

Personal Notes on Stochastic Processes

Adel Saleh

February 17, 2021

Abstract

The following document is a comprehensive treatment of the fundamentals of stochastic processes and stochastic analysis. It is compiled from several sources, spanning from personal favorite textbooks [1]–[9] to lecture notes of MATH338 given by Dr. Abbas al Hakim (American University of Beirut), to miscellaneous remarks and observations made by many contributors from the MathStackExchange and MathOverflow communities.

This self-study document is mainly written to shed light on the many possible viewpoints of stochastic processes and to build a solid theoretical framework that emphasizes on the construction of such processes and their sample spaces, thus allowing one to easily link stochastic analysis with other areas of mathematics. Though mainly concentrated on classical results in the field, some to be written sections will be dedicated to discussing more specialized topics such as the stochastic wave equation and stochastic Hasegawa-Mima equation.

Moreover, one large section on measure theory and probability theory is included, heavily inspired by the MATH303 course given by Dr. Bassam Shayya (American University of Beirut), providing a rigorous footing on which all subsequent sections will rely.

For me personally, this document is journal in which I keep track of all that I learn on stochastic analysis. It's a work in progress by nature and proofs for some results are to be written. As this document evolves, I hope it becomes a reference for myself and others too.

Contents

I	Fundamental Results in Stochastic Analysis	3
1	Measure Theory	4
1.1	General Lebesgue integral and the L^1 space	5
1.2	Measures on topological spaces. Borel σ -algebra and Radon Measures	15
1.3	Outer measures and product measures	16
1.4	Lebesgue-Stieltjes measure on \mathbb{R}^n	17
1.5	L^p Spaces	19
1.6	Absolute continuity and Radon-Nikodym theorem	20
2	Basic Probability Theory	21
2.1	Sample Spaces, Measurable Events and Probability Measures	21
2.2	Random Variables, Density Functions and the Push-Forward Measure	22
2.3	Conditioning over σ -algebras and Independence	25
2.4	Modes of convergence and fundamental theorems	26
3	Vector Valued Measures and Measure Valued Random Variables	28
3.1	Bochner Spaces: Measurability, Integration and Duality	28
3.2	Bochner Integral	28
3.3	Vector Measures	28
3.4	Introduction to Random Measures and Intensity	28
4	General Stochastic Processes	29
4.1	Definition, Existence and Measurability	29
4.2	Pathwise, Stochastic and Feller Continuity	32
4.3	Martingales and Stopping Times	33
4.4	Markov Processes and Feller Semi-Group	35
4.5	Gaussian Processes, Tempered Measures and White Noise	36
4.6	Weiner Process	40
4.7	Lévy and Jump Processes	45
5	Itô Calculus and Elementary Stochastic Differential Equations	46
5.1	Integration with respect to Brownian motion	46
5.2	Weakening defining conditions for Itô integral	50
5.3	Itô process and Itô-Doeblin Formula	51
5.4	SDE's and the Markov property	51
5.5	Feynman-Kac and Fokker-Planck Equations	54
5.6	Examples from finance and economics	54
6	General Stochastic Integration	54
6.1	Generalized Itô Integral	54
6.2	Itô-Doeblin formula for jump processes	54
6.3	*Functional Itô calculus and stochastic integral representation of martingales	54

II	Functional Analysis and Partial Differential Equations	55
7	Classical theory and fundamental equations	56
7.1	Fundamental existence theorems	56
7.2	Poisson equation	56
7.3	Diffusion equation	56
7.4	Wave equation	56
8	Hilbert Spaces	57
8.1	Elementary properties	57
8.2	Lax-Milgram	57
8.3	Reproducing kernel Hilbert spaces	57
9	Sobolev Spaces	58
9.1	Defintion, characterization, completeness and duality of $W^{m,p}(\Omega)$	58
9.2	Sobolev embeddings	58
9.3	The trace operator and fractional Sobolev spaces	58
9.4	Weak formulation of boundary value problems	58
9.5	*Weighted Sobolev Spaces and Non-Linear Potential Theory	58
9.6	*Sobolev spaces on manifolds	58
III	Special Topics	59
10	Stochastic Heat Equation	59
11	Stochastic Wave Equation	59
12	The Classical and Stochastic Hasegawa-Mima equation	59
13	Stochastic Integration in UMD spaces	59
14	Miscellaneous Remarks and Observations	59

Part I

Fundamental Results in Stochastic Analysis

1 Measure Theory

Measure theory, a field whose birth stemmed from the need for rigorous foundations of integration, is ubiquitous in modern day analysis. It was largely, but not solely, an answer to the deficiencies of the classical Riemann integral, which imposed conditions on integrable functions that were later deemed too restrictive. It's development is largely credited to the French mathematician Henri Lebesgue, but references go as far back as 19th century German mathematician Karl Weierstrass considered to be the father of modern day analysis.

The implications of Measure theory are far reaching in understanding geometric quantities such as areas and volumes, but one of its surprisingly intuitive features is its ability to quantify non-physical entities such as information and likelihood. For those reasons, it is the language choice for fields such as Probability Theory, Stochastic Processes, Harmonic Analysis and Partial Differential Equations. In fact, many real world problem could only be solved (or even understood!) when formulated in terms of measures, σ -algebras and Lebesgue integrals. One can dare say that even the most advanced topics in analysis are about showing the existence of certain measures and understanding their properties.

Inspired by the Math 303 course given by professor Bassam Shayya (American University of Beirut) and texts such [8], this large section seeks to explore measures from the ground up, aiming for maximum generality whenever it's possible.

1.1 General Lebesgue integral and the L^1 space

Definition (Measurable Space). Let X be any set. Suppose that there exists a collection of subsets \mathcal{M} of X with the following properties:

- (i) $\emptyset, X \in \mathcal{M}$.
- (ii) If $E \in \mathcal{M}$ then $E^c \in \mathcal{M}$.
- (iii) \mathcal{M} is closed under countable union.

Then we call \mathcal{M} a σ -algebra and the pair (X, \mathcal{M}) a measurable space. //

Definition. A function f from the measure space (X, \mathcal{F}) to the measure space (Y, \mathcal{G}) is said to be measurable if for all $E \in \mathcal{G}$ we have $f^{-1}(E) \in \mathcal{F}$. We denote $\mathcal{L}^0(X, Y)$ the set of all measurable functions from X to Y . //

Proposition 1.1. Let (X, \mathcal{M}_1) and (Y, \mathcal{M}_2) and (Z, \mathcal{M}_3) be measurable spaces and let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be measurable. Then $g \circ f$ is measurable.

Proof. ■

In many cases when the target space has an additional structure, such as topological or algebraic (or both), one asks whether measurability is compatible with these structures. For example if the target space has a topology, and a sequence of measurable functions $\{f_n\}$ converges pointwise to f , is f measurable? If the target space is a vector space equipped with a linear combination measurable functions also measurable? The answer is positive if $Y = \overline{\mathbb{R}}$, but even more generally when Y is a separable Banach space or if X is a perfect measure space and Y is any Banach space. In this section we only study the case when $Y = \overline{\mathbb{R}}$ and leave the more general case to Section 8.

First, let us equip the target space Y with a suitable σ -algebra that encodes the topological structure that it already has.

Definition (Borel σ -algebra). Let (X, \mathcal{T}) be a topological space. The Borel σ -algebra $\mathcal{B} = \mathcal{B}(X)$ is the smallest σ -algebra containing the topology \mathcal{T} . //

For the rest of the section, we will assume that the target space is always the extended real number line $\overline{\mathbb{R}}$ equipped with the Borel σ -algebra \mathcal{B} , and a function $f : X \rightarrow \overline{\mathbb{R}}$ is said to be Borel measurable if it's measurable w.r.t the Borel σ -algebra on $\overline{\mathbb{R}}$. When (X, Σ) is a measure space, we denote

$$\mathcal{M}(X) := \{f : X \rightarrow \overline{\mathbb{R}} \mid f \text{ is measurable w.r.t } \mathcal{B}\}.$$

Proposition 1.2. Let (X, Σ) be a measurable space. Then $\mathcal{M}(X)$ is a real vector space.

Lemma 1.3. The function $f \in \mathcal{M}(X)$ if and only if for every $x \in \mathbb{R}$ we have $f^{-1}((-\infty, x)) \in \mathcal{M}$.

Proof of Proposition 1.2. Let $f, g \in \mathcal{M}(X)$. Notice that we can write

$$\{f + g < b\} = \bigcup_{r \in \mathbb{Q}} \{f < r\} \cap \{g < b - r\}.$$

and therefore $f + g$ is measurable by the above lemma. Also, it is clear that for fixed $\alpha, x \in \mathbb{R}$ we have $\{\alpha f < x\} = \{f < x/\alpha\}$ is measurable since f is measurable. Therefore by induction we get that any finite combination of measurable functions is measurable. ■

Proposition 1.4. Let (X, Σ) be a measurable space and let $\{f_n\} \subset \mathcal{M}(X)$.

(i) The functions $f = \inf_n f_n$ and $g = \sup_n f_n$ are in $\mathcal{M}(X)$.

(ii) If $f = \limsup f_n$ or $f = \liminf f_n$ then $f \in \mathcal{M}(X)$.

Proof. ■

Lemma 1.5. To show that a function f is Borel measurable, it suffices to show that the set $f^{-1}((-\infty, a])$ is measurable for all $a \in \overline{\mathbb{R}}$.

Proof. ■

Definition. A function $f : (X, \mathcal{M}) \rightarrow (\mathbb{R}, \mathcal{B})$ is called simple if there are sets $E_1, \dots, E_n \in \mathcal{M}$ and constants $c_1, \dots, c_n \in \mathbb{R}$ such that

$$f(x) = \sum_{k=1}^n c_k \cdot \mathbf{1}_{E_k}(x).$$

We denote the space of all simple functions on X as $S(X)$. //

The essential property of measurable functions is that they are pointwise limits of simple functions. This alone helps in understanding and characterizing many other important properties of measurable functions, and most subsequent results in this section are due to this approximation property.

Proposition 1.6. Let $f : X \rightarrow \mathbb{R}$ be a non-negative measurable function. There is a sequence of sets $\{E_n\} \subset \mathcal{M}$ such that

$$f(x) = \sum_{n=1}^{\infty} \frac{\mathbf{1}_{E_n}(x)}{k}. \quad (1)$$

This implies that the space of simple functions $S(X)$ is dense in $L_+^0(X)$ with respect to the topology of pointwise convergence.

Intuition: A necessary condition for (1) to hold is that for all $x \in X$ and all $n \in \mathbb{N}$ we have

$$f_n(x) := \sum_{k=1}^n \frac{\mathbf{1}_{E_k}(x)}{k} \leq f(x).$$

This is the case since the sequence $\{f_n\}$ is increasing with f as it's pointwise limit. Also,

$$f_{n+1}(x) = \begin{cases} f_n(x) & \text{if } x \notin E_{n+1}, \\ f_n(x) + \frac{1}{n+1} & \text{if } x \in E_{n+1}. \end{cases}$$

First let $f_0(x) = 0$ for all $x \in X$. Let

$$E_1 = \{x \in X : f(x) \geq 1\}.$$

If $x \in E_1$ then $f(x) \geq 1$ and hence we define $f_1(x) = f_0(x) + 1 = 1$. If $x \notin E_1$, we set $f_1(x) = f_0(x)$. Hence

$$f_1(x) = f_0(x) + \mathbf{1}_{E_1}(x) = \mathbf{1}_{E_1}(x).$$

Now let

$$E_2 = \left\{ x \in X : f(x) \geq f_1(x) + \frac{1}{2} \right\}$$

If $x \in E_2$ then we set $f_2(x) = f_1(x) + 1/2$, otherwise we set $f_2(x) = f_1(x)$. Therefore we can write

$$f_2(x) = f_1(x) + \frac{1}{2} \mathbf{1}_{E_2}(x) = \mathbf{1}_{E_1} + \frac{1}{2} \mathbf{1}_{E_2}.$$

At this point we have that

$$f_2(x) = \begin{cases} 1 + \frac{1}{2} & \text{if } x \in E_1 \cap E_2, \\ 1 & \text{if } x \in E_1 \setminus E_2, \\ \frac{1}{2} & \text{if } x \in E_2 \setminus E_1, \\ 0 & \text{if } x \notin E_1 \cup E_2. \end{cases}$$

What f_2 is doing is checking that if $f_1(x) + 1/2$ exceeds $f(x)$ then keep $f_1(x)$ as is, otherwise add $1/2$ to $f_1(x)$.

Proof of Proposition 1.6. With E_1 defined as above, define recursively

$$E_n = \left\{ x \in X : f(x) \geq f_{n-1}(x) + \frac{1}{n} \right\} \quad \text{and} \quad f_n(x) = f_{n-1}(x) + \frac{1}{n} \mathbf{1}_{E_n}(x).$$

For each x , it is clear that the non-negative sequence $\{f_n(x)\}$ is non-decreasing and bounded from above by $f(x)$. To show that $f_n(x) \rightarrow f(x)$, it suffices to show that a subsequence of converges to $f(x)$. Let n_0 be the smallest integer such that $1/n_0 \leq f(x)$. Then let $m_0 \geq 1$ be the largest integer such that

$$\frac{1}{n_0} + \frac{1}{n_0 + 1} + \cdots + \frac{1}{n_0 + m_0} \leq f(x).$$

Then let $n_1 > n_0 + m_0$ be the smallest integer such that

$$\sum_{k=1}^{n_0+m_0} \frac{1}{k} + \frac{1}{n_1} \leq f(x),$$

and then m_1 be the largest integer such that

$$\sum_{k=1}^{m_0} \frac{1}{n_0 + k} + \sum_{k=1}^{m_1} \frac{1}{n_1 + k} \leq f(x) \quad \text{so that} \quad \sum_{k=1}^{m_0} \frac{1}{n_0 + k} + \sum_{k=1}^{m_1+1} \frac{1}{n_1 + k} \geq f(x).$$

Then let $n_2 \geq m_1 + n_1 + 1$ be the smallest integer such that

$$\sum_{k=1}^{m_0} \frac{1}{n_0 + k} + \sum_{k=1}^{m_1} \frac{1}{n_1 + k} + \frac{1}{n_2} \leq f(x).$$

We have that

$$f(x) - f_{n_2}(x) = f(x) - \sum_{k=1}^{m_0} \frac{1}{n_0 + k} - \sum_{k=1}^{m_1} \frac{1}{n_1 + k} - \frac{1}{n_2} \leq \frac{1}{n_1 + m_1 + 1} \leq \frac{1}{n_2}.$$

Proceeding in this fashion, we obtain a sequence of integers $n_0 \leq n_1 \leq n_2 \leq \dots \leq n_k$ and $m_0 \leq m_1 \leq \dots \leq m_k$ with $n_{j+1} \geq n_j + m_j + 1$ such that

$$\sum_{j=0}^{k-1} \sum_{i=0}^{m_j} \frac{1}{m_j + i} + \frac{1}{n_k} \leq f(x) \quad \text{and} \quad f(x) - f_{n_k}(x) \leq \frac{1}{n_k},$$

and therefore the sequence $f_{n_k}(x)$ converges to $f(x)$ as desired. ■

Proposition 1.7 (Another approximating sequence). *Let $f \in L_+^0(X)$ and for each $n \in \mathbb{N}$ define*

$$f_n(x) = \begin{cases} 2^{-n}(j-1) & \text{if } 2^{-n}(j-1) \leq f(x) < 2^{-n}j, \\ n & \text{if } f(x) \geq n \end{cases}.$$

Then we have that $f_n \nearrow f$. Furthermore, for any set $E \in \mathcal{M}$ on which f is bounded, the convergences is actually uniform.

Proof. ■

Now that we have established basic properties of real valued measurable functions, we can move on to define measures on a measurable space (X, \mathcal{M}) .

Definition. A measure on a measurable space (X, \mathcal{M}) is a function $\mu : \mathcal{M} \rightarrow [0, \infty]$ such that for $\mu(\emptyset) = 0$ and for any sequence $\{E_n\} \subset \mathcal{M}$ we have

$$\mu \left(\bigcup_{n=1}^{\infty} E_n \right) = \sum_{n=1}^{\infty} \mu(E_n). \quad (2)$$

The triple (X, \mathcal{M}, μ) is called a measure space. //

Theorem 1.8 (Continuity property). *Let (X, \mathcal{M}) be a measurable space and let $\mu : \mathcal{M} \rightarrow \mathbb{R}$ be a function such that $\mu(\emptyset) = 0$. Then μ is a measure if and only if the following hold.*

- (i) μ is finitely additive.
- (ii) For any increasing sequence of measurable sets $\{E_n\}$ we have

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

- (iii) In addition if $\mu < \infty$ then for any sequence of decreasing measurable sets $\{E_n\}$ we have

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

Proof. ■

Theorem 1.9 (Borel-Cantelli). *Let (X, \mathcal{M}, μ) be a measure space and let $\{E_n\} \subset \mathcal{M}$. Then*

$$\sum_{n=1}^{\infty} \mu(E_n) < \infty \quad \text{implies} \quad \mu\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right) = 0.$$

Proof. We have that

$$\mu\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{k=n}^{\infty} E_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} \mu(E_k) = 0.$$

The first equality is due to the continuity property of μ , the inequality is due to countable sub-additivity and the last equality is justified since the limit of the tail of convergent series is 0. ■

Proposition 1.10. *Let (X, \mathcal{M}, μ) be a measure space and $\{E_n\} \subset \mathcal{M}$ such that*

$$\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} E_k,$$

If E is the set of the above equality then $\lim_{n \rightarrow \infty} \mu(E_n) = \mu(E)$.

