A Survey of the Mathematical Foundations of Continuous-Time Finance

by

Peter Umberto Vinella

A thesis submitted in partial satisfaction of the
requirements for the degree of
Masters of Arts
in
Mathematics
in the
Graduate Division
of the
University of California, Berkeley

Committee in Charge:

Professor Lawrence C. Evans, Chair
Professor Alexandre Chorin
Professor Xin Guo

Fall 2015

Contents

Preface

This paper is a survey exploring the mathematics underlying continuous-time finance. Although the earliest results in continuous-time finance were derived using classical analysis, for the most part, I have chosen to also present them in a probabilistic setting to be consistent with current preferences. In the former setting, the price of financial instrument is given as the solution to a particular partial differential equation (PDE). In the latter case, the price of a financial instrument is the expressed as continuous-time stochastic process under some *risk-neutral* probability measure. Here, I will assume these processes are driven by Brownian motion which is generally sufficient for the finance theory under consideration.

Additionally, I have generally avoided presenting mathematical results in their most general setting to avoid technical issues that often obscure the heuristics. However, many of the results presented in the paper can be extended to a more general setting. Jeanblanc, Yor, and Chesney [**47**] and Shreve [**76**] are particularly good references for a discussion of this topic in the more general setting.

For the most part, I have chosen to present most of the results in a one-dimensional setting. While this is also consistent with most of the references cited herein, many of the results can be easily extended to higher dimensions. That disclaimer notwithstanding, in a few occasions, I state the result in higher dimensions and, in some cases, note that a result cannot be extended to higher dimensions or that such a generalization is non-trivial.

While discrete-time models are widely used in practice, I have chosen to limit the scope of this paper primarily to continuous-time processes for two basic reasons. First, the mathematics is richer and more interesting. Second, while many markets do not trade continuously, it is common industry practice to assume that, in fact, prices change continuously in time even during non-trading hours. However, I do make reference to discrete-time processes from time-to-time, primarily as motivation for the continuous case.

Being a survey, I do not generally provide rigorous proofs or derivations for results which are overly long or technical. Rather, I give an outline of such a proof or simply supply references to appropriate sources. And while I hope that I have presented some original insights, I do not claim to offer any original results. In this sense, my hope is that paper will serve as good reference to anyone interested in the mathematical justifications behind continuous-time finance in general.

I have generally followed Bingham and Kiesle [**9**], Jeanblanc, Yor, and Chesney [**47**], Pascucci [**66**], and Shreve [**76**] for results from mathematical finance. In terms of probability theory generally, I have relied primarily Billingsley [**8**], Çinlar [**14**], Durrett [**23**], Evans [**30**], Feller [**32**], and Kallenburg [**50**]. As for PDE-related results and analysis generally, I have relied on Evans [**29, 28**], Korn and

Korn [**54**], Pascucci [**66**], Wilmott [**89, 90**], and Yosida [**91**]. In terms of finance theory, I have generally relied on Duffie [**22**], Elton and Gruber [**25**], Ingersoll [**44**], Neftci [**63**], Samuelson [**72**], and Sharpe [**74**]. I have also provided references to specific topics and results in addition to a general bibliography.

Lastly, a quick word about notation and terminology. Unfortunately, as an amalgam of finance, analysis, and probability, there is no widely accepted standard for this subject. Given this, I have tried to follow a standard throughout the paper that is consistent with mathematical analysis. However, the notation and terminology used herein may not conform to that presented in original papers or other references.

Introduction

Although partial differential equations (PDE) played a central role in founding of continuous-time finance, with the advent of risk-neutral valuation based on martingale theory, much of the subsequent work has relied heavily on measure-theoretic probability theory. A great deal of this involves the analysis of relatively well-behaved, continuous-time stochastic processes known as *Itô processes* (*i.e.* continuous-time and continuous-price processes). Generally, such processes are used to describe the price dynamics of various financial instruments. Additionally, most of the common price processes are described as solutions to *stochastic differential equations* (SDE). As such, much of the mathematical foundation of continuous-time finance involves the extension of classical calculus to stochastic processes (in other words, *stochastic calculus*).

In this paper, I am primarily concerned with Itô processes and much of the mathematics presented herein involves results necessary to (a) work with Itô processes; (b) rigorously establish stochastic calculus with respect to Itô processes; and (c) subsequently develop theoretical foundation for rigorous solutions to SDE (*i.e.* Itô diffusions). I have also relied heavily on the well-known connection between a particular class of PDE and Itô diffusions. In fact, this was the approached used by Kolmogorov (*e.g.* [**75**]) and Feller (*e.g.* [**31**]) when such processes were first studied in earnest (although Bachelier stated some of the same key results in his landmark dissertation which laid the foundation for modern continuous-time finance (*cf.* [**3**, Ch. 2])).

In particular, I have followed the path of describing the evolution of price processes using *transition probability functions* which ultimately results in the *Feynman-Kac formula* in which the expected price of a financial instrument is expressed as the solution of a second-order PDE. I have also followed the modern probabilistic approach to continuous-time finance by emphasizing the use of *Brownian motion* to model price dynamics. Brownian motion possesses two key properties making it an indispensable tool in mathematical finance - it is *non-anticipative* and *memoryless*. Further, Brownian motion explicitly captures the fact that the information observable by investors (as captured by its *standard filtration*) evolves over time.

CHAPTER 1

General Probability Theory

At first glance, the price of a financial instrument appears to follow a random path that evolves continuously over time. Additionally, it is natural to think that these changes in price are due to information that becomes available to investors with the passage of time. In other words, information evolves over time and investors react to it. Thus, we need some rigorous framework in which the intuitive concepts of randomness, continuous-time paths (possibly with jumps), and information can be rigorously define. Measure-theoretic probability theory provides such a framework.

In fact, many of the following results presented herein hold for general measure spaces and are often presented as such in the literature. However, given the focus this paper, I have chosen to develop the material generally in the context of probability spaces. For a more general approach, see Billingsley **[8]**, Çinlar **[14]**, or Kallenberg **[50]**.

1.1. Probability Spaces

I shall begin with the basic concept of experiments, outcomes, and observable events.

1.1.1. Measurable Spaces. Assume that some experiment has been designed and Ω is the set of all possible outcomes. For instance, if the experiment consists of a tossing a single coin, then Ω can be thought of as the two faces of the coin where each face (*i.e.* ω_1 and ω_2) is an outcome (*i.e.* $\omega_1, \omega_2 \in \Omega$). If this experiment is perform over and over again a finite number of times, say n, the set of outcomes can be thought of as vectors $\omega = \omega_1, \ldots, \omega_n$ where $\omega \in \Omega^n := \Omega \times \cdots \times \Omega$. If the experiment is performed a countably infinite number of times, the sequence of outcomes can be thought of as $\Omega_\infty := \lim_{n\to\infty} \Omega^n$. On the other hand, the experiment might consist of picking some point at random in the interval $[0, 1]$. In other words, the set of outcomes can be countably finite, countably infinite, or uncountable.

Now, assume that the experiment consists of measuring the outside temperature. Since any thermometer has only limited accuracy, we can only *observe* the temperature based on the accuracy of the thermometer. In other words, we can only record the outcome of the experiment as measured by that particular thermometer and not the *actual* temperature itself.

These heuristics are captured in the following definition which mathematically captures the concept of *information*.

DEFINITION. Let Ω be an arbitrary, non-empty set. Then the set $\mathcal{F}$ consisting of subsets of Ω is called a σ-***algebra*** if

(1) $\emptyset, \Omega \in \mathcal{F}$;
(2) $A^c \in \mathcal{F}$ for all $A \in \mathcal{F}$ (*closed under complements*); and
(3) $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ with $A_k \in \mathcal{F}$ for all k (*closed under countable unions*).

THEOREM 1.1. *A σ-algebra $\mathcal{F}$ is closed under countable intersections, or*

$$\bigcap_{k=1}^{\infty} A_k \in \mathcal{F} \text{ with } A_k \in \mathcal{F} \text{ for all } k = 1, \ldots$$

PROOF. The proof follows directly from property (2) and DeMorgan's laws. □

DEFINITION. Given an arbitrary, non-empty set Ω and a σ-algebra $\mathcal{F}$ on Ω, the pair $(\Omega, \mathcal{F})$ is called a ***measurable space***. Ω is called the ***sample space*** and any $\omega \in \Omega$ is called a ***sample point*** or ***outcome***. Similarly, $\mathcal{F}$ is called the ***family of events*** and any $F \in \mathcal{F}$ is called an ***event***.

EXAMPLE. The ***Borel σ-algebra***, denoted as $\mathcal{B}$, is the smallest σ-algebra that contains all open intervals in $\mathbb{R}$. Its elements are called ***Borel sets***. The pair $(\mathbb{R}, \mathcal{B})$ forms a measurable space. □

DEFINITION. Given the measurable space $(\Omega, \mathcal{F})$, if Ω is countable, the space is said to be ***discrete***. Else, it is said to be ***infinite***.

