STOCHASTIC DIFFERENTIAL EQUATIONS

Adair Antonio da Silva Neto

March 6, 2023

Contents

Chapter 1

Introduction

An equation that models an evolution process that contains 'noise', with a certain randomness, is an equation of the form:

$$\frac{\mathrm{dX}}{\mathrm{d}t} = b(t, X_t) + \sigma(t, X_t) \cdot \text{'white noise'} \tag{1.1}$$

Problems like this appears naturally in biology (populational growth models), physics (charge in electrical circuits), engineering (filtering problems, like the Kalman's Filter exihibted in the Figure 1.1) and finance (optimal stopping, optimal portfolio and option pricing).
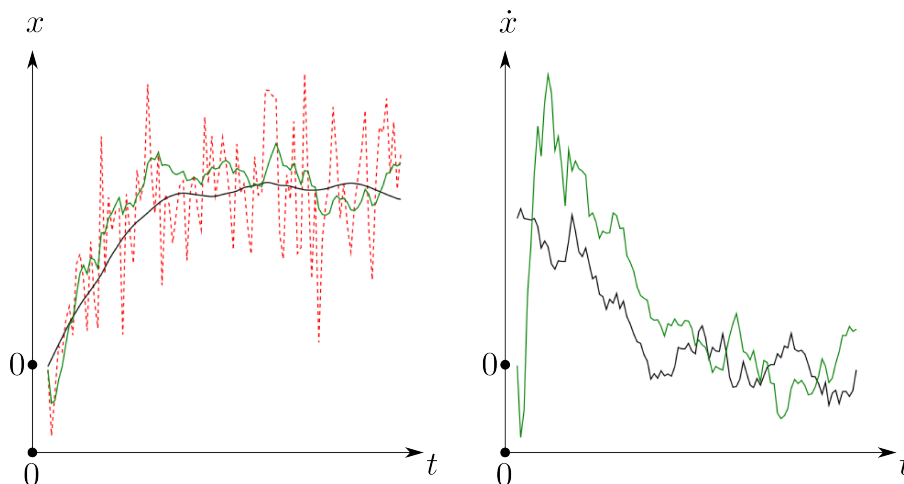
Figure 1.1: Kalman's Filter: ■ Observed data; ■ Filtered process; ■ Real data [Wik22].

From the mathematical viewpoint, we need to make sense of this kind of equations, since with the usual tools of calculus, it is not possible to treat them. Our goal then, in this work, is to give the foundations that allow us to treat this equations. In particular, firstly we're going to define the Itô's Integral and then we'll be ready to tackle stochastic differential equations.

Chapter 2

Probability Theory

2.1 Probability Spaces

A probability space starts with a set Ω called the **sample space**, which is, intuitively, a list of all possible outcomes of an experiment. Each $\omega \in \Omega$ is a **sample point**, and each subset $A \subseteq \Omega$ is an **event**.

To filter 'well-behaved' subsets, where it will be possible to measure a probability, the following definition is necessary.

> **Definition 2.1.1** (σ-algebra). If Ω is a set, then a σ**-algebra** $\mathfrak{F}$ on Ω is a family of subsets of Ω satisfying:
>
> 1. $\emptyset \in \mathfrak{F}$.
>
> 2. If $A \in \mathfrak{F}$, then $A^c \in \mathfrak{F}$.
>
> 3. If $A_1, A_2, \ldots \in \mathfrak{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{F}$.
>
> The pair $(\Omega, \mathfrak{F})$ is said to be a **measurable space**, and any subset $B \subseteq \Omega$ that also belongs to $\mathfrak{F}$ is called a **measurable set**.

> **Example 2.1.1** (Examples of σ-algebras). The following are σ-algebras.
>
> 1. The family $\{\emptyset, \Omega\}$ is called the **trivial σ-algebra**, and is the smallest one possible.
>
> 2. The power set $\mathfrak{P}(\Omega)$ is called the **discrete σ-algebra**, and is the largest one possible.
>
> 3. The family $\{\emptyset, \Omega, A, A^c\}$ is the σ**-algebra generated by the set** A and is usually denoted by $\mathfrak{F}_A$.

But the most important σ-algebra for probability theory is the Borel σ-algebra, which will be denoted by $\mathfrak{B}$. Taking $\Omega = \mathbb{R}$, the Borel σ-algebra is generated by the intersection of all σ-algebras containing the real line intervals.

Notice that the Borel σ-algebra contains all open sets, closed sets, and all their countable operations with union $\cup$, intersection $\cap$, and their complements ^{c}. This is the smallest σ-algebra containing all open subsets.

More generally, if U is the collection of all open subsets of a topological space Ω, then $\mathfrak{B}$ is called the **Borel σ-algebra** on Ω. The elements $B \in \mathfrak{B}$ are called **Borel sets**.

Given a measurable space, it's possible to assign each outcome to a probability.

Definition 2.1.2 (Probability Measure). Let $(\Omega, \mathfrak{F})$ be a measurable space. A **probability measure** P is the function

$$P : \mathfrak{F} \longrightarrow [0, 1]$$

satisfying

1. $P(\emptyset) = 0$ and $P(\Omega) = 1$.

2. (σ-additivity). If $A_1, A_2, \ldots \in \mathfrak{F}$ and $A_i \cap A_j = \emptyset$, for $i \neq j$ (i.e. are mutually exclusive), then

$$P\left(\bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} P(A_i)$$

And the triple $(\Omega, \mathfrak{F}, P)$ is called a **probability space**.

In order to restrict the functions to sets in the σ-algebra, the following definition will be needed.

Definition 2.1.3 ($\mathfrak{F}$-measurable function). Given a probability space $(\Omega, \mathfrak{F}, P)$, a function Y from the sample space Ω to $\mathbf{R}^n$ is called $\mathfrak{F}$-measurable if

$$Y^{-1}(U) = \{\omega \in \Omega : Y(\omega) \in U\} \in \mathfrak{F}$$

for all open sets $U \in \mathbf{R}^n$ (i.e. for all Borel sets $U \subseteq \mathbf{R}^n$). In other words, the inverse image of U is in the σ-algebra.

And to attach numerical values to each $\omega \in \Omega$, define the following function.

Definition 2.1.4 (Random variable). An $\mathfrak{F}$-measurable function $X : \Omega \longrightarrow \mathbf{R}^n$ is called a **random variable** on a complete probability space $(\Omega, \mathfrak{F}, P)$.

This definition means that if we know which event U in the $\mathfrak{F}$ has occurred, then we know which value of X has occurred.

Consider the measurable space $(\Omega, \mathfrak{P}(\Omega))$, and let X be a random variable with values x_i, $i = 1, 2, \ldots, k$. The sets

$$A_i = \{\omega \in \Omega : X(\omega) = x_i\} \subseteq \Omega$$

form a partition of Ω, and the σ-algebra generated by this partition is called the σ**-algebra generated by** X.

Notice that this is the smallest σ-algebra that contains all the sets A_i. This σ-algebra is often denoted by $\mathfrak{F}_X$. Intuitively, this σ-algebra represents all information available about the sample point ω by observing X.

2.2 Distribution and Density of a Random Variable

However, given only the probability measure P, it is not always immediate how to compute the probability of a given interval, set or value. For that purpose, the following three definitions will be useful.

Definition 2.2.1 (Distribution). The **distribution** of a random variable X is a function $\mu_X : \Omega \longrightarrow \mathbf{R}^n$ defined as

$$\mu_X(B) = P[X^{-1}(B)] = P[X \in B] = P(\{\omega : X(\omega) \in B\})$$

Definition 2.2.2 (Distribution Function). The **cumulative distribution function** (CDF) of a random variable X is defined as

$$F_X(x) = P[X \leq x] = P(\{\omega : X(\omega) \leq x\})$$

where $x \in \mathbf{R}$. Notice that F is non-decreasing, and right-continuous, and approaches 0 at $-\infty$ and 1 at $+\infty$.

Definition 2.2.3 (Joint Distribution). If X and Y are random variables, then their **joint distribution function** is

$$F_{X,Y}(x,y) = P[X \leq x, Y \leq y]$$

where $x, y \in \mathbf{R}$.

The distribution function gives the probability that the random variable X is on the interval $(-\infty, x]$. This also allows to look for an interval $[a, b]$ using $F_X(b) - F_X(a)$.

While the distribution of X returns the probability of an event $\{\omega : X(\omega) \in B\}$.

Definition 2.2.4 (Density function). A random variable X has a **probability density function** (PDF) $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ if f is a measurable function and

$$F_X(x) = P[X \leq x] = \int_{-\infty}^{x} f(y) \, dy$$

2.3 Expected Value and Variance

Another useful tool for any given random variable is to know its mean value and how much it varies. This motivates the following.

Definition 2.3.1 (Expectation). Let $(\Omega, \mathfrak{F}, P)$ be a probability space and X a random variable. If $\int_{\Omega} |X(\omega)| \, dP(\omega) < \infty$, then the **expected value** (or **mean value** of X with respect to P is

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) \, dP(\omega) = \int_{\mathbf{R}^n} x \, d\mu_X(x)$$

Notice that the expectation is linear. I.e., if X and Y are integrable and a and b are constants, then

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

If f is a Borel measurable function and $\int_\Omega |f(X)| \, dP(\omega) < \infty$, then

$$\mathbb{E}[f(X)] = \int_\Omega f(X(\omega)) \, dP(\omega) = \int_{\mathbf{R}^n} f(X) \, d\mu_X(x)$$

Theorem 2.3.1 (Chebychev's inequality).

$$P[\,|X| \geq \lambda\,] \leq \frac{1}{\lambda^p} \mathbb{E}[\,|X|^p\,]$$

for all $\lambda \geq 0$.

Definition 2.3.2 (Variance). Let $\mu := \mathbb{E}[X]$, i.e., the expected value of a random variable X. Then the **variance** of X is given by

$$\mathrm{Var}[X] = \mathbb{E}[\|X - \mu\|^2] = \int_\Omega |X - \mu|^2 \, dP(\omega)$$

Definition 2.3.3 (Covariance). Let X and Y be integral random variables. Let $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$. If XY is integrable, then the **covariance** of X and Y is:

$$\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mu_X \mu_Y$$

Notice that $\mathrm{Var}[X] = \mathrm{Cov}[X, X]$.

The variance of X can be computed more simply using the following theorem.

Theorem 2.3.2. Let X be a random variable. Then,

1. $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

2. $\mathrm{Var}[aX + b] = a^2 \mathrm{Var}[X]$.

3. $\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X]$.

4. $\mathrm{Cov}[aX + bY, Z] = a\mathrm{Cov}[X, Z] + b\mathrm{Cov}[Y, Z]$.

5. $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y]$.

Definition 2.3.4 (Independence). Two events $A, B \in \mathfrak{F}$ are **independent** if

$$P[A \cap B] = P[A] \cdot P[B]$$

More generally, any collection of events A_i, $i = 1, 2, \ldots$, is called independent if for any $n \in \mathbf{N}$ and any choice of indices i_k, $k = 1, 2, \ldots$,

$$P\left(\bigcap_{k=1}^n A_{i_k}\right) = \prod_{k=1}^n P\left(A_{i_k}\right)$$

Definition 2.3.5 (Conditional Expectation)**.** Let X and Y be random variables. Then, given that $f_Y(y) > 0$, the **conditional distribution** of X given $Y = y$ is given by the conditional density

$$f(x \mid y) = \frac{f(x,y)}{f_Y(y)}$$

And the **conditional expectation** of X given $Y = y$ is

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f(x \mid y) \, dx$$

The intuition behind this definition is to build an estimate of the random variable X given the information available in Y.

Some important and very useful properties of conditional expectation are listed below.

Theorem 2.3.3. Let $X, Y : \Omega \longrightarrow \mathbf{R}^n$ be random variables with $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|Y|] < \infty$, $\mathfrak{H} \subseteq \mathfrak{F}$ a σ-algebra, and let $a, b \in \mathbf{R}$.

1. $\mathbb{E}[aX + bY \mid \mathfrak{H}] = a\mathbb{E}[X \mid \mathfrak{H}] + b\mathbb{E}[Y \mid \mathfrak{H}]$.

2. $\mathbb{E}[\mathbb{E}[X \mid \mathfrak{H}]] = \mathbb{E}[X]$.

3. $\mathbb{E}[X \mid \mathfrak{H}] = X$ if X is $\mathfrak{H}$-measurable.

4. $\mathbb{E}[X \mid \mathfrak{H}] = \mathbb{E}[X]$ if X is independent of $\mathfrak{H}$.

5. $\mathbb{E}[Y \cdot X \mid \mathfrak{H}] = Y \cdot \mathbb{E}[X \mid \mathfrak{H}]$ if Y is $\mathfrak{H}$-measurable.

6. If $\mathfrak{G} \subseteq \mathfrak{H}$ is a σ-algebra, then

$$\mathbb{E}[X \mid \mathfrak{G}] = \mathbb{E}[\mathbb{E}[X \mid \mathfrak{H}] \mid \mathfrak{G}]$$

7. If $\varphi : \mathbf{R} \longrightarrow \mathbf{R}$ is convex and $\mathbb{E}[|\varphi(X)|] < \infty$, then

$$\varphi(\mathbb{E}[X \mid \mathfrak{H}]) \leq \mathbb{E}[\varphi(X) \mid \mathfrak{H}]$$

Definition 2.3.6 (Infinitely often)**.** If $A_1, A_2, \ldots, A_n, \ldots$ are events in the probability space, then the event

$$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\omega \in \Omega : \omega \text{ belongs to infinitely many of the } A_n\}$$

is called A_n **infinitely often**, or simply 'A_n **i.o.**'.

The next lemma helps us check if some sequence of events occurs infinitely often.

Lemma 2.3.4 (Borel-Cantelli)**.** If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then

$$P(A_n \text{ i.o.}) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) = 0$$

The most important distribution in this text is the **normal** (or **gaussian**) distribution.

Definition 2.3.7 (Normal Distribution). If random variable X has mean μ, variance σ^2 and a density function of the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

then X is said to have a **normal distribution**, which we denote by $X \sim N(\mu, \sigma^2)$.

Notice that X can be normalized (i.e. transformed into a distribution of the form $N(0,1)$) by taking $Z = \frac{X-\mu}{\sigma}$.

Definition 2.3.8 (Lipschitz and Hölder Conditions). A function f is said to be **Hölder continuous** of order α, $0 < \alpha \leq 1$, on an interval $[a, b] \subseteq \mathbf{R}$ if there exists a constant $K > 0$, such that for all $x, y \in [a, b]$ we have

$$|f(x) - f(y)| \leq K|x-y|^\alpha$$

If $\alpha = 1$, then the function f is said to be **Lipschitz continuous**.

Corollary 2.3.5. Let $X \in L^1(\mathbb{P})$, $\{\mathfrak{N}_k\}_{k=1}^\infty$ be an increasing family of σ-algebras, $\mathfrak{N}_k \subset \mathfrak{F}$ and define $\mathfrak{N}_\infty$ to be the σ-lagebra generated by $\{\mathfrak{N}_k\}_{k=1}^\infty$.

Then,

$$\mathbb{E}[X \mid \mathfrak{N}_k] \longrightarrow \mathbb{E}[X \mid \mathfrak{N}_\infty] \text{ as } k \to \infty$$

a.e. $\mathbb{P}$ and in $L^1(\mathbb{P})$.

2.4 Stochastic Processes

Definition 2.4.1 (Stochastic Process). A **stochastic process** is a collection of random variables $\{X(t)\}$ parametrized by time $t \in T$.

