Personal Notes on Stochastic Processes

Adel Saleh

March 9, 2021

Abstract

The following document is a comprehensive treatment of the fundamentals of stochastic processes and stochastic analysis on one hand, and functional analysis and partial differential equations on the other. It is compiled from several sources, spanning from personal favorite textbooks 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 MathOveflow 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

Part I

Fundamental Results in Stochastic Analysis

Chapter 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 developement 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 it's surprisingly intuitive features is it's 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 many excellent textbooks such as [1, 2, 3, 4] , this large section seeks to explore measures from the ground up, aiming for maximum generality whenever it's possible.

Bibliography

[1] G.B. Folland. *Real Analysis: Modern Techniques and Their Applications.* A Wiley-Interscience publication. Wiley, 1984.

[2] W. Rudin, W.A. RUDIN, and Tata McGraw-Hill Publishing Company. *Real and Complex Analysis.* Higher Mathematics Series. McGraw-Hill Education, 1987.

[3] R.G. Bartle and Karreman Mathematics Research Collection. *The Elements of Integration.* Wiley Classics Library. Wiley, 1966.

[4] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* Universitext. Springer New York, 2010.

[5] L.A. Steen and J.A. Seebach. *Counterexamples in Topology.* Dover books on mathematics. Dover Publications, 1995.

[6] R.B. Ash, C.D.D. Robert B. Ash, R. B, C.A. Doleans-Dade, Academic Press (Londyn)., and C. A. *Probability and Measure Theory.* Elsevier Science, 2000.

1.1 General Lebesgue Integral and the Space $L^1(\mu)$

Let X be any set. Measure theory, in it's most crude form, deals with the problem assigning a label (usually a real number) to some subsets of E of X, in a meaningful way. This label is called the *measure* of E and is denoted $\mu(E)$.' As the term might imply, a measure provides a way to "measure" a property of some (or all) subsets of X, and it has to at least satisfy the following intuitive requirements:

(i) $E \cap F = \emptyset$ implies $\mu(E \cup F) = \mu(E) + \mu(F)$. In particular, $\mu(E) + \mu(X \setminus E) = \mu(X)$.

(ii) $E \subset F$ implies $\mu(E) \leq \mu(F)$.

We say at least because an actual measure satisfies more (see definition 1.1.6). Here a few concrete examples of what might be a measure.

Example 1. Perhaps the simplest of all measures is the one that counts the number of elements in E, ie $\mu(E) = |E|$ with the understanding that $\mu(E) = \infty$ if E is infinite. In this case, all subsets of X are said to be measurable because they have a well defined measure, and thus we can regard μ as a mapping from 2^X to $\mathbb{N} \cup \{\infty\}$.

Example 2. Let Ω be the set of all possible outcomes of tossing a coin n times and $\mathbb{P} : \Omega \to [0, 1]$ be the function that assigns to each outcome it's probability. Then for an event $E \subset \Omega$ one can extend $\mathbb{P}$ to a measure by defining

$$\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\omega).$$

The new function $\mathbb{P} : 2^X \to [0, 1]$ is called a probability measure, for obvious reasons.

Example 3. If X the set of all atoms is an infinite metallic sheet and we cut a piece E of X, we can define a measure μ that assigns to the smaller sheet E it's weight. Then μ is a real valued function that tell us about the weigh distribution in the sheet.

Example 4. In fact, if X is countable and $f : X \to \mathbb{R}^+$ is any function, then one can easily obtain a way to measure a subset E of X by defining

$$\mu(E) = \begin{cases} 0 & \text{if } E = \emptyset, \\ \sum_{x \in E} f(x) & \text{otherwise}, \end{cases} \tag{1.1}$$

with the agreement that $\mu(E) = \infty$ if the sum diverges. This determines the measure of countable sets from the measure of singletons. It has the additional property:

$$\{E_n\}_{n=1}^{\infty} \subset 2^X \text{ s.t } \forall i, j \in \mathbb{N},\, E_i \cap E_j = \emptyset \implies \mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n). \tag{1.2}$$

This is called *countable additivity*, and is usually used as one of the defining properties of measures.

Things start becoming interesting when X is an uncoutable set, as technical limiations, or rather inconveniences start to arise. First off, let's see what happens if we change (1.2) to the following:

$$\{E_\alpha\}_{\alpha\in J} \subset 2^X \text{ s.t } \forall \beta,\gamma \in J,\ E_\beta \cap E_\gamma = \emptyset \implies \mu\Big(\bigcup_{\alpha\in J} E_n\Big) = \sum_{\alpha\in J} \mu(E_\alpha),$$

where J is any index set, possibly uncountable. This means that for any set E we have

$$\mu(E) = \mu\Big(\bigcup_{x\in E}\{x\}\Big) = \sum_{x\in E} \mu(\{x\}),$$

with the agreement that $\mu(E) = \infty$ if the sum above is infinite. This is a well defined measure in the sense that any E is measurable, ie has a measure. However, if we want to have $\mu(E) < \infty$, then for all but at most countably many elements $x \in E$ we will have $\mu(\{x\}) = 0$. Therefore, $\mu(E)$ is determined by $\mu(C)$ where C is an at most countable subset of E. This is inadequate if eventually we seek to use measures to quantify continuous quantities such as area and volume.

In fact, if we are given a function $\mu : 2^X \to [0,\infty]$ and want to look for the collection of sets for which (1.2) holds, then this collection need not all be 2^X (see Section 1.3). This and other reasons (to be revealed in later sections) require one to establish a notion of measurability, that is a notion of when is a collection of sets is the domain for a well defined non-trivial measure that satisfies (1.2).

Definition 1.1.1 (σ-algebra and measurability)**.** Let X be any set. Suppose that there exists a collection Σ of subsets of X with the following properties:

(i) $\emptyset, X \in \Sigma$.

(ii) If $E \in \Sigma$ then $E^c \in \Sigma$.

(iii) Σ is closed under countable unions, ie if $\{E_n\}_{n\in\mathbb{N}}$ is a sequence of sets in Σ then

$$\bigcup_{n\in\mathbb{N}} E_n \in \Sigma.$$

Then we call Σ a σ-algebra and the pair (X,Σ) a measurable space. A set $E \in \Sigma$ is called measurable.

Note. The motivation for σ-algebras that is provided here is technical. There is a rather intuitive and quite ingenius motivation for the definition of σ-algebras in terms of events, probabilty and conditional expectations due to Kolmogorov. This will be discussed in Section 1 and 3 of Chapter 2.

We will now adopt the convention that the domain of any measure, ie the collection of sets for which one can properly define a measure, is a σ-algebra. Before we introduce measures and the Lebesgue integral, we define an important class of functions that will be our candidates for integrability in general, and study some of their properties.

Definition 1.1.2 (Measurable function)**.** A function f from the measurable space (X,Σ_X) to the measurable space (Y,Σ_Y) is said to be measurable if for all $E \in \Sigma_Y$ we have $f^{-1}(E) \in \Sigma_X$. We denote $M(X,Y)$ the set of all measurable functions from X to Y.

In this abstract setting not much can be said about measurable functions, except for some obvious set theorotic properties.

Proposition 1.1.1. Let (X, Σ_X) and (Y, Σ_Y) be measurable spaces and let $f \in M(X, Y)$.

(i) The collection
$$\sigma(f) := \{f^{-1}(E) : E \in \Sigma_Y\},$$
is a σ-algebra on X and $\sigma(f) \subset \Sigma_X$. We say $\sigma(f)$ is the σ-algebra generated by f.

(ii) If (Z, Σ_Z) is a measure space and $g \in M(Y, Z)$ then $g \circ f \in M(X, Z)$.

We now equip the target space Y with an additional structure, such as topological or algebraic (or both), and ask whether measurability is compatible with these structures. So we would like to answer questions such as:

Q1. If Y has a topology $\mathcal{T}$, is there a σ-algebra on Y for which the space $M(X, Y)$ sequentially closed in the topology of pointwise convergence on Y^X?

Q2. If Y is a real vector space, is there a σ-algebra on Y for which $M(X, Y)$ is a real vector subspace of Y^X?

For the first question, if Y is metrizable then the answer is simple: the σ-algebra in question has to contain all open sets. The second question does not have a straight-forward answer given the minimal assumptions. The theory of Bochner measurability (Section 3.1) answers the second question in the case when Y is a separable Banach space, and also in this the σ-algebra on Y has to contain all open sets. For the purpose of the following section, one is interested only in the case when $Y = \overline{\mathbb{R}}$.[(i)] Even though $\overline{\mathbb{R}}$ is not a vector space, if the σ-algebra also contains all open sets, then $M(X, \overline{\mathbb{R}})$ is indeed a vector space.

Let us start by defining the σ-algebra containing all open sets. This is a special case of the following: often times one would like to make certain sets of relevance such as open sets,closed sets,singletons, etc.. measurable. This means we would like to have a σ-algebra containing those sets, without it being unecessarily large. In other words, we would like to find the 'smallest' σ-algebra making a collection $\mathcal{S}$ of sets measurable. This is indeed possible, as the next proposition shows.

Proposition 1.1.2. The abitrary intresction of σ-algebras on a set X is also a σ-algebra.

Definition 1.1.3. Let $\mathcal{S}$ be collection of subsets of X. We denote $\sigma(\mathcal{S})$ the smallest σ-algebra containing $\mathcal{S}$, ie
$$\sigma(\mathcal{S}) := \bigcap_{\alpha \in J} \Sigma_\alpha,$$
where $\{\Sigma_\alpha\}$ is the collection of all σ-algebras on X containing $\mathcal{S}$.

Definition 1.1.4 (Borel σ-algebra)**.** Let $(X, \mathcal{T})$ be a topological space. The Borel σ-algebra is defined as $\mathcal{B}_X := \sigma(\mathcal{T})$.

The Borel σ-algebra guarantees that all open sets, closed sets, their countable intersections and unions are measurable. Another desired property is that if X and Y are topological spaces and are each equipped with their Borel σ-algebras, any continuous function is immediately measurable. This fact will be exploited in subsequent sections to establish a relation

[(i)] The extended real number line $\overline{\mathbb{R}} := [-\infty, \infty] := \mathbb{R} \cup \{-\infty, \infty\}$ is the two point compactificatin of $\mathbb{R}$ and it's topology is generated by sets of the form $[-\infty, a)$, $(a, \infty]$ and (a, b) for $a, b \in \mathbb{R}$.

between arbitrary measurable functions and continuous functions. But in the current scope, we will focus on have a topology *only on the target space.*

With the Borel σ-algebra in hand, we can now answer Q1 when Y is a metric space, but we will need a lemma first.

Lemma 1.1.3. Let (X, Σ_X) be a measurable space and $(Y, \sigma(\mathcal{S}))$ be another measurable space where $\mathcal{S}$ is a collections of subsets of Y. If for all $E \in \mathcal{S}$ we have $f^{-1}(E) \in \Sigma_X$, then f is measurable.