DEFINITION. An event $A \in \mathcal{F}$ is considered to have ***occurred*** if an outcome $\omega \in A$ has occurred.

DEFINITION. An event $A \in \mathcal{F}$ is said to be ***indivisible*** if there are no non-empty sets $B, C \in \mathcal{F}$ such that $A = B \cup C$.

Clearly, the existence of a σ-algebra is not in doubt since every arbitrary non-empty set has at least two σ-algebras. The first is the ***trivial σ-algebra*** consisting of $\{\emptyset, \Omega\}$, which essentially contains no information (*i.e.* some outcome has occurred or no outcomes have), and the ***largest σ-algebra***, which is the collection of all subsets and contains all possible information (*i.e.* all outcomes are observable).

REMARK. Generally, the largest σ-algebra is much too large to be of any use. However, in the case of a countably finite sample space, the largest σ-algebra is known as the ***discrete σ-algebra*** and is denoted as 2^{Ω}.

□

As the names imply, the trivial σ-algebra is *smaller* than the largest σ-algebra since both its elements are contained in the latter. This notion of the relative size of two σ-algebras is captured in the following definition.

DEFINITION. Given two σ-algebras $\mathcal{F}$ and $\mathcal{E}$ defined on the same set Ω, $\mathcal{F}$ is said to be ***smaller*** than $\mathcal{E}$ if $A \in \mathcal{E}$ for each $A \in \mathcal{F}$. In this case, $\mathcal{F}$ is also said to be a ***sub-σ-algebra*** of $\mathcal{E}$ denoted as $\mathcal{F} \subseteq \mathcal{E}$.

Essentially, a σ-algebra represents all information about outcomes that are observable. A smaller σ-algebra contains fewer events and, consequently, contains less observable information about the state of the sample space while larger ones contain more. For instance, after conducting an experiment, only the events containing a given outcome can be observed and not the outcome itself. The greater the number of events in the σ-algebra containing that outcome, the more information we have about it. If the σ-algebra consists of all the events that contains that outcome, it will contain all information possible regarding that given event. This is captured in the next result.

THEOREM 1.2. *Given some set Ω and any set $A \subseteq \Omega$, let $\sigma(A)$ denote the intersection of all σ-algebras containing A. Then $\sigma(A)$ is a σ-algebra on Ω and there is only one such σ-algebra. Such a σ-algebra is called the* **smallest σ-algebra** *generated by A.*

PROOF. Let $A \subseteq \Omega$ and assume the set $\sigma(A)$ is the intersection of all σ-algebras containing A. Further, let $\mathcal{G}$ be any σ-algebra such that $A \in \mathcal{G}$. Clearly, Ω and $\emptyset \in \sigma(A)$ since they are contained in every such $\mathcal{G}$. Additionally, for any $B \in \sigma(A)$, $B \in \mathcal{G}$ since $\sigma(A)$ is the intersection of all such $\mathcal{G}$. By assumption, $\mathcal{G}$ is a σ-algebra, thus $B^c \in \mathcal{G}$ and $\bigcup_{i=1}^{\infty} B_i \in \mathcal{G}$ where $B_i \in \mathcal{G}$ for $i = 1, \ldots$. Since it is true for any such $\mathcal{G}$, it must be true all their intersection. Thus, $\sigma(A)$ is closed under complements and countable unions and, thus, it is a σ-algebra. It clearly is unique and the smallest σ-algebra by since it is the intersection of all σ-algebras containing A. □

REMARK. Based on the above theorem, the smallest σ-algebra generated by a given event contains all information about that event. If the σ-algebra is known, then the event is known and *vice versa*. Such a σ-algebra will play an important part in the concept of *conditional expectation* introduced later in the paper. □

In the context of continuous-time finance, the sample space represents the total breadth of information in a given market. However, not all information may be observable to investors at any point in time. Conversely, the σ-algebra represents observable information available to investors who subsequently determine the price level of a given instrument based on this information. However, since all information may not be observable, such prices may not accurately reflect the *true* price of the instrument (a result in finance known as the *efficient market theorem* expresses this phenomenon in economics terms (*cf.* [**25**, Ch. 17])).

1.1.2. Probability Measures. Generally, probability theory is concerned with measuring the likelihood that a given event will occur. This is captured in the following definition.

DEFINITION. Given a σ-algebra $\mathcal{F}$ on the set Ω, a mapping $\mathbb{P} : \mathcal{F} \to [0, 1]$ is called a ***probability measure*** if

(1) $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$ (*i.e.* a *finite measure*);
(2) $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$ (*non-negativity*); and
(3) $\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$ $\quad (A_k \in \mathcal{F},$ such that $A_i \cap A_j = \emptyset$ for $i \neq j)$ (*countable additivity*).

For the proof of the existence of such a measure and relevant properties of measures generally, see Kallenberg [**50**, pgs. 23-25].

REMARK. Any non-negative, finite measure that is countably subadditive can be transformed into probability measure by scaling (*i.e.* for such measure $\mu(\cdot)$, let $\mathbb{P}(\cdot) := \frac{1}{\mu(\Omega)}\mu(\cdot)$). □

The following result follows directly from the definitions.

THEOREM 1.3. *Let $\mathcal{F}$ be σ-algebra on the set Ω and that $\mathbb{P}$ is a probability measure. Further, assume that $\{A_k\}$ is a countable set of events in $\mathcal{F}$. Then $\mathbb{P}$ is* **countably subadditive** *or*

$$\mathbb{P}\left(\bigcup_j A_j\right) \leq \sum_j \mathbb{P}(A_j) \quad (\textit{for } A_j \in \{A_k\}).$$

PROOF. Let $\mathcal{F}$ be σ-algebra on the set Ω and assume that $\mathbb{P}$ is a probability measure. Since $\mathbb{P}$ is a probability measure then it is countably additive. Therefore, assume that A_1 and A_2 are any two non-empty events in $\mathcal{F}$ such that $A_1 \cap A_2 \neq \emptyset$. Now, let $B := A_1 \cap A_2$ and $C := A_1^c \cap A_2$. Then A_1 and C are disjoint sets. Hence, by countable additivity

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1 \cup C) = \mathbb{P}(A_1) + \mathbb{P}(C).$$

On the other hand, B and C are disjoint sets since A_1 and A_2 are non-empty. Thus

$$\mathbb{P}(B \cup C) = \mathbb{P}(B) + \mathbb{P}(C).$$

But, $B \cup C = A_2$. Therefore,

$$\mathbb{P}(A_2) = \mathbb{P}(B) + \mathbb{P}(C).$$

By assumption, $B \neq \emptyset$ and $\mathbb{P}(B) > 0$. Hence,

$$\mathbb{P}(A_2) > \mathbb{P}(C)$$

and

$$\mathbb{P}(A_1 \cup A_2) < \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

Again, the general result follows from induction. □

DEFINITION. Given a measurable space $(\Omega, \mathcal{F})$, the ***probability of an event*** A ***occurring*** (*a.k.a.* the ***probability mass of*** A) is denoted by

$$\mathbb{P}(A) := \mathbb{P}(\omega \in A).$$

REMARK. There is a slight abuse of notation here. There is an implicit assumption that there is some underlying selection process which causes an outcome to occur. Generally, this is viewed as some experiment which results in certain outcomes to become *true* and others to become *false*. Then the probability of event A occurring is the probability that *following* some experiment, an outcome contained in A will occur. □

EXAMPLE (Dirac measure). The mapping $\delta_\omega(\cdot)$ on the measurable space $(\Omega, \mathcal{F})$ defined as

$$\delta_\omega(A) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

is called the ***Dirac measure*** (*a.k.a.* ***Dirac point mass***) and expresses the fact that the total probability mass is located at ω (*i.e.* $\mathbb{P}(\omega) = 1$). □

REMARK. In the context of continuous-time finance, the probability that a specific price level will occur is based on the probability that observable information admitting such a price will occur. □

The next theorem follows directly from the definition and its proof is given in Durrett [**23**, pg. 2].

THEOREM 1.4. *Let $\mathbb{P}$ be a probability measure on the measurable space $(\Omega, \mathcal{F})$. Then*

(1) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *for all* $A \in \mathcal{F}$;
(2) $\mathbb{P}(A) \leq \mathbb{P}(B)$ *for all* $A, B \in \mathcal{F}$ *with* $A \subseteq B$ (*monotonicity*); and
(3) $\mathbb{P}(A_k) \uparrow \mathbb{P}(A)$ *if* $A_k \uparrow A$ *with* $A, A_k \in \mathcal{F}$ *for all* k. (*sequential continuity*).

1.1.3. Probability Spaces. For much of the sequel, I will be interested in a special class of measure spaces known as *probability spaces*.