For each point $\omega \in \Omega$, the mapping $t \longrightarrow X_t(\omega)$ is the respective **sample path**, also called **realization** or **trajectory**.

Definition 2.4.2 (Filtration). A filtration $\mathbf{F}$ is a collection of σ-algebras

$$\mathbf{F} = \{\mathfrak{F}_0, \mathfrak{F}_1, \dots, \mathfrak{F}_t, \dots, \mathfrak{F}_T\}$$

such that $\mathfrak{F}_t \subseteq \mathfrak{F}_{t+1}$.

The idea behind this definition is to model the flow of information. As the time t passes, the observer has more information, and does not lose any previous data. Notice that this implies finer partitions of the sample space Ω.

Given a stochastic process $\{X(t)\}$, let $\mathfrak{F}_t$ be the σ-algebra generated by the random variables X_s, $s = 0, \dots, t$. Since $\mathfrak{F}_t \subseteq \mathfrak{F}_{t+1}$, these σ-algebras form a filtration called the **natural filtration** of the process $\{X(t)\}$, and contain all available information of the process up to the time t.

Definition 2.4.3 (Martingale). A stochastic process $\{M_t\}$ on $(\Omega, \mathfrak{F}, P)$ is called a **martingale** with respect to a filtration $\mathfrak{M}_t$ if

1. M_t is $\mathfrak{M}_t$-measurable for all t.

2. $\mathbb{E}[\,|M_t|\,] < \infty$ for all t.

3. $\mathbb{E}[M_s \mid \mathfrak{M}_t] = M_t$ for all $s \geq t$.

Intuitively, a martingale is a stochastic process in which the future has no tendency to go up or down. In other words, the expected value of any time in the future is equal to the value of the process at the present time.

Theorem 2.4.1 (Doob's martingale inequality)**.** If M_t is a martingale such that the mapping $t \longrightarrow M_t(\omega)$ is continuous a.s., then for all $p \geq 1$, $T \geq 0$ and $\lambda > 0$,

$$P\left[\sup_{0 \leq t \leq T} |M_t| \geq \lambda\right] \leq \frac{1}{\lambda^p}\mathbb{E}[\,|M_T|^p\,]$$

2.5 Brownian Motion

In 1828 the Scottish botanist Robert Brown described the irregular motion of pollen grains suspended in fluid, that is now known as the Brownian Motion. To describe this mathematically, we use a stochastic process $B_t(\omega)$, which is the position at time t of the particle ω.

This can also be understood as the model for the cumulative effect of 'noise'. And is also called Wiener process, after N. Wiener, who formalized mathematically the Brownian motion.

Definition 2.5.1 (Brownian Motion)**.** A Brownian motion $\{B(t)\}$ is a stochastic process with the following properties:

1. (Independence of increments). $B(t) - B(s)$, for $t > s$, are independent. That means that the direction that the particle will go does not depend on the past.

2. (Normal increments). $B(t) - B(s)$ has normal distribution with mean 0 and variance $t - s$. Notice that by taking $s = 0$ we have $B(t) - B(0) \sim N(0, t)$.

3. (Continuity of paths). $B(t)$, $t \geq 0$, are continuous functions of t.

The existence and continuity of Brownian motion is proved using Kolmogorov's extension and continuity theorem [Oks13].

Theorem 2.5.1 (Quadratic variation)**.** If $t \geq s$,

$$\mathbb{E}[(B_t - B_s)^2] = t - s$$

More generally (in $\mathbf{R}^n$):

$$\mathbb{E}[(B_t - B_s)^2] = n(t - s)$$

Theorem 2.5.2.

$$\mathbb{E}[f(B_t)] = \frac{1}{\sqrt{2\pi t}} \int_{\mathbf{R}} f(x) e^{-\frac{x^2}{2t}} \, \mathrm{d}x$$

Theorem 2.5.3. The expected value of odd moments of B_t (other than one) are zero. And the even moments are given by

$$\mathbb{E}[B_t^{2k}] = \frac{(2k)!}{2^k k!} t^k$$

Last but not least, two important identities, useful for further calculations are:

$$B_j(B_{j+1} - B_j) = \frac{1}{2}[B_{j+1}^2 - B_j^2 - (B_{j+1} - B_j)^2]$$

$$B_j^2(B_{j+1} - B_j) = \frac{1}{3}(B_{j+1}^3 - B_j^3) - B_j(B_{j+1} - B_j)^2 - \frac{1}{3}(B_{j+1} - B_j)^3$$

2.6 Riemann-Stieltjes Integral

Definition 2.6.1 (Riemann-Stieltjes Integral). If φ is a continuous function on $[a, b]$ and F is a distribution function, we define the **Riemann-Stieltjes Integral** of φ on $[a, b]$ in relation to F (or weighted by F) as

$$\lim_{\|\Delta\| \to 0} \sum_{i=1}^{n} \varphi(y_i)[F(x_{i+1}) - F(x_i))] = \int_a^b \varphi(x) \, dF(x)$$

where $a = x_1 < x_2 < \ldots < x_{n+1} - b$, y_i is an arbitrary point of $[x_i, x_{i+1}]$, and $\|\Delta\| = \max_{1 \leq i \leq n}(x_{i+1} - x_i)$.

If F is a discrete random variable, then

$$\int_a^b \varphi \, dF = \sum_{i: a < x_i \leq b} \varphi(x_i) p(x_i)$$

And if F is a continuous random variable, then

$$\int_a^b \varphi \, dF = \int_a^b \varphi(x) f(x) \, dx$$

2.7 Modes of Convergence

Definition 2.7.1. A **space** $L^p(\Omega, \mathfrak{F}, P)$ is the space of measurable functions in which the p-th power of the absolute value is Lebesgue integrable. I.e., functions of the form

$$\|f\|_p := \left(\int_\Omega |f|^p \, d\mu \right)^{1/p} < \infty$$

We can also write this as

$$\|X\|_p := \sqrt[p]{\mathbb{E}[|X|^p]} < \infty$$

A sequence of random variable can converge in different ways. The following definitions are given in an increasing order of strength.

Definition 2.7.2 (Convergence in Distribution)**.** A sequence of random variables $\{X_n\}$ **converges in distribution** to X if their distribution functions $F_{X_n}(x)$ converge to the distribution function $F_X(x)$ at any point of continuity of F_X.

In other words, for any bounded function $f : \mathbf{R} \longrightarrow \mathbf{R}$ we have

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$$

I.e.,

$$\lim_{n\to\infty} \int_{\Omega_n} f(X_n)\,dP_n = \int_{\Omega} f(X)\,dP$$

Definition 2.7.3 (Convergence in Probability)**.** A sequence of random variables $\{X_n\}$ **converges in probability** to X if for any $\varepsilon > 0$, we have

$$\lim_{n\to\infty} P[\,|X_n(\omega) - X(\omega)| > \varepsilon\,] = 0$$

Definition 2.7.4 (Convergence Almost Surely)**.** A sequence of random variables $\{X_n\}$ **converges almost surely (a.s.)** to X if for any ω outside a set of probability zero we have that

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

Alternatively, we can write

$$P[X_n(\omega) \to X(\omega)] = 1$$

Definition 2.7.5 (L^p-Convergence)**.** A sequence of random variables $\{X_n\}$ **converges in** L^p to X if $\{X_n\} \subseteq L^p$, $p \in [1, \infty)$, and

$$\lim_{n\to\infty} \|X_n - X\|_p = 0$$

I.e., $\mathbb{E}\left[\,|X_n|^p\right] < \infty$ and

$$\lim_{n\to\infty} \mathbb{E}\left[\,|X_n - X|^p\right] = 0$$

Chapter 3

Stochastic Integrals

3.1   Formalizing the 'noise'

To study the equation (1.1), the first step is to give a formalization for the 'noise' that we are going to consider. In our case, we'll consider the 'white noise', which is a generalized stochastic process W_t with gaussian probability distribution with mean zero, finite variance such that $\mathbb{E}[W_s W_t] = 0$ whenever $t \neq s$.

In other words, the 'noise' will be represented by some stochastic process W_t and our goal is to satisfy the following properties:

- If $t_1 \neq t_2$, then W_{t_1} and W_{t_2} are independent.

- The collection $\{W_t\}$ is stationary, i.e., the joint distribution of any collections of W_t does not depend on t.

- $\mathbb{E}[W_t] = 0$ for all t.

However, this process W_t does not have reasonable properties (does not have continuous paths, and it is not measurable for finite time). So our first goal is to replace W_t with some convenient stochastic process.

Let $0 = t_0 < t_1 < \ldots < t_m = t$ and consider the discrete version

$$X_{k+1} - X_k = b(t_k, X_k) \Delta t_k + \sigma(t_k, X_k) W_k \Delta t_k$$

Now, replace $W_k \Delta t_k = \Delta V_k = V_{k+1} - V_k$, where V_t is a stochastic process. Since we want V_t to have stationary independent increments with mean zero, the only suitable process with continuous paths is the Brownian motion, represented in the Figure 3.1.

Hence, we take $V_t = B_t$, thus obtaining

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

Taking the limit as Δt_j goes to zero, we have

$$X_k = X_0 + \int_0^t b(s, X_s) \, ds + \int_0^t \sigma(s, X_s) \, dB_s$$

Notice that

$$B_t - B_s = W_s(t - s)$$

for all $t \geq s$, i.e., the 'white noise' is the derivative, with respect to time, of the Brownian motion.
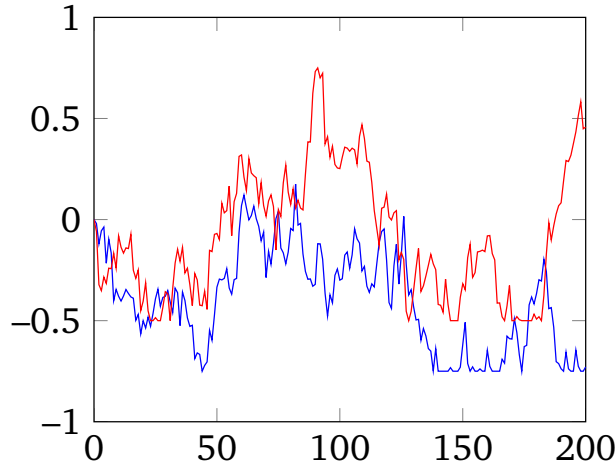

Figure 3.1: Two realizations of the Brownian motion [Jak]

3.2 Preparing the Terrain

Now we can start treating the equation. In order to do that, notice that in the determinist case, in which there isn't a 'noise', we have an equation of the form

$$\frac{dX_t}{dt} = b(t, X_t), \quad X_0 = x_0 \tag{3.1}$$

In this case, a solution is a function X_t such that

$$X_t - X_0 = \int_0^T b(s, X_s) \, ds \tag{3.2}$$

This function can be understood as a stochastic process over a space of probability given by a single point.

Using the formalism above, we propose as a solution to the equation (1.1) an identity of the form

$$X_t - X_0 = \int_0^t b(s, X_s) \, ds + \int_0^t \sigma(s, X_s) \, dB_s \tag{3.3}$$

where the left integral is a Lebesgue-Stieltjes' integral, but the right integral is what still needs to be formalized.

As said earlier, our task is to understand the following integral

$$\int_S^T f(t, \omega) \, dB_t(\omega) \tag{3.4}$$

where $f : [0, \infty] \times \Omega \longrightarrow \mathbf{R}$.

The idea will be the following: we'll first define the integral for the most elementary functions and, after that, using approximations and convergence criteria, we'll define it for the other classes of functions.

Since it is natural to approximate a given function $f(t, \omega)$ by

$$\sum_j f(t_j^*, \omega) \chi_{[t_j, t_{j+1}]}(t)$$

where $t_j^* \in [t_j, t_{j+1}]$, we could define the integral (3.4) as follows

$$\int_S^T f(t, \omega) \, dB_t(\omega) = \lim_{n \to \infty} f(t_j^*, \omega)[B_{t_{j+1}} - B_{t_j}](\omega)$$

However, the choice of points t_j^* make a difference here. If we take $t_j^* = t_j$, the left end point, then we obtain the **Itô integral**, denoted by

$$\int_S^T f(t, \omega) \, dB_t(\omega)$$

If, on the other hand, we take $t_j^* = (t_j + t_{j+1})/2$, the mid point, then we obtain the **Stratonovich integral**, denoted by

$$\int_S^T f(t, \omega) \circ dB_t(\omega)$$

One of the advantages of Itô integral is its application in Finance. Intuitively, by taking the leftmost point, one does not have to know the future, like whether a stock goes up or down.

Related to this intuition, and developing the Ito integral, we must restrict ourselves to functions f that only depends on the behaviour of $B_s(\omega)$ up to the time t_j.

> **Definition 3.2.1.** Let $B_t(\omega)$ be a Brownian motion. We define $\mathfrak{F}_t$ as the σ-algebra generated by the random variables B_s, where $s \leq t$. I.e., $\mathfrak{F}_t$ is the smallest σ-algebra containing all sets
>
> $$\{\omega : B_{t_1}(\omega) \in F_1, \ldots, B_{t_k}(\omega)\} \in F_k$$
>
> where $t_j \leq t$ and $F_j \subseteq \mathbf{R}^n$ are Borel sets, $j \leq k = 1, 2, \ldots$.

The intuition behind this definition is that $\mathfrak{F}_t$ can be thought as the 'history of B_s up to the time t'. A function $h(\omega)$ is $\mathfrak{F}_t$-measurable if h depends only on the values from B_0 up to B_t. I.e., it does not depend on the 'future'.

> **Definition 3.2.2** ($\mathfrak{N}_t$-Adapted). Let $\{\mathfrak{N}_t\}_{t \geq 0}$ be an increasing family of σ-algebras of subsets of Ω. A process $g(t, \omega) : [0, \infty) \times \Omega \longrightarrow \mathbf{R}^n$ is $\mathfrak{N}_t$-**adapted** if for each $t \geq 0$ the function $\omega \longrightarrow g(t, \omega)$ is $\mathfrak{N}_t$-measurable.

> **Example 3.2.1.** The process $h_1(t, \omega) = B_{t/2}(\omega)$ is $\mathfrak{F}_t$-measurable and, hence, $\mathfrak{F}_t$-adapted. Notice that this process only depends on previous information.
>
> However, the process $h_2(t, \omega) = B_{2t}(\omega)$ is not $\mathfrak{F}_t$-measurable and, hence, it is not $\mathfrak{F}_t$-adapted. That's because the process h_2 depends on future information.

3.3 Constructing the Itô Integral

The class of functions for which the Itô integral will be first defined is the following.

Definition 3.3.1. Let $\mathfrak{V}(S, T)$ be the class of functions

$$f(t, \omega) : [0, \infty) \times \Omega \longrightarrow \mathbf{R}$$

such that

1. The function $(t, \omega) \longrightarrow f(t, \omega)$ is $\mathfrak{B} \times \mathfrak{F}$-measurable, where $\mathfrak{B}$ denotes the Borel σ-algebra on $[0, \infty)$.

2. The function $f(t, \omega)$ is $\mathfrak{F}_t$-adapted.

3. $\mathbb{E}\left[\int_S^T f(t, \omega)^2 \, dt\right] < \infty$.

Given functions $f \in \mathfrak{V}$, in order to define the Ito integral

$$I[f](\omega) = \int_S^T f(t, \omega) \, dB_t(\omega)$$

we are going to define $I[\varphi]$ for a simple class of functions φ and then show how each $f \in \mathfrak{V}$ can be approximated by φ.

