

# An Introduction to Stochastic Differential Equations

Adair Antonio da Silva Neto

August 18, 2022

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Stochastic Integrals</b>	<b>2</b>
2.1	Formalizing the ‘noise’	2
2.2	Preparing the Terrain	3
2.3	Constructing the Itô Integral	5
2.4	Properties	9
2.5	Extensions	10
2.6	Stratonovich integral	10
<b>3</b>	<b>The Ito Formula and the Martingale Representation Theorem</b>	<b>12</b>
<b>4</b>	<b>Stochastic Differential Equations</b>	<b>13</b>
<b>A</b>	<b>Measure Theory</b>	<b>14</b>
A.1	Measurable Spaces	14
A.2	Measures	14
A.3	Measurable Functions	17
A.4	Integration	19
A.5	Product Spaces	23
A.6	The Monotone Class Lemma	24
A.7	Modes of Convergence	25
A.8	Conditional Expectation	26
<b>B</b>	<b>Probability Theory</b>	<b>28</b>
B.1	Probability Spaces	28
B.2	Distribution and Density of a Random Variable	30
B.3	Expected Value and Variance	31
B.4	Stochastic Processes	34
B.5	Brownian Motion	35
B.6	Riemann-Stieltjes Integral	36
B.7	Modes of Convergence	37
	<b>References</b>	<b>39</b>

# 1 Introduction

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

$$\frac{dX}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot \text{‘white noise’} \quad (1)$$

Problems like this appears naturally in biology (populational growth models), physics (charge in electrical circuits), engineering (filtering problems, like the Kalman’s Filter exhibited in the Figure 1) and finance (optimal stopping, optimal portfolio and option pricing).

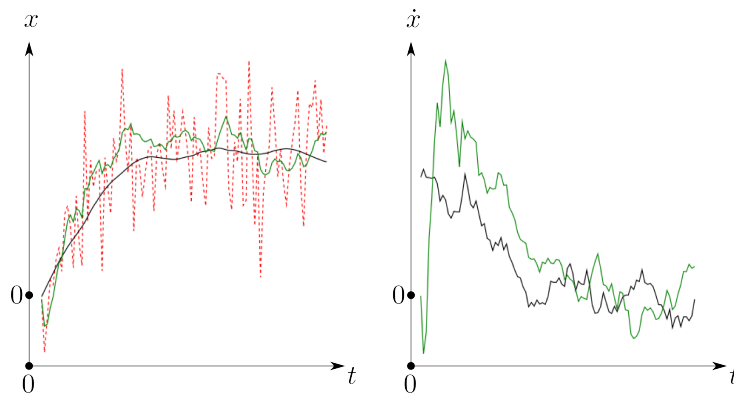


Figure 1: Kalman’s Filter: ■ Observed data; ■ Filtered process; ■ Real data [Wik22].

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.

## 2 Stochastic Integrals

### 2.1 Formalizing the ‘noise’

To study the equation (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  $0 = t_0 < t_1 < \dots < t_m = t$  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 2.

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  $\Delta t_j$  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.

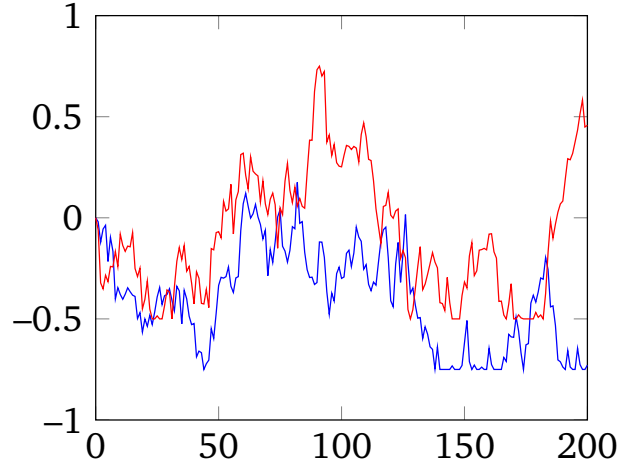


Figure 2: Two realizations of the Brownian motion [Jak]