Theorem 1.11 (Egorov). *Let (X, \mathcal{M}, μ) be a finite measure space. Suppose that $\{f_n\} \subset L^0(X)$ converges to $f \in L^0(X)$ almost everywhere. Then for every $\epsilon > 0$, there is a set $E \in \mathcal{M}$ such that $\mu(E) < \epsilon$ and $f_n \rightarrow f$ uniformly on $X \setminus E$.*

Proof. Let

$$E(n, k) = \left\{ x \in X : |f_n(x) - f(x)| \geq \frac{1}{k} \right\}.$$

Notice that if $x \in X$ is such that $f_n(x) \rightarrow f(x)$, then for any fixed $k \in \mathbb{N}$, x cannot be in infinitely many of the $E(n, k)$'s. Since convergence happens for almost all $x \in X$ this means that

$$\mu(\{x \in X : x \text{ is in infinitely many } E(n, k) \text{'s}\}) = \mu\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E(n, k)\right) = 0, \quad \text{for all } k \in \mathbb{N}.$$

Since $\mu(X) < \infty$ we have by part (iii) of Theorem 1.8 that

$$\lim_{m \rightarrow \infty} \mu\left(\bigcup_{n=m}^{\infty} E(n, k)\right) = 0, \quad \text{for all } k \in \mathbb{N},$$

and therefore for fixed $\epsilon > 0$ and fixed k , there is an integer m_k such that for all $m \geq m_k$ we have

$$\mu\left(\bigcup_{n=m}^{\infty} E(n, k)\right) < \frac{\epsilon}{2^k}.$$

Thus if we define

$$E = \bigcup_{k=1}^{\infty} \bigcup_{n=m_k}^{\infty} E(n, k), \quad \text{then} \quad \mu(E) \leq \sum_{k=1}^{\infty} \mu\left(\bigcup_{n=m_k}^{\infty} E(n, k)\right) < \sum_{k=1}^{\infty} \frac{\epsilon}{2^k} = \epsilon.$$

Also by the definition of E , we have have that for any $k \in \mathbb{N}$ there is an integer $m_k \in \mathbb{N}$ such that for all $x \in X \setminus E$ and all $m \geq m_k$ we have

$$|f_m(x) - f(x)| < \frac{1}{k} \quad \text{so that} \quad \sup_{x \in X \setminus E} |f_m(x) - f(x)| \leq \frac{1}{k},$$

and hence $f_n \rightarrow f$ uniformly on $X \setminus E$ as desired. ■

Now the we have finished setting up basic properties of measurable functions, we are in good shape to define the Lebesgue integral for \mathbb{R} valued functions. The approach would be to define the integral for simple functions and proving some of it's properties. Then using the density of simple functions to extend the definition to $L_+^0(X)$ and then use Motonone Convergence to extend these properties to functions in $L_+^0(X)$.

Definition (Lebesgue integral). Let (X, \mathcal{M}, μ) be a measure space. Define

$$\int_X f d\mu := \sum_{k=1}^n c_k \mu(E_k) \text{ for } f \in S(X).$$

Then use the above to define

$$\int_X f d\mu := \sup \left\{ \int_X s d\mu : s \in S(X) \text{ and } 0 \leq s \leq f \right\} \text{ for } f \in L_+^0(X).$$

We extend this definition for a specific set of functions in $L^0(X)$ namely

$$L^1(X) := \left\{ f \in L^0(X) : \int_X |f| d\mu < \infty \right\},^1$$

This guarantees that the following definition makes sense

$$\int_X f d\mu := \int_X f^+ d\mu - \int_X f^- d\mu \text{ for } f \in L^1(X). \quad (3)$$

If $A \in \mathcal{M}$ we define

$$\int_A f d\mu := \int_X \mathbf{1}_A \cdot f d\mu.$$

//

Remark. For functions $f \in L^0(X) \setminus L_+^0(X)$, if we have

$$f \in E := \left\{ f \in L^0(X) : \text{either } \int_X f^+ d\mu < \infty \text{ or } \int_X f^- d\mu < \infty \right\},$$

then we can use (3) to define the integral of f with the integral possibly being $\pm\infty$.

¹Equivalently equivalently

$$L^1(X) := \left\{ f \in L^0(X) : \int_X f^+ d\mu < \infty \text{ and } \int_X f^- d\mu < \infty \right\}.$$

Lemma 1.12. *For simple functions, the Lebesgue integral has the same operational properties of the Riemann integral and satisfies the same inequalities.*

Proof. Let f and g be simple functions on the measure space (X, \mathcal{M}, μ) and write

$$f = \sum_{i=1}^m c_i \mathbf{1}_{E_i} \quad \text{and} \quad g = \sum_{j=1}^n d_j \mathbf{1}_{F_j}.$$

We can always assume that each of the collections $\{E_i\}$ and $\{F_j\}$ partition of X . This will allows us to write

$$\mathbf{1}_{E_i} = \sum_{j=1}^n \mathbf{1}_{E_i \cap F_j}, \quad \text{so that} \quad f = \sum_{i=1}^m \sum_{j=1}^n c_i \mathbf{1}_{E_i \cap F_j},$$

and similarly that

$$\mathbf{1}_{F_j} = \sum_{i=1}^m \mathbf{1}_{E_i \cap F_j}, \quad \text{so that} \quad g = \sum_{i=1}^m \sum_{j=1}^n d_j \mathbf{1}_{E_i \cap F_j}.$$

We will use this to prove the lemma.

- (i) **(Monotonicity).** Suppose that $f \leq g$. Picking any element $x \in E_i \cap F_j$ tells us that $c_i = f(x) \leq g(x) = d_j$ for any $1 \leq i, j \leq m$ such that $E_i \cap F_j$ is non empty. Therefore,

$$\int_X f d\mu = \sum_{i=1}^m \sum_{j=1}^n c_i \mu(E_i \cap F_j) \leq \sum_{i=1}^m \sum_{j=1}^n d_j \mu(E_i \cap F_j) = \int_X g d\mu.$$

- (ii) **(Linearity).** We have that

$$\int_X (f + g) d\mu = \sum_{i,j=1}^{m,n} (c_i + d_j) \mathbf{1}_{E_i \cap F_j} = \sum_{i,j=1}^{m,n} c_i \mathbf{1}_{E_i \cap F_j} + \sum_{i,j=1}^{m,n} d_j \mathbf{1}_{E_i \cap F_j} = \int_X f d\mu + \int_X g d\mu.$$

- (iii) **(Absolute Value).**

$$\left| \int_X (f + g) d\mu \right| = \left| \sum_{i,j=1}^m (c_i + d_j) \mu(E_i \cap F_j) \right| \leq \sum_{i,j=1}^{m,n} |c_i + d_j| \mu(E_i \cap F_j) = \int_X |f + g| d\mu.$$

which completes the proof. ■

Lemma 1.13. *Let (X, \mathcal{M}, μ) be a measure space. Suppose that there are measurable functions f and g and sequences $\{f_n\}$ and $\{g_n\}$ that converge pointwise to f and g respectively. If $f < g$ and $G_n = \{f_n \leq g_n\}$, then $\lim_{n \rightarrow \infty} \mu(G_n) = \mu(X)$. In addition, if μ is finite then $\lim_{n \rightarrow \infty} \mu(E \setminus G_n) = 0$.*

Proof. For each $x \in X$, since $f(x) < g(x)$ and $f_n(x) \rightarrow f(x)$ and $g_n(x) \rightarrow g(x)$ then there is integer $N \in \mathbb{N}$ such that for all $n \geq N$ we have $f_n(x) < g_n(x)$. But this means that $x \in G_n$ for all $n \geq N$ and therefore

$$x \in E := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} G_k,$$

and thus $E = X$. By part (iii) of Theorem 1.8 we have that $\mu(E_n) \rightarrow \mu(X)$ as desired. ■

Proposition 1.14. *Let f and g be two positive measurable functions and let $E, F \subset \mathcal{M}$.*

(i) *If $f \leq g$ then*

$$\int_X f d\mu \leq \int_X g d\mu.$$

(ii) *If $E \subset F$ then*

$$\int_E f d\mu \leq \int_F f d\mu.$$

Proof. Let s be any simple function such that $0 \leq s \leq f$. Then $s \leq g$ and hence we have (i). Part (ii) follows by applying part (i) to $\mathbf{1}_E f$ and $\mathbf{1}_F f$. ■

Theorem 1.15 (Monotone Convergence Theorem). *Let $\{f_n\}$ be an increasing sequence of positive measurable functions that converge pointwise to f . Then*

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

Proof. Let s be any simple function such that $0 \leq s \leq f$ and let $0 < \alpha < 1$ be arbitrary. Let $G_n = \{f_n \geq \alpha s\}$ then it is clear that every $x \in X$ is eventually in G_n for all $n \geq N(x)$ and $G_n \subset G_{n+1}$. Therefore $\{G_n\}$ is an increasing sequence of sets whose union is X . Therefore,

$$\lim_{n \rightarrow \infty} \int_{G_n} s d\mu = \lim_{n \rightarrow \infty} \int_X \mathbf{1}_{G_n} \cdot s d\mu = \lim_{n \rightarrow \infty} \sum_{k=1}^m c_k \mu(E_k \cap G_n) = \sum_{k=1}^m c_k \mu(E_k) = \int_X s d\mu.$$

Also we have that

$$\int_{G_n} \alpha s d\mu \leq \int_{G_n} f_n d\mu \leq \int_X f_n d\mu.$$

And therefore by taking limits

$$\alpha \int_X s d\mu \leq \lim_{n \rightarrow \infty} \int_X f_n d\mu.$$

Since this is true for all $\alpha \in (0, 1)$, then by taking limit as $\alpha \rightarrow 1$ this inequality becomes true for $\alpha = 1$. Since s was arbitrary, we get

$$\int_X f d\mu \leq \lim_{n \rightarrow \infty} \int_X f_n d\mu,$$

The other reverse inequality follows immediately from monotonicity. ■

Corollary 1.15.1. *The Lebesgue integral for positive measurable functions has the same operational properties as the Riemann integral and satisfies the same inequalities.*

Proof. Let f, g be positive measurable functions on the measure space (X, \mathcal{M}, μ) . Let $\{f_n\}$ and $\{g_n\}$ be sequences of simple functions increasing to f and g respectively.

(i) **(Linearity).** Without loss of generality assume that f and g are non-negative.

$$\int_X (f + g) d\mu = \lim_{n \rightarrow \infty} \int_X (f_n + g_n) d\mu = \lim_{n \rightarrow \infty} \left[\int_X f_n d\mu + \int_X g_n d\mu \right] = \int_X f d\mu + \int_X g d\mu.$$

(ii) **(Absolute Value).**

$$\begin{aligned}
\left| \int_X (f + g) d\mu \right| &= \left| \lim_{n \rightarrow \infty} \left(\int_X f_n d\mu + \int_X g_n d\mu \right) \right| = \lim_{n \rightarrow \infty} \left| \int_X f_n d\mu + \int_X g_n d\mu \right| \\
&\leq \lim_{n \rightarrow \infty} \left| \int_X f_n d\mu \right| + \lim_{n \rightarrow \infty} \left| \int_X g_n d\mu \right| \\
&\leq \lim_{n \rightarrow \infty} \int_X |f_n| d\mu + \lim_{n \rightarrow \infty} \int_X |g_n| d\mu \\
&= \int_X |f| d\mu + \int_X |g| d\mu.
\end{aligned}$$

■

Corollary 1.15.2. *Suppose that $\{f_n\}$ is a non-negative sequence that increases to f almost everywhere on X . Then*

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

Here we have used the monotone convergence theorem to prove linearity of the Lebesgue integral, an approach similar to the one in [8]. This approach seems natural as the integral is defined as limit of integrals of simple functions, hence we extend the properties of the Lebesgue integral of simple functions to the Lebesgue integral of general measurable functions. However, some authors would argue that proving MCT before the algebraic properties of the Lebesgue integral is premature. Both points are valid, but the former is more suitable in the context of probability theory and more specifically in the construction of the Itô integral in Section 4.

Theorem 1.16 (Fatou's Lemma). *Let $\{f_n\}$ be a sequence of positive measurable functions. Then*

$$\int_X \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int_X f_n d\mu,$$

where both sides can be equal to $+\infty$.

Proof. Let $g_m = \inf_{k \geq m} \{f_k\}$ so that $g_m \leq f_n$ when $m \leq n$. This tells us that for all $m \leq n$ we have

$$\int_X g_m d\mu \leq \int_X f_n d\mu \quad \text{so that} \quad \int_X g_m d\mu \leq \liminf_{n \rightarrow \infty} \int_X f_n d\mu.$$

Now $\{g_m\}$ is an increasing sequence of measurable functions that converge pointwise to $\liminf_{n \rightarrow \infty} f_n$ and therefore by MCT we have

$$\lim_{m \rightarrow \infty} \int_X g_m d\mu = \int_X \lim_{m \rightarrow \infty} g_m d\mu = \int_X \liminf_{n \rightarrow \infty} f_n d\mu,$$

and the inequality is proved. ■

Corollary 1.16.1. *If f is non-negative measurable then*

$$f = 0 \text{ almost everywhere on } X \iff \int_X f d\mu = 0.$$

Proof. Assume that the integral of f is 0 and let

$$E_n = \left\{ x \in X : f(x) > \frac{1}{n} \right\}.$$

Then by definition $f \geq (1/n) \cdot \mathbf{1}_{E_n}$ and therefore

$$0 = \int_X f d\mu \geq \int_{E_n} f d\mu \geq \frac{1}{n} \mu(E_n),$$

and hence $\mu(E_n) = 0$. It follows that

$$\mu(\{x \in X : f(x) > 0\}) = \mu\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu(E_n) = 0,$$

and hence $f = 0$ almost everywhere. Conversely, suppose that $f = 0$ almost everywhere. This means that $\mu(\{f > 0\}) = 0$. Now let $f_n = n \cdot \mathbf{1}_{\{f > 0\}}$. It is clear that $f \leq \liminf f_n$ and

$$\int_X f_n d\mu = n\mu(E) = 0.$$

Then by Fatou's lemma we obtain

$$\int_X f d\mu \leq \int_X \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int_X f_n d\mu = 0,$$

which concludes the proof. ■

Theorem 1.17 (Dominated Convergence). *Suppose we are given sequences $\{f_n\} \subset L^0(X)$ and $\{g_n\} \subset L^0_+(X)$ such that*

- (i) $\lim_{n \rightarrow \infty} f_n = f$.
- (ii) $\lim_{n \rightarrow \infty} g_n = g \in L^1(X)$.
- (iii) $\lim_{n \rightarrow \infty} \int g_n d\mu = \int g d\mu$.
- (iv) $|f_n| \leq g_n$ for all $n \in \mathbb{N}$.

Under these assumptions we have

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

Proof. To start the proof, define

$$\varphi_n = g_n + g - |f_n - f|$$

Then φ_n is positive measurable since

$$|f_n| \leq g_n \implies |f| \leq g \implies \varphi_n \geq g_n + g - (|f_n| + |f|) \geq g_n + g - (g_n + g) = 0.$$

Also notice that $\varphi_n \rightarrow 2g$ as $n \rightarrow \infty$ almost everywhere on E . Now by Fatou's Lemma

$$\int 2g d\mu = \int 2 \lim_{n \rightarrow \infty} g_n d\mu = \int \lim_{n \rightarrow \infty} \varphi_n d\mu = \int \liminf_{n \rightarrow \infty} \varphi_n d\mu \leq \liminf_{n \rightarrow \infty} \int \varphi_n d\mu,$$

and

$$\varphi_n \leq g_n + g \implies \int \varphi_n d\mu \leq \int (g_n + g) d\mu \implies \limsup_{n \rightarrow \infty} \int \varphi_n d\mu \leq \int 2g d\mu,$$

and thus

$$\lim_{n \rightarrow \infty} \int \varphi_n d\mu = \int 2g d\mu.$$

Therefore we get

$$\int (g_n + g) d\mu = \int \varphi_n d\mu + \int |f_n - f| d\mu$$

Letting $n \rightarrow \infty$ we get the desired result. ■

1.2 Measures on topological spaces. Borel σ -algebra and Radon Measures

Definition (Borel σ -algebra, Borel measure). If (X, \mathcal{T}) is a topological space, we define the Borel algebra $\mathcal{B} := \mathcal{B}(X)$ to be the smallest σ -algebra containing \mathcal{T} . Any measure defined on \mathcal{B} is a Borel measure. //

Definition (Regularity). A measure μ is called outer regular on Borel-measurable E if

$$\mu(E) = \inf\{\mu(U) : E \subset U, U \text{ open}\}$$

μ is called inner regular on Borel-measurable E if

$$\mu(E) = \sup\{\mu(K) : K \subset E, K \text{ compact}\}.$$

μ is called regular if it is both inner and outer regular on all Borel sets. //

Definition (Radon measure). A Radon measure is a Borel measure that is locally finite and inner regular on open sets. //

Proposition 1.18. *Let X be a locally compact Hausdorff space. Then any Radon measure is outer regular on all Borel sets and finite on all compact sets.*

Proposition 1.19. *A Radon measure μ is inner regular on all σ -finite sets. This implies the following.*

- (i) *If μ is σ -finite then μ is regular.*
- (ii) *If X is σ -compact, then every Radon measure on X is regular.*

Proposition 1.20. *Let (X, \mathcal{B}, μ) be a σ -finite Radon measure space and let $E \in \mathcal{B}$.*

- (i) *For any $\epsilon > 0$, there is open set G and a closed set F such that $F \subset E \subset G$ and $\mu(G \setminus F) < \epsilon$.*
- (ii) *There is a G_δ set G and an F_σ set F such that $F \subset E \subset G$ and $\mu(G \setminus F) = 0$.*

Theorem 1.21 (Lusin). *Let X be a locally compact Hausdorff topological space equipped with the Borel σ -algebra and a Radon measure μ . Let $f \in L^0(X)$ such that $\mu(\{f \neq 0\}) < \infty$. For every $\epsilon > 0$, there a function $g \in C_c(X)$ such that*

- (i) $\mu(\{f \neq g\}) < \epsilon$.
- (ii) *If f is bounded then one can have $\|g\|_{L^\infty(\mu)} \leq \|f\|_{L^\infty(\mu)}$.*

1.3 Outer measures and product measures

Definition. An outer measure on a set X is a function $\mu^* : 2^X \rightarrow [0, \infty]$ such that $\mu^*(\emptyset) = 0$ and for any sequence $\{E_n\} \subset 2^X$ that cover a set $E \in 2^X$ we have

$$\mu^*(E) \leq \sum_{n=1}^{\infty} \mu^*(E_n). \quad (4)$$

This property is called countable sub-additivity. //

Outer measures are used in constructing some measures, such as the Lebesgue-Stieltjes measure on \mathbb{R}^n and general product measures. This is highlighted by the following theorem.

