process  $B_t(\cdot)$  is a Brownian motion on  $[0, 1]$ .*

**PROOF.** Now for  $n \geq 0$  and  $k \in \{1, \dots, 2^k\}$  define

$$Z_k^{(n)} = \sup_{t \in [(k-1)2^{-n}, k2^{-n}]} |W^{(n)}(t) - W^{(n+1)}(t)|$$

and observe that

$$Z_k^{(n)} = \left| \frac{1}{2} \eta_k^{(n)} - \eta_{2k-1}^{(n+1)} \right| = \left| \frac{1}{2} \xi_k^{(n)} \right|$$

Since  $\xi_k^{(n)}$  is normal with mean zero and variance  $2^{-(n+2)}$ , we have by Markov inequality that

$$\mathbb{P}[Z_k^{(n)} > \delta] \leq \frac{\mathbb{E}[|Z_k^{(n)}|^4]}{\delta^4} = \frac{3 \cdot 2^{-2(n+2)}}{\delta^4}.$$

In turn since the  $\{Z_k^{(n)} : k = 1, \dots, 2^n\}$  are mutually independent this implies that

$$\begin{aligned} \mathbb{P} \left[ \sup_{t \in [0,1]} |W^{(n)}(t) - W^{(n+1)}(t)| > \delta \right] &= \mathbb{P} \left[ \sup_k Z_k^{(n)} > \delta \right] = 2^n \mathbb{P}[Z_1^{(n)} > \delta] \\ &\leq 2^n \frac{3 \cdot 2^{-2(n+2)}}{\delta^4} := \psi(n, \delta). \end{aligned}$$

Since  $\psi(n, 2^{-n/5}) \sim c2^{-n/5}$  for some  $c > 0$ , we have that

$$\sum_{n=1}^{\infty} \mathbb{P} \left[ \sup_{t \in [0,1]} |W^{(n)}(t) - W^{(n+1)}(t)| > 2^{-n/5} \right] < \infty.$$

Hence the Borel-Cantelli lemma implies that with probability one there exists a random  $k(\omega)$  so that if  $n \geq k$  then

$$\sup_{t \in [0,1]} |W^{(n)}(t) - W^{(n+1)}(t)| \leq 2^{-n/5}.$$

In other words with probability one the  $\{W^{(n)}\}$  form a Cauchy sequence. Let  $B_t$  denote the limit. It is not hard to see that  $B_t$  has the properties that define Brownian motion. Furthermore since each  $W^{(n)}$  is uniformly continuous and converge in the supremum norm to  $B_t$ , we conclude that with probability one  $B_t$  is also uniformly continuous.  $\square$

## 5. Brownian motion has Rough Trajectories

In Section 4, we saw that Brownian motion could be seen as the limit of a ever roughening path. This leads us to wonder “how rough is Brownian motion?” We know it is continuous, but is it differentiable?

DEFINITION 5.1. *The (standard)  $p$ -th variation on the interval  $(s, t)$  of any continuous function  $f$  is defined to be*

$$V_p[f](s, t) = \sup_{\Gamma} \sum_k |f(t_{k+1}) - f(t_k)|^p \quad (3.4)$$

where the supremum is over all partitions

$$\Gamma = \{t_k\} : s = t_0 < t_1 < \dots < t_{n-1} < t_n = t, \quad (3.5)$$

for some  $n$ .

For a given partition  $\{t_k\}$  let us define the mesh width of the partition to be

$$|\Gamma| = \sup_{0 < k \leq N} |t_k - t_{k-1}|.$$

The variations of Brownian motion are finite only for a certain range of  $q$ :

PROPOSITION 5.2. *If  $B_t$  is a Brownian motion on the interval  $[0, T]$  with  $T < \infty$ , then*

$$V_p[B](0, T) < \infty \quad \text{a.s.} \quad \text{if and only if} \quad p > 2. \quad (3.6)$$

PROOF. See [17, 12] for details. □

The fact that large values of  $p$  imply boundedness of the quadratic variation may be surprising at first. However, this results from the fact that in order for the supremum in (3.4) to diverge for a continuous function we must consider a sequence  $\Gamma^N$  of partitions with diverging number of intervals. As these intervals become smaller, the variation that each of them captures becomes smaller. These small contribution become even smaller if they are raised to a power  $p > 1$ , whence the (possible) convergence. This concept is in close relation with the one of *Hölder continuity*.

Notice that if a function  $f$  has a nice bounded derivative on the interval  $[0, t]$  then  $V_1[f](0, t) < \infty$  since

$$|f(t_k) - f(t_{k+1})| = \left| \int_{t_k}^{t_{k+1}} f'(s) ds \right| \leq \left( \sup_{s \in [0, t]} |f'(s)| \right) |t_{k+1} - t_k|,$$

we see that

$$V_1[f](0, t) \leq \left( \sup_{s \in [0, t]} |f'(s)| \right) t.$$

Similar considerations hold if  $f$  is Lipschitz continuous in  $[0, T]$  with Lipschitz constant  $L$ :

$$V_1[f](0, t) = \sum_k |f(t_k) - f(t_{k+1})| \leq \sum_k L(t_{k+1} - t_k) = Lt.$$

Hence (3.6) implies that with probability one Brownian Motion can not have a bounded derivative on any interval. In fact something much stronger is true. With probability one, Brownian motion is nowhere differentiable as a function of time (see [12] for details).

From Proposition 5.2, we see that  $p = 2$  is the border case. It is quite subtle. On one hand the statement (3.6) is true, yet if one considers a *specific* sequence of *shrinking* partitions  $\Gamma^{(N)}$  (such that each successive partition contains the previous partition as a sub-partition) then

$$Q_N(T) := \sum_{\Gamma^{(N)}} |B(t_{k+1}^{(N)}) - B(t_k^{(N)})|^2 \rightarrow T \quad \text{a.s.}$$

Initially we will prove the following simpler statement.

**THEOREM 5.3.** *Let  $\Gamma^{(N)}$  be a sequence of partitions of  $[0, T]$  as in (3.5) with  $\lim_{N \rightarrow \infty} |\Gamma^{(N)}| \rightarrow 0$ . Then*

$$Q_N(T) := \sum_{\Gamma^{(N)}} |B(t_{k+1}^{(N)}) - B(t_k^{(N)})|^2 \xrightarrow{N \rightarrow \infty} T \quad (3.7)$$

in  $L^2(\Omega, \mathbb{P})$ .

**COROLLARY 5.4.** *Under the conditions of the above theorem we have  $\lim_{N \rightarrow \infty} Q_N(T) = T$  in probability.*

**PROOF.** We see that for any  $\varepsilon > 0$

$$\mathbb{P}[\omega : |Z_N(\omega) - T| > \varepsilon] \leq \frac{\mathbb{E}[|Z_N(\omega) - T|^2]}{\varepsilon^2} \rightarrow 0 \quad \text{as} \quad |\Gamma^{(N)}| \rightarrow 0.$$

□

**PROOF OF THEOREM 5.3.** Fix any sequence of partitions

$$\Gamma^{(N)} := \{\{t_k^{(N)}\} : 0 = t_1^{(N)} < t_2^{(N)} \dots < t_n^{(N)} = T\},$$

of  $[0, T]$  with  $|\Gamma^{(N)}| \rightarrow 0$  as  $N \rightarrow \infty$ . Defining

$$Z_N := \sum_{k=1}^{N-1} [B(t_{k+1}^{(N)}) - B(t_k^{(N)})]^2,$$

we need to show that

$$\mathbb{E}[Z_N - T]^2.$$

We have,

$$\mathbb{E}[Z_N - T]^2 = \mathbb{E}[Z_N]^2 - 2T\mathbb{E}[Z_N] + T^2 = \mathbb{E}[Z_N]^2 - T^2.$$

Using the convenient notation  $\Delta_k B := B(t_{k+1}^{(N)}) - B(t_k^{(N)})$  and  $\Delta_k t^{(N)} := |t_{k+1}^{(N)} - t_k^{(N)}|$  we have that

$$\begin{aligned} \mathbb{E}[Z_N]^2 &= \mathbb{E}\left[\sum_n (\Delta_n B)^2 \sum_k (\Delta_k B)^2\right] \\ &= \mathbb{E}\left[\sum_n (\Delta_n B)^4\right] + \mathbb{E}\left[\sum_{n \neq k} (\Delta_k B)^2 (\Delta_n B)^2\right] \\ &= 3 \sum_n (\Delta_n t^{(N)})^2 + \sum_{n \neq k} (\Delta_k t^{(N)}) (\Delta_n t^{(N)}) \end{aligned}$$

since  $\mathbb{E}(\Delta_k B)^2 = \Delta_k t$  and  $\mathbb{E}(\Delta_k B)^4 = (\Delta_k t^{(N)})^2$  because  $\Delta_k B$  is a Gaussian random variable with mean zero and variance  $\Delta_k t^{(N)}$ .

The limit of the first term equals 0 as the maximum partition spacing goes to zero since

$$\sum_n (\Delta_n t^{(N)})^2 \leq 3 \cdot \sup(\Delta_n t^{(N)}) T$$

Returning to the remaining term

$$\begin{aligned}
\sum_{n \neq k} (\Delta_k t^{(N)})(\Delta_n t^{(N)}) &= \sum_{k=1}^N \Delta_k t^{(N)} \left[ \sum_{n=1}^{k-1} \Delta_n t^{(N)} + \sum_{k+1}^N \Delta_n t^{(N)} \right] \\
&= \sum_{k=1}^N \Delta_k t^{(N)} (T - \Delta_k t^{(N)}) \\
&= T \sum_{k=1}^N \Delta_k t^{(N)} - \sum_{k=1}^N (\Delta_k t^{(N)})^2 \\
&= T^2 - 0
\end{aligned}$$

Summarizing, we have shown that

$$\mathbb{E}[Z_N - T]^2 \rightarrow 0 \quad \text{as} \quad N \rightarrow \infty$$

□

COROLLARY 5.5. *Under the conditions of Theorem 5.3, if  $\Gamma^{(N)} \subset \Gamma^{(N)}$  we have  $\lim_{N \rightarrow \infty} Q_N(T) = T$  almost surely.*

## 6. More Properties of Random Walks

We now return to the family of random walks constructed in Section 1. The collection of random walks  $B^{(n)}(t)$  constructed in (3.1) have additional properties which are useful to identify and isolate as the general structures will be important for our development of stochastic calculus.

Fixing an  $n \geq 0$ , define  $t_k = k2^{-n}$  for  $k = 0, \dots, 2^n$ . Then notice that for each such  $k$ ,  $B^{(n)}(t_k)$  is a Gaussian random variable since it is the sum of the mutually independent random variables  $\xi_j^{(n)}$ . Furthermore for any collection of  $(t_{k_1}, t_{k_2}, \dots, t_{k_m})$  with  $k_j \in \{1, \dots, 2^n\}$  we have that

$$(B^{(n)}(t_1), \dots, B^{(n)}(t_m))$$

is a multidimensional Gaussian vector.

Next notice that for  $0 \leq t_\ell < t_k \leq 1$  we have

$$\mathbb{E}[B_{t_k}^{(n)} | B_{t_\ell}^{(n)}] = \mathbb{E}[(B_{t_k}^{(n)} - B_{t_\ell}^{(n)}) + B_{t_\ell}^{(n)} | B_{t_\ell}^{(n)}] = \mathbb{E}[B_{t_k}^{(n)} - B_{t_\ell}^{(n)}] + B_{t_\ell}^{(n)} = B_{t_\ell}^{(n)} \quad (3.8)$$

since  $B_{t_k}^{(n)} - B_{t_\ell}^{(n)}$  is independent of  $B_{t_\ell}^{(n)}$  and  $\mathbb{E}[B_{t_k}^{(n)} - B_{t_\ell}^{(n)}] = 0$ . We will choose to view this single fact as the result of two finer grain facts. The first being that the distribution of the walk at time  $t_k$  given the values  $\{B_s^{(n)} : s < t_\ell\}$  is the same as the conditional distribution of the walk at time  $t_k$  given only  $B_{t_\ell}^{(n)}$ . In light of Definition 3.1, we can state this more formally by saying that for all functions  $f$

$$\mathbb{E}\left[f(B_{t_k}^{(n)}) \middle| \mathcal{F}_{t_\ell}\right] = \mathbb{E}\left[f(B_{t_k}^{(n)}) \middle| B_{t_\ell}^{(n)}\right]$$

where  $\mathcal{F}_t = \sigma(B_s^{(n)} : s \leq t)$ . This property is called the Markov property which states that the distribution of the future depends only on the past through the present value of the process.

There is a stronger version of this property called the strong Markov property which states that one can in fact restart the process and restart it from the current (random) value and run it for the remaining amount of time and obtain the same answer. To state this more precisely let us introduce the process  $X(t) = x + B^{(n)}(t)$  as the random walk starting from the point  $x$  and let  $\mathbb{P}_x$  be the probability distribution induced on  $C([0, 1]\mathbb{R})$  by the trajectory of  $X(t)$  for fixed initial  $x$ . Let  $\mathbb{E}_x$  be the expected value associated to  $\mathbb{P}_x$ . Of course  $\mathbb{P}_0$  is simply the random walk starting from 0

that we have been previously considering. Then the strong Markov property states that for any function  $f$

$$\mathbb{E}_0 f(X_{t_k}) = \mathbb{E}_0 F(X_{t_k - t_\ell}, t_k) \quad \text{where} \quad F(x, t) = \mathbb{E}_x f(X_t).$$

Neither of these Markov properties is solely enough to produce (3.8). We also need some fact about the mean of the process given the past. Again defining  $\mathcal{F}_t = \sigma(B_s^{(n)} : s \leq t)$ , we can rewrite (3.8) as

$$\mathbb{E}[B_{t_k}^{(n)} | \mathcal{F}_{t_\ell}] = B_{t_\ell}^{(n)}$$

by using the Markov property. This equality is the principle fact that makes a process what is called a martingale.

We now revisit these ideas making more general definitions which abstract these properties so we can talk about and use them in broader contexts.

## 7. More Properties of General Stochastic Processes

**Gaussian processes.** We begin by giving the general definition of a Gaussian process of which Brownian motion is an example.

**DEFINITION 7.1.**  $\{X_t\}$  is a Gaussian random process if all finite dimensional distributions of  $X$  are Gaussian random variables. I.e., for all  $t_1 < \dots < t_k \in \mathcal{T}$  and  $A_1, \dots, A_k \in \mathcal{B}$  (where here  $\mathcal{B}$  represents Borel sets on the real line) we have that there exists  $R$  a positive definite symmetric  $k \times k$  matrix and  $m \in \mathbb{R}^k$  so that

$$\mathbb{P}[X_{t_1} \in A_1 \text{ and } \dots \text{ and } X_{t_k} \in A_k] = \int_{A_1 \times \dots \times A_k} \frac{1}{(2\pi)^{k/2} \sqrt{\det R}} e^{-\frac{1}{2}(X-m)^T R^{-1}(X-m)}.$$

We have the associated definitions

$$\mu_t := \mathbb{E}[X_t], \quad R_{t,s} := \text{cov}(X_t X_s) = \mathbb{E}[(X_t - \mu_t) \cdot (X_s - \mu_s)].$$

**EXAMPLE 7.2.** By definition, Brownian motion is a Gaussian process with mean vector  $\mu_t = 0$  for all  $t > 0$  and covariance matrix

$$\begin{aligned} \text{Cov}(B_t, B_s) &= \text{Cov}(B_s + (B_t - B_s), B_s) = \mathbb{E}[B_s]^2 + \mathbb{E}[(B_t - B_s)B_s] \\ &= \mathbb{E}[B_s]^2 + \mathbb{E}[B_t - B_s] \mathbb{E}[B_s] = s. \end{aligned}$$

where we have assumed without loss of generality that  $s < t$  and in the third identity we have used the independence of increments of Brownian motion. This shows that for general  $t, s \geq 0$  we have

$$\text{Cov}(B_t, B_s) = \min\{t, s\}. \quad (3.9)$$

In fact, because the mean and covariance structure of a Gaussian process completely determine the properties of its marginals, if a Gaussian process has the same covariance and mean as a brownian motion then it is a Brownian Motion.

**THEOREM 7.3.** A Brownian motion is a Gaussian process with zero mean function, and covariance function  $\min(t, s)$ . Conversely, a Gaussian process with zero mean function, and covariance function  $\min(t, s)$  is a Brownian motion.

**PROOF.** Example 7.2 proves the forward direction. To prove the reverse direction, assume that  $X_t$  is a Gaussian process with zero mean and  $\text{Cov}(X_t, X_s) = \min(t, s)$ . Then the increments of the process, given by  $(X_t, X_{t+s} - X_t)$  are Gaussian random variables with mean 0. The variance of the increments  $X_{t+s} - X_t$  is given by

$$\text{Var}(X_{t+s} - X_t, X_{t+s} - X_t) = \text{Cov}(X_{t+s}, X_{t+s}) - 2\text{Cov}(X_t, X_{t+s} - X_t) + \text{Cov}(X_t, X_t) = (t+s) - 2t + t = s.$$

The independence of  $X_t$  and  $X_{t+s} - X_t$  follows immediately by

$$\text{Cov}(X_t, X_{t+s} - X_t) = \text{Cov}(X_t, X_{t+s}) - \text{Cov}(X_t, X_t) = t - t = 0.$$

□

**Martingales.** In order to introduce the concept of a martingale, we first adapt the concept of  $\sigma$ -algebra to the framework of stochastic processes. In particular, in the case of a stochastic process we would like to encode the idea of **history** of a process: by observing a process up to a time  $t > 0$  we have all the information on the behavior of the process before that time but none after it. Furthermore, as  $t$  increases we increase the amount of information we have on that process. This idea is the one that underlies the concept of *filtration*:

DEFINITION 7.4. *Given an indexing set  $T$ , a filtration of  $\sigma$ -algebras is a set of sigma algebras  $\{\mathcal{F}_t\}_{t \in T}$  such that for all  $t_1 < \dots < t_m \in T$  we have*

$$\mathcal{F}_{t_1} \subset \dots \subset \mathcal{F}_{t_m}.$$

We now define in which sense a filtration contains the information associated to a certain process

DEFINITION 7.5. *A stochastic process  $\{X_t\}$  is adapted to a filtration  $\{\mathcal{F}_t\}$  if its marginals are measurable with respect to the corresponding  $\sigma$ -algebras, i.e., if  $\sigma(X_t) \subseteq \mathcal{F}_t$  for all  $t \in T$ . In this case we say that the process is to the filtration  $\{\mathcal{F}_t\}$ .*

We also extend the concept of  $\sigma$ -algebras *generated* by a random variable to the case of a filtration. In this case, the filtration generated by a process  $\{X_t\}$  is the smallest filtration containing enough information about  $\{X_t\}$ .

DEFINITION 7.6. *Let  $\{X_t\}$  be a stochastic process on  $(\Omega, \mathcal{F}, \mathbb{P})$ . Then the filtration  $\{\mathcal{F}_t^X\}$  generated by  $X_t$  is given by*

$$\mathcal{F}_t^X = \sigma(X_s | 0 \leq s \leq t),$$

*which means the smallest  $\sigma$ -algebra with respect to which the random variable  $X_s$  is measurable, for all  $s \in [0, t]$ . Thus,  $\mathcal{F}_t^X$  contains  $\sigma(X_s)$  for all  $0 \leq s \leq t$ .*

The canonical example of this is the filtration generated by a discrete random process (i.e.  $T = \mathbb{N}$ ):

EXAMPLE 7.7 (Example 0.1 continued). *The filtration generated by the  $N$ -coin flip process for  $m < N$  is*

$$\mathcal{F}_m := \sigma(X_0, \dots, X_m).$$

Intuitively, we will think of  $\{\mathcal{F}_t^X\}$  as the **history** of  $\{X_t\}$  up to time  $t$ , or the information about  $\{X_t\}$  up to time  $t$ . Roughly speaking, an event  $A$  is in  $\{\mathcal{F}_t^X\}$  if its occurrence can be determined by knowing  $\{X_s\}$  for all  $s \in [0, t]$ . For example, if  $B_t$  is a Brownian motion consider the event

$$A = \{\omega \in \Omega : \max_{t \in (0, 1/2)} |B_s(\omega)| \leq 2\}.$$

It is clear that we have  $A \in \mathcal{F}_{1/2}$  as the history of  $B_t$  up to time  $t = 1/2$  determines whether  $A$  has occurred or not. However, we have that  $A \notin \mathcal{F}_{1/3}$  as the process may not yet have reached 2 at time  $t = 1/3$  but may do so before  $t = 1/2$ .

We now have all the tools to define the concept of a martingale:

DEFINITION 7.8.  *$\{X_t\}$  is a Martingale with respect to a filtration  $\mathcal{F}_t$  if for all  $t > s$  we have*

- i)  $X_t$  is  $\mathcal{F}_t$ -measurable,

- ii)  $\mathbb{E}[X_t|\mathcal{F}_s] = X_s,$
- iii)  $\mathbb{E}[|X_t|] < \infty.$

Condition iii) in Def. 7.8 involves a conditional expectation with respect to the  $\sigma$ -algebra  $\mathcal{F}_t$ . Recall that  $\mathbb{E}[X_{t+s}|\mathcal{F}_t]$  is an  $\mathcal{F}_t$ -measurable random variable which approximates  $X_{t+s}$  in a certain optimal way (and it is uniquely defined). Then Def. 7.8 states that, given the history of  $X_t$  up to time  $t$ , our best estimate of  $X_{t+s}$  is simply  $X_t$ , the value of  $\{X_t\}$  at the present time  $t$ . In other words, a martingale is the equivalent in stochastic calculus of a straight line.

EXAMPLE 7.9. *Brownian motion is a martingale WRT  $\{\mathcal{F}_t^B\}$ . Indeed, we have that*

$$\mathbb{E}[B_{t+s}|\mathcal{F}_t] = \mathbb{E}[B_t + (B_{t+s} - B_t)|\mathcal{F}_t] = \mathbb{E}[B_t|\mathcal{F}_t] + \mathbb{E}[B_{t+s} - B_t|\mathcal{F}_t] = B_t + 0,$$

by the independence of increments property.

The above strategy can be extended to general functions  $g(X)$ , as the only property that was used is independence of the increments: because  $g(X) \in \mathcal{F}^X$  this property implies that

$$\mathbb{E}[g(B_{t+s} - B_t)|\mathcal{F}_t^B] = \mathbb{E}[g(B_{t+s} - B_t)]. \quad (3.10)$$

EXAMPLE 7.10. *The process  $X_t := B_t^2 - t$  is a martingale WRT  $\{\mathcal{F}_t^B\}$ . Indeed we have that the process is obviously measurable WRT  $\{\mathcal{F}_t^B\}$  and that  $\mathbb{E}[|B_t^2|] = t < \infty$  verifying i) and ii) from Def. 7.8. For iii) we have*

$$\begin{aligned} \mathbb{E}[B_{t+s}^2|\mathcal{F}_t^B] &= \mathbb{E}[(B_t + B_{t+s} - B_t)^2|\mathcal{F}_t^B] \\ &= \mathbb{E}[B_t^2|\mathcal{F}_t^B] - 2\mathbb{E}[B_t]\mathbb{E}[B_{t+s} - B_t|\mathcal{F}_t^B] + \mathbb{E}[(B_{t+s} - B_t)^2|\mathcal{F}_t^B] = B_t^2 + s. \end{aligned}$$

subtracting  $t - s$  on both sides of the above equation we obtain  $\mathbb{E}[B_{t+s}^2 - (t + s)|\mathcal{F}_t^B] = B_t^2 - t$ .

EXAMPLE 7.11. *The process  $Y_t := \exp[\lambda B_t^2 - \lambda^2 t/2]$  is a martingale WRT  $\{\mathcal{F}_t^B\}$ . Again, the process is obviously measurable WRT  $\{\mathcal{F}_t^B\}$  and, computing the moment generating function of a gaussian random variable, we have that  $\mathbb{E}[\exp(\lambda B_t)] = \exp(t\lambda^2/2) < \infty$ , verifying i) and ii) from Def. 7.8. For iii) we have*

$$\begin{aligned} \mathbb{E}[\exp(\lambda B_{t+s})|\mathcal{F}_t^B] &= \mathbb{E}[\exp(\lambda(B_t + B_{t+s} - B_t))|\mathcal{F}_t^B] = \exp(\lambda B_t)\mathbb{E}[\exp(\lambda(B_{t+s} - B_t))|\mathcal{F}_t^B] \\ &= \exp(\lambda B_t)\mathbb{E}[\exp(\lambda B_s)] = \exp(s\lambda^2/2). \end{aligned}$$

multiplying by  $\exp[(t - s)\lambda^2/2]$  on both sides of the above equation we obtain

$$\mathbb{E}[\exp(\lambda B_{t+s} - \lambda^2(t + s)/2)|\mathcal{F}_t^B] = \exp(\lambda B_t - \lambda^2 t/2).$$

**Markov processes.** We now turn to the general idea of a Markov Process. As we have seen in the example above, this family of processes has the “memoryless” property, *i.e.*, their future depends on their past (their history, their *filtration*) only through their present (their state at the present time, or the  $\sigma$ -algebra generated by the random variable of the process at that time).

In the discrete time and countable sample space setting, this holds if given  $t_1 < \dots < t_m < t$ , we have that the distribution of  $X_t$  given  $(X_{t_1}, \dots, X_{t_m})$  equals the distribution of  $X_t$  given  $(X_{t_m})$ . This is the case if for all  $A \in \mathcal{B}(\mathbf{X})$  and  $s_1, \dots, s_m \in \mathbf{X}$  we have that

$$\mathbb{P}[X_t \in A | X_{t_1} = s_1, \dots, X_{t_m} = s_m] = \mathbb{P}[X_t \in A | X_{t_m} = s_m].$$

This property can be stated in more general terms as follows:

DEFINITION 7.12. *A random process  $\{X_t\}$  is called Markov with respect to a filtration  $\{\mathcal{F}_t\}$  when  $X_t$  is adapted to the filtration and, for any  $s > t$ ,  $X_s$  is independent of  $\mathcal{F}_t$  given  $X_t$ .*

The above definition can be restated in terms of Brownian motions as follows: For any set  $A \in (\mathbf{X})$  we have

$$\mathbb{P}[B_t \in A | \mathcal{F}_s^B] = \mathbb{P}[B_t \in A | B_s] \quad \text{a.s.} \quad (3.11)$$

Remember that conditional probabilities with respect to a  $\sigma$ -algebra are really random variables in the way that a conditional expectation with respect to *sigma*-algebra is a random variable. That is,

$$\mathbb{P}[B_t \in A | \mathcal{F}_s^B] = \mathbb{E}[\mathbb{1}_{B_t \in A}(\omega) | \mathcal{F}_s^B] \quad \text{and} \quad \mathbb{P}[B_t \in A | B_s] = \mathbb{E}[\mathbb{1}_{B_t \in A}(\omega) | \sigma(B_s)] .$$

EXAMPLE 7.13. *The fact that (3.11) holds can be shown directly by using characteristic functions. Indeed, to show that the distributions of the right and left hand side of (3.11) coincide it is enough to identify their characteristic functions. We compute*

$$\mathbb{E}\left[e^{i\vartheta B_t} | \mathcal{F}_s^B\right] = \mathbb{E}\left[e^{i\vartheta(B_s+B_t-B_s)} | \mathcal{F}_s^B\right] = e^{i\vartheta B_s} \mathbb{E}\left[e^{i\vartheta(B_t-B_s)} | \mathcal{F}_s^B\right] = e^{i\vartheta B_s} \mathbb{E}\left[e^{i\vartheta(B_t-B_s)}\right] ,$$

and similarly

$$\mathbb{E}\left[e^{i\vartheta B_t} | B_s\right] = \mathbb{E}\left[e^{i\vartheta(B_s+B_t-B_s)} | B_s\right] = e^{i\vartheta B_s} \mathbb{E}\left[e^{i\vartheta(B_t-B_s)} | B_s\right] = e^{i\vartheta B_s} \mathbb{E}\left[e^{i\vartheta(B_t-B_s)}\right] .$$

**Stopping times and Strong Markov property.** We now introduce the concept of stopping time. As the name suggests, a stopping time is a time at which one can stop the process. The accent in this sentence should be put on **can**, and is to be intended in the following sense: if someone is observing the process as it evolves, and is given the instructions on when to stop, I can stop the process given his/her/their observations. In other words, the observer does not need future information to know if the event triggering the stop of the process has occurred or not. We now define this concept formally:

DEFINITION 7.14. *For a measurable space  $(\Omega, \mathcal{F})$  and a filtration  $\{\mathcal{F}_t\}$  with  $\mathcal{F}_t \subseteq \mathcal{F}$  for all  $t \in \mathcal{T}$ , a random variable  $\{\tau\}$  is a stopping time WRT  $\{\mathcal{F}_t\}$  if  $\{\omega \in \Omega : \tau \leq t\} \in \mathcal{F}_t$  for all  $t \in \mathcal{T}$ .*

Classical examples of stopping times are *hitting times* such as the one defined in the following example

EXAMPLE 7.15. *The random time  $\tau_1 := \inf\{s > 0 : B_t \geq 1\}$  is a stopping time WRT the natural filtration of Brownian motion  $\{\mathcal{F}_t^B\}$ . Indeed, at any time  $t$  i can know if the event  $\tau_1$  has passed by looking at the past history of  $B_t$ . However, the random time  $\tau_0 := \sup\{s \in (0, t^*) : B_s = 0\}$  is NOT a stopping time WRT  $\{\mathcal{F}_t^B\}$  for  $t < t^*$ , as before  $t^*$  we cannot know for sure if the process will reach 0 again.*

The *strong* Markov property introduced below is a generalization of the Markov property to stopping times (as opposed to fixed times in Def. 7.12). More specifically, we say that a stochastic process has the strong Markov property if its future after any stopping time depends on its past only through the present (*i.e.*, its state at the stopping time).

DEFINITION 7.16. *The stochastic process  $\{X_t\}$  has the strong Markov property if for all finite stopping time  $\tau$  one has*

$$\mathbb{P}[X_{\tau+s} \in A | \mathcal{F}_\tau] = \mathbb{P}[X_{\tau+s} \in A | X_\tau] ,$$

where  $\mathcal{F}_\tau := \{A \in \mathcal{F}_t : \{\tau \leq t\} \cap A \in \mathcal{F}_t, \forall t > 0\}$ .

In the above definition, the  $\sigma$ -algebra  $\mathcal{F}_\tau$  can be interpreted as “all the information we have on the process up to time  $\tau$ ”.

THEOREM 7.17. *Brownian motion has the strong Markov property.*

We now use the above result to investigate some of the properties of Brownian motion:



EXAMPLE 7.18. For any  $t > 0$  define the maximum of Brownian motion in the interval  $[0, t]$  as  $M_t := \max_{s \in (0, t)} B_s$ . Similarly, for any  $m > 0$  we define the hitting time of  $m$  as  $\tau_m := \inf\{s \in [0, t] : B_s \geq m\}$ . Then, we write

$$\begin{aligned} \mathbb{P}[M_t \geq m] &= \mathbb{P}[\tau_m \leq t] = \mathbb{P}[\tau_m \leq t, B_t \geq m] + \mathbb{P}[\tau_m \leq t, B_t < m] \\ &= \mathbb{P}[\tau_m \leq t, B_t - B_{\tau_m} \geq 0] + \mathbb{P}[\tau_m \leq t, B_t - B_{\tau_m} < m]. \end{aligned}$$

Using the strong Markov property of Brownian motion we have that  $B_t - B_{\tau_m}$  is independent on  $\mathcal{F}_{\tau_m}$  and is a Brownian motion. So by symmetry of Brownian motion we have that

$$\mathbb{P}[\tau_m \leq t, B_t - B_{\tau_m} \geq 0] = \mathbb{P}[\tau_m \leq t, B_t - B_{\tau_m} \leq m] = \mathbb{P}[B_t \geq m],$$

and we conclude that

$$\mathbb{P}[M_t \geq m] = 2\mathbb{P}[B_t \geq m] = \frac{2}{\sqrt{2\pi t}} \int_m^\infty e^{-\frac{y^2}{2t}} dy.$$

This argument is called the reflection principle for Brownian motion. From the above argument one can also extract that

$$\lim_{T \rightarrow \infty} \mathbb{P}[\tau_m < T] = 1,$$

i.e., that the hitting times of Brownian motion are almost surely finite.

From the above example we can also derive the following formula

EXAMPLE 7.19. We will compute the probability density  $\rho_{\tau_m}(s)$  of the hitting time of level  $m$  by the Brownian motion, defined by  $\mathbb{P}[\tau_m \leq t] = \int_0^t \rho_{\tau_m}(s) ds$ . To do so we write

$$\mathbb{P}[\tau_m \leq t] = \mathbb{P}[M_t \geq m] = 2\mathbb{P}[B_t \geq m] = \sqrt{\frac{2}{\pi t}} \int_m^\infty e^{-\frac{y^2}{2t}} dy = \sqrt{2\pi} \int_{m/\sqrt{t}}^\infty e^{-u^2} du, \quad (3.12)$$

where in the last equality we made a change of variables  $u = y/\sqrt{t}$ . Now, differentiating (3.12) WRT  $t$  we obtain, by Leibnitz rule,

$$\rho_{\tau_m}(t) = \frac{d}{dt} \mathbb{P}[\tau_m \leq t] = \frac{|m|}{2\pi} t^{-3/2} e^{-\frac{m^2}{2t}}. \quad (3.13)$$

We immediately see from (3.13) that

$$\mathbb{E}[\tau_m] = \frac{|m|}{2\pi} \int_0^\infty s^{-1/2} e^{-\frac{m^2}{2s}} ds = \infty.$$

EXAMPLE 7.20. From the above computations we derive the distribution of zeros of Brownian motion in the interval  $[a, b]$  for  $0 < a < b$ . We start by computing the desired quantity on the interval  $[0, t]$  for an initial condition  $x$  which we assume WLOG to be  $x < 0$ :

$$\begin{aligned} \mathbb{P}[B_s = 0 \text{ for } s \in [0, t] | B_0 = x] &= \mathbb{P}\left[\max_{s \in [0, t]} B_s \geq 0 | B_0 = x\right] = \mathbb{P}\left[\max_{s \in [0, t]} B_s \geq -x | B_0 = 0\right] \\ &= \mathbb{P}[\tau_{-x} \leq t] = \frac{|x|}{2\pi} \int_0^t s^{-3/2} e^{-\frac{x^2}{2s}} ds. \end{aligned} \quad (3.14)$$

Since the above expression holds for all  $x$  we obtain the distribution of zeroes in the interval  $[a, b]$  by integrating (3.14) over all possible  $x$ , weighted by the probability of reaching  $x$  at time  $a$ :

$$\begin{aligned} \mathbb{P}[B_s = 0 \text{ for } s \in [a, b] | B_0 = 0] &= \int_{-\infty}^\infty \mathbb{P}[B_s = 0 \text{ for } s \in [a, b-a] | B_a = x] \mathbb{P}[B_a \in dx] \\ &= \int_{-\infty}^\infty \frac{|x|}{2\pi} \int_0^{b-a} s^{-3/2} e^{-\frac{x^2}{2s}} ds \sqrt{\frac{2}{\pi a}} e^{-\frac{x^2}{2a}} dx = \frac{2}{\pi} \arccos\left(\sqrt{\frac{a}{b}}\right). \end{aligned}$$

By taking the complement of the above we also obtain the probability that Brownian motion has no zeroes in the interval  $[a, b]$ :

$$\mathbb{P}[B_s \neq 0 \forall s \in [a, b]] = 1 - \mathbb{P}[B_s = 0 \text{ for } s \in [a, b]] = \frac{2}{\pi} \arcsin\left(\sqrt{\frac{a}{b}}\right).$$

The above result is referred to as the arcsine law for Brownian motion.