## 2.2 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

$$\frac{dX_t}{dt} = b(t, X_t), \quad X_0 = x_0 \quad (2)$$

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

$$X_t - X_0 = \int_0^T b(s, X_s) ds \quad (3)$$

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) an identity of the form

$$X_t - X_0 = \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s \quad (4)$$

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

$$\int_S^T f(t, \omega) dB_t(\omega) \quad (5)$$

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 (5) as follows

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

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 2.1.** Let  $B_t(\omega)$  be a Brownian motion. We define  $\mathfrak{F}_t$  as the  $\sigma$ -algebra generated by the random variables  $B_s$ , where  $s \leq t$ . I.e.,  $\mathfrak{F}_t$  is the smallest  $\sigma$ -algebra containing all sets

$$\{\omega : B_{t_1}(\omega) \in F_1, \dots, B_{t_k}(\omega) \in F_k\}$$

where  $t_j \leq t$  and  $F_j \subseteq \mathbf{R}^n$  are Borel sets,  $j \leq k = 1, 2, \dots$

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 2.2** ( $\mathfrak{N}_t$ -Adapted). Let  $\{\mathfrak{N}_t\}_{t \geq 0}$  be an increasing family of  $\sigma$ -algebras of subsets of  $\Omega$ . 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 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.

## 2.3 Constructing the Itô Integral

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

**Definition 2.3.** 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  $\sigma$ -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  $\varphi$  and then show how each  $f \in \mathfrak{V}$  can be approximated by  $\varphi$ .

**Definition 2.4** (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 2.1** (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  $\omega$ . Then there exists elementary functions  $\varphi_n \in \mathfrak{V}$  such that

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

as  $n \rightarrow \infty$ .

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  $\omega$  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  $n \rightarrow \infty$ .

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

Finally, we can define the Itô integral as follows.

**Definition 2.5** (The Itô integral). Let  $f \in \mathfrak{V}(S, T)$ , then the **Itô integral** of  $f$ , from  $S$  to  $T$  is

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

where the limit is in  $L^2(P)$  and  $(\varphi_n)$  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  $n \rightarrow \infty$ .

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

**Theorem 2.3** (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 2.4.** If  $f(t, \omega) \in \mathfrak{V}(S, T)$  and  $f_n(t, \omega) \in \mathfrak{V}(S, T)$  and

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

as  $n \rightarrow \infty$ , 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  $n \rightarrow \infty$ .

**Example 2.2.** 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,

$$\begin{aligned} \mathbb{E} \left[ \int_0^T (\varphi_n - B_s)^2 ds \right] &= \mathbb{E} \left[ \sum_j \int_{t_j}^{t_{j+1}} (B_j - B_s)^2 ds \right] = \sum_j \int_{t_j}^{t_{j+1}} (s - t_j)^2 ds \\ &= \sum_j \frac{1}{2} (t_{j+1} - t_j)^2 \longrightarrow 0 \text{ when } \Delta t_j \rightarrow 0 \end{aligned}$$

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

**Step 2.** Now notice that

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

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

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

**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

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

## 2.4 Properties

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

**Theorem 2.5.** Let  $f, g \in \mathfrak{V}(0, T)$  and let  $0 \leq S < U < T$ . Then

1.  $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$  for a.a.  $\omega$ .
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 2.6.** 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  $\lambda, T > 0$

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

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

## 2.5 Extensions

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

Considering the Definition 2.3, we can relax the condition 2. into

- 2.' There exists an increasing family of  $\sigma$ -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 weakening the condition 3. into

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

To 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

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

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

## 2.6 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<sup>+</sup>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) has a solution? Under which hypothesis? How can we study the behavior of the solution?

### **3 The Ito Formula and the Martingale Representation Theorem**

## 4 Stochastic Differential Equations

## A Measure Theory

### A.1 Measurable Spaces

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

**Definition A.1.** Let  $S$  be a set and  $\mathfrak{G}$  be a family of subsets of  $S$ . Then  $\mathfrak{G}$  is called a  $\sigma$ -**algebra** on  $S$  if

1.  $\emptyset, S \in \mathfrak{G}$ .
2. If  $A \in \mathfrak{G}$ , then  $A^c \in \mathfrak{G}$ .
3. If  $A_1, A_2, \dots \in \mathfrak{G}$ , then  $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{G}$ .