Proof. Consider the collection of sets

$$\Sigma_Y = \{E \in \sigma(\mathcal{S}) : f^{-1}(E) \in \Sigma_X\}.$$

It is clear that $\emptyset, X \in \Sigma$ and that if $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma_Y$ then

$$f^{-1}\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \bigcup_{n\in\mathbb{N}} f^{-1}(E_n) \in \Sigma_X,$$

since Σ_X is closed under countable unions. Therefore Σ_Y is closed under countable unions and is thus a σ-algebra. Since $\mathcal{S} \subset \Sigma_Y \subset \sigma(\mathcal{S})$ and $\sigma(\mathcal{S})$ is the smallest σ-algebra containing $\mathcal{S}$ then $\Sigma_Y = \sigma(\mathcal{S})$. ■

Theorem 1.1.4. Let (X, Σ_X) be a measurable space and Y be a metric space equipped with the Borel σ-algebra $\mathcal{B}_Y$. Then $M(X, Y)$ is a sequentially closed subset of Y^X w.r.t the topology of pointwise convergence.

Proof. Let $\{f_n\}$ be a sequence in $M(X, Y)$ that converges to $f \in Y^X$ (pointwise). To show that $f \in M(X, Y)$, it suffices to show that for all open sets E in Y we have that $f^{-1}(E) \in \Sigma_X$.

Indeed, let E be an open subset of Y. We need to write $f^{-1}(E)$ as countable unions and intersections of sets in Σ_X. This is done as follows. For each $k \in \mathbb{N}$, define

$$E_k = \{y \in Y : B(y, 1/k) \subset E\}.$$

Since E is open then E_k is eventually non-empty after some large enough k. Also, we have that E_k is closed. Indeed, let $\{y_n\}$ be a sequence in E_k converging to $y \in Y$ and let $z \in B(y, 1/k)$. Then there is a $n \in \mathbb{N}$ such that

$$d(y, y_n) < \frac{1}{k} - d(z, y) \quad \text{and therefore} \quad d(z, y_n) \leq d(z, y) + d(y, y_n) < \frac{1}{k}.$$

Thus $z \in B(y_n, 1/k)$, and since $y_n \in E_k$ then $z \in E_k$. Hence $B(y, 1/k) \subset E_k$ and therefore $y \in E_k$. This shows that E_k is closed and therefore $E_k \in \mathcal{B}_Y$ for all $k \in \mathbb{N}$. We also have that

$$f^{-1}(E) = \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} f_n^{-1}(E_k).$$

Since each f_n is measurable then $f_n^{-1}(E_k) \in \Sigma_X$ for all $k, n \in \mathbb{Z}$ and ttherefore $f^{-1}(E)$ being the countable union and intersection of measurable sets is also measurable. ■

For the rest of the section, we will say f is Borel measurable if f is $\overline{\mathbb{R}}$-valued and is measurable when $\overline{\mathbb{R}}$ is equipped with the Borel σ-algebra $\mathcal{B}(\overline{\mathbb{R}})$. We denote

$$M(X) := \{f \in \overline{\mathbb{R}}^X : f \text{ is measurable}\}, \quad M^+(X) := \{f \in M(X) : f \geq 0\} \tag{1.3}$$

Lemma 1.1.5. The collection of sets $\mathcal{S} = \{(-\infty, x) : x \in \mathbb{R}\}$, generates the Borel σ-algebra $\mathcal{B}(\overline{\mathbb{R}})$.

Proof. $\mathcal{S}$ being a subset of the topology on $\overline{\mathbb{R}}$, it is clear that $\sigma(\mathcal{S}) \subset \mathcal{B}(\overline{\mathbb{R}})$. Now

$$\{-\infty\} \cup [b, +\infty] = (-\infty, b)^c \in \sigma(\mathcal{S}).$$

Therefore

$$[-\infty, a) \cup [b, +\infty] = (-\infty, a) \cup \{-\infty\} \cup [b, +\infty] \in \sigma(\mathcal{S}), \quad (\forall a < b),$$

and thus

$$[a, b) = \big([-\infty, a) \cup [b, +\infty)\big)^c \in \sigma(\mathcal{S}).$$

The above is true for all $a, b \in \mathbb{R}$ with $a < b$. Now for any sequence $\{a_n\}$ and $\{b_n\}$ such that $a_n < b_n$ and $a_n \searrow a$ and $b_n \nearrow b$ we have

$$(a, b) = \bigcup_{n=1}^{\infty} [a_n, b_n) \in \sigma(\mathcal{S}).$$

This implies that

$$[-\infty, a] \cup [b, +\infty] = (a, b)^c \in \sigma(\mathcal{S}).$$

Therefore,

$$(a, b] = ([-\infty, b] \cup [d, +\infty]) \cap (a, c) \in \sigma(\mathcal{S}), \quad a < b < c < d.$$

Thus

$$(-\infty, a] \cup (b, +\infty] = (a, b]^c \in \sigma(\mathcal{S})$$

But

$$(-\infty, a] = \bigcap_{n=1}^{\infty} (-\infty, a_n) \in \sigma(\mathcal{S}), \;\; a_n \searrow a.$$

Therefore

$$(b, +\infty] \in \sigma(\mathcal{S}).$$

Hence $\sigma(\mathcal{S})$ contains all basis elements of the toplogy on $\overline{\mathbb{R}}$. Since all open sets are countable unions of basis elements, then $\sigma(\mathcal{S})$ contains the topology. Since $\mathcal{B}(\overline{\mathbb{R}})$ is the smallest σ-algebra containing the topology, we have that $\mathcal{B}(\overline{\mathbb{R}}) \subset \sigma(\mathcal{S})$ and the proof is complete. ■

Proposition 1.1.6. Let (X, Σ) be a measurable space. Then $M(X)$ is a real vector subspace of $\overline{\mathbb{R}}^X$.

Proof. Let $f \in M(X)$. 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 and so $\alpha f \in M(X)$. Now let $g \in M(X)$ be another function. To show that $f + g \in M(X)$, it suffices to show that for all $x \in \mathbb{R}$, the set $\{f + g < x\}$ is measurable by Lemmas 1.1.3 and 1.1.5. The trick here is to write

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

Since for all $r \in \mathbb{Q}$ the sets $\{f < r\}$ and $\{g < x - r\}$ are measurable, then the above set is measurable. Now proceed by induction to prove that any finite linear combination of functions in $M(X)$ is also measurable and the proof is complete. ■

Proposition 1.1.7. Let (X, Σ) be a measurable space and let $\{f_n\} \subset M(X)$.

(i) The functions f and g defined as

$$f(x) = \inf_{n\geq 1} f_n(x) \quad \text{and} \quad g(x) = \sup_{n\geq 1} f_n(x)$$

are in $M(X)$.

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

Definition 1.1.5. A function $f \in M(X)$ is called simple if there are sets $E_1, \ldots, E_n \in \Sigma_X$ and constants $c_1, \ldots, c_n \in \mathbb{R}$ 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 $S(X)$.

Simple functions are a generalization of step functions to abstract measurable spaces. 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 1.1.8. Let $f \in M^+(X)$. There is a sequence of sets $\{E_n\} \subset \Sigma$ such that

$$f(x) = \sum_{n=1}^{\infty} \frac{\mathbf{1}_{E_n}(x)}{k}. \tag{1.4}$$

This implies that the space of positive simple functions $\mathbf{S}^+(X)$ is dense in $M^+(X)$ with respect to the topology of pointwise convergence.

Intuition: A necessary condition for (1.4) 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

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

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 1.1.8. 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 $f_n(x) \to f(x)$, 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

$$\sum_{k=0}^{m_0} \frac{1}{n_0 + k} + \frac{1}{n_1} \leq f(x),$$

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 $n_0 \leq n_1 \leq n_2 \leq \cdots \leq n_k$ and $m_0 \leq m_1 \leq \cdots \leq m_k$ with $n_{j+1} \geq n_j + m_j + 1$ such that

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

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

Proposition 1.1.9 (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 $E \in M$ 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, Σ).

Definition 1.1.6 (Measure)**.** A measure on a measurable space (X, Σ) is a function $\mu : \Sigma \to [0, \infty]$ such that for $\mu(\emptyset) = 0$ and for any sequence $\{E_n\} \subset \Sigma$ we have

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n). \tag{1.5}$$

The triple (X, Σ, μ) is called a measure space.

The following theorem provides a necessary and sufficient condition for a function $\mu : \Sigma \to \overline{\mathbb{R}}$ to be a measure, and we equally it as a definition of measure.

Theorem 1.1.10 (Continuity property)**.** Let (X, Σ) be a measurable space and let $\mu : \Sigma \to \mathbb{R}$ 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

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

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

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

Theorem 1.1.11 (Borel-Cantelli)**.** Let (X, Σ, μ) be a measure space and let $\{E_n\} \subset \Sigma$. 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

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

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 1.1.12. Let (X, Σ, μ) be a measure space and $\{E_n\} \subset \Sigma$ 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 $\lim_{n\to\infty} \mu(E_n) = \mu(E)$.

Theorem 1.1.13 (Egorov)**.** Let (X, Σ, μ) be a finite measure space. Suppose that $\{f_n\} \subset M(X)$ that converges to $f \in M(X)$ almost everywhere. Then for every $\epsilon > 0$, there is a set $E \in \Sigma$ such that $\mu(E) < \epsilon$ and $f_n \to f$ 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 $f_n(x) \to f(x)$, 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

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

Since $\mu(X) < \infty$ we have by part (iii) of Theorem 1.1.10 that

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

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 $f_n \to f$ 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 $\overline{\mathbb{R}}$ valued functions. The approach would be to define the integral for simple functions and proving some of it's properties. Then, we use the density of simple functions to establish the integral and it's properties for functions in $M^+(X)$.

Definition 1.1.7 (Lebesgue integral)**.** Let (X, Σ, μ) be a measure space. Define

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

Then use the above to define

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

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

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

This guarantees that the following definition makes sense

$$\boxed{\int_X f d\mu := \int_X f^+ d\mu - \int_X f^- d\mu \text{ for } f \in M^1(X)}. \tag{1.6}$$

If $A \in M$ we define

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

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

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

then we can use (1.6) to define the integral of f with the integral possibly being $\pm\infty$.

Lemma 1.1.14. 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, 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.

(ii) Equivalently equivalently

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

(i) **(Monotonicity).** Suppopse 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,

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

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

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

(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 1.1.15. Let (X, 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 $\lim_{n\to\infty} \mu(G_n) = \mu(X)$. In addition, if μ is finite then $\lim_{n\to\infty} \mu(E \setminus G_n) = 0$.

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

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

and thus $E = X$. By part (iii) of Theorem 1.1.10 we have that $\mu(E_n) \to \mu(X)$ as desired. ∎

Proposition 1.1.16. Let f and g be two positive measurable functions and let $E, F \in \Sigma$.

(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 a simple 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 1.1.17 (Monotone Convergence Theorem)**.** Let $\{f_n\}$ be an increasing sequence of positive measurable functions that converge pointwise to f. Then

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

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,

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

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

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

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

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

The other reverse inequality follows immediately from monotonicity. ■

Corollary 1.1.17.1. The Lebesgue integral for positive measurable functions has the same operational properties as the Riemann integral and satisfies the same inequalities.