Theorem 1.22 (Caratheodory). *Suppose X has an outer-measure μ^* . Let \mathcal{M} be the collection of all subset A of X such that for all $E \in 2^X$ we have*

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A) \quad (5)$$

*Then \mathcal{M} is a σ -algebra and $\mu^*_{|\mathcal{M}}$ is a measure.*

Proof. ■

Condition (5) has an intuitive explanation in terms of events and probability, as will be explained in section 1.6.

Theorem 1.23 (Hahn-Kolmogorov). *Let (X, \mathcal{F}, μ) and (Y, \mathcal{G}, ν) be σ -finite measure spaces and let $\mathcal{F} \otimes \mathcal{G} := \sigma(\mathcal{F} \times \mathcal{G})$. If we define $\eta : \mathcal{F} \times \mathcal{G} \rightarrow \mathbb{R}$ as*

$$\eta(F \times G) = \mu(F)\nu(G).$$

then η extends to a unique measure on $\mathcal{F} \otimes \mathcal{G}$.

Proof. ■

Definition. If (X, \mathcal{F}, μ) and (Y, \mathcal{G}, ν) are σ -finite measure spaces then we denote

$$(X \times Y, \mathcal{F} \otimes \mathcal{G}, \mu \times \nu)$$

the product measure space constructed in Theorem 1.23. //

1.4 Lebesgue-Stieltjes measure on \mathbb{R}^n

Definition (Lebesgue-Stieltjes outer measure on \mathbb{R}). Let $F : \mathbb{R} \rightarrow \mathbb{R}^+$ be an increasing function. For an interval $I = (a, b)$ in \mathbb{R} define

$$\lambda^*(I) = F(b^-) - F(a^+).$$

Now let E be any subset of \mathbb{R} . We define

$$\lambda^*(E) = \inf \left\{ \sum_n \lambda^*(I_n) \mid \{I_n\} \text{ countable covering of } E \text{ with bounded open intervals} \right\},$$

where infimum can be $+\infty$. It is clear λ^* satisfies countable subadditivity. //

Notice that we do not assume that $\lambda^*(\emptyset) = 0$ yet since we can deduce it from countable subadditivity, as showcased in the following.

Lemma 1.24. *Let D_f be the set of discontinuities of a real function f . Then D_f is countable.*

Proof. Let D_f be the set of discontinuities of f . Then either we have $f(x^-) \leq f(x) < f(x^+)$ or $f(x^-) < f(x) \leq f(x^+)$. In the first case, there is a rational number $q(x) \in \mathbb{Q}$ such that $f(x) < q < f(x^+)$ and in the second case $f(x^-) < q < f(x)$. It is easy to see that $q : D \rightarrow \mathbb{Q}$ is injective. ■

Proposition 1.25. *Suppose $x \notin D_f$, then $\lambda^*(\{x\}) = 0$. Since $\emptyset \subset \{x\}$ this implies that $\lambda^*(\emptyset) = 0$ and hence λ^* is an outer measure.*

Proof. Since D_f is countable, then $\mathbb{R} \setminus D_f$ is dense. For $x \in \mathbb{R} \setminus D_f$, let $\{a_n\}$ and $\{b_n\}$ be sequences in $\mathbb{R} \setminus D_f$ such that $a_n \nearrow x$ and $b_n \searrow x$. By definition we will then have

$$\lambda^*(\{x\}) \leq \lambda^*((a_n, b_n)) = F(b_n) - F(a_n),$$

and since F is continuous at x then taking limits in the above equation completes the proof. ■

Definition. Let $(\mathbb{R}, \Sigma^{(1)}, \lambda)$ be the measure space obtained by restricting λ^* to measurable sets. When $F(x) = x$ we call λ the Lebesgue measure. //

Proposition 1.26. $\Sigma^{(1)}$ contains the Borel σ -algebra \mathcal{B} on \mathbb{R} .

Definition. For every $x \in \mathbb{R}$, pick an element $v \in x + \mathbb{Q}$ such that $v \in [0, 1]$. The collection of all such v 's is called a Vitali set. //

Proposition 1.27 (Non measurability of Vitali set). *We have that $V \notin \Sigma^{(1)}$.*

Lebesgue-Stieltjes product measure

We now proceed with constructing the Lebesgue measure on \mathbb{R}^n for $n \geq 2$. There are two approaches: one would be to use the Lebesgue measure λ defined on $(\mathbb{R}, \Sigma^{(1)})$ and use it to construct a product measure structure on \mathbb{R}^n inductively. This approach draws parallels with the one used to construct the *coin tossing space*, a fundamental and intuitive example of a *probability space* which is product of "smaller" coin tossing spaces. Another approach would be to construct a Lebesgue outer measure on \mathbb{R}^n similar to the Lebesgue outer measure on \mathbb{R} .

Definition (Lebesgue-Stieltjes product measure on \mathbb{R}^n). The Lebesgue-Stieltjes product measure is the measure obtained on the product measure space

$$(\mathbb{R}^n, \bigotimes_{k=1}^n \mathcal{L}(\mathbb{R}), \lambda^n),$$

as defined in the above. If λ is the standard Lebesgue measure on \mathbb{R} we simple call λ^n the Lebesgue measure. //

This measure generalizes the Lebesgue-Stieltjes measure on \mathbb{R} in a natural way so that for simple sets such as boxes $B = I_1 \times \cdots \times I_n$, one has that $\lambda^n(B) = \lambda(I_1) \cdots \lambda(I_n)$, as one generally defines area and volumes of boxes.

Lebesgue-Stieltjes outer-measure

Definition. Let \preceq be the partial order on \mathbb{R}^n defined as follows. If $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ then $x \preceq y$ if and only if $x_j \leq y_j$ for all $j = 1, \dots, n$. //

Definition (Increasing right-continuous function in \mathbb{R}^n). Let $F : \mathbb{R}^n \rightarrow \mathbb{R}$ be any function. Let $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$ be such that $a \preceq b$. For $1 \leq k \leq n$ let

$$S_k = \{(c_1, \dots, c_n) : c_j = a_j \text{ for exactly } k \text{ indices and } c_j = b_j \text{ for the other } n - k \text{ indices}\}.$$

We define

$$F((a, b]) := \sum_{k=0}^n (-1)^k \sum_{s \in S_k} F(s).$$

We say that F is increasing $F((a, b]) \geq 0$. //

Definition (Lebesgue-Stieltjes outer measure on \mathbb{R}^n for $n \geq 2$). Let $F : \mathbb{R}^n \rightarrow \mathbb{R}$ be an increasing right continuous function as in the above definition. For $a, b \in \mathbb{R}^n$ such that $a \prec b$ we define

$$(a, b] = \prod_{k=1}^n (a_k, b_k].$$

We also call $(a, b]$ a box for obvious reasons. Define the set function λ_n^* on boxes as

$$\lambda_n^*((a, b]) = F(a, b].$$

Now for $E \in 2^{\mathbb{R}^n}$ we define

$$\lambda^*(E) = \inf \left\{ \sum_{n=1}^{\infty} \lambda^*(B_n) : \{B_n\} \text{ collection of boxes s.t } E \subset \bigcup_{n=1}^{\infty} B_n \right\}.$$

Then λ_n^* is actually an outer measure. Denote $\mathcal{L}(\mathbb{R}^n)$ the σ -algebra on \mathbb{R}^n obtained by restricting the Lebesgue-Stieltjes outer measure as in the Caratheodory extension theorem 1.22 and λ_n to be the restriction of λ_n^* to $\mathcal{L}(\mathbb{R}^n)$. //

Proposition 1.28. *We have the inclusions*

$$\mathcal{B}(\mathbb{R}^n) \subsetneq \bigotimes^n \mathcal{L}(\mathbb{R}) \subsetneq \mathcal{L}(\mathbb{R}^n).$$

1.5 L^p Spaces

This section is inspired by [3], covering most of the fundamental properties of these spaces and some of their uses.

1.6 Absolute continuity and Radon-Nikodym theorem

Theorem 1.29 (Lebesgue-Radon-Nikodym). *Let μ and ν be finite measures on a measurable space (X, \mathcal{M}) . There is a function $f \in L^0(\mu) \cap L^0(\nu)$ and a μ -null set $F \in \mathcal{M}$ such that for all $E \in \mathcal{M}$ we have*

$$\nu(E) = \int_E f d\mu + \nu(E \cap F).$$

2 Basic Probability Theory

In many ways this section is inspired by excellent books [2], [9]. I personally took measure theory before taking any probability theory, and the definition of measurability was a bit arbitrary for me at first, especially that my source of intuition was always geometry and areas. The more mysterious equation to me was the Caratheodory condition (5), and how it was used to get a measure from an outer measure. This shroud around the notion of measurability was removed as soon as I took probability theory, and understood measurable sets from the point of view of sample spaces and events.

2.1 Sample Spaces, Measurable Events and Probability Measures

Definition. A probability space is a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ such that \mathbb{P} takes values in the interval $[0, 1]$. The set Ω is called a sample space and any measurable set is called an event. //

In more grounded terms, Ω contains all possible outcomes ω of an experiment that can be replicated. An event is therefore a collection of outcomes and events containing only one outcome are called simple events. Now let us say that the experiment was done that the outcome ω has been observed. If $\omega \in E$ then we say the event E happened. But this means that $\Omega \setminus E$ did not happen. Also, if we can tell whether $\omega \in E$ or $\omega \in F$ then the event $E \cup F$ happened, meaning that either or F happened. This suggests that following definition for the set of measurable events.

1. If E is measurable then E^c is measurable.
2. If E and F are measurable then $E \cup F$ is measurable.

The collection of all such events is called a σ -field. Is it still not a σ -algebra as we still need to have countable unions. However, suppose we further have

3. If $E_1 \subset E_2 \subset E_3 \subset \dots$ are measurable then $\cup E_n$ is measurable.
4. If $E_1 \supset E_2 \supset E_3 \supset \dots$ are measurable then $\cap E_n$ is measurable.

The the set of all measurable events becomes closed under countable unions. This makes it easier in some cases to deduce that a collection of sets is actually a σ -algebra.

Remark. Let $\{E_\alpha\}_{\alpha \in J}$ where J is uncountable be a collection of measurable events. From an intuitive point of view, it might seem reasonable to think that if we can tell whether an outcome $\omega \in E_\beta$ for some $\beta \in J$ then we can tell that $E = \bigcup_{\alpha \in J} E_\alpha$ happened (ie E is measurable). In that case, whether a set E is measurable or not is completely determined by whether it contains an outcomes ω such that $\{\omega\}$ is not measurable. There doesn't seem to be a problem at this stage. However, take the case when $\Omega = \mathbb{R}$ and let Σ is a σ -algebra containing the intervals that is also closed under arbitrary unions. It can be easily seen that $\Sigma = 2^{\mathbb{R}}$. But the Vitali set V becomes measurable, contradicting Proposition 1.27. In other words, allowing closure under uncountable unions prevents us from defining the Lebesgue measure on \mathbb{R} . In fact one can show that the only such measure on $\Sigma^{(1)}$ is the zero measure.

2.2 Random Variables, Density Functions and the Push-Forward Measure

Definition (Push-forward of a measure). Let $(\Omega_1, \Sigma_1, \mu)$ be a measure space and (Ω_2, Σ_2) be a measurable space. Let $X : \Omega_1 \rightarrow \Omega_2$ be measurable. The push-forward of μ , denoted by $X_*\mu$ is the a function on Σ_2 such defined by

$$X_*\mu(E) = \mu(\{X \in E\}), \quad \text{for all } E \in \Sigma_2.$$

It is clear that $X_*\mu$ is actually a measure on (Ω_2, Σ_2) . //

Theorem 2.1 (Change of variables). *Let $(\Omega_1, \Sigma_1, \mu)$ be a measure space and (Ω_2, Σ_2) be a measurable space. If $X : \Omega_1 \rightarrow \Omega_2$ is measurable and $X_*(\mu)$ is the push-forward measure of X then for any measurable function $g : \Omega_2 \rightarrow \mathbb{R}$ we have*

$$\int_{\Omega_2} g dX_*(\mu) = \int_{\Omega_1} g \circ X d\mu.$$

Definition (Random variable). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and equip \mathbb{R} with the Borel σ -algebra. A random variable is a measurable function on $X : \Omega \rightarrow \mathbb{R}$.

- (i) **(Expected Value)**. The expected value or mean of X is defined as

$$\mathbb{E}[X] := \int_{\Omega} X d\mathbb{P}.$$

- (ii) **(Variance)** We define variance of X to be

$$\text{Var}(X) := \mathbb{E} \left[(X - \mathbb{E}[X])^2 \right].$$

- (iii) **(Distribution Measure)** The distribution of measure of X is the push-forward measure

$$\mathbb{P}^X := X_*\mathbb{P}.$$

- (iv) **(C.D.F)** The cumulative distribution function of X is the function $F : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$F(x) := \mathbb{P}^X((-\infty, x]) = \mathbb{P}[X \leq x].$$

- (v) **(P.D.F)** If $\mathbb{P}^X \ll \lambda$ then the probability density function is the almost everywhere defined function

$$f_X := \frac{d\mathbb{P}^X}{d\lambda}.$$

Suppose now that Y is another random variable.

- (vi) **(Covariance)**. The covariance of X and Y is

$$\text{Cov}(X, Y) := \mathbb{E} [(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))].$$

(vii) **(Correlation)**. The correlation coefficient of X and Y is defined as

$$\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Note that all of the quantities above that involve expectation can very well be infinite. //

Proposition 2.2. *Let X be a random variable with μ_X and F_X defined as above. We have that*

(i) F_X is increasing.

(ii) F_X satisfies the following limits.

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{and} \quad \lim_{x \rightarrow \infty} F_X(x) = 1.$$

(iii) The Lebesgue-Stieltjes measure induced by F_X is μ_X and

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x d\mu_X = \int_{-\infty}^{\infty} x f_X(x) dx.$$

(iv) If F is differentiable then

$$\frac{dF}{dx} = f_X.$$

Proof. ■

Theorem 2.3 (Chebychev's inequality). *Let X be a random variable with mean μ and variance σ^2 . Then for any real positive constant k we have*

$$\mathbb{P}[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}.$$

Proof. We prove Markov's inequality first and use it to obtain our desired result. Markov's inequality states that if X is non-negative then

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

which follows from

$$\mathbb{E}[X] \geq \int_{\{X \geq a\}} X(\omega) d\mathbb{P}(\omega) \geq \int_{\{X \geq a\}} a d\mathbb{P}(\omega) = a \int_{\Omega} \mathbf{1}_{\{X \geq a\}}(\omega) d\mathbb{P}(\omega) = a \cdot \mathbb{P}[X \geq a].$$

Now we have

$$\mathbb{P}[|X - \mu| \geq k\sigma] = \mathbb{P}[(X - \mu)^2 \geq k^2\sigma^2] \leq \frac{\mathbb{E}[(X - \mu)^2]}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2},$$

as desired. ■

A random variable X induces a probability measure \mathbb{P}^X on \mathbb{R} . This measure is referred to as a probability law on \mathbb{R} . In many situations it is natural to identify the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the space $(\mathbb{R}, \mathcal{B}, \mathbb{P}^X)$ using X . This happens when one wants to study the properties of X that are irrelevant of the nature of the sample space Ω . So instead of looking at X itself, we study the induced objects such as \mathbb{P}^X , F_X or (when it exists) f_X .