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

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

Particularly, the smallest  $\sigma$ -algebra containing all of the open sets of  $\mathbf{R}$  is called the **Boreal algebra** for  $\mathbf{R}$  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  $\mathfrak{G}$ ?

**Definition A.2.** Let  $(S, \mathfrak{G})$  be a measurable space. A **measure** is an extended real-valued function  $\mu : \mathfrak{G} \longrightarrow \overline{\mathbf{R}}$  such that

1.  $\mu(\emptyset) = 0$ ;
2.  $\mu(A) \geq 0$ , for all  $A \in \mathfrak{G}$ ;
3. If  $(A_n)_{n=1}^{\infty}$  is a countable, disjoint sequence of subsets in  $\mathfrak{G}$ , 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  $\mu$  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  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$  for all  $x \in A^c$ .

If  $\mu(S) = 1$ , then  $\mu$  is a **probability measure** and  $(S, \mathfrak{S}, \mu)$  is called a **probability space**. Any measurable set  $A \in \mathfrak{S}$  is called an **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  $\mu(B \setminus A) = \mu(B) - \mu(A)$ .

**Theorem A.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

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

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

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

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

**Definition A.3.** 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.  $A_1, A_2, \dots, A_n \in \mathfrak{A}$  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  $\sigma$ -algebra.

**Definition A.4.** Let  $S$  be a set, and let  $\mathfrak{A}$  be an algebra of subsets of  $S$ . A **measure** is a real-valued function  $\mu : \mathfrak{A} \rightarrow \bar{\mathbf{R}}$  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  $\sigma$ -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  $\sigma$ -algebras.

**Theorem A.2** (Caratheodory Extension Theorem). Let  $S$  be a set,  $\mathfrak{A}$  an algebra of its subsets, and  $\mu$  a measure on  $\mathfrak{A}$ . Let  $\mathfrak{G}$  be the smallest  $\sigma$ -algebra containing  $\mathfrak{A}$ . Then there exists a measure  $\mu^*$  on  $\mathfrak{G}$  such that  $\mu(A) = \mu^*(A)$ , for all  $A \in \mathfrak{A}$ .

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

**Definition A.5.** Let  $S$  be a set,  $\mathfrak{A}$  an algebra of its subsets, and  $\mu$  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  $\mu$  is called  **$\sigma$ -finite**.

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

**Theorem A.3** (Hahn Extension Theorem). Let  $S$ ,  $\mathfrak{A}$ ,  $\mu$  and  $\mathfrak{G}$  be as specified in **Caratheodory Extension Theorem**. If  $\mu$  is  $\sigma$ -finite, then the extension  $\mu^*$  to  $\mathfrak{G}$  is unique.

**Definition A.6.** Let  $(S, \mathfrak{G}, \mu)$  be a measure space,  $A \in \mathfrak{G}$  be any set with measure zero, and let  $C \subseteq A$ . Denoting by  $\mathfrak{C}$  the family of such sets  $C$ , i.e.,

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

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

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

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

### A.3 Measurable Functions

**Definition A.7.** Given a measurable space  $(S, \mathfrak{G})$ , a real-valued function  $f : S \rightarrow \mathbf{R}$  is **measurable with respect to  $\mathfrak{G}$**  (or  **$\mathfrak{G}$ -measurable**) if

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

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  $f : \mathbf{R} \rightarrow \mathbf{R}$  is measurable with respect to Borel sets.
- If  $S$  is a countable set and  $\mathfrak{G}$  is a complete  $\sigma$ -algebra for  $S$  (i.e. contains all subsets of  $S$ ), then all functions  $f : S \rightarrow \mathbf{R}$  are  $\mathfrak{G}$ -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  $(S, \mathfrak{G})$  be a measurable space and consider the indicator function  $\chi_A : S \rightarrow \mathbf{R}$  of the form

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

Clearly,  $\chi_A$  is  $\mathfrak{G}$ -measurable iff.  $A \in \mathfrak{G}$ .