Proof. Let (X, M, μ) be a measure space and let $f, g \in M^+(X)$. Consider two sequences $\{f_n\}$ and $\{g_n\}$ of simple functions that increase to f and g respectively.

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

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

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

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

The proof is complete. ■

Here we have used the monotone convergence theorem to prove linearity of the Lebesgue integral, an approach similar to the one in [3]. 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 1.1.18 (Fatou's Lemma)**.** Let $\{f_n\}$ be a sequence in $M^+(X)$. Then

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

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

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

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

$$\lim_{n\to\infty} \int_X g_n d\mu = \int_X \lim_{n\to\infty} g_n d\mu = \int_X \liminf_{n\to\infty} f_n d\mu,$$

and the inequality is proved. ∎

Corollary 1.1.18.1. 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

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

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

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

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

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

which concludes the proof. ∎

Corollary 1.1.18.2. Suppose that $\{f_n\}$ is a sequence in $M^+(X)$ that converges to f *almost everywhere* on X. Then

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

Theorem 1.1.19 (Dominated Convergence Theorem)**.** Suppose we are given a sequence $\{f_n\}$ in $M^1(X)$ and $\{g_n\}$ in $M^+(X)$ that satisfy the following assumptions

(i) $f_n \to f$ almost everywhere on X.

(ii) $g_n \to g$ almost everywhere and $g \in M^1(X)$.

(iii) $|f_n| \leq g_n$ for all $n \in \mathbb{N}$.

(iv) $\displaystyle\int g_n d\mu \to \int g d\mu$ as $n \to \infty$.

Then we have that

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

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 $\varphi_n \to 2g$ as $n \to \infty$ almost everywhere on E. Now by Fatou's Lemma

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

and

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

and thus

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

Therefore we get

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

Letting $n \to \infty$ we get the desired result. ∎

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

Definition 1.2.1 (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 and the space $(X, \mathcal{B}, \mu)$ is called a Borel space.

Definition 1.2.2 (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.

The definitions above are motivated by the properties of the Lebesgue measure on $\mathbb{R}^d$.

Definition 1.2.3 (Radon measure, version 1)**.** A Radon measure μ is a Borel measure that satifies the following properties.

(i) μ inner regular on all open sets.

(ii) μ outer regular on all Borel sets.

(iii) μ is finite on all compact sets.

Given this definition, what lacks for a Radon measure to be regular is being inner regular on all Borel sets instead of just open sets. Propostion 1.2.2 provides a sufficient condition that remedies this lack.

Definition 1.2.4 (Radon measure, version 2)**.** A Radon measure is a Borel measure that satisfies the following.

(i) It is inner regular on all open sets.

(ii) It is locally finite, that is, every point has a neighbourhood of finite measure.

Any Borel space $(X, \mathcal{B})$ that equipped with a Radon measure is called a Radon space.

The two definition are *not* equivalent for arbitrary topological spaces. However, if X is locally compact Hausdorff (LCH) then there is a one-to-one correspondance between measures defined in definition 1.2.3, measures defined in definition 1.2.4, and positive linear functionals on $C_c(X)$[iii]. Also, *Urysohn's lemma* (Theorem 1.2.4) applies for these space. That is why we choose LCH as a minimum requirement on X. But LCH is still a weak condition. There are LCH metric spaces for which the above definitions are not equivalent. See exercise 7.12 of [1] for a an example of such spaces.

However, a reasonable topological property of that makes the two definitions equivalent is *σ-compactness*, that is, X is a countable union of compact sets. This property is enjoyed by important spaces such as second countable LCH spaces[iv]. Indeed, local compactness implies that the collection $\mathcal{C} := \{U \subset X : U \text{ is open}, \ \overline{U} \text{ is compact}\}$. forms a base for the topology on X. By second-countability, some countable subfamily $\mathcal{C}_0$ of $\mathcal{C}$ is itself a base for X. But then X is covered by $\mathcal{C}_0$ so X is σ-compact.

[iii] This is the content of the Riesz–Markov–Kakutani representation theorem.

[iv] Second countable LCH cannot be weakened to separable LCH. See counter-example 65 in [5].

Proposition 1.2.1. Let X be a locally compact σ-compact Hausdorff space. Let μ be a Borel measure that is inner regular on all open sets. Then μ is locally finite if and only if it is outer regular on all Borel sets and finite on compact sets.

We now introduce a desirable measure theoretic property of measure spaces.

Definition 1.2.5 (σ-finiteness)**.** Let (X, Σ, μ) be an arbitrary measure space. X is called σ-finite X is a countable union of sets with finite measure.

Proposition 1.2.2. A Radon measure μ is inner regular on all σ-finite sets. This implies that if μ is σ-finite then μ is regular.

Corollary 1.2.2.1. If X is σ-compact, then every Radon measure on X is regular.

Proposition 1.2.3. Let $(X, \mathcal{B}, \mu)$ 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 1.2.4 (Urysohn's lemma)**.**

Theorem 1.2.5 (Lusin)**.** Let X be a locally compact Hausdorff toplogical space equipped with the Borel σ-algebra and a Radon measure μ. Let $f \in M(X)$ such that $\mu(\{f \neq 0\}) < \infty$. For every $\epsilon > 0$, there a function $\varphi \in C_c(X)$ such that $\mu(\{f \neq \varphi\}) < \epsilon$. Furthermore, if f is bounded, then φ can be chosen so that $\mathrm{essup}(g) \leq \mathrm{essup}(f)$.

1.3 Outer Measures and Product Measures

Definition 1.3.1. An outer measure on a set X is a function $\mu^* : 2^X \to [0, \infty]$ such that $\mu^*(\emptyset) = 0$ and for any sequence $\{E_n\} \subset 2^X$ that cover a set $E \in 2^X$ we have

$$\mu^*(E) \leq \sum_{n=1}^{\infty} \mu^*(E_n). \tag{1.7}$$

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

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A) \tag{1.8}$$

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

Proof. ■

Condition (1.8) has an intuitive explanation in terms of events and probability, as will be explained in section 1.6.

Theorem 1.3.2 (Hahn-Kolmogorov)**.** Let $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ be σ-finite measure spaces and let $\mathcal{F} \otimes \mathcal{G} := \sigma(\mathcal{F} \times \mathcal{G})$. If we define $\eta : \mathcal{F} \times \mathcal{G} \to \mathbb{R}$ as

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

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

Proof. ■

Definition 1.3.2. If $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ 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 1.3.2.

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

Definition 1.4.1 (Lebesgue-Stieltjes outer measure on $\mathbb{R}$)**.** Let $F : \mathbb{R} \to \mathbb{R}^+$ 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

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

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 1.4.1. 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 $f(x^-) \le f(x) < f(x^+)$ or $f(x^-) < f(x) \le f(x^+)$. 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 $q : D \to \mathbb{Q}$ is injective. ∎

Proposition 1.4.2. 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 $x \in X \setminus D_f$, let $\{a_n\}$ and $\{b_n\}$ be sequences in $X \setminus D_f$ such that $a_n \nearrow x$ and $b_n \searrow x$. By definition we will then have

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

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

Definition 1.4.2. 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 1.4.3. $\Sigma^{(1)}$ contains the Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}$.

Definition 1.4.3. 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 1.4.4 (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 $n \ge 2$. 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 $\mathbb{R}^n$.

Definition 1.4.4 (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 Lebesgure 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

A rather unusual approach based on [6].

Definition 1.4.5. Let $\preceq$ be the partial order on $\mathbb{R}^n$ defined as follows. If $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ then $x \preceq y$ if and only if $x_j \leq y_j$ for all $j = 1, \ldots, n$.

Definition 1.4.6 (Increasing right-continuous function in $\mathbb{R}^n$)**.** Let $F : \mathbb{R}^n \to \mathbb{R}$ be any function Let $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ be such that $a \preceq b$. For $1 \leq k \leq n$ let

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

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 1.4.7 (Lebesgue-Stieltjes outer measure on $\mathbb{R}^n$ for $n \geq 2$)**.** Let $F : \mathbb{R}^n \to \mathbb{R}$ be an increasing right continuous function as in the above defintion. 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 1.3.1 and λ_n to be the restriction of λ_n^* to $\mathcal{L}(\mathbb{R}^n)$.

Proposition 1.4.5. 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 [4, 2], covering most of the fundamental properties of these spaces and some of their uses.

1.6 Absolute continuity and Radon-Nikodym theorem

Theorem 1.6.1 (Lebesgue-Radon-Nikodym)**.** Let μ and ν be finite measures on a measurable space (X, M). There is a function $f \in L^0(\mu) \cap L^0(\nu)$ and a μ-null set $F \in M$ such that for all $E \in M$ we have

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

Chapter 2

Basic Probability Theory

In many ways this section is inspired by excellent books [1, 2, 3].

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 (1.8), 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 information, rather than geometry.

Using measure theory to formalize probability is the notorious contribution of soviet Mathematician Andrey Nikolaevich Kolmogorov, which can be found in his book *The Foundations of the Theory of Probability*, originally published in German as *Grundbegriffe der Wahrscheifnlichkeitsrechnung* in 1933.

Bibliography

[1] J.B. Walsh. *Knowing the Odds: An Introduction to Probability.* Graduate studies in mathematics. American Mathematical Society, 2012.

[2] S.E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models.* Number v. 11 in Springer Finance Textbooks. Springer, 2004.

[3] R.M. Dudley, B. Bollobas, and W. Fulton. *Real Analysis and Probability.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002.

2.1 Sample Spaces, Measurable Events and Probability Measures

Definition 2.1.1. 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 definiton 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 $E_1 \subset E_2 \subset E_3 \subset \cdots$ are measurable then $\cup E_n$ is measurable.
4. If $E_1 \supset E_2 \supset E_3 \supset \cdots$ 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 1.4.4. 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 2.2.1 (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 $X : \Omega_1 \to \Omega_2$ be measurable. The push-forward is 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 2.2.1 (Change of variables)**.** Let Let $(\Omega_1, \Sigma_1, \mu)$ be a measure space and (Ω_2, Σ_2) be a measurable space. If $X : \Omega_1 \to \Omega_2$ is measurable and $X_*(\mu)$ is the push-forward measure of X then for any measurable function $g : \Omega_2 \to \mathbb{R}$ we have

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

Definition 2.2.2 (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 $X : \Omega \to \overline{\mathbb{R}}$. Each random variable defines the following.

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

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

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

$$\boxed{\mathbb{E}[X] := \int_\Omega X d\mathbb{P} = \int_{\mathbb{R}} x \, dX_*\mathbb{P}.}$$

(iii) **(Variance)** We define variance of X to be

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

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

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

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

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

Suppose now that $Y : \Omega \to \overline{\mathbb{R}}$ is another random variable.