Definition (Commonly occurring random variables). //

Definition (Random vector). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random vector \mathbf{X} is a mapping $\mathbf{X} : \Omega \rightarrow \mathbb{R}^d$ that is measurable with respect to the Borel σ -algebra $\mathcal{B}(\mathbb{R}^d)$. In particular, it is a vector (X_1, \dots, X_d) with each component being a random variable.

(i) **(Mean vector)** The mean of \mathbf{X} is defined as

$$\mu := (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d]).$$

(ii) **(Covariance matrix)** The covariance matrix of \mathbf{X} is defined as

$$\Sigma := [\text{Cov}(X_i, X_j)]_{i,j=1}^d.$$

(iii) **(Joint distribution measure)** The distribution measure of \mathbf{X} is defined as

$$\mathbb{P}^{\mathbf{X}} := \mathbf{X}_* \mathbb{P}.$$

(iv) **(Joint c.d.f)** The joint c.d.f of \mathbf{X} is the function $F_{\mathbf{X}} : \mathbb{R}^d \rightarrow \mathbb{R}$ defined as

$$F_{\mathbf{X}}(x_1, \dots, x_d) = \mathbb{P}^{\mathbf{X}} \left(\prod_{k=1}^d (-\infty, x_k] \right) = \mathbb{P}[X_1 \leq x_1, \dots, X_d \leq x_d].$$

//

Theorem 2.4 (n -dimensional Chebychev's inequality). *Let $X : \Omega \rightarrow \mathbb{R}^n$ be a random vector with mean μ and covariance matrix $C = [\text{Cov}(X_i, X_j)]_{i,j=1}^n$. If C is positive definite then for any $k \in \mathbb{R}$,*

$$\mathbb{P} \left[\sqrt{(X - \mu)^T C (X - \mu)} \geq k \right] \leq \frac{N}{k^2}.$$

2.3 Conditioning over σ -algebras and Independence

Definition (Conditional probability). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For any two events $A, B \in \mathcal{F}$ we define

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

A and B are called independent if $\mathbb{P}[A | B] = \mathbb{P}(A)$. //

Definition (Independence). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Two events A and B are called independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Let $\{\mathcal{G}_1, \dots, \mathcal{G}_n\}$ be a collection of sub σ -algebras of \mathcal{F} . then we call this collection independent if for all $A_1 \in \mathcal{G}_1, \dots, A_n \in \mathcal{G}_n$ we have

$$\mathbb{P}\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n \mathbb{P}(A_k).$$

A sequence $\{\mathcal{G}_n\}$ of sub σ -algebras of \mathcal{F} is called independent if \mathcal{G}_{n+1} and $\sigma(\mathcal{G}_1 \cup \dots \cup \mathcal{G}_n)$ are independent for all $n \in \mathbb{N}$. A sequence of random variables $\{X_n\}$ is called independent if $\sigma(X_n)$ is independent of $\sigma(X_1, \dots, X_n)$ are independent for all $n \in \mathbb{N}$. //

Definition (Conditional expectation). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and \mathcal{G} be a sub σ -algebra of \mathcal{F} . The conditional expectation $\mathbb{E}[X | \mathcal{G}]$ is defined as random variable having the following properties.

(i) $\mathbb{E}[X | \mathcal{G}]$ is \mathcal{G} -measurable.

(ii) For all $A \in \mathcal{G}$ we have that

$$\int_A \mathbb{E}[X | \mathcal{G}] d\mathbb{P} = \int_A X d\mathbb{P}.$$

Property (ii) is called partial averaging. //

Definition. Let X and Y be two random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that the joint p.m.f f_{XY} exists. Define

$$f_{X|Y}(x|y) := \frac{f_{XY}(x, y)}{f_Y(y)},$$

as the conditional density of X given Y . //

Proposition 2.5. *If X and Y are two jointly distributed random variables then*

(i) *If X and Y are discrete then*

$$p_X(x) = \sum_{y \in X(\Omega)} p_{X|Y}(x|y) p_Y(y), \quad \forall x \in X(\Omega).$$

(ii) *If X and Y are continuous then*

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dy.$$

2.4 Modes of convergence and fundamental theorems

Definition (Convergence in probability). Let $\{X_n\}$ be a sequence of random variables on a sample space. If there is a random variable X such that for every $\epsilon > 0$ one has

$$\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| \geq \epsilon] = 0,$$

then one says $\{X_n\}$ converges to X in probability. //

Theorem 2.6. *The function*

$$d(X, Y) = \mathbb{E}[\min(|X - Y|, 1)],$$

is complete metric on $L^0(\Omega)$ and $X_n \rightarrow X$ in probability if and only if $d(X_n, X) \rightarrow 0$.

Theorem 2.7 (Weak law of large numbers). *Let $\{X_n\}$ be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that for some $\mu, \sigma \in \mathbb{R}$ we have $\mathbb{E}[X_n] = \mu$ and $\text{Var}[X_n] = \sigma^2$ for all $n \in \mathbb{N}$. We have that*

$$\lim_{n \rightarrow \infty} \bar{X}_n = \lim_{n \rightarrow \infty} \frac{X_1 + \cdots + X_n}{n} = \mu, \quad \text{in probability.}$$

Proof. Since the X_n 's are independent then

$$\sigma_n^2 := \text{Var}[\bar{X}_n] = \text{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\text{Var}[X_1] + \cdots + \text{Var}[X_n]}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Let $\epsilon > 0$ be given. We have by Chebychev's inequality that

$$\mathbb{P}[|\bar{X}_n - \mu| \geq \epsilon] \leq \sigma_n^2 \epsilon^{-2} = \frac{\sigma^2}{n\epsilon^2},$$

which gives the desired result. ■

Definition (Convergence in distribution). Let $\{X_n\}$ be a sequence of random variables and for each $n \in \mathbb{N}$ define $F_n := F_{X_n}$. Let X be a random variable with c.d.f $F := F_X$. We say $\{X_n\}$ converges to X in distribution if

$$\lim_{n \rightarrow \infty} F_n(x) = F(x),$$

for all $x \in \mathbb{R}$ such that F is continuous at x . //

The proof of the following claim is trivial.

Theorem 2.8 (Continuous mapping theorem). *Let $\{X_n\}$ be a sequence of random vectors in \mathbb{R}^n and let $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a continuous function. If $\{X_n\}$ converges to X almost surely, in probability or in distribution then $\{g(X_n)\}$ converges to $g(X)$ in the same way $\{X_n\}$ converges to X .*

Theorem 2.9. *Suppose that $\{X_n\}$, $\{A_n\}$ and $\{B_n\}$ are sequences of random vectors in \mathbb{R}^m , \mathbb{R}^m and \mathbb{R}^{mn} respectively. Furthermore suppose that*

- (i) $\{X_n\}$ converges in distribution to X .
- (ii) $\{A_n\}$ converges in probability to a random vector A .

(iii) $\{B_n\}$ converges in probability to a non-random vector B .

Then we have

$$\lim_{n \rightarrow \infty} A_n X + B_n = AX + B, \quad \text{in distribution.}$$

Theorem 2.10 (Central limit theorem). *Let $\{\mathbf{X}_n\}$ be a sequence of independent random vectors in \mathbb{R}^d with common mean vector μ and covariance matrix Σ . Then*

$$\sqrt{d} \left(\overline{\mathbf{X}}_n - \mu \right) \rightarrow N(0, \Sigma) \quad \text{in distribution,}$$

or equivalently

$$\sqrt{d} \cdot \Sigma^{-\frac{1}{2}} \left(\overline{\mathbf{X}}_n - \mu \right) \rightarrow N(0, I_d) \quad \text{in distribution.}$$

Theorem 2.11 (Generalized Central Limit Theorem). *Under the assumptions of Theorem 2.10, if $f : \mathbb{R}^d \rightarrow \mathbb{R}^m$ is a continuously differentiable function with Jacobian matrix $J(\mathbf{x})$ then*

$$\sqrt{n} \left(f(\overline{\mathbf{X}}_n) - f(\mu) \right) \rightarrow N \left(0, J(\mu) \Sigma J(\mu)^T \right) \quad \text{in distribution.}$$

3 Vector Valued Measures and Measure Valued Random Variables

3.1 Bochner Spaces: Measurability, Integration and Duality

Based on [4].

3.2 Bochner Integral

3.3 Vector Measures

3.4 Introduction to Random Measures and Intensity

4 General Stochastic Processes

4.1 Definition, Existence and Measurability

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A stochastic process is a collection $\{X_t\}_{t \in T}$ of random variables $X_t : \Omega \rightarrow \mathbb{R}^n$ indexed by T . For fixed $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is called a path or a realization of the process. Consider the following notation.

- (i) $X(t) = X(t, \cdot) = X_t$ is the random variable $\omega \mapsto X_t(\omega)$.
- (ii) $X(\omega) = X(\cdot, \omega) = X_\omega$ is the path $t \mapsto X_t(\omega)$.

It will be often useful to view $\{X_t\}$ as a

1. function of two variables $X : T \times \Omega \rightarrow \mathbb{R}^n$ defined by $X(t, \omega) = X_t(\omega)$.
2. a random variable $X : \Omega \rightarrow (\mathbb{R}^n)^T$ such that $\omega \mapsto X(\omega)$.

We alternate freely between the above notation and viewpoints depending on convenience. //

Analogously to real valued random variables, a stochastic process X has distribution space as a $(\mathbb{R}^n)^T$ random variable. Indeed, consider first the Borel σ -algebra \mathcal{B} generated by the product topology on $(\mathbb{R}^n)^I$. Sets of the form

$$S(t, U) := \{f \in (\mathbb{R}^n)^I : f(t) \in U\}, \quad U \text{ open in } \mathbb{R}^n, t \in T,$$

are a subbasis for the product topology and hence a basis for the product topology contains sets of the form,

$$\bigcap_{j=1}^k S(t_j, F_j) = \{f \in (\mathbb{R}^n)^I : f(t_1) \in F_1, \dots, f(t_k) \in F_k\}, \quad F_j \in \mathcal{B}(\mathbb{R}^n). \quad (6)$$

It is clear that $\omega \mapsto X(\cdot, \omega)$ is measurable from (Ω, \mathcal{F}) to $((\mathbb{R}^n)^I, \mathcal{B})$. Hence we can define its distribution measure

$$\mu_X(F) := \mathbb{P}[X \in F] = \mathbb{P}\{\omega \in \Omega : X(\cdot, \omega) \in F\} \quad \text{where } F \in \mathcal{B}((\mathbb{R}^n)^I).$$

For Borel sets B of the form (6) we have

$$\mu_X(B) = \mathbb{P}[X \in B] = \mathbb{P}(\{\omega \in \Omega : X(\cdot, \omega) \in B\}) = \mathbb{P}[X_{t_1}(\omega) \in F_1, \dots, X_{t_k}(\omega) \in F_k].$$

Therefore we can identify Ω with a subset of $(\mathbb{R}^n)^I$ and X can be viewed as a probability measure or probability law μ_X on the measure space $((\mathbb{R}^n)^I, \mathcal{B})$. This is useful in studying properties of stochastic processes when the nature of sample space is not relevant.

Definition (Distinguishing between stochastic processes). Let $\{X\}_{t \in T}$ and $\{Y\}_{t \in T}$ be two stochastic processes.

- (i) The processes have the same finite dimensional distributions if for all $t_1, \dots, t_k \in T$ the random vectors $(X_{t_1}, \dots, X_{t_k})$ and $(Y_{t_1}, \dots, Y_{t_k})$ are equal in distribution.
- (ii) X is called a version of Y if for all $t \in T$ we have $\mathbb{P}[X(t) = Y(t)] = 1$.
- (iii) The two processes are called indistinguishable if for almost all $\omega \in \Omega$ we have $X(\cdot, \omega) = Y(\cdot, \omega)$.

It is immediate that (iii) \implies (ii) \implies (i). //

Constructing stochastic processes using finite dimensional distributions.

The next theorem allows us to construct stochastic processes from a probability law on $(\mathbb{R}^n)^T$ equipped with the Borel σ -algebra \mathcal{B} , given that the law satisfies natural consistency properties.

Theorem 4.1 (Kolmogorov's extension theorem). *Suppose for every $k \in \mathbb{N}$ and every $t_1, \dots, t_k \in T$ we are given a measure ν_{t_1, \dots, t_k} on \mathbb{R}^n . Suppose furthermore that*

$$\nu_{t_{\sigma(1)}, \dots, t_{\sigma(k)}}(F_1 \times \dots \times F_k) = \nu_{t_1, \dots, t_k}(F_{\sigma^{-1}(1)} \times \dots \times F_{\sigma^{-1}(k)})$$

and for all $m \in \mathbb{N}$

$$\nu_{t_1, \dots, t_k}(F_1 \times \dots \times F_k) = \nu_{t_1, \dots, t_k, \dots, t_{k+m}}(F_1 \times \dots \times F_k \times \mathbb{R}^{mn}).$$

Then there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process X such that $t_1, \dots, t_k \in T$ we have

$$\mathbb{P}[X_{t_1} \in F_1, \dots, X_{t_k} \in F_k] = \nu_{t_1, \dots, t_k}(F_1 \times \dots \times F_k), \quad F_j \in \mathcal{B}(\mathbb{R}^n).$$

In other words, there is a probability measure μ_X on $(\mathbb{R}^n)^I$ such that if B is a Borel set in $(\mathbb{R}^n)^I$ of the form (6) then

$$\mu_X(B) = \nu_{t_1, \dots, t_k}(F_1 \times \dots \times F_k).$$

We will prove a slightly more general result.

Definition. A separable metric space (X, d) is universally measurable (u.m.) iff for every law \mathbb{P} on the completion \overline{X} of S there are Borel sets A and B in \overline{X} with $A \subset S \subset B$ and $\mathbb{P}(A) = \mathbb{P}(B)$, so that S is measurable for the (measure-theoretic) completion of P . //

Theorem 4.2. *Let T be any set and let $\{(S_t, B_t)\}_{t \in T}$ be a family of universally measurable spaces. Suppose that for every finite subset S of T we have a probability law \mathbb{P}_S on the measurable space*

$$(S_S, B_S) := \left(\prod_{t \in S} S_t, \bigotimes_{t \in S} B_t \right)$$

and that the collection $\{\mathbb{P}_F : F \subset T, |F| < \infty\}$ is consistent. Then there is a probability measure \mathbb{P} on

$$\left(\prod_{t \in T} S_t, \bigotimes_{t \in T} B_t \right)$$

such that \mathbb{P} restricted to S_F is equal to \mathbb{P}_F for any finite set $F \subset T$.

Proof.

...

Now with $\Omega = \mathbb{R}^T$, defined

$$X(t, \omega) = \omega(t), \quad \omega \in \Omega.$$

Is is then clear that for Borel sets in $F_1, \dots, F_n \subset \mathbb{R}$ we have that

$$\begin{aligned} \mathbb{P}^T[X_{t_1} \in F_1, \dots, X_{t_n} \in F_n] &= \mathbb{P}^T\{\omega \in \Omega : \omega(t_1) \in F_1, \dots, \omega(t_n) \in F_n\} \\ &= \mathbb{P}^T \circ \pi_{TG}^{-1}(F_1 \times \dots \times F_n) \\ &= \mathbb{P}^F(F_1 \times \dots \times F_n). \end{aligned}$$

so that the random vector $\{X_{t_1}, \dots, X_{t_n}\}$ has probability distribution \mathbb{P}^F on \mathbb{R}^F . ■

Measurability of stochastic processes on $T = \mathbb{R}^+$.

In the following we let $T = [0, \infty)$ and equip T with the Borel σ -algebra $\mathcal{B}(T)$.

Definition (Measurability and adaptability). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Equip $T \times \Omega$ and \mathbb{R}^n with the σ -algebras $\mathcal{B}(T) \otimes \mathcal{F}$ and $\mathcal{B}(\mathbb{R}^n)$ respectively. Suppose we are given a filtration $\{\mathcal{F}_t\}_{t \in T}$. A stochastic process $\{X_t\}_{t \in T}$ is called

- (i) measurable if the function $X : T \times \Omega \rightarrow \mathbb{R}^n$ defined by $X(t, \omega) = X_t(\omega)$ is measurable.
- (ii) adapted if for all $t \in T$ we have $\sigma(X_t) \subset \mathcal{F}_t$.
- (iii) progressively measurable for each $t \in T$ we have that $X : T \times \Omega \rightarrow \mathbb{R}^n$ is measurable when the domain and target space are equipped with the σ -algebras $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$ and $\mathcal{B}(\mathbb{R}^n)$ respectively.

We denote the natural filtration of X to be $\{\mathcal{F}_t^X\}$ where $\mathcal{F}_t^X = \sigma(X_s; 0 \leq s \leq t)$. //

Proposition 4.3 (Chung and Doob, 1965). *If a stochastic process X is measurable and adapted to a filtration $\{\mathcal{F}_t\}$ then X is progressively measurable.*

Characterizing square integrable processes using mean and covariance functions.