Definition 3.3.2 (Elementary functions). A function $\varphi \in \mathfrak{V}$ is **elementary** if it has the form

$$\varphi(t, \omega) = \sum_j e_j(\omega) \chi_{[t_j, t_{j+1})}(t)$$

Since $\varphi \in \mathfrak{V}$, each function e_j is $\mathfrak{F}_j$-measurable. And, as we did before, we define

$$\int_S^T \varphi(t, \omega) \, dB_t(\omega) = \sum_{j \geq 0} e_j(\omega)[B_{j+1} - B_j](\omega)$$

To further our development of the Ito integral, we need the following result.

Theorem 3.3.1 (Itô isometry). If $\varphi(t, \omega)$ is bounded and elementary, then

$$\mathbb{E}\left[\left(\int_S^T \varphi(t, \omega) \, dB_t(\omega)\right)^2\right] = \mathbb{E}\left[\int_S^T \varphi(t, \omega)^2 \, dt\right]$$

Using this isometry, it is possible to extended the previous definition to functions in $\mathfrak{V}$ by the following steps.

1. Let $g(\omega) \in \mathfrak{V}$ be bounded and continuous for each ω. Then there exists elementary func-

tions $\varphi_n \in \mathfrak{V}$ such that

$$\mathbb{E}\left[\int_S^T (g - \varphi_n)^2 \, dt\right] \longrightarrow 0$$

as $n \to \infty$.

2. Let $h \in \mathfrak{V}$ be bounded, then there exists bounded functions $g_n \in \mathfrak{V}$ such that $g_n(\omega)$ are continuous for all ω and n, and

$$\mathbb{E}\left[\int_S^T (h - g_n)^2 \, dt\right] \longrightarrow 0$$

3. Let $f \in \mathfrak{V}$. Then there exists a sequence of functions $(h_n) \subseteq \mathfrak{V}$ such that h_n is bounded for each n and

$$\mathbb{E}\left[\int_S^T (f - h_n)^2 \, dt\right] \longrightarrow 0$$

as $n \to \infty$.

Please note that we started from bounded and continuous functions, generalized into bounded functions and then generalized further into any function in our class $\mathfrak{V}$.

These steps can be summarized in the following result, which guarantees that the elementary functions are dense in $\mathfrak{V}(S, T)$.

Theorem 3.3.2. If $g \in \mathfrak{V}(S, T)$, then there exists a sequence of elementary functions $\varphi_n \in \mathfrak{V}(S, T)$ such that $\varphi_n \to g$ in $L^2(P)$

Finally, we can define the Itô integral as follows.

Definition 3.3.3 (The Itô integral). Let $f \in \mathfrak{V}(S, T)$, then the **Itô integral** of f, from S to T is

$$\int_S^T f(t, \omega) \, dB_t(\omega) = \lim_{n \to \infty} \int_S^T \varphi_n(t, \omega) \, dB_t(\omega)$$

where the limit is in $L^2(P)$ and (φ_n) is a sequence of elementary functions such that

$$\mathbb{E}\left[\int_S^T (f(t, \omega) - \varphi_n(t, \omega))^2 \, dt\right] \longrightarrow 0$$

as $n \to \infty$.

This new definition induces a generalized form of the Ito isometry.

Theorem 3.3.3 (Ito isometry). For all $f \in \mathfrak{V}(S, T)$,

$$\mathbb{E}\left[\left(\int_S^T f(t, \omega) \, dB_t(\omega)\right)^2\right] = \mathbb{E}\left[\int_S^T f(t, \omega)^2 \, dt\right]$$

Corollary 3.3.4. If $f(t, \omega) \in \mathfrak{V}(S, T)$ and $f_n(t, \omega) \in \mathfrak{V}(S, T)$ and

$$\mathbb{E}\left[\left(\int_S^T f_n(t, \omega) - f(t, \omega)\right)^2 dt\right] \longrightarrow 0$$

as $n \to \infty$, then

$$\int_S^T f_n(t, \omega) \, dB_t(\omega) \longrightarrow \int_S^T f(t, \omega) \, dB)t(\omega)$$

in $L^2(P)$ as $n \to \infty$.

Example 3.3.1. We wish to compute the integral

$$\int_0^T B(t) \, dB(t)$$

Step 1. Let

$$\varphi_n(t) = \sum_{i=0}^{n-1} B_n(t_i^n)[B_n(t_{i+1}^n) - B_n(t_i^n)]$$

be a sequence of elementary functions. Then,

$$\mathbb{E}\left[\int_0^T (\varphi_n - B_s)^2 \, ds\right] = \mathbb{E}\left[\sum_j \int_{t_j}^{t_{j+1}} (B_j - B_s)^2 \, ds\right] = \sum_j \int_{t_j}^{t_{j+1}} (s - t_j)^2 \, ds$$

$$= \sum_j \frac{1}{2}\left(t_{j+1} - t_j\right)^2 \longrightarrow 0 \text{ when } \Delta t_j \to 0$$

I.e., $\varphi_n(t)$ converges to $B(t)$ almost surely as $\max_i(t_{i+1}^n - t_i^n) \to 0$ by the continuity of $B(t)$.

Step 2. Now notice that

$$B_n(t_i)[B_n(t_{i+1}) - B_n(t_i)] = B_n(t_{i+1})B_n(t_i) - B_n^2(t_i) + B_n^2(t_{i+1}) - B_n^2(t_{i+1})$$

$$= \frac{1}{2}[B_n^2(t_{i+1}) - B_n^2(t_i) - (B_n(t_{i+1}) - B_n(t_i))^2]$$

Step 3. With these two results, we can write the original integral as

$$\int_0^T X_n(t) \, dB(t) = \frac{1}{2}\sum_{i=0}^{n-1}\left[B_n^2(t_{i+1}) - B_n^2(t_i)\right] - \frac{1}{2}\sum_{i=0}^{n-1}[B_n(t_{i+1}) - B_n(t_i)]^2$$

Step 4. Since the first sum is a telescopic sum and the second one converges in probability to T by the quadratic variation of Brownian motion, we have

$$\int_0^T X_n(t) \, dB(t) = \frac{1}{2}B^2(t) - \frac{1}{2}B^2(0) - \frac{1}{2}T = \frac{1}{2}B^2(t) - \frac{1}{2}T$$

Hence, the integral converges in $L^2(P)$ to

$$\int_0^T B(t)\, dB(t) = \lim_{n\to\infty} \int_0^T X_n(t)\, dB(t) = \frac{1}{2}B^2(t) - \frac{1}{2}T$$

3.4 Properties

Before heading to more theoretical and important results, let us notice some natural facts of the Itô integral.

Theorem 3.4.1. Let $f, g \in \mathfrak{V}(0, T)$ and let $0 \le S < U < T$. Then

1. $\int_S^T f\, dB_t = \int_S^U f\, dB_t + \int_U^T f\, dB_t$ for a.a. ω.

2. $\int_S^T (cf + g)\, dB_t = c\int_S^T f\, dB_t + \int_S^T g\, dB_t$ where $c \in \mathbf{R}$.

3. $\mathbb{E}\left[\int_S^T f\, dB_t\right] = 0$.

4. $\int_S^T f\, dB_t$ is $\mathfrak{F}_T$-measurable.

With the Doob's martingale inequality, it can be proved that the Itô integral can be chosen to depend continuously on t. A proof of this fact is on the third chapter of [Oks13].

Now an important result is that the Itô integral is a martingale.

Theorem 3.4.2. Let $f(t, \omega) \in \mathfrak{V}(0, T)$ for all T. Then

$$M_t(\omega) = \int_0^t f(s, \omega)\, dB_s$$

is a martingale w.r.t. $\mathfrak{F}_t$ and for $\lambda, T > 0$

$$P\left[\sup_{0 \le t \le T} |M_t| \ge \lambda\right] \le \frac{1}{\lambda^2}\mathbb{E}\left[\int_0^T f(s, \omega)^2\, ds\right]$$

Summarizing, the Itô integral is continuous, adapted, linear, a martingale, satisfies the Itô isometry and has Quadratic variation.

3.5 Extensions

Using the concept of martingales, we can generalize the Itô integral for a larger class of functions than $\mathfrak{V}$.

Considering the Definition 3.3.1, we can relax the condition 2. into

2.' There exists an increasing familiy of σ-algebras $\mathfrak{H}_t$ such that B_t is a martingale w.r.t. $\mathfrak{H}_t$ and f_t is $\mathfrak{H}_t$-adapted.

The idea here is that B_t must remain a martingale with respect to the history of f_s.

Another way of extending the Itô integral definition is by weaking the condition 3. into

3.' $P\left[\int_S^T f(s, \omega)^2 \, ds < \infty\right] = 1$

Too understand why this works, let $\mathfrak{W}_{\mathfrak{H}}(S, T)$ denote the class of processes satisfying conditions 1, 2' and 3' above. Then, in the 1-dimensional Brownian motion, for all t there exists step functions $f_n \in \mathfrak{W}_{\mathfrak{H}}$ such that

$$\int_0^t |f_n - f|^2 \, ds \longrightarrow 0$$

in probability.

Now, since $\int_0^t f_n(s, \omega) \, dB_s(\omega)$ converges in probability to a random variable and the limit depends only on f, we define

$$\int_0^t f(s, \omega) \, dB_s(\omega) = \lim_{n \to \infty} \int_0^t f_n(s, \omega) \, dB_s(\omega)$$

where the limit is in probability and $f \in \mathfrak{W}_{\mathfrak{H}}$.

3.6 Stratonovich integral

As we saw, the Itô integral is one possible interpretation of the integral

$$\int_S^T f(t, \omega) \, dB_t(\omega)$$

The Stratonovich integral is another possibility and, in general, leads to different results (except when the integrating function has a derivative). In some cases, the Stratonovich definition may be adequate.

In other cases, Itô's feature of 'not looking in the future' (as [Oks13] puts it), justifies its use in biology and finance, for example. Moreover, Itô integral is a martingale, and Stratonovich's is not. This gives an important computational advantage to Itô's definition.

And, to quote [S+04],

> However, it [Stratonovich's integral] is inappropriate for finance. In finance, the integrand represents a position in an asset and the integrator represents the price of that asset. We cannot decide at 1:00 p.m. which position we took at 9:00 a.m. We must decide the position at the beginning of each time interval, and the Ito integral is the limit of the gain achieved by that kind of trading as the time between trades approaches zero.

However, Stratonovich integral is used in manifolds and is of particular interest to physicists for obeying rules of classical calculus, such as the chain rule.

In this section, we presented a formalism for modeling evolution process with noise using the Brownian motion. We saw that the classical notion of solution doesn't work and we proposed a

new interpretation. As a result, it was necessary to construct a new theory of integration and, with that, the Itô and Stratonovich integrals. We hence studied the properties of these integrals.

Given that, we can advance on more elaborated questions, for example: does the equation (1.1) has a solution? Under which hypothesis? How can we study the behavior of the solution?

Chapter 4

The Itô Formula and the Martingale Representation Theorem

4.1 1-Dimensional Itô Processes

Itô formula: Integral version of the Chain Rule

Idea: introduce **Itô processes (stochastic integrals)** as sums of a dB_s– and a ds-integral. This family of integrals is stable under smooth maps (what does this mean?)

> **Definition 4.1.1** (1-dimensional Itô processes)**.** B_t 1-dimensional Brownian motion on $(\Omega, \mathfrak{F}, \mathbb{P})$. A 1-dimensional Itô process (or stochastic integral) is a stochastic process X_t on $(\Omega, \mathfrak{F}, \mathbb{P})$ of the form
>
> $$X_t = X_0 + \int_0^t u(s, \omega) \, ds + \int_0^t v(s, \omega) \, dB_s \tag{4.1}$$
>
> where $v \in \mathfrak{W}_{\mathfrak{H}}$ and
>
> $$\mathbb{P}\left[\int_0^t v(s, \omega)^2 \, ds < \infty \text{ for all } t \geq 0\right] = 1$$
>
> and u is $\mathfrak{H}_t$-adapted (as in the condition 2') and
>
> $$\mathbb{P}\left[\int_0^t |u(s, \omega)| \, ds < \infty \text{ for all } t \geq 0\right] = 1$$

If X_t is an Itô process of this form, then the equation (4.1) is written in the differential form

$$dX_t = u dt + v dB_t$$

Hence, we can write

$$d\left(\frac{1}{2}B_t^2\right) = \frac{1}{2}dt + B_t dB_t$$

> **Theorem 4.1.1** (The 1-dimensional Itô formula)**.** X_t be an Itô process given by
>
> $$dX_t = u dt + v dB_t$$

and $g(t,x) \in \mathcal{C}^2([0,\infty) \times \mathbf{R})$. Then

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

is an Itô process, and

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)(dX_t)^2 \qquad (4.2)$$

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

$$dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0, \ dB_t \cdot dB_t = dt$$

Proof. Notice that the claim is equivalent to the expression

$$g(t, X_t) = g(0, X_0) + \int_0^t \left(\frac{\partial g}{\partial s}(s, X_s) + u_s \frac{\partial g}{\partial x}(s, X_s) + \frac{1}{2}v_s^2\frac{\partial^2 g}{\partial x^2}(s, X_s) \right) ds + \int_0^t v_s \frac{\partial g}{\partial x}(s, X_s) \ dB_s$$

where $u_s = u(s, \omega)$ and $v_s = v(s, \omega)$.

We'll assume that g and its derivatives above are bounded. In the general case, we can approximate g by $\mathcal{C}^2$ functions that converge uniformly to g.

Then write the Taylor expansion with respect to the arguments (in this case, t and X_t), taking the first-order terms for all arguments with zero quadratic variation (here, t) and up to the second order for every argument with non zero quadratic variation (here, X_t).

Notice that the terms with crossed Δt_j and ΔB_j go to zero by Itô isometry.

Also notice that the terms with crossed $(\Delta B_i)^2 - \Delta t_i$ and $(\Delta B_j)^2 - \Delta t_j$, $i \neq j$, vanish because they are indepedent. For the similar terms with $i = j$, the sum converges to a dt integral, showing that $(dB_t)^2 = dt$.

Since the remainder goes to zero as $\Delta t_j \to 0$, we complete the proof. □

Example 4.1.1. Consider again the following integral

$$I = \int_0^t B_s \ dB_s$$

Choose $X_t = B_t$ and $g(t,x) = \frac{1}{2}x^2$, then

$$Y_t = g(t, B_t) = \frac{1}{2}B_t^2$$

Then, by Itô's formula,

$$\begin{aligned} dY_t &= \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dB_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(dB_t)^2 \\ &= B_t dB_t + \frac{1}{2}(dB_t)^2 \\ &= B_t dB_t + \frac{1}{2}dt \end{aligned}$$

I.e.,

$$d\left(\frac{1}{2}B_t^2\right) = B_t dB_t + \frac{1}{2}dt$$

Put it another way,

$$\frac{1}{2}B_t^2 = \int_0^t B_s \ dB_s + \frac{1}{2}t$$

Example 4.1.2. Now consider

$$\int_0^t s \ dB_s$$

Put

$$g(t,x) = tx \text{ and } Y_t = g(t, B_t) = tB_t$$

By Itô's formula,

$$dY_t = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dB_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(dB_t)^2$$

i.e.

$$d(tB_t) = dY_t = B_t dt + t dB_t + \cdot 0$$

In the integral form,

$$tB_t = \int_0^t B_s \ ds + \int_0^t s \ dB_s$$

or

$$\int_0^t s \ dB_s = tB_t - \int_0^t B_s \ ds$$

This example is generalized in the next theorem.