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

$$\boxed{\mathrm{Cov}(X, Y) := \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right].}$$

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

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

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

A random variable is thus an $\overline{\mathbb{R}}$-valued function X with random input. It is called a *discrete random variable* if it is of the $\sum_{k=1}^{\infty} c_k \mathbf{1}_{E_k}$ with $E_k \in \mathcal{F}$. It is called *continuous random variable* if it has continuous c.d.f, which is equivalent to saying that $\mathbb{P}[X = x] = 0$ for all $x \in \overline{\mathbb{R}}$. It is called *mixed* if it is neither.

Proposition 2.2.2. 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 is right-continuous.

(iii) F_X satisfies the following limits:

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

(iv) The Lebesgue-Stieltjes measure induced by F_X is $X_*\mathbb{P}$.

(v) If F is continuous then f_X exists almost everywhere and

$$\int_{\mathbb{R}} f_X(x)dx = 1, \quad \mathbb{P}^X(a,b) = \int_a^b f_X(x)dx \text{ for all } a,b \in \mathbb{R}\,.$$

(vi) If F is differentiable then f_X exists everywhere, f_X is the derivative of F, and

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

Proof. If $x \le y$ and $\omega \in \{X \le x\}$ then $X(\omega) \le x \le y$ and hence $\omega \in \{X \le y\}$. Therefore $\{X \le x\} \subset \{X \le y\}$ and

$$F(x) = \mathbb{P}(\{X \le x\}) \le \mathbb{P}(\{X \le y\}) = F(y),$$

which proves (i).

We will use the fact that $\mathbb{P}$ satisfies the continuity condition. Now fix $x \in \mathbb{R}$ and let $\{x_n\}$ be any sequence converging to x and $x_n \ge x$ for all n. We want to show that $F(x_n) \to F(x)$. First, define the sequence

$$s_n = \sup_{k\ge n} x_k,$$

then clearly $x \le x_n \le s_n$ for all n. In addition, s_n is decreasing and hence (i) implies that

$$F(x) \le F(x_n) \le F(s_n). \tag{2.1}$$

Furthermore, $\{s_n\}$ is a subsequence of $\{x_n\}$ and thus converges to the same limit as $\{x_n\}$. Now we will construct a decreasing sequence of sets using $\{s_n\}$. Let $E_n := \{X \leq s_n\}$, then one clearly has that $E_n \subset E_{n+1}$ (since $s_{n+1} \leq s_n$) and that

$$\bigcap_{n=1}^{\infty} E_n = \bigcap_{n=1}^{\infty}\{X \leq s_n\} = \{X \leq x\}.$$

We can thus the continuity property of $\mathbb{P}$ to get

$$F(x) = \mathbb{P}(\{X \leq x\}) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mathbb{P}(E_n) = \lim_{n\to\infty} F(s_n).$$

Hence, by taking limits in (2.1) on gets (ii).

Now suppose that $x_n \nearrow +\infty$ then the sequence of sets $\{X \leq x_n\}$ is an increasing sequence of sets with

$$\bigcup_{n=1}^{\infty}\{X \leq x_n\} = \{X \in \mathbb{R}\} = \Omega,$$

and hence by the continuity property of $\mathbb{P}$ we get

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mathbb{P}(\{X \leq x_n\}) = \mathbb{P}\left(\bigcup_{n=1}^{\infty}\{X \leq x_n\}\right) = \mathbb{P}(\Omega) = 1,$$

and (iii) is proved. Similarly, if $x_n \searrow -\infty$ then $\{X \leq x_n\}$ is a decreasing sequence of sets with

$$\bigcap_{n=1}^{\infty}\{X \leq x_n\} = \{X = -\infty\} = \emptyset,$$

and hence

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mathbb{P}(\{X \leq x_n\}) = \mathbb{P}\left(\bigcap_{n=1}^{\infty}\{X \leq x_n\}\right) = \mathbb{P}(\emptyset) = 0,$$

which finishes the proof. ∎

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. We call $\mathbb{P}^X$ the *probability law* induced by X.

Theorem 2.2.3 (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

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

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

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

as desired. ■

Commonly occuring random variables.

Definition 2.2.3. A random variable $X : \Omega \to \overline{\mathbb{R}}$ is said to be normal there are numbers $\mu, \sigma \in \mathbb{R}$ such that the p.d.f of X is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma}}.$$

μ is called the mean of X and σ is called the standard deviation.

Random Vectors.

Definition 2.2.4 (Random vector)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random vector $\mathbf{X}$ is a mapping $\mathbf{X} : \Omega \to \mathbb{R}^d$ that is measurable with respect to the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$. In particular, it is a vector $(X_1, \ldots, X_d)$ with each component being a random variable.

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

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

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

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

(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 $F_{\mathbf{X}} : \mathbb{R}^d \to \mathbb{R}$ defined as

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

Theorem 2.2.4 (n-dimensional Chebychev's inequality)**.** Let $X : \Omega \to \mathbb{R}^n$ be a random vector with mean μ and covariance matrix $C = [\mathrm{Cov}(X_i, X_j)]_{i,j=1}^n$. 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 2.3.1 (Conditional probability)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Fix an event $B \in \mathcal{F}$ such that $\mathbb{P}(B) > 0$. We define the meaure $\mathbb{P}\left[\cdot \mid B\right]$

$$\mathbb{P}\left[A \mid B\right] := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \text{for } A \in \mathcal{F}.$$

The above quanitity is called the conditional probability of A given B. The events A and B are called independent if $\mathbb{P}\left[A \mid B\right] = \mathbb{P}(A)$.

Definition 2.3.2 (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 $\{\mathcal{G}_1, \cdots, \mathcal{G}_n\}$ be a collection of sub σ-algebras of $\mathcal{F}$. then we call this collection independent if for all $A_1 \in \mathcal{G}_1, \ldots, A_n \in \mathcal{G}_n$ 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 independent of $\mathcal{G}_{n+1}$ and $\sigma(\mathcal{G}_1 \cup \cdots \cup \mathcal{G}_n)$ 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 $\sigma(X_1, \ldots, X_n)$ are independent for all $n \in \mathbb{N}$.

Definition 2.3.3 (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 $\mathbb{E}[X \mid \mathcal{G}]$ is defined as random variable having the following properties.

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

(ii) For all $A \in \mathcal{G}$ we have that $\mathbb{E}\Big[\mathbf{1}_A \cdot \mathbb{E}\left[X \mid \mathcal{G}\right]\Big] = \mathbb{E}[\mathbf{1}_A \cdot X]$.

Property (ii) is usually called partial averaging.

Theorem 2.3.1. Suppose that X is a random variable and $\mathcal{G}$ is a sub σ-algebra of $\mathcal{F}$. Then $\mathbb{E}\left[X \mid \mathcal{G}\right]$ exists.

Proof. Suppose that $X \in L^1(\mathbb{P})$. Define the measure ν on $\mathcal{G}$ as

$$\nu(A) = \int_A X d\,\mathbb{P}, \quad \text{for } A \in \mathcal{G}.$$

It is clear that $\nu << \mathbb{P}$. Therefore, by Radon-Nikodym theorem there is a $\mathcal{G}$-measurable function, which we call $\mathbb{E}\left[X \mid \mathcal{G}\right]$ such that

$$\nu(A) = \int_A \mathbb{E}\left[X \mid \mathcal{G}\right] d\,\mathbb{P} = \int_A X d\,\mathbb{P},$$

and this function is unique up to a set of $\mathbb{P}$-measure 0. ∎

Suppose that $A, B \in \mathcal{F}$ and consider the conditional expectation $\mathbb{E}[\mathbf{1}_A \mid \sigma(B)]$. Conditions (i) and (ii) of the above definition then imply that if $\mathbb{P}(B) \neq 0$ then

$$\mathbb{E}\left[\mathbf{1}_A \mid \sigma(B)\right](\omega) = \begin{cases} \mathbb{P}\left[A \mid B\right] & \text{if } \omega \in B, \\ \mathbb{P}\left[A \mid B^c\right] & \text{if } \omega \in B^c. \end{cases}$$

So we can define conditional probability as a random variable $\mathbb{P}\left[A \mid B\right] := \mathbb{E}\left[\mathbf{1}_A \mid \sigma(B)\right]$. It is also well defined even if $\mathbb{P}(B) = 0$ but then $\mathbb{P}\left[A \mid B\right]$ equals zero on B and $\mathbb{P}(A \cap B^c)$ on B^c.

Definition 2.3.4. 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 2.3.2. 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 2.4.1 (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

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

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

Theorem 2.4.1. The function

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

is complete metric on $\mathcal{M}(\Omega)$ and $X_n \to X$ in probability if and only if $d(X_n, X) \to 0$.

Theorem 2.4.2 (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 $\mathrm{Var}[X_n] = \sigma^2$ for all $n \in \mathbb{N}$. We have that

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

Proof. Since the X_n's are independent then

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

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

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

which gives the desired result. ■

Definition 2.4.2 (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

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

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

The proof of the following claim is trivial.

Theorem 2.4.3 (Continuous mapping theorem)**.** Let $\{X_n\}$ be a sequence of random vectors in $\mathbb{R}^n$ and let $g : \mathbb{R}^n \to \mathbb{R}^m$ 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 2.4.4. 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

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

Theorem 2.4.5 (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

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

or equivalently

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

Theorem 2.4.6 (Generalized Central Limit Theorem)**.** Under the assumptions of Theorem 2.4.5, if $f : \mathbb{R}^d \to \mathbb{R}^m$ is a continuously differentiable function with Jacobian matrix $J(\mathbf{x})$ then

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

2.5 Accumulating Information using Discrete Filrations.

2.6 Basic Statistics Problems in the Language of Probability Theory

Remark. We have that for $|x| < 1$ that

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \implies \frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} n x^{n-1} \implies \frac{x}{(1-x)^2} = \sum_{n=1}^{\infty} n x^n.$$

Remark. Suppose $(\Omega, \mathcal{F}, \mathbb{P}) = \left(\prod \Omega_n, \bigotimes \mathcal{F}_n, \prod \mathbb{P}_n\right)$ is the infinite coin tossing space (corresponding to a fair coin). Define

$$E := \{\omega \in \Omega : \exists n \in \mathbb{N} \text{ s.t } \omega_{n-1} = \omega_n = H\}.$$

Let $X : \Omega \to \mathbb{N}$ be defined as

$$X(\omega) = \begin{cases} \min\{n : \omega_{n-1} = \omega_n = H\} & \text{if } \omega \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Compute $\mathbb{E}[X]$.