## 8. A glimpse of the connection with PDEs

The Gaussian

$$\rho(t, z) = \frac{e^{-\frac{z^2}{2t}}}{\sqrt{2\pi t}}$$

is the fundamental solution to the heat equation. By direct calculation one sees that if  $t > 0$  then  $\frac{\partial \rho}{\partial t} = \frac{1}{2} \frac{\partial^2 \rho}{\partial x^2}$ . There is a small problem at zero, namely  $\rho$  blows up. However, for any “nice” function  $\phi(x)$  (smooth with compact support),

$$\lim_{t \rightarrow 0} \int \rho(t, x) \phi(x) dx = \phi(0).$$

This is the definition of the “delta function”  $\delta(x)$ . (If this is uncomfortable to you, look at [16].) Hence we see that  $\rho(t, x)$  is the (weak) solution to

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \frac{1}{2} \frac{\partial^2 \rho}{\partial x^2} \\ \rho(0, x) &= \delta(x) \end{aligned}$$

To see the connection to probability, we set  $p(t, x, y) = \rho(t, x - y)$  and observe that for any function  $f$  we have

$$\mathbb{E}\{f(B_t) | B_0 = x\} = \int_{-\infty}^{\infty} f(y) p(t, x, y) dy$$

We will write  $\mathbb{E}_x f(B_t)$  for  $\mathbb{E}\{f(B_t) | B_0 = x\}$ . Now notice that if  $u(t, x) = \mathbb{E}_x f(B_t)$  then  $u(t, x)$  solves

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{1}{2} \frac{\partial^2 u}{\partial x^2} \\ u(0, x) &= f(x) \end{aligned}$$

This is Kolmogorov’s Backward equation. We can also write it in terms of the transition density  $p(t, x, y)$

$$\begin{aligned} \frac{\partial p}{\partial t} &= \frac{1}{2} \frac{\partial^2 p}{\partial x^2} \\ p(0, x, y) &= \delta(x - y) \end{aligned}$$

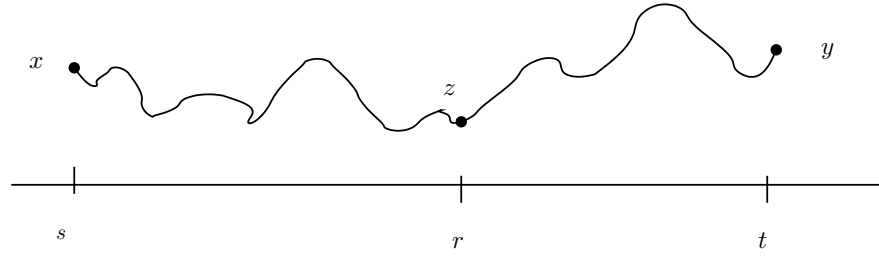
From this we see why it is called the “backwards” equation. It is a differential equation in the  $x$  variable. This is the “backwards” equation in  $p(t, x, y)$  in that it gives the initial point. This begs a question. Yes, there is also a forward equation. It is written in terms of the forward variable  $y$ .

$$\begin{aligned} \frac{\partial p}{\partial t} &= \frac{1}{2} \frac{\partial^2 p}{\partial y^2} \\ p(0, x, y) &= \delta(x - y) \end{aligned}$$

In this case it is identical to the backwards equation. In general it will not be.

We make one last observation: the  $p(t, s, x, y) = p(t - s, x, y) = \rho(t - s, x - y)$  satisfy the Chapman-Kolmogorov equation (the semi-group property). Namely, for any  $s < r < t$  and any  $x, y$  we have

$$p(s, t, x, y) = \int_{-\infty}^{\infty} p(s, r, x, z) p(r, t, z, y) dz$$



This also suggests the following form for the Kolmogorov forward equation. If we write an equation for  $p(s, t, x, y)$  evolving in  $s$  and  $y$ , then we get an equation with a final condition instead of an initial condition. Namely, for  $s \leq t$

$$\frac{\partial p}{\partial s} = -\frac{1}{2} \frac{\partial^2 p}{\partial y^2}$$

$$\rho(t, t, x, y) = \delta(x - y)$$

Hence, we are solving backwards in time.



## CHAPTER 4

### Itô Integrals

#### 1. Properties of the noise Suggested by Modeling

If we want to model a process  $X_t$  which was subject to random noise we might think of writing down a differential equation for  $X_t$  with a “noise” term regularly injecting randomness like the following:

$$\frac{dX_t}{dt} = f(X_t) + g(X_t) \cdot (\text{noise})_t$$

The way we have written it suggests the “noise” is generic and is shaped to the specific state of the system by the coefficient  $g(X_t)$ .

It is instructive to write this in integral form over the interval  $[0, t]$

$$X_t = X_0 + \int_0^t f(X_s) ds + \int_0^t g(X_s) \cdot (\text{noise})_s ds \quad (4.1)$$

it is reasonable to take the “noise” term to be a pure noise, independent of the structure of  $X_t$  and leave the “shaping” of the noise to a particular setting to the function  $g(X_t)$ . Since we want all moments of time to be the same it is reasonable to assume that distribution of “noise” is stationary in time. We would also like the noise at one moment to be independent of the noise at different moment. Since in particular both of these properties should hold when the function  $g \equiv 1$  we consider that simplified case to gain insight.

Defining

$$V_t = \int_0^t (\text{noise})_s ds$$

stationarity translates to the distribution of  $V_{t+h} - V_t$  being independent of  $t$ . Independence translates to

$$V_{t_1} - V_0, V_{t_2} - V_{t_1}, \dots, V_{t_{n+1}} - V_{t_n}$$

being a collection of mutually independent random variables for any collection of times

$$0 < t_1 < t_2 < \dots < t_n.$$

Rewriting (4.1) with  $g \equiv 1$  produces

$$X_t = X_0 + \int_0^t f(X_s) ds + V_t$$

we see that if further decide that we would like to model processes  $X_t$  which are continuous in time we need to require that  $t \mapsto V_t$  is almost surely a continuous process.

Clearly from our informal definition,  $V_0 = 0$ . Collecting all of the properties we desire of  $V_t$ :

- i)  $V_0 = 0$
- ii) Stationary increments
- iii) Independent increments
- iv)  $t \mapsto V_t$  is almost surely continuous.

Comparing this list with Theorem 3.6 we see that  $V_t$  must be a Brownian motion. We will chose it to be standard Brownian motion to fix a normalization.

Hence we are left to make sense of the integral equation

$$X_t = X_0 + \int_0^t f(X_s) dS + \int_0^t g(X_s) \frac{dB_s}{ds} ds$$

Of course this leads to its own problems since we saw in Section 5 that  $B_s$  is nowhere differentiable. Formally canceling the “ $ds$ ” maybe we can make sense of the integral

$$\int_0^t g(X_s) dB_s.$$

There is a well established classical theory of integrals in this form called “RiemannStieltjes” integration. We will briefly sketch this theory in the next section. However, we will see that even this theory is not applicable to the above integral. This will lead us to consider a new type of integration theory designed explicitly for random functions like Brownian motion. This named the Itô Integral after Kiyoshi Itô who developed the modern version though earlier version exist (notably in the work of Paley, Wiener and Zygmund).

## 2. Riemann-Stieltjes Integral

Before we try to understand how to integrate against Brownian motion, we recall the classical Riemann-Stieltjes integration theory. Given two continuous functions  $f$  and  $g$ , we want to define

$$\int_0^T f(t) dg(t). \quad (4.2)$$

We begin by considering a piecewise function  $\phi$  function defined by

$$\phi(t) = \begin{cases} a_0 & \text{for } t \in [t_0, t_1] \\ a_k & \text{for } t \in (t_k, t_{k+1}], k = 1, \dots, n-1 \end{cases} \quad (4.3)$$

for some partition

$$0 = t_0 < t_1 < \dots < t_{n-1} < t_n = T$$

and constants  $a_k \in \mathbb{R}$ . For such a function  $\phi$  it is intuitively clear that

$$\int_0^T \phi(t) dg(t) = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \phi(s) dg(s) = \sum_{k=0}^{n-1} a_k \int_{t_k}^{t_{k+1}} dg(s) = \sum_{k=0}^{n-1} a_k [g(t_{k+1}) - g(t_k)]. \quad (4.4)$$

because  $\int_{t_k}^{t_{k+1}} dg(s) = g(t_{k+1}) - g(t_k)$  by the fundamental theorem of Calculus (since  $\int_{t_k}^{t_{k+1}} dg(s) = \int_{t_k}^{t_{k+1}} g'(s) ds$  if  $g$  is differentiable).

The basic idea of defining (4.2) is to approximate  $f$  by a sequence of step functions  $\{\phi_n(t)\}$  each of the form given in (4.3) so that

$$\sup_{t \in [0, T]} |f(t) - \phi_n(t)| \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (4.5)$$

A natural choice of partition for the  $n$ th level is  $t_k^{(n)} = Tk2^{-n}$  for  $k = 0, \dots, 2^n$  and then define the  $n$ th approximating function by

$$\phi_n(t) = \begin{cases} f(T) & \text{if } t = T \\ f(t_k^{(n)}) & \text{if } t \in [t_k^{(n)}, t_{k+1}^{(n)}) \end{cases}$$

If  $f$  is continuous, it easy to see that (4.5) holds.

We then are left to show that there exists a constant  $\alpha$  so that

$$\left| \int_0^T \phi^{(n)}(t) dg(t) - \alpha \right| \longrightarrow 0 \quad \text{as } n \rightarrow \infty$$

We would then define the integral  $\int_0^T f(t) dg(t)$  to be equal to  $\alpha$ . One of the keys to proving this convergence is a uniform bound on the approximating integrals  $\int_0^T \phi^{(n)}(t) dg(t)$ . Observe that

$$\left| \int_0^T \phi^{(n)}(t) dg(t) \right| \leq \|f\|_\infty \sum_{k=0}^{n-1} |g(t_k) - g(t_{k+1})| \leq \|f\|_\infty V_1[g](0, T)$$

where

$$\|f\|_\infty = \sum_{t \in [0, T]} |f(t)|.$$

This uniform bound implies that the  $\int_0^T \phi^{(n)}(t) dg(t)$  say in a compact subset of  $\mathbb{R}$ . Hence there must be a limit point  $\alpha$  of the sequence and a subsequence which converges to it. It is not then hard to show that this limit point is unique. That is to say that if any other subsequence converges it must also converge to  $\alpha$ . We define the value of  $\int_0^T f(t) dg(t)$  to be  $\alpha$ . Hence it seems sufficient for  $g$  to have  $V_1[g](0, T) < \infty$  if when  $g$  and  $f$  are continuous. It can also be shown to be necessary for a reasonable class of  $f$ .

It is further possible to show using essentially the same calculations that the limit  $\alpha$  is independent of the sequence of partitions as long as the maximal spacing goes to zero and independent of the choice of point at which to evaluate the integrand  $f$ . In the above discussion we chose the left hand endpoint  $t_k$  of the intervals  $[t_k, t_{k+1}]$ . However we were free to choose any point in the interval.

While the compactness argument above is a standard path in mathematics is often more satisfying to explicitly show that the  $\{\phi^{(n)}\}$  are a Cauchy sequence by showing that for any  $\varepsilon > 0$  there exists and  $N$  so that if  $n, m > N$  then

$$\left| \int_0^T \phi^{(m)}(t) dg(t) - \int_0^T \phi^{(n)}(t) dg(t) \right| < \varepsilon$$

Since  $\phi^{(m)} - \phi^{(n)}$  is again a step function of the form (4.3), the integral  $\int_0^T [\phi^{(m)} - \phi^{(n)}](t) dg(t)$  is well defined given by a sum of the form (4.4). Hence we have

$$\left| \int_0^T \phi^{(m)}(t) dg(t) - \int_0^T \phi^{(n)}(t) dg(t) \right| = \left| \int_0^T [\phi^{(m)} - \phi^{(n)}](t) dg(t) \right| \leq \|\phi^{(m)} - \phi^{(n)}\|_\infty V_1[g](0, T)$$

Since  $f$  is continuous and the partition spacing is going to zero it is not hard so see that the  $\{\phi^{(n)}\}$  form a Cauchy sequence under the  $\|\cdot\|_\infty$  norm which completes the proof that integrals of the step functions form a Cauchy sequence.

### 3. A motivating example

We begin by considering the example

$$\int_0^T B_s dB_s \tag{4.6}$$

where  $B$  is a standard Brownian motion. Since  $V_1[B](0, T) = \infty$  almost surely we can not entirely follow the prescription of the Riemann-Stieltjes integral given in Section 2. However, it still seems reasonable to approximate the integrand  $B_s$  by a sequence of step functions of the form (4.3). However, since  $B$  is random, the  $a_k$  from (4.3) will have to be random variables.

Fixing a partition  $0 = t_0 < t_1 < \dots < t_N = T$ , we define two different sequences of step function approximations of  $B$ . For  $t \in [0, T]$ , we define

$$\begin{aligned}\phi^N(t) &= B(t_k) && \text{if } t \in [t_k, t_{k+1}) \\ \hat{\phi}^N(t) &= B(t_{k+1}) && \text{if } t \in (t_k, t_{k+1}]\end{aligned}$$

Just as in the Riemann-Stieltjes setting (see (4.4)), for such step functions it is clear that one should define the respective integrals in the following manner:

$$\begin{aligned}\int_0^T \phi^N(t) dB_s &= \sum_{i=0}^{N-1} B(t_k)[B(t_{k+1}) - B(t_k)] \\ \int_0^T \hat{\phi}^N(t) dB_s &= \sum_{i=0}^{N-1} B(t_{k+1})[B(t_{k+1}) - B(t_k)]\end{aligned}$$

In the Riemann-Stieltjes setting, the two are the same. But in this case, the two have very different properties as the following calculation shows. Since  $B(t_k)$  and  $B(t_{k+1}) - B(t_k)$  are independent we have  $\mathbb{E}B(t_k)[B(t_{k+1}) - B(t_k)] = \mathbb{E}[B(t_k)]\mathbb{E}[B(t_{k+1}) - B(t_k)] = 0$ . So

$$\begin{aligned}\mathbb{E}\left[\int_0^T \phi^N dB_s\right] &= \sum_{k=0}^{N-1} \mathbb{E}B(t_k)[B(t_{k+1}) - B(t_k)] \\ &= \sum_{k=0}^{N-1} \mathbb{E}[B(t_k)]\mathbb{E}[B(t_{k+1}) - B(t_k)] = 0\end{aligned}$$

While since  $\mathbb{E}[B(t_{k+1}) - B(t_k)]^2 = t_{k+1} - t_k$ , we have

$$\begin{aligned}\mathbb{E}\left[\int_0^T \hat{\phi}^N dB_s\right] &= \mathbb{E}\sum_{i=0}^{N-1} B(t_{k+1})[B(t_{k+1}) - B(t_k)] + \mathbb{E}\sum_{k=0}^{N-1} [B(t_{k+1}) - B(t_k)]^2 \\ &= 0 + \sum_{k=0}^{N-1} [t_{k+1} - t_k] = T\end{aligned}$$

Hence, how we construct our step functions will be important in our analysis. The choice of the endpoint used in  $\phi^N(t)$  leads to what is called the *Itô integral*. The choice used in  $\hat{\phi}^N(t)$  is called the *Klimontovich Integral*. While if the midpoint is chosen, this leads to the *Stratonovich integral*. The question on which to use is a modeling question and is dependent on the problem being studied. We will see that it is possible to translate between all three in most cases. We will concentrate on the Itô integral since it has some nice additional properties which make the analysis attractive.

#### 4. Itô integrals for a simple class of step functions

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and  $B_t$  be a standard Brownian motion. Let  $\mathcal{F}_t$  be a filtration on the probability space to which  $B_t$  is adapted. For example, one could have  $\mathcal{F}_t = \sigma(B_s : s \leq t)$  be the filtration generated by the Brownian motion  $B_t$ .

**DEFINITION 4.1.**  $\phi(t, \omega)$  is an elementary stochastic process if there exists a collection of bounded, disjoint intervals  $\{I_k\} = \{[t_0, t_1), [t_1, t_2), \dots, [t_{N-1}, t_N)\}$  associated to a partition  $0 \leq t_0 < t_1 < \dots < t_N$  and a collection of random variables  $\{\alpha_k : k = 0, \dots, N\}$  so that

$$\phi(t, \omega) = \sum_{k=0}^N \alpha_k(\omega) \mathbf{1}_{I_k}(t),$$



the random variable  $\alpha_k(\omega)$  is measurable with respect to  $\mathcal{F}_{t_k}$ , and  $\mathbb{E}[|\alpha_k(\omega)|^2] < \infty$ . We denote the space of all elementary functions as  $\mathcal{S}_2$ .

To be precise, stochastic integrals are defined on the class of *progressively measurable* processes, defined below:

DEFINITION 4.2. A stochastic process  $\{X_t\}_{t \geq 0}$  on  $(\Omega, \mathcal{F}, \mathbb{P}, \{F_t\})$  is called *progressively measurable*, if for any  $t \geq 0$ ,  $X_t(\omega)$ , viewed as a function of two variables  $(t, \omega)$  is  $\mathcal{B}_{[0,t]} \times \mathcal{F}_t$ -measurable, where  $\mathcal{B}_{[0,t]}$  is the Borel  $\sigma$ -algebra on  $[0, t]$ .

FACT 1. Every adapted right continuous with left limits (“cadlag”) or left continuous with right limits process is progressively measurable. The reason to assume progressive measurability is to ensure that the expectation and the integral can be interchanged (by Fubini’s theorem).

FACT 2. Any progressively measurable process  $X = \{X_t\}_{t \in [0, T]}$  can be approximated by a sequence of simple processes  $X^n = \{X_t^n\}_{t \in [0, T]} \in \mathcal{S}_2$  in the  $L_2$ -sense, that is

$$\mathbb{E} \left[ \int_0^T |X_t^n - X_t|^2 dt \right] = \int_0^T \mathbb{E} [|X_t^n - X_t|^2] dt \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

The proof of this approximation can be found in [13].

Because of the above result, we will be able to extend the notion of integral to adapted processes, and we restrict our attention to such processes for the rest of the chapter.

Next, we define a functional  $I$  which will be our integral operator. That is to say  $I(\phi) = \int_0^\infty \phi dB$ . Just as for Riemann-Stieltjes integral, if  $\phi$  is an elementary stochastic process, it is relatively clear what we should mean by  $I(\phi)$ , namely

DEFINITION 4.3. The stochastic integral operator of an elementary stochastic process is given by

$$I(\phi) := \sum_k \alpha_k [B(t_{k+1}) - B(t_k)].$$

We first observe that  $I$  satisfies one of the standard properties of an integral in that it is a linear functional. In other words, if  $\lambda \in \mathbf{R}$ , and  $\phi$  and  $\psi$  are elementary stochastic processes then

$$I(\lambda\phi) = \lambda I(\phi) \quad \text{and} \quad I(\psi + \phi) = I(\psi) + I(\phi) \quad (4.7)$$

Thanks to our requirement that  $\alpha_k$  are measurable with respect to the filtration associated to the left endpoint of the interval  $[t_k, t_{k+1})$  we have the following properties which will play a central role in what follows and should be compared to the calculations in Section 3.

LEMMA 4.4. If  $\phi$  is an elementary stochastic processes then

$$\begin{aligned} \mathbb{E}I(\phi) &= 0 \quad (\text{mean zero}) \\ \mathbb{E}[I(\phi)^2] &= \int_0^\infty \mathbb{E}|\phi(t)|^2 ds \quad (\text{Itô Isometry}) \end{aligned}$$

REMARK 4.5. An isometry is a map between two spaces which preserves distance (i.e. the norm). If we consider

$$I(\mathcal{S}_2) = \{I(\phi) : \phi \in \mathcal{S}_2\}$$

then according to Lemma 4.4 the map  $I(\phi) \mapsto \phi$  is an isometry between the space of random variables

$$L^2(I(\mathcal{S}_2), \mathbb{P}) = \{X \in I(\mathcal{S}_2) : \|X\| = \sqrt{\mathbb{E}(X^2)} < \infty\}$$

and the space of elementary stochastic processes  $L^2(\mathcal{S}_2, \mathbb{P}[d\omega] \times dt)$  equipped with the norm

$$\|\phi\| = \left( \int_0^\infty \mathbb{E}\phi^2(t, \omega) dt \right)^{\frac{1}{2}}$$

PROOF OF LEMMA 4.4. We begin by showing  $I(\phi)$  is mean zero.

$$\begin{aligned}\mathbb{E}[I(\phi)] &= \sum_k \mathbb{E}[\alpha_k(B(t_{k+1}) - B(t_k))] \\ &= \sum_k \mathbb{E}[\mathbb{E}[\alpha_k(B(t_{k+1}) - B(t_k)) | \mathcal{F}_{t_k}]] \\ &= \sum_k \mathbb{E}[\alpha_k \mathbb{E}[B(t_{k+1}) - B(t_k) | \mathcal{F}_{t_k}]] = 0\end{aligned}$$

Turning to the Itô isometry

$$\begin{aligned}\mathbb{E}[I(\phi)^2] &= \left[ \mathbb{E} \left( \sum_k \alpha_k (B(t_{k+1}) - B(t_k)) \right) \right] \cdot \left[ \mathbb{E} \left( \sum_j \alpha_j (B(t_{j+1}) - B(t_j)) \right) \right] \\ &= \sum_{j,k} \mathbb{E}[\alpha_k \alpha_j [B(t_{k+1}) - B(t_k)][B(t_{j+1}) - B(t_j)]] \\ &= 2 \sum_{j < k} \mathbb{E}[\alpha_k \alpha_j [B(t_{k+1}) - B(t_k)][B(t_{j+1}) - B(t_j)]] \\ &\quad + \sum_k \mathbb{E}[\alpha_k^2 [B(t_{k+1}) - B(t_k)]^2]\end{aligned}$$

Next, we examine each component separately: Recall for  $t_{k+1} \leq t_j$

$$\begin{aligned}\mathbb{E}[\alpha_k \alpha_j [B(t_{k+1}) - B(t_k)][B(t_{j+1}) - B(t_j)]] \\ &= \mathbb{E}[\mathbb{E}[\alpha_k \alpha_j [B(t_{k+1}) - B(t_k)][B(t_{j+1}) - B(t_j)] | \mathcal{F}_{t_j}]] \\ &= \mathbb{E}[\alpha_k \alpha_j [B(t_{k+1}) - B(t_k)] \mathbb{E}[[B(t_{j+1}) - B(t_j)] | \mathcal{F}_{t_j}]] \\ &= 0\end{aligned}$$

since  $\mathbb{E}[[B(t_{j+1}) - B(t_j)] | \mathcal{F}_{t_j}] = 0$ . Similarly

$$\begin{aligned}\sum_k \mathbb{E}[\alpha_k^2 [B(t_{k+1}) - B(t_k)]^2] &= \sum_k \mathbb{E}[\mathbb{E}[\alpha_k^2 [B(t_{k+1}) - B(t_k)]^2 | \mathcal{F}_{t_k}]] \\ &= \sum_k \mathbb{E}[\alpha_k^2 (t_{k+1} - t_k)]\end{aligned}$$

Hence, we have:

$$\mathbb{E}[I(\phi)^2] = 0 + \sum_k \mathbb{E}[\alpha_k^2] (t_{k+1} - t_k) = \int \mathbb{E}[\phi^2(s)] ds$$

□

So far we have just defined the Itô Integral on the whole positive half line  $[0, \infty)$ . For any  $0 \leq s < t \leq \infty$ , we make the following definition

$$\int_s^t \phi_r dB_r = I(\phi \mathbf{1}_{[s,t]})$$

We can now talk about the stochastic process

$$M_t = I(\phi \mathbf{1}_{[0,t]}) = \int_0^t \phi_s dB_s \tag{4.8}$$

associated to a given elementary stochastic process  $\phi_t$ , where the last two expressions are just different notation for the same object.

We now state a few simple consequences of our definitions.

LEMMA 4.6. Let  $\phi \in \mathcal{S}_2$  and  $0 < s < t$  then

$$\int_0^t \phi_r dB_r = \int_0^s \phi_r dB_r + \int_s^t \phi_r dB_r$$

and

$$M_t = \int_0^t \phi_s dB_s$$

is measurable with respect to  $\mathcal{F}_t$ .

PROOF OF LEMMA 4.6. Clearly  $M_t$  is measurable with respect to  $\mathcal{F}_t$  since  $\{\phi_s : s \leq t\}$  are by assumption and the construction of the integral only uses the information from  $\{B_s : s \leq t\}$ . The first property follows from  $\phi \mathbf{1}_{[0,t]} = \phi \mathbf{1}_{[0,s]} + \phi \mathbf{1}_{[s,t]}$  and hence

$$I(\phi \mathbf{1}_{[0,t]}) = I(\phi \mathbf{1}_{[0,s]}) + I(\phi \mathbf{1}_{[s,t]})$$

by (4.7). □

LEMMA 4.7. Let  $M_t$  be as in (4.8) for an elementary process  $\phi$  and a Brownian motion  $B_t$  both adapted to a filtration  $\{\mathcal{F}_t : t \geq 0\}$ . Then  $M_t$  is a martingale with respect to the filtration  $\mathcal{F}_t$ .

PROOF OF LEMMA 4.7. Looking at Definition 7.8, there are three conditions we need to verify. The measurability is contained in Lemma 4.6. The fact that  $\mathbb{E}[M_t^2] < \infty$  follows from the Itô isometry since

$$\mathbb{E}[M_t^2] = \int_0^t \mathbb{E}|\phi_s|^2 ds \leq \int_0^\infty \mathbb{E}|\phi_s|^2 ds$$

because the last integral is assumed to be finite in the definition of an elementary stochastic process.

All that remains is to verify that for  $s < t$ .

$$\mathbb{E}[M_t(\phi) - M_s(\phi) | \mathcal{F}_s] = 0.$$

There are a few cases. Taking one case, say  $s$  and  $t$  are in the disjoint intervals  $[t_k, t_{k+1})$  and  $[t_j, t_{j+1})$ , respectively. We have that

$$M_t(\phi) - M_s(\phi) = \alpha_k [B(t_{k+1}) - B_s] + \left( \sum_{n=k+1}^{j-1} \alpha_n [B(t_{n+1}) - B(t_n)] \right) + \alpha_j [B_t - B(t_j)]$$

Next, take repeated expectations with respect to the filtrations  $\{\mathcal{F}_a\}$ , where  $a \in \{t_j, t_{j-1}, \dots, t_{k+1}, s\}$ . This would then imply that for each  $a$

$$\begin{aligned} \mathbb{E}[\alpha_a [B(t_{a+1}) - B(t_a)] | \mathcal{F}_a] &= \alpha_a \mathbb{E}[B(t_{a+1}) - B(t_a) | \mathcal{F}_a] \\ &= \alpha_a \mathbb{E}[B(t_{a+1}) - B(t_a)] \\ &= 0 \end{aligned}$$

Hence,  $\mathbb{E}[M_t(\phi) - M_s(\phi) | \mathcal{F}_s] = 0$ . The other cases can be done similarly. And the conclusion immediately follows. □

LEMMA 4.8. In the same setting as Lemma 4.7,  $M_t$  is a continuous stochastic process. (That is to say, with probability one the map  $t \mapsto M_t$  is continuous for all  $t \in [0, \infty)$ ).

PROOF OF LEMMA 4.8. We begin by noticing that if  $\phi(t, \omega)$  is a simple process with

$$\phi(t, \omega) = \sum_{k=1}^{N-1} \alpha_k(\omega) \mathbf{1}_{[t_k, t_{k+1})}(t)$$

then if  $t \in (t_{k^*}, t_{k^*+1})$  then

$$M_t(\phi, \omega) = \sum_{k=1}^{k^*-1} \alpha_k(\omega)[B(t_{k+1}, \omega) - B(t_k, \omega)] + \alpha_{k^*}(\omega)[B(t, \omega) - B(t_{k^*}, \omega)]$$

Hence it is clear that  $M_t(\phi, \omega)$  is continuous if  $\phi$  is a simple function since the Brownian motion  $B_t$  is continuous.  $\square$

## 5. Extension to the Closure of Elementary Processes

We will denote by  $\bar{\mathcal{S}}_2$  the closure in  $L^2(\mathbb{P} \times dt)$  of the square integrable elementary processes  $\mathcal{S}_2$ . Namely,

$$\bar{\mathcal{S}}_2 = \left\{ \text{all stochastic processes } f(t, \omega) : \text{there exist a sequence } \phi_n(t, \omega) \in \mathcal{S}_2 \right. \\ \left. \text{so that } \int_0^\infty \mathbb{E}(f(t) - \phi_n(t))^2 dt \rightarrow 0 \text{ as } n \rightarrow \infty \right\}$$

Also recall that we define

$$L^2(\Omega, \mathbb{P}) := \left\{ f : \Omega \rightarrow \mathbb{R} : \int_\Omega f(\omega)^2 \mathbb{P}[d\omega] < \infty \right\} \\ = \{ \text{Random Variables } X \text{ with } \mathbb{E}(X^2) < \infty \}$$

and

$$L^2(\Omega \times [0, \infty), \mathbb{P} \times dt) := \left\{ f : \Omega \times [0, \infty) : \int_0^\infty \int_\Omega f(t, \omega)^2 \mathbb{P}[d\omega] dt < \infty \right\} \\ = \left\{ \text{Stochastic Processes } X_t \text{ with } \int_0^\infty \mathbb{E}(X_t^2) dt < \infty \right\}$$

In general, in the interest of brevity we will write  $L^2(\Omega)$  for the first and  $L^2(\Omega \times [0, \infty))$  for the second space above. We will occasionally write  $\|X\|_{L^2(\Omega)}$  for  $\sqrt{\mathbb{E}(X^2)}$  and  $\|X\|_{L^2(\Omega \times [0, \infty))}$  for  $(\int_0^\infty \mathbb{E}(X_t^2) dt)^{\frac{1}{2}}$  though we will simply write  $\|\cdot\|_{L^2}$  when the context is clear.

Recalling that our definition of  $\mathcal{S}_2$  and hence  $\bar{\mathcal{S}}_2$  had a filtration  $\mathcal{F}_t$  in the background, we define  $L_{\text{ad}}^2(\Omega \times [0, \infty))$  to be the subset of  $L^2(\Omega \times [0, \infty))$  in which all of the stochastic processes are adapted to the filtration  $\mathcal{F}_t$ . We have the following simple observation

LEMMA 5.1. *The closure of the space of elementary processes is contained in the space of square integrable, adapted processes, i.e.,  $\bar{\mathcal{S}}_2 \subset L_{\text{ad}}^2(\Omega \times [0, \infty))$ .*

PROOF. The adaptedness follows from the adaptedness of  $\mathcal{S}_2$ . The fact that elements in  $\bar{\mathcal{S}}_2$  are square integrable follows from the following calculation. Fixing an  $X \in \bar{\mathcal{S}}_2$  and a sequence  $\phi_n \in \bar{\mathcal{S}}_2$  so that  $\|\phi_n - X\| \rightarrow 0$  as  $n \rightarrow \infty$  we fix an  $n$  so that  $\|\phi_n - X\| \leq 1$ . Then since for any real numbers  $a$  and  $b$  and  $p \geq 1$  one has  $|a + b|^2 \leq 2^{p-1}|a|^2 + 2^{p-1}|b|^p$  we see that

$$\|f\|_{L^2}^2 = \|f_n + (f - \phi_n)\|_{L^2}^2 \leq 2\|\phi_n\|_{L^2}^2 + 2\|f - \phi_n\|_{L^2}^2 \leq 2(\|\phi_n\|_{L^2}^2 + 1) < \infty$$

where the term on the far right is finite since for every  $n$ ,  $\phi_n \in \mathcal{S}_2$  (observe that  $\|\phi_n\|_{L^2}^2 = \sum \mathbb{E}(\alpha_k^2)$ ).  $\square$

In fact,  $\bar{\mathcal{S}}_2$  is *exactly*  $L_{\text{ad}}^2(\Omega \times [0, \infty))$ ,

THEOREM 5.2.  $\bar{\mathcal{S}}_2 = L_{\text{ad}}^2(\Omega \times [0, \infty))$

PROOF.  $\bar{\mathcal{S}}_2 \subset L_{\text{ad}}^2(\Omega \times [0, \infty))$  was just proven in Lemma 5.1. For the other direction, see the proof of [13, Theorem 3.1.5] or [11]  $\square$

We are now ready to state the main theorem of this section. This result shows that given a Cauchy sequence of elementary stochastic processes  $\{\phi_n\} \in \mathcal{S}_2$  converging in  $L^2(\Omega, [0, \infty))$  to a process  $f \in \bar{\mathcal{S}}_2$ , there exists a random variable  $X \in L^2(\Omega)$  to which the stochastic integrals  $I(\phi_n)$  converge in  $L^2(\Omega)$ . We will then define this variable  $X$  to be the stochastic integral  $I(f) = \int_0^\infty f_s dB_s$ . The concept of the construction is summarized in the following scheme:

$$\begin{array}{ccc}
\{\phi_n(t)\}_{t \in \mathcal{T}} \in \mathcal{S}^2 & \xleftarrow{I(\cdot)} & I(\mathcal{S}^2) \ni \int \phi_n dB_s \\
\int_0^\infty \mathbb{E}[(\phi_n - f)^2] dt \rightarrow 0 & \begin{array}{c} \downarrow n \rightarrow \infty \\ \downarrow n \rightarrow \infty \end{array} & \mathbb{E}[(I(\phi_n) - X)^2] \rightarrow 0 \\
\{f(t)\}_{t \in \mathcal{T}} \in \bar{\mathcal{S}}^2 & \xleftarrow{I(\cdot)} & L^2(\Omega) \ni X =: \int f dB_s
\end{array}$$

**THEOREM 5.3.** *For every  $f \in \bar{\mathcal{S}}_2$ , there exists a random variable  $X \in L^2(\Omega)$  so that if  $\{\phi_n : n = 1, \dots, \infty\}$  is a sequence of elementary stochastic processes (i.e. elements of  $\mathcal{S}_2$ ) converging to the stochastic process  $f$  in  $L^2(\Omega \times [0, \infty))$  then  $I(\phi_n)$  converges to  $X$  in  $L^2(\Omega)$ .*

**DEFINITION 5.4.** *For any  $f \in \bar{\mathcal{S}}_2$ , we define the Itô integral*

$$I(f) = \int_0^\infty f_t dB_t$$

*to be the random variable  $X \in L^2(\omega)$  given by Theorem 5.3.*

**PROOF OF THEOREM 5.3.** We start by showing that  $I(\phi_n)$  is a Cauchy sequence in  $L^2(\Omega)$ . By the linearity of the map  $I$  (see (4.7)),  $I(\phi_n) - I(\phi_m) = I(\phi_n - \phi_m)$ . Hence by the Itô isometry for the elementary stochastic processes (see lemma 4.4), we have that

$$\mathbb{E}[(I(\phi_n) - I(\phi_m))^2] = \mathbb{E}[I(\phi_n - \phi_m)^2] = \int_0^\infty \mathbb{E}[(\phi_n(t) - \phi_m(t))^2] dt \quad (4.9)$$

Since the sequence  $\phi_n$  converges to  $f$  in  $L^2(\Omega \times [0, \infty))$ , we know that it is a Cauchy sequence in  $L^2(\Omega \times [0, \infty))$ . By the above calculations we hence have that  $\{I(\phi_n)\}$  is a Cauchy sequence in  $L^2(\Omega)$ . It is a classical fact from real analysis that this space is complete which implies that every Cauchy sequence converges to a point in the same space. Let  $X \in L^2(\Omega)$  denote the limit.