Theorem 4.1.2 (Integration by parts)**.** Suppose that $f(s, \omega) = f(s)$ only depends on s and that f is continuous and of bounded variation in $[0, t]$. Then

$$\int_0^t f(s) \ dB_s = f(t)B_t - \int_0^t B_s \ df_s$$

4.2 The Multi-Dimensional Itô Formula

Definition 4.2.1 (n-dimensional Itô process)**.** Let $B(t, \omega) = (B_{t_1}(t, \omega), \ldots, B_m(t, \omega))$ denote m-dimensional Brownian motion. If each $u_i(t, \omega)$ and $v_{ij}(t, \omega)$, for $1 \leq i \leq n$ and $1 \leq j \leq m$, satisfies the conditions given in the definition of Itô process, then we can form n Itô processes

$$\begin{cases} dX_1 = u_1 dt + v_{11} dB_1 + \ldots + v_{1m} dB_m \\ \quad \vdots \\ dX_n = u_n dt + v_{n1} dB_1 + \ldots + v_{nm} dB_m \end{cases}$$

Or simply

$$dX(t) = u\,dt + v\,dB(t)$$

where

$$X(t) = \begin{bmatrix} X_1(t) \\ \vdots \\ X_n(t) \end{bmatrix}, \ u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \ v = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nm} \end{bmatrix}, \ dB(t) = \begin{bmatrix} dB_1(t) \\ \vdots \\ dB_m(t) \end{bmatrix}$$

The process $X(t)$ is called an n-**dimensional Itô process**.

Theorem 4.2.1 (General Itô Formula). Let

$$dX(t) = u\,dt + v\,dB(t)$$

be an n-dimensional Itô process. And let $g(t,x) = (g_1(t,x),\ldots,g_p(t,x))$ be a $\mathcal{C}^2$ map from $[0,\infty) \times \mathbf{R}^n$ into $\mathbf{R}^p$. Then the process

$$Y(t,\omega) = g(t,X(t))$$

is an Itô process, whose k-th component is given by

$$dY_k = \frac{\partial g_k}{\partial t}(t,X)\,dt + \sum_i \frac{\partial g_k}{\partial x_i}(t,X)\,dX_i + \frac{1}{2}\sum_{i,j} \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t,X)\,dX_i\,dX_j$$

where $dB_i\,dB_j = \delta_{ij}dt$, $dB_i\,dt = dt\,dB_i = 0$.

4.3 The Martingale Representation Theorem

Let $B(t)$ be an n-dimensional Brownian motion. We saw that if $v \in \mathfrak{V}^n$, then the Itô integral

$$X_t = X_0 + \int_0^t v(s,\omega)\,dB(s), \ t \geq 0$$

is always a martingale w.r.t. filtration $\mathfrak{F}_t^n$ and the probability measure $\mathbb{P}$.

In this section, we show that the converse is also true: Any $\mathfrak{F}_t^n$-martingale (w.r.t to $\mathbb{P}$) can be represented as an Itô integral. This result is called the martingale representation theorem, and is important in many applications, such as mathematical finance.

Lemma 4.3.1. Fix $T > 0$. The set of random variables

$$\{\varphi(B_{t_1},\ldots,B_{t_n})\}$$

where $t_i \in [0,T]$, $\varphi \in C_0^\infty(\mathbf{R}^n)$, and $n = 1,2,\ldots$ is dense in $L^2(\mathfrak{F}_T,\mathbb{P})$.

Lemma 4.3.2. The linear span of random variables of the type

$$\exp\left\{ \int_0^T h(t) \, dB_t(\omega) - \frac{1}{2} \int_0^T H^2(t) \, dt \right\}$$

where $h \in L^2[0, T]$ is dense in $L^2(\mathfrak{F}_T, \mathbb{P})$.

Suppose that $B(t)$ is n-dimensional. If $v(s, \omega) \in \mathfrak{V}^n(0, T)$, then the random variable

$$V(\omega) := \int_0^T v(t, \omega) \, dB(t)$$

is $\mathfrak{F}_t^n$-measurable and by the Itô isometry

$$\mathbb{E}[V^2] = \int_0^T \mathbb{E}[v^2(t, \cdot)] \, dt < \infty$$

so $V \in L^2(\mathfrak{F}_T, \mathbb{P})$.

Theorem 4.3.3 (Itô Representation Theorem)**.** Let $F \in L^2(\mathfrak{F}_T, \mathbb{P})$. Then there exists a unique stochastic process $f(t, \omega) \in \mathfrak{V}^n(0, T)$ such that

$$F(\omega) = \mathbb{E}[F] + \int_0^T f(t, \omega) \, dB(t)$$

Theorem 4.3.4 (The Martingale Representation Theorem)**.** Let $B(t)$ be an n-dimensional Brownian motion. Suppose M_t is an $\mathfrak{F}_t^{(n)}$-martingale w.r.t. $\mathbb{P}$ and that $M_t \in L^2(\mathbb{P})$ for all $t \geq 0$.

Then there exists a unique stochastic process $g(s, \omega)$ such that $g \in \mathfrak{V}^{(n)}(0, t)$ for all $t \geq 0$ and

$$M_t(\omega) = \mathbb{E}[M_0] + \int_0^t g(s, \omega) \, dB(s)$$

almost surely, for all $t \geq 0$.

Chapter 5

Stochastic Differential Equations

Stochastic differential equations model evolution processes with continuous random noise, and are of the form

$$\frac{dX_t}{dt} = b_t(X_t) + \sigma_t(X_t)W_t \tag{5.1}$$

where $b_t(x), \sigma_t(x) \in \mathbf{R}$ and W_t is the one-dimensional 'white noise'.

Inspired by the deterministic case, we propose as a solution a stochastic process X_t which satisfies the following identity

$$X_t = X_0 + \int_0^t b_s(X_s) \, ds + \int_0^t \sigma_s(X_s) \, dB_s \tag{5.2}$$

I.e.,

$$dX_t = b_t(X_t) \, dt + \sigma_t(X_t) \, dB_t \tag{5.3}$$

Notice that to get from (5.1) to (5.3) we replace W_t by $\frac{dB_t}{dt}$ and multiply by dt.

In this chapter, we address the following questions. When does a solution exists, and when is it unique? What are its properties? How to solve such an equation?

5.1 Solution Methods

We start this section by the following example, which shows an 'Itô correction' to the exponential.

> **Example 5.1.1** (Stochastic Exponential). Consider
>
> $$dU_t = U_t dX_t, \quad U_0 = 1 \iff U_t = 1 + \int_0^t U_s \, dX_s \tag{5.4}$$
>
> U is called the **stochastic exponential of** X. If X_t have finite variation, then the solution of the equation (5.4) is given by $U_t = e^{X_t}$. The proof follows imediatelly from the Itô's formula (see [Kle12]).

For the Brownian motion B_t, the stochastic exponential is given by

$$U_t = e^{B_t - \frac{1}{2}t} \tag{5.5}$$

Let $Y_t = B_t - \frac{1}{2}t$ and $g(t,x) = x - \frac{1}{2}t$. Then

$$dY_t = -\frac{1}{2}dt + dB_t$$

Noting that $U_t = e^{Y_t}$, define $g(t,x) = e^x$. Then $U_t = g(t,Y_t)$ and

$$dU_t = e^{Y_t}dY_t + \frac{1}{2}e^{Y_t}(dY_t)^2 = U_t\left(-\frac{1}{2}dt + dB_t\right) + \frac{1}{2}U_tdt$$
$$= U_tdB_t$$

A linear SDE in one dimension is of the form

$$dX_t = (\alpha_t + \beta_tX_t)dt + (\gamma_t + \delta_tX_t)dB_t \tag{5.6}$$

where $\alpha, \beta, \gamma, \delta$ are adapted processes and continuous functions of t.

Example 5.1.2 (Stochastic Exponential SDE)**.** If $\alpha_t = 0 = \gamma_t$, we have

$$dU_t = \beta_tU_tdt + \delta_tU_tdB_t$$

The solution for this equation is

$$U_t = U_0 \exp\left(\int_0^t \left(\beta_s - \frac{1}{2}\delta_s^2\right)ds + \int_0^t \delta_s\, dB_s\right)$$

To verify this, suppose for simplicity that $U_0 = 1$, and let

$$g(t,x) = \exp\left(\int_0^t \left(\beta_s - \frac{1}{2}\delta_s^2\right)ds + \int_0^t \delta_s\, dx\right)$$

Note that $U_t = g(t,B_t)$ and compute

$$\frac{\partial g}{\partial t} = \left(\beta_t - \frac{1}{2}\delta_t^2\right)U_t, \quad \frac{\partial g}{\partial x} = \delta_tU_t, \quad \frac{\partial^2 g}{\partial x^2} = \delta_t^2U_t$$

By Itô's formula,

$$dU_t = \left(\beta_t - \frac{1}{2}\delta_t^2\right)U_tdt + \delta_tU_tdB_t + \frac{1}{2}\delta_t^2U_tdt$$
$$= \beta_tU_tdt + \delta_tU_tdB_t$$

Another way to look at this equation is by remarking that it is the stochastic exponential of Y_t, i.e.,

$$dU_t = U_tdY_t \tag{5.7}$$

in which Y_t is an Itô process defined by

$$dY_t = \beta_t dt + \delta_t dB_t$$

Therefore,

$$U_t = U_0 \exp\left(Y_t - Y_0 - \frac{1}{2}[Y,Y](t)\right)$$

$$= U_0 \exp\left(\int_0^t \beta_s \, ds + \int_0^t \delta_s \, dB_s - \frac{1}{2}\int_0^t \delta_s^2 \, ds\right)$$

$$= U_0 \exp\left(\int_0^t \left(\beta_s - \frac{1}{2}\delta_s^2\right) ds + \int_0^t \delta_s \, dB_s\right)$$

Example 5.1.3 (Stochastic Logarithm)**.** How can we compute the integral $\int_0^t \frac{dX_s}{X_s}$? Applying Itô formula for $g(t,x) = \ln x$, where $x > 0$,

$$d(\ln X_t) = \frac{1}{X_t}dX_t + \frac{1}{2}\left(-\frac{1}{X_t^2}\right)(dX_t)^2$$

$$= \frac{dX_t}{X_t} - \frac{1}{2X_t^2}(dX_t)^2$$

Hence,

$$\frac{dX_t}{X_t} = d(\ln X_t) + \frac{1}{2X_t^2}(dX_t)^2 \tag{5.8}$$

Example 5.1.4 (Population Growth)**.** Consider the following Stochastic Differential Equation (SDE):

$$\frac{dN_t}{dt} = a_t N_t, \quad N_0 \in \mathbf{R}$$

where $a_t = r_t + \alpha W_t$.

If r_t is a constant r, by our previous discussion we have

$$\frac{dN_t}{N_t} = rdt + \alpha dB_t \iff dN_t = rN_t dt + \alpha N_t dB_t$$

In the integral form,

$$\int_0^t \frac{dN_s}{N_s} = rt + \alpha B_t$$

Using (5.8)

$$\frac{dN_t}{N_t} = d(\ln N_t) + \frac{1}{2N_t^2}\left(\alpha^2 N_t^2 dt\right) = d(\ln N_t) + \frac{1}{2}\alpha^2 dt$$

Hence,

$$\ln\left(\frac{N_t}{N_0}\right) = \left(r - \frac{1}{2}\alpha^2\right)t + \alpha B_t \iff N_t = N_0 \exp\left(\left(r - \frac{1}{2}\alpha^2\right)t + \alpha B_t\right)$$

Remark that if we used the Stratonovich interpretation, then the solution would be

$$N_t = N_0 \exp(rt + \alpha B_t)$$

Example 5.1.5 (Stock Prices). Let $P(t)$ denote the price of a stock at time t. We may assume that the relative change of price evolves according to the SDE

$$\frac{dP}{P} = \mu dt + \sigma dB_t \iff dP = \mu P dt + \sigma P dB_t$$

where $\mu > 0$ is the **drift**, and σ is the **volatility** of the stock.

By Itô's formula,

$$d(\ln P) = \frac{dP}{P} - \frac{1}{2}\frac{\sigma^2 P^2}{P^2} dt$$
$$= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dB_t$$

Hence,

$$P(t) = p_0 \exp\left(\sigma B_t + \left(\mu - \frac{\sigma^2}{2}\right)t\right)$$

Since the integral form of the SDE is

$$P(t) = p_0 + \int_0^t \mu P \, ds + \int_0^t \sigma P \, dB_t$$

and

$$\mathbb{E}\left(\int_0^t \sigma P \, dB_t\right) = 0$$

we have that

$$\mathbb{E}(P(t)) = p_0 + \int_0^t \mu \mathbb{E}(P(s)) \, ds$$

Hence,

$$\mathbb{E}(P(t)) = p_0 e^{\mu t} \quad \text{for } t \geq 0$$

The above example is a generalization of Example 5.1.4, where B_t is independent of N_0, i.e., there's no noise in a_t.

In this case, we obtain

$$\mathbb{E}[N_t] = \mathbb{E}[N_0]e^{rt}$$

For the Stratonovich solution,

$$\mathbb{E}[N_t] = \mathbb{E}[N_0]e^{\left(r + \frac{1}{2}\alpha^2\right)t}$$

Notice that for Itô solution we have

1. If $r > \frac{1}{2}\alpha^2$, then $N_t \to \infty$ as $t \to \infty$ a.s.

2. If $r < \frac{1}{2}\alpha^2$, then $N_t \to 0$ as $t \to \infty$ a.s.

3. If $r = \frac{1}{2}\alpha^2$, then N_t will fluctuate between arbitrary large and small values as $t \to \infty$ a.s.

Definition 5.1.1 (Geometric Brownian Motion)**.** Processes of the type

$$X_t = X_0 \exp(\mu t + \alpha B_t)$$

where μ, α are constants, are called **geometric Brownian motions**.

Example 5.1.6 (Electric Circuit)**.** Let $Q(t)$ be the charge at time t at a fixed point in a electric circuit. Then the charge satisfies the following differential equation

$$LQ''(t) + RQ'(t) + \frac{1}{C}Q(t) = F(t), \quad Q(0) = Q_0,\ Q'(0) = I_0$$

In which L is inductance, R is resistance, C is capacitance and $F(t)$ the potential source.

It may be the case that some coefficient, say F_t, are not deterministic, but of the form

$$F(t) = G(t) + \alpha W(t)$$

We reduce this second order equation to a system of first order equations by defining

$$X(t, \omega) = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} Q_t \\ Q'_t \end{bmatrix}$$

and then obtaining

$$\begin{cases} X'_1 = X_2 \\ LX'_2 = -RX_2 - \frac{1}{2}X_1 + G_t + \alpha W_t \end{cases} \tag{5.9}$$

Letting

$$dX = \begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 \\ -\frac{1}{CL} & -\frac{R}{L} \end{bmatrix}, \quad H_t = \begin{bmatrix} 0 \\ \frac{1}{L}G_t \end{bmatrix}, \quad K = \begin{bmatrix} 0 \\ \frac{\alpha}{L} \end{bmatrix}$$

we can write the system (5.9) in a matricial form

$$dX_t = AX_t dt + H_t dt + K dB_t \tag{5.10}$$

Now, rewrite (5.10) as

$$\exp(-At)dX_t - \exp(-At)AX_t dt = \exp(-At)[H_t dt + K dB_t] \tag{5.11}$$

Applying Itô's formula to $d(\exp(-At)X_t)$,

$$d(\exp(-At)X_t) = (-A)\exp(-At)X_t dt + \exp(-At)dX_t$$

and replacing it into (5.11) and using integration by parts,

$$\exp(-At)X_t - X_0 = \int_0^t \exp(-As)H_s\, ds + \int_0^t \exp(-As)K\, dB_s$$

i.e.,

$$X_t = \exp(At)\left(X_0 + \exp(-At)KB_t + \int_0^t \exp(-As)[H_s + AKB_s]\,ds\right)$$

For more solution methods, see [Eva12], [Gar88], and [Kle12].