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  $\varphi$  is measurable iff. each  $A_i \in \mathfrak{G}$ .

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.4** (Pointwise convergence preserves measurability). Let  $(S, \mathfrak{G})$  be a measurable space, and let  $(f_n)$  be a sequence of  $\mathfrak{G}$ -measurable functions converging pointwise to  $f$ , i.e.,

$$\lim_{n \rightarrow \infty} f_n(s) = f(s), \forall s \in S$$

Then  $f$  is also  $\mathfrak{G}$ -measurable.

Second, any measurable function  $f$  can be expressed as the pointwise limit of a sequence  $(\varphi_n)$  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.5** (Approximation of measurable functions by simple functions). Let  $(S, \mathfrak{G})$  be a measurable space. If  $f : S \rightarrow \mathbf{R}$  is  $\mathfrak{G}$ -measurable, then there exists a sequence of measurable simple functions  $(\varphi_n)$  such that  $\varphi_n \rightarrow f$  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  $\varphi_n \rightarrow f$  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  $\mathfrak{G}$ -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.8.** 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.9** (Measurable selection). Let  $(S, \mathfrak{S})$  and  $(T, \mathfrak{T})$  be measurable spaces, and let  $\Gamma$  be a correspondence of  $S$  into  $T$ . Then the function  $h : S \longrightarrow T$  is a **measurable selection from  $\Gamma$**  if  $h$  is measurable and  $h(s) \in \Gamma(s)$ , for all  $s \in S$ .

**Theorem A.6** (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  $\Gamma$ .

## 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.10.** 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  $\varphi$  with respect to  $\mu$**  is

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

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

**Definition A.11.** For  $f \in M^+(S, \mathfrak{S})$ , the **integral of  $f$  with respect to  $\mu$**  is

$$\int_A f(s) \mu(ds) = \int_S f(s) \chi_A(s) \mu(ds)$$

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  $(\varphi_n)$  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.7.** If  $\varphi \in M^+(S, \mathfrak{S})$  is a simple function and  $\lambda : \mathfrak{S} \rightarrow \mathbf{R}$  is defined by

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

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

**Theorem A.8** (Monotone Convergence Theorem). If  $(f_n)$  is a monotone increasing sequence of functions in  $M^+(S, \mathfrak{S})$  converging pointwise to  $f$  then

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

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  $\lambda : \mathfrak{S} \rightarrow \mathbf{R}_+$ , defined as  $\lambda(A) = \int_A f \, d\mu$  for all  $A \in \mathfrak{S}$ , then  $\lambda$  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.9** (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.12.** 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  $\mu$ , then  $f$  is integrable and the **integral of  $f$  with respect to  $\mu$**  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  $\mu$  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  $\mu$ . 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  $\mu$ -integrable.
2.  $f$  is  $\mu$ -integrable iff.  $|f|$  is  $\mu$ -integrable. In that case,  $|\int f \, d\mu| \leq \int |f| \, d\mu$ .
3. If  $f$  is  $\mathfrak{S}$ -measurable,  $g$  is  $\mu$ -integrable, and  $|f| \leq |g|$ , then  $f$  is  $\mu$ -integrable and  $\int |f| \, d\mu \leq \int |g| \, d\mu$ .
4. If  $f$  and  $g$  are  $\mu$ -integrable and  $\alpha \in \mathbf{R}$ , then  $\alpha f$  and  $f + g$  are  $\mu$ -integrable, satisfying the standard rules of calculus.

**Theorem A.10** (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.13.** Let  $\lambda$  and  $\mu$  be finite measures on  $(S, \mathfrak{S})$ . If

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

then  $\lambda$  is **absolutely continuous** with respect to  $\mu$ , written  $\lambda \ll \mu$ .

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

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



**Theorem A.11** (Radon-Nikodym Theorem). Let  $\lambda$  and  $\mu$  be  $\sigma$ -finite positive measures on  $(S, \mathfrak{S})$  with  $\lambda \ll \mu$ . 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  $\lambda$  with respect to  $\mu$ .