Theorem 4.4 (Karhunen-Mercer-Loève). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $T = [a, b]$. Consider a square integrable stochastic process $\{X_t\}_{t \in T}$ with zero mean and continuous covariance function C which is positive semi-definite. Define the operator $T : L^2([a, b]) \rightarrow L^2([a, b])$ as*

$$T(f)(s) = \int_a^b C(s, t) f(t) dt.$$

Let $\{e_n\}$ be an orthonormal basis for $L^2([a, b])$ formed by the eigenfunctions of T and let $\{\lambda_n\}$ be it's eigenvalues. For each n , define the random variable

$$Z_t^{(n)} = \sum_{k=1}^n e_k(t) \int_a^b X_s e_k(s) ds.$$

Then $Z_t^{(n)} \rightarrow X_t$ in $L^2(\Omega)$ and for fixed $\omega \in \Omega$ we have $Z_t^{(n)}(\omega) \rightarrow X_t(\omega)$ in $L^\infty([a, b])$.

4.2 Pathwise, Stochastic and Feller Continuity

It is often asked whether a stochastic process has continuous paths for almost all $\omega \in \Omega$. This requires a topological structure on T .

Definition (Continuity of a stochastic process). Let T be any topological space and let $\{X_t\}_{t \in T}$ be a stochastic process.

- (i) X continuous at $t_0 \in T$ if for almost all $\omega \in \Omega$ we have

$$\lim_{t \rightarrow t_0} X_t(\omega) - X_{t_0}(\omega) = 0.$$

It is continuous if the above holds for all $t_0 \in T$.

- (ii) X is continuous in mean at t_0 if

$$\lim_{t \rightarrow t_0} \mathbb{E}[X_t - X_{t_0}] = 0.$$

It is continuous in mean if (iii) holds for all $t_0 \in T$.

- (iii) continuous in probability at t_0 if $\lim_{t \rightarrow t_0} \mathbb{P}[X_t - X_{t_0}] = 0$.

- (iv) continuous if (v) holds for all $t_0 \in T$.

- (v) Feller continuous if for every Borel function φ the function $t \mapsto \mathbb{E}[\varphi(X_t)]$ is continuous.

Condition (ii) is often stated as X has continuous sample paths

//

Theorem 4.5 (Kolmogorov's continuity condition). Let T be a closed cube in \mathbb{R}^n . Suppose the stochastic process $\{X_t\}_{t \in T}$ satisfies the following condition: there are positive constants $C, p \in \mathbb{R}$ and $\gamma > N$ such that

$$\mathbb{E}[|X_t - X_s|^p] \leq C|t - s|^\gamma, \quad \forall s, t \in T.$$

Then there is a continuous version of $\{X_t\}$. In addition, if we call $\{\tilde{X}_t\}_{t \in T}$ this modification and θ is chosen so that $1 \leq \theta < (\gamma - N)/p$ then

$$\sup_{s \neq t} \frac{|X_s - X_t|}{|t - s|^\theta} \in L^p(\Omega).$$

Proof.

■

4.3 Martingales and Stopping Times

Random, optional, and stopping times.

Definition (Stopping time). Let $(\Omega, \mathcal{F}, \mathbb{P})$. A stopping time w.r.t a filtration $\{\mathcal{F}_t\}$ is a random variable $\tau : \Omega \rightarrow \mathbb{R}$ such that

$$\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t, \quad \text{for all } t \in T.$$

The random variable τ is said to be a stopping time for a stochastic process $\{X_t\}$ if it is a stopping time w.r.t $\{X_t\}$'s natural filtration. //

Martingales and convergence theorems.

Theorem 4.6 (Doob's martingale inequality). *Let $\{M_t\}$ be a right-continuous sub-martingale and let $[s, t] \subset \mathbb{R}^+$ be a bounded interval. Then*

$$\mathbb{P} \left[\sup_{s \leq u \leq t} M_u \geq \lambda \right] \leq \frac{\mathbb{E}[M_t^+]}{\lambda}.$$

Proof. ■

Theorem 4.7 (Doob's martingale inequality). *Let $\{M_t\}$ be a continuous martingale and let $[0, t] \subset \mathbb{R}^+$ be a bounded interval. Then for all $p \geq 1$ and $\lambda \in \mathbb{R}^+ \setminus \{0\}$ we have*

$$\mathbb{P} \left[\sup_{0 \leq s \leq t} M_s \geq \lambda \right] \leq \frac{\mathbb{E}[|M_t|^p]}{\lambda^p}.$$

Proof. ■

Definition (Upcrossing). //

Theorem 4.8 (Doob's upcrossing inequality).

Theorem 4.9 (Martingale convergence theorem, version 1). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\{X_t\}$ be a right-continuous sub-martingale with respect to a filtration $\{\mathcal{F}_t\}$. If $\sup_{t \geq 0} \mathbb{E}[X_t^+] < \infty$ then X_t converges pointwise almost surely to a random variable $X \in L^1(\Omega)$.*

Theorem 4.10 (Martingale convergence theorem, version 2). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\{X_t\}$ be a uniformly integrable martingale with respect to a filtration $\{\mathcal{F}_t\}$. There is a random variable $X \in L^1(\Omega)$ such that*

$$X_t = \mathbb{E}[X | \mathcal{F}_t], \text{ (a.s.) and } X_t \rightarrow X \text{ as } t \rightarrow \infty \text{ in } L^1(\Omega).$$

Conversely for any $X \in L^1(\Omega)$ then the process $\{\mathbb{E}[X | \mathcal{F}_t]\}$ is a uniformly integrable martingale.

Optional sampling.

Definition (Stopped process). A process $\{Y_t\}$ is said to be a stopped process if there is a stochastic process $\{X_t\}$ with stopping time τ such that

$$Y(t, \omega) = X(\tau(\omega) \wedge t, \omega), \quad \text{for all } (t, \Omega) \in T \times \Omega.$$

Sometimes the process $\{Y_t\}$ is denoted as $\{X_t^\tau\}$.

//

Theorem 4.11 (Optional sampling theorem). *Let $\{X_t\}$ be a stochastic process with filtration $\{\mathcal{F}_t\}$ and stopping time τ . Suppose that $\{X_t\}$ is a Martingale and let $\{X_t^\tau\}$ be the stopped process obtained from $\{X_t\}$. Then $\{X_t^\tau\}$ is also a Martingale and $\mathbb{E}[X_t^\tau] = \mathbb{E}[X_0]$ for all $t \in T$.*

Theorem 4.12 (Doob's decomposition Theorem).

4.4 Markov Processes and Feller Semi-Group

4.5 Gaussian Processes, Tempered Measures and White Noise

This section is based on [5], [10]–[12].

Definition. Let T be an index set. A process $\{X_t\}_{t \in T}$ is called a Gaussian process if for all $t_1, \dots, t_n \in T$ the random vector $(X(t_1), \dots, X(t_n))$ is a k -dimensional Gaussian random vector. //

Theorem 4.13 ((Some version of) Bochner-Minlos). *Let T be any set. Let $m : T \rightarrow \mathbb{R}$ be any function and $C : T \times T \rightarrow \mathbb{R}$ be any positive definite kernel² with $C(s, t) = C(t, s)$ for all $s, t \in T$. There is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Gaussian process $\{W_t\}_{t \in T}$ with mean function m and covariance C .*

Proof. For any finite $F = \{t_1, \dots, t_n\} \subset T$, let C_F be the matrix of $[C(t_i, t_j)]_{i,j=1}^n$. Consider the normal law $N(0, C_F)$ on \mathbb{R}^F . If $G \subset F$, and t is a linear function on \mathbb{R}^G , or equivalently a point of \mathbb{R}^G with the usual inner product, then $t \circ f_{FG}$ on \mathbb{R}^F is the linear form with the coordinates of t on G and 0 on $F \setminus G$. Thus we have equality of inner products

$$(C_F(t \circ f_{FG}), t \circ f_{FG}) = (C_G(t), t).$$

Then $N(0, C_F) \circ f_{FG}^{-1} = N(0, C_G)$ since each has the characteristic function $\exp(-(C_G(t), t)/2)$. So the family of probability laws

$$\{N(0, C_F) : \text{for all finite } F \subset T\}$$

is consistent and Kolmogorov's theorem applies. Hence there is a probability measure \mathbb{P}_T on $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$ such that $\mathbb{P}^T \circ \pi_{TF}^{-1} = \mathbb{P}_F$ for all finite subsets F of T . ■

Corollary 4.13.1. *Let (M, \mathcal{G}, σ) be a σ -finite measure space. Define the functions $m : \mathcal{G} \rightarrow \mathbb{R}$ and $C : \mathcal{G} \times \mathcal{G} \rightarrow \mathbb{R}$ as*

$$m(A) = 0, \quad C(A, B) = \sigma(A \cap B), \quad \text{for all } A, B \in \mathcal{G}.$$

Then there is a Gaussian process $\{W_A^{(\sigma)}\}_{A \in \mathcal{G}}$ on that space with mean m and covariance C .

Definition. The Schwartz space of real valued rapidly decreasing smooth functions on \mathbb{R}^d is defined as

$$\mathcal{S}(\mathbb{R}^d) = \left\{ f \in C^\infty(\mathbb{R}^d) : \sup_{x \in \mathbb{R}} (1 + |\mathbf{x}|^k) |\partial^\alpha f(\mathbf{x})| < \infty, \text{ for all integers } k \text{ and multi-indices } \alpha \right\}.$$

It is a Fréchet space³ with topology generated by the countable family of semi-norms

$$\mathcal{P} = \left\{ \|f\|_{k,\alpha} := \sup_{x \in \mathbb{R}} (1 + |\mathbf{x}|^k) |\partial^\alpha f(\mathbf{x})| < \infty, \text{ for all integers } k \text{ and multi-indices } \alpha \right\}.$$

²A function $C : T \times T \rightarrow \mathbb{R}$ is called a positive definite kernel if for any finite set $F \subset T$ then the matrix $[C(s, t)]_{s,t \in F}$ is non-negative definite.

³If X is any real vector space and \mathcal{P} is a countable collection of semi-norms such that

- If $x \in X$ and $\|x\| = 0$ for all $\|\cdot\| \in \mathcal{P}$ then $x = 0$,
- If $\{x_n\}$ is a Cauchy for every $\|\cdot\| \in \mathcal{P}$ then there is an $x \in X$ such that for all $\|\cdot\| \in \mathcal{P}$ we have $\|x_n - x\| \rightarrow 0$,

then one can define a complete metric on X as

$$d(x, y) = \sum_{n \in \mathbb{N}} 2^{-n} \frac{\|x - y\|_n}{1 + \|x - y\|_n}, \quad x, y \in X.$$

The resulting metric space is called a Fréchet space.

It's dual space $\mathcal{S}'(\mathbb{R}^d)$ is called the space of tempered distributions and is equipped with the weak* topology.⁴ It is equipped with the Borel σ -algebra $\mathcal{B}(\mathcal{S}')$ which can be shown to be generated by *cylinder sets*, which are sets of the form

$$\{\xi \in \mathcal{S}' : \langle \xi, \varphi_1 \rangle \in F_1, \dots, \langle \xi, \varphi_n \rangle \in F_n\}$$

where $\varphi_1, \dots, \varphi_n \in \mathcal{S}(\mathbb{R}^d)$ and $F_1, \dots, F_n \in \mathcal{B}(\mathbb{R})$. //

Theorem 4.14 (Bochner-Minlos). *Let $\mathcal{S}' := \mathcal{S}'(\mathbb{R}^d)$. There is a probability measure \mathbb{P} on $(\mathcal{S}', \mathcal{B}(\mathcal{S}'))$ such that*

$$\mathbb{E} \left[e^{i\langle \cdot, \varphi \rangle} \right] = \int_{\mathcal{S}'} e^{i\langle \xi^*, \varphi \rangle} d\mathbb{P}(\xi^*) = e^{-\frac{1}{2}\|\varphi\|^2}.$$

Proposition 4.15. *Let $\varphi_1, \dots, \varphi_n \in \mathcal{S}(\mathbb{R}^d)$ be n orthonormal functions⁵ in $L^2(\mathbb{R}^d)$ and let μ_n be the measure on \mathbb{R}^n defined as*

$$d\mu_n := e^{-\frac{n}{2}} e^{-\frac{1}{2}|\mathbf{x}|^2} d\mathbf{x} = e^{-\frac{n}{2}} e^{-\frac{1}{2}|(x_1, \dots, x_n)|^2} dx_1 \dots dx_n,$$

Then the random vector

$$\xi \mapsto (\langle \xi, \varphi_1 \rangle, \dots, \langle \xi, \varphi_k \rangle), \quad \text{for all } \xi \in \mathcal{S}',$$

has distribution measure μ_n .

Proof. To prove the above theorem, it suffices to show that for all $f \in L^1(\mu_k)$ we have

$$\int_{\mathcal{S}'} f(\langle \xi, \varphi_1 \rangle, \dots, \langle \xi, \varphi_k \rangle) d\mathbb{P}(\xi) = \int_{\mathbb{R}^n} f(\mathbf{x}) d\mu_n(\mathbf{x}).$$

Start with $f \in C_c^\infty(\mathbb{R}^n) \subset L^1(\mu_n)$. If \hat{f} is the Fourier transform of f then we have the following equality⁶

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} \hat{f}(\mathbf{y}) e^{i\mathbf{x} \cdot \mathbf{y}} d\mathbf{y}, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

Therefore we have that

$$\begin{aligned} \int_{\mathcal{S}'} f(\langle \xi, \varphi_1 \rangle, \dots, \langle \xi, \varphi_k \rangle) d\mathbb{P}(\xi) &= \int_{\mathcal{S}'} (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} \hat{f}(\mathbf{y}) e^{i(\langle \xi, \varphi_1 \rangle, \dots, \langle \xi, \varphi_n \rangle) \cdot \mathbf{y}} d\mathbf{y} d\mathbb{P}(\xi) \\ &= (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} \hat{f}(\mathbf{y}) \int_{\mathcal{S}'} e^{i\langle \xi, \sum_{j=1}^n y_j \varphi_j \rangle} d\mathbf{y} d\mathbb{P}(\xi) \\ &= (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} \hat{f}(\mathbf{y}) \exp \left(-\frac{1}{2} \left\| \sum_{j=1}^n y_j \varphi_j \right\|_{L^2(\mathbb{R}^d)}^2 \right) d\mathbf{y} \\ &= (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} \hat{f}(\mathbf{y}) e^{-\frac{1}{2}|\mathbf{y}|^2} d\mathbf{y}, \end{aligned}$$

⁴It is the smallest topology on \mathcal{S}' that ensures all evaluation maps $f_\varphi : \mathcal{S}' \rightarrow \mathbb{R}$ of the form

$$f_\varphi(\xi) = \langle \xi, \varphi \rangle, \quad \text{for some } \varphi \in \mathcal{S}(\mathbb{R}^d),$$

are continuous

⁵Meaning that we have

$$\int_{\mathbb{R}^d} \varphi_i(\mathbf{x}) \varphi_j(\mathbf{x}) d\mathbf{x} = \delta_{ij}, \quad i, j = 1, \dots, k.$$

⁶This means that f is the inverse Fourier transform of it's Fourier tranform. This is due to the fact that $C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$ and the Fourier transform in an automorphism on $\mathcal{S}(\mathbb{R}^n)$.

where the third inequality is justified by the Bochner-Minlos theorem and in the last inequality we have used orthonormality of the φ_j 's. Now we have

$$\begin{aligned}
(2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} \hat{f}(\mathbf{y}) e^{-\frac{1}{2}|\mathbf{y}|^2} d\mathbf{y} &= (2\pi)^{-n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(\mathbf{x}) e^{-\frac{1}{2}|\mathbf{y}|^2 - i\mathbf{x} \cdot \mathbf{y}} d\mathbf{x} d\mathbf{y} \\
&= (2\pi)^{-n} \int_{\mathbb{R}^n} f(\mathbf{x}) \int_{\mathbb{R}^n} e^{-\frac{1}{2}|\mathbf{y}|^2 - i\mathbf{x} \cdot \mathbf{y}} d\mathbf{y} d\mathbf{x} \\
&= (2\pi)^{-n} \int_{\mathbb{R}^n} f(\mathbf{x}) \cdot (2\pi)^{\frac{n}{2}} e^{-\frac{1}{2}|\mathbf{x}|^2} d\mathbf{x}^8 \\
&= (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f(\mathbf{x}) e^{-\frac{1}{2}|\mathbf{x}|^2} d\mathbf{x} \\
&= \int_{\mathbb{R}^n} f(\mathbf{x}) d\mu_n(\mathbf{x}).
\end{aligned}$$

Since this is true for all $f \in C_c^\infty(\mathbb{R}^n)$ then by density this same equality holds for all $f \in L^1(\mu_n)$. \blacksquare