DEFINITION. A ***probability space*** is the triplet $(\Omega, \mathcal{F}, \mathbb{P})$ where $(\Omega, \mathcal{F})$ is a measurable space and $\mathbb{P}$ is a probability measure on $\mathcal{F}$.

REMARK. A measurable space generally admits numerous probability measures. Often these are determined from empirical observation. However, in the context of continuous-time finance, there is a canonical measure, often called the *risk-neutral measure*, which expresses the price level of a given instrument as the *present value* of its expected cash flows under that measure (*cf.* Section 4.6). □

EXAMPLE. $((0,1), \mathcal{B}(0,1), \mathcal{L})$ is probability space where $\mathcal{B}(0,1)$ is the collection of Borel sets on $(0,1)$ and $\mathcal{L}$ is the Lebesgue measure. □

The following definition helps characterizes *discrete* probability spaces.

DEFINITION. Given a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, an outcome $\omega \in \Omega$ is said to be an ***atom*** (*a.k.a.* ***point mass***) of the probability measure $\mathbb{P}$ if it is an event (*i.e.* $\{\omega\} \in \mathcal{F}$) and $\mathbb{P}\{\omega\} > 0$. $\mathbb{P}$ is said to be ***diffuse*** on the set $\Gamma := \{\omega \in \Omega \mid \{\omega\} \in \mathcal{F} \text{ and } \mathbb{P}(\{\omega\}) = 0\}$. If $\Gamma = \emptyset$ (*i.e.* there are only atoms), then $\mathbb{P}$ is said to be ***purely atomic*** or ***discrete***. If $\Gamma = \Omega$ (*i.e.* there are no atoms), then $\mathbb{P}$ is said to be ***diffuse*** or ***continuous***.

REMARK. In other words, a probability measure is continuous if has no atoms. □

EXAMPLE. The Dirac measure is purely atomic and Lebesgue measure is diffuse. □

Motivation. In the sequel, I will frequently rely on the property *almost surely* which holds that a given proposition is true except on sets with probability zero (*cf.* on page 17). However, there is no guarantee that all sets of measure zero are observable. For instance, assume that $F \in \mathcal{F}$ for some σ-algebra $\mathcal{F}$ and that $\mathbb{P}(F) = 0$. Further assume that $A \subset \Omega$. It is quite possible that $A \subset F$, but $A \notin \mathcal{F}$. Hence, A is not observable and it is not clear how to apply "almost surely" to such sets. Fortunately, the following results show that it is generally necessary to only consider probability spaces in which all sets with probability zero are observable. □

DEFINITION. Given a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, any arbitrary subset of an event with probability zero is called a $\mathbb{P}$-***null set*** or simply a ***null set*** (*a.k.a.* a ***negligible set***). The probability space is said to be ***complete*** if all null sets are in $\mathcal{F}$ (*i.e.*).

REMARK. In other words, all impossible and all certain outcomes are observable (*i.e.* those with probability zero and probability one, respectively)

THEOREM 1.5. *Any probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *can be uniquely extended to a complete probability space* $\left(\Omega, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}\right)$ *such that*

$$\tilde{\mathcal{F}} := \{A \cup N \mid A \in \mathcal{F} \text{ and } N \in \mathcal{N}\}$$

where $\mathcal{N}$ *be the collection of all null sets of* $(\Omega, \mathcal{F}, \mathbb{P})$ *and*

$$\tilde{\mathbb{P}}\left(\tilde{A}\right) := \mathbb{P}(A) \quad \left(\text{for all } \tilde{A} \in \tilde{\mathcal{F}} \text{ where } \tilde{A} = A \cup N \text{ for some } N \in \mathcal{N}\right).$$

PROOF. Let $\mathcal{N}$ be the collection of null sets of $(\Omega, \mathcal{F}, \mathbb{P})$. Further, define $\tilde{\mathcal{F}}$ as

$$\tilde{\mathcal{F}} = \{F \cup N \mid F \in \mathcal{F} \text{ and } N \in \mathcal{N}\}.$$

First, it follows that $\tilde{\mathcal{F}}$ is a σ-algebra by DeMorgan's laws since $\mathcal{F}$ is a σ-algebra and $\mathcal{N}$ is closed under countable unions.

Further, the mapping $\tilde{\mathbb{P}}$ defined as

$$\tilde{\mathbb{P}}\left(\tilde{A}\right) := \mathbb{P}(A) \quad \left(\text{for all } \tilde{A} \in \tilde{\mathcal{F}} \text{ where } \tilde{A} = A \cup N \text{ for some } N \in \mathcal{N}\right)$$

is a probability measure since $\mathbb{P}$ is a probability measure and $\mathbb{P}(A \cup N) = \mathbb{P}(A)$. □

Given the above theorem, I will assume in the sequel that all probability spaces are complete.

1.1.4. Finite Product Spaces. Product spaces and product measures play an important part in continuous-time finance and probability theory generally. For instance, the price of a portfolio can be represented as a (random) vector in $(\mathbb{R}^n; \mathcal{B}^n)$ where $\mathcal{B}^n$ is the σ-algebra of Borel sets in $\mathbb{R}^n$ (*i.e.* $\mathcal{B}^n = \mathcal{B}(\mathbb{R}^n)$). As much of this subject involves somewhat deep results from measure theory, I will only provide a rather brief introduction generally following Williams [**88**, ch. 8].

Motivation. Assume that I want to conduct an experiment to determine the probability of tossing two consecutive heads. This can be modeled as a single coin tossed twice. Now, I can represent the outcome of a single toss of a (fair) coin by the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = \{\{H\}, \{T\}\}$, $\mathcal{F} = (\emptyset, \Omega, \{H\}, \{T\})$ and

$$\mathbb{P}(A) = \begin{cases} 1/2, & A = \{H\} \\ 1/2, & A = \{T\}. \end{cases}$$

Then it seems reasonable to assume that the experiment of two consecutive coin tosses can modeled using the product space $(\Omega, \mathcal{F}) \times (\Omega, \mathcal{F})$ along with the mapping $\tilde{\mathbb{P}} : \mathcal{F} \times \mathcal{F} \to [0, 1]$ defined as

$$\tilde{\mathbb{P}}(A, B) := \mathbb{P}(A)\,\mathbb{P}(B) \quad (\text{for any } A, B \in \mathcal{F}).$$

The question becomes: *is this product space with this mapping (a) a probability space and (b) does it correctly model the tossing of a single coin twice*?

Conversely, this experiment can also be modeled as an experiment in which two identical coins are tossed simultaneously. In this case, the probability space, $(\Lambda, \mathcal{E}, \mathbb{Q})$, can be defined as

$$\Lambda = \{\{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}\} \text{ and } \mathcal{E} = 2^{\Lambda}$$

with

$$\mathbb{Q}(A) = \begin{cases} 1/4, & A = \{H, H\} \\ 1/4, & A = \{H, T\} \\ 1/4, & A = \{T, H\} \\ 1/4, & A = \{T, T\}. \end{cases}$$

Since these two experiments are essentially equivalent, we should expect that $(\Omega, \mathcal{F}) \times (\Omega, \mathcal{F}) = (\Lambda, \mathcal{E})$ and $\tilde{\mathbb{P}} = \mathbb{Q}$ (whatever that might mean). The following results answer these questions. □

I will begin by showing that the product of two σ-algebras is, in fact, a σ-algebra.

Definition. Let $(F, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces. For any $A \subseteq F$ and $B \subseteq E$, the (***Cartesian***) ***product of*** A ***and*** B is denoted as $A \times B$ and defined as

$$A \times B := \{(x, y) \mid \text{ for all } x \in A \text{ and } y \in B\}.$$

If $A \in \mathcal{F}$ and $B \in \mathcal{E}$, then $A \times B$ is said to be a ***measurable rectangle***.

Theorem 1.6. *Let $(F, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces and let $\mathcal{G}$ be the collection of all measurable rectangles. Then $\mathcal{G}$ is a σ-algebra on the product set $F \times E$.*

PROOF. Let $(F, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces and let $\mathcal{G}$ be the collection of all measurable rectangles of $F \times E$. By definition, $\emptyset := (\emptyset, \emptyset)$ and $F \times E$ are contained in $\mathcal{G}$. Additionally, any element of $\mathcal{G}$ is of the form $(A \times B)$ where $A \in \mathcal{F}$ and $B \in \mathcal{E}$. Since $\mathcal{F}$ and $\mathcal{E}$ are σ-algebras,

$$\begin{aligned}(A \times B)^c &= \{(x, y) \mid \text{ for all } x \in A^c \text{ and } y \in B^c\} \\ &= A^c \times B^c \in \mathcal{G}.\end{aligned}$$

Therefore, $\mathcal{G}$ is closed under complements.

Now, I must show that $\mathcal{G}$ is closed under countable unions. Therefore, let $\{(A \times B)_k\}_{k=1}^{\infty}$ be an arbitrary countable collection of elements of $\mathcal{G}$. Then for any k,

$$(A \times B)_k = \{(x, y) \mid \text{ for all } x \in A_k \text{ and } y \in B_k\}$$

where $A_k \in \mathcal{F}$ and $B_k \in \mathcal{E}$.