**Lemma A.12.** Let  $\lambda_1$  and  $\lambda_2$  be finite measures on  $(S, \mathfrak{S})$ . Then there exists a triple of measures  $\gamma, \alpha_1, \alpha_2$  such that

$$\lambda_i = \gamma + \alpha_i$$

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  $\sigma$ -algebra of subsets of  $Z$  that is a natural product of  $\mathfrak{X}$  and  $\mathfrak{Y}$ .

**Definition A.14** (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.15** (Product Space). Let  $\mathfrak{F} = \mathfrak{X} \times \mathfrak{Y}$  be the  $\sigma$ -algebra generated by  $\mathfrak{E}$ . The measurable space  $(Z, \mathfrak{F})$  is called the **product space**.

**Theorem A.13.** Let  $(X, \mathfrak{X})$ ,  $(Y, \mathfrak{Y})$ ,  $\mathfrak{C}$ , and  $\mathfrak{E}$  be as specified above. Let  $\mu : \mathfrak{C} \rightarrow \mathbb{R}_+$  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  $\mu$  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.16.** Let  $(X, \mathfrak{X})$  and  $(Y, \mathfrak{Y})$  be measurable spaces, and let  $(Z, \mathfrak{Z})$  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  $f : Z \rightarrow \mathbf{R}$  and let  $x \in X$ . The  **$x$ -section of  $f$**  is the function  $f_x : Y \rightarrow \mathbf{R}$  defined by  $f_x(y) = f(x, y)$ . And the  **$y$ -section of  $f$**  is the function  $f_y : X \rightarrow \mathbf{R}$  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.14.** Let  $(X, \mathfrak{X})$  and  $(Y, \mathfrak{Y})$  be measurable spaces, and let  $(Z, \mathfrak{Z})$  be the product space.

1. If the set  $E$  in  $Z$  is  $\mathfrak{Z}$ -measurable, then every section of  $E$  is measurable.
2. If the function  $f : Z \rightarrow \mathbf{R}$  is  $\mathfrak{Z}$ -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 property  $P$  holds at a given  $\sigma$ -algebra.

**Definition A.17** (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  $A_1 \subseteq A_2 \subseteq \dots$  of sets in  $\mathfrak{M}$ .
2. The intersection of every nested decreasing sequence  $A_1 \supseteq A_2 \supseteq \dots$  of sets in  $\mathfrak{M}$ .

Some important facts:

1. Every  $\sigma$ -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  $\sigma$ -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  $\sigma$ -algebra.

**Lemma A.15** (Monotone class). Let  $S$  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  $\sigma$ -algebra  $\mathfrak{G}$  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  $\sigma$ -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  $L^\infty$  norm if, for every  $\varepsilon > 0$ , there exists  $N$  such that for every  $n \geq N$ ,

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

for  $\mu$ -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  $n \rightarrow \infty$ .

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  $n \rightarrow \infty$ .

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.18** (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  $P(B|A)$  to denote this probability.

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

**Definition A.19** (Conditional Expectation). The **conditional expectation** of  $f$  given  $A$  is defined as

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

**Definition A.20** (Measurable Partition). A family  $A_{\eta}_{\eta \in H}$  of subsets of  $\Omega$  is a **measurable partition** of  $\Omega$  if the following conditions hold:

1.  $A_{\eta} \in \mathfrak{F}, \forall \eta \in H$ .
2.  $\bigcup_{\eta \in H} A_{\eta} = H$ .
3.  $A_{\eta} \cap A_{\eta'} = \emptyset, \forall \eta \neq \eta'$ .