Note that X is nothing the number of coin tosses needed until two heads are observed. Here are some heuristics first. If we toss and get T on the first throw, then we repeat. By independence, the expected number of throws until we get HH is still the same, but we have tossed at least once.

$$\mathbb{E}[X] = 2p^2 + (1 + \mathbb{E}[X])(1-p) + (2 + \mathbb{E}[X])p(1-p).$$

Therefore

$$\mathbb{E}[X] = \frac{1-p^2}{p^2(1-p)}.$$

Solution. We only treat the case $p = q = 1/2$. Since $X(\Omega) = 2 + \mathbb{N}$ we have

$$\mathbb{E}[X] = \sum_{n=2}^{\infty} n \cdot \mathbb{P}[X = n] = 2p^2 + \sum_{n=3}^{\infty} n \cdot \mathbb{P}[X = n].$$

Now for $n \geq 3$ we have that $\omega \in \{X = n\}$ if $\omega_{n-1} = \omega_n = H$ and $\omega_1 \cdots \omega_{n-2}$ does not contain consecutive heads. This also forces that $\omega_{n-3} = T$ (or else we would have had $X(\omega) = n-1$). Now the number of outcomes $\omega \in \Omega_n$ with no consecutive heads equals

$$\begin{cases} b_n = b_{n-1} + b_{n-2} & \text{if } n \geq 2, \\ b_0 = 1, \ b_1 = 2. \end{cases}$$

By solving the above recurrence explicitely, it can be shown that

$$b_n = \frac{1}{2}\left(1 + \frac{3}{\sqrt{5}}\right)\left(\frac{1+\sqrt{5}}{2}\right)^n + \frac{1}{2}\left(1 - \frac{3}{\sqrt{5}}\right)\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

Therefore, we have that

$$\mathbb{P}(\{X = n\}) = \frac{b_{n-3}}{2^n} = \frac{(1-c)}{\phi^3}\left(\frac{\phi}{2}\right)^n + \frac{c}{\phi_*^3}\left(\frac{\phi_*}{2}\right)^n.$$

Now

$$\sum_{n=3}^{\infty} n\left(\frac{\phi}{2}\right)^n = \frac{\phi}{2-2\phi-\phi^2/2} - \frac{\phi}{2} - 2\left(\frac{\phi}{2}\right)^2.$$

Therefore

$$\frac{(1-c)}{\phi^3}\sum_{n=3}^{\infty} n\left(\frac{\phi}{2}\right)^n = (1-c)\left[\frac{2}{\phi^2(2-\phi)^2} - \frac{1}{2\phi^2} - \frac{1}{2\phi}\right]$$

and similarly

$$\frac{c}{\phi_*^3}\sum_{n=3}^{\infty} n\left(\frac{\phi_*}{2}\right)^n = c\left[\frac{2}{\phi_*^2(2-\phi_*)^2} - \frac{1}{2\phi_*^2} - \frac{1}{2\phi_*}\right]$$

so that after a computation we get

$$\mathbb{E}[X] = \frac{1}{2} + \frac{(1-c)}{\phi^3}\sum_{n=3}^{\infty} n\left(\frac{\phi}{2}\right)^n + \frac{c}{\phi_*^3}\sum_{n=3}^{\infty} n\left(\frac{\phi_*}{2}\right)^n = \frac{1}{2} + \frac{11}{2} = 6.$$

■

Solution. Let E_n be event that H was observed on the n'th toss, ie $E_n = \{\omega : \omega_n = H\}$. By the law of total expectation, we have that

$$\mathbb{E}[X] = \mathbb{P}(E_1)\mathbb{E}[X \mid E_1] + (1-\mathbb{P}(E_n))\mathbb{E}[X \mid E_n^c],$$

Then we also get

$$\mathbb{E}[X \mid E_1] = \mathbb{P}(E_2)\cdot\mathbb{E}[X \mid E_1 \cap E_2] + (1-\mathbb{P}(E_2))\mathbb{E}[X \mid E_1 \cap E_2^c].$$

■

Chapter 3

Vector Valued Measures and Measure Valued Random Variables

Bibliography

[1] Tuomas Hytonen, Jan van Neerven, Mark Veraar, and Lutz Weis. *Analysis in Banach Spaces, Volume I: Martingales and Littlewood-Paley Theory*. 12 2016.

[2] J. Diestel and J.J. Uhl. *Vector Measures*. Mathematical surveys and monographs. American Mathematical Society, 1977.

[3] O. Kallenberg. *Random Measures*. Elsevier Science & Technology Books, 1983.

3.1 Bochner Spaces: Measurability, Integration and Duality

Based on [1].

3.2 Bochner Integral

Based on [1].

3.3 Vector Measures

Based on [2].

3.4 Introduction to Random Measures and Intensity

Based on [3].

Chapter 4

General Stochastic Processes

One often hears of stochastic processes when attempting to model phenomena that involves randomness in time such as speech signals, weather, the stock market, behavior of particles in fluid, radioactive decy, and many more. But the intuition of stochastic processes originated from much more grounded sources such as coin tossing and any game of chance in general. In fact the term *Martingale*, formally introduced J. Ville [1] and adopted by pioneers in the field such as P. Lévy, J. Doob and E. Borel and used to describe a wide class of important processes, has origins dating back to the early 18th century. It describes a strategy used by gamblers in which "a gambler doubles his stake at each each loss, in order to quit with a sure profit, provided that he wins once" [2].

Another source of motivation for stochastic processes, perhaps the most well known, was the observation made by scottish botanist Robert Brown in 1827, on the movement of Pollen inside a fluid. This kind of movement was later dubbed *Brownian motion*, and it's discovery influenced the work of many physicists such as Albert Einstein. The mathematical formulation of Brownian motion provided both evidence for the existence of atoms and insight on how to compute their size.

In mathematical terms, a stochastic process is a collection $\{X_t\}_{t\in T}$ of S-valued random variables, where T is some index set (usually positive time) and S is some state space (usually $\mathbb{R}^d$). It is also a function whose input is random and output is a function from T to S. Therefore, it is no surprise that the theory stochastic processes is laborious and quite demanding, requiring results from many other fields of mathematical analysis.

The existence of sample spaces on which one can define certain processes is a consequence of results in Infinite Dimensional Measure Theory and Functional Analysis, such as the Kolmogorov extension theorem and the Bochner-Minlos theorem. Furthermore, the theory of stochastic processes links up quite admirably with the theory of partial differential. One fundamental fundamental relation between Brownian motion and ellpitic differential equations is established in this chapter, and the matter is discussed thouroughly in later chapters.

Bibliography

[1] J. Ville. *Étude Critique de la Notion de Collectif.* Collection des monographies des probabilités. Gauthier-Villars, 1939.

[2] R. Mansuy. *Histoire de martingales.* Journal Electronique de l'Histoire des Probabilités et de la Statistique. Mathématiques et sciences humaines, 2005.

[3] I. Karatzas, I.K.S. Shreve, S. Shreve, and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics (113) (Book 113). Springer New York, 1991.

[4] R. Dalang, D. Khoshnevisan, F. Rassoul-Agha, C. Mueller, D. Nualart, and Y. Xiao. *A Minicourse on Stochastic Partial Differential Equations*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2008.

[5] Daniel Alpay, Palle Jorgensen, and David Levanony. On the equivalence of probability spaces, 2016.

[6] H. Holden, B. Oksendal, and J. Uboe. *Stochastic Partial Differential Equations*. Birkhauser Boston, 2014.

[7] R.M. Dudley, B. Bollobas, and W. Fulton. *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002.

4.1 Definition, Existence and Measurability

A stochastic process can be viewed in any one three equivalent ways, and we will often frequently switch between those viewpoints.

Definition 4.1.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let T be any set. An $\mathbb{R}^d$ valued stochastic process is defined in the one of the following ways.

(i) It is a function $X : T \times \Omega \to \mathbb{R}^d$ such that $X(t, \cdot) : \Omega \to \mathbb{R}^d$ is measurable for each t.

(ii) It is a collection of $\mathbb{R}^d$-valued random variables $\{X_t\}$ indexed by T.

(iii) It is a random variable $X : \Omega \to (\mathbb{R}^n)^T$, where $(\mathbb{R}^n)^T$ is equipped with the Borel σ-algebra generated by the product topology.

If T has a partial order then a stochastic process is called a random field.

Generally, we will define a stochstic process using (i) as it is the most notationally convenient. But then the reader should immeadiately understand that X_t or $X(t)$ is just the radom variable $X(t, \cdot) : \Omega \to \mathbb{R}^d$, and that we will switch between these notations quite frequently, sometimes in the same context. Therefore, it should always be clear that

$$X_t(\omega) = X(t)(\omega) = X(t, \omega), \quad \text{for all } t \in T,\ \omega \in \Omega.$$

On rarer occasions we use the notation $X(\omega)$ or X_ω to mean the function $X(\cdot, \omega) : T \to \mathbb{R}^d$. The aformentioned function is also called a *path* or *realization* of the process. It should also be understood that for all $t \in T$ and $\omega \in \Omega$ we have

$$X_\omega(t) = X(\omega)(t) = X(t, \omega), \quad \text{for all } t \in T,\ \omega \in \Omega.$$

We now elaborate on part (iii) of the above definition. Let us first define the product topology on $(\mathbb{R}^n)^T$ and the Borel σ-algebra $\mathcal{B}$ generated by the product topology.

Definition 4.1.1 (Product topology on $(\mathbb{R}^d)^T$)**.** For an open set U in $\mathbb{R}^d$ and $t \in T$ define the collection

$$\mathcal{S} := \{\pi_t^{-1}(U) : U \text{ open in } \mathbb{R}^d,\ t \in T\} \quad \text{where } \pi_t^{-1}(U) := \{f \in (\mathbb{R}^d)^T : f(t) \in U\}.$$

The collection $\mathcal{S}$ of all such sets is a subasis for the product topology. Hence we can form a basis $\mathcal{C}$ for the product topology by taking finite instersections of elements in $\mathcal{S}$ as follows

$$\bigcap_{j=1}^{n} S(t_j, U_j) = \pi_{t_1}^{-1}(U_1) \cap \cdots \cap \pi_{t_n}^{-1}(U_n),$$

where each U_j is open in $\mathbb{R}^d$. Then the product toplogy is the unique topology on $(\mathbb{R}^d)^T$ generated by the basis $\mathcal{C}$. //

Definition 4.1.2. We denote $\mathcal{B}$ the Borel σ-algebra generated by the product topology on $(\mathbb{R}^n)^T$. Equivalently, it is the smallest σ-algebra containing *cylinder* sets, ie sets of the form

$$\pi_{t_1}^{-1}(B_1) \cap \cdots \cap \pi_{t_n}^{-1}(B_n) = \{f \in (\mathbb{R}^d)^T : f(t_1) \in B_1, \ldots, f(t_n) \in B_n\}, \tag{4.1}$$

where $B_1, \ldots, B_n$ are Borel subsets of $\mathbb{R}^d$. //

Definition 4.1.2. Let X be a stochastic process. Each random variable defines the following.

(i) **(Distribution Measure).** The probabilty law induced by X on $(\mathbb{R}^d)^T$ is the push-forward measure

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

(ii) **(FIDIs).** For each $n \in \mathbb{N}$ and $t_1, \dots, t_n \in T$, we define a measure on $\mathbb{R}^{dn}$ called a *finite dimensional distribution* of X as

$$\boxed{\nu_{t_1,\dots,t_n}(B_1 \times \cdots \times B_n) = \mathbb{P}^X\left(\pi_{t_1}^{-1}(B_1) \cap \cdots \cap \pi_{t_n}^{-1}(B_n)\right),}$$

where $B_1, \dots, B_n$ are Borel subsets of $\mathbb{R}^d$.