To see that  $X$  does not depend on the sequence  $\{\phi_n\}$ , let  $\{\tilde{\phi}_n\}$  be another sequence converging to  $f$ . The same reasoning as above ensures the existence of a  $\tilde{X} \in L^2(\Omega)$  so that  $I(\tilde{\phi}_n) \rightarrow \tilde{X}$  in  $L^2(\Omega)$ . On the other hand

$$\begin{aligned}
\mathbb{E}[(I(\phi_n) - I(\tilde{\phi}_n))^2] &= \mathbb{E}[(I(\phi_n - \tilde{\phi}_n))^2] = \int_0^\infty \mathbb{E}[(\phi_n(t) - \tilde{\phi}_n(t))^2] dt \\
&\leq 2 \int_0^\infty \mathbb{E}[(\phi_n(t) - f)^2] dt + 2 \int_0^\infty \mathbb{E}[(f - \tilde{\phi}_n(t))^2] dt
\end{aligned}$$

where in the inequality we have again used the fact that  $(\phi_n - \tilde{\phi}_n)^2 = [(\phi_n - f) + (f - \tilde{\phi}_n)]^2 \leq 2(\phi_n - f)^2 + 2(f - \tilde{\phi}_n)^2$ . Since both  $\phi_n$  and  $\tilde{\phi}_n$  converge to  $f$  in  $L^2(\Omega \times [0, \infty))$  we have that the last two terms on the righthand side go to 0. This in turn implies that  $\mathbb{E}(X - \tilde{X})^2 = 0$  and that the two random variables are the same.  $\square$

REMARK 5.5. While the above construction might seem like magic since it is so soft, it is an application of a general principle in mathematics. If one can define a linear map on a dense subset of elements in a space in such a way that the map is an isometry then the map can be uniquely extend to the whole space. This approach is beautifully presented in our current context in [10] using the following lemma

LEMMA 5.6. (Extension Theorem) Let  $\mathcal{B}_1$  and  $\mathcal{B}_2$  be two Banach spaces. Let  $\mathcal{B}_0 \subset \mathcal{B}_1$  be a linear space. If  $L : \mathcal{B}_0 \rightarrow \mathcal{B}_2$  is defined for all  $b \in \mathcal{B}_0$  and  $|Lb|_{\mathcal{B}_2} = |Lb|_{\mathcal{B}_1}, \forall b \in \mathcal{B}_0$ . Then there exists a unique representation of  $L$  to  $\overline{\mathcal{B}_0}$  (closure of  $\mathcal{B}_0$ ) called  $\overline{L}$  with  $Lb = \overline{L}b, \forall b \in \mathcal{B}_0$ .

EXAMPLE 5.7. We use the above theorem to show that

$$I_T(B_s) = \int_0^T B_s dB_s = \frac{1}{2}(B_T^2 - T).$$

To do so, we show that the sequence  $\phi_n = \sum_{j=1}^N \mathbb{1}_{[t_j^{(n)}, t_{j+1}^{(n)})} B_{t_j^{(n)}} \Delta B_j^n$  with  $\Delta t_j^n = t_{j+1}^{(n)} - t_j^{(n)} \rightarrow 0$  converges in  $L^2(\Omega, [0, T])$  to  $\{B_t\}$ . Indeed we have that

$$\begin{aligned} \int_0^T \mathbb{E} [|\phi_n - B_s|^2] dt &= \sum_{j=1}^N \int_{t_j^{(n)}}^{t_{j+1}^{(n)}} (B_{t_j^{(n)}} - B_s)^2 ds = \sum_{j=1}^N \int_{t_j^{(n)}}^{t_{j+1}^{(n)}} (t_{j+1}^{(n)} - s) ds \\ &= \frac{1}{2} \sum_{j=1}^N (t_{j+1}^{(n)} - t_j^{(n)})^2 \rightarrow 0. \end{aligned}$$

Then, by Theorem 5.3 we have that

$$\int_0^T B_s dB_s = \lim_{n \rightarrow \infty} I(\phi_n) = \lim_{n \rightarrow \infty} \sum_j B_{t_j^{(n)}} \Delta B_j^n.$$

Now, writing  $\Delta B_j^n := B_{t_j^{(n)}} - B_{t_{j+1}^{(n)}}$  we have

$$\Delta(B_j^2) := B_{t_j^{(n)}}^2 - B_{t_{j+1}^{(n)}}^2 = (B_{t_j^{(n)}} - B_{t_{j+1}^{(n)}})^2 + 2B_{t_j^{(n)}}(B_{t_j^{(n)}} - B_{t_{j+1}^{(n)}}) = (\Delta B_j^n)^2 + 2B_{t_j^{(n)}} \Delta B_j^n$$

and therefore

$$\sum_j B_{t_j^{(n)}} \Delta B_j^n = \frac{1}{2} \left( \sum_j \Delta(B_j^2) - \sum_j (\Delta B_j^n)^2 \right) = \frac{1}{2} \left( B_T^2 - \sum_j (\Delta B_j^n)^2 \right).$$

The term on the RHS converges to  $(B_T^2 - T)/2$  in  $L^2(\Omega)$ , which therefore corresponds to  $\int_0^T B_s dB_s$ .

## 6. Properties of Itô integrals

PROPOSITION 6.1. Let  $f, g \in L_{\text{ad}}^2(\Omega, [0, \infty))$ ,  $\lambda \in \mathbb{R}$ , then

i) *Linearity:*

$$\int_0^\infty (\lambda f_s + g_s) dB_s = \lambda \int_0^\infty f_s dB_s + \int_0^\infty g_s dB_s, \quad (4.10)$$

ii) *Separability:* for all  $S > 0$ ,

$$\int_0^\infty f_s dB_s = \int_0^S f_s dB_s + \int_S^\infty f_s dB_s,$$

iii) *Mean 0:*

$$\mathbb{E} \left[ \int_0^\infty f_s dB_s \right] = 0, \quad (4.11)$$

iv) *Itô isometry:*

$$\mathbb{E} \left[ \left( \int_0^\infty f_s dB_s \right)^2 \right] = \int_0^\infty \mathbb{E} [f_s^2] dt. \quad (4.12)$$

PROOF. For the proof of points i) and ii) we refer to [9].

To prove iv), for  $f \in L_{\text{ad}}^2(\Omega, [0, \infty))$ , let  $\{\phi_n\} \in \mathcal{S}_2$  be a sequence of elementary stochastic processes converging to  $f$  in  $L_{\text{ad}}^2(\Omega, [0, \infty))$ . By Theorem 5.3, there exists  $X \in L_2(\Omega)$  with  $\mathbb{E} [(X - I(\phi_n))^2] \rightarrow 0$  as  $n \rightarrow \infty$ . Since  $\mathbb{E} [X^2], \mathbb{E} [I(\phi_n)^2] < \infty$  using Cauchy-Schwartz (or Hölder) inequality we write

$$\begin{aligned} |\mathbb{E} [I(\phi_n)^2] - \mathbb{E} [X^2]| &= |\mathbb{E} [(X - I(\phi_n))(X + I(\phi_n))]| \\ &\leq |\mathbb{E} [(X - I(\phi_n))I(\phi_n)] + \mathbb{E} [(X - I(\phi_n))X]| \\ &\leq \sqrt{\mathbb{E} [X^2] + \mathbb{E} [I(\phi_n)^2]} \sqrt{\mathbb{E} [(X - I(\phi_n))^2]}, \end{aligned}$$

with the term on the right hand side  $\sqrt{\mathbb{E} [(X - I(\phi_n))^2]} = \|X - I(\phi_n)\|_{L^2(\omega)} \rightarrow 0$  as  $n \rightarrow \infty$ . This implies that in the limit  $n \rightarrow \infty$  we have

$$\mathbb{E} [I(\phi_n)^2] \rightarrow \mathbb{E} [X^2].$$

At the same time,

$$\mathbb{E} I(\phi_n)^2 = \int_0^\infty \mathbb{E} (\phi_n(t))^2 dt \rightarrow \int_0^\infty \mathbb{E} (f(t)^2) dt.$$

Combining these facts produces

$$\mathbb{E} (X^2) = \int_0^\infty \mathbb{E} (f(t)^2) dt$$

as desired. The exact same logic produces  $\mathbb{E} X = 0$ .  $\square$

## 7. A continuous in time version of the the Itô integral

In this section, we consider the

DEFINITION 7.1. *For any  $f \in L_{\text{ad}}(\Omega \times [0, \infty))$  and any  $t \geq 0$ , we define the Itô integral process  $\{I_t(f)\}$  as*

$$I_t(f) = \int_0^t f_s(\omega) dB_s(\omega) := I(f \mathbf{1}_{(0,t]}).$$

Note that the process introduced above is well defined and adapted to  $\mathcal{F}_t$ . In fact since

$$\mathbb{E} I_t(f)^2 = \int_0^t \mathbb{E} f_t^2 dt \leq \int_0^\infty \mathbb{E} f_t^2 dt < \infty$$

we see that  $I_t(f)$  is an adapted stochastic process whose second moment is uniformly bounded in time. It is not immediately clear that we can fix the realization  $\omega$  (which in turn fixes the realization of  $f$  and  $W$ ) and then change the time  $t$  in  $I_t(f)(\omega) = \int_0^t f_s(\omega) dB_s(\omega)$ . We built the integral as some limit at each time  $t$ . Changing the time  $t$  requires us to repeat the limiting process. This would be fine except that we built the Itô integral through an  $L^2$ -limit. Hence at each time it is only defined up to sets of probability zero. If we only want to define  $I_t$  for some countable sequence of time then this still would not be a problem because the countable union of measure zero sets is still measure zero. However, we would like to define  $I_t(f)$  for all  $t \in [0, T]$ . This is a problem and more work is required.

THEOREM 7.2. *Let  $f \in L_{\text{ad}}^2(\Omega, [0, \infty))$ , then there exists a continuous version of  $I_t(f)$ , i.e., there exists a  $t$ -continuous stochastic process  $\{J_t\}$  on  $(\Omega, \mathcal{F}, \mathbb{P})$  such that for all  $t \in \mathcal{T}$*

$$\mathbb{P} [J_t = I_t(f)] = 1.$$

To prove this result, we state a very useful theorem which we will prove later.

**THEOREM 7.3** (Doob's Martingale Inequality). *Let  $M_t$  be a continuous-time martingale with respect to the filtration  $\mathcal{F}_t$ . Then*

$$\mathbf{P}\left(\sup_{0 \leq t \leq T} |M_t| \geq \lambda\right) \leq \frac{\mathbb{E}[|M_T|^p]}{\lambda^p}$$

where  $\lambda \in \mathbf{R}^+$ , and  $p \geq 1$ .

**PROOF OF THEOREM 7.2.** By Lemma 4.8 we know that  $I_t(\phi, \omega)$  is continuous if  $\phi$  is a simple function. If we had a sequence of simple functions  $\phi_n$  converging to  $f$  in  $L^2$ , we would like to “transfer” the continuity of  $\phi$  to  $f$ . To do so we use the following fact: “The uniform limit of continuous functions is continuous”. In other words, if  $f_n$  is continuous and  $\sup_{t \in [0, T]} |f_n(t) - f(t)| \rightarrow 0$  as  $n \rightarrow \infty$  then  $f$  is continuous.

To do so, let  $I_t(\phi_n) = \int_0^t \phi_n(s) dB_s$  and consider a Cauchy sequence  $\{\phi_n\}$ , for which we have

$$\begin{aligned} \mathbb{P}\left[\sup_{0 \leq t \leq T} |I_t(\phi_n) - I_t(\phi_m)| > \varepsilon\right] &\leq \frac{1}{\varepsilon^2} \mathbb{E}[|I_T(\phi_n) - I_T(\phi_m)|^2] \\ &\leq \frac{1}{\varepsilon^2} \mathbb{E}\left[\int_0^T (\phi_n(s) - \phi_m(s))^2 ds\right] \end{aligned}$$

The last term goes to zero as  $n, m \rightarrow \infty$ . Hence we can find a subsequence  $\{n_k\}$  so that

$$\mathbb{P}\left[\sup_{0 \leq t \leq T} |I_t(\phi_{n_{k+1}}) - I_t(\phi_{n_k})| > \frac{1}{k^2}\right] \leq \frac{1}{2^k}$$

If we set

$$A_k = \left\{ \sup_{0 \leq t \leq T} |I_t(\phi_{n_{k+1}}) - I_t(\phi_{n_k})| > \frac{1}{k^2} \right\}$$

then  $\sum_k \mathbb{P}[A_k] \leq \sum_k 2^{-k} < \infty$ . Hence the Borel-Cantelli lemma tells us that there is an random  $N(\omega)$  so that

$$n > N(\omega) \implies \sup_{0 \leq t \leq T} |I_t(\phi_{n_{k+1}}) - I_t(\phi_{n_k})| \leq \frac{1}{k^2}$$

If we set  $J_t^{(k)} = I_t(\phi_{n_k})$  then the  $\{J_t^{(k)}\}$  form a Cauchy sequence in the sup norm ( $|f|_{\text{sup}} = \sup_{t \in [0, T]} |f(t)|$ ). Since the convergence is uniform in  $t$  and each  $J_t^{(k)}$  is continuous, we know that for almost every  $\omega$  the limit point  $\lim_{k \rightarrow \infty} J_t^{(k)} = J_t$  is also continuous in  $t$ . Finally, since by assumption we also have  $J_t^{(k)} \rightarrow I_t(f) = \int_0^t f_s(\omega) dB_s(\omega)$  in  $L^2$ , we have that

$$\int_0^t f_s(\omega) dB_s(\omega) = J_t(\omega) \quad \text{a.s.},$$

as required. □

## 8. An Extension of the Itô Integral

Up until now we have only considered the Itô integral for integrands  $f$  such that  $\mathbb{E} \int_0^T f^2 ds < \infty$ . However it is possible to make sense of  $\int_0^T f_s(\omega) dB_s(\omega)$  if we only know that

$$\mathbb{P}\left[\int_0^T |f_s(\omega)|^2 ds < \infty\right] = 1.$$



Most of the previous properties hold. In particular  $\int_0^T f_s(\omega)dB_s(\omega)$  is a perfectly fine random variable which is almost surely finite. However, it is not necessarily true that  $\mathbb{E}[\int_0^T f_s(\omega)dB_s(\omega)] = 0$ . Which in turn means that  $\int_0^T f_s(\omega)dB_s(\omega)$  need not be a martingale. (In fact it is what is called a local martingale.) By obvious reasons, the Itô isometry property (4.12) may not hold in this case.

EXAMPLE 8.1.

## 9. Itô Processes

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be the canonical probability space and let  $\mathcal{F}_t$  be the  $\sigma$ -algebra generated by the Brownian motion  $B_t(\omega)$ .

DEFINITION 9.1.  $X_t(\omega)$  is an Itô process if there exist stochastic processes  $f(t, \omega)$  and  $\sigma(t, \omega)$  such that

- i)  $f(t, \omega)$  and  $\sigma(t, \omega)$  are  $\mathcal{F}_t$ -measurable,
- ii)  $\int_0^t |f| ds < \infty$  and  $\int_0^t |\sigma|^2 ds < \infty$  almost surely,
- iii)  $X_0(\omega)$  is  $\mathcal{F}_0$ -measurable,
- iv) With probability one the following holds

$$X_t(\omega) = X_0(\omega) + \int_0^t f_s(\omega) ds + \int_0^t \sigma_s(\omega)dB_s(\omega) \quad (4.13)$$

The processes  $f(t, \omega)$  and  $\sigma(t, \omega)$  are referred to as drift and diffusion coefficients of  $X_t$ .

For brevity, one often writes (4.13) as

$$dX_t(\omega) = f_t(\omega) dt + \sigma_t(\omega)dB_t(\omega)$$

But this is just notation for the integral equation above!



## CHAPTER 5

### Stochastic Calculus

This section introduces the fundamental tools for the computation of stochastic integrals. Indeed, similarly to what is done in classical calculus, stochastic integrals are rarely computed by applying the definition of Itô integral from the previous chapter. In the case of classical calculus, instead of applying the definition of Riemann integral one usually computes  $\int f(x)dx$  by applying the fundamental theorem of calculus and choosing  $\frac{d}{dx}F(x) = f(x)$  such that

$$\int f(x)dx = \int \frac{d}{dx}F(x)dx = F(x). \quad (5.1)$$

Even though, as we have seen in the previous section differentiation in this framework is not possible, it is possible to obtain a similar result for Itô integrals. In the following chapter we will introduce such a formula (called the Itô formula) allowing for rapid computation of stochastic integrals.

#### 1. Itô's Formula for Brownian motion

We first introduce the Itô formula for the Brownian motion process.

**THEOREM 1.1.** *Let  $f \in C^2(\mathbb{R})$  (the set of twice continuously differentiable functions on  $\mathbb{R}$ ) and  $B_t$  a standard Brownian motion. Then for any  $t > 0$ ,*

$$f(B_t) = f(0) + \int_0^t f'(B_s)dB_s + \frac{1}{2} \int_0^t f''(B_s) ds.$$

To prove this theorem, we first state the following partial result, proven at the end of the section.

**LEMMA 1.2.** *Let  $g$  be a continuous function and  $\Gamma^n := \{t_k^n : k = 1, \dots, N(n)\}$  be a sequence of partitions of  $[0, t]$  such that  $t_0 = 0$  and  $t_N = t$*

$$|\Gamma^n| := \sup_i |t_{i+1}^n - t_i^n| \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty. \quad (5.2)$$

Then

$$\sum_{k=0}^{N-1} g(\xi_k^n) \left( B_{t_k^n} - B_{t_{k+1}^n} \right)^2 \rightarrow \int_0^t g(B_s) ds,$$

for any choice of  $\xi_k^n \in (B_{t_k^n}, B_{t_{k+1}^n})$ .

**PROOF OF THEOREM 1.1.** Without loss of generality we can assume that  $f$  and its first two derivatives are bounded. After establishing the result for such functions we can approximate any function by such a sequence and pass to the limit to obtain the general result.

Let  $\{t_k^n : k = 1, \dots, N(n)\}$  be a sequence of partitions of  $[0, t]$  such that  $t_0 = 0$  and  $t_N = t$

$$|\Gamma^n| := \sup_i |t_{i+1}^n - t_i^n| \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

Now for any level  $n$ ,

$$f(B_t) - f(0) = \sum_{k=1}^{N(n)} \left( f(B_{t_k}) - f(B_{t_{k-1}}) \right). \quad (5.3)$$

Taylor's Theorem implies

$$f(B_{t_k}) - f(B_{t_{k-1}}) = f'(B_{t_{k-1}})(B_{t_k} - B_{t_{k-1}}) + \frac{1}{2}f''(\xi_k)(B_{t_k} - B_{t_{k-1}})^2,$$

for some  $\xi_k \in [B_{t_{k-1}}, B_{t_k}]$ . Returning to (5.3), we have

$$f(B_t) - f(0) = \sum_{k=1}^{N(n)} f'(B_{t_{k-1}})(B_{t_k} - B_{t_{k-1}}) + \frac{1}{2} \sum_{k=1}^{N(n)} f''(\xi_k)(B_{t_k} - B_{t_{k-1}})^2. \quad (5.4)$$

By the construction of the Itô integral, for the first term on the right hand side of (5.4) we have

$$\sum_{k=1}^{N(n)} f'(B_{t_{k-1}})(B_{t_k} - B_{t_{k-1}}) \longrightarrow \int_0^t f'(B_s) dB_s$$

in  $L^2$  as  $n \rightarrow \infty$ . Combining this with the result of Lemma 1.2 (proven below) with  $g = f''$  for the second term on the right hand side of (5.4) we conclude the proof.  $\square$

PROOF OF LEMMA 1.2. We now want to show that

$$A_n := \sum_{k=1}^{N(n)} g(\xi_k)(B_{t_k} - B_{t_{k-1}})^2 \longrightarrow \int_0^t g(B_s) ds$$

in probability as  $n \rightarrow \infty$ . We begin by showing that

$$C_n := \sum_{k=1}^{N(n)} g(B_{t_{k-1}})(B_{t_k} - B_{t_{k-1}})^2 \longrightarrow \int_0^t g(B_s) ds, \quad (5.5)$$

in probability as  $n \rightarrow \infty$ . First since  $g(B_s)$  is continuous, we have that

$$D_n := \sum_{k=1}^{N(n)} g(B_{t_{k-1}})(t_k - t_{k-1}) \longrightarrow \int_0^t g(B_s) ds.$$

as  $n \rightarrow \infty$ . Therefore, we obtain (5.5), by showing that the term  $|C_n - D_n|$  converges to 0 in  $L^2(\Omega)$  as this directly implies convergence in probability. To that end observe that

$$\begin{aligned} \mathbb{E}[(D_n - C_n)^2] &= \mathbb{E} \left[ \sum_{k=1}^{N(n)} g(B_{t_{k-1}})^2 (\Delta_k t - \Delta_k^2 B)^2 \right] \\ &\quad + 2\mathbb{E} \left[ \sum_{j < k} g(B_{t_{k-1}})g(B_{t_{j-1}}) (\Delta_k t - \Delta_k^2 B) (\Delta_j t - \Delta_j^2 B) \right] \end{aligned} \quad (5.6)$$

where  $\Delta_k := t_k - t_{k-1}$  and  $\Delta_k^2 B := (B_{t_k} - B_{t_{k-1}})^2$ . Now since

$$\mathbb{E}[(\Delta_k t - \Delta_k^2 B)^2] = (\Delta_k t)^2 - 2(\Delta_k t)^2 + 3(\Delta_k t)^2 = 2(\Delta_k t)^2$$

Considering the first term in (5.6) we have

$$\mathbb{E} \left[ \sum_{k=1}^{N(n)} g(B_{t_{k-1}})^2 (\Delta_k t - \Delta_k^2 B)^2 \right] = 2 \sum_{k=1}^{N(n)} \mathbb{E} [g(B_{t_{k-1}})^2 (\Delta_k t)^2] \leq 2|\Gamma^N| \sum_{k=1}^{N(n)} \mathbb{E} [g(B_{t_{k-1}})^2 (\Delta_k t)]$$

Since the sum converges to

$$\int_0^t \mathbb{E} [g(B_s)^2] ds < \infty$$

as  $n \rightarrow \infty$  and  $|\Gamma^N| \rightarrow 0$ , the product goes to zero. All that remains is the second sum in (5.6). Since

$$\begin{aligned} \mathbb{E} [g(B_{t_{k-1}})g(B_{t_{j-1}})(\Delta_k t - \Delta_k^2 B)(\Delta_j t - \Delta_j^2 B)] \\ = \mathbb{E} [g(B_{t_{k-1}})g(B_{t_{j-1}})(\Delta_j t - \Delta_j^2 B)\mathbb{E} [\Delta_k t - \Delta_k^2 B | \mathcal{F}_{t_{k-1}}]]] \end{aligned}$$

and  $\mathbb{E} [\Delta_k t - \Delta_k^2 B | \mathcal{F}_{t_{k-1}}] = 0$ , we see that the second sum is in fact zero.

All that remains is to show that  $A_n$  converges to  $C_n$ . Now

$$\begin{aligned} |C_n - A_n| &\leq \sum_{k=1}^{N(n)} |g(\xi_k) - g(B_{t_k})| (B_{t_k} - B_{t_{k-1}})^2 \\ &\leq \left( \sup_k |g(\xi_k) - g(B_{t_k})| \right) \sum_{k=1}^{N(n)} (B_{t_k} - B_{t_{k-1}})^2. \end{aligned}$$

Since the first term goes to zero in probability as  $n \rightarrow \infty$  by the continuity and boundedness of  $g(B_s)$  and the second term converges to the quadratic variation of  $B_t$  (which equals  $t$ ), we conclude that the product converges to zero in probability.  $\square$

**1.1. A second look at Itô's Formula.** Looking back at (5.4), one sees that the Itô integral term in Itô's Formula comes from the sum against the increments of Brownian motion. This term results directly from the first order Taylor expansion of  $f$ , and can be identified with the first order derivative term that we are used to see in the fundamental theorem of calculus (5.1). The second sum

$$\sum_{k=1}^{N(n)} f''(\xi_k) (B_{t_k} - B_{t_{k-1}})^2,$$

which contains the squares of the increments of Brownian motion, results from the second order term in the Taylor expansion and is *absent* in the classical calculus formulation. However, since the sum

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

converges in probability to the quadratic variation of the Brownian motion  $B_t$ , which according to Lemma 5.3 is simply  $t$ , this term gives a nonzero contribution in the limit  $n \rightarrow \infty$  and should be considered in this framework. We refer to this term as the *Itô correction term*. In light of this remark, if we let  $[B]_t$  denote the quadratic variation of  $B_t$ , then one can reinterpretate the Itô correction term

$$\frac{1}{2} \int_0^t f''(B_s) ds \quad \text{as} \quad \frac{1}{2} \int_0^t f''(B_s) d[B]_s. \quad (5.7)$$

We wish to derive a more general version of Itô's formula for a general Itô process  $X_t$  defined by

$$X_t = X_0 + \int_0^t f_s ds + \int_0^t g_s dB_s.$$

Beginning in the same way as before, we write the expression analogous to (5.4), namely

$$f(X_t) - f(X_0) = \sum_{k=1}^{N(n)} f'(X_{t_{k-1}})(X_{t_k} - X_{t_{k-1}}) + \frac{1}{2} \sum_{k=1}^{N(n)} f''(\xi_k)(X_{t_k} - X_{t_{k-1}})^2 \quad (5.8)$$

for some twice continuously differentiable function  $f$  and some partition  $\{t_k\}$  of  $[0, t]$ . It is reasonable to expect the first and second sums to converge respectively to the integrals

$$\int_0^t f'(X_s) dX_s \quad \text{and} \quad \frac{1}{2} \int_0^t f''(X_s) d[X]_s. \quad (5.9)$$

The first is the Itô stochastic integral with respect to an Itô process  $X_t$  while the second is an integral with respect to the differential of the quadratic variation  $[X]_t$  of the process  $X_t$ . So far this discussion has proceeded mainly by analogy to the simple Brownian motion. To make sense of the two terms in (5.9) we need to better understand the quadratic variation of an Itô process  $X_t$  and to define the concept of stochastic integral against  $X_t$ . While the former point is covered in the following section, the latter is quickly clarified by this intuitive definition:

**DEFINITION 1.3.** *Given an Itô process  $\{X_t\}$  with differential  $dX_t = f_t dt + \sigma_t dB_t$  and an adapted stochastic process  $\{h_t\}$  such that*

$$\int_0^\infty |h_s f_s| ds < \infty \quad \text{and} \quad \int_0^\infty (h_s \sigma_s)^2 ds < \infty \quad a.s.$$

*then we define the integral of  $h_t$  against  $X_t$  as*

$$\int_0^t h_s dX_s := \int_0^t h_s f_s ds + \int_0^t h_s \sigma_s dB_s. \quad (5.10)$$

## 2. Quadratic Variation and Covariation

We generalize the definition (3.7) of quadratic variation to the one of *quadratic covariation*

**DEFINITION 2.1.** *Let  $X_t, Y_t$  be two adapted, stochastic processes. Their quadratic covariation is defined as*

$$[X, Y]_t := \lim_{N \rightarrow \infty}^p \sum_{j=0}^{j^N} \left( X_{t_{j+1}^N} - X_{t_j^N} \right) \left( Y_{t_{j+1}^N} - Y_{t_j^N} \right)$$

*where  $\lim^p$  denotes a limit in probability and  $\{t_j^N\}$  is a set partitioning the interval  $[0, t]$  defined by*

$$\Gamma^N := \{ \{t_j^N\} : 0 = t_0^N < t_1^N < \dots < t_{j^N}^N = t \} \quad (5.11)$$

*with  $|\Gamma^N| := \sup_j |t_{j+1}^N - t_j^N| \rightarrow 0$  as  $N \rightarrow \infty$ . Furthermore, we define the quadratic variation of  $X_t$  as*

$$[X]_t := [X, X]_t. \quad (5.12)$$

We can also speak about the quadratic variation on an interval different than  $[0, t]$ . For  $0 \leq s < t$  we will write respectively  $[X]_{s,t}$  and  $[X, Y]_{s,t}$  for the quadratic variation and cross-quadratic variation on the interval  $[s, t]$ .

Just from the algebraic form of the pre-limiting object the quadratic variation satisfies a number of properties.

**LEMMA 2.2.** *Assuming all of the objects are defined, then for any adapted, continuous stochastic processes  $X_t, Y_t$ ,*

i) *for any constant  $c \in \mathbb{R}$  we have*

$$[cX]_t = c^2[X]_t$$

ii) *for  $0 < s < t$  we have*

$$[X]_{0,s} + [X]_{s,t} = [X]_{0,t} \quad (5.13)$$

iii) we have that

$$0 \leq [X]_{0,s} \leq [X]_{0,t} \quad (5.14)$$

for  $t > s \geq 0$ . In other words, the map  $t \mapsto [X]_t$  is nondecreasing a.s. .

iv) we can write

$$[X \pm Y]_t = [X]_t + [Y]_t \pm 2[X, Y]_t. \quad (5.15)$$

Consequently, quadratic covariations can be written in terms of quadratic variations as

$$[X, Y]_t = \frac{1}{2} \left( [X + Y]_t - [X]_t - [Y]_t \right) = \frac{1}{4} \left( [X + Y]_t - [X - Y]_t \right). \quad (5.16)$$

PROOF. Parts i) and ii) are a direct consequence of Def. 2.1, while part iii) results from ii) the fact that  $[X, Y]_t$  is defined as a sum of squares and is therefore nonnegative:  $[X]_{0,t} = [X]_{0,s} + [X]_{s,t} \geq [X]_{0,s}$ . Part iv) is obtained by noticing that

$$\begin{aligned} [X \pm Y]_t &= \lim_{N \rightarrow \infty} \sum_{j=0}^{j^N} \left( (X_{t_{j+1}^N} - X_{t_j^N}) \pm (Y_{t_{j+1}^N} - Y_{t_j^N}) \right)^2 \\ &= \lim_{N \rightarrow \infty} \sum_{j=0}^{j^N} \left( (X_{t_{j+1}^N} - X_{t_j^N})^2 \pm 2(X_{t_{j+1}^N} - X_{t_j^N})(Y_{t_{j+1}^N} - Y_{t_j^N}) + (Y_{t_{j+1}^N} - Y_{t_j^N})^2 \right) \\ &= [X]_t \pm 2[X, Y]_t + [Y]_t, \end{aligned}$$

while (5.16) is obtained by rearranging the terms of the above result (for the first inequality) and by using it to compute  $[X + Y]_t + [X - Y]_t$  (for the second).  $\square$

In the following sections we will see that the quadratic variation of Itô integrals acquires a particularly simple form. We will show this by first considering quadratic variations of Itô integrals and then extend this result to Itô processes.

### 2.1. Quadratic Variation of an Itô Integral.

LEMMA 2.3. Let  $\sigma_t$  be a process adapted to the foltration  $\{\mathcal{F}_t^B\}$  and such that  $\int_0^\infty \sigma_s^2 ds < \infty$  a.s.. Then defining  $M_t := I_t(\sigma) = \int_0^t \sigma_s dB_s$  we have that

$$[M]_t = \int_0^t \sigma_s^2 ds \quad (5.17)$$

or in differential notation  $d[M]_t = \sigma_t^2 dt$ .

PROOF OF LEMMA 2.3. It is enough to prove (5.17) when  $\sigma_s$  is an elementary stochastic process in  $\mathcal{S}_2$ . The general case can then be handled by approximation as in the proof of the Itô isometry. Hence, we assume that

$$\sigma_t = \sum_{j=1}^K \alpha_{j-1} \mathbf{1}_{[t_{j-1}, t_j)}(t) \quad (5.18)$$

where the  $\alpha_k$  satisfy the properties required by  $\mathcal{S}_2$  and  $K$  is some integer. Without loss of generality we can assume that  $t$  is the right endpoint of our interval so that the partition takes the form

$$0 = t_0 < t_1 < \dots < t_K = t$$

Now observe that if  $[s, r] \subset [t_{j-1}, t_j]$  then

$$\int_s^r \sigma_\tau dB_\tau = \alpha_{j-1} (B_s - B_r)$$

Hence  $\{s_\ell^{(n)}\}$  is a sequence of partitions of the interval  $[t_{j-1}, t_j]$  so that

$$t_{j-1} = s_0^{(n)} < s_1^{(n)} < \dots < s_{N(n)}^{(n)} = t_j$$

and  $|\Gamma^N| = \sup_\ell |s_{\ell-1}^{(n)} - s_\ell^{(n)}| \rightarrow 0$  as  $n \rightarrow \infty$ . Then the quadratic variation of  $M_t$  on the interval  $[t_{j-1}, t_j]$  is the limit  $n \rightarrow \infty$  of

$$\sum_{\ell=1}^{N(n)} (M_{s_{\ell-1}} - M_{s_\ell})^2 = \alpha_{j-1}^2 \sum_{\ell=1}^{N(n)} (B_{s_{\ell-1}} - B_{s_\ell})^2.$$

Since the summation on the right hand side limits to the quadratic variation of the Brownian motion  $B$  on the interval  $[t_{j-1}, t_j]$  which we know to be  $t_j - t_{j-1}$  we conclude that

$$[M]_{t_{j-1}, t_j} = \alpha_{j-1}^2 (t_j - t_{j-1}).$$

Since the quadratic variation on disjoint intervals adds, we have that

$$[M]_t = \sum_{j=1}^K [M]_{t_{j-1}, t_j} = \sum_{j=1}^K \alpha_{j-1}^2 (t_j - t_{j-1}) = \int_0^t \sigma_s^2 ds$$

where the last equality follows from the fact that  $\sigma_s$  takes the form (5.18). As mentioned at the start, the general form follows from this calculation by approximation by functions in  $\mathcal{S}_2$ .  $\square$

REMARK 2.4. *The proof of the above result in Klebaner (Theorem 4.14 on pp. 106) has a subtle issue. When bounding*

$$2 \sum_{i=0}^{n-1} g^2(B_{t_i^n})(t_{i+1}^n - t_i^n)^2 \leq 2\delta_n \mathbb{E} \left[ \sum_{i=0}^{n-1} g^2(B_{t_i^n})(t_{i+1}^n - t_i^n) \right],$$

Klebaner asserts that as  $n \rightarrow \infty$ ,  $\delta_n = |\Gamma_n| \rightarrow 0$ ,  $\sum_{i=0}^{n-1} \mathbb{E}[g^2(B_{t_i^n})(t_{i+1}^n - t_i^n)]$  would stay finite, and thus their product would go to 0. However, the finiteness of  $\sum_{i=0}^{n-1} \mathbb{E}[g^2(B_{t_i^n})(t_{i+1}^n - t_i^n)]$  is unjustified. In fact, if it were finite, it must converge to  $\int_0^t \mathbb{E}[g^2(B_s)] ds$  (Riemann sum). However, this integral might be infinity for certain choice of  $g$ , for example,  $g(x) = e^{x^2}$  (see Example 4.5 on pp. 99 of [Klebaner]). The proof here uses the same computation of second moment but only for “nice” functions (i.e., those with compact support). The convergence in probability (note: this is weaker than convergence in  $L^2$ ) for general continuous functions is established using approximation. The stopping rules for Itô integral are needed here, but we defer it to the later part of the course.