Hence,

$$\begin{aligned}\bigcup_{k=1}^{\infty} (A \times B)_k &= \bigcup_{k=1}^{\infty} \{(x, y) \mid \text{ for all } x \in A_k \text{ and } y \in B_k\} \\ &= \left\{(x, y) \mid \text{ for all } x \in \bigcup_{k=1}^{\infty} A_k \text{ and } y \in \bigcup_{k=1}^{\infty} B_k\right\} \\ &= \left(\bigcup_{k=1}^{\infty} A_k \times \bigcup_{k=1}^{\infty} B_k\right) \in \mathcal{G}.\end{aligned}$$

Thus, $\mathcal{G}$ is closed under countable unions and $\mathcal{G}$ must be a σ-algebra. □

DEFINITION. Let $(F, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces. Then the *σ-algebra* generated by the collection of all measurable rectangles in $F \times E$ is called the ***product σ-algebra of $\mathcal{F}$ and $\mathcal{E}$*** and denoted as $\mathcal{F} \otimes \mathcal{E}$. The measurable space consisting of the product set $F \times E$ along with $\mathcal{F} \otimes \mathcal{E}$ is called the ***product (measurable) space of*** $(F, \mathcal{F})$ ***and*** $(E, \mathcal{E})$ and is denoted as $(F \times E, \mathcal{F} \otimes \mathcal{E})$ or $(F, \mathcal{F}) \times (E, \mathcal{E})$.

EXAMPLE. It is possible to construct the two dimensional product space of the reals as follows

$$(\mathbb{R}, \mathcal{B}) \times (\mathbb{R}, \mathcal{B}) = (\mathbb{R} \times \mathbb{R}, \mathcal{B} \otimes \mathcal{B}) = \left(\mathbb{R}^2, \mathcal{B}^2\right).$$

□

REMARK. The above example suggests that $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}^2(\mathbb{R})$. In fact, this is generalized below. □

The following result, which is due to Fubini, shows that a product measure can be similarly constructed. A formal proof can be found in Williams [**88**, pgs. 77-78].

THEOREM 1.7. *Let* $(F, \mathcal{F}, \mu)$ *and* $(E, \mathcal{E}, \nu)$ *be measure spaces with finite measures. Then the product space,* $(F \times E, \mathcal{F} \otimes \mathcal{E}, \lambda)$ *is a measure space under the measure*

$$\lambda(A, B) := \mu(A)\,\nu(B) \quad (\textit{for } A \in \mathcal{F} \textit{ and } B \in \mathcal{E}).$$

Moreover, λ *is unique.*

DEFINITION. Let $(F, \mathcal{F}, \mu)$ and $(E, \mathcal{E}, \nu)$ be measure spaces with finite measures. Then the measure λ defined as

$$\lambda(A, B) := \mu(A)\,\nu(B) \quad (\text{for } A \in \mathcal{F} \text{ and } B \in \mathcal{E})$$

is called the ***product measure of*** μ ***and*** ν and is denoted as $\lambda = \mu \times \lambda$. The measure space consisting of the product set $F \times E$ along with $F \otimes \mathcal{E}$ and λ is called the ***product*** (***measure***) ***space of*** $(F, \mathcal{F}, \mu)$ ***and*** $(E, \mathcal{E}, \nu)$ and is denoted as $(F \times E, \mathcal{F} \otimes \mathcal{E}, \mu \times \lambda)$ or $(F, \mathcal{F}, \mu) \times (E, \mathcal{E}, \nu)$.

If fact, these results can be extended to any finite product of measure spaces using induction.

COROLLARY 1.8. *Let* $\{(F_k, \mathcal{F}_k, \mu_k)\}$ *be a finite family of measure spaces, then*

$$(F, \mathcal{F}, \mu) := \mathsf{X}_{k=1}^{n} (F_k, \mathcal{F}_k, \mu_k)$$

is a measure space where

$$F = \mathsf{X}_{k=1}^{n} F_k, \qquad \mathcal{F} = \bigotimes_{k=1}^{n} \mathcal{F}_k, \quad \mu = \mathsf{X}_{k=1}^{n} \mu_k.$$

In particular,

$$(\mathbb{R}^n, \mathcal{B}^n, \mathcal{L}^n) = \mathsf{X}_{k=1}^{n} (\mathbb{R}, \mathcal{B}, \mathcal{L}) \tag{1.1.1}$$

where $\mathcal{L}$ is the Lebesgue measure on the Borel sets.

The case of infinite products of measure spaces will be considered in the context of stochastic processes below (*cf.* Section 2.2.3).

1.2. Random Variables

In the discussion so far, I have made remarkably few assumptions regarding the nature of a probability space. Generally, such a space can be a very abstract set with little or no mathematical structure other than a set algebra and a finite measure. In other words, not much with which to work. A special class of functions known as *random variables* help solve this problem.

As functions mapping one probability space into another, random variables allow us to work in probability spaces possessing much richer mathematical properties (such as topologies and geometries) while still preserving the notion of observable events in the original probability space. Importantly, under the rather weak condition of *measurability*, they also preserve the probability measure of the original space (in fact, the probability measure on the target space is simply the *pull-back measure*

on the original space). As such, it is common to ignore the original probability space and simply concentrate on the richer space.

These ideas are captured in the following definitions.

DEFINITION. Given the two measurable spaces $(F, \mathcal{F})$ and $(E, \mathcal{E})$, a ***random variable*** (or more precisely, a ***random element***) $X(\cdot)$ is the measurable mapping from F into E such that $\omega \mapsto X(\omega)$. Further, the range of a random element is called the ***state space***. If the random variable takes on only a finite number of values, it is said to be ***simple***.

REMARK. Under this definition a random variable is a simply a function from one measurable space to another. The mathematical structure of the target space is often more important (and interesting) than the random variable itself. In actually, the term "random variable" should only refer to a mapping where the state space is $\mathbb{R}$. Other important state spaces are $\mathbb{R}^n$ (*i.e.* random *vectors*), spaces of functions such as L^p, and spaces of measures. However, I will follow the standard convention and refer to all such mappings from some sample space to a state space as a "random variable" unless otherwise noted. □

REMARK. A common abuse notation in probability theory is to refer to the random variable $X(\cdot)$ simply as X, thereby confusing the function $X(\cdot)$ with its value $X(\omega)$ for a particular $\omega \in \Omega$. Similarly, it is common to use the short-hand notation, $\{X = x\}$, for $\{\omega \in \Omega \mid X(\omega) = x\}$. □

In the context continuous-time finance, the sample space often represents the state of all information (both observable and unobservable) in the market while the state space represents all possible price levels for all the financial instruments that trade in that market. A random variable then is used to set the price level of some financial asset based on that information. As such, I generally will assume that the price of a financial instrumental is a real-valued random variable mapping the sample space to the reals equipped with the σ-algebra of Borel sets $\mathcal{B}$ (*i.e.* $X : (\Omega; \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$). In some cases, the price of an instrument will be considered in the context of a a portfolio consisting of a finite number of instruments. Here, the price of the constituents of the portfolio can be thought of as vector in $\mathbb{R}^n$. This gives rise to the following definition.

DEFINITION. Given two measurable spaces $(F, \mathcal{F})$ and $(E, \mathcal{E})$, a ***random vector*** is a mapping $X : (F; \mathcal{F})^n \to (E, \mathcal{E})^n$ denoted by

$$X(\omega) = \left(X^1(\omega), \ldots, X^n(\omega)\right).$$

EXAMPLE. For any $A \in \mathcal{F}$, the ***indicator function*** (of A) defined as

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

:

is a random variable. □

REMARK. For any $A \in \mathcal{F}$,

$$\delta_\omega(A) = 1_A(\omega).$$

In other words, the indicator function is a *random measure.* □

EXAMPLE. Given the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\{A_k\}_{k=1}^n$ be a finite disjoint partition of the sample space Ω (*i.e.* $\bigcup_{k=1}^n A_k = \Omega$ and $A_i \cap A_j = \emptyset$ for $i \neq j$). Then the ***simple function*** $\varphi(\cdot)$ defined as the finite sum of indicator functions

$$\varphi(\omega) := \sum_{k=1}^n a_k 1_{A_k}(\omega)$$

where the a_k are real-valued constants is a random variable. □

EXAMPLE. Standard step functions are examples of simple functions and, thus, are random variables. □

An important property of random variables is *measurability* which ensures that events in the state space are probabilistically equivalent (*i.e.* have the same probability of occurring) to those in the original space.