5.2 Existence and Uniqueness

Theorem 5.2.1. Let $T > 0$ and $b : [0, T] \times \mathbf{R}^n \longrightarrow \mathbf{R}^n$, $\sigma : [0, T] \times \mathbf{R}^n \longrightarrow \mathbf{R}^{n \times m}$ are measurable functions.

If the following conditions are satisfied

1. Coefficients satisfy the linear growth condition

$$|b(t,x)| + |\sigma(t,x)| \leq C(1 + |x|), \quad x \in \mathbf{R}^n, \, t \in [0, T]$$

for some constant C;

2. Coefficients are locally Lipschitz in x uniformly in t

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq D|x-y|, \quad x,y \in \mathbf{R}^n, \, t \in [0, T]$$

for some constant D;

3. Z is a random variable which is independent of the σ-algebra $\mathfrak{F}_\infty^{(m)}$ generated by B_s ($s \geq 0$) and such that $\mathbb{E}[|Z|^2] < \infty$.

Then the stochastic differential equation

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dB_t, \quad 0 \leq t \leq T, \, X_0 = Z \tag{5.12}$$

has a unique t-continuous solution $X_t(\omega)$ with the property that $X_t(\omega)$ is adapted to the filtration $\mathfrak{F}_t^Z$ generated by Z and B_s ($s \leq t$), and

$$\mathbb{E}\left[\int_0^T |X_t|^2\,dt\right] < \infty$$

Proof. Existence: successive approximations (Picard iterations). By Gronwall's lemma, the Lipschitz condition and Itô isometry imply uniqueness. □

5.3 Weak and Strong Solutions

The solution that we found is called a **strong solution**, in which the version of Brownian motion is given, and the solution is $\mathfrak{F}_t^Z$-adapted.

However, what if we're to find a pair of processes $((\bar{X}_t, \bar{B}_t), \mathfrak{H}_t)$ on a probability space such that the differential form (5.12) holds? In this case, the solution $(\bar{X}_t, \bar{B}_t)$ is called a **weak solution**. Put another way, 'weak solutions are solutions in distribution' [Kle12, 137].

The uniqueness of the Theorem 5.2.1 is **strong** or **pathwise** uniqueness. **Weak uniqueness** means that any two solutions are identical in law (have the same finite-dimensional distributions).

Notice that any strong solution is a weak solution (a result from Yamada and Watanabe) and any solution is weakly unique. The next example shows that it is possible to have no strong solutions, but a weakly unique weak solution.

Example 5.3.1 (The Tanaka Equation). Consider

$$dX_t = \text{sign}(X_t)dB_t, \quad X_0 = 0 \tag{5.13}$$

Since $\sigma(t,x) = \text{sign}(x)$ is not continuous, it is not Lipschitz and hence the Theorem 5.2.1 does not apply. In fact, (5.13) has no strong solution.

Let X_t be any Brownian motion. Then

$$Y_t = \int_0^t \text{sign}(X_s)\, dX_s \iff dY_t = \text{sign}(X_t)dX_t$$

Hence,

$$dX_t = \text{sign}(X_t)dY_t$$

We'll see in Levy's theorem that Y_t is a Brownian motion (more than that, any weak solution is a Brownian motion) and X_t is a weakly unique weak solution.

Chapter 6

Diffusion Processes

This chapter studies the properties of solutions of stochastic differential equations. Since these solutions can be interpreted as in the physical Brownian motion, such stochastic processes are called **(Itô) diffusions**.

6.1 Basic Definitions

Throughout this chapter, consider a stochastic differential equation

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t \tag{6.1}$$

with $X_t \in \mathbf{R}^n$, $\mu(t,x) \in \mathbf{R}^n$, $\sigma(t,x) \in \mathbf{R}^{n \times m}$ and B_t is m-dimensional Brownian motion. We call μ the **drift coefficient** and σ (or $\frac{1}{2}\sigma\sigma^T$) the **diffusion coefficient**.

> **Definition 6.1.1** (Itô Diffusion). A (time-homogeneous) **Itô diffusion** is a stochastic process
>
> $$X_t(\omega) : [0, \infty] \times \Omega \longrightarrow \mathbf{R}^n$$
>
> satisfying a stochastic differential equation of the form
>
> $$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \quad t \geq s, \quad X_s = x$$
>
> satisfying the conditions of the Theorem 5.2.1, which in this case simplify to:
>
> $$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq D|x-y|, \quad x,y \in \mathbf{R}^n$$
>
> with $|\sigma|^2 = \sum |\sigma_{ij}|^2$.
>
> We denote the solution as $X_t = X_t^{s,x}$ and, for $s = 0$, we write $X_t^x = X_t^{0,x}$.

Notice that the process $X_t(\omega)$ has the property of being **time-homogeneous**, i.e.

$$X_{s+h}^{s,x} = x + \int_s^{s+h} \mu(X_u^{s,x}) \, du + \int_s^{s+h} \mu(X_u^{s,x}) \, dB_u$$

$$= x + \int_0^h \mu(X_{s+v}^{s,x}) \, dv + \int_0^h \mu(X_{s+v}^{s,x}) \, d\overline{B}_v$$

with $u = s + v$, $\overline{B}_v = B_{s+v} - B_s$, $v \geq 0$.

Taking $s = 0$,

$$X_h^{0,x} = x + \int_0^h \mu(X_v^{0,x}) \, dv + \int_0^h \mu(X_v^{0,x}) \, dB_v$$

Since the processes $\{\overline{B}_v\}$ and $\{B_v\}$ have the same P^0-distribution, by weak uniqueness of the solution, we have that $\{X_{s+h}^{s,x}\}_{h \geq 0}$ and $\{X_h^{0,x}\}_{h \geq 0}$ have the same P^0-distributions. Thus, $\{X_t\}_{t \geq 0}$ is time-homogeneous.

Let us now introduce new probability laws Q^x that give the distribution of $\{X_t\}_{t \geq 0}$ assuming that $X_0 = x$. Let $\mathfrak{M}_\infty$ be the σ-algebra generated by the random variables $\omega \longrightarrow X_t(\omega) = X_t^y(\omega)$, with $t \geq 0$ and $y \in \mathbf{R}^n$. We define Q^x on $\mathfrak{M}$ as

$$Q^x[X_{t_1} \in E_1, \ldots, X_{t_k} \in E_k] = P^0[X_{t_1}^x \in E_1, \ldots, X_{t_k}^x \in E_k]$$

where $E_1 \subset \mathbf{R}^n$ are Borel sets and $1 \leq i \leq k$.

Also let $\mathfrak{F}_t^{(m)}$ be the σ-algebra generated by $\{B_r : r \leq t\}$ and $\mathfrak{M}_t$ be the σ-algebra generated by $\{X_r : r \leq t\}$. Since X_t is measurable w.r.t. $\mathfrak{F}_t^{(m)}$, it follows that $\mathfrak{M}_t \subseteq \mathfrak{F}_t^{(m)}$.

6.2 The Markov Property

Another feature of the solution X_t is the Markov property, which says that, given the present state, the future is independent of the past. Put another way, given what happened up to time t, the future of the process is the same as the behavior obtained by starting the process at X_t. More precisely,

> **Definition 6.2.1** (Markov Property). For any $0 \leq s < t$,
>
> $$P[X_t \leq y \mid \mathfrak{F}_s] = P[X_t \leq y \mid X_s] \quad \text{a.s.}$$

The next results states that a solution to an SDE X_t is a Markov process w.r.t. $\{\mathfrak{F}_t^{(m)}\}_{t \geq 0}$.

> **Theorem 6.2.1** (Markov Property for Itô Diffusions). Let $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ be a bounded Borel function. For $t, h \geq 0$,
>
> $$\mathbb{E}^x[f(X_{t+h}) \mid \mathfrak{F}_t^{(m)}](\omega) = \mathbb{E}^{X_t(\omega)}[f(X_h)] \tag{6.2}$$
>
> where $\mathbb{E}^x$ denotes the expected value w.r.t. the probability measure Q^x, i.e., $\mathbb{E}^x[f(X_h)] = [f(X_h^x)]$ w.r.t. the probability measure P^0.

> **Proof.** By uniqueness (of SDE solutions), $X_r(\omega) = X_r^{t,X_t}(\omega)$.
>
> Define
>
> $$F(x, t, r, \omega) = X_r^{t,x}(\omega), \quad r \geq t$$
>
> and use this to write $X_r(\omega) = F(X_t, t, r, \omega)$.
>
> Using that $\omega \longrightarrow F(x, t, r, \omega)$ is $\mathfrak{F}_t^{(m)}$-independent and the time-homogeneity, we have that
>
> $$\mathbb{E}[f(F(X_t, t, t+h, \omega)) \mid \mathfrak{F}_t^{(m)}] = \mathbb{E}[f(F(X_t, 0, h, \omega))]$$

Let $g(t, \omega) = f \circ F(x, t, t+h, \omega)$ and notice that it is measurable. Thus, we can approximate g pointwise boundedly by

$$\sum_{k=1}^{m} \varphi_k(x) \psi_k(\omega)$$

Compute

$$\mathbb{E}[g(X_t, \omega) \mid \mathfrak{F}_t^{(m)}] = \mathbb{E}[g(X_t, \omega)]$$

Using that $\{X_r\}$ is time-homogeneous, the result follows. $\qquad\square$

Remark. Since $\mathfrak{M}_t \subseteq \mathfrak{F}_t^{(m)}$, we have that X_t is also a Markov process w.r.t. $\{\mathfrak{M}_t\}_{t \geq 0}$. With the fact $\mathbb{E}^{X_t}[f(X_h)]$ is $\mathfrak{M}_t$-measurable and the Theorem 2.3.3 items 3 and 6, it follows that

$$\mathbb{E}^x[f(X_{t+h}) \mid \mathfrak{M}_t] = \mathbb{E}^x[\mathbb{E}^x[f(X_{t+h}) \mid \mathfrak{F}_t^{(m)}] \mid \mathfrak{M}_t]$$
$$= \mathbb{E}^x[\mathbb{E}^{X_t}[f(X_h)] \mid \mathfrak{M}_t]$$
$$= \mathbb{E}^{X_t}[f(X_h)]$$

What happens if we replace the time t in (6.2) by a random time $\tau(\omega)$, i.e., a stopping time? The Strong Markov Property states that the equation (6.2) continues to hold, but before proving it, we need to make precise some terms.

To define a stopping time, the intuition is that we should be able to verify whether or not $\tau \leq t$ has occurred using the information contained in an increasing family of σ-algebras.

Definition 6.2.2 (Stopping time). Let $\{\mathfrak{N}_t\}$ be an increasing family of σ-algebras of subsets of Ω. A (strict) **stopping time** (also called a **Markov time**) w.r.t. $\{\mathfrak{N}_t\}$ is a function $\tau : \Omega \longrightarrow [0, \infty]$ such that $\{\omega : \tau(\omega) \leq t\} \in \mathfrak{N}_t$ for all $t \geq 0$.

Definition 6.2.3 (First Exit Time). Given $H \subset \mathbf{R}^n$, we define the **first exist time** from H as

$$\tau_H = \inf\{t > 0 : X_t \notin H\}$$

Remark. The family $\{\mathfrak{M}_t\}$ is right-continuous and τ_H is a stopping time for any Borel set H (assuming that the sets of measure zero are included in $\mathfrak{M}_t$).

Definition 6.2.4. Let τ be a stopping time w.r.t. $\{\mathfrak{N}_t\}$, and let $\mathfrak{N}_\infty$ be the smallest σ-algebra containing $\mathfrak{N}_t$ for all $t \geq 0$. Then the σ-algebra $\mathfrak{N}_t$ consists of all sets $N \in \mathfrak{N}_\infty$ such that

$$N \bigcap \{\tau \leq t\} \in \mathfrak{N}_t, \quad \forall\, t \geq 0$$

Remark. If $\mathfrak{N}_t = \mathfrak{M}_t$, we can describe $\mathfrak{M}_t$ as the σ-algebra generated by $\{X_{\min(s,\tau)} : s \geq 0\}$.

If $\mathfrak{N}_t = \mathfrak{F}_t^{(m)}$, we can describe $\mathfrak{F}_t^{(m)}$ as the σ-algebra generated by $\{B_{s \wedge \tau} : s \geq 0\}$.

Theorem 6.2.2 (Strong Markov Property for Itô Diffusions). Let f be a bounded Borel function defined on $\mathbf{R}^n$, τ a stopping time w.r.t. $\mathfrak{F}_t^{(m)}$, $\tau < \infty$ a.s. then

$$\mathbb{E}^x[f(X_{\tau+h}) \mid \mathfrak{F}_t^{(m)}]_{(\omega)} = \mathbb{E}^{X_\tau(\omega)}[f(X_h)], \quad \forall\, h \geq 0 \qquad (6.3)$$

Proof. The idea here is to imitate the proof of the Theorem 6.2.1. □

The equation (6.3) can be extended as follows.

Corollary 6.2.3. If $f_1, \ldots, f_k$ are bounded Borel functions on $\mathbf{R}^n$, τ a stopping time w.r.t. $\mathfrak{F}_t^{(m)}$, $\tau < \infty$ a.s. then

$$\mathbb{E}^x[f_1(X_{\tau+h_1})f_2(X_{\tau+h_2})\cdots f_k(X_{\tau+h_k}) \mid \mathfrak{F}_t^{(m)}] = \mathbb{E}^{X_\tau}[f_1(X_{\tau+h_1})f_2(X_{\tau+h_2})\cdots f_k(X_{\tau+h_k})] \tag{6.4}$$

Proof. By induction on k. □

6.2.1 Hitting distribution and Harmonic measure

We end this section by applying the concept of the shift operator to the expected hitting time to a boundary and the probabilistic representation of Harmonic functions.

Definition 6.2.5 (Shift Operator). Let $\mathfrak{H}$ be the set of all real $\mathfrak{M}_\infty$-measurable functions, and $t \geq 0$. The **shift operator**

$$\theta_t : \mathfrak{H} \longrightarrow \mathfrak{H}$$

as follows.

If $\eta = g_1(X_{t_1})\cdots g_k(X_{t_k})$, with each g_i Boreal measurable and $t_i \geq 0$, we define

$$\theta_t \eta = g_1(X_{t_1+t})\cdots g_k(X_{t_k+t})$$

And we extend to all functions in $\mathfrak{H}$ in a natural way: by taking limits of sums of such functions.

It follows from (6.4) that

$$\mathbb{E}^x[\theta_\tau \eta \mid \mathfrak{F}_\tau^{(m)}] = \mathbb{E}^{X_\tau}[\eta] \tag{6.5}$$

for all stopping times τ and all bounded $\eta \in \mathfrak{H}$, with

$$(\theta_\tau \eta)(\omega) = (\theta_t \eta)(\omega), \quad \text{if } \tau(\omega) = t$$

Theorem 6.2.4. Let $H \subset \mathbf{R}^n$ be measurable, τ_H be the first exit time from H for an Itô diffusion X_t, α another stopping time, g a bounded continuous function on $\mathbf{R}^n$ and define

$$\eta = g(X_{\tau_H})\chi_{\{\tau_H < \infty\}}, \quad \tau_H^\alpha = \inf\{t > \alpha : X_t \notin H\}$$

Then

$$\theta_\alpha \eta \chi_{\{\alpha < \infty\}} = g(X_{\tau_H^\alpha})\chi_{\{\tau_H^\alpha < \infty\}}$$

Proof. First approximate η by functions $\eta^{(k)}$.