We now consider the quadratic covariation of two Itô integrals with respect to independent Brownian motions.

LEMMA 2.5. *Let  $B_t, W_t$  two independent Brownian motions, and  $f_s, g_s$  two stochastic processes, all adapted to the underlying filtration  $\mathcal{F}_t$  and such that  $\int_0^\infty f_s^2 ds, \int_0^\infty g_s^2 ds < \infty$  almost surely. We define*

$$M_t := \int_0^t f_s dB_s \quad \text{and} \quad N_t := \int_0^t g_s dW_s.$$

Then, for all  $t \geq 0$  one has

$$[N, M]_t = 0. \tag{5.19}$$

PROOF OF (5.19). Again without lost of generality it is enough to prove the result for  $\sigma_t$  and  $g_s$  in  $\mathcal{S}_2$ . We can further assume that both functions are defined with respect to the same partition

$$0 = t_0 < t_1 < \dots < t_K = t$$



Since as observed in (5.13), the quadratic variation on disjoint intervals adds, we need only show that  $[N, M]_{t_{j-1}, t_j} = 0$  on any of the partition intervals  $[t_{j-1}, t_j]$ .

Fixing such an interval  $[t_{j-1}, t_j]$ , we see that  $[N, M]_{t_{j-1}, t_j} = \sigma_{t_{j-1}} g_{t_{j-1}} [W, B]_{t_{j-1}, t_j}$ . The easiest way to see this is to use the ‘‘polarization’’ equality (5.16)

$$2[W, B]_{t_{j-1}, t_j} = \left[ \frac{W+B}{\sqrt{2}}, \frac{W+B}{\sqrt{2}} \right]_{t_{j-1}, t_j} - \left[ \frac{W-B}{\sqrt{2}}, \frac{W-B}{\sqrt{2}} \right]_{t_{j-1}, t_j} = (t_j - t_{j-1}) - (t_j - t_{j-1}) = 0$$

since  $\frac{W+B}{\sqrt{2}}$  and  $\frac{W-B}{\sqrt{2}}$  are standard Brownian motions and hence have quadratic variation the length of the time interval.  $\square$

REMARK 2.6. *One can also prove the above result more directly by following the same argument as in Theorem 5.3 of Chapter 3. The key calculation is to show that the expected value of the approximating quadratic variation is 0 and not the length of the time interval as in the proof of Theorem (5.3) of Chapter 3. For any partitions  $\{s_\ell\}$  of  $[t_{j-1}, t_j]$  one has to zero we have*

$$\mathbb{E} \sum_{\ell} (B_{s_\ell} - B_{s_{\ell-1}})(B_{s_\ell} - B_{s_{\ell-1}}) = \sum_{\ell} \mathbb{E}(B_{s_\ell} - B_{s_{\ell-1}}) \mathbb{E}(B_{s_\ell} - B_{s_{\ell-1}}) = 0$$

**2.2. Quadratic Variation of an Itô Process.** In this section, using the results presented above, we finally obtain a simple expression for the quadratic variation of an Itô process.

LEMMA 2.7. *If  $X_t$  is an Itô process with differential  $dX_t = \mu_t dt + \sigma_t dB_t$ , then*

$$[X]_t = [I(\sigma^2)]_t = \int_0^t \sigma_s^2 ds, \quad (5.20)$$

or equivalently  $d[X]_t = \sigma_t^2 dt$ .

By comparing this result with (5.17) we notice that the only contribution to the quadratic variation process comes from the Itô integral. The following result will be useful in the proof of (5.20).

LEMMA 2.8. *Let  $X_t$  and  $Y_t$  be adapted stochastic processes, such that  $X_t$  is continuous a.s. and  $Y_t$  has trajectories with finite first variation ( $V_1[Y](t) < \infty$ ) then  $[X, Y]_t = 0$  a.s..*

Before we give the proof of Lemma 2.8 we observe that it immediately yields (5.20).

PROOF OF LEMMA 2.7. Defining  $F_t = \int_0^t \mu_s ds$  and  $M_t = \int_0^t \sigma_s dB_s$ , observe that  $X_t = F_t + M_t$  and that  $F_t$  is continuous and of finite first variation almost surely. Hence  $[F]_t = 0$ . Since  $M_t$  is continuous a.s., we have that  $[M, F]_t = 0$  almost surely. Hence

$$[X]_t = [F]_t + 2[F, M]_t + [M]_t = [M]_t = \int_0^t \sigma_s^2 ds.$$

$\square$

PROOF OF LEMMA 2.8. Let  $\Gamma^N := \{t_i^N : i = 0, \dots, i_N\}$  be a sequence of partitions of  $[0, t]$  such that  $|\Gamma^N| = \sup_i |t_{i+1}^N - t_i^N| \rightarrow 0$  as  $N \rightarrow \infty$ . Now

$$\left| \sum_{i=1}^{i_N} (X_{t_i} - X_{t_{i-1}})(Y_{t_i} - Y_{t_{i-1}}) \right| \leq \left( \sup_i |X_{t_i} - X_{t_{i-1}}| \right) \sum_{i=1}^{i_N} |Y_{t_i} - Y_{t_{i-1}}|$$

The summation on the right hand side is bounded from above by the first variation of  $Y_t$  which by assumption is finite a.s. On the other hand, as  $n \rightarrow \infty$  the supremum goes to zero since  $|\Gamma^N| \rightarrow 0$  and  $X_t$  is a.s. continuous.  $\square$

REMARK 2.9. Similarly to the formal considerations in Section 1.1, we may think of the differential of the quadratic variation process  $d[X]_t$  as the limit of the difference term  $(X_{t_{k+1}^n} - X_{t_k^n})^2$ , which in turn is the square of the differential of brownian motion  $(dX_t)^2$ . Therefore, formally speaking, we can obtain the result of the previous lemma by writing

$$\begin{aligned} d[X]_t &= (dX_t)^2 = (\mu_t dt + \sigma_t dB_t)^2 \\ &= \mu_t^2 (dt)^2 + 2\mu_t \sigma_t (dt)(dB_t) + \sigma_t^2 (dB_t)^2 \\ &= \sigma_t^2 dt. \end{aligned}$$

where we have applied  $(dt)^2 = (dt)(dB_t) = 0$  (cfr. Lemma 2.8) and  $(dB_t)^2 = dt$  (cfr. Lemma 2.3). These formal multiplication rules are summarized in the following table: By the same formal

$\times$	$dt$	$dB_t$
$dt$	$0$	$0$
$dB_t$	$0$	$dt$

arguments, such rules apply to the computation of the quadratic covariation of two Itô processes  $Y_t = \mu'_t dt + \sigma'_t dB_t$ :

$$\begin{aligned} d[X, Y]_t &= (dX_t)(dY_t) = (\mu_t dt + \sigma_t dB_t)(\mu'_t dt + \sigma'_t dB_t) \\ &= \mu_t \mu'_t (dt)^2 + \mu_t \sigma'_t (dt)(dB_t) + \mu'_t \sigma_t (dt)(dB_t) + \sigma_t \sigma'_t (dB_t)^2 \\ &= \sigma_t \sigma'_t dt. \end{aligned} \tag{5.21}$$

This result can be verified by going through the steps of the proof of the above lemmas.

### 3. Itô's Formula for an Itô Process

Using the lessons learned in the previous sections, we now proceed to compute the infinitesimal of (5.8) as

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2. \tag{5.22}$$

Of course this is just notation for the integral equation

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)dX_s + \frac{1}{2} \int_0^t f''(X_s)(dX_s)^2.$$

The first integral is simply the integral against an Itô process as we have already discussed. In light of Remark 2.9, we should interpret  $d[X]_t = (dX_t)^2 = (\mu_t dt + \sigma_t dB_t)^2 = \sigma_t^2 dt$ . Hence

$$\frac{1}{2} \int_0^t f''(X_s)(dX_s)^2 = \frac{1}{2} \int_0^t f''(X_s)\sigma_s^2 ds$$

This formal calculation (which is correct) leads us to suggest the following general Itô formula.

**THEOREM 3.1.** *Let  $X_t$  be the Itô process given by*

$$dX_t = \mu_t dt + \sigma_t dB_t$$

*If  $f$  is a  $C^2$  function then*

$$\begin{aligned} f(X_t) &= f(X_0) + \int_0^t f'(X_s)dX_s + \frac{1}{2} \int_0^t f''(X_s)d[X]_s \\ &= f(X_0) + \int_0^t f'(X_s)\mu_s ds + \int_0^t f'(X_s)\sigma_s dB_s + \frac{1}{2} \int_0^t f''(X_s)\sigma_s^2 ds \end{aligned}$$

PROOF OF THEOREM 3.1. Without loss of generality, we can assume that both  $\mu_t$  and  $\sigma_t$  are adapted elementary stochastic processes satisfying the assumptions required by an Itô process. Furthermore, by inserting partition points if needed, we can assume that they are both defined on the same partition

$$0 = t_0 < t_1 < \dots < t_N = t.$$

Since

$$f(X_t) - f(X_0) = \sum_{\ell=1}^N f(X_{t_\ell}) - f(X_{t_{\ell-1}})$$

we need only prove Itô's formula for  $f(X_{t_\ell}) - f(X_{t_{\ell-1}})$ . Now since  $X_t$  is constant for  $t \in [t_{\ell-1}, t_\ell]$  we can take  $\xi = X_{t_{\ell-1}}$  and for  $[r, s] \subset [t_{\ell-1}, t_\ell]$  we have

$$X_s - X_r = \int_r^s \mu_\tau d\tau + \int_r^s \sigma_\tau dB_\tau = \mu_{t_{\ell-1}}(s - r) + \sigma_{t_{\ell-1}}(B_s - B_r).$$

Let  $\{s_i^{(n)} : i = 0, \dots, K(n)\}$  be a sequence of partitions of  $[t_{\ell-1}, t_\ell]$  such that  $|\Gamma^N| = \sup_i |s_i^{(n)} - s_{i-1}^{(n)}| \rightarrow 0$  as  $n \rightarrow \infty$ . Now using Taylor's theorem we have

$$f(X_{s_j}) - f(X_{s_{j-1}}) = f'(X_{s_{j-1}})(X_{s_j} - X_{s_{j-1}}) + \frac{1}{2}f''(\xi_j)(X_{s_j} - X_{s_{j-1}})^2.$$

for some  $\xi_j \in (X_{s_{j-1}}, X_{s_j})$ . Hence we have

$$\begin{aligned} f(X_{t_\ell}) - f(X_{t_{\ell-1}}) &= \sum_{j=1}^{K(n)} f(X_{s_j}) - f(X_{s_{j-1}}) \\ &= \sum_{j=1}^{K(n)} f'(X_{s_{j-1}})(X_{s_j} - X_{s_{j-1}}) + \frac{1}{2} \sum_{j=1}^{K(n)} f''(\xi_j)(X_{s_j} - X_{s_{j-1}})^2 = (I) + \frac{1}{2}(II) \end{aligned}$$

Since

$$(X_{s_j} - X_{s_{j-1}})^2 = \mu_{t_{\ell-1}}^2 (s_j - s_{j-1})^2 + 2\mu_{t_{\ell-1}}\sigma_{t_{\ell-1}}(s_j - s_{j-1})(B_{s_j} - B_{s_{j-1}}) + \sigma_{t_{\ell-1}}^2 (B_{s_j} - B_{s_{j-1}})^2$$

we have

$$(I) = \mu_{t_{\ell-1}} \sum_{j=1}^{K(n)} f'(X_{s_{j-1}})(s - r) + \sigma_{t_{\ell-1}} \sum_{j=1}^{K(n)} f'(X_{s_{j-1}})(B_s - B_r) = (Ia) + (Ib)$$

and

$$\begin{aligned} (II) &= \mu_{t_{\ell-1}}^2 \sum_{j=1}^{K(n)} f''(\xi_j)(s_j - s_{j-1})^2 + 2\mu_{t_{\ell-1}}\sigma_{t_{\ell-1}} \sum_{j=1}^{K(n)} f''(\xi_j)(s_j - s_{j-1})(B_{s_j} - B_{s_{j-1}}) \\ &\quad + \sigma_{t_{\ell-1}}^2 \sum_{j=1}^{K(n)} f''(\xi_j)(B_{s_j} - B_{s_{j-1}})^2 = (IIa) + (IIb) + (IIc) \end{aligned}$$

As  $n \rightarrow \infty$ , it is clear that

$$(Ia) \longrightarrow \int_{t_{\ell-1}}^{t_\ell} f'(X_s) b_s ds, \quad \text{and} \quad (Ib) \longrightarrow \int_{t_{\ell-1}}^{t_\ell} f'(X_s) \sigma_s dB_s.$$

Using the same arguments as in Theorem 1.1 we see that

$$(IIc) \longrightarrow \sigma_{t_{\ell-1}}^2 \int_{t_{\ell-1}}^{t_\ell} f''(X_s) ds = \int_{t_{\ell-1}}^{t_\ell} \sigma_s^2 f''(X_s) ds.$$

All that remains is to show that (IIa) and (IIb) converge to zero as  $n \rightarrow \infty$ . Observe that

$$|(IIa)| \leq \mu_{t_{\ell-1}}^2 |\Gamma^N| \sum_{j=1}^{K(n)} f''(\xi_j)(s_j - s_{j-1}) \quad \text{and} \quad |(IIb)| \leq 2\mu_{t_{\ell-1}} \sigma_{t_{\ell-1}} |\Gamma^N| \sum_{j=1}^{K(n)} f''(\xi_j)(B_{s_j} - B_{s_{j-1}}).$$

Since the two sums converge to  $\int f''(X_s) ds$  and  $\int f''(X_s) dB_s$  respectively the fact that  $|\Gamma^N| \rightarrow 0$  implies that (IIa) and (IIb) converge to zero as  $n \rightarrow \infty$ . Putting all of these results together produces the quoted result.  $\square$

REMARK 3.2. Notice that the “multiplication table” given in Remark 2.9 is reflected in the details of the proof of Theorem 3.1. Each of the terms in  $(dX_t)^2$  correspond to one of the terms labeled (II) which came from  $(X(s_j) - X(s_{j-1}))^2$  in the Taylor’s theorem expansion. The term (IIa) which corresponds to  $(dt)^2$  limits to zero as the multiplication table indicates. The term (IIb) which corresponds to  $(dt)(dB_t)$  also tends to zero again as the table indicates. Lastly, (IIc) which corresponds to  $(dB_t)^2$  limits to an integral against  $(dt)$  as indicated in the table.

EXAMPLE 3.3. Consider the stochastic process with differential

$$dX_t = \frac{1}{2}X_t dB_t + X_t dB_t.$$

The above process is an example of a geometric Brownian motion, a process widely used in finance to model the price of a stock. We apply Itô formula to the function  $f(x) := \log x$ . Using that  $\partial_x f(x) = x^{-1}$  and  $\partial_{xx}^2 f(x) = -x^{-2}$  we obtain

$$d \log X_t = \frac{1}{X_t} dX_t - \frac{1}{2} \frac{1}{X_t^2} d[X]_t = \frac{1}{X_t} \left( \frac{1}{2} X_t dB_t + X_t dB_t \right) - \frac{1}{2} \frac{1}{X_t^2} (X_t^2 dt) = dB_t.$$

In the integral form the above can be written as  $\log X_t = \log X_0 + B_t$  and therefore  $X_t = X_0 e^{B_t}$ .

#### 4. Full Multidimensional Version of Itô Formula

We now give the full multidimensional version of Itô’s formula. We will include the possibility that the function depends on time and that there is more than one Brownian motion. We begin with the definition of a multidimensional Itô process.

DEFINITION 4.1. A stochastic process  $X_t = (X_1(t), \dots, X_d(t)) \in \mathbb{R}^d$  is an Itô process if each of the coordinate process  $X_i(t)$  is a one-dimensional Itô process.

Let  $\mu_i(t)$  and  $\sigma_{ij}(t)$  be adapted stochastic processes and  $\{B_j(t)\}_{j=1}^m$  be  $m$  mutually independent, standard Brownian motions such that for  $i = 1, \dots, d$  we can write each component aof the  $d$ -dimensional Itô process as

$$dX_i(t) = \mu_i(t) dt + \sum_{j=1}^m \sigma_{ij}(t) dB_j(t). \quad (5.23)$$

If we collect the Brownian motions into one  $m$ -dimensional brownian motion  $B_t = (B_1(t), \dots, B_m(t))$  and define the  $\mathbb{R}^d$ -valued process  $\mu_t = (\mu_1(t), \dots, \mu_d(t))$  and the matrix valued process  $\sigma_t$  whose matrix elements are the  $\sigma_{ij}(t)$  then we can write

$$dX_t = \mu_t dt + \sigma_t dB_t. \quad (5.24)$$

While this is nice and compact, it is perhaps more suggestive to define the  $\mathbb{R}^d$ -valued processes  $\sigma^{(j)} = (\sigma_{1,j}, \dots, \sigma_{n,j})$  for  $j = 1, \dots, m$  and write

$$dX_t = \mu_t dt + \sum_{j=1}^d \sigma_t^{(j)} dB_j(t). \quad (5.25)$$

This emphasizes that the process  $X_t$  at each moment of time is pushed in the direction which  $\mu_t$  points and given  $m$  random kicks, in the directions the  $\sigma_t^{(j)}$  point, whose magnitude and sign are dictated by the Brownian motions  $B_j(t)$ .

We now want to derive the process which describes evolution of  $F(X_t)$  where  $F : \mathbb{R}^d \rightarrow \mathbb{R}$ . In other words, the multidimensional Itô formula.

We begin by developing some intuition. Recall Lemma 2.5 stating that the cross-quadratic variation of independent Brownian motions is zero. Hence if  $B_t$  and  $W_t$  are independent standard Brownian motions then the multiplication table for  $(dX_t)^2$  and  $(dY_t)(dX_t)$  if  $X_t$  and  $Y_t$  are two Itô processes is given in the following table.

$\times$	$dt$	$dB_t$	$dW_t$
$dt$	0	0	0
$dB_t$	0	$dt$	0
$dW_t$	0	0	$dt$

TABLE 1. Formal multiplication rules for differentials of two independent Brownian motions

**THEOREM 4.2.** *Let  $F : \mathbb{R}^d \rightarrow \mathbb{R}$  be a function such that  $F(x) \in C^2$  in  $x \in \mathbb{R}^d$ . If  $X_t$  is as above then*

$$dF(X_t) = \sum_{i=1}^n \frac{\partial F}{\partial x_i}(X_t) dX_i(t) + \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^n \frac{\partial^2 F}{\partial x_i \partial x_k}(X_t) d[X_i, X_k]_t \quad (5.26)$$

Furthermore one has

$$\sum_{i=1}^d \sum_{k=1}^n \frac{\partial^2 F}{\partial x_i \partial x_k}(X_t) d[X_i, X_k](t) = \sum_{i=1}^n \sum_{k=1}^n \frac{\partial^2 F}{\partial x_i \partial x_k}(X_t) a_{ik}(t) dt \quad (5.27)$$

where

$$a_{ik}(t) = \sum_{j=1}^d \sigma_{ij}(t) \sigma_{kj}(t)$$

The matrix  $a(t)$  can be written compactly as  $\sigma(t)\sigma(t)^T$ . The matrix  $a$  is often called the diffusion matrix.

We will only sketch the proof of this version of Itô formula since it follows the same logic of the others already proven. Proofs can be found in many places including [13, 7, 3].

**SKETCH OF PROOF.** Similarly to the proof of Theorem 3.1 we introduce the family of partitions  $\Gamma_N$  of the interval  $[0, t]$  as in (8.7) with  $\lim_{n \rightarrow \infty} |\Gamma_N| = 0$  and expand in Taylor the function  $f$  in each of these intervals:

$$F(X_t) - F(X_0) = \sum_{\ell=1}^N \left\{ \sum_{i=1}^d \frac{\partial}{\partial x_i} F(X_{s_{\ell-1}}) (X_i(s_\ell) - X_i(s_{j-1})) \right.$$

$$\begin{aligned}
& + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} F(\xi_\ell) (X_i(s_\ell) - X_i(s_{\ell-1})) (X_j(s_\ell) - X_j(s_{\ell-1})) \Big\} \\
& = \sum_{\ell=1}^N \left\{ (I)_\ell + \frac{1}{2} (II)_\ell \right\},
\end{aligned}$$

for  $\xi_\ell \in \times_{i=1}^d [X_i(s_\ell), X_i(s_{\ell+1})]$ . For the first order term, it is straightforward to generalize the proof of Theorem 3.1 to obtain that

$$\lim_{N \rightarrow \infty} \sum_{\ell=1}^N (I)_\ell = \int_0^t \sum_{i=1}^d \frac{\partial}{\partial x_i} F(X_s) dX_i(s).$$

We formally recover the expression of the second order term by combining (5.21) with the rules of Table 1:

$$\begin{aligned}
\lim_{N \rightarrow \infty} \sum_{\ell=1}^N (II)_\ell & = \int_0^t \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} F(X_s) (dX_i(s))(dX_j(s)) \\
& = \int_0^t \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} F(X_s) \sum_{k,l=1}^m (\sigma_{ik}(t) dB_k(s)) (\sigma_{jl}(t) dB_l(s)) \\
& = \int_0^t \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} F(X_s) \sum_{k=1}^m \sigma_{ik}(t) \sigma_{jk}(t) dt,
\end{aligned}$$

where in the second equality we have used that  $(dt)(dB_j(t)) = 0$  and in the third that  $(dB_i(t))(dB_j(t)) = 0$  for  $i \neq j$ . Note that one should check that when taking the limit in the first equality  $F(\xi_\ell)$  can be replaced by  $F(X_{s_\ell})$ . This can be done by reproducing the proof of Lemma 1.2.  $\square$

**REMARK 4.3.** *The fact that (5.27) holds requires that  $B_i$  and  $B_j$  are independent if  $i \neq j$ . However, (5.26) holds even when they are not independent.*

Theorem 4.2 is for a scalar valued function  $F$ . However by applying the result to each coordinate function  $F_i : \mathbb{R}^d \rightarrow \mathbb{R}$  of the function  $F : \mathbb{R}^d \rightarrow \mathbb{R}^p$  with  $F = (F_1, \dots, F_p)$  we obtain the full multidimensional version. Instead of writing it again in coordinates, we take the opportunity to write a version aligned with the perspective and notation of (5.26). Recalling that the directional derivative of  $F$  in the direction  $\nu \in \mathbb{R}^d$  at the point  $x \in \mathbb{R}^d$  is

$$DF(x)[\nu] = \lim_{\varepsilon \rightarrow 0} \frac{F(x + \varepsilon\nu) - F(x)}{\varepsilon} = (\nabla F \cdot \nu)(x) = \sum_{k=1}^p \sum_{i=1}^d \frac{\partial F_k}{\partial x_i} \nu_i e_k$$

where  $e_k$  is the  $k$ -th unit vector of  $\mathbb{R}^p$ . Similarly, the second directional derivative at the point  $x \in \mathbb{R}^d$  in the directions  $\nu, \eta \in \mathbb{R}^d$  is given by

$$D^2 F(x)[\nu, \eta] = \lim_{\varepsilon \rightarrow 0} \frac{DF(x + \varepsilon\eta)[\nu] - DF(x)[\nu]}{\varepsilon} = \sum_{k=1}^p \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 F_k}{\partial x_i \partial x_j} \nu_i \eta_j e_k$$

Then in the notation of (5.26), Theorem 4.2 can be rewritten as

$$\begin{aligned}
dF(X(t)) & = DF(X(t))[f(t)] dt + \sum_{i=1}^d DF(X(t))[\sigma^{(i)}(t)] dB_i(t) \\
& \quad + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d D^2 F(X(t))[\sigma^{(i)}(t), \sigma^{(j)}(t)] d[B_i, B_j](t)
\end{aligned}$$

If  $B_i$  and  $B_j$  are assumed to be independent if  $i \neq j$ , then

$$dF(X(t)) = DF(X(t))[f(t)] dt + \sum_{i=1}^d DF(X(t))[\sigma^{(i)}(t)] dB_i(t) + \frac{1}{2} \sum_{i=1}^d D^2F(X(t))[\sigma^{(i)}(t), \sigma^{(i)}(t)] dt.$$

We now consider special cases of Theorem 3.1 that will be helpful in practice. The first describes evolution of a function  $F(x, t)$  that depends explicitly on time:

**COROLLARY 4.4.** *Let  $F : \mathbb{R}^d \times [0, \infty) \rightarrow \mathbb{R}$  such that  $F(x, t)$  is  $C^2$  in  $x \in \mathbb{R}^d$  and  $C^1$  in  $t \in [0, \infty)$ . Furthermore, let  $X_t$  be a  $d$ -dimensional Itô process as in (5.23). Then*

$$dF(X_t, t) = \frac{\partial F}{\partial t}(X_t, t) dt + \sum_{i=1}^n \frac{\partial F}{\partial x_i}(X_t, t) dX_i(t) + \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^n \frac{\partial^2 F}{\partial x_i \partial x_k}(X_t, t) d[X_i, X_k]_t. \quad (5.28)$$

where  $a$  is defined as in (5.27).

**PROOF.** In this proof we will make use of the “multiplication table” at the beginning of this section. Consider the  $d + 1$ -dimensional process  $\bar{X}_t := (X_1(t), \dots, X_d(t), Y_t)$  for  $Y_t$  given by  $dY_t = dt + 0 dB_t$ . Then by applying Itô’s Formula Theorem 5.26 and the fact that  $Y_t$  is of finite variation (so  $d[Z, Y]_t = 0$  for continuous  $Z_t$ ) we obtain

$$\begin{aligned} dF(X_t) &= \sum_{i=1}^d \partial_i F(X_t, t) dX_i(t) + \partial_t F(X_t, t) (dt + 0 dB_t) \\ &\quad + \frac{1}{2} \left( \sum_{i,j=1}^d \partial_{ij}^2 F(X_t, t) d[X_i, X_j]_t + 2 \sum_{i=1}^d \partial_i \partial_t F(X_t, t) d[X_i, Y_t] + \partial_t^2 F(X_t, t) d[Y_t, Y_t] \right) \\ &= \sum_{i=1}^d \partial_i F(X_t, t) dX_i(t) + \partial_t F(X_t, t) dt + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 F(X_t, t) d[X_i, X_j]_t. \end{aligned}$$

Note that the existence of the second derivative in  $t$  is not needed in the above formula and can therefore be dropped.  $\square$

**COROLLARY 4.5.** *Let  $X_t, Y_t$  be two Itô processes. Then*

$$d(X_t Y_t) = Y_t dX_t + X_t dY_t + d[X, Y]_t. \quad (5.29)$$

*This result is known as stochastic integration by parts formula.*

**PROOF.** Let  $F : \mathbb{R}^2 \rightarrow \mathbb{R}$  with  $F(x, y) = x \cdot y$ , then since

$$\partial_x F(x, y) = y, \quad \partial_y F(x, y) = x, \quad \partial_{xx}^2 F(x, y) = \partial_{yy}^2 F(x, y) = 0, \quad \partial_{xy}^2 = 1,$$

by Itô’s Formula Theorem 5.26 we have

$$d(X_t Y_t) = dF(X_t, Y_t) = Y_t dX_t + X_t dY_t + 1 d[X, Y]_t. \quad (5.30)$$

$\square$

**EXAMPLE 4.6.** *We compute the stochastic integral*

$$\int_0^t s dB_s.$$

*Applying the integration by parts formula (5.30) for  $dX_t = dB_t$ ,  $dY_t = dt$  we obtain*

$$d(tB_t) = t dB_t + B_t dt + d[B, t]_t.$$

Since  $Y_t$  is of finite variation we have  $d[B, Y]_t = 0$ , and by integrating and rearranging the terms we obtain

$$\int_0^t s dB_s = \int_0^t d(sB_s) - \int_0^t B_s ds = tB_t - \int_0^t B_s ds.$$

EXAMPLE 4.7. Assume that  $f(x, t) \in C^{2,1}(\mathbb{R}, \mathbb{R}_+)$  satisfies the PDE

$$\frac{\partial}{\partial t} f(x, t) + \frac{\partial^2}{\partial x^2} f(x, t) = 0,$$

and  $\mathbb{E}[f(B_t, t)^2] < \infty$ , then have that

$$\begin{aligned} df(B_t, t) &= \partial_t f(B_t, t) dt + \partial_x f(B_t, t) dB_t + \frac{1}{2} \partial_{xx}^2 f(B_t, t) d[B]_t \\ &= \left( \partial_t + \frac{1}{2} \partial_{xx}^2 \right) f(B_t, t) dt + \partial_x f(B_t, t) dB_t = \partial_x f(B_t, t) dB_t. \end{aligned}$$

Therefore  $f(B_t, t) = f(0, 0) + \int_0^t \partial_x f(B_s, s) dB_s$  is a martingale.

## 5. Collection of the Formal Rules for Itô's Formula and Quadratic Variation

We now recall some of the formal calculations, bringing them all together in one place. We consider a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with a filtration  $\mathcal{F}_t$ . We assume that  $B_t$  and  $B_t$  are independent standard Brownian motions adapted to the filtration  $\mathcal{F}_t$ .

For any  $\rho \in [0, 1]$ ,  $Z_t = \rho B_t + \sqrt{1 - \rho^2} B_t$  is again a standard Brownian motion. Furthermore

$$\begin{aligned} [Z]_t &= [Z, Z]_t = \rho^2 [B, B]_t + 2\rho\sqrt{1 - \rho^2} [B, B]_t + (1 - \rho^2) [B, B]_t \\ &= \rho^2 t + 0 + (1 - \rho^2)t = t \end{aligned}$$

Or in the formal differential notation,  $d[Z]_t = dt$ . This result can be understood by using the formal multiplication table for differentials which formally states:

$$\begin{aligned} d[Z]_t &= (dZ_t)^2 = (\rho dB_t + \sqrt{1 - \rho^2} dB_t)^2 = \rho^2 (dB_t)^2 + 2\rho\sqrt{1 - \rho^2} dB_t dB_t + (1 - \rho^2) (dB_t)^2 \\ &= \rho^2 dt + 0 + (1 - \rho^2) dt = dt \end{aligned}$$

Similarly, one has

$$\begin{aligned} d[Z, B]_t &= (dZ_t)(dB_t) = \rho(dB_t)^2 + \sqrt{1 - \rho^2} (dB_t)(dB_t) = \rho dt + 0 = \rho dt \\ d[Z, B]_t &= (dZ_t)(dB_t) = \rho(dB_t)(dB_t) + \sqrt{1 - \rho^2} (dB_t)^2 = 0 + \sqrt{1 - \rho^2} dt = \sqrt{1 - \rho^2} dt \end{aligned}$$

Now let  $\sigma_t$  and  $g_t$  be adapted stochastic processes (adapted to  $\mathcal{F}_t$ ) with

$$\int_0^t \sigma_s^2 ds < \infty \quad \text{and} \quad \int_0^t g_s^2 ds < \infty$$

almost surely. Now define

$$\begin{aligned} dM_t &= \sigma_t dB_t & dN_t &= g_t dB_t \\ dU_t &= \sigma_t dB_t & dV_t &= \sigma_t dZ_t \end{aligned}$$

Of course these are just formal expression. For example,  $dM_t = \sigma_t dB_t$  means  $M_t = M_0 + \int_0^t \sigma_s dB_s$ . Using the multiplication table from before we have

$$\begin{aligned} d[M]_t &= (dM_t)^2 = \sigma_t^2 (dB_t)^2 = \sigma_t^2 dt & d[U]_t &= (dU_t)^2 = \sigma_t^2 (dB_t)^2 = \sigma_t^2 dt \\ d[N]_t &= (dN_t)^2 = g_t^2 (dB_t)^2 = g_t^2 dt & d[V]_t &= (dV_t)^2 = \sigma_t^2 (dZ_t)^2 = \sigma_t^2 dt \end{aligned}$$



and the cross-quadratic variations

$$d[M, N]_t = (dM_t)(dN_t) = \sigma_t g_t (dB_t)^2 = \sigma_t g_t dt$$

$$d[M, U]_t = (dM_t)(dU_t) = \sigma_t^2 (dB_t)(dB_t) = 0$$

$$d[M, Z]_t = (dM_t)(dZ_t) = \rho \sigma_t^2 (dB_t)(dB_t) + \sqrt{1 - \rho^2} \sigma_t^2 (dB_t)^2 = \sqrt{1 - \rho^2} \sigma_t^2 dt$$

Next we define

$$dH_t = \mu_t dt \quad \text{and} \quad dK_t = f_t dt$$

and observe that since  $H_t$  and  $K_t$  have finite first variation we have that

$$d[H]_t = (dH_t)^2 = \mu_t^2 (dt)^2 = 0 \quad \text{and} \quad d[K]_t = (dK_t)^2 = f_t^2 (dt)^2 = 0$$

Furthermore if  $X_t = H_t + M_t$  and  $Y_t = K_t + N_t$  then using the previous calculations

$$d[X]_t = d[X, X]_t = d[H + M, H + M]_t = d[H]_t + d[M]_t + 2d[H, M]_t = \sigma_t^2 dt$$

$$d[X, Y]_t = d[H + M, K + N]_t = d[H, K + N]_t + d[M, K + N]_t = d[M, N]_t = \sigma_t g_t dt$$

or using the formal algebra

$$d[X]_t = (dX_t)^2 = \mu_t^2 (dt)^2 + 2\mu_t \sigma_t (dt)(dB_t) + \sigma_t^2 (dB_t)^2 = 0 + 0 + \sigma_t^2 dt = \sigma_t^2 dt$$

$$\begin{aligned} d[X, Y]_t &= (dX_t)(dY_t) = \mu_t f_t (dt)^2 + \sigma_t f_t (dt)(dB_t) + g_t f_t (dt)(dB_t) + \sigma_t g_t (dB_t)^2 \\ &= 0 + 0 + 0 + \sigma_t g_t dt = \sigma_t g_t dt \end{aligned}$$



## Stochastic Differential Equations

### 1. Definitions

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space equipped with a filtration  $\{\mathcal{F}_t\}_{\mathcal{T}}$ . Let  $B_t = (B_1(t), \dots, B_m(t)) \in \mathbb{R}^m$  be a  $m$ -dimensional Brownian motion with  $\{B_j(t)\}_{j=1}^m$  a collection of mutually independent Brownian motions such that  $B_j(t) \in \mathcal{F}_t$  and for any  $0 \leq s < t$ ,  $B_j(t) - B_j(s)$  is independent of  $\mathcal{F}_s$ . Obviously, these conditions are satisfied by the natural filtration  $\{\mathcal{F}_t^B\}_{\mathcal{T}}$ .