DEFINITION. Given two measurable spaces $(F, \mathcal{F})$ and $(E, \mathcal{E})$, a random variable $X : (F; \mathcal{F}) \to (E; \mathcal{E})$ is said to be $\mathcal{F}$-***measurable*** (***relative to*** $\mathcal{E}$) if the inverse image of an event in the state space is an event in the sample space, or

$$X^{-1}(A) \in \mathcal{F} \quad (\text{for all } A \in \mathcal{E})$$

where

$$X^{-1}(A) := \{X \in A\} = \{\omega \in \Omega \mid X(\omega) \in A\} \quad (\text{for each } A \in \mathcal{E}).$$

The following result is helpful in the sequel.

THEOREM 1.9. *Let X be a measurable, real-valued random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and f some real-valued continuous function on $\mathbb{R}$. Then $f(X)$ is a measurable random variable.*

PROOF. Let X be a measurable, real-valued random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and f some real-valued, continuous function on $\mathbb{R}$. Further, define the random variable $Y : \Omega \to \mathbb{R}$ as

$$Y(\cdot) := f \circ X(\cdot) = f(X(\cdot)).$$

For any $B \in \mathcal{B}$, I have to show that $Y^{-1}(B) \in \mathcal{F}$. Or, in other words, I have to show that $(f \circ X)^{-1}(B) = X^{-1} \circ f^{-1}(B) \in \mathcal{F}$.

Therefore, let $A := f^{-1}(B) := \{x \in \mathbb{R} \mid f(x) \in B\}$ for some $B \in \mathcal{B}$. Now, B is open by definition and f is continuous. Hence, A is open and $A \in \mathcal{B}$ since all open sets in $\mathbb{R}$ are in $\mathcal{B}$. However, X is measurable. Thus, $X^{-1}(A) \in \mathcal{F}$. Hence, $X^{-1} \circ f^{-1}(B) \in \mathcal{F}$. □

DEFINITION. Given two measurable spaces $(F, \mathcal{F})$ and $(E, \mathcal{E})$ and a random variable $X : (F; \mathcal{F}) \to (E; \mathcal{E})$, the set $\sigma(X) := \{X^{-1}(A) \mid \text{ for all } A \in \mathcal{E}\}$ is called the ***σ-algebra generated by*** X.

The following result proves that such a σ-algebra exists

THEOREM 1.10. *Given two measurable spaces $(F, \mathcal{F})$ and $(E, \mathcal{E})$ and a measurable random variable $X : (F; \mathcal{F}) \to (E; \mathcal{E})$, there exists a unique smallest sub-σ-algebra of $\mathcal{F}$ with respect to which X is measurable.*

PROOF. Let $(F, \mathcal{F})$ and $(E, \mathcal{E})$ be two measurable spaces and let $X(F; \mathcal{F}) \to (E; \mathcal{E})$ be measurable. Further, let

$$\sigma(X) := \left\{X^{-1}(A) \mid A \in \mathcal{E}\right\}$$

Define $\mathcal{G}$ to be the collection of sets in F of the form

$$\{X \in A\} := \{\omega \in \Omega \mid X(\omega) \in A\} \quad (\text{for each } A \in \mathcal{E}).$$

Now, for any countable collection $\{A_k\} \in \mathcal{E}$, $\cup_{k=1}^{\infty} A_k \in \mathcal{E}$ and $A_k^c \in \mathcal{E}$ for all k since $\mathcal{E}$ is a σ-algebra. Hence,

$$\cup_{k=1}^{\infty} \{X \in A_k\} = \{X \in \cup_{k=1}^{\infty} A_k\} \in \mathcal{G}$$

and

$$\{X \in A_i\}^c = \{X \in A_i^c\} \in \mathcal{G}.$$

Therefore, $\mathcal{G}$ is closed under countable unions and complements and, thus, it is a σ-algebra since $\emptyset$ and F are also in $\mathcal{G}$. By definition, X is $\sigma(X)$-measurable. Hence, $\mathcal{G} \subseteq \sigma(X)$. By construction, X is $\mathcal{G}$-measurable. Hence, $\sigma(X) \subseteq \mathcal{G}$. Thus, $\sigma(X) = \mathcal{G}$. □

REMARK. The σ-algebra generated by some random variable X contains all the *information* about X. For instance, if f is some *nice* function, then the random variable, $Y = f(X)$ is $\sigma(X)$-measurable. In other words, if we *know* X, we know Y. □

The above result can easily be extended to countable collection of random variables directly from the definition of σ-algebras.

COROLLARY 1.11. *Let $\{X_k : (F; \mathcal{F}) \to (E; \mathcal{E})\}$ be a countable collection of random variables, then there exists the unique smallest sub-σ-algebra with respect to which the X_k are measurable.*

1.2.1. Distributions. Through random variables, an outcome in the sample space (which may a very abstract set like the faces on a die) can be expressed in terms of the state space, a presumably nicer space such as $\mathbb{R}^n$. Likewise, a random variable's *distribution* provides the means of expressing the probability of an event occurring in the sample space in terms of an event occurring in the state space.

Motivation. Given a measurable random variable $X : (\Omega; \mathcal{F}) \to (E; \mathcal{E})$, every event in the state space $A \in \mathcal{E}$ corresponds to the event in the sample space $B = \{\omega \in \Omega \mid X(\omega) \in A\}$. This leads naturally to the existence of a measure on the state space $\mu(\cdot)$ induced by X which is simply $\mu(A) = \mathbb{P}(B)$ (*i.e* the *pullback measure*). This is formalized in the following definition. □

DEFINITION. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measurable space $(E, \mathcal{E})$, the (***probability***) ***distribution*** (*a.k.a.* ***law***) of a random variable $X : (\Omega; \mathcal{F}) \to (E; \mathcal{E})$ is the measure, $\mathbb{P}_X : \mathcal{E} \to \Omega$, defined as

$$\mathbb{P}_X(A) := \mathbb{P} \circ X^{-1}(A) = \mathbb{P}\{\omega \in \Omega \mid X(\omega) \in A\} \quad (\text{for all } A \in \mathcal{E}).$$

NOTATION. $\mathbb{P}_X(\cdot)$ is sometimes denoted as $\mu(\cdot)$ to stress that it is a measure. □

EXAMPLE (Poisson distribution). A real-valued random variable X is said to have ***Poisson distribution with rate*** λ if, for $\lambda > 0$ and any $k \in \mathbb{N}$,

$$\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

□

EXAMPLE (Binomial distribution). A real-valued random variable X is said to have ***Binomial distribution*** if, for some $n > 0$ and any $k = 1, \ldots, n$

$$\mathbb{P}(X = k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}$$

where $0 \leq p \leq 1$ (p is often referred to as the probability of a *successful trial*).

In particular, if $p = 1/2$, then

$$\mathbb{P}(X = k) = \frac{1}{2^n} \begin{pmatrix} n \\ k \end{pmatrix}.$$

□

The next result follows directly from the above definition.

THEOREM 1.12. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ and $(E, \mathcal{E})$ be a probability space and a measurable space respectively. Then for a measurable random variable $X : (\Omega; \mathcal{F}) \to (E; \mathcal{E})$, its distribution is a probability measure on $(E, \mathcal{E})$ and it is the unique probability measure generated by X.*

The distribution of a real-valued random variable leads naturally to the following definition of a *distribution function*.

DEFINITION. If $X : (\Omega; F) \to (\mathbb{R}; \mathcal{B})$ is a measurable, real-valued random variable, then the function, $F_X : \mathbb{R} \to [0, 1]$, defined by

$$F_X(x) := \mathbb{P}_X(X \leq x) = \mathbb{P}\{\omega \in \Omega \mid X(\omega) \leq x\} \quad (x \in \mathbb{R})$$

is called the (***cumulative***) ***distribution function*** (*a.k.a.* ***cdf***) of X.

REMARK. This is an abuse of notation since the distribution function depends on X *and* the probably measure $\mathbb{P}$. □

NOTATION. Unless it is necessary to avoid confusion, it is customary to drop the subscript indicating the random variable and simply denote the distribution and distribution function of a real-valued random variable as

$$F(x) := \mathbb{P}(X \leq x).$$

□

Generally, distribution functions arising in continuous-time finance have sufficient regularity (*i.e. absolute continuity*) to admit a (strong) derivative. This gives rise to the following definition.

DEFINITION. Given a real-valued, measurable random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, if its distribution function F has the form

$$F(x) = \int_{-\infty}^{x} f(y)\, dy < \infty,$$

then the real-valued function f is called the (***probability***) ***density function of*** F (*a.k.a.* ***pdf***).