Definition (White noise process). Let d be integer greater than or equal to 1 and let $T = \mathcal{S}(\mathbb{R}^d)$ and $\Omega := \mathcal{S}'(\mathbb{R}^d)$ and $\mathcal{F} = \mathcal{B}(\mathcal{S}'(\mathbb{R}^d))$. Let \mathbb{P} be the probability measure on (Ω, \mathcal{F}) obtained from the Bochner-Minlos theorem. The Gaussian process $W : T \times \Omega \rightarrow \mathbb{R}$ defined as

$$W(t, \omega) := W(\varphi, \xi) := \langle \xi, \varphi \rangle, \quad \text{for } (t, \omega) = (\varphi, \xi) \in T \times \Omega = \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}'(\mathbb{R}^d).$$

is called a white noise process. //

Definition (Smoothed white noise process). Let $\varphi \in L^2(\mathbb{R}^d)$ and for $\mathbf{x} \in \mathbb{R}$ let $\varphi_{\mathbf{x}}(y) := \varphi(\mathbf{y} - \mathbf{x})$ for $y \in \mathbb{R}^d$. We define the smoothed white noise process $W_\varphi : \mathbb{R}^d \times \mathcal{S}'(\mathbb{R}^d) \rightarrow \mathbb{R}$ as

$$W_\varphi(\mathbf{x}, \omega) := W(\varphi_{\mathbf{x}}, \omega) = \langle \omega, \varphi_{\mathbf{x}} \rangle,$$

where W is the white noise process introduced in the above definition. //

Proposition 4.16. *Let W_φ be smoothed white noise. We have the following properties.*

- (i) *For $\mathbf{x} \in \mathbb{R}^d$ the random variable $W_\varphi(\mathbf{x})$ is normally distributed with mean 0 and variance $\|\varphi\|_{L^2(\mathbb{R}^d)}^2$.*
- (ii) *If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ are chosen so that $\text{supp } \varphi_{\mathbf{x}} \cap \text{supp } \varphi_{\mathbf{y}} = \emptyset$ then $W_\varphi(\mathbf{x})$ and $W_\varphi(\mathbf{y})$ are independent.*
- (iii) *For any $h \in \mathbb{R}^d$ and all $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$ we have that*

$$(W_\varphi(\mathbf{x}_1 + h), \dots, W_\varphi(\mathbf{x}_n + h)) \stackrel{d}{=} (W_\varphi(\mathbf{x}_1), \dots, W_\varphi(\mathbf{x}_n))$$

Theorem 4.17 (Bochner-Minlos with tempered measure on \mathbb{R}). *Let σ be a tempered measure on \mathbb{R} . Let $\mathcal{S}' := \mathcal{S}'(\mathbb{R})$ be as in the above definition. Then there is a probability measure $\mathbb{P}^{(\sigma)}$ on \mathcal{S}' and a real valued Gaussian process $\{W_\varphi^{(\sigma)}\}_{\varphi \in \mathcal{S}}$ such that for all $\varphi \in \mathcal{S}$ we have*

⁷We have use the equality

$$\int_{-\infty}^{\infty} e^{-bx^2 + iax} dx = \sqrt{\frac{\pi}{b}} e^{-a^2/4b}.$$

- (i) $W_{\varphi}^{(\sigma)}(\xi) = \langle \xi, \varphi \rangle$ for all $\xi \in \mathcal{S}'$.
- (ii) $\mathbb{E}[W_{\varphi}^{(\sigma)}] = 0$.
- (iii) $\mathbb{E}[\exp(iW_{\varphi}^{(\sigma)})] = \exp\left(-\frac{1}{2} \int_{\mathbb{R}} |\hat{\varphi}(u)|^2 d\sigma(u)\right)$.

In the above $\hat{\varphi}$ is the Fourier transform of φ with respect to the Lebesgue measure.

Corollary 4.17.1.

4.6 Wiener Process

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $T = \mathbb{R}^+$. A Wiener process with starting point $\mathbf{x} \in \mathbb{R}^n$ is a Gaussian stochastic process $X : T \times \Omega \rightarrow \mathbb{R}^n$ such that

- (i) $X_0 = \mathbf{x}$ and $\mathbb{E}[X_t] = \mathbf{x}$ for all $t \in T$.
- (ii) $\text{Cov}(X_s - \mathbf{x}, X_t - \mathbf{x}) = \mathbb{E}[(X_s - \mathbf{x}) \cdot (X_t - \mathbf{x})] = n \min(s, t)$ for all $s, t \in T$.
- (iii) $X(\cdot, \omega)$ is continuous for almost all $\omega \in \Omega$.

This process is also referred to as Brownian motion. The pushforward measure $X_*\mathbb{P}$ on $(\mathbb{R}^n)^T$ is sometimes denoted as $\mathbb{P}^{\mathbf{x}}$ to emphasize the starting point of the process (i.e $X_0 = \mathbf{x}$). //

Brownian motion as a probability law.

Proposition 4.18. *There is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Gaussian process $\{W_t\}$ on that space satisfying (i) and (ii).*

Proof. A Gaussian process $\{W_t\}$ having the properties (i) and (ii) of the above definition has finite dimensional distributions measures

$$\nu_{t_1, \dots, t_k}(F_1 \times \dots \times F_k) = \int_{F_1 \times \dots \times F_k} \prod_{j=1}^k (2\pi \Delta t_j)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \frac{|\Delta \mathbf{x}_j|^2}{\Delta t_j}\right) d\mathbf{x}_1 \dots d\mathbf{x}_k,$$

with $x_0 = \mathbf{x}$, $t_0 = 0$, $\Delta \mathbf{x}_j = \mathbf{x}_j - \mathbf{x}_{j-1}$, $\Delta t_k = t_k - t_{k-1}$ and $F_1, \dots, F_k \in \mathcal{B}(\mathbb{R}^n)$. Furthermore, these measures satisfy the consistency conditions for the Kolmogorov extension theorem. ■

Proposition 4.19. *For any $0 \leq t_1 < \dots < t_n \leq T$ we have that*

$$B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}},$$

are independent normal random variables with $\mathbb{E}[B_{t_{i+1}} - B_{t_i}] = 0$ and $\text{Var}[B_{t_{i+1}} - B_{t_i}] = t_{i+1} - t_i$.

Proof. ■

Proposition 4.20. *There is a modification $\{B_t\}$ of the stochastic process $\{W_t\}$ obtained in Proposition 4.18. that has almost surely continuous paths.*

Proof. We will prove that for all $s, t \in T$ we have

$$\mathbb{E}\left[|W(t) - W(s)|^4\right] = n(n+2)|t-s|^2,$$

and then the result follows from the Kolmogorov continuity theorem. ■

Proposition 4.21 (Hitting time). *Let $m \in \mathbb{R}$ and let τ_m be the hitting time of the one dimensional Brownian motion $\{B_t\}_{t \in T}$ ie*

$$\tau_m(\omega) = \inf\{t \in T : B_t(\omega) = m\}.$$

Then τ_m satisfies the reflection equality which says that for all $w \in \mathbb{R}$ we have

$$\mathbb{P}\{\tau_m \leq t, B_t \leq w\} = \mathbb{P}\{B_t \geq 2m - w\}.$$

This implies that the probability density function f_{τ_m} of τ_m is

$$f_{\tau_m}(t) = |m|(2\pi)^{-1/2} t^{-3/2} \exp(-m^2/2t).$$

We have all the following properties.

1. $\{B_t\}$ is a Gaussian process i.e $(B_{t_1}, \dots, B_{t_k})$ is multi-normal on \mathbb{R}^{nk} .
2. There is a continuous version of $\{B_t\}$.
3. Each component of $B_t = (B_t^{(1)}, \dots, B_t^{(n)})$ is standard Brownian motion on \mathbb{R} .
4. $\langle B, B \rangle_t = \lim_{n \rightarrow \infty} \sum_{t_i \in \Pi_n} |B_{t_i} - B_{t_{i-1}}|^2 = t$ almost surely. +
5. For any $t_0 \geq 0$, $\{B_{t_0+t} - B_{t_0}\}$ is Brownian motion.
6. If $UU^T = I$, then $\{UB_t\}$ is a Brownian motion.
7. For $c \in \mathbb{R}$, $\{c^{-1}B_{c^2t}\}$ is also a Brownian motion.
8. $E[\exp(\lambda(B_s - B_t))] = \exp(\lambda^2(s - t)/2)$.
9. If $\{B_t\}$ is standard one dimensional Brownian motion then $\int_0^t B_s dB_s = \frac{1}{2}B_t^2 - t$.
10. For all $m > 0$ and $w \leq m$ we have $\mathbb{P}\{\tau_m \leq t, B_t \leq w\} = \mathbb{P}\{B_t \geq 2m - w\}$.⁹
11. We have $f_{\tau_m}(t) = |m|(2\pi)^{-1/2}t^{-3/2} \exp(-m^2/2t)$.
12. We have the joint density $f_{M_t, B_t}(m, w) = 2(2m - w)(2\pi)^{-1/2}t^{-3/2} \exp(-(2m - w)^2/2t)$.

Brownian motion as a limit of random walks.

Brownian motion as a special case of white noise.

Proposition 4.22. *Let $\varphi \in L^2(\mathbb{R}^d)$ and suppose $\{\varphi_n\}$ is sequence in $\mathcal{S}(\mathbb{R}^d)$ that converges to φ in $L^2(\mathbb{R}^d)$. For each n , define the function $f_n : \mathcal{S}' \rightarrow \mathbb{R}$ as $f_n(\xi) := \langle \xi, \varphi_n \rangle$. Then $\{f_n\}$ has a limit $f \in L^2(\mathcal{S}', \mathbb{P})$ and this limit is independent of choice of the sequence $\{\varphi_n\}$.*

⁹Hitting time $\tau_m(\omega) = \inf\{t \geq 0 : B_t(\omega) = m\}$.

Representing solutions to elliptic PDEs using Brownian motion

There is an inherent connection between elliptic partial differential equations and Brownian motion.

Lemma 4.23. *Suppose that $f \in C_c^2(\mathbb{R})$. Then*

$$\lim_{t \rightarrow 0} \frac{\mathbb{E}^x[f(B_t)] - f(x)}{t} = \frac{1}{2}f''(x).$$

Proof. We have that

$$\begin{aligned} \frac{\mathbb{E}^x[f(B_t)] - f(x)}{t} &= \frac{1}{t} \left(\int_{\mathbb{R}} f(y) \frac{1}{\sqrt{2\pi t}} e^{-|x-y|^2/2t} dy - f(x) \right) \\ &= \frac{1}{t} \left(\int_{\mathbb{R}} f(x+y) \frac{1}{\sqrt{2\pi t}} e^{-y^2/2t} dy - f(x) \right) \quad (\text{variable change } y \mapsto y-x) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{f(x+y) - f(x)}{t\sqrt{t}} e^{-y^2/2t} dy \quad (\text{since } \int_{\mathbb{R}} (2\pi)^{1/2} e^{-y^2/2} dy = 1) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{f(x + \sqrt{t} \cdot y) - f(x)}{t} e^{-y^2} dy \quad (\text{variable change } y^2/2t \mapsto y-x) \end{aligned}$$

Now Taylor's theorem tells us that there is a $\xi \in [0, 1]$ such that

$$f(x + \sqrt{t}y) - f(x) = f'(x)\sqrt{t} \cdot y + \frac{1}{2}f''(x + \xi\sqrt{t} \cdot y)ty^2.$$

Therefore

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{\mathbb{E}^x[f(B_t)] - f(x)}{t} &= \lim_{t \rightarrow 0} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{1}{t} \left(f'(x)\sqrt{t} \cdot y + \frac{1}{2}f''(x + \xi\sqrt{t} \cdot y)ty^2 \right) e^{-y^2} dy \\ &= \lim_{t \rightarrow 0} \frac{1}{\sqrt{2\pi}} \left(t^{-1/2}f'(x) \int_{\mathbb{R}} ye^{-y^2} dy + \frac{1}{2} \int_{\mathbb{R}} f''(x + \xi\sqrt{t} \cdot y)y^2 e^{-y^2} dy \right) \\ &= \lim_{t \rightarrow 0} \frac{1}{2} \int_{\mathbb{R}} f''(x + \xi\sqrt{t} \cdot y) \frac{y^2 e^{-y^2}}{\sqrt{2\pi}} dy \\ &= \frac{1}{2} \int_{\mathbb{R}} \lim_{t \rightarrow 0} f''(x + \xi\sqrt{t} \cdot y) \frac{y^2 e^{-y^2}}{\sqrt{2\pi}} dy \\ &= \frac{1}{2} \int_{\mathbb{R}} \frac{y^2 e^{-y^2}}{\sqrt{2\pi}} f''(x) dy = \frac{1}{2} f''(x) \int_{\mathbb{R}} \frac{y^2 e^{-y^2}}{\sqrt{2\pi}} dy \\ &= \frac{1}{2} f''(x), \end{aligned}$$

as desired. ■

Lemma 4.24. *Suppose that $f \in C_c^2(\mathbb{R})$. The process*

$$M_t = f(B_t) - \int_0^t f''(B_s) ds,$$

is a Martingale w.r.t the natural filtration $\{\mathcal{F}_t\}$ of Brownian motion.

Proof. It suffices to show that for all $s, t \in T$ with $s < t$ we have

$$\mathbb{E} \left[f(B_t) - f(B_s) - \frac{1}{2} \int_s^t f''(B_u) du \middle| \mathcal{F}_s \right] = 0,$$

which is equivalent to showing that for all $x \in \mathbb{R}$ and $t \in T$ we have

$$\mathbb{E}^x \left[f(B_t) - f(x) - \frac{1}{2} \int_0^t f''(B_s) ds \right] = 0.$$

Now define for $x \in \mathbb{R}$ the mean function $m_x : T \rightarrow \mathbb{R}$ as

$$m_x(t) = \mathbb{E}^x[f(B_t)].$$

We have that by iterated conditioning that

$$m'_x(t)^+ := \lim_{h \rightarrow 0^+} \frac{\mathbb{E}^x[f(B_{t+h})] - \mathbb{E}^x[f(B_t)]}{h} = \lim_{h \rightarrow 0} \mathbb{E}^x \left[\mathbb{E}^x \left[\frac{f(B_{t+h}) - f(B_t)}{h} \middle| \mathcal{F}_t \right] \right].$$

First we notice that

$$\left| \mathbb{E}^x \left[\frac{f(B_{t+h}) - f(B_t)}{h} \middle| \mathcal{F}_t \right] \right| \leq \frac{1}{2} \|f''\|_\infty.$$

and then

$$\begin{aligned} \lim_{h \rightarrow 0} \mathbb{E}^x \left[\frac{f(B_{t+h}) - f(B_t)}{h} \middle| \mathcal{F}_t \right] &= \lim_{h \rightarrow 0} \mathbb{E}^x \left[\frac{f(B_{t+h}) - f(B_t)}{h} \middle| \sigma(B_t) \right] && \text{(Markov property)} \\ &= \lim_{h \rightarrow 0} \mathbb{E}^{B_t} \left[\frac{f(B_h) - f(B_0)}{h} \right] && \text{(strong Markov property)} \\ &= \frac{1}{2} f''(B_t). && \text{(by above lemma)} \end{aligned}$$

Therefore by using dominated convergence we obtain

$$m'_x(t)^+ = \mathbb{E}^x \left[\lim_{h \rightarrow 0^+} \mathbb{E}^x \left[\frac{f(B_{t+h}) - f(B_t)}{h} \middle| \mathcal{F}_t \right] \right] = \mathbb{E}^x \left[\frac{1}{2} f''(B_t) \right].$$

We can work in a similar fashion to obtain that

$$m'_x(t)^- := \lim_{h \rightarrow 0^-} \frac{\mathbb{E}^x[f(B_{t+h})] - \mathbb{E}^x[f(B_t)]}{h} = \mathbb{E}^x \left[\frac{1}{2} f''(B_t) \right].$$

Therefore $m'_x(t)$ is well defined for all t . Hence we can conclude that

$$\mathbb{E}^x \left[f(B_t) - f(x) - \frac{1}{2} \int_0^t f''(B_s) ds \right] = m_x(t) - m_x(0) - \int_0^t m'_x(t) dt = 0,$$

as desired. ■

Theorem 4.25. Consider the boundary value problem

$$\begin{cases} u''(x) = g(x), & \text{for all } x \in [a, b], \quad g \in C([a, b]), \\ u(a) = A, \quad u(b) = B, & \text{for } A, B \in \mathbb{R}. \end{cases}$$

Let $\{B_t\}_{t \in T}$ be Brownian motion with $B_0 = x \in [a, b]$ and let τ be the exit time random variable defined as

$$\tau(\omega) := \inf_{t \in T} \{t \in T : B_s(\omega) = a \text{ or } B_s(\omega) = b\}.$$

Then

$$u(x) = \mathbb{E}^x \left[A \cdot \mathbf{1}_{\{W_\tau = a\}} + B \cdot \mathbf{1}_{\{W_\tau = b\}} - \frac{1}{2} \int_0^\tau g(B_s) ds \right].$$

Proof. Suppose that u solves the above boundary value problem. Consider the stochastic process

$$M_t = u(B_t) - \frac{1}{2} \int_0^t u''(B_s) ds.$$

By the above lemma, M_t is a martingale. Therefore we can apply optional sampling (Theorem 4.11) to M_t to obtain that

$$\mathbb{E}^x[M_{t \wedge \tau}] = \mathbb{E}^x[M_0] = \mathbb{E}^x[u(B_0)] = \mathbb{E}[u(x)] = u(x), \quad \text{for all } t \in T,$$

In particular, for $t = \tau$ we obtain

$$u(x) = \mathbb{E}^x[M_\tau] = \mathbb{E}^x \left[u(B_\tau) - \int_0^\tau u''(B_s) ds \right].$$

Since $u(B_\tau) = A$ when $B_\tau = a$ and $u(B_\tau) = B$ when $B_\tau = b$, and $g(x) = u''(x)$ then the result follows. ■

Theorem 4.26. Consider the boundary value problem

$$\begin{cases} -\Delta u = g, & \text{in } \Omega \subset \mathbb{R}^2, \\ u(x) = f(x), & \text{on } \partial\Omega. \end{cases}$$

Let $\{B_t\} = \{(B_t^{(1)}, B_t^{(2)})\}_{t \in T}$ be two dimensional Brownian motion with $B_0 = \mathbf{x} \in \Omega$ and let τ be the exit time random variable defined as

$$\tau(\omega) := \inf_{t \in T} \{t \in T : B_s(\omega) \in \partial\Omega\}.$$

Then

$$u(x) = \mathbb{E}^{\mathbf{x}} \left[f(B_\tau) + \frac{1}{2} \int_0^\tau g(B_s) ds \right].$$

4.7 Lévy and Jump Processes

Definition. A stochastic process X is called a Lévy process if

- (i) $X(0) = 0$ almost surely.
- (ii) X has independent increments.
- (iii) X has stationary increments.
- (iv) X is stochastically continuous in the sense that $\lim_{s \rightarrow t} \mathbb{P}(|X(t) - X(s)| > \epsilon) = 0$.