DEFINITION 1.1. Let  $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma_j: \mathbb{R}^d \rightarrow \mathbb{R}^d$  for  $j = 1, \dots, m$  be fixed functions. An equation of the form

$$dX_t = \mu(X_t) dt + \sum_{j=1}^m \sigma_j(X_t) dB_j(t) \quad (6.1)$$

representing the integral equation

$$X_t = x + \int_0^t \mu(X_s) ds + \sum_{j=1}^m \int_0^t \sigma_j(X_s) dB_j(s). \quad (6.2)$$

where  $X_t$  is an unknown process is a Stochastic Differential Equation (SDE) driven by the Brownian motion  $\{B_t\}$ . The functions  $\mu(x)$ ,  $\sigma(x)$  are called the drift and diffusion coefficients, respectively.

It is more compact to introduce the matrix

$$\sigma(x) = \begin{pmatrix} | & & | \\ \sigma_1(x) & \cdots & \sigma_m(x) \\ | & & | \end{pmatrix} \in \mathbb{R}^{m \times d}$$

and write

$$dX_t = b(X_t) dt + \sigma(X_t) dB_t. \quad (6.3)$$

There are different concepts of solution for a SDE. The most natural is the one of *strong* solution:

DEFINITION 1.2. A stochastic process  $\{X_t\}$  is a strong solution to the SDE (6.1) driven by the Brownian motion  $B_t$  with (possibly random) initial condition  $X_0 \in \mathbb{R}$  if the following holds

- i)  $\{X_t\}$  is adapted to  $\{\mathcal{F}_t\}$ ,
- ii)  $\{X_t\}$  is continuous,
- iii)  $X_t = X_0 + \int_0^t \mu(X_s) ds + \sum_{j=1}^m \int_0^t \sigma_j(X_s) dB_j(s)$  almost surely.

REMARK 1.3. Often, the choice of Brownian motion in the above definition is implicit. However, it is important that the strong solution of an SDE depends on the chosen Brownian motion driving it. A conceptually useful way to restate strong existence, say for all  $t \geq 0$ , is that there exists a measure map  $\Phi: (t, B) \mapsto X_t(B)$  from  $[0, \infty) \times C([0, \infty), \mathbb{R}^d) \rightarrow \mathbb{R}^d$  such that  $X_t = \Phi(t, B)$  solves (6.2) and  $X_t$  is measurable with respect to the filtration generated by  $B_t$ .

DEFINITION 1.4. We say that a strong solution to (6.1) (driven by a Brownian motion  $B_t$ ) is strongly unique if for any two solutions  $X_t, Y_t$  with the same initial condition  $X_0$  of (6.1) we have that

$$\mathbb{P}[X_t = Y_t \text{ for all } t \in [0, T]] = 1.$$

REMARK 1.5. By definition, the strong solution of a SDE is continuous. For this reason, to prove strong uniqueness it is sufficient to prove that two solutions  $X_t, Y_t$  satisfy

$$\mathbb{P}[X_t = Y_t] = 1 \quad \text{for all } t \in [0, T].$$

Indeed, assuming that  $X_t$  is a version of  $Y_t$ , by countable additivity, the set  $A = \{\omega: X_t(\omega) = Y_t(\omega) \text{ for all } t \in \mathbb{Q}_+\}$  has probability one. By right-continuity (resp. left-continuity) of the sample paths, it follows that  $X$  and  $Y$  have the same paths for all  $\omega \in A$ .

## 2. Examples of SDEs

We now consider a few useful examples of SDEs that have a strong solution.

**2.1. Geometric Brownian Motion.** The geometric Brownian motion, or Black Scholes model in finance, is a stochastic process  $X_t$  that solves the SDE

$$dX_t = \mu X_t dt + \sigma X_t dB_t, \tag{6.4}$$

where  $\mu, \sigma \in \mathbb{R}$  are constants. This model can be used to describe the evolution of the price  $X_t$  of a stock, which is assumed to have a mean increase and fluctuations with variance that depend both linearly on the stock price  $X_t$ . The coefficients  $\mu, \sigma$  are called the percentage drift and percentage volatility, respectively. We see immediately that (6.4) has a solution by Itô's formula: letting  $f(x) = \log x$  we have that

$$d \log X_t = \frac{1}{X_t} (\mu X_t dt + \sigma X_t dB_t) - \frac{1}{2} \frac{1}{X_t^2} \sigma^2 X_t^2 dt = (\mu - \frac{1}{2} \sigma^2) dt + \sigma dB_t,$$

Therefore, by integrating and exponentiating, that the solution of the equation reads

$$X_t = X_0 \exp \left[ \left( \mu - \frac{1}{2} \sigma^2 \right) t + \sigma B_t \right].$$

The uniqueness of this solution will be proven shortly.

**2.2. Stochastic Exponential.** Let  $X_t$  be an Itô process with differential  $dX_t = \mu_t dt + \sigma_t dB_t$ . We consider the following SDE:

$$dU_t = U_t dX_t \tag{6.5}$$

with initial condition  $U_0 = u_0 \in \mathbb{R}$ . Note that often one chooses  $u_0 = 1$ . Since the above SDE is analogous to the ODE  $df = f dt$  whose solution is given by the exponential function  $f(t) = \exp(t)$ , the process  $U_t$  solving (6.5) is often called the *stochastic exponential* and one writes  $X_t = \mathcal{E}(X)_t$ . The following result ensures that this process exists and is unique:

PROPOSITION 2.1. The SDE (6.5) has a unique strong solution, given by

$$U_t = \mathcal{E}(X)_t := U_0 \exp \left[ X_t - X_0 - \frac{1}{2} [X]_t \right] = U_0 \exp \left[ \int_0^t \left( \mu_s - \frac{1}{2} \sigma_s^2 \right) ds + \int_0^t \sigma_s dB_s \right]. \tag{6.6}$$

PROOF. If  $u_0 = 0$ , then it is immediate by (6.5) that  $U_t \equiv 0$  for all  $t \geq 0$  and it is the only solution. Now suppose  $u_0 \neq 0$ . We start by proving existence. The proposed solution is

clearly adapted, so defining  $V_t = X_t - X_0 - \frac{1}{2}[X]_t$  only have to verify that  $e^{V_t}$  satisfies (6.5), *i.e.*,  $d(e^{V_t}) = e^{V_t} dX_t$ . First, using Itô's formula, we get

$$dV_t = dX_t - \frac{1}{2} d[X]_t = dX_t - \sigma_t^2 dt \quad (6.7)$$

$$d(e^{V_t}) = e^{V_t} \left( dV_t + \frac{1}{2} d[V]_t \right). \quad (6.8)$$

Note that (6.7) implies  $[V]_t = [X]_t$ , because the  $\int \sigma_t^2 dt$  part is of finite variation. Therefore  $dV_t = dX_t - \frac{1}{2} d[V]_t$ . Plugging it back to (6.8), we have the desired equality.

We next check that the solution is *unique*. Suppose we have another solution  $\tilde{U}$  that satisfies (6.5). Notice that  $U$  is nonzero when  $u_0 \neq 0$ , we can compute  $d(\tilde{U}_t/U_t)$

$$\begin{aligned} d(\tilde{U}_t/U_t) &= \tilde{U}_t d(1/U_t) + \frac{1}{U_t} d\tilde{U}_t + d[\tilde{U}, U^{-1}]_t \\ &= \tilde{U}_t \left[ -\frac{dU_t}{U_t^2} + \frac{d[U]_t}{U_t^3} \right] + \frac{\tilde{U}_t dX_t}{U_t} + d[\tilde{U}, U^{-1}]_t \\ &= \frac{\tilde{U}_t \sigma_t^2 dt}{U_t} + d[\tilde{U}, U^{-1}]_t. \end{aligned}$$

We need the stochastic differential of  $U^{-1}$ .

$$\begin{aligned} d(U^{-1}) &= -\frac{dU_t}{U_t^2} + \frac{d[U]_t}{U_t^3} = -\frac{dX_t}{U_t} + \frac{\sigma_t^2 dt}{U_t} \\ \therefore d[\tilde{U}, U^{-1}]_t &= -\sigma_t^2 \tilde{U}_t U_t^{-1} dt \end{aligned}$$

Plugging this back gives  $d(\tilde{U}_t/U_t) = 0$ . So the ratio of the two solution stays a constant for all  $t \geq 0$ . Since both solutions start at  $u_0$ , we conclude  $\mathbb{P}[\tilde{U}_t = U_t, \forall t \geq 0] = 1$ .  $\square$

Similarly to the above definition. we introduce the *stochastic logarithm*  $X_t = \mathcal{L}(U)_t$  of a process  $U_t$  with a stochastic differential  $dU_t = \mu'_t dt + \sigma'_t dB'_t$  and  $U_t \neq 0$  as the solution to the following SDE:

$$dX_t = \frac{dU_t}{U_t} \text{ and } X_t = 0. \quad (6.9)$$

Again we have that the solution to the above SDE exists and is unique.

**PROPOSITION 2.2.** *Under the conditions listed above, the SDE (6.9) has a unique solution given by*

$$\mathcal{L}(U)_t = \log \left( \frac{U_t}{U_0} \right) + \int_0^t \frac{d[U]_s}{U_s^2}.$$

Furthermore, as suggested by the framework of stochastic calculus, the above operators are inverse wrt each other:

**PROPOSITION 2.3.** *If  $u_0 = 1$  we have  $\mathcal{L}(\mathcal{E}(X))_t = X_t$  and if  $U_t \neq 0$  then  $\mathcal{E}(\mathcal{L}(U))_t = U_t$ .*

**PROOF.** It is enough to check that  $dX_t = d(\mathcal{E}(X)_t)/\mathcal{E}(X)_t$  and  $dU_t = \mathcal{L}(U)_t d\mathcal{L}(U)_t$ .  $\square$

**2.3. Linear SDEs.** Let  $\{\alpha_t\}, \{\beta_t\}, \{\gamma_t\}, \{\delta_t\}$  be given (*i.e.*, independent on  $X_t$ ) continuous stochastic processes adapted to the filtration  $\{\mathcal{F}_t\}$ . We consider the family of SDEs given by

$$dX_t = (\alpha_t + \beta_t X_t) dt + (\gamma_t + \delta_t X_t) dB_t. \quad (6.10)$$

We proceed to solve such family of SDEs, which includes as special cases some of the examples treated previously in this course. We do so in two steps:

i) First we consider the case where  $\alpha_t = \gamma_t \equiv 0$ . In this case we should solve the SDE

$$dU_t = \beta_t U_t dt + \delta_t U_t dB_t,$$

which by Proposition 2.1, choosing  $U_0 = 1$  and defining  $dY = \beta_t dt + \delta_t dB_t$  has the unique solution

$$U_t = \mathcal{E}(Y)_t = \exp \left[ \int_0^t \left( \beta_s - \frac{1}{2} \delta_s^2 \right) ds + \int_0^t \delta_s dB_s \right].$$

ii) We now proceed to consider the full SDE (6.10), and make the ansatz of a separable solution, *i.e.*, assume that  $X_t = U_t V_t$  where  $dV_t = a_t dt + b_t dB_t$  for unknown processes  $\{a_t\}, \{b_t\}$ . Then we compute

$$\begin{aligned} dX_t &= U_t dV_t + V_t dU_t + d[U, V](t) \\ &= U_t a_t dt + U_t b_t dB_t + V_t \beta_t U_t dt + V_t \delta_t U_t dB_t + b_t \delta_t U_t dt \\ &= (a_t U_t + b_t \delta_t U_t + \beta_t X_t) dt + (b_t U_t + \delta_t X_t) dB_t. \end{aligned}$$

We notice that the above expression coincides with the RHS of (6.10) if

$$a_t = \frac{\alpha_t - \delta_t \gamma_t}{U_t} \quad \text{and} \quad b_t = \frac{\gamma_t}{U_t}.$$

This uniquely defines the process  $V_t$  (whose initial condition is fixed by the fact that  $U_0 = 1$  to be  $V_0 = X_0$ ) as

$$V_t = X_0 + \int_0^t \frac{\alpha_s - \delta_s \gamma_s}{U_s} ds + \int_0^t \frac{\gamma_s}{U_s} dB_s,$$

which in turn defines the solution to (6.10) as  $X_t = U_t \cdot V_t$ .

EXAMPLE 2.4. Letting  $a, b \in \mathbb{R}$ , consider the following SDE on  $t \in (0, T)$ :

$$dX_t = \frac{b - X_t}{T - t} dt + dB_t \quad \text{with } X_0 = a. \quad (6.11)$$

It is clear that this is a linear SDE with

$$\alpha_t = \frac{b}{T - t}, \quad \beta_t = \frac{1}{T - t}, \quad \gamma_t = 1, \quad \delta_t = 0.$$

Therefore, the solution to (6.11) is given by

$$X_t = a \left( 1 - \frac{t}{T} \right) + b \frac{t}{T} + (T - t) \int_0^t \frac{1}{T - s} ds. \quad (6.12)$$

Since  $\int_0^t (T - s)^{-1} ds < \infty$  for all  $t < T$  the Itô integral in (6.12) is a martingale. Furthermore, as we have proven in Homework 2 it is a gaussian process. Hence,  $X_t$  is also a gaussian process, with  $\mathbb{E}[X_t] = a + (a - b)t/T$  and covariance structure

$$\begin{aligned} \text{Cov}(X_s, X_t) &= (T - t)(T - s) \text{Cov} \left( \int_0^{\min(s,t)} \frac{1}{T - q} dB_q, \int_0^{\min(s,t)} \frac{1}{T - q} dB_q + \int_{\min(s,t)}^{\max(s,t)} \frac{1}{T - q} dB_q \right) \\ &= (T - \max(t, s))(T - \min(t, s)) \text{Var} \left( \int_0^{\min(s,t)} \frac{1}{T - q} dB_q \right) = \min(s, t) - \frac{st}{T} \end{aligned}$$

where in the second equality we have used that the covariance of independent random variables is 0. The above expression suggests that the variance of the process  $X_t$  is 0 at  $t = 0$  and  $t = T$ , and maximized at  $t = T/2$ , while the expected value of the process  $X_t$  is on the line interpolating between  $a$  and  $b$ . Hence the name *Brownian Bridge*: the process above can be seen as a *Brownian Motion*

with initial condition  $B_0 = a$  and conditioned on its final value  $B_T = b$ . Indeed, one can prove (cfr Klebaner, Example 5.11) that

$$\lim_{t \rightarrow T} (T - t) \int_0^t \frac{1}{T - s} dB_s = 0 \quad \text{a.s.}$$

### 3. Existence and Uniqueness for SDEs

In this section we prove a theorem giving sufficient conditions on the coefficients  $\mu(\cdot), \sigma(\cdot)$  for the existence and uniqueness of the solution to the associated SDE. We will consider solutions to the equation

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dB_t \quad (6.13)$$

Note that we have assumed explicit dependence on time  $t$  of the drift and diffusion coefficients as this will allow us to weaken the conditions of the following theorem:

**THEOREM 3.1.** *Fix a terminal time  $T > 0$ . Assume that  $\sigma(t, x)$  and  $\mu(t, x)$  are globally Lipschitz continuous, i.e., that there is a positive constant  $K$  so that for any  $t \in [0, T]$  and any  $x$  and  $y$  we have  $|\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq K|x - y|$ . Then the SDE (6.13) with initial condition  $X_0 = x$  has a solution*

$$X_t = x + \int_0^t \mu(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s,$$

and this solution is strongly unique. Furthermore the solution satisfies is in  $L^2(\Omega \times [0, T])$ , i.e.,

$$\mathbb{E} \left[ \int_0^t X_s^2 ds \right] < \infty.$$

It is clear that in order for the solution  $X_t$  to be well defined we need that

$$\int_0^t |\mu(s, X_s)| ds < \infty \quad \text{and} \quad \int_0^t \sigma(s, X_s)^2 ds < \infty \quad \text{a.s.}$$

However, the assumptions of Theorem 3.1, while being easier to check, are more strict than the ones above. It is useful to recall that these assumptions are needed even for ODEs to have existence and uniqueness: we remind in the following examples how the non-Lipschitz character of the drift can

**EXAMPLE 3.2 (Existence).** *The ODE*

$$\frac{dx}{dt} = x^2 \quad \text{with} \quad x(0) = 1$$

has a drift coefficient  $\mu(x) = x^2$  that is not uniformly Lipschitz continuous (although it is locally Lipschitz continuous) because it grows faster than linearly. This ODE has a solution  $x(t) = \frac{1}{1-t}$ . However, it is clear that this solution is only well defined for all  $t \in (0, 1)$  and diverges for  $t \rightarrow 1$ . In other words, the solution to this ODE does not exist beyond  $t = 1$ .

**EXAMPLE 3.3 (Uniqueness).** *The ODE*

$$\frac{dx}{dt} = 2\sqrt{|x|}$$

is not locally Lipschitz continuous at  $x = 0$ . The solution of this ODE with initial condition  $x_0 = 0$  is not unique. Indeed, it is immediate to check that

$$x_{1,t} = 0 \quad \text{and} \quad x_{2,t} = t^2,$$

are both solution to this equation with the given initial condition.

#### 3.1. Proof of Theorem 3.1.

**Preparatory results.** We start by proving two very useful lemmas:

LEMMA 3.4. (*Gronwall's inequality*) Let  $y(t)$  be a nonnegative function such that

$$y(t) \leq A + D \int_0^t y(s) \, ds \quad (6.14)$$

for nonnegative  $A, D \in \mathbb{R}$ . Then  $y(t)$  satisfies

$$f(t) \leq A \exp(Dt)$$

PROOF OF LEMMA 3.4. By repeatedly iterating the (6.14) we obtain

$$\begin{aligned} y(t) &\leq A + \int_0^t D y(s) \, ds \\ &\leq A + \int_0^t D \left( A + \int_0^s D y(q) \, dq \right) \, ds \\ &\leq A + A \int_0^t D \, ds + \int_0^t D \int_0^s D y(q) \, dq \, ds \\ &\leq A + ADt + D^2 \int_0^t \int_0^s \left( A + D \int_0^r y(\tau) \, d\tau \right) \, dq \, ds \\ &\leq A + ADt + AD^2 \int_0^t \int_0^s \, dq \, ds + D^3 \int_0^t \int_0^s \int_0^q y(\tau) \, d\tau \, dq \, ds \leq \dots \\ &\leq A + ADt + A \frac{D^2 t^2}{2} + A \frac{D^3 t^3}{3!} + D^4 \int_0^t \int_0^s \int_0^q \int_0^r y(\tau) \, d\tau \, dr \, dq \, ds. \end{aligned}$$

We notice that repeating the above procedure  $k$  times we will obtain the first  $k$  terms of the Taylor expansion of  $A \exp Dt$  plus a remainder term resulting from an integral iterated  $k + 1$  times. For finite  $T$  we can bound such integral by defining the constants

$$C := \int_0^T y(s) \, ds < \infty \quad \text{and} \quad G := A + DC,$$

so that  $y(t) \leq G$ . Consequently, we can bound the remainder term by  $Gt^{k+1}D^{k+1}/k!$  which vanishes exponentially fast in the limit  $k \rightarrow \infty$ , uniformly in  $t \in [0, T]$ .

Alternative proofs assuming the existence and uniqueness of solutions to ODEs can be found in any good ODE or dynamics book. For instance [6] or [5].  $\square$

LEMMA 3.5. Let  $\{y_n(t)\}$  be a sequence of functions satisfying

- $y_0(t) \leq A,$
- $y_{n+1}(t) \leq D \int_0^t y_n(s) \, ds < \infty \quad \forall n > 0, t \in [0, T],$

for positive constants  $A, D \in \mathbb{R}$ , then  $y_n(t) \leq CD^n t^n / n!$ .

PROOF. the proof of this result goes by induction: the first step is trivial, while for the induction step we have

$$y_{n+1}(t) \leq D \int_0^t y_n(s) \, ds \leq D \int_0^t C \frac{D^n t^n}{n!} \, ds = C \frac{D^{n+1} t^{n+1}}{(n+1)!}.$$

$\square$





where the second inequality uses the fact that the coefficients are growing no faster than linearly. Second, similar computation as for the uniqueness yields

$$y_{n+1}(t) \leq 2K^2(1+T) \int_0^t y_n(s) ds \quad \forall t \in [0, T], \quad n = 0, 1, 2, \dots$$

which is finite by induction. Lemma 3.5 implies

$$y_n(t) = \mathbb{E} [(X_{n+1}(t) - X_n(t))^2] \leq C \frac{(4K^2 + 4K^2T)^n}{n!},$$

which goes to zero uniformly for all  $t \in [0, T]$ . We thus conclude  $X_n(t)$  converges in  $L^2$  uniformly and their  $L^2$ -limit,  $X(t) \in L^2(\Omega, \mathcal{F}, P)$ .

It remains to show that the limit process  $X(t)$  solves (6.13). Since  $X_n \xrightarrow{L^2} X$ , we have

$$\begin{aligned} & \mathbb{E} [\mu(t, X_n(t)) - \mu(t, X(t))]^2 + \mathbb{E} [\sigma(t, X_n(t)) - \sigma(t, X(t))]^2 \\ & \leq K^2 \mathbb{E} [(X_n(t) - X(t))^2] + K^2 \mathbb{E} [(X_n(t) - X(t))^2] \\ & \rightarrow 0, \quad \text{uniformly in } t \end{aligned}$$

By Itô's isometry and Fubini:

$$\begin{aligned} \mathbb{E} \left[ \left( \int_0^t \sigma(s, X_n(s)) dB_s - \int_0^t \sigma(s, X(s)) dB_s \right)^2 \right] &= \mathbb{E} \left[ \left( \int_0^t (\sigma(s, X_n(s)) - \sigma(s, X(s))) dB_s \right)^2 \right] \\ &= \int_0^t \mathbb{E} [(\sigma(s, X_n(s)) - \sigma(s, X(s)))^2] ds \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Similarly, by Cauchy-Schwarz inequality we have that :

$$\begin{aligned} \mathbb{E} \left[ \left( \int_0^t \mu(s, X_n(s)) dB_s - \int_0^t \mu(s, X(s)) ds \right)^2 \right] &= \mathbb{E} \left[ \left( \int_0^t (\mu(s, X_n(s)) - \mu(s, X(s))) ds \right)^2 \right] \\ &= t \int_0^t \mathbb{E} [(\mu(s, X_n(s)) - \mu(s, X(s)))^2] ds \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

We thus have

$$X(t) = x + \int_0^t \mu(s, X(s)) ds + \int_0^t \sigma(s, X(s)) dB_s,$$

i.e.,  $X(t)$  solves (6.13).

**REMARK 3.6.** *Looking through the proof of Theorem 3.1 we see that the assumption of global Lipschitz continuity can be weakened to the following assumption*

- i)  $|\mu(t, x)| + |\sigma(t, x)| < C(1 + |x|)$  (necessary for existence),
- ii)  $|\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq C|x - y|$  (necessary for uniqueness).

#### 4. Weak solutions to SDEs

Until now we have studied strong solutions to SDEs, *i.e.*, solutions for which a Brownian motion (and a probability space) is given *in advance*, and that are constructed based on such Brownian motion. If we are only given some functions  $\mu(x)$  and  $\sigma$  *without* fixing a Brownian motion, we may be able to construct a *weak solution* to an SDE of the form (6.1). Such solutions allow to choose a convenient Brownian motion (and consequently a probability space!) for the solution  $X_t$  to satisfy the desired SDE.

DEFINITION 4.1. A weak solution of the stochastic differential equation (6.1) with initial condition  $X_0 \sim \rho_0$  for a given probability distribution  $\rho_0$  is a continuous stochastic process  $X_t$  defined on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  such that for some Brownian motion  $B_t$  and some filtration  $\{\mathcal{F}_t\}$  such that  $B_j(t) \in \mathcal{F}_t$  and for any  $0 \leq s < t$ ,  $B_j(t) - B_j(s)$  is independent of  $\mathcal{F}_s$ , the process  $X_t$  is adapted and satisfies the stochastic integral equation (6.2).

In other words, in the case of a weak solution we are free to choose *some* convenient Brownian motion that allows  $X_t$  to be a solution. In this sense, these solutions are also distributional solutions, *i.e.*, solutions that have the “right” marginals.

Because two solutions  $X_t, Y_t$  may live on different probability spaces, we cannot compare their paths as in the case of strong solutions. Instead, we weaken the concept of strong uniqueness to the one of *weak uniqueness*, *i.e.*, uniqueness in law of the solution process:

DEFINITION 4.2. The weak solution of a SDE is said to be weakly unique if any two solutions  $X_t, Y_t$  have the same law, *i.e.*, for all  $\{t_i \in [0, T]\}, \{A_i \in \mathcal{B}\}$  we have

$$\mathbb{P} \left[ \bigcap_i \{X_{t_i} \in A_i\} \right] = \mathbb{P} \left[ \bigcap_i \{Y_{t_i} \in A_i\} \right].$$

EXAMPLE 4.3. Consider the SDE  $dY_t = dB_t$  with initial condition  $Y_0 = 0$ . This SDE has clearly a strong solution, which is  $Y_t = B_t$ . If we let  $W_t$  be another Brownian motion (possibly defined on another probability space) then  $W_t$  will not be, in general, a strong solution to the above SDE (in the case that the two probability spaces are different, the two solutions cannot even be compared). It will, however, be a weak solution to the SDE, as being a Brownian motion completely determines the marginals of the process.

We will now consider an example for which there exists a weak solution, but not a strong solution:

EXAMPLE 4.4 (Tanaka’s SDE). For certain  $\mu$  and  $\sigma$ , solutions to (6.1) may exist for some Brownian motion and some admissible filtrations but not for others. Consider the SDE

$$dX_t = \text{sign}(X_t)dB_t, \quad X_0 = 0; \tag{6.15}$$

where  $\sigma(t, x) = \text{sign}(x)$  is the sign function

$$\text{sign}(x) = \begin{cases} +1, & \text{if } x \geq 0 \\ -1, & \text{if } x < 0. \end{cases}$$

The function  $\sigma(x)$  is not continuous and thus not Lipschitz. A strong solution does not exist for this SDE, with the filtration  $\mathbb{F} = (\mathcal{F}_t)$  chosen to be  $\mathcal{F}_t := \sigma(B_s, 0 \leq s \leq t)$ . Suppose  $X_t$  is a strong solution to Tanaka’s SDE, then we must have

$$\tilde{\mathcal{F}}_t := \sigma(X_s, 0 \leq s \leq t) \subseteq \mathcal{F}_t. \tag{6.16}$$

Notice that for any  $T \geq 0$ ,  $\int_0^T \mathbb{E} [\text{sign}(X_t)^2] ds < \infty$ , the Itô integral  $\int_0^t \text{sign}(X_t)dB_s$  is well defined and  $X_t$  is a martingale. Moreover, the quadratic variation of  $X_t$  is

$$[X]_t = \int_0^t [\text{sign}(X_t)]^2 ds = \int_0^t 1 \cdot ds = t,$$

thus  $X_t$  must be a Brownian motion (by Lévy’s characterization, to be proved later). We may denote  $X_t = \tilde{B}_t$  to emphasize that it is a Brownian motion. Now multiplying both sides of (6.15) by  $\text{sign}(X_t)$ , we obtain

$$dB_t = \text{sign}(\tilde{B}_t)d\tilde{B}_t. \tag{6.17}$$

and thus  $B_t = \int_0^t \text{sign}(\tilde{B}_s) d\tilde{B}_s$ . By Tanaka's formula (to be shown later), we then have

$$B_t = |\tilde{B}_t| - \tilde{L}_t$$

where  $\tilde{L}_t$  is the local time of  $\tilde{B}_t$  at zero. It follows that  $B_t$  is  $\sigma(|\tilde{B}_s|, 0 \leq s \leq t)$ -measurable. This leads to a contradiction to (6.16), because it would imply that

$$\mathcal{F}_t \subseteq \sigma(|\tilde{B}_s|, 0 \leq s \leq t) \subsetneq \sigma(\tilde{B}_s, 0 \leq s \leq t) = \tilde{\mathcal{F}}_t.$$

Still, as we have seen above, choosing  $X_t = \tilde{B}_t$  there exists a Brownian motion  $B_s$  such that Tanaka's SDE holds. Such pair of Brownian motions forms a weak solution to Tanaka's equation.

## 5. Markov property of Itô diffusions

The solutions (weak or strong) to stochastic differential equations are referred to as *diffusion processes* (or Itô diffusions).

DEFINITION 5.1. An Itô process  $dX_t = \mu_t dt + \sigma_t dB_t$  is an Itô diffusion if  $\mu_t, \sigma_t$  are measurable wrt the filtration  $\{\mathcal{F}_t^X\}$  generated by  $X_t$  for all  $t \in [0, T]$ , i.e.,

$$\mu_t, \sigma_t \in \mathcal{F}_t^X.$$

REMARK 5.2. It is clear that solution  $\{X_t\}$  to the SDE  $dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dB_t$  for continuous functions  $\mu, \sigma$  is an Itô diffusion. For this reason, such SDEs are called of diffusion-type.

Recall Def. 7.12 that a Markov process is a process whose future depends on its past only through its present value, while if this property holds also for stopping times, the process is said to have the strong Markov property (cfr Def. 7.16).

THEOREM 5.3. The solution  $\{X_t\}$  to the SDE

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dB_t, \tag{6.18}$$

has the strong Markov property.

While we do not present the proof of this result, which can be found in [13], it should be intuitively clear why solutions to (6.18) have the Markov property. Indeed, we see that the drift and diffusion coefficients of the above SDE only depend on the time and on the value of  $X_t$  at that time (and not on its past value). This fact, combined with the independence of the increments of Brownian motion results in the Markov (and the strong Markov) property of such solutions.

## PDEs and SDEs: The connection

Throughout this chapter, except when specified otherwise, we let  $\{X_t\}$  be a solution to the SDE

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dB_t. \quad (7.1)$$

As we have seen in the last chapter, solutions to the above equation are referred to as diffusion processes. This name comes from the fact that Brownian motion, the archetypical diffusion process, was invented to model the diffusion of a dust particle in water. Similarly, in the world of partial differential equations, diffusion equations model precisely the same type of phenomenon. In this chapter we will see that the correspondence between these two domains goes well beyond this point.

### 1. Infinitesimal generators

Having seen in the previous chapter that solutions to SDEs possess the strong Markov property, we introduce the following operator to study the evolution of their finite-dimensional distributions:

**DEFINITION 1.1.** *The infinitesimal generator for a continuous time Markov process  $X_t$  is an operator  $A$  such that for any function  $f$ ,*

$$A_t f(x) := \lim_{dt \downarrow 0} \frac{\mathbb{E}[f(X_{t+dt}) | X_t = x] - f(x)}{dt}, \quad (7.2)$$

*provided the limit exists. The set of function for which the above limit exists is called the domain  $\mathcal{D}(A_t)$  of the generator.*

This operator encodes the infinitesimal change in the probability distribution of the process  $X_t$ . One way of seeing this is by choosing  $f(x) = \mathbb{1}_A(x)$  for a set  $A \in \mathbb{R}^d$ .

We now look at some examples where we find the explicit form of the generator for Itô diffusions:

**EXAMPLE 1.2.** *The infinitesimal generator for a standard one-dimensional Brownian motion  $B_t$  is*

$$A = \frac{1}{2} \frac{d^2}{dx^2}$$

*for all  $f$  that are  $C^2$  with compact support. To derive this, we first apply Itô's formula to any  $f \in C^2$  and write*

$$\begin{aligned} f(B_t) &= f(B_0) + \int_0^t \frac{d^2}{dx^2} f(B_s) dB_s + \int_0^t \frac{1}{2} \frac{d^2}{dx^2} f(B_s) ds \\ &= f(B_0) + \int_0^t f'(B_s) dB_s + \int_0^t \frac{1}{2} \frac{d^2}{dx^2} f(B_s) ds \end{aligned}$$

*Apply this formula to two time points  $t$  and  $t+r$ , we have*

$$f(B_{t+r}) = f(B_t) + \int_t^{t+r} f'(B_s) dB_s + \int_t^{t+r} \frac{1}{2} \frac{d^2}{dx^2} f(B_s) ds$$

When  $f$  has compact support,  $f'(x)$  is bounded and suppose  $|f'(x)| \leq K$  and thus for each  $t$ ,

$$\int_0^t \mathbb{E}(f'^2(B_s)) ds \leq \int_0^t K^2 ds = K^2 t < \infty.$$

Hence the first integral has expectation zero, due to Itô's isometry. It follows that

$$\begin{aligned} \mathbb{E}[f(B_{t+r})|B_t = x] &= f(x) + \mathbb{E} \left\{ \int_t^{t+r} f'(B_s) dB_s + \int_t^{t+r} \frac{1}{2} \frac{d^2}{dx^2} f(B_s) ds \middle| B_t = x \right\} \\ &= f(x) + \mathbb{E} \left\{ \int_t^{t+r} f'(B_s) dB_s + \int_t^{t+r} \frac{1}{2} \frac{d^2}{dx^2} f(B_s) ds \right\} \\ &= f(x) + \mathbb{E} \int_t^{t+r} \frac{1}{2} \frac{d^2}{dx^2} f(B_s) ds \end{aligned}$$

where the second equality is due to the independence of the post- $t$  process  $B_{t+s} - B_t$  and  $B_t$ . Subtracting  $f(x)$ , dividing by  $r$  and letting  $r \rightarrow 0$  on both sides, we obtain

$$\begin{aligned} Af(x) &= \lim_{r \downarrow 0} \frac{\mathbb{E}[f(B_{t+r})|B_t = x] - f(x)}{r} = \lim_{r \downarrow 0} \frac{\mathbb{E} \left[ \int_t^{t+r} \frac{1}{2} \frac{d^2}{dx^2} f(B_s) ds \right]}{r} \\ &= \frac{d \left( \mathbb{E} \left[ \int_t^{t+r} \frac{1}{2} \frac{d^2}{dx^2} f(B_s) ds \right] \right)}{dr} \stackrel{(*)}{=} \frac{d \int_t^{t+r} \frac{1}{2} \mathbb{E} \left[ \frac{d^2}{dx^2} f(B_s) \right] ds}{dr} \\ &= \frac{1}{2} \frac{d^2}{dx^2} f(B_s) \Big|_{s=t} = \frac{1}{2} \frac{d^2}{dx^2} f(B_t) = \frac{1}{2} \frac{d^2}{dx^2} f(x) \end{aligned}$$

In the above calculation, we inverted the order of the integrals using Fubini-Tonelli's theorem.

REMARK 1.3. Here we omit the subscript  $t$  in the generator  $A$  because Brownian motion is time-homogeneous, i.e.,

$$\frac{\mathbb{E}[f(B_{t+dt})|B_t = x] - f(x)}{dt} = \frac{\mathbb{E}[f(B_{s+dt})|B_s = x] - f(x)}{dt}$$

and thus  $A_t f(x) = A_s f(x)$ . The generator  $A = \frac{1}{2} \frac{d^2}{dx^2}$  does not change with time.