Then compute $\theta_t \chi_{[t_J, t_{j+1})}(\tau_H) = \chi_{[t_j+t, t_{j+1}+t]}(\tau_H^t)$.

Finally, notice that $\theta_t \eta = g(X_{\tau_H^t})\chi_{\{\tau_H^t < \infty\}}$. □

In particular, if $\alpha = \tau_G$, with $G \subset\subset H$ (i.e. $\overline{G}$ is compact and $\overline{G} \subset H$) measurable, and $< \infty$ a.s. w.r.t. Q^x, then $\tau_H^\alpha = \tau_H$ and

$$\theta_{\tau_G} g(X_{\tau_H}) = g(X_{\tau_H}) \tag{6.6}$$

Thus, if f is a bounded measurable function, from (6.5) and (6.6) we have that

$$\mathbb{E}^x[f(X_{\tau_H})] = \mathbb{E}^x[\mathbb{E}^{X_{\tau_G}}[f(X_{\tau_H})]] = \int_{\partial G} \mathbb{E}^y[f(X_{\tau_H})]Q^x[X_{\tau_G} \in dy]$$

for $x \in G$.

> **Definition 6.2.6** (Harmonic Measure). The **harmonic measure** of X on ∂G, denoted by μ_G^x, is defined as
>
> $$\mu_G^x(F) = Q^x[X_{\tau_G} \in F], \quad \text{for } F \subset \partial G, \ x \in G$$

The function

$$\varphi(x) = \mathbb{E}^x[f(X_{\tau_H})]$$

satisfies the **mean value property**:

$$\varphi(x) = \int_{\partial G} \varphi(y) \, d\mu_G^x(y), \quad \forall \, x \in G \tag{6.7}$$

for all Borel sets $G \subset\subset H$.

6.3 The Generator of an Itô Diffusion

In this section, we associate a second order partial differential operator to an Itô diffusion.

> **Definition 6.3.1** (Generator). Let $\{X_t\}$ be a time-homogeneous Itô diffusion in $\mathbf{R}^n$. The (infinitesimal) **generator** A of X_t is defined by
>
> $$A f(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t}, \quad x \in \mathbf{R}^n \tag{6.8}$$
>
> The set of functions $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ such that the limit exists at x is denoted by $\mathfrak{D}_A(x)$, and $\mathfrak{D}_A$ denotes the set of functions for which the limit exists for all $x \in \mathbf{R}^n$.

Now consider the differential operator associated with the SDE (6.1)

$$L f(x) = \sum_i \mu_i(x) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} (\sigma \sigma^\mathrm{T})_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}$$

But how are the generator and the coefficients of the stochastic differential equation related? We'll prove that A and L coincide on $\mathcal{C}_0^2(\mathbf{R}^n)$. To show this, we need the following lemma.

> **Lemma 6.3.1.** Let $Y_t = Y_t^x$ be an Itô process in $\mathbf{R}^n$ of the form
>
> $$Y_t^x(\omega) = x + \int_0^t u(s, \omega) \, ds + \int_0^t v(s, \omega) \, dB_s(\omega)$$
>
> where B is m-dimensional.

And let $f \in \mathcal{C}_0^2(\mathbf{R}^n)$ (notice that this means that f has compact support), τ be a stopping time w.r.t. $\{\mathfrak{F}_t^{(m)}\}$ and assume that $\mathbb{E}^x[\tau] < \infty$. Also assume that $u(t, \omega)$ and $v(t, \omega)$ are bounded on the set of (t, ω) such that $Y(t, \omega)$ belongs to the support of f. Then

$$\mathbb{E}^x[f(Y_\tau)] = f(x) + \mathbb{E}^x\left[\int_0^\tau \left(\sum_i u_i(s, \omega)\frac{\partial f}{\partial x_i}(Y_s) + \frac{1}{2}\sum_{i,j}(vv^{\mathrm{T}})_{i,j}(s, \omega)\frac{\partial^2 f}{\partial x_i \partial x_j}(Y_s)\right) ds\right]$$

where $\mathbb{E}^x$ is the expectation w.r.t. the probability law R^x for Y_t starting at x:

$$R^x[Y_{t_1} \in F_1, \ldots, Y_{t_k} \in F_k] = P^0[Y_{t_1}^x \in F_1, \ldots, Y_{t_k}^x \in F_k], \quad F_i \text{ are Borel sets}$$

Proof. First, apply Itô's formula to $f(Y)$.

Notice that $(vdB)_i(vdB)_j = (vv^{\mathrm{T}})_{ij}dt$ and use this to write an expression for $f(Y_t)$.

Compute the expected value of $f(Y_\tau)$.

Using that g is a bounded Borel function, notice that the expected value of the dB part goes to zero (separate the measurable functions and write the integral as the difference of the measurable and non-measurable parts). $\qquad\square$

Theorem 6.3.2 (Formula for the generator)**.** Let X_t be the Itô diffusion

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$$

If $f \in \mathcal{C}_0^2(\mathbf{R}^n)$, then $f \in \mathcal{D}_A$ and

$$Af(x) = \sum_i \mu_i(x)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j}(\sigma\sigma^{\mathrm{T}})_{i,j}(x)\frac{\partial^2 f}{\partial x_i \partial x_j} \qquad (6.9)$$

Proof. Follows from the Lemma 6.3.1. $\qquad\square$

Example 6.3.1 (Standard Brownian Motion)**.** The standard n-dimensional Brownian motion B_t, which satisfies the SDE $dX_t = dB_t$, i.e., with $\mu = 0$ and $\sigma = I_n$, has the generator

$$Af = \frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x_i^2} = \frac{1}{2}\Delta, \quad f = f(x_1, \ldots, x_n) \in \mathcal{C}_0^2(\mathbf{R}^n)$$

Example 6.3.2 (Graph of Brownian Motion)**.** Consider the two-dimensional process X_t satisfying

$$dX_t = \begin{pmatrix} dt \\ dB_t \end{pmatrix}$$

i.e., with $\mu = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\sigma = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

X_t may be interpreted as the graph of Brownian motion. Its generator is given by

$$Af = \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}, \quad f = f(t, x) \in \mathcal{C}_0^2(\mathbf{R}^n)$$

6.4 Dynkin's Formula

Using the theory we already have, we can easily draw a useful formula.

Theorem 6.4.1 (Dynkin's Formula). Let $f \in \mathcal{C}_0^2(\mathbf{R}^n)$ and suppose that τ is a stopping time with $\mathbb{E}^x[\tau] < \infty$. Then

$$\mathbb{E}^x[f(X_\tau)] = f(x) + \mathbb{E}^x\left[\int_0^\tau \mathrm{A}f(X_s)\, ds\right] \tag{6.10}$$

Proof. Follows from combining Lemma 6.3.1 with (6.9). □

Notice that, if τ is the first exit time of a bounded set, then the result holds for any function $f \in \mathcal{C}^2$.

Example 6.4.1. Consider an n-dimensional Brownian motion $\mathrm{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_n)$ starting at $a = (a_1, \ldots, a_n)$ with $|a| < \mathrm{R}$. Let us compute the expected value of the first exit time τ_K of B from the ball

$$\mathrm{K} = \mathrm{K}_\mathrm{R} = \{x \in \mathbf{R}^n : |x| < \mathrm{R}\}$$

Choose an integer k and apply Dynkin's formula with $\mathrm{X} = \mathrm{B}$, $\tau = \sigma_k = \min(k, \tau_K)$, and $f \in \mathcal{C}_0^2$ such that $f(x) = |x|^2$ for $|x| \leq \mathrm{R}$:

$$\mathbb{E}^a[f(\mathrm{B}_{\sigma_k})] = f(a) + \mathbb{E}^a\left[\int_0^{\sigma_k} \frac{1}{2}\Delta f(\mathrm{B}_s)\, ds\right]$$

$$= |a|^2 + \mathbb{E}^a\left[\int_0^{\sigma_k} n\, ds\right]$$

$$= |a|^2 + n\mathbb{E}^a[\sigma_k]$$

Since $\mathbb{E}^a[f(\mathrm{B}_{\sigma_k})] \leq \mathrm{R}^2$,

$$\mathbb{E}^a[\sigma_k] \leq \frac{1}{n}(\mathrm{R}^2 - |a|^2), \quad \forall\, k$$

Taking $k \to \infty$, we have that $\tau_k = \lim \sigma_k < \infty$ a.s. and

$$\mathbb{E}^a[\tau_K] = \frac{1}{n}(\mathrm{R}^2 - |a|^2)$$

Now assume that $n \geq 2$ and $|b| > \mathrm{R}$. We'll compute the probability that B starting at b ever hits the ball K.

Let α_k be the first exit time from the annulus

$$\mathrm{A}_k = \{x : \mathrm{R} < |x| < 2^k\mathrm{R}\}, \quad k = 1, 2, \ldots$$

and put

$$\mathrm{T}_K = \inf\{t > 0 : \mathrm{B}_t \in \mathrm{K}\}$$

We define $f = f_{n,k} \in \mathcal{C}^2$ with compact support such that

$$\mathrm{R} \le |x| \le 2^k \mathrm{R} \implies f(x) = \begin{cases} -\log |x|, & \text{when } n = 2 \\ |x|^{2-n}, & \text{when } n > 2 \end{cases}$$

Since $\Delta f = 0$ in A_k, by Dynkin's formula,

$$\mathbb{E}^b[f(\mathrm{B}_{\alpha_k})] = f(b), \quad \forall\, k \tag{6.11}$$

Define

$$p_k = \mathrm{P}^b[\,|\mathrm{B}_{\alpha_k}| = \mathrm{R}] \quad \text{and} \quad q_k = \mathrm{P}^b[\,|\mathrm{B}_{\alpha_k}| = 2^k \mathrm{R}]$$

In the case $n = 2$, by (6.11) and the definition of f,

$$-\log \mathrm{R} \cdot p_k - (\log \mathrm{R} + k \log 2) \cdot q_k = -\log |b|, \quad \forall\, k$$

Therefore, $\lim_{k \to \infty} q_k = 0$, and thus

$$\mathrm{P}^b[\mathrm{T_K} < \infty] = 1$$

i.e., Brownian motion is **recurrent** in $\mathbf{R}^2$.

In the case $n > 2$,

$$p_k \mathrm{R}^{2-n} + q_k (2^k \mathrm{R})^{2-n} = |b|^{2-n}$$

Since $q_k \in [0, 1]$,

$$\lim_{k \to \infty} p_k = \mathrm{P}^b[\mathrm{T_K} < \infty] = \left(\frac{|b|}{\mathrm{R}} \right)^{2-n}$$

i.e., Brownian motion is **transient** in $\mathbf{R}^n$ for $n > 2$.

6.5 The Characteristic Operator

We now introduce an operator closely related to the generator, which is often more suitable, such as in the Dirichlet problem.

Definition 6.5.1 (Characteristic Operator)**.** Let $\{\mathrm{X}_t\}$ be an Itô diffusion. The **characteristic operator** $\mathfrak{A} = \mathfrak{A}_\mathrm{X}$ of $\{\mathrm{X}_t\}$ is defined by

$$\mathfrak{A} f(x) = \lim_{\mathrm{U} \downarrow x} \frac{\mathbb{E}^x[f(\mathrm{X}_{\tau_\mathrm{U}})] - f(x)}{\mathbb{E}^x[\tau_\mathrm{U}]}, \quad x \in \mathbf{R}^n \tag{6.12}$$

where each U is an open set U_k decreasing to the point x in the sense that $\mathrm{U}_{k+1} \subset \mathrm{U}_k$ and $\bigcap_k \mathrm{U}_k = \{x\}$, and $\tau_\mathrm{U} = \inf\{t > 0 : \mathrm{X}_t \notin \mathrm{U}\}$ is the first exit time from U for X_t.

The set of functions $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ such that the limit (6.12) exists for all $x \in \mathbf{R}^n$ and all $\{\mathrm{U}_k\}$ is denoted by $\mathfrak{D}_{\mathfrak{A}}$.

If $\mathbb{E}^x[\tau_\mathrm{U}] = \infty$ for all open sets $\mathrm{U} \ni x$, we define $\mathfrak{A} f(x) = 0$.

In fact, $\mathfrak{D}_A \subseteq \mathfrak{D}_{\mathfrak{A}}$ and

$$A f = \mathfrak{A} f \quad \forall f \in \mathfrak{D}_A$$

i.e., the characteristic operator is a generalization of the generator.

We will prove that $\mathfrak{A}_X$ and L_X coincide on $\mathcal{C}^2$. Before that, we need a property of exit times.

> **Definition 6.5.2** (Trap)**.** A point $x \in \mathbf{R}^n$ is called a **trap** for $\{X_t\}$ if
>
> $$Q^x[\{X_t = t \text{ for all } t\}] = 1$$
>
> Put another way, x is a trap iff. $\tau_{\{x\}} = \infty$ a.s. w.r.t. Q^x.

For example, if $\mu(x_0) = \sigma(x_0) = 0$, then x_0 is a trap for X_t by strong uniqueness of X_t. The intuition here is that the process never leaves the point.

> **Lemma 6.5.1.** If x is not a trap for X_t, then there exists an open set U containing x such that $\mathbb{E}^x[\tau_U] < \infty$.

In words, if the point is not a trap, there exists an open set containing it such that the process starting at it leaves it in finite time.

> **Theorem 6.5.2.** If $f \in \mathcal{C}^2$, then $f \in \mathfrak{D}_A$ and $\mathfrak{A} f = L f$.

> **Proof.** First, we need to show that if x is a trap for $\{X_t\}$, then $\mathfrak{A} f(x) = 0$. To do that, choose a bounded open set containing x. Modifying f to f_0 outside V such that $f_0 \in \mathcal{C}_0^2(\mathbf{R}^n)$, this claim follows.
>
> Now, if x is not a trap, choose a bounded open set U containing x such that $\mathbb{E}^x[\tau_U] < \infty$. Applying Dynkin's formula and using that $L f$ is continuous, we have the result. □

> **Remark.** From the discussions in this chapter, we obtained that an Itô diffusion is
>
> 1. Continuous;
>
> 2. A strong Markov process;
>
> 3. Such that the domain of definition of its characteristic operator includes $\mathcal{C}^2$.
>
> This means that an Itô diffusion is a diffusion in the sense of Dynkin.

Appendix A

Measure Theory

A.1 Measurable Spaces

Given a set S, on what collection $\mathfrak{S}$ of subsets of S are suitable to be a domain of measure?

Definition A.1.1. Let S be a set and $\mathfrak{S}$ be a family of subsets of S. Then $\mathfrak{S}$ is called a σ-**algebra** on S if

1. $\emptyset, S \in \mathfrak{S}$.

2. If $A \in \mathfrak{S}$, then $A^c \in \mathfrak{S}$.

3. If $A_1, A_2, \ldots \in \mathfrak{S}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{S}$.

The pair $(S, \mathfrak{S})$ is said to be a **measurable space**, and any subset $A \subseteq \mathfrak{S}$ is called an $\mathfrak{S}$-**measurable set**.