//

Theorem 4.27 (Lévy-Khintchine).

Theorem 4.28. *Any Lévy process X then admits a cadlag modification.*

Proof.

■

Theorem 4.29 (Lévy-Itô decomposition). *Let X be a Lévy process. Then X admits a decomposition*

$$X(t) = \gamma t + \sigma B(t) + X^P(t) + X^M(t),$$

where

- (i) $B(t)$ is standard Brownian motion.
- (ii) $X^P(t)$ is a compound Poisson process.
- (iii) $X^M(t)$ square integrable pure jump process.

Proof.

■

5 Itô Calculus and Elementary Stochastic Differential Equations

5.1 Integration with respect to Brownian motion

But in many applications one asks how much does a function/process of Brownian motion change if Brownian motion changes by ΔB_t . This change depends on the path of Brownian motion, we are looking for a suitable quantity

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

that somehow encodes information about the change of f with respect to B_t . Since Brownian motion paths are of unbounded variation, defining the integral of a function with respect to Brownian motion is out of the question for most processes f , unless f is a piecewise constant; in this case we are only interested in change of B_t at the discrete jumps of f .

Notice that I is a stochastic process, so $I(t)$ is a random variable, and this provides a work around the limitation of unbounded variation: we can define $I(t)$ as an element in $L^2(\mathbb{P})$ for each t (instead of path by path) and hope to establish the desired properties of regular integration.

Definition (Simple/Elementary processes). A process $X = X(\omega, t)$ is called simple for each $\omega \in \Omega$, there is a sequence of positive numbers $\{t_n\}_{n \geq 0}$ increasing to infinity with $t_0(\omega) = 0$ and a sequence $\{c_n(\omega)\}_{n \geq 0}$ of real random variables such that

$$X(t, \omega) = \sum_{n=1}^{\infty} \mathbf{1}_{[t_{n-1}, t_n)}(t) \cdot c_{n-1}(\omega).$$

A simple process is called elementary if it is adapted to the natural filtration of Brownian motion. If X is elementary then c_0 becomes non-random, ie $c_0(\omega)$ is the same for all ω . //

Requiring that X be adapted to the natural filtration of Brownian motion provides both a practical advantage and a theoretical one. The former is that the value of $X(t)$ can be determined at time t by the information in \mathcal{F}_t , and the Markov property further implies that this value depends only on $\sigma(B_t)$. The latter is stated in Proposition 5.2.

Similarly to simple functions in an arbitrary measure space, defining the Itô integral is straightforward and can be done path by path as follows.

Definition. Let $\{X_t\}$ be an elementary process. We define the Itô integral of X as

$$\int_0^t X_s dB_s = c_n(B_t - B_{t_n}) + \sum_{k=1}^n c_{k-1}(B_{t_k} - B_{t_{k-1}})$$

where n is the (random) index for which $t \in [t_n, t_{n+1})$. //

Proposition 5.1 (Properties of Itô integral for elementary integrands).

Similarly to the Lebesgue integral, one would like to approximate general processes with simple ones and define the Itô integral as the limit of Itô integrals of simple processes. It turns out that if X satisfies some integration and measurability properties then X is indeed the limit of simple processes in an appropriate sense.

Definition. Let $I = [a, b] \subset \mathbb{R}^+$ be an interval and let $\{X_t\}_{t \in I}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that the following holds.

- (i) $\{X_t\}$ is progressively measurable with respect to the natural filtration of $\{B_t\}$.
- (ii) The random variable $Y(\omega) = \int_0^t X(\omega, s)^2 ds$ has finite expectation.

The space of all processes satisfying the above is denoted $\mathcal{V} = \mathcal{V}(I)$. //

\mathcal{V} will be our standard space for integration. It is clear that \mathcal{V} contains all elementary processes and in fact we have more.

Proposition 5.2. *Let $X \in \mathcal{V}(I)$. Then there is a sequence of elementary processes $\{\Delta_n\} \subset \mathcal{V}$ such that*

$$\lim_{n \rightarrow \infty} \int_{\Omega} \int_a^b (X(\omega, t) - \Delta_n(\omega, t))^2 dt d\omega = 0, \quad \text{ie} \quad \mathbb{E}[\|X - \Delta_n\|_{L^2(I)}] \rightarrow 0.$$

Proof. The proof is divided onto three steps.

Step 1: Suppose first that X is continuous and $|X(t, \omega)| \leq M$ for all t and all ω . Define the sequence of simple processes $\{X_n\}$ as

$$X_n(t, \omega) = \sum_{k=1}^n X(t_{k-1}, \omega) \cdot \mathbf{1}_{[t_{k-1}, t_k)}(t), \quad t_k = (1 - \frac{k}{2^n})a + \frac{k}{2^n}b, \quad k = 0, 1, 2, \dots$$

It is clear that X_n is elementary. Furthermore, $X_n(\cdot, \omega) \rightarrow X(\cdot, \omega)$ uniformly for each ω and this implies that

$$\lim_{n \rightarrow \infty} \int_a^b (X_n(t, \omega) - X(t, \omega))^2 dt = 0.$$

Also we have that

$$\int_a^b (X_n(t, \omega) - X(t, \omega))^2 dt \leq 4M^2(b - a), \quad \text{for all } \omega \in \Omega,$$

and therefore by dominated convergence we have that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int_a^b (X_n(t, \cdot) - X(t, \cdot))^2 dt \right] = 0.$$

Step 2: Suppose that X is bounded. For each n , let φ_n be a non-negative continuous function such that

$$\text{supp } \varphi_n = \left[-\frac{1}{n}, 0 \right] \quad \text{and} \quad \int_{\mathbb{R}} \varphi_n(x) dx = 1,$$

and define the convolution of the path $X(\cdot, \omega)$ with φ_n as

$$X_n(t, \omega) = (X(\cdot, \omega) * \varphi_n)(t) = \int_0^t \varphi_n(s - t) X(s, \omega) ds.$$

One can show that that $X_n \in \mathcal{V}$ for all $n \in \mathbb{N}$. Since $\{X_n\}$ is an approximate identity¹⁰ we have that for each ω

$$\lim_{n \rightarrow \infty} \int_a^b (X_n(t, \omega) - X(t, \omega))^2 dt = 0.$$

¹⁰

and therefore by dominated convergence

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int_a^b (X_n(t, \cdot) - X(t, \cdot))^2 dt \right] = 0.$$

Step 3: Now let X be any element in \mathcal{V} . Let

$$X_n(t, \omega) = \begin{cases} -n & \text{if } X(t, \omega) < -n, \\ X(t, \omega) & \text{if } |X(t, \omega)| \leq n, \\ n & \text{if } X(t, \omega) > n. \end{cases}$$

It is clear that $X_n \in \mathcal{V}$ for all n and that

$$\lim_{n \rightarrow \infty} \int_a^b (X_n(t, \omega) - X(t, \omega))^2 dt = 0.$$

Also we have for all $\omega \in \Omega$ that

$$\int_a^b (X_n(t, \omega) - X(t, \omega))^2 dt \leq 2 \int_a^b X_n(t, \omega)^2 dt + 2 \int_a^b X(t, \omega)^2 dt \leq 4 \int_a^b X(t, \omega)^2 dt.$$

and therefore by dominated convergence

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int_a^b (X_n(t, \cdot) - X(t, \cdot))^2 dt \right] = 0.$$

Now one can easily conclude the desired result by approximating X with bounded processes (Step 3), then approximate those bounded processes with continuous ones (Step 2), and finally approximate continuous processes using elementary processes (Step 1). \blacksquare

Using the above proposition, we can now define the Itô integral for processes in \mathcal{V} .

Definition. Let $X \in \mathcal{V} = \mathcal{V}(I)$ where $I = [a, b]$ and let Δ_n be a sequence of simple processes converging to X as in proposition 5.2. The Itô isometry for elementary processes says that

$$\mathbb{E} \left[\left(\int_a^b \Delta_n(t, \cdot) - \Delta_m(t, \cdot) dB_t \right)^2 \right] = \mathbb{E} \left[\int_a^b (\Delta_n(t, \cdot) - \Delta_m(t, \cdot))^2 dt \right],$$

and therefore $\{ \int_a^b \Delta_n dB_t \}$ is a Cauchy sequence in $L^2(\Omega)$. We define the Itô integral of $\{X_t\}$ as

$$\int_a^b X(\cdot, t) dB_t := \lim_{n \rightarrow \infty} \int_a^b \Delta_n(\cdot, t) dB_t,$$

where the limit is taken in $L^2(\mathbb{P})$. //

Lemma 5.3. *The definition of \mathcal{I} is independent of the choice of simple processes. More precisely, if $\{\Delta_n\}$ is any sequence of simple process satisfying the requirements of the above definition then*

$$\left\| \int_a^b \Delta_n(t) dB_t - \int_a^b X(t) dB_t \right\|_{L^2(\Omega)} \rightarrow 0.$$

Lemma 5.4. *The Itô integral $\mathcal{I}(\omega, t)$ of a simple process $\Delta(\omega, t)$ is a continuous Martingale.*

Proof. For all $\omega \in \Omega$ and $t \in I$ write

$$\Delta(t, \omega) = \sum_k e_k(\omega) \cdot \mathbf{1}_{[t_k, t_{k+1})}(t), \quad \mathcal{I}(t, \omega) = \sum_{t_k \leq t} e_k(\omega) (B(t_{k+1}, \omega) - B(t_k, \omega)).$$

Assume that $t \in (t_k, t_{k+1})$ for some k and let $h \in \mathbb{R}$ such that $t+h \in (t_k, t_{k+1})$

$$|\mathcal{I}(t+h, \omega) - \mathcal{I}(t, \omega)| = |e_k(\omega)| |B(t+h, \omega) - B(t, \omega)|,$$

and therefore $\mathcal{I}(\cdot, \omega)$ is continuous at t . On the other hand,

$$|\mathcal{I}(t_k+h, \omega) - \mathcal{I}(t_k, \omega)| = \begin{cases} |e_{k-1}| |B(t_k+h, \omega) - B(t_k, \omega)| & \text{if } h < 0, \\ |e_k| |B(t_k+h, \omega) - B(t_k, \omega)| & \text{if } h > 0. \end{cases}$$

Therefore, $\mathcal{I}(\cdot, \omega)$ is continuous at t_k .

Now \mathcal{I} is a Martingale. Indeed, let $t \in I$ and $h > 0$ and let ℓ be the index for which $t \in [t_\ell, t_{\ell+1})$ then we have

$$\begin{aligned} \mathbb{E} [\mathcal{I}(t+h) \mid \mathcal{F}_t] &= \mathbb{E} \left[\int_0^t \Delta(s) dB_s + \int_t^{t+h} \Delta(s) dB_s \mid \mathcal{F}_t \right] \\ &= \int_0^t \Delta(s) dB_s + \mathbb{E} \left[\sum_{t \leq t_k \leq t_{k+1} \leq t+h} e_k (B(t_{k+1}) - B(t_k)) \mid \mathcal{F}_t \right] \\ &= \int_0^t \Delta(s) dB_s + \mathbb{E} [e_\ell (B(t_{\ell+1}) - B(t)) \mid \mathcal{F}_t] \\ &\quad + \sum_{t_{\ell+1} \leq t_k \leq t+h} e_k \mathbb{E} [B(t_{k+1}) - B(t_k)] \\ &= \int_0^t \Delta(s) dB_s = \mathcal{I}(t), \end{aligned}$$

as desired. ■

Theorem 5.5 (Properties). *The Itô integral $\mathcal{I}(\omega, t)$ of a stochastic process $X(\omega, t)$ that is adapted to the natural filtration $\{\mathcal{F}_t\}_{t \in I}$ of Brownian motion satisfies the following.*

- (i) *For each $t \in I$, $\mathcal{I}(t, \cdot)$ is \mathcal{F}_t -measurable.*
- (ii) *\mathcal{I} satisfies the Itô isomerty.*
- (iii) *\mathcal{I} is a Martingale.*
- (iv) *Almost all paths $\mathcal{I}(\cdot, \omega)$ can be chosen to be continuous.*
- (v) *The quadratic variation of the Itô process is given by*

$$[\mathcal{I}, \mathcal{I}](t, \omega) = \int_0^t X(t, \omega) dB_t(\omega).$$

Proof. Part (i)-(iii) follow directly from the fact that $\mathcal{I}(t)$ is a pointwise limit of \mathcal{F}_t -measurable functions. For part (iv), let $\{\Delta_n\}$ be a sequence of simple processes such that

$$\lim_{n \rightarrow \infty} \|X - \Delta_n\|_{L^2(I), L^1(\Omega)} := \lim_{n \rightarrow \infty} \int_{\Omega} \int_I |\Delta_n(\omega, t) - X(\omega, t)|^2 dt d\omega = 0.$$

We will show that there is a subsequence $\{\Delta_{n_k}\}$ such that for almost all $\omega \in \Omega$

$$\|\mathcal{I}_{n_{k+1}}(\omega, \cdot) - \mathcal{I}_{n_k}(\omega, \cdot)\|_{L^\infty(I)} = \sup_{t \in [0, T]} \left| \int_0^t \Delta_{n_{k+1}}(s, \omega) dB_s - \int_0^t \Delta_{n_k}(s, \omega) dB_s \right| \rightarrow 0.$$

Hence for almost all $\omega \in \Omega$, the sequence of continuous functions $\{\mathcal{I}_{n_k}(\omega, \cdot)\}_{k \in \mathbb{N}}$ is Cauchy in $L^\infty(I)$ and therefore converges to a continuous element $\mathcal{J}(\omega, \cdot) \in L^\infty(I)$, which by Lemma 4.1 will be almost surely equal to $\mathcal{I}(\omega, \cdot)$.

By Doob's martingale inequality applied on $\mathcal{I}_n - \mathcal{I}_m$ with $p = 2$ we have that for any $\epsilon > 0$ that

$$\begin{aligned} \mathbb{P} \left[\|\mathcal{I}_n(\cdot, \omega) - \mathcal{I}_m(\cdot, \omega)\|_{L^\infty(I)} \geq \epsilon \right] &\leq \frac{1}{\epsilon^2} \|\mathcal{I}_n(T) - \mathcal{I}_m(T)\|_{L^2(\Omega)}^2 \\ &= \frac{1}{\epsilon^2} \mathbb{E} \left[\int_0^T (\Delta_n(t) - \Delta_m(t))^2 dt \right] \quad (\text{by It\^o isometry}) \\ &= \frac{1}{\epsilon^2} \|\Delta_n - \Delta_m\|_{L^2(I), L^1(\Omega)}^2 \xrightarrow{m, n \rightarrow \infty} 0. \end{aligned}$$

Therefore there is a subsequence $\{\Delta_{n_k}\}$ such that

$$\mathbb{P} \left[\|\mathcal{I}_{n_{k+1}} - \mathcal{I}_{n_k}\|_{L^\infty(I)} \geq 2^{-k} \right] < 2^{-k},$$

so that by the Borel-Cantelli lemma

$$\mathbb{P} \left\{ \omega \in \Omega : \|\mathcal{I}_{n_{k+1}}(\cdot, \omega) - \mathcal{I}_{n_k}(\cdot, \omega)\|_{L^\infty(I)} \geq 2^{-k} \text{ for infinitely many } k \right\} = 0.$$

Therefore for almost all $\omega \in \Omega$ there is an integer N_ω such that for all $k \geq N_\omega$ we have

$$\|\mathcal{I}_{n_{k+1}}(\omega, \cdot) - \mathcal{I}_{n_k}(\omega, \cdot)\|_{L^\infty(I)} < 2^{-k},$$

and thus $\{\mathcal{I}_{n_k}(\cdot, \omega)\}$ is Cauchy for almost all ω as desired. ■

5.2 Weakening defining conditions for It\^o integral

Definition. Let $I = [0, T]$ be an interval and $\mathcal{W} = \mathcal{W}(I)$ be the set of all stochastic processes $X(t, \omega)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying the following conditions.

- (i) $X : I \times \Omega \rightarrow \mathbb{R}$ is measurable.
- (ii) There is filtration $\{\mathcal{H}_t\}$ such that $B(t, \omega)$ is Martingale with respect to this filtration and $X(t, \cdot)$ is \mathcal{H}_t -adapted.
- (iii) For almost all $\omega \in \Omega$, $X(\cdot, \omega) \in L^2(I)$.