REMARK. The name arises from the fact that the probability *mass* of any interval $[a, b]$ is given by

$$\mathbb{P}(a \leq X \leq b) = F(b) - F(a) = \int_{a}^{b} f(y)\, dy.$$

In particular,

$$\int_{-\infty}^{\infty} f(y)\, dy = 1.$$

□

REMARK. From the definition, a distribution function admits a density only if it is absolutely continuous (*cf.* [**8**, pg. 439]). As such, only the distribution function of a continuous distribution can admit a density. □

EXAMPLE (Normal distribution). Since a Poisson distribute is discrete, it does not have a density function. On the other, the ***standard normal*** (*a.k.a.* ***Gaussian***) ***distribution***, denoted as

$$X \sim N(0, 1),$$

has a density function of the form

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}. \tag{1.2.1}$$

Thus,

$$\mathbb{P}(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-y^2}{2}}\, dy.$$

Similarly, a ***normal*** (*a.k.a.* ***Gaussian***) ***distribution with mean*** μ ***and variance*** σ^2, denoted as

$$X \sim N(\mu, \sigma^2),$$

has a density function of the form

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

Then

$$\mathbb{P}(X \leq x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{\frac{-(y-\mu)^2}{2\sigma^2}} dy.$$

□

REMARK. As I will show in sequel, normal random variables play a central role in continuous-time finance since most price processes are assumed to normal or log normal. □

The concept of a distribution can be extended to multiple random variables.

DEFINITION. Given a probability space and a measurable space $(\Omega, \mathcal{F}, \mathbb{P})$ and $(E, \mathcal{E})$, if $X_1, \ldots, X_m$ are random variables from $(\Omega; \mathcal{F}) \to (E; \mathcal{E})$, their ***joint*** (***probability***) ***distribution*** (*a.k.a.* ***joint probability measure***) is defined as

$$\begin{aligned} \mathbb{P}_{X_1,\ldots,X_m}(A_1, \ldots, A_m): &= \mathbb{P}(X_1 \in A_1, \ldots, X_m \in A_m). \\ &= \mathbb{P}(X_1 \in A_1 \cap \ldots \cap X_m \in A_m) \end{aligned}$$

where $A_1, \ldots A_m \in \mathcal{E}$.

If $X_1, \ldots, X_m$ are random variables from $(\Omega; F) \to (\mathbb{R}; \mathcal{B})$, then the function, $F_{X_1,\ldots,X_m} : \mathbb{R}^m \to [0, 1]$, defined by

$$F_{X_1,\ldots,X_m}(x_1, \ldots, x_m) := \mathbb{P}_{X_1,\ldots,X_m}(X_1 \leq x_1, \ldots, X_m \leq x_m) \quad (\text{for all } x_1, \ldots, x_m \in \mathbb{R}).$$

is called the ***joint distribution function*** of $\mathbb{P}_{X_1,\ldots,X_m}$.

EXAMPLE. Let X and Y be real-valued random variables. Then for $A = (A_1, A_2) \in \mathcal{B}^2$,their joint probability distribution is

$$\begin{aligned} \mathbb{P}_{X,Y}(A) &= \mathbb{P}((X, Y) \in A) \\ &= \mathbb{P}(X \in A_1, X \in A_2) \\ &= \mathbb{P}(X \in A_1 \cap X \in A_2). \end{aligned}$$

□

DEFINITION. Given the real-valued random variables $X_1, \ldots, X_m$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, if the joint distribution function F has the form

$$F(x_1, \ldots, x_m) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f_1(y) \cdots f_m(y)\, dy_1 \cdots dy_m < \infty,$$

then the real-valued function f is called the (***probability***) ***density function of*** F.

EXAMPLE (Multivariate normal distribution). Let $X = (X^1, \ldots, X^n)$ be a n-dimensional real-valued random vector with a joint density of the form

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\mu)^{\mathrm{T}} \Sigma^{-1} (x-\mu)} \quad (x \in \mathbb{R}^n)$$

for some $\mu \in \mathbb{R}^n$ and some positive definitive $n \times n$ symmetric matrix Σ. Then X is said to have ***multivariate normal*** (*a.k.a.* ***Gaussian***) ***distribution*** (***with mean*** μ ***and covariance matrix*** Σ) and

$$X \sim N(\mu, \Sigma).$$

□

1.2.2. Uniqueness. Since random variables are functions, there are a number of ways to consider *equality*. Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is probability space and that $(E, \mathcal{E})$ is some measurable space.

DEFINITION. If $X, Y : (\Omega; F) \to (E; \mathcal{E})$ are two measurable random variables, they are said to be (point-wise) ***equal***, denoted as $X = Y$, if

$$X(\omega) = Y(\omega) \quad (\text{for all } \omega \in \Omega).$$

Similarly, they are said to be ***equal in distribution***, denoted as $X \overset{d}{=} Y$, if

$$\mathbb{P}_X(A) = \mathbb{P}_Y(A) \quad (\text{for all } A \in \mathcal{E}).$$

In the latter case, X and Y are said to ***identically distributed***.

The next result follows directly from the definitions. Note that the converse is false .

THEOREM 1.13. *Let $X, Y : (\Omega; F) \to (E; \mathcal{E})$ be two measurable random variables such that $X = Y$, then $X \overset{d}{=} Y$.*

Viewed as functions, it is natural to apply the measure theoretic property, *almost everywhere*, to random variables.

DEFINITION (Almost surely). We say that a property holds ***almost surely*** or ***a.s.*** if the probability of it being true is one (or, equivalently, it has probability zero of not being true). If particular, an event $A \in \mathcal{F}$ is said to ***happen almost surely*** if

$$\mathbb{P}(A) = 1.$$

DEFINITION. The measurable random variables $X, Y : (\Omega; F) \to (E; \mathcal{E})$ are said to ***equal almost surely***, denoted $X = Y$ *a.s.*, if

$$\mathbb{P}(\omega \in \Omega \mid X(\omega) = Y(\omega)) = 1.$$

NOTATION. Typically the set $\{\omega \in \Omega \mid X(\omega) = Y(\omega)\}$ is denoted as $\{X = Y\}$. □

Two random variables on a complete probability space are equal almost surely if they agree everywhere except on a set of outcomes with probability zero. In other words, they are *indistinguishable* from a probabilistic perspective. This creates an equivalence class much like functions that agree almost everywhere in measure theory. This is captured in the next result.

THEOREM 1.14. *The property "almost surely" creates an equivalence class of random variables.*

PROOF. Reflexivity and symmetry follow directly from the definition. Consequently, I need only prove transitivity (*i.e.* $\mathbb{P}(X = Z) = 1$). Therefore, let $X, Y, Z : (\Omega; F) \to (E; \mathcal{E})$ such that $X = Y$ *a.s.* and $Y = Z$ *a.s.*

First note that

$$\{X = Y\} \cap \{Y = Z\} \subseteq \{X = Z\} \tag{1.2.2}$$

Now, I claim that $\{X = Z\} \subseteq \{X = Y\}$. If not, then there is some $A \in \{X = Z\}$ such that $A \cap \{X = Y\} = \emptyset$. However, by the additivity property of probability measures (*cf.* Theorem 1.3),

$$\begin{aligned} \mathbb{P}(A \cup \{X = Y\}) &= \mathbb{P}(A) + \mathbb{P}\{X = Y\} \\ &> 1 \end{aligned}$$

which is impossible since $\mathbb{P}$ is a probability measure (*i.e.* $0 \leq \mathbb{P} \leq 1$).

Similarly, $\{X = Z\} \subseteq \{Y = Z\}$. Given these results

$$\{X = Z\} \subseteq \{X = Y\} \cap \{Y = Z\}.$$

Combining this with (1.1.1)

$$\{X = Z\} = \{X = Y\} \cap \{Y = Z\}.$$

If I claim that $\{X = Y\} = \{Y = Z\}$. If not, then $\{X = Y\}^c \cap \{Y = Z\} \neq \emptyset$ or $\{X = Y\}^c \cap \{Y = Z\} \neq \emptyset$. Without loss of generally, assume $\{X = Y\}^c \cap \{Y = Z\} \neq \emptyset$. Then

$$\begin{aligned} \mathbb{P}(\{X = Y\} \cup \{Y = Z\}) &= \mathbb{P}(\{X = Y\} \cup (\{X = Y\}^c \cap \{Y = Z\})) \\ &= \mathbb{P}(\{X = Y\}) + \mathbb{P}(\{X = Y\}^c \cap \{Y = Z\}) \quad (\perp) \\ &> 1. \end{aligned}$$

Therefore, $\{X = Y\} = \{Y = Z\} = \{X = Z\}$ and $\mathbb{P}(X = Z) = 1$. □

Clearly, equality is a stronger property than almost surely equality since two equal random variables are also equal *a.s.*, but not vice versa. However, two random variables that are equal a..s. have the same distributions since they only differ on a set of probability zero. This is captured in the following result.