For any set S and any collection $\mathfrak{A}$ of subsets of S there is at least one σ-algebra containing $\mathfrak{A}$: the family of all subsets of S. Taking the intersection of all the σ-algebras containing $\mathfrak{A}$, we obtain the smallest σ-algebra containing $\mathfrak{A}$, which is called the σ-**algebra generated by** $\mathfrak{A}$.

Particularlly, the smallest σ-algebra containing all of the open sets of **R** is called the **Boreal algebra** for **R** and is denoted by $\mathfrak{B}$. Any set in $\mathfrak{B}$ is called a **Borel set**.

A.2 Measures

Now, how can we assign a size (or a probability) to all the sets in $\mathfrak{S}$?

Definition A.2.1. Let $(S, \mathfrak{S})$ be a measurable space. A **measure** is an extended real-valued function $\mu : \mathfrak{S} \longrightarrow \overline{\mathbf{R}}$ such that

1. $\mu(\emptyset) = 0$;

2. $\mu(A) \geq 0$, for all $A \in \mathfrak{S}$;

3. If $(A_n)_{n=1}^\infty$ is a countable, disjoint sequence of subsets in $\mathfrak{S}$, then $\mu(\bigcup_{n=1}^\infty A_n) = \sum_{n=1}^\infty \mu(A_n)$.

The triple $(S, \mathfrak{S}, \mu)$ is called a **measure space**.

A measure is nonnegative, assigns zero to the null set, and is **countably additive**. If $\mu(S) < \infty$, then μ is finite.

We say that a proposition holds **almost everywhere (a.e.)** if there exists a set $A \in \mathfrak{S}$ with $\mu(A) = 0$, such that the proposition holds on the complement of A. Intuitively, the proposition holds everywhere except on sets of measure zero.

For example, a sequence of functions (f_n) on S converges a.e. to a function f if there exists $A \in \mathfrak{S}$ with $\mu(A) = 0$ such that $\lim_{n \to \infty} f_n(x) = f(x)$ for all $x \in A^c$.

If $\mu(S) = 1$, then μ is a **probability measure** and $(S, \mathfrak{S}, \mu)$ is called a **probability space**. Any measurable set $A \in \mathfrak{S}$ is called and **event** and $\mu(A)$ is the **probability of the event** A. In a probability space, the phrase **almost surely (a.s.)** is used interchangeably with 'almost everywhere'.

Notice that if $A, B \in \mathfrak{S}$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$. If $\mu(A) < \infty$, then $\mu(B \backslash A) = \mu(B) - \mu(A)$.

Theorem A.2.1. Let $(S, \mathfrak{S}, \mu)$ be a measure space.

1. If $(A_n)_{n=1}^\infty$ is an increasing sequence in $\mathfrak{S}$ (i.e. if $A_n \subseteq A_{n+1}$ for all n), then

$$\mu\left(\bigcup_{n=1}^\infty A_n \right) = \lim_{n \to \infty} \mu(A_n)$$

2. If $(B_n)_{n=1}^\infty$ is a decreasing sequence in $\mathfrak{S}$ (i.e. if $B_{n+1} \subseteq B_n$ for all n) and if $\mu(B_m) < \infty$ for some m, then

$$\mu\left(\bigcap_{n=1}^\infty B_n \right) = \lim_{n \to \infty} \mu(B_n)$$

First, we define a measure on a small family of sets and then present an extension theorem.

Definition A.2.2. Let S be a set and $\mathfrak{A}$ be a family of subsets of S. Then $\mathfrak{A}$ is called an **algebra** if

1. $\emptyset, S \in \mathfrak{A}$;

2. $A \in \mathfrak{A}$ implies $A^c \in \mathfrak{A}$;

3. $A_1, A_2, \ldots, A_n \in \mathfrak{A}$ implies $\bigcup_{i=1}^n A_i \in \mathfrak{A}$.

I.e., an algebra is closed under complementation and finite union. On an algebra, the idea of measure is very much similar to the definition on a σ-algebra.

Definition A.2.3. Let S be a set, and let $\mathfrak{A}$ be an algebra of subsets of S. A **measure** is a real-valued function $\mu : \mathfrak{A} \longrightarrow \overline{\mathbf{R}}$ such that

1. $\mu(\emptyset) = 0$;

2. $\mu(A) \geq 0$, for all $A \in \mathfrak{A}$;

3. If $(A_n)_{n=1}^{\infty}$ is a disjoint sequence of sets in $\mathfrak{A}$ with $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$, then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.

Notice that last item is different from the definition on a σ-algebra. Here, we require that the union is contained in the algebra.

Although it is easier to define measures on algebras, it is more convenient to work with σ-algebras.

Theorem A.2.2 (Caratheodory Extension Theorem). Let S be a set, $\mathfrak{A}$ an algebra of its subsets, and μ a measure on $\mathfrak{A}$. Let $\mathfrak{S}$ be the smallest σ-algebra containing $\mathfrak{A}$. Then there exists a measure μ^* on $\mathfrak{S}$ such that $\mu(A) = \mu^*(A)$, for all $A \in \mathfrak{A}$.

To rule out the possibility of more than one extension of μ to $\mathfrak{S}$, we'll need the following definition

Definition A.2.4. Let S be a set, $\mathfrak{A}$ an algebra of its subsets, and μ a measure on $\mathfrak{A}$. If there is a countable sequence of sets $(A_i)_{i=1}^{\infty} \in \mathfrak{A}$ with $\mu(A_i) < \infty$ for all i, and $S = \bigcup_{i=1}^{\infty} A_i$, then μ is called σ-**finite**.

By definition, any probability measure is σ-finite. And the next theorem shows that the extension of a σ-finite measure is unique.

Theorem A.2.3 (Hahn Extension Theorem). Let S, $\mathfrak{A}$, μ and $\mathfrak{S}$ be as specified in Caratheodory Extension Theorem. If μ is σ-finite, then the extension μ^* to $\mathfrak{S}$ is unique.

Definition A.2.5. Let $(S, \mathfrak{S}, \mu)$ be a measure space, $A \in \mathfrak{S}$ be any set with measure zero, and let $C \subseteq A$. Denoting by $\mathfrak{C}$ the family of such sets C, i.e.,

$$\mathfrak{C} = \{C \subset S : C \subseteq A \text{ for some } A \in \mathfrak{S} \text{ with } \mu(A) = 0\}$$

The **completion** of $\mathfrak{S}$ is the family $\mathfrak{S}'$ constructed by starting with any set $B \in \mathfrak{S}$, and then adding and subtracting from it sets in $\mathfrak{C}$. That is

$$\mathfrak{S}' = \{B' \subseteq S : B' = (B \cup C_1) \setminus C_2, B \in \mathfrak{S}, C_1, C_2 \in \mathfrak{C}\}$$

Intuitively, $\mathfrak{S}'$ consists of all the subsetes of S that differ from a set in $\mathfrak{S}$ by a setset of a set of μ-measure zero. Using this definition, a measure μ on $(S, \mathfrak{S})$ can be extended to $(S, \mathfrak{S}')$.

A.3 Measurable Functions

Definition A.3.1. Given a measurable space $(S, \mathfrak{S})$, a real-valued function $f : S \longrightarrow \mathbf{R}$ is **measurable with respect to $\mathfrak{S}$** (or $\mathfrak{S}$-**measurable**) if

$$\{s \in S : f(s) \le a\} \in \mathfrak{S}, \forall a \in \mathbf{R}$$

If the space is a probability space, then f is called a **random variable**.

Some cases in which measurability can be easily verified:

- Any monotone or continuous function $f : \mathbf{R} \longrightarrow \mathbf{R}$ is measurable with respect to Borel sets.

- If S is a countable set and $\mathfrak{S}$ is a complete σ-algebra for S (i.e. contains all subsets of S), then all functions $f : S \longrightarrow \mathbf{R}$ are $\mathfrak{S}$-measurable.

In more general cases, the measurability of a function is established by showing that it is the limit of a sequence of 'simpler' functions. We'll do this progressively.

First, let $(S, \mathfrak{S})$ be a measurable space and consider the indicator function $\chi_A : S \longrightarrow \mathbf{R}$ of the from

$$\chi_A(s) = \begin{cases} 1, & \text{if } s \in A \\ 0, & \text{if } s \notin A \end{cases}$$

Clearly, χ_A is $\mathfrak{S}$-measurable iff. $A \in \mathfrak{S}$.

Now, consider the finite weighted sums of indicator functions:

$$\varphi(s) = \sum_{i=1}^{n} a_i \chi_{A_i}(s)$$

where $(A_i)_{i=1}^{n}$ is a sequence of subsets of S, and $(a_i)_{i=1}^{n}$ is a sequence of real numbers. Functions like this are called **simple functions**.

If the sets (A_i) form a partition of S and if all of the values a_i are distinct, then the previous equation is the **standard representation** of the function. If this is the case, then φ is measurable iff. each $A_i \in \mathfrak{S}$.

As we'll see in the next two theorems, the set of all measurable functions consists of those that are pointwise limits of measurable simple functions.

First, any function f that is pointwise limit of a sequence (f_n) of measurable function is itself measurable.

> **Theorem A.3.1** (Pointwise convergence preserves measurability)**.** Let $(S, \mathfrak{S})$ be a measurable space, and let (f_n) be a sequence of $\mathfrak{S}$-measurable functions converging pointwise to f, i.e.,
>
> $$\lim_{n \to \infty} f_n(s) = f(s), \ \forall s \in S$$
>
> Then f is also $\mathfrak{S}$-measurable.

Second, any measurable function f can be expressed as the pointwise limit of a sequence (φ_n) of measurable simple functions. If f is nonnegative, the sequence can be chosen to be strictly increasing. And if f is bounded, the sequence can be chosen to converge uniformly.

The idea is that we can construct a class of measurable functions by taking the class of measurable simple function and then closing this under pointwise convergence. The next result shows that this set contains all the measurable functions.

> **Theorem A.3.2** (Approximation of measurable functions by simple functions)**.** Let $(S, \mathfrak{S})$ be a measurable space. If $f : S \longrightarrow \mathbf{R}$ is $\mathfrak{S}$-measurable, then there exists a sequence of measurable simple functions (φ_n) such that $\varphi_n \longrightarrow f$ pointwise. If $0 \leq f$, then the sequence can be chosen so that
>
> $$0 \leq \varphi_n \leq \varphi_{n+1} \leq f$$
>
> for all n. If f is bounded, then the sequence can be chosen so that $\varphi_n \longrightarrow f$ uniformly.

With these two theorems, we see that a function is measurable iff. it is the pointwise limit of a sequence of measurable simple functions. The standard way to prove that a function is measurable is to find such an approximation.

Some important properties follow. Suppose that f and g are $\mathfrak{S}$-measurable functions on S and $c \in \mathbf{R}$. Then,

- The functions $f + g$, $f \cdot g$, $|f|$ and $c \cdot f$ are $\mathfrak{S}$-measurable.

- If (f_n) is a sequence of $\mathfrak{S}$-measurable functions, then $\inf f_n$, $\sup f_n$, $\liminf f_n$ and $\limsup f_n$ are all $\mathfrak{S}$-measurable.

- All continuous functions on $\mathbf{R}^l$ are $\mathfrak{B}^l$-measurable.

- The composition of Borel measurable functions are Borel measurable. However, this is not true for Lebesgue measurable functions.

Definition A.3.2. Let $(S, \mathfrak{S})$ and $(T, \mathfrak{T})$ be measurable spaces. Then the function $f : S \longrightarrow T$ is **measurable** if the inverse image of every measurable set is measurable, i.e.,

$$\{s \in S : f(s) \in A\} \in \mathfrak{S}, \ \forall A \in \mathfrak{T}$$

Hence, if $(S, \mathfrak{S})$, $(T, \mathfrak{T})$, and $(U, \mathfrak{U})$ are measurable spaces, and $f : S \longrightarrow T$ and $g : T \longrightarrow U$ are measurable functions, then $h : S \longrightarrow U$ defined by $h(s) = g \circ f(s)$ is a measurable function.

Definition A.3.3 (Measurable selection)**.** Let $(S, \mathfrak{S})$ and $(T, \mathfrak{T})$ be measurable spaces, and let Γ be a correspondence of S into T. Then the function $h : S \longrightarrow T$ is a **measurable selection from** Γ if h is measurable and $h(s) \in \Gamma(s)$, for all $s \in S$.

Theorem A.3.3 (Measurable Selection)**.** Let $S \subseteq \mathbf{R}^l$ and $T \subseteq \mathbf{R}^m$ be Borel sets, with their Borel subsets $\mathfrak{S}$ and $\mathfrak{T}$. Let $\Gamma : S \longrightarrow T$ be a nonempty compact-valued and upper hemi-continuous (UHC) correspondence. Then there exists a measurable selection from Γ.

A.4 Integration

Let $(S, \mathfrak{S}, \mu)$ be a fixed measure space and $M(S, \mathfrak{S})$ be the space of measurable, extended real-valued functions on S.

For a nonnegative, measurable and simple function, we have the following definition.

Definition A.4.1. Let $\varphi \in M^+(S, \mathfrak{S})$ be a measurable simple function, with the standard representation $\varphi(s) = \sum_{i=1}^{n} a_i \chi_{A_i}(s)$. Then the **integral of** φ **with respect to** μ is

$$\int_S \varphi(s) \, \mu(\mathrm{d}s) = \sum_{i=1}^{n} a_i \mu(A_i)$$

To extended from simple functions to all $M^+(S, \mathfrak{S})$, we define

Definition A.4.2. For $f \in M^+(S, \mathfrak{S})$, the **integral of** f **with respect to** μ is

$$\int_A f(s) \, \mu(\mathrm{d}s) = \int_S f(s) \chi_A(s) \mu(\mathrm{d}s)$$

We also denote the integrals above as $\int f \, d\mu$ and $\int_A f \, d\mu$.

Since any function $f \in M^+(S, \mathfrak{S})$ can be expressed as the limit of an increasing sequence (φ_n) of simple functions in $M^+(S, \mathfrak{S})$. However, we cannot define the integral as the limit of such a sequence because it would leave open the possibility that the limit depends on the particular sequence chosen. The Monotone Convergence Theorem shows that the limit value is unique.

Before that, the following result shows that a simple function on a measure space can be used to define a new measure on the space.

Lemma A.4.1. If $\varphi \in M^+(S, \mathfrak{S})$ is a simple function and $\lambda : \mathfrak{S} \longrightarrow \mathbf{R}$ is defined by

$$\lambda(A) = \int_A \varphi \, d\mu$$

for all $A \in \mathfrak{S}$, then λ is a measure on $\mathfrak{S}$.

Theorem A.4.2 (Monotone Convergence Theorem)**.** If (f_n) is a monotone increasing sequence of functions in $M^+(S, \mathfrak{S})$ converging pointwise to f then

$$\int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu$$

Some important properties:

1. If $f, g \in M^+(S, \mathfrak{S})$ and $c \geq 0$, then

$$\int (f + g) \, d\mu = \int f \, d\mu + \int g \, d\mu$$

and

$$\int c \cdot f \, d\mu = c \int f \, d\mu$$

2. If $f, g \in M^+(S, \mathfrak{S})$ and $f \leq g$, then

$$\int f \, d\mu \leq \int g \, d\mu$$

3. If $f \in M^+(S, \mathfrak{S})$ and $A, B \in \mathfrak{S}$ with $A \subseteq B$, then

$$\int_A f \, d\mu \leq \int_B f \, d\mu$$

4. If $f \in M^+(S, \mathfrak{S})$ and $A \in \mathfrak{S}$ with $\mu(A) = 0$, then

$$\int_A f \, d\mu = 0$$

5. If $f \in M^+(S, \mathfrak{S})$, $\int_S f \, d\mu < \infty$, and $A = \{s \in S : f(s) = +\infty\}$, then $\mu(A) = 0$.