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

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

**Definition A.21** (Conditional Expectation). Let  $\mathcal{A} \subset \mathfrak{F}$  be a  $\sigma$ -algebra and let  $f : \Omega \longrightarrow \mathbf{R}$  be an integrable function. Then the **conditional expectation of  $f$  relative to  $\mathcal{A}$**  is an  $\mathcal{A}$ -measurable function  $\mathbb{E}[f|\mathcal{A}] : \Omega \longrightarrow \mathbf{R}$  such that

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

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

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

$$\mathbb{E}[\chi_B|\mathcal{A}](\hat{\omega})\mu(A_\eta) = \int_{A_\eta} \chi_B d\mu = \mu(B \cap A_\eta)$$

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

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

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

Summarizing this fact,

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

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

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

$$\begin{aligned} \int_C \mathbb{E}[\mathbb{E}(f|\mathcal{A}_2)|\mathcal{A}_1] d\mu &= \int_C \mathbb{E}(f|\mathcal{A}_2)(\omega) d\mu \\ &= \int_C f(\omega) d\mu \\ &= \int_C \mathbb{E}[f|\mathcal{A}_1](\omega) d\mu, \forall C \in \mathcal{A}_1 \end{aligned}$$

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

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

## B Probability Theory

### B.1 Probability Spaces

A probability space starts with a set  $\Omega$  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 B.1** ( $\sigma$ -algebra). If  $\Omega$  is a set, then a  $\sigma$ -**algebra**  $\mathfrak{F}$  on  $\Omega$  is a family of subsets of  $\Omega$  satisfying:

1.  $\emptyset \in \mathfrak{F}$ .
2. If  $A \in \mathfrak{F}$ , then  $A^c \in \mathfrak{F}$ .
3. If  $A_1, A_2, \dots \in \mathfrak{F}$ , 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 B.1** (Examples of  $\sigma$ -algebras). The following are  $\sigma$ -algebras.

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

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

Notice that the Borel  $\sigma$ -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  $\sigma$ -algebra containing all open subsets.

More generally, if  $\mathcal{U}$  is the collection of all open subsets of a topological space  $\Omega$ , then  $\mathfrak{B}$  is called the **Borel  $\sigma$ -algebra** on  $\Omega$ . 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 B.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. ( $\sigma$ -additivity). If  $A_1, A_2, \dots \in \mathfrak{F}$  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  $\sigma$ -algebra, the following definition will be needed.

**Definition B.3** ( $\mathfrak{F}$ -measurable function). Given a probability space  $(\Omega, \mathfrak{F}, P)$ , a function  $Y$  from the sample space  $\Omega$  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  $\sigma$ -algebra.

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

**Definition B.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  $x_i, i = 1, 2, \dots, k$ . The sets

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

form a partition of  $\Omega$ , and the  $\sigma$ -algebra generated by this partition is called the  **$\sigma$ -algebra generated by  $X$** .

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

## B.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 B.5** (Distribution). The **distribution** of a random variable  $X$  is a function  $\mu_X : \Omega \rightarrow \mathbf{R}^n$  defined as

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

**Definition B.6** (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 B.7** (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 B.8** (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

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

### B.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 B.9** (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 B.1** (Chebychev's inequality).

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

for all  $\lambda \geq 0$ .

**Definition B.10** (Variance). Let  $\mu := \mathbb{E}[X]$ , i.e., the expected value of a random variable  $X$ . Then the **variance** of  $X$  is given by

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

**Definition B.11** (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:

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

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

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

**Theorem B.2.** Let  $X$  be a random variable. Then,

1.  $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ .
2.  $\text{Var}[aX + b] = a^2 \text{Var}[X]$ .
3.  $\text{Cov}[X, Y] = \text{Cov}[Y, X]$ .
4.  $\text{Cov}[aX + bY, Z] = a \text{Cov}[X, Z] + b \text{Cov}[Y, Z]$ .
5.  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \text{Cov}[X, Y]$ .

**Definition B.12** (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  $A_i$ ,  $i = 1, 2, \dots$ , is called independent if for any  $n \in \mathbb{N}$  and any choice of indices  $i_k$ ,  $k = 1, 2, \dots$ ,

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

**Definition B.13** (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

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

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

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

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 B.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  $\sigma$ -algebra, and let  $a, b \in \mathbf{R}$ .