THEOREM 1.15. *Let $X, Y : (\Omega; F) \to (E; \mathcal{E})$ be two measurable random variables such that $X = Y$ a.s., then $X \stackrel{d}{=} Y$.*

PROOF. Let $A = \{\omega \in \Omega \mid X(\omega) \neq Y(\omega)\}$. Since $X = Y$ a.s., then $\mathbb{P}(A) = 0$. Therefore, for any $B \in \mathcal{F}$, $\mathbb{P}_X(A \cap B) = \mathbb{P}_Y(A \cap B) = 0$ and

$$\begin{aligned}
\mathbb{P}_X(B) &= \mathbb{P}_X((B \cap A) \cup (B \cap A^c)) \\
&= \mathbb{P}_X(B \cap A) + \mathbb{P}_X(B \cap A^c) \\
&= \mathbb{P}_X(B \cap A^c) \\
&= \mathbb{P}\{\omega \in \Omega \mid X(\omega) \in B \cap A^c\} \\
&= \mathbb{P}\{\omega \in \Omega \mid Y(\omega) \in B \cap A^c\} \quad (X = Y \text{ a.s. on } A^c) \\
&= \mathbb{P}_Y(B \cap A^c) \\
&= \mathbb{P}_Y(B \cap A) + \mathbb{P}_Y(B \cap A^c) \\
&= \mathbb{P}_Y((B \cap A) \cup (B \cap A^c)) \\
&= \mathbb{P}_Y(B).
\end{aligned}$$

□

1.3. Conditional Probability and Independence

Often, we are interested in the knowing the probability of an event occurring predicated by the occurrence of one or more other events (*e.g.* what is the probability of tossing 10 heads in a row). Mathematically, this is captured by concept of *conditional probability*.

Assume in this section that $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space and that all random variables are real-valued and $\mathcal{F}$-measurable unless otherwise noted.

DEFINITION. Let $A, B \in \mathcal{F}$ such that $\mathbb{P}(B) > 0$. Then the ***conditional of probability of*** A ***given*** B is

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{1.3.1}$$

The next result follows directly from the definitions and helps illustrate the nature of conditional probability.

THEOREM 1.16 (Bayes theorem). *Let $A, B \in \mathcal{F}$ such that $\mathbb{P}(A), \mathbb{P}(B) > 0$. Then*

$$\mathbb{P}(B \mid A) = \mathbb{P}(A \mid B) \frac{\mathbb{P}(B)}{\mathbb{P}(A)}.$$

Motivation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then given some event $B \in \mathcal{F}$, the probability conditioned on B can be thought of as the restriction of $\mathbb{P}$ to a new measurable space $\left(\tilde{\Omega}, \tilde{\mathcal{F}}\right)$ "generated" by B, where

$$\tilde{\Omega} = B \text{ and } \tilde{\mathcal{F}} := \{A \cap B \mid A \in \mathcal{F}\}.$$

This is formalized in the following result. □

THEOREM 1.17. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then for any $B \in \mathcal{F}$, the subspace $\left(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}\right)$ is a probability space, where*

$$\tilde{\Omega} = B,\ \tilde{\mathcal{F}} := \{A \cap B \mid A \in \mathcal{F}\},\ \textit{and}\ \tilde{\mathbb{P}}(\,\cdot\,) := \mathbb{P}(\,\cdot\,|B)$$

PROOF. The fact that $\left(\tilde{\Omega}, \tilde{\mathcal{F}}\right)$ is a measurable space follows directly from the definitions. I need only to show that $\tilde{\mathbb{P}}$ is a probability measure on $\left(\tilde{\Omega}, \tilde{\mathcal{F}}\right)$. However, $\tilde{\mathbb{P}}$ agrees with $\mathbb{P}$ on $\tilde{\mathcal{F}}$ except for the scaling factor, $1/\mathbb{P}(B)$. Hence, it is non-negative and countability subadditive. Also,

$$\tilde{\mathbb{P}}(\emptyset) = 0 \quad \text{and} \quad \tilde{\mathbb{P}}(B) = 1.$$

Thus, $\tilde{\mathbb{P}}$ is a probability measure. □

Motivation. Recall

$$\mathbb{P}(A \cap B) = \mathbb{P}(\omega \in A \text{ and } \omega \in B).$$

After rearranging terms, (1.3.1) yields the probability of A and B occurring, or

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B). \tag{1.3.2}$$

Recall that $\mathbb{P}(A|B)$ is the probability of A occurring given the fact that B occurs. However, if A and B are *independent* events, the impact of B occurring should have no impact on the probability of A occurring. Therefore, if A and B are independent events, is seems reasonable that

$$\mathbb{P}(A \mid B) \;=\; \mathbb{P}(A)$$

which, by (1.3.2), implies

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

This is formalized as follows. □

DEFINITION. Given $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, A is said to be ***independent of*** B (or $A \perp B$) if

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

This definition can be extended to collections of multiple events, random variables, and σ-algebras.

DEFINITION. Let $A = \{A_k\}$ be a countable collection of events such that $A_k \in \mathcal{F}$ for all $k = 1, \ldots$. Then the A_k are said to be ***independent*** if for all indexes, $1 \leq k_1 < k_2, \ldots, < k_m$,

$$\mathbb{P}\left(\bigcap_{j=1}^{m} A_{k_j}\right) = \prod_{j=1}^{m} \mathbb{P}\left(A_{k_j}\right).$$

DEFINITION. Let $\{X_k\}_{k=1}^{\infty}$ be a countable collection of measurable, real-valued random variables. Then the X_k are said to be ***independent*** if for all integers $m > 2$ and choices of Borel sets,

$B_1, \ldots, B_m \in \mathcal{B}$,

$$\mathbb{P}\left(X_1 \in B_1, \ldots, X_m \in B_m\right) = \prod_{j=1}^{m} \mathbb{P}\left(X_j \in B_j\right).$$

The next result follows directly from the definitions.

THEOREM 1.18. *Let $X_1, \ldots, X_m$ be measurable, real-valued random variables. Then the following statements are equivalent:*

(1) *The X_k are independent for $k = 1, \ldots, m$;*
(2) $\mathbb{P}_{X_1,\ldots,X_m}\left(B_1, \ldots, B_m\right) = \prod_{k=1}^{m} \mathbb{P}_{X_k}\left(B_k\right), \quad B_1, \ldots, B_m \in \mathcal{B}$;
(3) $F_{X_1,\ldots,X_m}\left(x_1, \ldots, x_m\right) = \prod_{k=1}^{m} F_{X_k}\left(x_k\right), \quad x_1, \ldots, x_m \in \mathbb{R}$; *and*
(4) *If the random variables have densities f_{X_k} for all $k = 1, \ldots m$, then*

$$f_{X_1,\ldots,X_m}\left(x\right) = \prod_{k=1}^{m} f_{X_k}\left(x\right) \quad \left(x \in \mathbb{R}\right).$$

DEFINITION. Two σ-algebras $\mathcal{F}$ and $\mathcal{G}$ on the same sample space are said to be ***independent*** if F and G are independent for all $F \in \mathcal{F}$ and $G \in \mathcal{G}$, or

$$\mathbb{P}\left(F \cap G\right) = \mathbb{P}\left(F\right)\mathbb{P}\left(G\right) \quad \left(\text{for all } F \in \mathcal{F} \text{ and } G \in \mathcal{G}\right).$$

The next two results follow directly from the above definition.

THEOREM 1.19. *Let X and Y be measurable random variables, then $\sigma\left(X\right)$ and $\sigma\left(Y\right)$ are independent if and only X and Y are independent.*

THEOREM 1.20. *Let $\{X_j\}_{j=1}^{n+m}$ be a collection of random variables and suppose that $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}$. Then*

$$Y := f\left(X_1, \ldots, X_n\right) \quad \textit{and} \quad Z := g\left(X_{n+1}, \ldots, X_{n+m}\right)$$

are independent.

In terms of continuous-time finance, if the price of an asset is a random variable that is independent from a σ-algebra $\mathcal{F}$, then the information contained in $\mathcal{F}$ has no bearing on the price. On the other hand, if the random variable is $\mathcal{F}$-measurable, then $\mathcal{F}$ contains all the information necessary to determine the price level.

1.4. Integration with Respect to Probability Measures

Integration plays a key role in continuous-time finance and probability theory generally. However, since the sample space can be an arbitrary set without a geometric structure, standard Riemann integration is not viable. The solution is to use measure-theoretic integration that is similar to Lebesgue integration.

Assume in this section that $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space and that all random variables are real-valued and $\mathcal{F}$-measurable unless otherwise noted.

Motivation. Assume for the moment that $X : (\Omega; F) \to (\mathbb{R}; \mathcal{B})$ is a measurable random variable and let $\mathbb{R}$ be partitioned into a countable disjoint open intervals $\{I_k\}$ such that $\mathbb{R} = \cup_{k=1}^{\infty} I_k$. Now, I wish to define the expression

$$\int_{\Omega} X(\omega)\, d\mathbb{P}(\omega)$$

in terms of the $\{I_k\}$ whatever that might mean.

Since $I_k \in \mathcal{B}$ for all k and X is measurable, then each I_k has finite mass (which may be zero) under the pull-back probability measure $\mathbb{P}_X$. Then it seems reasonable to define integration under a probability measure as the sum of the appropriately weighted (probability) masses of the intervals, or

$$\int_{\Omega} X(\omega)\, d\mathbb{P}(\omega) = \lim_{|I_k| \to 0} \sum_{k=1}^{\infty} x_k \mathbb{P}_X(I_k)$$

for some x_k contained in the image of X on I_k.