6. If $f \in M^+(S, \mathfrak{S})$ and $\lambda : \mathfrak{S} \longrightarrow \mathbf{R}_+$, defined as $\lambda(A) = \int_A f \, d\mu$ for all $A \in \mathfrak{S}$, then λ is a measure on $(S, \mathfrak{S})$.

7. Let (g_i) be a sequence of functions in $M^+(S, \mathfrak{S})$. Then

$$\int \left(\sum_{i=1}^{n} g_i \right) d\mu = \sum_{i=1}^{n} \int g_i \, d\mu$$

Lemma A.4.3 (Fatou's Lemma). If (f_n) is a sequence of functions in $M^+(S, \mathfrak{S})$, then

$$\int (\liminf f_n) \, d\mu \leq \liminf \int f_n \, d\mu$$

To generalize to functions that take on both positive and negative values, we begin by defining the positive and negative parts of a function f^+ and f^-.

$$f^+(s) = \begin{cases} f(s), & \text{if } f(s) \geq 0 \\ 0, & \text{if } f(s) < 0 \end{cases}$$

and

$$f^-(s) = \begin{cases} -f(s), & \text{if } f(s) \leq 0 \\ 0, & \text{if } f(s) > 0 \end{cases}$$

Hence, if f is measurable, f^+ and f^- are both in $M^+(S, \mathfrak{S})$ and $f = f^+ - f^-$. This motivates the following generalization.

Definition A.4.3. Let $(S, \mathfrak{S}, \mu)$ be a measure space, and let f be a measurable real-valued function on S. If f^+ and f^- both have finite integrals with respect to μ, then f is integrable and the **integral of f with respect to μ** is

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu$$

In probability, if $(S, \mathfrak{S}, \mu)$ is a probability space and f is a measurable function, then f is a random variable. And if f is integrable, the integral of f with respect to μ is called the **expected value of f** and is denoted by

$$\mathbb{E}[f] = \int f \, d\mu$$

Now let $L(S, \mathfrak{S}, \mu)$ denote the set of all $\mathfrak{S}$-measurable, real-valued functions on S that are integrable with respect to μ. How can we determine whether a function is integrable?

1. If f is bounded, measurable, real-valued function on S, and $\mu(S) < \infty$, then f is μ-integrable.

2. f is μ-integrable iff. $|f|$ is μ-integrable. In that case, $|\int f \, d\mu| \leq \int |f| \, d\mu$.

3. If f is $\mathfrak{S}$-measurable, g is μ-integrable, and $|f| \leq |g|$, then f is μ-integrable and $\int |f| \, d\mu \leq \int |g| \, d\mu$.

4. If f and g are μ-integrable and $\alpha \in \mathbf{R}$, then αf and $f + g$ are μ-integrable, satisfying the standard rules of calculus.

Theorem A.4.4 (Lebesgue Dominated Convergence Theorem)**.** Let $(S, \mathfrak{S}, \mu)$ be a measure space, and let (f_n) be a sequence of integrable functions that converges almost everywhere to a measurable function f. If there exists an integrable function g such that $|f_n| \leq g$ for all n, then f is integrable and

$$\int f \, d\mu = \lim \int f_n \, d\mu$$

Definition A.4.4. Let λ and μ be finite measures on $(S, \mathfrak{S})$. If

$$\mu(A) = 0 \implies \lambda(A) = 0 \, \forall A \in \mathfrak{S}$$

then λ is **absolutely continuous** with respect to μ, written $\lambda << \mu$.

If there is $A \in \mathfrak{S}$ such that $\lambda(B) = \lambda(A \cap B)$, for all $B \in \mathfrak{S}$, then λ is **concentrated on** A.

If there are disjoint sets $A, B \in \mathfrak{S}$ such that λ is concentrated on A and μ is concentrated on B, then λ and μ are **mutually singular**, written $\lambda \perp \mu$.

Theorem A.4.5 (Radon-Nikodym Theorem)**.** Let λ and μ be σ-finite positive measures on $(S, \mathfrak{S})$ with $\lambda << \mu$. Then there exists an integrable function h such that

$$\lambda(A) = \int_A h(s) \, \mu(ds), \text{ all } A \in \mathfrak{S}$$

And the function h is called the **Radon-Nikodym derivative** of λ with respect to μ.

Lemma A.4.6. Let λ_1 and λ_2 be finite measures on $(S, \mathfrak{S})$. Then there exists a triple of measures $\gamma, \alpha_1, \alpha_2$ such that

$$\lambda_i = \gamma + \alpha_1$$

where $i = 1, 2$, and $\alpha_1 \perp \alpha_2$.

A.5 Product Spaces

Let $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$ be measurable spaces, and let $Z = X \times Y$.

Our first task is to define a σ-algebra of subsets of Z that is a natural product of $\mathfrak{X}$ and $\mathfrak{Y}$.

Definition A.5.1 (Measurable rectangle)**.** A set $C = A \times B \subseteq Z$ is a **measurable rectangle** if $A \in \mathfrak{X}$ and $B \in \mathfrak{Y}$.

We denote by $\mathfrak{C}$ the set of all measurable rectangles and $\mathfrak{E}$ the set of all finite unions of measurable rectangles. Notice that $\mathfrak{E}$ is an algebra and that every set in $\mathfrak{E}$ can be written as the finite union of disjoint measurable rectangles.

Definition A.5.2 (Product Space)**.** Let $\mathfrak{F} = \mathfrak{X} \times \mathfrak{Y}$ be the σ-algebra generated by $\mathfrak{E}$. The measurable space $(Z, \mathfrak{F})$ is called the **product space**.

Theorem A.5.1. Let $(X, \mathfrak{X})$, $(Y, \mathfrak{Y})$, $\mathfrak{C}$, and $\mathfrak{E}$ be as specified above. Let $\mu : \mathfrak{C} \longrightarrow \mathbf{R}_+$ have the following properties:

- $\mu(\emptyset) = 0$;

- If $(C_i) = ((A_i \times B_i))_{i=1}^{\infty}$ is a sequence of disjoint sets in $\mathfrak{C}$ and $\bigcup_{i=1}^{\infty} C_i$ is in $\mathfrak{C}$, then $\mu\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \mu(C_i)$.

Then there is a measure on $\mathfrak{E}$ that coincides with μ on $\mathfrak{C}$.

This theorem can be naturally extended to any space that is the product of a finite number of measurable spaces.

Definition A.5.3. Let $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$ be measurable spaces, and let $(Z, \mathfrak{F})$ be the product space. And let $E \subseteq Z$ and $x \in X$. Then the *x*-**section of** E is the set $E_x = y \in Y : (x, y) \in E$. And the *y*-**section of** E is the set $E_y = x \in X : (x, y) \in E$.

Let $f : Z \longrightarrow \mathbf{R}$ and let $x \in X$. The *x*-**section of** f is the function $f_x : Y \longrightarrow \mathbf{R}$ defined by $f_x(y) = f(x, y)$. And the *y*-**section of** f is the function $f_y : X \longrightarrow \mathbf{R}$ defined by $f_y(x) = f(x, y)$.

Intuitively, the *x*-section of a function is found by fixing x and viewing f only as a function of y.

Theorem A.5.2. Let $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$ be measurable spaces, and let $(Z, \mathfrak{F})$ be the product space.

1. If the set E in Z is $\mathfrak{F}$-measurable, then every section of E is measurable.

2. If the function $f : Z \longrightarrow \mathbf{R}$ is $\mathfrak{F}$-measurable, then every section of f is measurable.

A.6 The Monotone Class Lemma

The main question that this section aims to answer is whether a propery P holds at a given σ-algebra.

Definition A.6.1 (Monotone Class). A **monotone class** is a nonempty collection $\mathfrak{M}$ of sets such that it contains

1. The union of every nested increasing sequence $A_1 \subseteq A_2 \subseteq \ldots$ of sets in $\mathfrak{M}$.

2. The intersection of every nested decreasing sequence $A_1 \supseteq A_2 \supseteq \ldots$ of sets in $\mathfrak{M}$.

Some important facts:

1. Every σ-algebra is a monotone class.

2. Given a nonempty collection of subsets of S, denoted by $\mathfrak{A}$, there exists a smallest monotone class containing $\mathfrak{A}$, which is called the monotone class generated by $\mathfrak{A}$.

3. The σ-algebra generated by $\mathfrak{A}$ contains the monotone class generated by $\mathfrak{A}$.

4. If a monotone class is an algebra, then it is a σ-algebra.

> **Lemma A.6.1** (Monotone class). Leet S be a set and let $\mathfrak{A}$ be an algebra of subsets of S. Then the monotone class $\mathfrak{M}$ generated by $\mathfrak{A}$ is the same as the σ-algebra $\mathfrak{S}$ generated by $\mathfrak{A}$.

Using this lemma, our original task in this section is solved. To establish that P holds for all sets in a product σ-algebra, it suffices to show that

1. P holds for all finite unions of disjoint measurable rectangles.

2. The family of sets $\mathfrak{E}$ for which P holds is a monotone class.

A.7 Modes of Convergence

To the traditional ideas of convergence **pointwise** and **uniformly**, learned in an undergraduate course in Analysis, we'll add new forms of convergence, as the functions f_n acquire an infinite and inequivalent number of ways to approach their limit f.

Let $(X, \mathfrak{B}, \mu)$ be a measure space, and the functions f_n (and their limit f) measurable with respect to this space. Then, we say that f_n converges to f

1. **pointwise almost everywhere** if, for $(\mu-)$almost everywhere $x \in X$, $f_n(x)$ converges to $f(x)$.

2. **uniformly almost everywhere, essentially uniformly**, or **in L^∞ norm** if, for every $\varepsilon > 0$, there exists N such that for every $n \geq N$,

$$|f_n(x) - f(x)| \leq \varepsilon$$

   for μ-almost every $x \in X$.

3. **almost uniformly** if, for every $\varepsilon > 0$, there exists a set $E \in \mathfrak{B}$ of measure $\mu(E) \leq \varepsilon$ such that f_n converges uniformly to f on the complement of E.

4. **in L^1 norm** if the quantity

$$\|f_n - f\|_{L^1(\mu)} = \int_x |f_n(x) - f(x)| \, d\mu$$

   converges to 0 as $n \to \infty$.

5. **in measure** if, for every $\varepsilon > 0$, the measures

$$\mu(\{x \in X : |f_n(x) - f(x)| \geq \varepsilon\})$$

   converge to zero as $n \to \infty$.

In probability theory, if f_n and f are random variables, convergence in L^1 is usually referred to as **convergence in mean**, pointwise convergence almost everywhere is referred to as **almost sure convergence**, and convergence in measure is often referred to as **convergence in probability**.

Recall that a property $P(x)$ is said to hold **almost everywhere** if the set of x for which $P(x)$ fails has Lebesgue measure zero.

A.8 Conditional Expectation

Let $(\Omega, \mathfrak{F}, \mu)$ be a probability space.

Definition A.8.1 (Conditional Probability)**.** For a measurable set A satisfying $\mu(A) > 0$ and for any measurable set B we define the **conditional probability** of B given A as

$$\mu_A(B) := \frac{\mu(B \cap A)}{\mu(A)}$$

We'll also use $P(B \mid A)$ to denote this probability.

Now notice that $\mu_A : \mathfrak{F} \longrightarrow [0, 1]$, i.e., it is also a probability measure.

Definition A.8.2 (Conditional Expectation)**.** The **conditional expectation** of f given A is defined as

$$\mathbb{E}[f \mid A] = \int f \, d\mu_A$$

Definition A.8.3 (Measurable Partition)**.** A family $A_{\eta \, \eta \in H}$ of subsets of Ω is a **measurable partition** of Ω if the following conditions hold:

1. $A_\eta \in \mathfrak{F}, \forall \eta \in H$.

2. $\bigcup_{\eta \in H} A_\eta = H$.

3. $A_\eta \cap A_{\eta'} = 0, \forall \eta \neq \eta'$.

If the index set H is countable, then we call the measurable partition **countable**.

To formalize it better, extending the concept to σ-algebras not generated by a countable partition, we'll introduce the following.

Definition A.8.4 (Conditional Expectation)**.** Let $A \subset \mathfrak{F}$ be a σ-algebra and let $f : \Omega \longrightarrow \mathbf{R}$ be an integrable function. Then the **conditional expectation of f relative to** A is an $\mathfrak{A}$-measurable function $\mathbb{E}[f \mid \mathfrak{A}] : \Omega \longrightarrow \mathbf{R}$ such that

$$\int_C \mathbb{E}[f \mid \mathfrak{A}](\omega) \mu \, (d\omega) = \int_C f(\omega) \mu \, (d\omega), \forall C \in \mathfrak{A}$$

It can be proved that this function exists and is unique in the sense that if g also satisties the condition above, then $\mathbb{E}[f \mid \mathfrak{A}] = g$.

Notice that to compute conditional probability $P(B \mid A_\eta)$ we may take f to be the indicator function χ_B. Then

$$\mathbb{E}[\chi_B \mid \mathfrak{A}](\hat{\omega}) \mu(A_\eta) = \int_{A_\eta} \chi_b \, d\mu = \mu(B \cap A_\eta)$$

for all $\hat{\omega} \in A_\eta$, all $\eta \in H$.

If $\mu(A_\eta) > 0$, then

$$\mathbb{E}[\chi_B \mid \mathfrak{A}](\hat{\omega}) = \frac{\mu(B \cap A_\eta)}{\mu(A_\eta)}$$

Summarizing this fact,

$$P(B \mid A_\eta) = \mathbb{E}[\chi_B \mid \mathfrak{A}](\hat{\omega})$$

for all $\hat{\omega} \in A_\eta$, all $\eta \in H$.

We end this section with an important property of conditional expectations. Let $\mathfrak{A}_1 \subseteq \mathfrak{A}_2 \subseteq \mathfrak{F}$. Then we have

$$\int_C \mathbb{E}[\mathbb{E}(f \mid \mathfrak{A}_2) \mid \mathfrak{A}_1]\, d\mu = \int_c \mathbb{E}(f \mid \mathfrak{A}_2)(\omega)\, d\mu$$

$$= \int_C f(\omega)\, d\mu$$

$$= \int_C \mathbb{E}[f \mid \mathfrak{A}_1](\omega)\, d\mu, \ \forall C \in \mathfrak{A}_1$$

This is known as the **law of the iterated expectation**. To state it in another way,

$$\int_C \mathbb{E}[f - \mathbb{E}(f \mid \mathfrak{A}_2) \mid \mathfrak{A}_1]\, d\mu = 0, \ \forall C \in \mathfrak{A}_1$$

Bibliography

[Dur19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[Eva12] Lawrence C Evans. *An introduction to stochastic differential equations*, volume 82. American Mathematical Soc., 2012.

[Gar88] Thomas C Gard. *Introduction to Stochastic Differential Equations*. Dekker, Inc, 1988.

[Jak]   Jake. How to draw brownian motions in tikz/pgf. https://tex.stackexchange.com/a/59934. [Acesso em: 07/07/2022].

[Kle12] Fima C Klebaner. *Introduction to stochastic calculus with applications*. World Scientific Publishing Company, 2012.

[Oks13] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[S$^+$04] Steven E Shreve et al. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer, 2004.

[Wik22] Wikipedia. Kalman filter — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Kalman%20filter&oldid=1091166052, 2022. [Online; accessed 06-July-2022].