The procedure to obtain the infinitesimal generator of Brownian motion can be straightforwardly generalized to the case of Itô diffusions:

EXAMPLE 1.4. Assume that  $X_t$  satisfies the SDE (7.1), then its generator  $A_t$  is

$$A_t f(x) = \mu(t, x) \frac{d}{dx} f(x) + \frac{\sigma^2(t, x)}{2} \frac{d^2}{dx^2} f(x) \quad (7.3)$$

for all  $f \in C^2$  with compact support. The computation is similar to the Brownian motion case. First apply Itô's formula to  $f(X_t)$  and get

$$f(X_{t+r}) = f(X_t) + \int_t^{t+r} \left\{ \mu(s, X_s) \frac{d}{dx} f(X_s) + \frac{\sigma^2(s, X_s)}{2} \frac{d^2}{dx^2} f(X_s) \right\} ds + \int_t^{t+r} \sigma(s, X_s) \frac{d}{dx} f(X_s) dB_s$$

Then using the fact that  $f \in C^2$  with compact support, the last integral has expectation zero. Conditioning on  $X_t = x$ , computing  $\frac{\mathbb{E}[f(X_{t+r})|X_t=x] - f(x)}{r}$ , exchanging integrals by Fubini-Tonelli and taking  $r \rightarrow 0$ , we conclude that the generator has the form (7.3).

The above example can be further generalized to the case when the function  $f$  also depends on time:

EXAMPLE 1.5. Consider the two dimensional process  $(t, X_t)$ , where the first coordinate is deterministic and the second coordinate  $X_t$  satisfies (7.1). We treat it as a process  $Y_t = (t, X_t) \in \mathbb{R}^2$ . In this case, the generator of  $Y_t$ , according to the definition in (7.2) is given by

$$\begin{aligned} A_t f(t, x) &:= \lim_{dt \downarrow 0} \frac{\mathbb{E}[f(Y_{t+dt}) | Y_t = (t, x)] - f(t, x)}{dt} \\ &= \mu(t, x) \frac{\partial}{\partial x} f(t, x) + \frac{\sigma^2(t, x)}{2} \frac{\partial^2}{\partial x^2} f(t, x) + \frac{\partial}{\partial t} f(t, x) \end{aligned} \quad (7.4)$$

for any  $f \in C^{1,2}$  that has compact support.

Formally speaking, if  $A_t$  is the generator of  $X_t$ , what  $A_t$  does to  $f$  is to map it to the “drift coefficient” in the stochastic differential of  $f(X_t)$ , i.e.,

$$df(X_t) = A_t f(X_t) \cdot dt + (\text{something}) \cdot dB_t$$

REMARK 1.6. The notation here is slightly different from [Klebaner], where Klebaner always uses  $L_t$  to denote the operator on functions  $f \in C^{1,2}$  so that

$$L_t f(t, x) = \mu(t, x) \frac{\partial}{\partial x} f(t, x) + \frac{\sigma^2(t, x)}{2} \frac{\partial^2}{\partial x^2} f(t, x) \quad (7.5)$$

and call such  $L_t$  the “generator of  $X_t$ ”. Comparing this form with (7.3) and (7.4), and since  $L_t$  acts on  $C^{1,2}$  functions, we can relate  $L_t$  to the generator  $A_t$  of  $(t, X_t)$ , i.e.,

$$L_t f(t, x) + \frac{\partial}{\partial t} f(t, x) = A_t f(t, x)$$

When we look at martingales constructed from the generators,  $A_t$  will give a more compact (and maybe more intuitive) notation.

**Exercise.** Find the generator  $A_t$  of  $(X_t, Y_t)$ , where  $X_t$  and  $Y_t$  satisfies

$$\begin{aligned} dX_t &= \mu(t, X_t)dt + \sigma(t, X_t)dB_t \\ dY_t &= \alpha(t, Y_t)dt + \beta(t, Y_t)dB_t \end{aligned}$$

What if  $X_t$  and  $Y_t$  are driven by two independent Brownian motions?

## 2. Martingales associated with diffusion processes

Suppose  $X_t$  solves (7.1) and  $A_t$  is its generator (see (7.3)). For  $f \in C^2$ , we know from Itô’s formula that

$$f(X_t) = f(X_0) + \int_0^t A_s f(X_s) dt + \int_0^t \sigma(s, X_s) f'(X_s) dB_s$$

Under proper conditions, the third term on the right is a well-defined Itô integral and also a martingale. We can construct martingales by isolating this integral, i.e., let

$$M_t := f(X_t) - f(X_0) - \int_0^t A_s f(X_s) ds \quad \left( = \int_0^t \sigma(s, X_s) f'(X_s) dB_s \right), \quad (7.6)$$

and we will see that for certain  $\mu$ ,  $\sigma$  and  $f$  functions,  $M_t$  will be a martingale. First of all, if  $X_t$  is a solution to (7.1) (either weak or strong), then, by definition,  $M_t$  is always  $\mathcal{G}_t := \sigma(X_s, 0 \leq s \leq t)$  measurable. So from now on, for the purpose of constructing martingales, we will only say “ $X_t$  solves the SDE (7.1)” without specifying whether  $X_t$  is a strong or a weak solution. Recall that if

$$\int_0^t \mathbb{E}[\sigma^2(s, X_s) f'^2(X_s)] ds < \infty, \quad (7.7)$$

then the Itô integral  $\int_0^t \sigma(s, X_s) f'(X_s) dB_s$  is a martingale. Therefore, the usual technical step is to prove (7.7) in order to conclude that  $M_t$  is a martingale.

Theorems 6.2 in [Klebaner] gives a set of conditions for  $M_t$  to be a martingale:

CONDITION 1. *Let the following assumptions hold:*

- i)  $\mu(t, x)$  and  $\sigma(t, x)$  are locally Lipschitz in  $x$  with a Lipschitz constant independent of  $t$  and are growing at most linearly in  $x$ ; and
- ii)  $f \in C^2$  and  $f'$  is bounded,

Condition (i) implies that (7.1) has a strong solution, but more importantly, it controls the speed of growth of the solution  $X_t$  (see Theorem 5.4 and also the proof in Theorem 6.2 of [Klebaner] for more details); (ii) controls the magnitude of  $f$ , which together with (i) ensure the finiteness of the integral in (7.7). [9, Theorem 6.3] gives an alternative set of conditions to Condition 1, however, the proof follows the same idea. The above result can be summarized in the following theorem

THEOREM 2.1. *Let  $\{X_t\}$  be a solution to (7.1),  $f$  a function such that Condition 1 holds, then  $M_t$  defined in (7.6) is a martingale.*

We now generalize the above result to the case when  $f$  is time-dependent. Let  $X_t$  solve (7.1). If  $A_t$  is the generator of the two dimensional process  $(t, X_t)$  (see the expression in (7.4)), then for any function  $f(t, x) \in C^{1,2}$

$$M_t := f(t, X_t) - f(0, X_0) - \int_0^t A_s f(s, X_s) ds \quad (7.8)$$

can be a martingale if  $\mu, \sigma$  and  $f$  satisfy certain conditions. Again, using Itô's formula, we see that

$$M_t = \int_0^t \sigma(s, X_s) \frac{\partial}{\partial x} f(s, X_s) dB_s$$

The approach to show that  $M_t$  is a martingale is the same as above. For example, if Condition 1 (ii) above is modified to

CONDITION 2. *Let the following assumptions hold:*

- i)  $\mu(t, x)$  and  $\sigma(t, x)$  are locally Lipschitz in  $x$  with a Lipschitz constant independent of  $t$  and are growing at most linearly in  $x$ ; and
- ii)  $f \in C^{1,2}$  and  $\frac{\partial}{\partial x} f(t, x)$  is bounded for all  $t$  and  $x$ .

then, we can conclude that  $M_t$  defined in (7.8) is a martingale:

THEOREM 2.2. *Let  $\{X_t\}$  be a solution to (7.1),  $f$  a function such that Condition 2 holds, then  $M_t$  defined in (7.8) is a martingale.*

One advantage of using  $A_t$  instead of  $L_t$  is that we can express  $M_t$  in the same form, that is,  $M_t := f(X_t) - f(X_0) - \int_0^t A_s f(X_s) ds$ , provided  $A_t$  is chosen to be the generator of  $X_t$  (which might be high-dimensional). The following are a few immediate consequences, stated under Condition 2. However, one should keep in mind that there are other conditions, under which these claims are also true.

COROLLARY 2.3 (Dynkin's formula). *Suppose that  $X_t$  solves (7.1) and that Condition 2 holds. Let  $A_t$  be the generator of  $(t, X_t)$  (see (7.4)). Then for any  $t \in [0, T]$ ,*

$$\mathbb{E}[f(t, X_t)] = \mathbb{E}[f(0, X_0)] + \mathbb{E} \left[ \int_0^t A_s f(s, X_s) ds \right],$$

*The result is also true if  $t$  is replaced by a bounded stopping time  $\tau \in [0, T]$ .*



COROLLARY 2.4. Assume that  $X_t$  solves (7.1) and that Condition 2 holds. If  $f$  solves the following PDE

$$(A_t f(t, x) =) \quad \mu(t, x) \frac{\partial}{\partial x} f(t, x) + \frac{\sigma^2(t, x)}{2} \frac{\partial^2}{\partial x^2} f(t, x) + \frac{\partial}{\partial t} f(t, x) \equiv 0,$$

then  $f(t, X_t)$  is a martingale.

EXAMPLE 2.5. Consider  $X_t = B_t$ , then  $\sigma \equiv 1$  and  $\mu \equiv 0$ , which satisfies Condition 2 (i). Then, for any  $f(t, x)$  that satisfies Condition 2 (ii') (or conditions in Theorem 6.3 of Klebaner) and solves

$$\frac{1}{2} \frac{\partial^2}{\partial x^2} f(t, x) + \frac{\partial}{\partial t} f(t, x) \equiv 0,$$

$f(t, B_t)$  is a martingale. For example,  $f(t, x) = x, x^2 - t, x^3 - 3tx, e^{t/2} \sin(x)$ , or  $e^{x-t/2}$ .

### 3. Connection with PDEs

In the previous section we have seen that under proper conditions the solution  $f$  of some PDE can be used to construct a martingale. In this section, we will see that the solutions of certain PDEs may be represented by the expectation of the solution of the SDE.

Throughout this section we will assume that  $X_t$  solves the SDE (7.1) whose coefficients satisfy Condition 2 (i), and  $A_t$  is the generator of  $(t, X_t)$  as given in (7.4). Furthermore, we assume that  $f$  satisfies Condition 2 (ii'). Note that other conditions, under which  $M_t$  defined in (7.8) is a martingale, would also work.

#### 3.1. Kolmogorov Backwards Equation.

THEOREM 3.1. Under the standing assumptions, if  $f(t, x)$  solves the PDE

$$\begin{cases} A_t f = 0 & \text{for all } t \in (0, T) \\ f(T, x) = g(x) \end{cases}, \quad (7.9)$$

for some function  $g$  such that  $\mathbb{E}[|g(X_T)|] < \infty$ . Then,

$$f(t, x) = \mathbb{E}(g(X_T) | X_t = x), \quad \text{for all } t \in [0, T].$$

PROOF. Under the standing assumptions and the fact that  $A_t f = 0$ , we know that  $f(t, X_t)$  is a martingale, due to Corollary 2.4. Then for any  $t \in [0, T]$ ,

$$\mathbb{E}[f(T, X_T) | \mathcal{F}_t] = f(t, X_t)$$

Using the boundary condition, we have  $f(T, X_T) = g(X_T)$ . The result then follows from the Markov property of the solution to the diffusion-type SDE, i.e.,

$$f(t, X_t) = \mathbb{E}[g(X_T) | \mathcal{F}_t] = \mathbb{E}[g(X_T) | X_t].$$

□

REMARK 3.2. Note that the above theorem, assuming that  $f(t, x)$  solves the given PDE, represents expectation values of the process  $X_t$  in terms of such solutions. Under suitable regularity conditions on the coefficients of the SDE (7.1) and on the boundary condition  $g(x)$  one can show that such expected value is the unique solution to the PDE (7.9). These results, however, go beyond the scope of this course and will not be presented here. We refer the interested reader to, e.g., [13].

DEFINITION 3.3. For a Itô diffusion  $\{X_t\}$  solving (7.1) the PDE (7.9) is called the Kolmogorov Backwards equation.

The name of the above PDE is due to the fact that it has to be solved *backwards* in time, *i.e.*, the boundary condition in (7.9) is fixed at the *end* of the time interval of interest. This may seem at first counterintuitive. One possible way to interpret this fact is that bringing the time derivative on the other side of the equality we obtain  $-\partial_t f = L_t f$ , where  $L_t$  is the generator defined in (7.5). In this form, the “arrow of time” is given by the fact that  $\sigma^2$  has nonnegative values and corresponds to the second derivative “widening the support” of  $f$ , and the negative sign in front of the time derivative corresponds to an evolution in the “reverse” direction. Another way of understanding the fact that the direction of time is reversed in (7.9) is the following: in order to establish an expectation WRT a certain function *in the future* (*e.g.*, the value of an option at expiration), an operator that evolves such function should project it *backwards* to the information we have at the moment, *i.e.*, the value of  $X_t$

EXAMPLE 3.4. Letting  $g(x) = \mathbb{1}_A(x)$ , we have that being able to solve (7.9) is equivalent to knowing

$$\mathbb{E}[\mathbb{1}_A(X_T)|X_t = x] = \mathbb{P}[X_T \in A|X_t = x].$$

EXAMPLE 3.5. Letting  $X_t$  be the solution to the Black Scholes model  $dX_t = \mu X_t dt + \sigma X_t dB_t$  and  $g(x) = V(x)$  some value function of an option at time  $T$ , then being able to solve (7.9) is equivalent to knowing the expected value of that option at expiration:  $\mathbb{E}[V(X_T)|X_t = x]$ .

We now state an extension of Theorem 3.1 which deals with the case where the right side of the PDE is nonzero.

THEOREM 3.6. Under the standing assumptions, if  $f(t, x)$  solves the PDE

$$\begin{cases} A_t f(t, x) = -\phi(x) & \text{for all } t \in (0, T) \\ f(T, x) = g(x) \end{cases}$$

for some bounded function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  and  $g$  such that  $\mathbb{E}[|g(X_T)|] < \infty$ . Then,

$$f(t, x) = \mathbb{E}\left(g(X_T) + \int_t^T \phi(X_s) ds \middle| X_t = x\right), \quad \text{for all } t \in [0, T].$$

PROOF. By Theorem 2.2 we have that

$$M_t := f(t, X_t) - f(0, X_0) - \int_0^t A_s f(s, X_s) ds$$

is a martingale. Plugging in  $A_t f(t, x) = -\phi(x)$  and taking conditional expectation of  $M_T | \mathcal{F}_t$ , since  $M_t = \mathbb{E}(M_T | \mathcal{F}_t)$  we get

$$f(t, X_t) - f(0, X_0) - \int_0^t A_s f(s, X_s) ds = \mathbb{E}(f(T, X_T) | \mathcal{F}_t) - f(0, X_0) - \mathbb{E}\left(\int_0^T A_s f(s, X_s) ds \middle| \mathcal{F}_t\right)$$

which can be rewritten as

$$f(t, X_t) = \mathbb{E}\left[g(X_T) + \int_t^T \phi(X_s) ds \middle| \mathcal{F}_t\right]$$

Finally, by the Markov property of Itô diffusions, we obtain

$$f(t, x) = \mathbb{E}\left[g(X_T) + \int_t^T \phi(X_s) ds \middle| X_t = x\right]$$

□

**3.2. Feynman-Kac formula.** Theorem 3.1 can be generalized even further:

**THEOREM 3.7 (Feynman-Kac Formula).** *Under the standing assumptions, if  $f(t, x)$  solves*

$$\begin{cases} A_t f(t, x) = r(t, x)f(t, x) & \text{for all } t \in [0, T] \\ f(T, x) = g(x) \end{cases}, \quad (7.10)$$

where  $r(t, x)$  and  $g(x)$  are some bounded functions, then

$$f(t, x) = \mathbb{E} \left( e^{-\int_t^T r(s, X_s) ds} g(X_T) \middle| X_t = x \right).$$

Furthermore,  $f(t, x)$  above is the unique solution to (7.10).

**PROOF.** The proof of uniqueness of the solution goes beyond the scope of this lecture and we do not prove it here. As in the previous cases, we want to show that the content of the expectation value is a martingale. Therefore, consider

$$M_\tau := e^{-\int_t^\tau r(s, X_s) ds} f(\tau, X_\tau).$$

Defining  $U_\tau = e^{-\int_t^\tau r(s, X_s) ds}$ ,  $Y_\tau = f(\tau, X_\tau)$  we apply Itô's formula to obtain

$$dM_\tau = d(U_\tau Y_\tau) = U_\tau d(Y_\tau) + Y_\tau d(U_\tau) + d[U, Y]_\tau.$$

Recall from the chapter on the stochastic exponential that  $U_\tau = \mathcal{E}(r(\tau, X_\tau))$  and that therefore  $dU_\tau = r(\tau, X_\tau)U_\tau d\tau$ . Furthermore, we recognize that  $U_\tau$  has finite variation, so  $d[U, Y] = 0$ . Combining these observations we obtain by Itô's formula for  $f$ , that

$$\begin{aligned} dM_\tau &= U_\tau \left( \left[ \partial_\tau f(\tau, X_\tau) + \mu(\tau, X_\tau) \partial_x f(\tau, X_\tau) + \frac{1}{2} \sigma(\tau, X_\tau)^2 \partial_{xx}^2 f(\tau, X_\tau) \right] d\tau \right. \\ &\quad \left. + \sigma(\tau, X_\tau) \partial_x f(\tau, X_\tau) dB_\tau - r(\tau, X_\tau) f(\tau, X_\tau) d\tau \right) \\ &= U_\tau ((A_\tau - r(\tau, X_\tau)) f(\tau, X_\tau) + \sigma(\tau, X_\tau) \partial_x f(\tau, X_\tau) dB_\tau). \end{aligned}$$

We immediately realize that the drift term in the above formula vanishes by assumption, and that the Itô integral term is a martingale by the standing assumptions on  $f$ ,  $\sigma$  and  $\mu$ . Consequently, the expected value of the martingale is constant and we have that

$$f(t, x)e^0 = M_t = \mathbb{E}[M_T | \mathcal{F}_t] = \mathbb{E} \left[ e^{-\int_t^T r(s, X_s) ds} g(X_T) \middle| X_t = x \right]$$

□

**EXAMPLE 3.8 (Example 3.5 continued).** *Let us consider the Black Scholes model i.e.,  $dX_t = \mu X_t dt + \sigma X_t dB_t$  for  $\sigma, \mu \in \mathbb{R}$ . We consider the case where one can cash his/her option and obtain a risk-free interest that satisfies the ODE*

$$dX_t = rX_t dt,$$

for a positive constant  $r \in \mathbb{R}$ . Then, one needs to factor such possible, risk-free earning in the value  $V(t, X_t)$  of the asset  $X_t$  (the underlying), i.e., compare the expected value at future time  $T$ ,  $V(X_T) = V^*(X_T)$  with the projected risk-free value today:

$$e^{r(T-t)} V(t, X_t) = \mathbb{E} [V^*(X_T) | X_t = x]$$

or, in other words,

$$V(t, X_t) = \mathbb{E} \left[ e^{-r(T-t)} V(X_T) \middle| X_t = x \right].$$

The above is an example of the expected value in Theorem 3.7, and therefore obeys the PDE

$$\begin{cases} \partial_t V(t, x) + \mu x \partial_x V(t, x) + \frac{1}{2} \sigma^2 x^2 \partial_{xx}^2 V(t, x) - r V(t, x) = 0 & \text{for all } t \in [0, T] \\ V(T, x) = V^*(x) \end{cases},$$

which is called the Black Scholes equation.

#### 4. Time-homogeneous Diffusions

In this section we now consider a class of diffusion processes whose drift and diffusion coefficients do *not* depend explicitly on time:

DEFINITION 4.1. *If  $X_t$  solves*

$$dX_t = \mu(X_t) dt + \sigma(X_t) dB_t \quad (7.11)$$

*then  $X_t$  is a time-homogeneous Itô diffusion process.*

Intuitively, the evolution of such processes does not depend on the time at which the process is started, *i.e.*,  $\mathbb{P}[X_t \in A | X_0 = x_0] = \mathbb{P}[X_{t+s} \in A | X_s = x_0]$  for all  $A \in \mathcal{B}(R)$ ,  $s \in \mathcal{T}$  s.t.  $t + s \in \mathcal{T}$ . In other words, their evolution is invariant wrt translations in time, whence the name *time-homogeneous*.

DEFINITION 4.2. *Given a SDE with a unique solution we define the associated Markov semigroup  $P_t$  by*

$$(P_t \phi)(x) = \mathbb{E}_x \phi(X_t)$$

To see that this definition satisfies the semigroup property observe that the Markov property states that

$$(P_{t+s} \phi)(x) = \mathbb{E}_x \phi(X_{t+s}) = \mathbb{E}_x \mathbb{E}_{X_s} \phi(X_t) = \mathbb{E}_x (P_t \phi)(X_s) = (P_s P_t \phi)(x).$$

Note that for the class of processes introduced above the infinitesimal generator is time-independent, *i.e.*, we have  $A_t f = A f$ . As a further consequence of the translation invariance (in time) of the SDE (7.11), the fact that the final condition of the backward Kolmogorov equation is at a specific time  $T$  is not relevant in this framework. This enables us to “store” the time-reversal in the function itself and look at the backward Kolmogorov equation as a forward equation as we explain below.

Let  $f_-(x, t)$  be a bounded,  $C^{2,1}$  function satisfying

$$\begin{aligned} \frac{\partial f_-}{\partial t} &= \mathcal{L} f_- \\ f_-(x, 0) &= g(x) \end{aligned} \quad (7.12)$$

where  $L$  is the generator defined in (7.5). For simplicity we also assume that  $g$  is bounded and continuous. Then we have the analogous result to Theorem 3.1

THEOREM 4.3. *under  $M_t = f_-(X_t, T - t)$  is a martingale for  $t \in [0, T]$ .*

PROOF. The proof is identical to the Brownian case. We start by applying Itô’s formula

$$\begin{aligned} f_-(X_s, T - s) - f_-(X_0, T) &= \int_0^s \left[ -\frac{\partial f_-}{\partial t}(X_\gamma, T - \gamma) + (\mathcal{L} f_-)(X_\gamma, T - \gamma) \right] d\gamma \\ &\quad + \int_0^s \frac{\partial f_-}{\partial x}(X_\gamma, T - \gamma) dB_\gamma. \end{aligned}$$

As before the integrand of the first integral is identically zero because  $\frac{\partial f_-}{\partial t} = \mathcal{L} f_-$ . Hence only the stochastic integral is left on the right-hand side.  $\square$

And as before we have

COROLLARY 4.4. *In the above setting*

$$f_-(x, t) = \mathbb{E}[g(X_t)|X_0 = x]$$

The restriction to bounded and continuous  $g$  is not needed.

PROOF OF COR. 4.4. Since  $s \mapsto f_-(X_s, T - s) - f_-(X_0, T)$  is a martingale,

$$\mathbb{E}[f_-(X_T, 0)|X_0 = x] = \mathbb{E}[u(X_0, T)|X_0 = x]$$

because at  $s = 0$  we see that  $f_-(X_s, T - s) - u(X(0), T) = 0$ . Since  $\mathbb{E}[u(X_0, T)|X_0 = x] = u(x, T)$  and  $\mathbb{E}[f_-(X_T, 0)|X_0 = x] = \mathbb{E}[g(X_T)|X_0 = x]$ , the proof is complete.  $\square$

For a more detailed discussion of Poisson and Dirichlet problems we refer to [13].

## 5. Stochastic Characteristics

To better understand Theorem 4.3 and Corollary 4.4, we begin by considering the deterministic case

$$\begin{aligned} \frac{\partial f_-}{\partial t} &= (b \cdot \nabla) f_- \\ f_-(x, 0) &= f(x) \end{aligned} \tag{7.13}$$

We want to make an analogy between the method of characteristics used to solve (7.13) and the results in Theorem 4.3 and Corollary 4.4. The method of characteristics is a method of solving (7.13) which in this simple setting amounts to finding a collection of curves (“characteristic curves”) along which the solution is constant. Let us call these curves  $x(t)$  where  $t$  is the parameterizing variable. Mathematically, we want  $f_-(\xi(t), T - t)$  to be a constant independent of  $t \in [0, T]$  for some fixed  $T$ . Hence the constant depends only on the choice of  $\xi(t)$ . We will look for  $\xi(t) = (\xi_1(t), \dots, \xi_d(t))$  which solve an ODE and thus we can parameterize the curves  $\xi(t)$  by their initial condition  $\xi(0) = x$ . It may seem odd (and unneeded) to introduce the final time  $T$ . This is done so that  $f_-(T, x) = f_-(T, \xi(0))$  and to keep the analogy close to what is traditionally done in SDEs. Differentiating  $f_-(\xi(t), T - t)$  with respect to  $t$ , we see that maintaining constant amounts

$$\sum_{i=1}^d \frac{\partial f_-}{\partial x_i}(\xi(t), T - t) \frac{d\xi_i}{dt}(t) = \frac{\partial f_-}{\partial t}(\xi(t), T - t) = \sum_{i=1}^d b_i(\xi(t)) \frac{\partial f_-}{\partial x_i}(\xi(t), T - t)$$

where the last equality follows from (7.13). We conclude that for this equality to hold in general we need

$$\frac{d\xi}{dt} = b(\xi(t)) \quad \text{and} \quad \xi(0) = x.$$

Since  $f_-(\xi(t), T - t)$  is a constant we have

$$f_-(\xi(0), T) = f_-(\xi(T), 0) \implies f_-(x, T) = f(\xi(T)) \tag{7.14}$$

which provides a solution to (7.13) to all points which can be reached by curves  $\xi(T)$ . Under mild assumptions this is all of  $\mathbb{R}^d$ .

Looking back at Theorem 4.3, we notice that differently from the *ode* case we did *not* find a SDE  $X_t$  which keeps  $f_-(X_t, T - t)$  constant in the fully fledged sense. However, we have obtained something very close to it: We chose  $t \mapsto f_-(X_t, T - t)$  to be a martingale, *i.e.*, a process that is constant *on average!* This is the content of Theorem 4.3 and Corollary 4.4 (putting the accent on the *expectation* part of the result), which mimics the result of (7.14), only with the addition of expected values. Hence we might be provoked to make the following fanciful statement.

*Stochastic differential equations are the method of characteristics for diffusions. Rather than follow a single characteristic back to the initial value to find the current value, we trace an infinite collection of stochastic curves each back to its own initial value which we then average weighting with the probability of the curve.*

## 6. A fundamental example: Brownian motion and the Heat Equation

We now consider the simple but fundamental case of standard brownian motion.

Let us consider a compact subset  $D \subset \mathbb{R}^2$  with a smooth boundary  $\partial D$  and a  $f(x)$  defined on  $\partial D$ .

**The Dirichlet problem:** We are looking for a function  $u(x)$  such that

$$\Delta u = \frac{\partial^2 u}{\partial^2 y_1} + \frac{\partial^2 u}{\partial^2 y_2} = 0 \text{ for } y = (y_1, y_2) \text{ inside } D.$$

$$\lim_{y \rightarrow x} u(y) = f(x) \text{ for all } x \in \partial D.$$

Let  $B(t, \omega) = (B_1(t, \omega), B_2(t, \omega))$  be a two dimensional Brownian Motion. Define the stopping time

$$\tau = \inf\{t > 0 : B(t) \notin D\}$$

Let  $\mathbb{E}_y$  be the expectation with respect to the Wiener measure for a Brownian Motion starting from  $y$  at time  $t = 0$ . Let us define  $\phi(x) = \mathbb{E}_y f(B(\tau))$ . We are going to show that  $\phi(x)$  solves the Dirichlet problem.

LEMMA 6.1. *With probability 1,  $\tau < \infty$ . In fact,  $\mathbb{E}\tau^r < \infty$  for all  $r > 0$ .*

PROOF.

$$\begin{aligned} \mathbb{P}\{\tau \geq n\} &\leq \mathbb{P}\{|B(1) - B(0)| \leq \text{diam}D, |B(2) - B(1)| \leq \text{diam}D, \dots, |B(n) - B(n-1)| \leq \text{diam}D\} \\ &\leq \prod_{k=1}^n \mathbb{P}\{|B(k) - B(k-1)| \leq \text{diam}D\} = \alpha^n \text{ where } \alpha \in (0, 1) \end{aligned}$$

Hence  $\sum_{n=1}^{\infty} \mathbb{P}\{\tau \geq n\} < \infty$  and the Borel-Cantelli lemma says that  $\tau$  is almost surely finite. Now lets look at the moments.

$$\mathbb{E}\tau^r = \int x^r \mathbb{P}\{\tau \in dx\} \leq \sum_{n=1}^{\infty} n^r \mathbb{P}\{\tau \in (n-1, n]\} \leq \sum_{n=1}^{\infty} n^r \mathbb{P}\{\tau \geq n-1\} \leq \sum_{n=1}^{\infty} n^r \alpha^n < \infty$$

□

Lets fix a point  $y$  in the interior of  $D$ . Lets put a small circle of radius  $\rho$  around  $y$  so that the circle is contained completely in  $D$ . Let  $\tau_{\rho, y}$  be the first moment of time  $B(t)$  hits the circle of radius  $\rho$  centered at  $y$ .

Because the law of Brownian Motion is invariant under rotations, we see that  $B(\tau_{\rho, y})$  is distributed uniformly on the circle centered at  $y$ . (Lets call this circle  $S_{\rho}(y)$ .)

THEOREM 6.2.  *$\phi(x)$  solves the Laplace equation.*

PROOF. i) We start by proving the mean value property. To do so we invoke the Strong Markov property of  $B(t)$ . Let  $\tau_S = \inf\{t : B(t) \in S_{\rho}(y)\}$  and  $z_{\vartheta} = (\rho \cos \vartheta, \rho \sin \vartheta)$  be the point on  $S_{\rho}(y)$  at angle  $\vartheta$ . We notice that any path from  $y$  to the boundary of  $D$  must pass through  $S_{\rho}(y)$ . Thus we can think of  $\phi(y)$  as the weighted average of  $\mathbb{E}\{f(B(\tau)) | B(\tau_S) = z_{\vartheta}\}$  where  $\vartheta$  moves around the circle  $S_{\rho}(y)$ . Each entry in this average is weighted by the chance of hitting that point on the sphere starting from  $y$ . Since this chance is uniform (all points are equally likely), we simply get the factor of  $\frac{1}{2\pi}$  to normalize things to be a probability measure.

$$\phi(y) = \frac{1}{2\pi} \int_0^{2\pi} d\vartheta \mathbb{E}\{f(B(\tau)) | B(\tau_S) = z_{\vartheta}\} = \frac{1}{2\pi} \int_0^{2\pi} d\vartheta \phi(z_{\vartheta}) \quad (7.15)$$

ii)  $\phi(x)$  is infinitely differentiable. This can be easily shown but let us just assume it since we are doing this exercise in explicit calculation to improve our understanding, not to prove every detail of the theorems.

iii) Now we see that  $\phi$  satisfies  $\Delta\phi = \frac{\partial^2\phi}{\partial y_1^2} + \frac{\partial^2\phi}{\partial y_2^2} = 0$ . We expand about a point  $y$  in the interior of  $D$ .

$$\begin{aligned} \phi(z) = & \phi(y) + \frac{\partial\phi}{\partial y_1}(z_1 - y_1) + \frac{\partial\phi}{\partial y_2}(z_2 - y_2) \\ & + \frac{1}{2} \left[ \frac{\partial^2\phi}{\partial y_1^2}(z_1 - y_1)^2 + \frac{\partial^2\phi}{\partial y_2^2}(z_2 - y_2)^2 + \frac{\partial^2\phi}{\partial y_1\partial y_2}(z_1 - y_1)(z_2 - y_2) \right] + O(|z - y|^3) \end{aligned}$$

Now we integrate this over a circle  $S_\rho(y)$  centered at  $y$  of radius  $\rho$ . We take  $\rho$  to be sufficiently small so that the entire disk is in the domain  $D$ . By direct calculation we have

$$\int_{S_\rho(y)} (z_1 - y_1) dz = 0, \quad \int_{S_\rho(y)} (z_2 - y_2) dz = 0, \quad \int_{S_\rho(y)} (z_1 - y_1)(z_2 - y_2) dz = 0$$

and

$$\int_{S_\rho(y)} (z_1 - y_1)^2 dz = (\text{const})\rho^2, \quad \int_{S_\rho(y)} (z_2 - y_2)^2 dz = (\text{const})\rho^2.$$

Since by the mean value property,

$$\phi(y) = (\text{const}) \int_{S_\rho(y)} \phi(z) dz$$

we see that

$$0 = (\text{const})\rho^2 \left( \frac{\partial^2\phi}{\partial y_1^2} + \frac{\partial^2\phi}{\partial y_2^2} \right) + O(\rho^3).$$

And thus,

$$\Delta\phi = \frac{\partial^2\phi}{\partial y_1^2} + \frac{\partial^2\phi}{\partial y_2^2} = 0$$

□





## Martingales and Localization

This chapter is dedicated to a more in-depth study of martingales and their properties. Some of the results exposed here are fairly general, and their proof in full generality required tools that are more advanced than the ones we have at our disposal. For this reason, some of the proofs will be given in a simplified setting/under stronger assumptions together with a reference for the more general result.

### 1. Martingales & Co.

We recall the definition of a martingale given at the beginning of the course, extending it slightly.

DEFINITION 1.1.  $\{X_t\}$  is a Martingale with respect to a filtration  $\mathcal{F}_t$  ( $\mathcal{F}_t$ -martingale for short) if for all  $t > s$  we have

- i)  $X_t$  is  $\mathcal{F}_t$ -measurable,
- ii)  $\mathbb{E}[|X_t|] < \infty$ ,
- iii)  $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$  a.s..

Similarly,  $X_t$  is a  $\mathcal{F}_t$ -supermartingale [ $\mathcal{F}_t$ -submartingale] if it satisfies conditions i) and ii) above, and

$$\mathbb{E}[X_t|\mathcal{F}_s] \leq X_s, \quad [\mathbb{E}[X_t|\mathcal{F}_s] \geq X_s] \quad \text{a.s..}$$

When the filtration is clear from the context we simply say that a process is a [super/sub-]martingale.

Super- and Submartingales extend the idea of a process that is constant in expectation to processes that are, respectively, nonincreasing and nondecreasing in expectation. It is clear that a martingale is both a supermartingale and a submartingale, while a supermartingale that is also a submartingale is a martingale.

PROPOSITION 1.2. A supermartingale [submartingale]  $M_t$  is a martingale on  $[0, T]$  if and only if  $\mathbb{E}[M_T] = \mathbb{E}[M_0]$ .

PROOF. xx □

REMARK 1.3. By Jensen's inequality on conditional expectations we have for any convex function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , a martingale  $M_t$  satisfies

$$\mathbb{E}[g(M_t)|\mathcal{F}_t] \geq g(\mathbb{E}[M_t|\mathcal{F}_s]) = g(M_s),$$

so application of a convex [concave] map to a martingale makes it a submartingale [supermartingale].