(iii) **(Mean Function).** The mean function of X is the mapping $m : T \to \overline{\mathbb{R}}$ defined as

$$\boxed{m_X(t) : \mathbb{E}[X_t].}$$

(iv) **(Covariance Function).** The covariance function of X is the function $C : T \times T \to \overline{\mathbb{R}}$ defind as

$$\boxed{C_X(s,t) := \mathrm{Cov}(X_s, X_t).}$$

For a real and vector valued random variable X, it is the distribution measure that determine the type of the variable. The domain Ω on which it is defined is irrelevant, and X is usually identified with it's probability distribution measure, ie the probability law $\mathbb{P}^X$ that it induces on it's state space $\mathbb{R}^d$. The same concept is used for a stochastic process X. It is usually identified with it's probability distribution measure $\mathbb{P}^X$, ie the probability law it induces on $(\mathbb{R}^d)^T$.

Another reason for this identification is that once a probability measure $\mathbb{P}$ on $((\mathbb{R}^d)^T, \mathcal{B})$ is given, there is a canonical way to construct a process X having $\mathbb{P}$ as it's distribution measure. Indeed, we can simply define $(\Omega, \mathcal{F}, \mathbb{P})$ to be $((\mathbb{R}^d)^T, \mathcal{B}, \mathbb{P})$ and then $X : T \times \Omega \to \mathbb{R}^d$ as

$$X(t, \omega) := \omega(t), \quad \omega \in \Omega = (\mathbb{R}^d)^T.$$

Then the random variable $X : \Omega \to (\mathbb{R}^d)^T$ is just the identity map and hence it's distribution measure $\mathbb{P}^X$ is the same as $\mathbb{P}$. Therefore, for the large part of this section, instead of studying stochastic processes, we will study probability measures on $((\mathbb{R}^d)^T, \mathcal{B})$.

Let $\mathbb{P}$ be a probability measure on $(\mathbb{R}^d)^T$. Then this measure completely determines a family $\mathscr{F}$ of probability measures called called the finite dimensional measures of $\mathbb{P}$ defined by

$$\mathscr{F} := \{\nu_{t_1,\dots,t_n} : \text{ for all } t_1, \dots, t_n \in T \text{ and for all } n \in \mathbb{N}\}. \tag{4.2}$$

where each $\nu_{t_1,\dots,t_n}$ is a measure on $\mathbb{R}^{nd}$ that is defined by

$$\nu_{t_1,\dots,t_n}(B_1 \times \cdots \times B_n) = \mathbb{P} \circ \pi^{-1}(B_1 \times \cdots \times B_n), \tag{4.3}$$

where $B_1, \dots, B_n$ are Borel sets in $\mathbb{R}^d$ and $\pi : (\mathbb{R}^d)^T \to (\mathbb{R}^d)^n$ is the natural projection map. This familiy satifies what are called *natural consistency conditions*, that is, for all Borel sets $B_1, \dots, B_n$ in $\mathbb{R}^d$, for all permutations $\sigma \in S_n$, and for all $m \in \mathbb{N}$ one has that

$$\nu_{t_{\sigma(1)},\dots,t_{\sigma(n)}}(B_1 \times \cdots \times B_n) = \nu_{t_1,\dots,t_n}(B_{\sigma^{-1}(1)} \times \cdots \times B_{\sigma^{-1}(n)}), \tag{4.4}$$

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

$$\nu_{t_1,\dots,t_n}(B_1 \times \cdots \times B_n) = \nu_{t_1,\dots t_n,t_{n+1},\dots,t_{n+m}}(B_1 \times \cdots \times B_n \times \mathbb{R}^{md}). \tag{4.5}$$

The next theorem due to Kolmogorov, states that if conversely, we are given a family $\mathscr{F}$ of probability measures that satisfies consistency conditions (4.4) and (4.5), then there is a unique probability measure $\mathbb{P}$ on $(\mathbb{R}^d)^T$ for which (4.3) holds. We will prove a slightly more general result, but we introduce some notation first.

Definition 4.1.3. Let $\{(S_t, \Sigma_t)\}_{t\in T}$ be a colletion of measurable spaces. For each *ordered* subset F of T we write

$$(S_F, \Sigma_F) := \Big(\prod_{t\in F} S_t, \bigotimes_{t\in F} \Sigma_t\Big).$$

A family of probability measures $\mathscr{F} := \{\nu_F : F \subset T, |F| < \infty\}$ is called *consistent with respect to* $\{(S_t, \Sigma_t)\}$ if for every finite set F we have that $\nu_F \in \mathscr{F}$ is a probability measure on (S_F, Σ_F) and these measures satisfy (4.4) and (4.5) (of course, with the Borel sets taken in the S_t instead of $\mathbb{R}^d$).

Definition 4.1.4. A separable metric space (S, d) is universally measurable (u.m.) iff for every law $\mathbb{P}$ on the completion $\overline{S}$ 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 $\mathbb{P}$.

Theorem 4.1.1 (Kolmogorov)**.** Let T be any set and let $\{(S_t, \mathcal{B}_t)\}_{t\in T}$ be a collection of universally measurable metric spaces. Consider a family $\mathscr{F} := \{\nu_F : F \subset T, |F| < \infty\}$ of measures that is consistent with respect to this is collection. Then there is a unique probability measure $\mathbb{P}$ on $(S_T, \mathcal{B}_T)$ such that for any finite subset F ot T, $\mathbb{P}$ restricted to S_F is ν_F.

Corollary 4.1.1.1. Under the same assumptions as theorem 4.1.1, there is stochastic process X such that $\mathscr{F}$ contains exactly all the FIDI's of X.

Proof. ■

Definition 4.1.5 (Distinguishing between stochastic processes)**.** Let $\{X\}_{t\in T}$ and $\{Y\}_{t\in T}$ be two stochastic processes.

(i) The processes are called equal in distribution if they have the same finite dimensional distributions.

(ii) The processes are a modification of one another if $\mathbb{P}[X(t) = Y(t)] = 1$ for all $t \in T$.

(iii) The two processes are called indistinguishable if $X(\omega) = Y(\omega)$ for almost all $\omega \in \Omega$.

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

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 4.1.6 (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 $X : T \times \Omega \to \mathbb{R}^n$ 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 $X : T \times \Omega \to \mathbb{R}^n$ 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 4.1.2 (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.

Definition 4.1.7. A process X is called square integrable on $[a,b] \subset \mathbb{R}^+$ if

$$\mathbb{E}\left[\int_a^b X(t)^2 dt\right] < \infty \quad \text{or} \quad \int_\Omega \int_a^b X(t,\omega)^2 dt d\omega < \infty.$$

Theorem 4.1.3 (Karuhen-Mercer-Loève)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $T = [a,b]$. Consider a square integrable stochastic process X with zero mean and continuous covariance function C which is positive semi-definite. Define the operator $T : L^2([a,b]) \to L^2([a,b])$ as

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

Let $\{e_n\}$ be an orthonormal basis for $L^2([a,b])$ formed by the eigenfunctions of T. For each n, define the process X_n as

$$X_n(t,\omega) = \sum_{k=1}^n e_k(t) \int_a^b X(s,\omega) e_k(s) ds.$$

Then for all $t \in [a,b]$ we have $X_n(t) \to X(t)$ in $L^2(\Omega)$, and for all $\omega \in \Omega$ we have $X_n(\omega) \to X(\omega)$ 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 4.2.1 (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

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

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

(ii) X is continuous in mean at t_0 if

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

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

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

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

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

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

Theorem 4.2.1 (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

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

Then there is a continuous version of $\{X_t\}$. In addition, if we call 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. ■

Cadlag processes.

Proposition 4.2.2. Let X be a cadlag process with natural filtration $\{\mathcal{F}_t\}$. Let $t_0 \in [0,\infty)$ be given. Then the event

$$E := \{\omega \in \Omega : X(\cdot,\omega) \text{ is continuous on } [0,t_0)\},$$

is measurable w.r.t $\mathcal{F}_{t_0}$.

Proof. We will use the notation $\omega(t) = X(t,\omega)$ so that $\omega : \mathbb{R}^+ \to \mathbb{R}$ becomes a function of t. First of all it is clear that since ω is continuous on $[0, t_0)$ then one can write

$$E = \bigcup_k E_k = \bigcup_k \left\{ \omega \in E : \omega \text{ is continuous on } \left[0, t_0 - \frac{1}{k}\right] \right\}.$$

We would like to show that each $E_k \in \mathcal{F}_{t_0}$, for this would prove that $E \in \mathcal{F}_{t_0}$. First for $k, n \in \mathbb{N}$ define

$$S_{kn} := \left\{ (p,q) \in \mathbb{Q}^2 : 0 \leq p, q \leq \frac{1}{k} \text{ and } |p - q| < \frac{1}{n} \right\}.$$

Then for $m \in \mathbb{N}$ and $(p,q) \in S_{nk}$ define

$$E_{mpq} := \left\{ \omega \in \Omega : |\omega(p) - \omega(q)| < \frac{1}{m} \right\}.$$

It is clear that $E_{mpq} \in \mathcal{F}_{t_0}$ since $\mathcal{F}_{t_0}$ contains $\sigma(X_t)$ for all $t \in [0, t_0)$. Now the claim is that

$$E_k = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{(p,q)\in S_{kn}} E_{mpq}.$$

Indeed, let $\omega \in E_k$. Then ω is uniformly continuous on $[0, t_0 - 1/k]$ and therefore for each $m \in \mathbb{N}$, there is an $n \in \mathbb{N}$ such that for all $p, q \in \mathbb{Q} \cap [0, t_0 - 1/k]$ with $|p - q| < 1/n$ implies $|\omega(p) - \omega(q)| < 1/m$. This implies that ω is in the R.H.S of the above equality. On the other hand, if ω is in the R.H.S of the above equality, then one has that $\omega : \mathbb{Q} \cap [0, t_0 - 1/k) \to \mathbb{R}$ is uniformly continuous. Hence, ω extends uniquely to a continuous function $\hat{\omega} : [0, t_0) \to \mathbb{R}$. But since ω is right continuous and agrees with the continuous function $\hat{\omega}$ on a dense set, then $\omega = \hat{\omega}$ and this would imply that $\omega \in E_k$ as desired. Therefore, $E_k \in \mathcal{F}_{t_0}$ and the proof is complete. ■

4.3 Martingales and Stopping Times

Based mainly on [3].

Random, optional, and stopping times.

Definition 4.3.1. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space.

(i) A random time is a random variable $\tau : \Omega \to \overline{\mathbb{R}}$.

(ii) A radom time τ is said to be an optional time if $\{\tau < t\} \in \mathcal{F}_t$ for all t.

(iii) An optional time τ is said to be a stopping time if $\{\tau = t\} \in \mathcal{F}_t$ for all t.

Let X be an adapted process and $\Gamma \in \mathcal{B}(\mathbb{R}^d)$.

(iv) The random time $\tau(\omega) = \inf\{t \geq 0 : X(t, \omega) \in \Gamma\}$, is called a hitting time.