1.  $\mathbb{E}[aX + bY | \mathfrak{H}] = a\mathbb{E}[X | \mathfrak{H}] + b\mathbb{E}[Y | \mathfrak{H}]$ .
2.  $\mathbb{E}[\mathbb{E}[X | \mathfrak{H}]] = \mathbb{E}[X]$ .
3.  $\mathbb{E}[X | \mathfrak{H}] = X$  if  $X$  is  $\mathfrak{H}$ -measurable.
4.  $\mathbb{E}[X | \mathfrak{H}] = \mathbb{E}[X]$  if  $X$  is independent of  $\mathfrak{H}$ .
5.  $\mathbb{E}[Y \cdot X | \mathfrak{H}] = Y \cdot \mathbb{E}[X | \mathfrak{H}]$  if  $Y$  is  $\mathfrak{H}$ -measurable.
6. If  $\mathfrak{G} \subseteq \mathfrak{H}$  is a  $\sigma$ -algebra, then

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

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

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

**Definition B.14** (Infinitely often). If  $A_1, A_2, \dots, A_n, \dots$  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 B.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 B.15** (Normal Distribution). If random variable  $X$  has mean  $\mu$ , variance  $\sigma^2$  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 B.16** (Lipschitz and Hölder Conditions). A function  $f$  is said to be **Hölder continuous** of order  $\alpha$ ,  $0 < \alpha \leq 1$ , 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**.

## B.4 Stochastic Processes

**Definition B.17** (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 B.18** (Filtration). A filtration  $\mathbf{F}$  is a collection of  $\sigma$ -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  $\Omega$ .

Given a stochastic process  $\{X(t)\}$ , let  $\mathfrak{F}_t$  be the  $\sigma$ -algebra generated by the random variables  $X_s$ ,  $s = 0, \dots, t$ . Since  $\mathfrak{F}_t \subseteq \mathfrak{F}_{t+1}$ , these  $\sigma$ -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 B.19** (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.  $\mathbb{E}[M_s | \mathfrak{M}_t] = M_t$  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 B.5** (Doob's martingale inequality). If  $M_t$  is a martingale such that the mapping  $t \rightarrow M_t(\omega)$  is continuous a.s., then for all  $p \geq 1$ ,  $T \geq 0$  and  $\lambda > 0$ ,

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

## B.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  $\omega$ .

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 B.20** (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 B.6** (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 B.7.**

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

**Theorem B.8.** 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:

$$\begin{aligned} B_j(B_{j+1} - B_j) &= \frac{1}{2}[B_{j+1}^2 - B_j^2 - (B_{j+1} - B_j)^2] \\ B_j^2(B_{j+1} - B_j) &= \frac{1}{3}(B_{j+1}^3 - B_j^3) - B_j(B_{j+1} - B_j)^2 - \frac{1}{3}(B_{j+1} - B_j)^3 \end{aligned}$$

## B.6 Riemann-Stieltjes Integral

**Definition B.21** (Riemann-Stieltjes Integral). If  $\varphi$  is a continuous function on  $[a, b]$  and  $F$  is a distribution function, we define the **Riemann-Stieltjes Integral** of  $\varphi$  on  $[a, b]$  in relation to  $F$  (or weighted by  $F$ ) as

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

where  $a = x_1 < x_2 < \dots < x_{n+1} = b$ ,  $y_i$  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

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

## B.7 Modes of Convergence

**Definition B.22.** 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 B.23** (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  $f : \mathbf{R} \rightarrow \mathbf{R}$  we have

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

I.e.,

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

**Definition B.24** (Convergence in Probability). A sequence of random variables  $\{X_n\}$  **converges in probability** to  $X$  if for any  $\varepsilon > 0$ , we have

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

**Definition B.25** (Convergence Almost Surely). A sequence of random variables  $\{X_n\}$  **converges almost surely (a.s.)** to  $X$  if for any  $\omega$  outside a set of probability zero we have that

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

Alternatively, we can write

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

**Definition B.26** ( $L^p$ -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

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

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

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0$$



## References

- [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.
- [Jak] Jake. How to draw brownian motions in tikz/pgf. <https://tex.stackexchange.com/a/59934>. [Acesso em: 07/07/2022]. 3
- [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. 9, 11, 35
- [S<sup>+</sup>04] Steven E Shreve et al. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer, 2004. 11
- [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]. 1