For such functions one can show that there is a sequence of simple processes $\{\Delta_n\} \subset \mathcal{W}$ that converge to X in probability for each $t \in [0, T]$. By defining the integral of simple processes in the usual way, defines

$$\int_0^T X(t)dB_t := \lim_{n \rightarrow \infty} \int_0^T \Delta_n(t)dB_t \quad \text{in probability.}$$

//

Theorem 5.6. *The Itô integral of functions in $\mathcal{W}(I)$ has the same properties of the integral for functions $\mathcal{V}(I)$, except that it's not a Martingale but rather a local Martingale.*

5.3 Itô process and Itô-Doeblin Formula

Definition (Itô Process). Let $\{\mathcal{F}_t\}$ be the natural filtration for Brownian motion. Suppose there are adapted processes $\{\alpha(t, \cdot)\}$ and $\{\sigma(t, \cdot)\}$ and a stochastic process $\{X(t, \cdot)\}$ such that

$$X(t) = X(0) + \int_0^t \alpha(s, \omega)ds + \int_0^t \sigma(s, \omega)dB_s.$$

Then X is called an Itô process.

//

Proposition 5.7 (Quadratic variation of Itô process). *If X is an Itô process then*

$$[X, X](t, \omega) = \int_0^t \sigma(s, \omega)dB_s(\omega).$$

Theorem 5.8 (Itô-Doeblin Formula). *Let $\{X(t, \cdot)\}$ be an Itô process and $g \in C^2([0, \infty) \times \mathbb{R})$. Define the stochastic process $Y(t, \omega) := g(t, X(t, \omega))$. Then $\{Y(t, \cdot)\}$ is an Itô process and for all $T \geq 0$*

$$Y(t) = Y(0) + \int_0^t \frac{\partial g}{\partial t}(u, X(u))du + \int_0^t f_x(u, X(u))dX(u) + \frac{1}{2} \int_0^t \frac{\partial^2 g}{\partial x^2}(u, X(u))d[X, X](u).$$

5.4 SDE's and the Markov property

Definition. Let $I \subset \mathbb{R}^+$ be any closed interval (possibly unbounded). Let $\mu, \sigma : I \times \mathbb{R} \rightarrow \mathbb{R}$ be Borel functions and $\{W_t\}$ be a Wiener process. A first order linear stochastic differential equation is a relation of the form

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \quad t \in I, \quad (7)$$

In other words if $t_0 = \min I$ then

$$X(t) = X(t_0) + \int_{t_0}^t \mu(u, X(u))du + \int_{t_0}^t \sigma(u, X(u))dW(u), \quad t \in I. \quad (8)$$

The function μ is called the *drift* and the function $\sigma^2/2$ is called the *diffusion coefficient*.

//

Definition (Strong solution). Let $\{W_t\}_{t \geq 0}$ be a Brownian motion with admissible filtration $\{\mathcal{F}_t\}_{t \geq 0}$ on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let ξ^* be a random variable. A progressively measurable process $\{X_t\}$ is a strong solution with initial condition $X(0) = \xi^*$ if (7) holds almost surely.

//

Definition (Weak solution). A stochastic process (X_t, \mathcal{F}_t) on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a weak solution with initial distribution μ if there exists a Brownian motion $\{B_t\}_{t \geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $(\mathcal{F}_t)_{t \geq 0}$ is an admissible filtration and $\mathbb{P}(X_0 \in \cdot) = \mu(\cdot)$ and (7) holds almost surely for all $t \geq 0$. //

Proposition 5.9 (Uniqueness of strong solutions). *Let $I = [0, T]$. Suppose that the functions μ and σ are Lipschitz. If X and Y are two strong solutions to (7) then X and Y are indistinguishable.*

Proof. Since both X and Y satisfy (7) and $X(0) = Y(0)$ then for all $t \in I$ we have that

$$X(t) - Y(t) = \int_0^t \mu(u, X(u)) - \mu(u, Y(u)) du + \int_0^t \sigma(u, X(u)) - \sigma(u, Y(u)) dW(u).$$

Squaring both sides we get

$$(X(t) - Y(t))^2 \leq 2 \left(\int_0^t \mu(u, X(u)) - \mu(u, Y(u)) du \right)^2 + 2 \left(\int_0^t \sigma(u, X(u)) - \sigma(u, Y(u)) dW(u) \right)^2.$$

Taking expectations we obtain

$$\begin{aligned} \mathbb{E} \left[(X(t) - Y(t))^2 \right] &\leq 2 \mathbb{E} \left[\left(\int_0^t \mu(u, X(u)) - \mu(u, Y(u)) du \right)^2 \right] \\ &\quad + 2 \mathbb{E} \left[\left(\int_0^t \sigma(u, X(u)) - \sigma(u, Y(u)) dW(u) \right)^2 \right]. \end{aligned}$$

On one hand, we have by the Cauchy-Schwarz inequality that

$$\left(\int_0^t \mu(u, X(u)) - \mu(u, Y(u)) du \right)^2 \leq T \int_0^t (\mu(u, X(u)) - \mu(u, Y(u)))^2 du,$$

and therefore by taking expectations and using the Lipschitz continuity of μ we get

$$\begin{aligned} \mathbb{E} \left[\left(\int_0^t \mu(u, X(u)) - \mu(u, Y(u)) du \right)^2 \right] &\leq T \int_0^t \mathbb{E} \left[(\mu(u, X(u)) - \mu(u, Y(u)))^2 \right] du \\ &\leq TK \int_0^t \mathbb{E} \left[(X(u) - Y(u))^2 \right] du. \end{aligned}$$

On the other hand, by the Itô isometry we have that

$$\left(\int_0^t \sigma(u, X(u)) - \sigma(u, Y(u)) dW(u) \right)^2 = \int_0^t (\sigma(u, X(u)) - \sigma(u, Y(u)))^2 du,$$

and therefore by the taking expectations and using the Lipschitz continuity of σ we get

$$\begin{aligned} \mathbb{E} \left[\left(\int_0^t \sigma(u, X(u)) - \sigma(u, Y(u)) dW(u) \right)^2 \right] &= \int_0^t \mathbb{E} \left[(\sigma(u, X(u)) - \sigma(u, Y(u)))^2 \right] du \\ &\leq M \int_0^t \mathbb{E} \left[(X(u) - Y(u))^2 \right] du. \end{aligned}$$

All of the above imply that for all $t \in [0, T]$ we have

$$\mathbb{E} \left[(X(t) - Y(t))^2 \right] \leq 2(TK + M) \int_0^t \mathbb{E} \left[(X(u) - Y(u))^2 \right] du.$$

Therefore, by Gronwall's inequality we have

$$\mathbb{E} \left[(X(t) - Y(t))^2 \right] = 0, \quad \text{for all } t \in [0, T].$$

This means $X(t) = Y(t)$ almost surely. Now let

$$F = \{ \omega \in \Omega : X_r(\omega) = Y_r(\omega), \forall r \in \mathbb{Q} \cap [0, T] \}.$$

For each $r \in \mathbb{Q} \cap [0, T]$ we have that the event

$$E_r = \{ \omega \in \Omega : X_r(\omega) \neq Y_r(\omega) \},$$

has probability 0. This means that

$$\mathbb{P}(F) = \mathbb{P} \left(\Omega \setminus \bigcup_{r \in \mathbb{Q} \cap [0, T]} E_r \right) = 1.$$

This means that almost surely $X(r) - Y(r) = 0$ for all $r \in \mathbb{Q} \cap [0, T]$. Since for almost all $\omega \in \Omega$ we have $X(\cdot, \omega)$ and $Y(\cdot, \omega)$ are continuous then by density we have that almost surely $X(t) - Y(t) = 0$ and therefore X and Y are indistinguishable. \blacksquare

Theorem 5.10 (Uniqueness of weak solutions). *Suppose that μ and σ are Lipschitz functions and that X and Y be two weak solutions to (7) such that X_0 and Y_0 induce the same probability law $\mu = \mathbb{P}^{X_0} = \mathbb{P}^{Y_0}$. Then X and Y have the same finite dimensional distributions.*

Theorem 5.11 (Existence of strong solutions). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $I = [0, T]$. Suppose we have the following*

(i) *Two functions $\mu, \sigma : I \times \mathbb{R} \rightarrow \mathbb{R}$ and constants $C, D \in \mathbb{R}$ such that for all $x, y \in \mathbb{R}$ and $t \in [0, T]$ we have*

$$(i)a. \quad |\mu(t, x)| + |\sigma(t, x)| \leq C(1 + |x|),$$

$$(i)b. \quad |\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D|x - y|.$$

(ii) *$\{B_t\}$ is a Brownian motion with natural filtration $\{\mathcal{F}_t\}$.*

(iii) *$\xi \in L^2(\Omega)$ and is independent of $\mathcal{F}_\infty = \bigcup \mathcal{F}_t$.*

(iv) *$\{\mathcal{F}_t^\xi\}$ is the filtration generated by ξ and $\{\mathcal{F}_t\}$.¹¹*

Under these assumption, equation (7) has a solution a strong solution $\{X_t\}$ that is adapted to the filtration $\{\mathcal{F}_t^\xi\}$ and $\|X^2\|_{L^1(I)} \in L^1(\Omega)$ ¹².

¹¹This filtration is defined as follows: $\mathcal{F}_0^\xi = \sigma(\xi)$ and

$$\mathcal{F}_t^\xi = \sigma \left(\bigcup_{s \in [0, t]} \mathcal{F}_s \right), \quad t \in [0, T].$$

¹²This means that

$$\int_\Omega \int_0^T |X(t, \omega)|^2 dt d\omega < \infty$$

Definition (Markov property). Let $I = [t, T]$ and $x \in \mathbb{R}$ be given. Suppose X is a stochastic process that solves (7) with initial condition $X(t) = x$. Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a Borel function. We define

$$g(x, t) = \mathbb{E} [h(X(T)) \mid X(t) = x] .$$

//

5.5 Feynman-Kac and Fokker-Planck Equations

Theorem 5.12 (Fokker-Planck). Let $\{X_t\}$ be a real valued stochastic process satisfying the following stochastic differential equation

$$dX(u) = \mu(u, X(u)) + \sigma(u, X(u))dW(u),$$

where W is any Weiner process. If $p(x, t)$ is the p.d.f of $X(t)$ then

$$\frac{\partial}{\partial t} p(x, t) = -\frac{\partial}{\partial x} [\mu(x, t)p(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [\sigma^2(x, t)p(x, t)].$$

Theorem 5.13 (Feynman-Kac). Let X be a real valued stochastic process satisfying the following stochastic differential equation

$$dX(u) = \beta(u, X(u)) + \gamma(u, X(u))dW(u).$$

Let h be a real Borel function. Fix $T > 0$ and let $t \in [0, T]$. Let

$$g(x, t) = \mathbb{E}^{t, x} [h(X(T))] = \mathbb{E} [h(X(T)) \mid X(t) = x] .$$

Then the function g satisfies the following

$$\begin{cases} \frac{\partial g}{\partial t} = -\beta \frac{\partial g}{\partial x} - \frac{1}{2} \gamma^2 \frac{\partial^2 g}{\partial x^2}, & \text{for all } (x, t) \in \mathbb{R} \times [0, T], \\ g(x, T) = h(x), & \text{for all } x \in \mathbb{R}. \end{cases} \quad (9)$$

5.6 Examples from finance and economics

6 General Stochastic Integration

6.1 Generalized Itô Integral

6.2 Itô-Doeblin formula for jump processes

6.3 *Functional Itô calculus and stochastic integral representation of martingales

Part II

Functional Analysis and Partial Differential Equations

7 Classical theory and fundamental equations

7.1 Fundamental existence theorems

7.2 Poisson equation

7.3 Diffusion equation

7.4 Wave equation

8 Hilbert Spaces

8.1 Elementary properties

8.2 Lax-Milgram

8.3 Reproducing kernel Hilbert spaces

9 Sobolev Spaces

9.1 Defintion, characterization, completeness and duality of $W^{m,p}(\Omega)$

9.2 Sobolev embeddings

9.3 The trace operator and fractional Sobolev spaces

9.4 Weak formulation of boundary value problems

9.5 *Weighted Sobolev Spaces and Non-Linear Potential Theory

9.6 *Sobolev spaces on manifolds

Part III

Special Topics

10 Stochastic Heat Equation

11 Stochastic Wave Equation

12 The Classical and Stochastic Hasegawa-Mima equation

13 Stochastic Integration in UMD spaces

14 Miscellaneous Remarks and Observations

Understanding $dB_t \cdot dB_t = dt$ and $dB_t dt = 0$. Not to be understood in the sense of

$$(B_{t_{j+1}} - B_{t_j})^2 \simeq t_{j+1} - t_j.$$

First of all, we have

$$\begin{aligned} \left| \sum_{j=1}^{n-1} (B_{t_{j+1}} - B_{t_j})(t_{j+1} - t_j) \right| &\leq n \cdot \sup_{0 \leq j \leq n-1} |B_{t_{j+1}} - B_{t_j}| |t_{j+1} - t_j| \\ &\leq T \cdot \sup_{0 \leq j \leq n-1} |B_{t_{j+1}} - B_{t_j}|, \end{aligned}$$

which goes to zero since B_t is continuous. Second,

$$[B, B](t) = \lim_{|\Pi| \rightarrow 0} \sum_{j=0}^{n-1} (B_{t_{j+1}} - B_{t_j})^2 = t \quad (\text{a.s.}),$$

which follows from the inequality

$$\left\| t - \sum_{j=0}^{n-1} (B_{t_{j+1}} - B_{t_j})^2 \right\|_{L^2(\Omega)} \leq 2|\Pi|t.$$

If $\Pi_n = \{0, t/n, 2t/n, \dots, t\}$ and we define

$$Z_{j+1} = \frac{B_{t_{j+1}} - B_{t_j}}{\sqrt{t_{j+1} - t_j}} = \sqrt{\frac{n}{t}} (B_{t_{j+1}} - B_{t_j}),$$

then it can be shown using the law of large numbers on the independent random variables $\{Z_{j+1}^2\}$ with common mean $\bar{\mu} = 1$ that

$$\frac{1}{t} \sum_{j=0}^{n-1} (B_{t_{j+1}} - B_{t_j})^2 = \sum_{j=0}^{n-1} \frac{Z_{j+1}^2}{n} \longrightarrow \bar{\mu} = 1, \quad (\text{a.s.}).$$

Remark on definition of Itô integral.

References

- [1] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, ser. Hochschultext / Universitext. Springer, 2003. [Online]. Available: <https://books.google.com.lb/books?id=kXw9hB4EEpUC>.
- [2] S. Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*, ser. Springer Finance Textbooks v. 11. Springer, 2004. [Online]. Available: <https://books.google.com.lb/books?id=08kD1NwQBsqC>.
- [3] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, ser. Universitext. Springer New York, 2010. [Online]. Available: <https://books.google.com.lb/books?id=GAA2Xq0IIGoC>.
- [4] T. Hytonen, J. van Neerven, M. Veraar, and L. Weis, *Analysis in Banach Spaces, Volume I: Martingales and Littlewood-Paley Theory*. Dec. 2016, ISBN: 978-3-319-48519-5 (print), 978-3-319-48520-1 (online). DOI: 10.1007/978-3-319-48520-1.
- [5] H. Holden, B. Øksendal, and J. Ubøe, *Stochastic Partial Differential Equations*. Birkhauser Boston, 2014. [Online]. Available: <https://books.google.com.lb/books?id=IBsgswEACAAJ>.
- [6] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, ser. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2004. [Online]. Available: <https://books.google.com.lb/books?id=1m195FLM5koC>.
- [7] I. Karatzas, I. Shreve, S. Shreve, and S. Shreve, *Brownian Motion and Stochastic Calculus*, ser. Graduate Texts in Mathematics (113) (Book 113). Springer New York, 1991. [Online]. Available: https://books.google.com.lb/books?id=ATNy%5C_Zg3PSsC.
- [8] R. Bartle and K. M. R. Collection, *The Elements of Integration*, ser. Wiley Classics Library. Wiley, 1966. [Online]. Available: <https://books.google.com.lb/books?id=UrLvAAAAMAAJ>.
- [9] J. Walsh, *Knowing the Odds: An Introduction to Probability*, ser. Graduate studies in mathematics. American Mathematical Society, 2012. [Online]. Available: <https://books.google.com.lb/books?id=4uC0uEXpvyoC>.
- [10] R. Dalang, D. Khoshnevisan, F. Rassoul-Agha, C. Mueller, D. Nualart, and Y. Xiao, *A Mini-course on Stochastic Partial Differential Equations*, ser. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2008, ISBN: 9783540859932. [Online]. Available: <https://books.google.com.lb/books?id=EBeVxgEACAAJ>.
- [11] D. Alpay, P. Jorgensen, and D. Levanony, *On the equivalence of probability spaces*, 2016. arXiv: 1601.00639 [math.PR].
- [12] R. Dudley, B. Bollobas, and W. Fulton, *Real Analysis and Probability*, ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002. [Online]. Available: <https://books.google.com.lb/books?id=7UuT7UZViNOc>.