(v) If $X_0 = x \in \Gamma$ then $\tau(\omega) := \inf\{t \geq 0 : X(t, \omega) \notin \Gamma\}$ is called an exit time.

A radom time is usually used to sample randomly from a process X. Indeed, if τ is a *finite* radnom time we define the radom sampling variable $X_\tau : \Omega \to \mathbb{R}^d$ as

$$X_\tau(\omega) = X(\tau(\omega), \omega).$$

It is clearly a radom variable. As for optional time,

Martingales and convergence theorems.

Theorem 4.3.1 (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 4.3.2 (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 4.3.2 (Upcrossing)**.**

Theorem 4.3.3 (Doob's upcrossing inequality)**.**

Theorem 4.3.4 (Martingale convergence theorem, version 1)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probabilty 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 4.3.5 (Martingale convergence theorem, version 2)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probabilty 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

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

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

Optional sampling.

Definition 4.3.3 (Stopped process)**.** A process $\{Y_t\}$ is said to be a stopped process if there is a stochatic process $\{X_t\}$ with stopping time τ such that

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

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

Theorem 4.3.6 (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 4.3.7 (Doob's decomposition Theorem)**.**

4.4 Markov Processes and Feller Semi-Group

Definition 4.4.1. A transition kernel on a measurable space (S, Σ) is a map $N : S \times \Sigma \to \overline{\mathbb{R}^+}$ such that for each $s \in S$, the map $A \mapsto N(s, A)$ is a measure and for each $A \in \mathcal{F}$ the map $s \mapsto N(s, A)$ is measurable. If $N(s, S) = 1$ for all s then N is called a transition probability.

Definition 4.4.2. A collection $\{\mathbb{P}_t\}_{t \geq 0}$ of transition probabilities is called a *homogeneous transition function* if for all $s, t \geq 0$ we have $\mathbb{P}_{t+s} = \mathbb{P}_t \mathbb{P}_s$.

Definition 4.4.3. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ be a filtered probability space. Let X be a stochastic process with state space (S, Σ). The X is called a Markov process if it is adapted and for all $f \in \mathcal{M}(S)$ we have

$$\mathbb{E}[f(X_t) \mid \mathcal{F}_s] = \mathbb{P}_{t-s} f(X_s) = \int_{\mathbb{R}} f(x) \mathbb{P}_{t-s}(X_s, dx).$$

for all $s \leq t$.

4.5 Gaussian Processes, Tempered Measures and White Noise

This section is based on [4, 5, 6, 7].

Definition 4.5.1. 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 $(X(t_1),\dots,X(t_k))$ is a k-dimensional Gaussian random vector vector.

Theorem 4.5.1 ((Some version of) Bochner-Minlos)**.** Let T be any set. Let $m : T \to \mathbb{R}$ be any function and $C : T\times T \to \mathbb{R}$ be any positive definite kernel[(i)] 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 4.5.1.1. Let $(M,\mathcal{G},\sigma)$ be a σ-finite measure space. Define the functions $m : \mathcal{G} \to \mathbb{R}$ and $C : \mathcal{G}\times\mathcal{G} \to \mathbb{R}$ as

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

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

Definition 4.5.2. The Shwartz space of real valued rapidly decreasing smooth functions on $\mathbb{R}^d$ is defined as

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

It is a Fréchét space [(ii)] with topology generated by the countable family of semi-norms

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

[(i)]A function $C : T\times T \to \mathbb{R}$ 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.

[(ii)]If 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 $\|x_n - x\| \to 0$,

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 Fréchét space.

It's dual space $\mathcal{S}'(\mathbb{R}^d)$ is called the space of temepered distributions and is equipped with the weak* topology.[(iii)]. 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

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

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

Theorem 4.5.2 (Bochner-Minlos). Let $\mathcal{S}' := \mathcal{S}'(\mathbb{R}^d)$. There is a probability measure $\mathbb{P}$ on $\left(\mathcal{S}', \mathcal{B}(\mathcal{S}')\right)$ 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 4.5.3. Let $\varphi_1, \ldots, \varphi_n \in \mathcal{S}(\mathbb{R}^d)$ be n orthonormal functions[(iv)] in $L^2(\mathbb{R}^d)$ and let μ_n be the measure on $\mathbb{R}^n$ defined as

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

Then the random vector

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

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

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

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 equality[(v)]

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

Therefore we have that

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

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

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

are continuous

(iv) Meaning that we have

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

(v) This 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

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

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

Definition 4.5.3 (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 $(\Omega, \mathcal{F})$ obtained from the Bochner-Minlos theorem. The Gaussian process $W : T \times \Omega \to \mathbb{R}$ defined as

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

is called a white noise process.

Definition 4.5.4 (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 $W_\phi : \mathbb{R}^d \times \mathcal{S}'(\mathbb{R}^d) \to \mathbb{R}$ 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 defintion.

Proposition 4.5.4. 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 $\|\varphi\|_{L^2(\mathbb{R}^d)}$.

(ii) If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ are chosen to that $\operatorname{supp} \varphi_{\mathbf{x}} \cap \operatorname{supp} \varphi_{\mathbf{y}} = \emptyset$ then $W_\varphi(\mathbf{x})$ and $W_\varphi(\mathbf{y})$ are independent.

(iii) For any $h \in \mathbb{R}^d$ and all $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ we have that

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

Theorem 4.5.5 (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

---

(vi) We 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) $\mathbb{E}[\exp(iW_\varphi^{(\sigma)})] = \exp\Bigg(-\dfrac{1}{2} \displaystyle\int_{\mathbb{R}} |\hat{\varphi}(u)|^2 d\sigma(u) \Bigg)$.

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

Corollary 4.5.5.1.

4.6 Weiner Process

Definition 4.6.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $T = \mathbb{R}^+$. A Weiner process with starting point $\mathbf{x} \in \mathbb{R}^n$ is a Gaussian stochastic process $X : T \times \Omega \to \mathbb{R}^n$ such that

(i) $X_0 = \mathbf{x}$ and $\mathbb{E}[X_t] = \mathbf{x}$ for all $t \in T$.

(ii) $\mathrm{Cov}(X_s - \mathbf{x}, X_t - \mathbf{x}) = \mathbb{E}[(X_s - \mathbf{x}) \cdot (X_t - \mathbf{x})] = n \min(s,t)$ 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 4.6.1. 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

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

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 $F_1, \ldots, F_k \in \mathcal{B}(\mathbb{R}^n)$. Futheremore, these measures satisfy the consistency conditions for the Kolomogorov extension thoerem. ■

Proposition 4.6.2. For any $0 \leq t_1 < \cdots < t_n \leq T$ we have that

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

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

Proof. ■

Proposition 4.6.3. There is a modification $\{B_t\}$ of the stochastic process $\{W_t\}$ obtained in Proposition 4.6.1. 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 4.6.4 (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 $(B_{t_1}, \ldots, B_{t_k})$ is multi-normal on $\mathbb{R}^{nk}$.
2. There is a continuous version of $\{B_t\}$.
3. Each component of $B_t = (B_t^{(1)}, \ldots, B_t^{(n)})$ is standard Brownian motion on $\mathbb{R}$.
4. $\langle B, B\rangle_t = \lim_{n\to\infty} \sum_{t_i \in \Pi_n} |B_{t_i} - B_{t_{i-1}}|^2 = t$ 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 $c \in \mathbb{R}$, $\{c^{-1}B_{c^2 t}\}$ 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 $\int_0^t B_s dB_s = \frac{1}{2}B_t - t$.
10. For all $m > 0$ and $w \leq m$ we have $\mathbb{P}\{\tau_m \leq t,\, B_t \leq w\} = \mathbb{P}\{B_t \geq 2m - w\}$.(viii)
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 of $\{M_t\}$ and $\{B_t\}$

$$f_{M_t,B_t}(m,w) = 2(2m-w)(2\pi)^{-1/2}t^{-3/2}\exp(-(2m-w)^2/2t).$$

Brownian motion as a limit of random walks.

Brownian motion as a special case of white noise.

Proposition 4.6.5. 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 $f_n : \mathcal{S}' \to \mathbb{R}$ 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 $\{\varphi_n\}$ converging to φ.

Definition 4.6.2 (d-parameter Brownian motion)**.** The stochastic process $B : \mathbb{R}^d \times \mathcal{S}'(\mathbb{R}^d) \to \mathbb{R}$ defined by

$$B(\mathbf{x}, \omega) := \langle \omega, \mathbf{1}_{[0,x_1],\ldots,[0,x_d]}\rangle, \qquad \mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d, \;\; \omega \in \mathcal{S}'(\mathbb{R}^d),$$

is called the d-parameter Brownian motion in dimension one.

Representing solutions to elliptic PDEs using Brownian motion.

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

Lemma 4.6.6. Suppose that $f \in C_c^2(\mathbb{R})$. Then

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

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

Proof. We have that

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

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

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

as desired. ■

Lemma 4.6.7. 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 Brwonian 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 $m_x:T\to\mathbb{R}$ as

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

We have that by iterated conditioning that

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

First we notice that

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

and then

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

Therefore by using dominated convergence we obtain

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

We can work in a similar fashion to obtain that

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

Therefore $m_x'(t)$ is well defined for all t. Hence we can conlude 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 4.6.8. Consider the boundary value problem

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

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

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

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

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

By the above lemma, M_t is a martingale. Therefore we can apply optional sampling (Theorem 4.3.6) 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 4.6.9. 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 a dimenstional 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

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

4.7 Lévy and Jump Processes

Definition 4.7.1. 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 $\lim_{s \to t} \mathbb{P}(|X(t) - X(s)| > \epsilon) = 0$.

Theorem 4.7.1 (Lévy-Khintchine)**.**

Theorem 4.7.2. Any Lévy process X then admits a cadlag modification.

Proof. ∎

Theorem 4.7.3 (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. ∎

4.8 Point Processes*

Definition 4.8.1. Let E be separable Banach space equipped with the Borel σ-algebra $\mathcal{B}$. We define the set of all *locally finite point configurations* S as

$$S := \{F \subset E : F \cap B \text{ is finite for every bounded set } B \subset E\},$$

equipped with with the σ-algebra

$$\Sigma_S = \sigma\big(\{F \in S : F \cap B = m\};\ B \in \mathcal{B}_0,\ m \in \mathbb{N}\big),$$

where $\mathcal{B}_0$ is the collection of all bounded Borel sets.

Definition 4.8.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and E be a separable Banach. Let (S, Σ_S) be the space of all locally finite point configurations of E. A point process is a random variable $X : \Omega \to S$.

Proposition 4.8.1. A point process X is measurable if and only the function $N_B : \Omega \to \mathbb{Z}$ defined as $N_B(\omega) := \#X(\omega) \cap B$ is measurable for every $B \in \mathcal{B}_0$.

Chapter 5

Itô Calculus and Elementary Stochastic Differential Equations

This chapter is heavily inspired by the textbooks [1, 2, 3, 4].

Bibliography

[1] S.E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Number v. 11 in Springer Finance Textbooks. Springer, 2004.