Recall that a random variable  $X$  is [square-]integrable if  $\mathbb{E}[|X|] < \infty$  [ $\mathbb{E}[X^2] < \infty$ ]. The condition of simple integrability of a random variable  $X$  can be equivalently stated as the condition

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X|\mathbb{1}_{|X|>n}] = 0. \tag{8.1}$$

Indeed, on one hand as  $\lim_{n \rightarrow \infty} |X|\mathbb{1}_{|X|>n} = 0$  a.s. and  $\mathbb{E}[|X|\mathbb{1}_{|X|>n}] \leq \mathbb{E}[|X|] < \infty$  we have by the dominated convergence theorem<sup>1</sup> that  $\lim_{n \rightarrow \infty} \mathbb{E}[|X|\mathbb{1}_{|X|>n}] = \mathbb{E}[0] = 0$ . On the other hand,

<sup>1</sup>See Theorem 0.1 in the appendix for a reminder of this theorem

we have that

$$\mathbb{E}[|X|] = \mathbb{E}[|X|\mathbb{1}_{|X|\leq n}] + \mathbb{E}[|X|\mathbb{1}_{|X|>n}] \leq n + \mathbb{E}[|X|\mathbb{1}_{|X|>n}], \quad (8.2)$$

and by using that both summands on the right hand side are bounded (the first by definition and the second by assumption) we obtain that  $\mathbb{E}[|X|] < \infty$ .

We now generalize the definitions above to stochastic processes.

DEFINITION 1.4. A stochastic process  $X_t$  on  $\mathcal{T} = [0, T]$  (where possibly  $T = \infty$ ) is

- i) integrable if  $\sup_{t \in \mathcal{T}} \mathbb{E}[|X_t|] < \infty$ ,
- ii) square integrable if  $\sup_{t \in \mathcal{T}} \mathbb{E}[X_t^2] < \infty$  (i.e., the second moments are uniformly bounded),
- iii) uniformly integrable if  $\lim_{n \rightarrow \infty} \sup_{t \in \mathcal{T}} \mathbb{E}[|X_t|\mathbb{1}_{|X_t|>n}] = 0$

The introduction of (8.1) allows to separate the concepts of simple and uniform integrability for stochastic processes as in the latter definition the limit is taken *after* the supremum. As one would expect, uniform integrability is stronger than simple integrability: similarly to (8.2) one has

$$\sup_{t \in \mathcal{T}} \mathbb{E}[|X_t|] = \sup_{t \in \mathcal{T}} \mathbb{E}[|X_t|\mathbb{1}_{|X_t|\leq n}] + \sup_{t \in \mathcal{T}} \mathbb{E}[|X_t|\mathbb{1}_{|X_t|>n}] \leq n + \sup_{t \in \mathcal{T}} \mathbb{E}[|X_t|\mathbb{1}_{|X_t|>n}] < \infty.$$

For the converse result we need stronger assumptions. We give below examples of such results:

PROPOSITION 1.5. A stochastic process  $\{X_t\}$  is uniformly integrable if, either

- i) It is dominated by a random variable  $Y$  defined on the same probability space, i.e.,  $X_t(\omega) \leq Y(\omega)$  such that  $\mathbb{E}[|Y|] < \infty$ ,
- ii) There exists some positive function  $G(x)$  on  $(0, \infty)$  with  $\lim_{x \rightarrow \infty} G(x)/x = \infty$  such that

$$\sup_{t \in \mathcal{T}} \mathbb{E}[G(|X_t|)] < \infty.$$

PROOF. We only prove the first result, for which we have that

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathcal{T}} \mathbb{E}[|X_t|\mathbb{1}_{|X_t|>n}] \leq \lim_{n \rightarrow \infty} \mathbb{E}[|Y|\mathbb{1}_{|Y|>n}] < \infty.$$

For the proof of the second result we refer *e.g.*, to [15]. □

In ii) of the above theorem we see that we need something *slightly* better than simple integrability to have uniform integrability. In particular we see that  $G(x) = x^{1+\varepsilon}$  for any  $\varepsilon > 0$  satisfies condition ii) of the above theorem. In particular, all square integrable martingales are uniformly integrable.

THEOREM 1.6. Let  $Y$  be an integrable random variable on a filtered probability space  $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$ , then

$$M_t := \mathbb{E}[Y|\mathcal{F}_t] \quad (8.3)$$

is a uniformly integrable martingale.

PROOF. xx □

We define a martingale such as the one in (8.3) as *closed* by the random variable  $Y$ . In particular, for any finite time interval  $[0, T]$  by definition every martingale is closed by its value at  $T$  since  $\mathbb{E}[M_T|\mathcal{F}_t] = M_t$  and we have the following corollary.

COROLLARY 1.7. Any martingale  $M_t$  on a finite time interval is uniformly integrable.

The above results can be extended to infinite time intervals.

THEOREM 1.8 (Martingale convergence theorem). Let  $M_t$  on  $\mathcal{T} = [0, \infty)$  be an integrable [sub/super]-martingale. Then there exists an almost sure (i.e., pointwise) limit  $\lim_{t \rightarrow \infty} M_t = Y$ , and  $Y$  is an integrable random variable.

The above theorem does *not* establish a correspondence between the random variables in terms of expected values. In particular, we may have cases where the theorem above applies but we have  $\lim_{t \rightarrow \infty} \mathbb{E}[M_t] \neq \mathbb{E}[Y]$ :

EXAMPLE 1.9. Consider the martingale  $M_t = \exp[B_t - t/2]$ . Because it is positive, we have that

$$\sup_{t \in \mathcal{T}} \mathbb{E}[|M_t|] = \sup_{t \in \mathcal{T}} \mathbb{E}[M_t] = \mathbb{E}[M_0] = 1 < \infty,$$

so it converges almost surely to a random variable  $Y$  by Theorem 1.8. However, we see that by the law of large numbers for Brownian Motion  $B_t/t \rightarrow 0$  a.s. and therefore

$$\mathbb{E}[Y] = \mathbb{E}\left[\lim_{t \rightarrow \infty} M_t\right] = \mathbb{E}\left[\lim_{t \rightarrow \infty} e^{t(B_t/t - 1/2)}\right] = 0,$$

which differs from  $\lim_{t \rightarrow \infty} \mathbb{E}[M_t] = 1$ .

The above observation means in particular that the conditions of Theorem 1.8 do not guarantee convergence in the  $L^1$  norm. Under the stronger condition of *uniform* integrability of the process  $X_t$  one obtains the same result with convergence in  $L^1$  norm and consequently the closedness of the martingale:

THEOREM 1.10. Let  $M_t$  be a uniformly integrable martingale on  $\mathcal{T} = [0, \infty)$ , then it converges as  $t \rightarrow \infty$  in  $L^1$  and a.s. to a random variable  $Y$ . Conversely, if  $M_t$  converges in  $L^1$  to an integrable random variable  $Y$  then it is square integrable and converges almost surely. In both cases  $M_t$  is closed by  $Y$ .

## 2. Optional stopping

After studying martingales *per se*, we consider their relation with stopping times. In particular, we will see that martingales behave nicely with respect to stopping times. To be more explicit, given a stochastic process  $X_t$  and recalling the definition Def. 7.14 of a stopping time  $\tau$ , we denote by  $\tau \wedge t = \min(\tau, t)$  and define the *stopped process*

$$X_t^\tau := X_{\tau \wedge t} = \begin{cases} X_t & \text{if } t < \tau \\ X_\tau & \text{else} \end{cases}.$$

The following theorem gives an example of the nice relationship between martingales and stopping times: it says that the martingale property is maintained by a process when such process is stopped.

THEOREM 2.1. For a  $\mathcal{F}_t$ -martingale  $M_t$  and any stopping time  $\tau$ , the process  $M_{\tau \wedge t}$  is a  $\mathcal{F}_t$ -martingale (and therefore a  $\mathcal{F}_{\tau \wedge t}$ -martingale), so

$$\mathbb{E}[M_{\tau \wedge t}] = \mathbb{E}[M_0] \quad \text{for all } t > 0. \tag{8.4}$$

PROOF. xx □

Martingales are often thought of as *fair games* because of their property of conserving their expected value: It is impossible, on average, to make positive gains by playing such game. Under this interpretation, Theorem 2.1 states that even if a player is given the possibility of quitting the game use any betting strategy, he/she will not be able to make net gains at time  $t$  provided that his/her strategy only depends on past information (cfr. Def. 7.14 of stopping time). However, the above property is lost if the player is patient enough, as the following example shows:

EXAMPLE 2.2. Let  $B_t$  be a Standard Brownian Motion (a martingale, hence an example of a “fair game”: you can think of it as a continuous version of betting one dollar on every coin flip) and define  $\tau_1 := \inf\{t : B_t \geq 1\}$  (the strategy of stopping as soon as you have a net gain of 1\$). Then by definition we have that  $B_\tau = 1 \neq 0 = \mathbb{E}[B_0]$ .

A similar situation to the one described above holds when considering the “martingale” betting strategy of doubling your bet every time you lose a coin flip. This strategy leads to an almost sure net win of 1\$ if one is patient enough (and has enough money to bet). As the examples above shows, stopped martingales may lose the property of conserving the expected value in the limit  $t \rightarrow \infty$ . The following theorem gives sufficient conditions for the martingale property to hold in this limit, *i.e.*, for the expected value of a game to be conserved at a stopping time  $\tau$ :

**THEOREM 2.3** (Optional stopping theorem). *Let  $M_t$  be a martingale,  $\tau$  a stopping time, then we have  $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$  if either of the following conditions holds:*

- *The stopping time  $\tau$  is bounded a.s, *i.e.*,  $\exists K < \infty: \tau \leq K$ ,*
- *The martingale  $M_t$  is uniformly integrable,*
- *The stopping time is finite a.s. (*i.e.*,  $\mathbb{P}[\tau = \infty] = 0$ ),  $M_t$  is integrable and*

$$\lim_{t \rightarrow \infty} \mathbb{E}[M_t \mathbb{1}_{\tau > t}] = 0.$$

PROOF. □

Under the gaming interpretation of above, we see that a game is “fair”, *i.e.*, it is impossible to make net gains, on average, using only past information, if any of the conditions i)-iii) hold. In particular, in the case of coin-flip games (or casino games) we see that a winning strategy does not exist as condition ii) holds: there is only a finite amount of money in the world, so the martingale is uniformly bounded, and in particular uniformly integrable. A simplified example of such a situation is given next:

**EXAMPLE 2.4.** *Let  $B_t$  be a Standard Brownian Motion on the interval  $a < 0 < b$  and define the stopping time  $\tau = \tau_{ab} = \inf\{t \in [0, \infty) : B_t \notin (a, b)\}$ . The stopped process  $B_\tau \wedge t$  is uniformly bounded and in particular uniformly integrable. Hence, by Theorem 2.3 we have that  $\mathbb{E}[B_\tau] = \mathbb{E}[B_0] = 0$ . However, we also have that  $B_\tau = b$  with probability  $p$  and  $B_\tau = a$  with probability  $1 - p$ , therefore*

$$0 = \mathbb{E}[B_\tau] = a \cdot (1 - p) + b \cdot p \quad \Rightarrow \quad \mathbb{P}[B_\tau = b] = p = \frac{-a}{b - a},$$

*which we have concluded based on considerations based on the martingale properties of  $B_t$  and therefore extends to any martingale for which  $\tau_{ab}$  is finite a.s..*

We conclude the chapter by presenting the converse of Theorem 2.3:

**PROPOSITION 2.5.** *Let  $X_t$  be a stochastic process such that for any stopping time  $\tau$ ,  $X_\tau$  is integrable and  $\mathbb{E}[X_0] = \mathbb{E}[X_\tau]$ . Then  $X_t$  is a martingale.*

PROOF. □

### 3. Localization

This section is devoted to the use of stopping times for the study of the properties of stochastic processes. As we have seen, the stopped process may have some properties that the original process did not have (*e.g.*, uniform integrability on  $[0, \infty)$  in Example 2.4). One can generalize such situation to a *sequence* of stopping times, such as the following example:

**EXAMPLE 3.1.** *Consider, similarly to Example 2.4, a Standard Brownian Motion  $B_t$  on the interval  $(-n, n)$  for  $n \in \mathbb{N}$ . Then we can define the stopping times  $\tau_n := \inf\{t : B_t \notin (-n, n)\}$ . For each  $n > 0$ , the process is uniformly integrable.*

In the above example, by taking the limit  $n \rightarrow \infty$  one would approach the original setting of unbounded Brownian Motion by approximating it with uniformly bounded stopped processes. This process can be extremely useful to obtain stronger results as the ones obtained previously in the course, as we will see later in this section, and justifies the following definition:

**DEFINITION 3.2.** *A property of a stochastic process  $X_t$  is said to hold locally if there exists a sequence  $\{\tau_n\}$  of stopping times with the property  $\lim_{n \rightarrow \infty} \tau_n(\omega) = \infty$  a.s. such that the stopped process  $X_{\tau_n \wedge t}$  has such property. In this case, the sequence  $\{\tau_n\}$  is called the localizing sequence.*

A particularly useful example is the one of the martingale property:

**DEFINITION 3.3.** *An adapted process  $M_t$  is a local martingale if there exists a sequence of stopping times  $\{\tau_n\}$  such that  $\lim_{n \rightarrow \infty} \tau_n(\omega) = \infty$  a.s. and the stopped process  $M_{\tau_n \wedge t}$  is a martingale for all  $n$ .*

It is clear that if a property holds in the original sense, then it also holds locally: one just has to take  $\tau_n = n > t$ . On the contrary a local martingale is in general not a martingale:

**EXAMPLE 3.4.** *Consider the Itô integral  $\int_0^t \exp[B_s^2] dB_s$ , for  $t < 1/4$  and define  $\tau_n := \inf\{t > 0 : \exp[B_s^2] = n\}$ . The process  $M_{\tau_n \wedge t}$  is a martingale, since we can write it as*

$$M_{\tau_n \wedge t} = U_t = \int_0^t \exp[B_s^2] \mathbb{1}_{\exp[B_s^2] \leq n} dB_s$$

is square integrable by Itô isometry. However, we have that

$$\mathbb{E}[\exp[2B_t^2]] = \frac{1}{2\pi t} \int_{-\infty}^{\infty} e^{2x^2} e^{-x^2/(2t)} dx$$

which diverges for  $t > 1/4$ , implying that  $M_t$  is not integrable.

We now list some results that, besides allowing to practice the use of localization methods, give sufficient conditions for a local martingale to be a martingale.

**PROPOSITION 3.5.** *Let  $M_t$  be a local martingale such that  $|M_t| \leq Y$  for an integrable random variable  $Y$ , then  $M_t$  is a uniformly integrable martingale.*

**PROOF.**  $\square$

**PROPOSITION 3.6.** *A non-negative local martingale  $M_t$ , for  $t \in [0, T]$  is a supermartingale.*

**PROOF.** Let  $\{\tau_n\}$  be the localizing sequence of  $M_t$ . Then for any  $t$  we have that  $\lim_{n \rightarrow \infty} \tau_n \wedge t = t$  a.s. and therefore that  $\lim_{n \rightarrow \infty} M_{\tau_n \wedge t} = M_t$ . Consequently, by Fatou's lemma on conditional expectations we have

$$\mathbb{E}[M_t | \mathcal{F}_s] = \mathbb{E}\left[\liminf_{n \rightarrow \infty} M_{\tau_n \wedge t} | \mathcal{F}_s\right] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[M_{\tau_n \wedge t} | \mathcal{F}_s] = \liminf_{n \rightarrow \infty} M_{\tau_n \wedge s} = M_s \quad \text{a.s.},$$

where in the second equality we have used that the limit exists. In particular, we have that  $\mathbb{E}[M_t] \leq \mathbb{E}[M_0] < \infty$ .  $\square$

**COROLLARY 3.7.** *A non-negative local martingale  $M_t$  on  $\mathcal{T} = [0, T]$  for  $T < \infty$  is a martingale if and only if  $\mathbb{E}[M_T] = M_0$*

**PROOF.** This is a direct result of Proposition 1.2 and Proposition 3.6.  $\square$

**REMARK 3.8.** *As explained in [9] there exists a necessary and sufficient condition for a local martingale to be a martingale: for the local martingale to be of "Dirichlet Class", i.e., such that such that the collection of random variables*

$$\mathbf{X} = \{X_\tau : \tau \text{ is a finite stopping time}\}$$

is uniformly integrable, i.e.,  $\sup_{X \in \mathbf{X}} \lim_{n \rightarrow \infty} \mathbb{E}[|X| \mathbb{1}_{|X| > n}] = 0$ .

We now give some slightly more advanced examples of the use of localization procedure. We begin by revisiting the problem of proving moment bounds for Itô integrals.

**Moment Bounds for Itô Integrals.** We let  $I_t = \int_0^t \sigma_s dB_s$ . We want to prove the moment bounds

$$\mathbb{E}|I_t|^{2p} \leq (2p-1)(2p-3)\cdots 3 \cdot 1 \cdot M^{2p}t^p,$$

under the assumption that  $|\sigma_s| \leq M$  a.s.

The case  $p = 1$  follows from the Itô isometry. Therefore, we now proceed to prove the induction step. Let us assume the inequality for  $p - 1$  and use it to prove the inequality for  $p$ . For any  $N > 0$ , we define

$$\tau_N = \inf\{t \geq 0 : \int_0^t |I_s|^{4p-2} \sigma_s^2 ds \geq N\}$$

Applying Itô formula to  $x \mapsto |x|^{2p}$  and evaluating at the time  $t \wedge \tau_N$  produces

$$|I_{t \wedge \tau_N}|^{2p} = p(2p-1) \int_0^{t \wedge \tau_N} |I_s|^{2(p-1)} \sigma_s^2 ds + 2p \int_0^{t \wedge \tau_N} |I_s|^{2p-1} \sigma_s dB_s = (I) + (II)$$

now by the induction hypothesis

$$\begin{aligned} \mathbb{E}(I) &\leq p(2p-1) \int_0^t \mathbb{E}|I_s|^{2(p-1)} \sigma_s^2 ds \leq (2p-1)(2p-3)\cdots 3 \cdot 1 \cdot pM^{2p} \int_0^t s^{p-1} ds \\ &\leq (2p-1)(2p-3)\cdots 3 \cdot 1 \cdot M^{2p}(t \wedge \tau_N)^p \leq (2p-1)(2p-3)\cdots 3 \cdot 1 \cdot M^{2p}t^p \end{aligned}$$

If we define

$$U_t = \int_0^t |I_s|^{2p-1} \sigma_s \mathbf{1}_{s \leq \tau_N} dB_s$$

then  $U_t$  is a martingale since

$$\int_0^t |I_s|^{4p-2} |\sigma_s|^2 \mathbf{1}_{s \leq \tau_N} ds = \int_0^{t \wedge \tau_N} |I_s|^{4p-2} |\sigma_s|^2 ds \leq N.$$

Since  $t \wedge \tau_N$  is a bounded stopping time, the optional stopping lemma says that  $\mathbb{E}U_{t \wedge \tau_N} = 0$ . However as noted above

$$\mathbb{E}(II) = \mathbb{E}U_{t \wedge \tau_N}$$

so one obtains

$$\mathbb{E}|I_{t \wedge \tau_N}|^{2p} \leq (2p-1)(2p-3)\cdots 3 \cdot 1 \cdot M^{2p}t^p.$$

Since  $|I_s|$  is almost surely finite, we know that  $\tau_N$  is finite with probability one. Hence  $|I_{t \wedge \tau_N}|^{2p} \rightarrow |I_t|^{2p}$  almost sure. Then by Fatou's lemma we have

$$\mathbb{E}|I_t|^{2p} \leq \liminf_{N \rightarrow \infty} \mathbb{E}|I_{t \wedge \tau_N}|^{2p} \leq (2p-1)(2p-3)\cdots 3 \cdot 1 \cdot M^{2p}t^p$$

**SDEs with Superlinear Coefficients.** Let  $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma^{(i)}: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be such that for any  $R > 0$  there exists a  $C$  such that

$$\begin{aligned} |b(x) - b(y)| + \sum_{i=1}^m |\sigma^{(i)}(x) - \sigma^{(i)}(y)| &\leq C|x - y| \\ |b(x)| + |\sigma(x)| &\leq C \end{aligned}$$

for any  $x, y \in \mathcal{B}_0(R)$ , where  $\mathcal{B}_0(r) := \{x \in \mathbb{R}^d : \|x\|_2 < r\}$ .

Consider the SDE

$$dX_t = b(X_t) dt + \sum_{i=1}^m \sigma^{(i)}(X_t) dB_t^{(i)} \quad (8.5)$$

For any  $R$  let  $b_R$  and  $\sigma_R^{(i)}$  are functions such that  $b_R(x) = b(x)$  and  $\sigma_R(x) = \sigma(x)$  in  $\mathcal{B}_0(R)$ , they are globally bounded and globally Lipschitz in  $\mathbb{R}^d$ .

Since  $b_R$  and  $\sigma_R$  satisfy the existence and uniqueness assumptions of Chapter 2.3, there exists a solution  $X_t^{(R)}$  to the equation

$$dX_t^{(R)} = b_R(X_t^{(R)}) dt + \sum_{i=1}^m \sigma_R^{(i)}(X_t^{(R)}) dB_t^{(i)} \quad (8.6)$$

For any  $N > 0$  and  $R > 0$  we define the stopping time

$$\tau_R := \inf\{t \geq 0 : |X_t^{(R)}| > R\}$$

**THEOREM 3.9.** *If*

$$\mathbb{P} \left[ \lim_{R \rightarrow \infty} \tau_R = \infty \right] = 1$$

*there there exists a unique strong solution to (8.5).*

**PROOF.** Fix a  $T > 0$ . For  $R \in \mathbb{N}$  let  $\Omega_R = \{\tau_R < T < \tau_{R+1}\}$ . By the assumption

$$\mathbb{P} \left[ \bigcup_{R=1}^{\infty} \Omega_R \right] = 1.$$

Also notice that the  $\Omega_R$  are disjoint for  $R = 1, 2, \dots$  and we can define the process

$$X_t(\omega) = X_t^{(R)}(\omega) \text{ for } t \in [0, T] \text{ if } \omega \in \Omega_R.$$

Since  $\sup |X_t^{(R)}(\omega)| < R$ , we know that  $b(X_t^{(R)}(\omega)) = b^{(R)}(X_t^{(R)}(\omega))$  and  $\sigma(X_t^{(R)}(\omega)) = \sigma^{(R)}(X_t^{(R)}(\omega))$  for all  $t \in [0, T]$ . Hence  $X_t$  as defined solves the original equation. Uniqueness comes from the fact that solutions to (8.6) are unique.  $\square$

#### 4. Quadratic variation for martingales

Recall the definition of quadratic variation of a stochastic process:

**DEFINITION 4.1.** *The quadratic variation of an adapted stochastic process  $X_t$  is defined as*

$$[X]_t := \lim_{N \rightarrow \infty}^p \sum_{j=0}^{j^N} \left( X_{t_{j+1}^N} - X_{t_j^N} \right)^2$$

where  $\lim^p$  denotes a limit in probability and  $\{t_j^N\}$  is a set partitioning the interval  $[0, t]$  defined by

$$\Gamma^N := \{ \{t_j^N\} : 0 = t_0^N < t_1^N < \dots < t_{j^N}^N = t \} \quad (8.7)$$

with  $|\Gamma^N| := \sup_j |t_{j+1}^N - t_j^N| \rightarrow 0$  as  $N \rightarrow \infty$ .

The process defined above is a sum of positive contributions and is therefore nondecreasing in  $t$  a.s..

Now let  $M_t$  be a [local] martingale. In light of Remark 1.3 we know that  $M_t^2$  is a [local] submartingale. Hence, we would like to know if we can transform  $M_t^2$  back to a martingale, for example by subtracting a ‘‘compensation process’’ removing the nondecreasing part of the squared

process. It turns out that such process exists and is precisely the quadratic variation process. The intuition behind this result comes from the following computation: assume that  $s < t$ , then we have

$$\mathbb{E} [M_t M_s] = \mathbb{E} [M_s \mathbb{E} [M_t | \mathcal{F}_s]] = \mathbb{E} [M_s^2]$$

where in the second equality we have used the martingale property. As a consequence of this we can write

$$\mathbb{E} [(M_t - M_s)^2] = \mathbb{E} [M_t^2] - 2\mathbb{E} [M_t M_s] + \mathbb{E} [M_s^2] = \mathbb{E} [M_t^2] - \mathbb{E} [M_s^2]. \quad (8.8)$$

In particular this implies that the summands in the definition of quadratic variation can be expressed, on expectation, as differences of expectation values that cancel telescopically, leading to (part of) the following theorem.

**THEOREM 4.2.** *This theorem can be stated in the martingale and local martingale version:*

- i) *Let  $M_t$  be a square-integrable martingale, then the quadratic variation process  $[M]_t$  exists and  $M_t^2 - [M]_t$  is a martingale.*
- ii) *Let  $M_t$  be a local martingale, then the quadratic variation process  $[M]_t$  exists and  $M_t^2 - [M]_t$  is a local martingale.*

**PROOF.** We only prove point i) of the theorem above. Point ii) follows for locally square integrable martingales by localization, *i.e.*, by substituting  $t \rightarrow \tau_n \wedge t$  where  $\tau_n$  is the localizing sequence. Repeating the calculation leading to (8.8) with conditional expectations we obtain

$$\mathbb{E} [M_t^2 - M_s^2 | \mathcal{F}_s] = \mathbb{E} \left[ \sum_{j=1}^{j^N} (M_{t_j^N} - M_{t_{j-1}^N})^2 | \mathcal{F}_{t_{j-1}^N} \right].$$

Now, taking the limit in probability of the right hand side (we do not prove that such limit exists here, but we refer to [15]) and rearranging we obtain that  $\mathbb{E} [M_t^2 - [M]_t | \mathcal{F}_s] = M_s^2 - [M]_s$  as desired.  $\square$

We conclude this section by proving a surprising result about martingales with finite first variation.

**LEMMA 4.3.** *Let  $M_t$  be a continuous local martingale with finite first variation. Then  $M_t$  is almost surely constant.*

The intuition behind the above result is quite simple: considering a continuum time interval, constraining a continuous martingale on behaving “nicely” in order to have finite first variation (for example monotonically or in a differentiable way) the martingale would somehow have to be “consistent with its trend at  $t_-$ ” (except of course in a set of measure 0) and could therefore not respect the constant conditional expectation property. In other words, martingales with finite first variation are too “stiff” to be different from the identity function.

**PROOF OF LEMMA 4.3.** We assume for this proof that  $M_t$  is a [locally] bounded martingale. We will eventually show that the variance of  $M_t$  is zero and hence  $M_t$  is constant. Picking some partition of time  $0 = t_0 < t_1 < \dots < t_k = t$ , recalling (8.8) we consider the variance at time  $t$

$$\begin{aligned} \mathbb{E} M_t^2 &= \mathbb{E} \sum (M_{t_n}^2 - M_{t_{n-1}}^2) = \mathbb{E} \sum (M_{t_n} - M_{t_{n-1}})^2 \\ &\leq \mathbb{E} \sup_{t_n} |M_{t_n} - M_{t_{n-1}}| \sum |M_{t_n} - M_{t_{n-1}}| \end{aligned}$$

Since the first variation  $V(t) = \lim_{\Delta T \rightarrow 0} \sum |M_{t_n} - M_{t_{n-1}}|$  was assumed to be finite we obtain

$$\mathbb{E} M_t^2 \leq (\text{const}) \mathbb{E} \lim_{\Delta T \rightarrow 0} \sup_{t_n} |M_{t_n} - M_{t_{n-1}}|$$



this limit is zero because  $M$  was assumed to be continuous.

Hence the variance of  $M_t$  is zero and thus  $M_t$  is constant almost surely. Thus  $M_t$  is constant for any countable collection of times. Use the rational numbers and then continuity to conclude that it is constant and the same constant for all times.  $\square$

## 5. Lévy-Doob Characterization of Brownian-Motion

THEOREM 5.1 (Lévy-Doob). *If  $X(t)$  is a continuous martingale such that*

- i)  $X(0) = 0$ ,
- ii)  $X(t)$  is a square integrable-martingale with respect to the filtration it generates,
- iii)  $X(t)^2 - t$  is a square integrable-martingale with respect to the filtration it generates

then  $X(t)$  is a standard Brownian motion.

It is important the  $X(t)$  be continuous. For example if  $N_t$  is a jump process  $N_t - t$  and  $(N_t - t)^2 - t$  are both martingales but  $N_t - t$  is quite different from browninan motion.

PROOF. Our proof essentially follows that of Doob found in [2], which approaches the problem as a central limit theorem, proved through a clever computational trick. Fix a positive integer  $N$  and an  $\varepsilon > 0$ . Define

$$\tau(\varepsilon, N) = \inf\{s > 0 : \sup_{\substack{s_1 < s_2 < s \\ |s_1 - s_2| < 1/N}} |X(s_1) - X(s_2)| = \varepsilon\} .$$

If there is no such time  $s$ , set  $\tau = \infty$ . Fix a time  $t$ . We what to show that the random variable  $X(t)$  has the same Gaussian distribution as  $B_t$ . To do this it is enough to show that  $\mathbb{E}e^{i\alpha X(t)} = e^{-\alpha^2 t/2}$ , that is show that they both have the same characteristic functions (Fourier transform). It is a standard result in basic probability that if a sequence of random variables have characteristic functions which converge for each  $\alpha$  to  $e^{-\alpha^2 t/2}$  then the sequence of random variables has a limit and it is Gaussian. See [1] for a nice discussion of characteristic functions and convergence of probability measures. Hence, we will show that

$$\mathbb{E} \left\{ e^{i\alpha X(t \wedge \tau)} \right\} \rightarrow e^{-\frac{\alpha^2 t}{2}} + O(\varepsilon) \text{ as } N \rightarrow \infty \text{ for any } \varepsilon > 0.$$

Since  $\varepsilon$  will be arbitrary and the lefthand side is independent of  $\varepsilon$ , this will imply the result.

Partition the interval  $[0, t]$  with point  $t_k = \frac{kt}{N}$ . Set

$$I = \left| \mathbb{E} \left\{ \prod_{j=1}^N e^{i\alpha(X_j - X_{j-1})} - \prod_{j=1}^N e^{-\frac{\alpha^2}{2}(t_j - t_{j-1})} \right\} \right|$$

where  $X_j := X(t_j \wedge \tau)$ . In general, observe that the following identity holds

$$\begin{aligned} A_1 A_2 A_3 \cdots A_N - B_1 B_2 \cdots B_N &= A_1 A_2 \cdots A_{N-1} (A_N - B_N) \\ &\quad + A_1 A_2 \cdots A_{N-2} (A_{N-1} - B_{N-1}) B_N \\ &\quad + A_1 A_2 \cdots A_{N-3} (A_{N-2} - B_{N-2}) B_{N-1} B_N \\ &\quad \vdots \\ &\quad + (A_1 - B_1) B_2 B_3 \cdots B_N . \end{aligned}$$

Hence

$$I = \left| \mathbb{E} \left\{ \sum_{k=1}^N \prod_{j=1}^{N-k-1} e^{i\alpha(X_j - X_{j-1})} \left( e^{i\alpha(X_{N-k} - X_{N-k+1})} - e^{-\frac{\alpha^2}{2}(t_{N-k} - t_{N-k+1})} \right) \prod_{j=N-k+1}^N e^{-\frac{\alpha^2}{2}(t_j - t_{j-1})} \right\} \right|$$

All of the terms in the first product have modulus one and all of the terms in the second product are less than one. Hence

$$I \leq \mathbb{E} \left\{ \sum_{k=1}^N \left| \mathbb{E} \left\{ e^{i\alpha(X_{N-k} - X_{N-k+1})} - e^{-\frac{\alpha^2}{2}(t_{N-k} - t_{N-k+1})} \middle| \mathcal{F}_{t_{N-k} \wedge \tau} \right\} \right| \right\}$$

Now observe that by Taylor's theorem

$$e^{i\alpha\Delta_k X} - e^{-\frac{\alpha^2}{2}\Delta_k t} = i\alpha\Delta_k X - \frac{\alpha^2}{2}(\Delta_k X)^2 + \frac{\alpha^2}{2}\Delta_k t + O(\Delta_k X)^3 + O(\Delta_k t)^2$$

where  $\Delta_k X = X_k - X_{k-1}$  and  $\Delta_k t = t_k - t_{k-1}$ . The constants implied by  $O(\Delta_k X)^3$  and  $O(\Delta_k t)^2$  can be taken to be uniformly bounded for  $\varepsilon \in (0, \varepsilon_0]$  and  $N \in [N_0, \infty)$  for some  $\varepsilon_0 > 0$  and  $N_0 < \infty$ . Observe that by using the martingale assumptions on  $X$  and the optional stopping lemma, we have that

$$\begin{aligned} \mathbb{E}\{\Delta_{N-k} X | \mathcal{F}_{t_{N-k} \wedge \tau}\} &= 0 \\ \mathbb{E}\{(\Delta_{N-k} X)^2 | \mathcal{F}_{t_{N-k} \wedge \tau}\} &= \Delta_{N-k}(t \wedge \tau) \leq t_{N-k} - t_{N-k-1}. \end{aligned}$$

Here  $\Delta_k(t \wedge \tau) = t_k \wedge \tau - t_{k-1} \wedge \tau$ . By our definition of  $\tau$ ,  $|\Delta_{N-k} X| \leq \varepsilon$ . So we have

$$\mathbb{E}\{|\Delta_{N-k} X|^3 | \mathcal{F}_{t_{N-k} \wedge \tau}\} \leq \left( \sup_{k, \omega} |\Delta_{N-k} X| \right) \mathbb{E}\{(\Delta_{N-k} X)^2 | \mathcal{F}_{t_{N-k} \wedge \tau}\} \leq \varepsilon \mathbb{E}\{(\Delta_{N-k} X)^2 | \mathcal{F}_{t_{N-k} \wedge \tau}\}.$$

And thus,

$$\begin{aligned} I &\leq \mathbb{E} \left\{ \sum_{k=1}^N \left| \mathbb{E}\{i\alpha\Delta_k X | \mathcal{F}_{t_{N-k}}\} - \mathbb{E}\left\{ \frac{\alpha^2}{2}(\Delta_k X)^2 | \mathcal{F}_{t_{N-k}} \right\} + \frac{\alpha^2}{2}\Delta_k t + C(\Delta_k t)^2 + C\mathbb{E}\{(\Delta_k X)^3 | \mathcal{F}_{t_{N-k} \wedge \tau}\} \right| \right\} \\ &\leq \left[ -N \left( \frac{t}{N} \right) \frac{\alpha^2}{2} + N \left( \frac{t}{N} \right) \frac{\alpha^2}{2} + CN \left( \frac{t}{N} \right)^2 + C\varepsilon N \left( \frac{t}{N} \right) \right] = C \frac{t}{N} + \varepsilon Ct \end{aligned}$$

Observe that  $\tau \rightarrow \infty$  as  $N \rightarrow \infty$  for any fixed  $\varepsilon$ . Hence we have that

$$\left| \mathbb{E} \left\{ e^{i\alpha X(t)} \right\} - e^{-\frac{\alpha^2 t}{2}} \right| \leq \lim_{N \rightarrow \infty} I \leq \varepsilon Ct$$

Notice that the left hand side is independent of  $\varepsilon$ . Since  $C$  and  $t$  are fixed and  $\varepsilon$  was any arbitrary number in  $(0, \varepsilon_0]$ , we conclude that

$$\mathbb{E} \left\{ e^{i\alpha X(t)} \right\} = e^{-\frac{\alpha^2 t}{2}}$$

□

We now give a slightly different formulation of the Levy-Doob theorem. Let  $M_t$  be a continuous martingale. Then by Theorem 4.2 if  $[M]_t = t$  condition ii) of Theorem 5.1 is satisfied and we obtain the following result.