[2] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Hochschultext / Universitext. Springer, 2003.

[3] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2004.

[4] I. Karatzas, I.K.S. Shreve, S. Shreve, and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics (113) (Book 113). Springer New York, 1991.

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

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

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 expected to be a stochastic process, and so $I(t)$ is a radom variable. This idea will provide a work around the limitation of unbounded variation: we can define I at each time t instead of each path ω. More specifically, we will define $I(t)$ as an element in $L^2(\mathbb{P})$ for each t and hope to establish the desired properties of regular integration.

Definition 5.1.1 (Simple/Elementary processes)**.** A process $X = X(\omega, t)$ is called simple for each $\omega \in \Omega$, there is a sequence of postitive 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 theortical 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 5.1.2.

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

Definition 5.1.2. 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 5.1.1 (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 satsfies some integration and measurability properties then X is indeed the limit of simple processes in an appropriate sense.

Definition 5.1.3. 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 5.1.2. Let $I = [a, b] \subset \mathbb{R}^+$ be any interval (possibly unbounded) and et $X \in \mathcal{V}(I)$. Then there is a sequence of elementary processes $\{\Delta_n\} \subset \mathcal{V}(I)$ such that

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

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

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

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

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

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

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

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

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

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 indentity[(i)] we have that for each ω

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

and therefore by dominated convergence

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

(i)

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

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

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

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

Now one can easliy 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). ■

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

Definition 5.1.4. 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 5.1.2. 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 $\left\{\int_a^b \Delta_n dB_t\right\}$ is a Cauchy sequence in $L^2(\Omega)$. We define the Itô integral of $\{X_t\}$ as

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

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

Lemma 5.1.3. The definition of $\hat{I}$ 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

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

Lemma 5.1.4. The Itô integral $\hat{I}(\omega,t)$ of a simple process $\Delta(\omega,t)$ is a continuous Martingale.

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

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

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

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

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

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

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

Now $\hat{I}$ 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

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

as desired. ■

Theorem 5.1.5 (Properties)**.** The Itô integral $\hat{I}(\omega,t)$ 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 $t \ge 0$, $\hat{I}(t,\cdot)$ is $\mathcal{F}_t$-measurable.

(ii) $\hat{I}$ satisfies the Itô isomerty.

(iii) $\hat{I}$ is a Martingale.

(iv) Almost all paths $\hat{I}(\cdot,\omega)$ can be chosen to be continuous.

(v) The quadratic variation of the Itô integral is given by

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

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

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

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

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

Hence for almost all $\omega \in \Omega$, the sequence of continuous functions $\{\hat{I}_{n_k}(\omega,\cdot)\}_{k\in\mathbb{N}}$ 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 $\hat{I}(\omega,\cdot)$.

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

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

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

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

so that by the Borel-Cantelli lemma

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

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

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

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

Weakening defining conditions for Itô integral

Definition 5.1.5. 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) $X : I \times \Omega \to \mathbb{R}$ 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

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

Theorem 5.1.6. 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.

5.2 Itô process and Itô-Doeblin Formula

Definition 5.2.1 (Itô Process)**.** Let $\{\mathcal{F}_t\}$ be the natural filtration for Brownian motion. Suppose there are adapted processes $\alpha, \sigma : [0, \infty) \times \Omega \to \mathbb{R}$ with $\sigma \in \mathcal{V}$ such that and a stochastic process X such that

$$X(t, \omega) = X(0, \omega) + \int_0^t \alpha(s, \omega) ds + \int_0^t \sigma(s, \omega) dB_s(\omega). \tag{5.1}$$

or for shorthand

$$dX_t = \alpha dt + \sigma dB_t.$$

Then X is called an Itô process.

Proposition 5.2.1 (Quadratic variation of Itô process)**.** If X is an Itô process then

$$[X, X](t, \omega) = \int_0^t \sigma(s, \omega)^2 ds.$$

Proof. ■

Thus $[X, X]$ is continuous and increasing each ω and therefore we can properly define for each ω the integral of a function f with respect to $[X, X]$ as

$$\int_0^t f(s, \omega) d[X, X](s, \omega) = \int_0^t f(s, \omega)\sigma(s, \omega)^2 ds.$$

This allows to formally (and correctly) replace $d[X, X]_t$ by $\sigma^2 dt$ in integrals. Note that the above proof also justifies the following replacements

$$(dt)^2 = dt\, dB_t = dB_t\, dt = 0 \quad \text{and} \quad (dB_t)^2 = dt.$$

Definition 5.2.2 (Itô integral w.r.t Itô process)**.** Let $X, Y : [0, \infty) \times \Omega \to \mathbb{R}$ be two processes such that X an Itô process. We define

$$\int_0^t Y(s, \omega) dX_s(\omega) := \int_0^t Y(s, \omega)\alpha(s, \omega) dt + \int_0^t Y(s, \omega)\sigma(s, \omega) dB_s(\omega) \tag{5.2}$$

Of course, this assumes that $\sigma Y \in \mathcal{V}(\mathbb{R}^+)$ and that for all $\omega \in \Omega$ we have $\alpha(\cdot, \omega)Y(\cdot, \omega) \in L^1([0, t])$ for all $t \in \mathbb{R}^+$ so that the above integrals are well defined.

Theorem 5.2.2 (Itô-Doeblin Formula)**.** Let X be an Itô process and $g \in C^2([0, \infty) \times \mathbb{R})$. Define the stochastic process Y as

$$Y(t, \omega) := f(t, X(t, \omega)).$$

Then Y is also an Itô process and for all $t \geq 0$ we have

$$Y_t = Y_0 + \int_0^t \frac{\partial g}{\partial t}(u, X_u) du + \int_0^t f_x(u, X_u) dX_u + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(u, X_u) d[X, X]_u, \tag{5.3}$$

with the usual shorthand

$$dY_t = \partial_t f(t, X_t) dt + \partial_x f(t, X_t) dX_t + \partial_{xx} f(t, X_t) d[X, X](t).$$

5.3 SDE's and the Markov property

Suppose that one is given an Itô process X satisfying the following equation

$$dX_t = \alpha X_t dt + \sigma X_t dB_t. \tag{5.4}$$

with $X_0 = x_0$, where $\alpha, \sigma, x_0 \in \mathbb{R}$ are constants. Applying the Itô-Doeblin formula (5.3) to the process $Y_t = \ln X_t$, one obtains

$$\ln X_t - \ln X_0 = \int_0^t \frac{1}{X_t} dX_t - \frac{1}{2}\sigma^2 t.$$

By the definition of Itô integral (5.2) with respect to Itô processes one has that

$$\int_0^t \frac{1}{X_t} dX_t = \int_0^t \alpha X_t \frac{1}{X_t} dt + \int_0^t \sigma \frac{1}{X_t} X_t dB_t = \alpha t + \sigma B_t.$$

Therefore one obtains that

$$X_t = x_0 \exp\left((\alpha - \sigma^2/2)t + \sigma B_t\right). \tag{5.5}$$

This process is called called *geometric Brownian motion.* Equation (5.4) is called a stochastic differential equations, because X_t is written as a sum of regular integral and an Itô integral, both of which having as integrands a funtion of X_t.

Definition 5.3.1. Let $I \subset \mathbb{R}^+$ be any closed interval (possibly unbounded). Let $\mu, \sigma : I \times \mathbb{R} \to \mathbb{R}$ be Borel functions and $\{W_t\}$ be Brownian motion. A first order linear stochastic differential equation is a relation of the form

$$dX(t) = \mu(t, X(t))du + \sigma(t, X(t))dW(t), \;\; t \in I, \tag{5.6}$$

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

$$X(t) = X(t_0) + \int_{t_0}^t \mu(u, X(u))du + \int_{t_0}^t \sigma(u, X(u))dW(u), \quad t \in I. \tag{5.7}$$

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

Definition 5.3.2 (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 radom variable. A progressively measurable process $\{X_t\}$ is a strong solution with initial condition $X(0) = \xi^*$ if (5.6) holds almost surely.

Definition 5.3.3 (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 (5.6) holds almost surely for all $t \geq 0$.

Proposition 5.3.1 (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 (5.6) then X and Y are indistinguishable.

Proof. Since both X and Y satisfy (5.6) 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 thererfore X and Y are indistinguishable. ■

Theorem 5.3.2 (Uniquness of weak solutions)**.** Suppose that μ and σ are Lipschitz functions and that X and Y be two weak soltuions to (5.6) such that X_0 and Y_0 induce the same probabiliy law $\mu = \mathbb{P}^{X_0} = \mathbb{P}^{Y_0}$. Then X and Y have the same finite dimensional distributions.

Theorem 5.3.3 (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 $\mu, \sigma : I \times \mathbb{R} \to \mathbb{R}$ and constants $C, D \in \mathbb{R}$ such that that for all $x, y \in \mathbb{R}$ and $t \in [0, T]$ we have

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

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

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

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

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

Under these assumption, equation (5.6) has a solution a strong solution $\{X_t\}$ that is adapted to the filtration $\{\mathcal{F}_t^\xi\}$ and $\|X^2\|_{L^1(I)} \in L^1(\Omega)$[(iii)].

Definition 5.3.4 (Markov property)**.** Let $I = [t, T]$ and $x \in \mathbb{R}$ be given. Suppose X is a stochastic process that solves (5.6) with initial condition $X(t) = x$. Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function. We define

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

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

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

[(iii)] This means that

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

5.4 Feynman-Kac and Fokker-Planck Equations

Theorem 5.4.1 (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 5.4.2 (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

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

Then the function g satisfies the following

$$\begin{cases} \dfrac{\partial g}{\partial t} = -\beta\dfrac{\partial g}{\partial x} - \dfrac{1}{2}\gamma^2\dfrac{\partial^2 g}{\partial x^2}, & \text{for all } (x,t) \in \mathbb{R} \times [0,T], \\ g(x,T) = h(x), & \text{for all } x \in \mathbb{R}. \end{cases} \tag{5.8}$$

5.5 Examples from finance and economics

Chapter 6

General Stochastic Integration

Bibliography

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

Chapter 7

Classical theory and fundamental equations

Bibliography

7.1 Fundamental existence theorems

7.2 Poisson equation

7.3 Diffusion equation

7.4 Wave equation

Chapter 8

Hilbert Spaces

Bibliography

8.1 Elementary properties

8.2 Lax-Milgram

8.3 Reproducing kernel Hilbert spaces

Chapter 9

Sobolev Spaces

Bibliography

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

Chapter 10

Stochastic Heat Equation

Bibliography

Chapter 11

Stochastic Wave Equation

Bibliography

Chapter 12

The Classical and Stochastic Hasegawa-Mima equation

Bibliography

Chapter 13

Stochastic Integration in UMD spaces

Bibliography

Chapter 14

Miscellaneous Remarks and Observations

Bibliography

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

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

which goes to zero since B_t is continuous. Second,

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

which follows from the inequality

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

If $\Pi_n = \{0, t/n, 2t/n, \ldots, t\}$ 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 $\overline{\mu} = 1$ that

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

Remark on definition of Itô integral.