**THEOREM 5.2 (Levy-Doob theorem).** *If  $M_t$  is a continuous martingale with  $[M]_t = t$  and  $M_0 = 0$  then  $M_t$  is standard Brownian motion.*

## 6. Random time changes

Let  $M_t$  be a contiguous martingale with respect to the filtration  $\mathcal{F}_t$ . Since the quadratic variation map  $t \mapsto [M]_t$  is non-decreasing, we can define its left-inverse by

$$\tau_t = \inf\{s \geq 0 : [M]_s \geq t\} \quad (8.9)$$

and the limiting value

$$[M]_\infty = \lim_{t \rightarrow \infty} [M]_t$$

**THEOREM 6.1** (Dambis-Dubins-Schwartz). *If  $[M]_\infty > T$  then  $B_t = M_{\tau_t}$  is a Brownian motion on the interval  $[0, T]$  with respect to the filtration  $\mathcal{G}_t = \mathcal{F}_{\tau_t}$ . Conversely, there exists a standard Brownian motion  $B_t$  such that  $M_t = B_{[M]_t}$  for  $t \geq 0$ .*

**REMARK 6.2.** *Theorem 6.1 shows that any continuous martingale is just the time change of Brownian motion with  $[M]_t$  giving the rate at which randomness is injected into the system.*

**THEOREM 6.1.** Now  $[B]_t = [M]_{\tau_t} = t$  since  $\tau_t$  is the inverse of  $t \mapsto [M]_t$ . Hence  $B_t^2 - t$  is martingale. Since

$$\mathbb{E}(B_t | \mathcal{G}_s) = \mathbb{E}(M_{\tau_t} | \mathcal{F}_{\tau_s}) = M_{\tau_s} = B_s$$

we see that  $B_t$  is also a martingale. Hence the by the Levy-Doob Theorem (Theorem 5.2),  $B_t$  is a standard Brownian motion. The second theorem follows from the first. Since  $W_{[M]_t}$  has the same distribution as  $B_{[M]_t}$  we see that  $B_{[M]_t} = M_{\tau_{[M]_t}} = M_t$  since  $\tau_{[M]_t} = t$ .  $\square$

It is useful to explicitly see the above construction for an Itô process

$$dM_t = \sigma_t dB_t$$

where  $\sigma_t > 0$  for all  $t \geq 0$ . Then

$$\partial_t [M]_t = \sigma_t^2$$

Since  $\tau_t$ , define in (8.9), is the inverse of  $t \mapsto [M]_t$  we have that

$$\partial_t \tau_t = \frac{1}{\sigma_t^2}.$$

## 7. Time Change for an SDE

Consider the one dimensional SDE

$$dX_t = \sigma(X_t) dB_t$$

with  $\sigma(x) > 0$  and  $\int_0^\infty \sigma^2(X_s) ds = \infty$  almost surely. (A simple condition which ensures this is  $\sigma(x) \geq c > 0$  for all  $x$ .) Then as before

$$[X]_t = \int_0^t \sigma^2(X_s) ds \quad \text{and} \quad \tau_t = \int_0^t \frac{1}{\sigma^2(X_s)} ds = \inf\{s : [X]_s > t\}$$

Hence if we define  $M_t = X_{\tau_t}$  then  $[M]_t = [X]_{\tau_t} = t$  so  $M_t$  is a Brownian motion and

$$dM_t = \frac{1}{\sigma^2(X_{\tau_t})} dX_{\tau_t} = \frac{1}{\sigma(X_{\tau_t})} dB_{\tau_t} = \frac{1}{\sigma(M_t)} dB_{\tau_t}$$

Now consider the SDE

$$dX_t = \sigma^2(X_t) f(X_t) dt + \sigma(X_t) g(X_t) dB_t$$

where  $\sigma : \mathbb{R} \rightarrow (0, \infty)$ ,  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $g : \mathbb{R}^d \times \mathbb{R}^{m,d}$  and  $B_t$  is a  $m$ -dimensional Brownian motion.

Then with the same definitions as above define  $dM_t = \sigma(X_t) dB_t$  and  $B_t = M_{\tau_t}$ . Since  $[M]_t = \int_0^t \sigma^2(X_s) dB_s$  we see that  $B_t$  is a standard Brownian motion. Observe that  $d\tau_t = \sigma^{-2}(X_{\tau_t}) dt$  and

$$dB_t = \frac{1}{\sigma(X_{\tau_t})} dB_{\tau_t}$$

Defining  $Y_t = X_{\tau_t}$ , we have that

$$\begin{aligned} dY_t &= \frac{1}{\sigma^2(X_{\tau_t})} dX_{\tau_t} = \frac{1}{\sigma^2(X_{\tau_t})} (\sigma^2(X_{\tau_t}) f(X_{\tau_t}) dt + \sigma(X_{\tau_t}) g(X_{\tau_t}) dB_{\tau_t}) \\ &= f(Y_t) dt + g(Y_t) \frac{1}{\sigma(X_{\tau_t})} dB_{\tau_t} = f(Y_t) dt + g(Y_t) dB_t \end{aligned}$$

Hence  $Y_t$  solves the equation

$$dY_t = f(Y_t) dt + g(Y_t) dB_t$$

Note that it is only a weak solution since the Brownian motion  $B_t$  was constructed at the same time as the solution  $Y_t$ . A strong solution required that the Brownian motion be specified in advance.

REMARK 7.1. *The same argument shows that if*

$$dX_t = \sigma_t^2 f(X_t) dt + \sigma_t g(X_t) dB_t$$

for some positive, adapted stochastic process  $\sigma_t$ . If  $\tau_t = \int_0^t \sigma_s^{-2} ds$  and  $Y_t = X_{\tau_t}$  then

$$dY_t = f(Y_t) dt + g(Y_t) dB_t$$

for the standard Brownian motion  $B_t = M_{\tau_t}$  where  $M_t = \int_0^t \sigma_s^2 dB_s$ .

## 8. Martingale representation theorem

### 9. Martingale inequalities

THEOREM 9.1 (Doob's Martingale Inequality). *Let  $M_t$  be a continuous-time martingale with respect to the filtration  $\mathcal{F}_t$ . Then*

$$\mathbf{P}\left(\sup_{0 \leq t \leq T} |M_t| \geq \lambda\right) \leq \frac{\mathbb{E}[|M_T|^p]}{\lambda^p}$$

where  $\lambda \in \mathbf{R}^+$ , and  $p \geq 1$ .

We now prove a limited version of the very useful Burkholder-Davis-Gundy inequalities. For the complete versions see [4, 8, 14].

THEOREM 9.2 (Burkholders Inequality). *Let  $X_t = \int_0^t f_s(\omega) dB_s$  then for any  $p \geq 1$*

$$\mathbb{E} \sup_{0 \leq s \leq t} |X_s|^{2p} \leq C_p \mathbb{E} \langle X_t \rangle^p$$

where  $C_p$  is a constant independent of the process, depending only on  $p$ .

PROOF. Doob's  $L^p$  maximal inequality combined with Itô's formula implies that

$$\mathbb{E} \sup_{0 \leq s \leq t} |X_s|^{2p} \leq q^p \mathbb{E} |X_t|^{2p} \tag{8.10}$$

$$= q^p \mathbb{E} \left\{ \frac{2p(2p-1)}{2} \int_0^t |X_s|^{2(p-1)} f(s)^2 ds \right\} + q^p \mathbb{E} \left\{ 2p \int_0^t |X_s|^{2p-1} f(s) dB_s \right\} \tag{8.11}$$

where  $\frac{1}{p} + \frac{1}{q} = 1$ . Next we introducing the stopping time

$$\tau_N = \inf\{t \geq 0 : \int_0^t |X_s|^{4p-2} |f_s|^2 ds \geq N\}$$

Let  $I_N(t) = 2p \int_0^t |X_{s \wedge \tau_N}|^{2p-1} f_{s \wedge \tau_N} dB_s$ . Since the integrand is bounded by the construction of  $\tau_N$  we have that  $\mathbb{E}I_N(t) = 0$ . Notice that

$$\mathbb{E} \left\{ 2p \int_0^{t \wedge \tau_N} |X_s|^{2p-1} f(s) dB_s \right\} = \mathbb{E}I_N(t \wedge \tau_N) = 0$$

where the last equality follows from the Optional Stopping Theorem. Next observe that

$$\mathbb{E} \left\{ \int_0^t |X_s|^{2(p-1)} f(s)^2 ds \right\} \leq \mathbb{E} \left\{ \sup_{0 \leq s \leq t} |X_s|^{2(p-1)} \int_0^t f(s)^2 ds \right\} = \mathbb{E} \left\{ \sup_{0 \leq s \leq t} |X_s|^{2(p-1)} \langle X_t \rangle \right\}$$

then by Holders inequality we have that

$$\leq \mathbb{E} \left( \sup_{0 \leq s \leq t} |X_s|^{2p} \right)^{1-\frac{1}{p}} (\mathbb{E} \langle X_t \rangle^p)^{\frac{1}{p}}$$

Putting everything together produces

$$\mathbb{E} \sup_{0 \leq s \leq t \wedge \tau_N} |X_s|^{2p} \leq C \mathbb{E} \left( \sup_{0 \leq s \leq t \wedge \tau_N} |X_s|^{2p} \right)^{1-\frac{1}{p}} \left( \mathbb{E} \langle X_{t \wedge \tau_N} \rangle^p \right)^{\frac{1}{p}}$$

By the definition of the stopping time everything is finite, hence we can divide thought by the first term on the right to obtain

$$\mathbb{E} \left( \sup_{0 \leq s \leq t \wedge \tau_N} |X_s|^{2p} \right)^{\frac{1}{p}} \leq C \left( \mathbb{E} \langle X_{t \wedge \tau_N} \rangle^p \right)^{\frac{1}{p}}$$

raising everything to the power  $p$  and then removing the stopping time by taking the limit as  $N \rightarrow \infty$  gives the result.  $\square$



## Girsanov's Theorem

### 1. An illustrative example

We begin with a simple example. We will frame it in a rather formal way as this will make the analogies with later examples clearer. Let us consider the probability space  $(\omega, \mathbb{P}, \mathcal{F})$  where  $\Omega = \mathbb{R}$  and  $\mathbb{P}$  is the standard Gaussian with mean zero on variance one. (For completeness let  $\mathcal{F}$  be the Borel  $\sigma$ -algebra on  $\mathbb{R}$ .)

Now consider two random variable  $Z$  and  $\tilde{Z}$  on this probability space. As always, a real valued random variable is a function from  $\Omega$  into  $\mathbb{R}$ . Let us define

$$Z(\omega) = \omega \quad \text{and} \quad \tilde{Z}(\omega) = \omega + \mu$$

for some fixed constant  $\mu$ . Since  $\omega$  is drawn under  $\mathbb{P}$  with respect the  $N(0, 1)$  measure on  $\mathbb{R}$  we have that  $Z$  is also distributed  $N(0, 1)$  and  $\tilde{Z}$  is distributed  $N(\mu, 1)$ .

Now let us introduce the density function associated to  $\mathbb{P}$  as

$$\phi(\omega) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\omega^2}{2}\right)$$

Now we introduce the function

$$\rho_\mu(\omega) = \frac{\phi(\omega + \mu)}{\phi(\omega)} = \exp\left(-\omega\mu - \frac{\mu^2}{2}\right)$$

Since  $\rho_\mu$  is a function from  $\Omega$  to  $\mathbb{R}$  is can be viewed as a random variable and we have

$$\mathbb{E}_{\mathbb{P}}\rho_\mu = \int_{\omega} \rho_\mu(\omega)\mathbb{P}(d\omega) = \int_{-\infty}^{\infty} \rho_\mu(\omega)\phi(\omega)d\omega = \int_{-\infty}^{\infty} \phi(\omega - \mu)d\omega = 1$$

since  $\phi(\omega - \mu)$  is the density of a  $N(\mu, 1)$  random variable. Hence  $\rho_\mu$  is a  $L^1(\Omega, \mathbb{P})$  radnom variable. Hence we can define a new measure  $\mathbf{Q}$  on  $\Omega$  by

$$\mathbf{Q}(d\omega) = \rho_\mu(\omega)\mathbb{P}(d\omega).$$

This means that for any random variable  $X$  on  $\Omega$  we have that the expected vlaue with respect to the  $\mathbf{Q}$ , denoted by  $\mathbb{E}_{\mathbf{Q}}$  is define by

$$\mathbb{E}_{\mathbf{Q}}[X] = \mathbb{E}_{\mathbb{P}}[X\rho_\mu]$$

Furthermore observe that for any bounded  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\begin{aligned} \mathbb{E}_{\mathbf{Q}}f(\tilde{Z}) &= \mathbb{E}_{\mathbb{P}}[f(\tilde{Z})\rho_\mu] = \int_{-\infty}^{\infty} f(\tilde{Z}(\omega))\rho_\mu(\omega)\phi(\omega)d\omega = \int_{-\infty}^{\infty} f(\omega + \mu)\phi(\omega + \mu)d\omega \\ &= \int_{-\infty}^{\infty} f(\omega)\phi(\omega)d\omega = \mathbb{E}_{\mathbb{P}}f(Z) \end{aligned}$$

Which implies that

$$\begin{aligned} \text{Under the measure } \mathbf{Q} : & \quad \tilde{Z} \text{ is dsitributed } N(0, 1) \\ \text{Under the measure } \mathbb{P} : & \quad \tilde{Z} \text{ is dsitributed } N(\mu, 1) \end{aligned}$$

Now let's consider a higher dimensional version. Let  $\Omega = \mathbb{R}^n$  and let  $\mathbb{P}$  be  $n$ -dimensional Wiener measure with covariance  $\sigma^2 I$  where  $\sigma > 0$  and  $I$  is the  $n \times n$  dimensional covariance matrix. In analogy to before, we define for  $\omega = (\omega_1, \dots, \omega_n) \in \mathbb{R}^n$

$$\phi(\omega) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \omega_i^2\right)$$

and for  $\mu = (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$

$$\rho_\mu(\omega) = \frac{\phi(\omega + \mu)}{\phi(\omega)} = \exp\left(-\frac{1}{\sigma^2} \sum_{i=1}^n \omega_i \mu_i - \frac{1}{2\sigma^2} \sum_{i=1}^n \mu_i^2\right)$$

Then if we define the  $\mathbb{R}^n$  valued random variables  $Z(\omega) = (Z_1(\omega), \dots, Z_n(\omega))$  and  $\tilde{Z}(\omega) = (\tilde{Z}_1(\omega), \dots, \tilde{Z}_n(\omega)) = Z(\omega) + \mu$ . Then if we define  $\mathbf{Q}(d\omega) = \rho_\mu(\omega)\mathbb{P}(d\omega)$  then following the same reasoning as before that

Under the measure  $\mathbf{Q}$ :  $\tilde{Z}$  is distributed  $N(0, \sigma^2 I)$

Under the measure  $\mathbb{P}$ :  $\tilde{Z}$  is distributed  $N(\mu, \sigma^2 I)$

**EXAMPLE 1.1.** *Importance Sampling*

Let  $f: \mathbf{R} \rightarrow \mathbf{R}$ , and  $X$  be distributed  $N(0, 1)$ . We have:

$$\begin{aligned} \mathbb{E}f(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-\frac{(x-\mu)^2}{2}} e^{\frac{\mu^2}{2} - \mu x} dx \end{aligned}$$

for some  $\mu \in \mathbf{R}$ .

For  $n$  large and  $\{X_i\}_{i=1}^n$  iid  $N(0, 1)$ , we have:

$$\mathbb{E}[f(X)] \approx \frac{1}{n} \sum_{i=1}^n f(X_i)$$

Let  $Y$  be distributed  $N(\mu, 1)$ . We have:

$$\begin{aligned} \mathbb{E}[f(X)] &= \mathbb{E}\left[f(Y) e^{-\mu Y + \frac{\mu^2}{2}}\right] \\ &\approx \frac{1}{n} \sum_{i=1}^n f(Y_i) e^{-\mu Y_i + \frac{\mu^2}{2}} \end{aligned}$$

for  $\{Y_i\}_{i=1}^n$  iid  $N(\mu, 1)$ .

## 2. Tilted Brownian motion

Let  $X_t = B_t + \mu t$  where  $B_t$  is standard Brownian Motion,  $\mu \in \mathbf{R}$ . Let  $0 = t_0 < t_1 < \dots < t_n \leq T$ . Let  $f, g: \mathbf{R}^n \rightarrow \mathbf{R}$  such that  $f(x_1, x_2, \dots, x_n) = g(x_1, x_2 - x_1, \dots, x_n - x_{n-1})$ .

We have:



$$\begin{aligned} & \mathbf{E} [g(X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}})] \\ &= \int_{\Omega \times \dots \times \Omega} \left( \frac{g(x_1, x_2 - x_1, \dots, x_n - x_{n-1})}{(2\pi)^{\frac{n}{2}} t_1^{\frac{1}{2}} (t_2 - t_1)^{\frac{1}{2}} \dots (t_n - t_{n-1})^{\frac{1}{2}}} \right) \prod_{i=1}^n e^{-\frac{[(x_i - x_{i-1}) - \mu(t_i - t_{i-1})]^2}{2(t_i - t_{i-1})}} \mathbf{P}(X_{t_1} \in dx_1, \dots, X_{t_n} \in dx_n) \end{aligned}$$

Now:

$$\begin{aligned} \prod_{i=1}^n e^{-\frac{[(x_i - x_{i-1}) - \mu(t_i - t_{i-1})]^2}{2(t_i - t_{i-1})}} &= \prod_{i=1}^n e^{-\frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}} \prod_{i=1}^n e^{\mu(x_i - x_{i-1}) - \frac{1}{2}\mu^2(t_i - t_{i-1})} \\ &= \prod_{i=1}^n e^{-\frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}} e^{\mu x_n - \frac{1}{2}\mu^2 t_n} \end{aligned}$$

And so:

$$\begin{aligned} \mathbf{E} [f(X_{t_1}, \dots, X_{t_n})] &= \mathbf{E} \left[ f(B_{t_1}, \dots, B_{t_n}) e^{\mu B_{t_n} - \frac{1}{2}\mu^2 t_n} \right] \\ &= \mathbf{E} [f(B_{t_1}, \dots, B_{t_n}) M_t] \end{aligned}$$

where  $M_t$  is the exponential martingale  $e^{\mu B_t - \frac{1}{2}\mu^2 t}$ .

### 3. Shifting Wiener Measure a Simple Cameron Martin Theorem

xx

**3.1. Some calculations.** Let  $Z_1, \dots, Z_n$  be iid  $N(0,1)$  on  $(\Omega, \mathcal{F}, \mathbf{P})$ . Also let  $f : \mathbf{R}^n \rightarrow \mathbf{R}$ . Then:

$$\begin{aligned} \mathbb{E}_{\mathbf{P}} f(Z_1, \dots, Z_n) &= \int_{\Omega \times \dots \times \Omega} f(\zeta_1, \dots, \zeta_n) \mathbf{P}(Z_1 \in d\zeta_1, \dots, Z_n \in d\zeta_n) \\ &= \int_{\Omega \times \dots \times \Omega} f(\zeta_1, \dots, \zeta_n) \left( \frac{1}{2\pi^{\frac{n}{2}}} \right) e^{-\frac{1}{2} \sum_{i=1}^n \zeta_i^2} d\zeta_1 \dots d\zeta_n \end{aligned}$$

Now define a new measure  $\tilde{\mathbf{P}}$ :

$$d\tilde{\mathbf{P}}(d\omega) = Y d\mathbf{P}(d\omega)$$

$$\text{where } Y = e^{(\sum_{i=1}^n \mu_i Z_i(\omega) - \frac{1}{2} \sum_{i=1}^n \mu_i^2)}$$

$$\text{and } \mu = (\mu_1, \dots, \mu_n) \in \mathbf{R}^n$$

The expectation of  $f$  under  $\tilde{\mathbf{P}}$  is given by:

$$\begin{aligned}\mathbf{E}_{\tilde{\mathbf{P}}}f(Z_1, \dots, Z_n) &= \int_{\Omega \times \dots \times \Omega} f(\zeta_1, \dots, \zeta_n) e^{(\sum_{i=1}^n \mu_i \zeta_i - \frac{1}{2} \sum_{i=1}^n \mu_i^2)} \mathbb{P}(Z_1 \in d\zeta_1, \dots, Z_n \in d\zeta_n) \\ &= \int_{\Omega \times \dots \times \Omega} f(\zeta_1, \dots, \zeta_n) \left( \frac{1}{(2\pi)^{\frac{n}{2}}} \right) e^{-\frac{1}{2} \sum_{i=1}^n (\zeta_i - \mu_i)^2} d\zeta_1 \dots d\zeta_n\end{aligned}$$

Now let  $\tilde{Z} = Z - \mu$ . We see that:

$$\begin{aligned}\mathbf{E}_{\tilde{\mathbf{P}}}Z_i &= \mu_i, & \mathbf{E}_{\tilde{\mathbf{P}}}(Z_i - \mu_i)^2 &= 1 \\ \mathbf{E}_{\mathbf{P}}Z_i &= 0, & \mathbf{E}_{\mathbf{P}}(Z_i - \mu_i)^2 &= 1\end{aligned}$$

We also have:

$$\begin{aligned}\mathbf{E}_{\mathbf{P}}f(Z_1, \dots, Z_n) &= \mathbf{E}_{\tilde{\mathbf{P}}}f(\tilde{Z}_1, \dots, \tilde{Z}_n) \\ &= \mathbf{E}_{\mathbf{P}}Yf(\tilde{Z}_1, \dots, \tilde{Z}_n) \\ &= \mathbf{E}_{\mathbf{P}}Yf(Z_1 - \mu_1, \dots, Z_n - \mu_n)\end{aligned}$$

#### 4. Girsanov's Theorem for SDES

##### 4.1. Girsanov (1D).

DEFINITION 4.1. Given two measures  $\mu, \nu$ , we say  $\nu \ll \mu$  ( $\nu$  is absolutely continuous wrt  $\mu$ ) if:

$$\mu(A) = 0 \Rightarrow \nu(A) = 0 \text{ for all measurable sets } A.$$

LEMMA 4.2. Let  $\mu$  and  $\nu$  be probability measures on  $(\Omega, \mathcal{G})$  with  $d\nu(\omega) = f(\omega)d\mu(\omega)$  for some  $f \in L^1(\mu)$ . Let  $X$  be a random variable with:

$$\mathbf{E}_{\nu}|X| = \int |X(\omega)|d\nu(\omega) = \int |X(\omega)|f(\omega)d\mu(\omega)$$

If  $\mathcal{H} \subset \mathcal{G}$  is a  $\sigma$ -algebra, then:

$$\mathbf{E}_{\nu}[X|\mathcal{H}] \mathbf{E}_{\mu}[f|\mathcal{H}] = \mathbf{E}_{\mu}[fX|\mathcal{H}]$$

Let  $dY_t = a_t(\omega) + dB_t$  Define:  $M_t = e^{-\int_0^t a_s dB_s - \frac{1}{2} \int_0^t a_s^2 ds}$  ( $\star$ ) Assume  $\mathbf{E}|M_t| < \infty$  so that  $M_t$  is a martingale. ( $\star$ ) Holds if any of the following hold:

- i)  $\exists c(t)$  such that  $|a_s(\omega)| < c(t) \forall 0 < s \leq t$  a.s.
- ii) Novikov's Condition:

$$\mathbf{E}_{\mathbf{P}} \left[ e^{\frac{1}{2} \int_0^t a^2 ds} \right] < \infty$$

CLAIM 4.3. If  $d\mathbf{Q}(\omega) = M_T(\omega)d\mathbf{P}(\omega)$  for  $t \leq T$ , then  $Y_t$  is a SBM wrt  $\mathbf{Q}$ .

*Proof:* We want to show that  $Y_t$  is a SBM wrt  $\mathbf{Q}$ . We are going to show:

- i)  $Y_t$  is a martingale wrt  $\mathbf{Q}$
- ii)  $Y_t^2 - t$  is a martingale wrt  $\mathbf{Q}$

Assuming  $Y_0 = 0$ , Levy-Doob then implies that  $Y_t$  is a SBM wrt  $\mathbf{Q}$ . Let  $K_t = M_t Y_t = g(M_t, Y_t)$ , where  $g(x, y) = xy$ . Then,  $dK_t = M_t dY_t + Y_t dM_t + dY_t dM_t$

$$\begin{aligned} dM_t &= -M_t a_t dB_t \\ \therefore dK_t &= M_t(a_t dt + dB_t) - Y_t M_t a_t dB_t - M_t a_t dB_t^2 \\ &= M_t(a_t dt + dB_t) - Y_t M_t a_t dB_t - M_t a_t dt \\ &= M_t(1 - Y_t a_t) dB_t \end{aligned}$$

And so  $K_t$  is a martingale wrt  $\mathbb{P}$ . Now,

$$\begin{aligned} \mathbf{E}_{\mathbf{Q}} [Y_t | \mathcal{F}_s] &= \frac{\mathbf{E}_{\mathbf{P}} [M_t Y_t | \mathcal{F}_s]}{\mathbf{E}_{\mathbf{P}} [M_t | \mathcal{F}_s]} \\ &= \frac{K_s}{M_s} \\ &= Y_s \end{aligned}$$

And so  $Y_s$  is a martingale wrt  $\mathbf{Q}$ .

THEOREM 4.4. *Let:*

$$\begin{aligned} dX_t &= b(X_t) dt + \sigma(X_t) dB_t \\ dY_t &= [\gamma(t, \omega) + b(Y_t)] dt + \sigma(Y_t) dB_t \\ X(0) &= Y(0) = x \end{aligned}$$

Assume  $\exists u(t, \omega)$  with:

- i)  $\sigma(Y(t, \omega))u(t, \omega) = \gamma(t, \omega)$  a.s.
- ii)  $\mathbf{E} \left[ e^{\frac{1}{2} \int_0^T u_s^2 ds} \right] < \infty$

If  $M_T = e^{-\int_0^T u_s dB_s - \frac{1}{2} \int_0^T u_s^2 ds}$  and  $d\mathbf{Q} = M_T d\mathbf{P}$  With  $\mathbf{Q}$  the Law of  $Y_t$ ,  $\mathbf{P}$  the Law of  $X_t$ , and  $\hat{B}_t = \int_0^t u(s, \omega) ds + B_t$  Then  $Y_t$  solves:

$$dY_t = b(Y_t) dt + \sigma(Y_t) d\hat{B}_t$$

and  $\hat{B}_t$  is a BM wrt  $\mathbf{Q}$

Let  $X_t, Y_t$  be defined by:

$$\begin{aligned} dX_t &= a(X_t) dt + \sigma(X_t) dB_t \\ dY_t &= b(Y_t) dt + \sigma(Y_t) dB_t \\ X_0 &= Y_0 = x \end{aligned}$$

Define  $u(y)$  by:  $\sigma(y)u(y) = b(y) - a(y)$  and let:

$$M_t = e^{-\int_0^t u(Y_s) dB_s - \frac{1}{2} \int_0^t |u(Y_s)|^2 ds}$$

$d\mathbf{Q} = M_t d\mathbf{P}$

The  $\mathbf{Q}$  Law of  $Y_t$  is the same as the  $\mathbf{P}$  Law of  $X_t$ .

$$\hat{B}_t = B_t + \int_0^t u(Y_s) ds$$

Under  $\mathbf{Q}$ ,  $\hat{B}_t$  is a standard BM.

$$\begin{aligned} dY_t &= a(Y_t) dt + \sigma(Y_t) d\hat{B}_t \\ &= a(Y_t) dt + \sigma(Y_t) [u(Y_t) dt + dB_t] \\ &= a(Y_t) dt + b(Y_t) dt - a(Y_t) dt + \sigma(Y_t) dB_t \\ &= b(Y_t) dt + \sigma(Y_t) dB_t \end{aligned}$$

And so the Law of  $Y_t$  on  $C(0, T; \mathbf{R}^n)$  is equivalent to the Law of  $X_t$  on  $C(0, T; \mathbf{R}^n)$ .

EXAMPLE 4.5. *Brownian motion Tracking a Continuous Function:*

Let  $B_t$  be a BM starting at 0. Let  $g(t)$  be a continuous function with  $g(0) = 0$ .

We would like to establish the truth of the following statement:

$$\mathbf{P} \left( \sup_{0 \leq t \leq 1} |B_t - g(t)| < \varepsilon \right) > 0$$

$\exists h \in C^1$  such that  $|g(s) - h(s)| \leq \frac{\varepsilon}{2}$  for  $0 < s \leq 1$ .

By the triangle inequality:

$$|B_t - g(t)| \leq |B_t - h(t)| + |h(t) - g(t)|$$

And so we need only show that:

$$\mathbf{P} \left( \sup_{0 \leq t \leq 1} |B_t - h(t)| < \frac{\varepsilon}{2} \right) > 0$$

Let the event  $G$  be given by:

$$\begin{aligned} G &= \left\{ |B_s - h(s)| < \frac{\varepsilon}{2}, s \in [0, 1] \right\} \\ &= \left\{ |X_s| < \frac{\varepsilon}{2}, s \in [0, 1] \right\} \end{aligned}$$

Where we define  $X_s = B_s - h(s)$ . We then see that:

$$dX_s = dB_s - h'(s) ds$$

and

$$M_t = e^{\left( \int_0^t h'(s) dB_s - \frac{1}{2} \int_0^t |h'(s)|^2 ds \right)}$$

With  $d\mathbf{Q} = M_t d\mathbf{P}$ , we see that:

- i) Under  $\mathbf{Q}$ ,  $X_t$  is a standard BM.
- ii)  $\mathbf{P} \sim \mathbf{Q}$  (i.e.,  $\mathbf{Q}(G) > 0 \Rightarrow \mathbf{P}(G) > 0$ ).

$$\begin{aligned} \mathbf{Q}(G) &= \int \left( \frac{d\mathbf{Q}}{d\mathbf{P}} \right) \mathbf{1}_G d\mathbf{P} \\ &\leq \left( \int \left( \frac{d\mathbf{Q}}{d\mathbf{P}} \right)^2 d\mathbf{P} \right)^{\frac{1}{2}} \mathbf{P}(G)^{\frac{1}{2}} \end{aligned}$$

and so:

$$\mathbf{P}(G) \geq \frac{\mathbf{Q}(G)^2}{\int \left( \frac{d\mathbf{Q}}{d\mathbf{P}} \right)^2 d\mathbf{P}}$$

## Bibliography

1. Leo Breiman, *Probability*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992, Corrected reprint of the 1968 original. MR 93d:60001
2. J. L. Doob, *Stochastic processes*, John Wiley & Sons Inc., New York, 1953. MR 15,445b
3. Richard Durrett, *Stochastic calculus*, Probability and Stochastics Series, CRC Press, Boca Raton, FL, 1996, A practical introduction. MR 1398879 (97k:60148)
4. ———, *Stochastic calculus, a practical introduction*, CRC Press, 1996.
5. John Guckenheimer and Philip Holmes, *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Applied Mathematical Sciences, vol. 42, Springer-Verlag, New York, 1990, Revised and corrected reprint of the 1983 original.
6. Philip Hartman, *Ordinary differential equations*, second ed., Birkhäuser, Boston, Mass., 1982.
7. Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR 1121940 (92h:60127)
8. ———, *Brownian motion and stochastic calculus*, second ed., Springer-Verlag, New York, 1991. MR 92h:60127
9. Fima C. Klebaner, *Introduction to stochastic calculus with applications*, third ed., Imperial College Press, London, 2012. MR 2933773
10. N. V. Krylov, *Introduction to the theory of random processes*, Graduate Studies in Mathematics, vol. 43, American Mathematical Society, Providence, RI, 2002. MR 1885884 (2003d:60001)
11. Hui-Hsiung Kuo, *Introduction to stochastic integration*, Universitext, Springer, New York, 2006. MR 2180429 (2006e:60001)
12. Peter Mörters and Yuval Peres, *Brownian motion*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2010, With an appendix by Oded Schramm and Wendelin Werner. MR 2604525 (2011i:60152)
13. Bernt Øksendal, *Stochastic differential equations*, fifth ed., Universitext, Springer-Verlag, Berlin, 1998, An introduction with applications. MR 1619188 (99c:60119)
14. Philip Protter, *Stochastic integration and differential equations: a new approach*, Springer-Verlag, 1990.
15. Philip E. Protter, *Stochastic integration and differential equations*, Stochastic Modelling and Applied Probability, vol. 21, Springer-Verlag, Berlin, 2005, Second edition. Version 2.1, Corrected third printing. MR 2273672 (2008e:60001)
16. Walter A. Strauss, *Partial differential equations*, John Wiley & Sons Inc., New York, 1992, An introduction. MR 92m:35001
17. S. J. Taylor, *Exact asymptotic estimates of brownian path variation*, Duke Mathematical Journal **39** (1972), no. 2, 219241, Mathematical Reviews number (MathSciNet) MR0295434, Zentralblatt MATH identifier 0241.60069.



## APPENDIX A

### Some Results from Analysis

Recalling that, given a probability space  $(\Omega, \Sigma, \mathbb{P})$  and a random variable  $X$  on such a space we define the expectation of a function  $f$  as the integral

$$\mathbb{E}[f(X)] = \int_{\Omega} f(\omega) \mathbb{P}(d\omega),$$

where  $\mathbb{P}$  denotes the (probability) measure against which we are integrating. The following results are stated for a general measure  $\mu$  (i.e., not necessarily a probability measure).

**THEOREM 0.1** (Lebesgue's Dominated Convergence theorem). *Let  $\{f_n\}$  be a sequence of measurable functions on a measure space  $(\Omega, \Sigma, \mu)$ . Suppose that the sequence converges pointwise to a function  $f$  and is dominated by some integrable function  $g$  in the sense that*

$$|f_n(x)| \leq g(x)$$

for all numbers  $n$  in the index set of the sequence and all points  $x \in S$ . Then  $f$  is integrable and

$$\lim_{n \rightarrow \infty} \int_{\Omega} |f_n - f| d\mu = 0$$

which also implies

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu$$

**THEOREM 0.2** (Fatou's Lemma). *Given a measure space  $(\Omega, \Sigma, \mu)$  and a set  $X \in \Sigma$ , let  $\{f_n\}$  be a sequence of  $(\Sigma, \mathcal{B}_{\mathbb{R}_{\geq 0}})$ -measurable non-negative functions  $f_n : X \rightarrow [0, +\infty]$ . Define the function  $f : X \rightarrow [0, +\infty]$  by setting*

$$f(x) = \liminf_{n \rightarrow \infty} f_n(x),$$

for every  $x \in X$ . Then  $f$  is  $(\Sigma, \mathcal{B}_{\mathbb{R}_{\geq 0}})$ -measurable, and

$$\int_{\Omega} f d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} f_n d\mu.$$

where the integrals may be finite or infinite.

**REMARK 0.3.** *The above theorem can in particular be used when  $f$  is the indicator function  $\mathbb{1}_{A_n}$  for a sequence of sets  $\{A_n\} \in \Sigma$ , obtaining*

$$\mu(\liminf_{n \rightarrow \infty} A_n) = \int_{\Omega} \liminf_{n \rightarrow \infty} \mathbb{1}_{A_n} d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} \mathbb{1}_{A_n} d\mu = \liminf_{n \rightarrow \infty} \mu(A_n).$$