Going forward, my plan is to generally follow the construction of general Lebesgue integral and establish integration under a probability measure first for simple functions and then extend this construct to functions of greater and greater complexity. This is formally proved in Durrett [**23**, ch 1.4]. □

Definition. Let $\{A_i\}_{i=1}^m$ be a finite, disjoint partition of Ω and let f be an almost surely bounded, non-negative simple function on Ω of the form,

$$f(\omega) = \sum_{i=1}^{m} a_i 1_{A_i}(\omega)$$

where $0 \leq a_i < \infty$ for all $i = 1, \ldots, m$.

Then its ***integral with respect to*** $\mathbb{P}$ is defined as

$$\int_{\Omega} f(\omega)\, d\mathbb{P}(\omega) := \sum_{i=1}^{m} a_i \mathbb{P}(\omega \in A_i).$$

This definition is reasonable since the right-hand side is well-defined (in fact, it is bounded since it is the finite sum of bounded summands). Moreover, the following result shows this integral is unique.

Theorem 1.21. *Let f be an almost surely bounded, non-negative, simple function. Then $\int_{\Omega} f(\omega)\, d\mathbb{P}(\omega)$ as defined above is independent of the partition and is therefore unique.*

PROOF. Let $\{A_i\}_{i=1}^k$ and $\{B_j\}_{j=1}^l$ be two finite, disjoint partitions of Ω. Additionally, let f be an almost surely bounded, non-negative, simple function such that

$$f(\omega) = \sum_{i=1}^{m} a_i 1_{A_i}(\omega) \quad 0 \leq a_i < \infty \quad (\text{for all } i = 1, \ldots, m).$$

and

$$f(\omega) = \sum_{j=1}^{l} b_j 1_{B_j}(\omega) \quad 0 \leq b_j < \infty \quad (\text{for all } j = 1, \ldots, l).$$

Since $\mathbb{P}$ is countably additive over disjoint sets (*cf.* Theorem 1.3)

$$\begin{aligned} \sum_{i=1}^{k} a_i \mathbb{P}(\omega \in A_i) &= \sum_{i=1}^{k} a_i \sum_{j=1}^{l} \mathbb{P}(\omega \in A_i \cap B_j) \\ &= \sum_{i=1}^{k} \sum_{j=1}^{l} a_i \mathbb{P}(\omega \in A_i \cap B_j) \end{aligned}$$

and

$$\begin{aligned} \sum_{j=1}^{l} b_j \mathbb{P}(\omega \in B_j) &= \sum_{j=1}^{l} b_j \sum_{i=1}^{k} \mathbb{P}(\omega \in A_i \cap B_j) \\ &= \sum_{j=1}^{l} \sum_{i=1}^{k} b_j \mathbb{P}(\omega \in A_i \cap B_j). \end{aligned}$$

Therefore,

$$\sum_{i=1}^{k} a_i \mathbb{P}(\omega \in A_i) - \sum_{j=1}^{l} b_j \mathbb{P}(\omega \in B_j) = \sum_{i=1}^{k} \sum_{j=1}^{l} (a_i - b_j) \mathbb{P}(\omega \in A_i \cap B_j).$$

Now, if $A_i \cap B_j = \emptyset$ for all i and j, then the right-hand side is zero and the result is proved.

Therefore, assume that $A_i \cap B_j \neq \emptyset$ for some i and j. Then there exists a $\xi \in \Omega$ such that $\xi \in A_i \cap B_j$. Thus, $f(\xi) = a_i$ and $f(\xi) = b_j$ since f is simple. Consequently, $a_i = b_j$. Since this is true for all such i and j, the result holds. □

NOTATION. The integral

$$\int_\Omega f(\omega) \, d\mathbb{P}(\omega)$$

is commonly written as

$$\int_\Omega f(\omega) \, \mathbb{P}(d\omega) \text{ or } \int_\Omega f \, d\mathbb{P}.$$

□

We now relax the assumption that f is simple and extend the definition to general bounded, non-negative functions as follows.

DEFINITION.

Assume that f is real-valued, bounded and non-negative a.s. Then its ***integral with respect to*** $\mathbb{P}$ is defined as

$$\int_{\Omega} f \, d\mathbb{P} := \sup_{\substack{g \leq f \\ g \text{ simple}}} \int_{\Omega} g \, d\mathbb{P}.$$

Based on the previous result, it follows from the above theorem that the right-hand side is well-defined, unique, non-negative and it is bounded since f is bounded. Hence, the left-hand side is well defined.

DEFINITION. If $\int_{\Omega} | f | d\mathbb{P} < \infty$, f is said to be ***summable***.

Now, I can relax the assumption that f is bounded and extend the definition to non-negative functions as follows.

DEFINITION.

Assume that f is non-negative a.s., then its ***integral with respect to*** $\mathbb{P}$ is defined as

$$\int_{\Omega} f \, d\mathbb{P} := \sup_{\substack{0 \leq g \leq f \\ g \text{ summable}}} \int_{\Omega} g \, d\mathbb{P}.$$

Here, we allow $\int_{\Omega} f \, d\mathbb{P}(\omega) = \infty$.

Based on the previous result, it follows that the right-hand side is well-defined (although possibly ∞), unique, and non-negative. Hence, again, the left-hand side is well defined.

To complete the derivation of integration with respect to a probability measure, note that any real-valued function f can be expressed as difference of two non-negative functions as

$$f(x) = f^{+}(x) - f^{-}(x) \quad (\text{for any } x \in \mathbb{R})$$

where

$$f^{+}(x) = f(x) \vee 0 \text{ and } f^{-}(x) = (-f(x)) \vee 0.$$

Using this, integration of any general function with respect to a probability measure can be defined as the difference of two integrals with non-negative integrands as follows.

DEFINITION. If f is any real-valued, measurable function, then its ***integral with respect to*** $\mathbb{P}$ is defined as

$$\int_{\Omega} f \, d\mathbb{P} := \int_{\Omega} f^{+} d\mathbb{P} - \int_{\Omega} f^{-} d\mathbb{P}$$

provided that one of the two integrals on the right-hand side is finite.

As constructed above, integration with respect to a probability measure possesses many of the properties of Lebesgue integration as captured in the following result which is proved in Durrett [**23**, ch. 1.4-1.5].

THEOREM 1.22. *Suppose X, Y are real-valued random variables. Then*

(1) $\int_\Omega a(X+Y)\,d\mathbb{P} = a\int_\Omega X d\mathbb{P} + a\int_\Omega Y d\mathbb{P}$ *for all* $a \in \mathbb{R}$ (*linearity*);
(2) *If* $Y \leq X$ *a.s., then* $\int_\Omega Y d\mathbb{P} \leq \int_\Omega X d\mathbb{P}$ (*comparison*);
(3) *If* $X \geq 0$ *a.s., then* $\int_\Omega X d\mathbb{P} \geq 0$*;*
(4) *If* $Y = X$ *a.s., then* $\int_\Omega Y d\mathbb{P} = \int_\Omega X d\mathbb{P}$*; and*
(5) $\left|\int_\Omega X d\mathbb{P}\right| \leq \int_\Omega |X|\, d\mathbb{P}$.

Generally, it is easier to compute the integral with respect to a probability measure in terms of its distribution using the following result which is proved in Durrett [**23**, pg. 30-32].

THEOREM 1.23 (Change of variables). *Let X be a real-valued random variable and let $f \in L^1(\mathbb{R})$ be a measurable function. Then*

$$\int_\Omega f(X(\omega))\, d\mathbb{P}(\omega) = \int_{\mathbb{R}} f(x)\, \mathbb{P}_X(dx) \tag{1.4.1}$$

where $\mathbb{P}_X(dx) = \mathbb{P}_X(X(\omega) \in dx)$.

REMARK. The right-hand expression is integration in the sense of the *Lebesgue-Stieltjes* integral (*cf.* [**5**, pg. 65]). □

REMARK. To understand the name of the theorem, recall that $\mathbb{P}_X(\cdot) = \mathbb{P} \circ X^{-1}(\cdot)$ and let $h(\omega) = X(\omega)$. Then

$$\begin{aligned} \int_\Omega f(h(\omega))\, d\mathbb{P}(\omega) &= \int_{\mathbb{R}} f(x)\, d\left(\mathbb{P} \circ h^{-1}(x)\right) \\ &= \int_{\mathbb{R}} f(x)\, \mathbb{P} \circ h^{-1}(dx) \end{aligned}$$

which is the general change of variable formula. □

NOTATION. If X is a real-valued random variable, $\mathbb{P}_X(dx)$ is sometimes written as $\mu(dx)$ to emphasize that it is integration with respect to a measure. □

Motivation. It also seems reasonable to define integration with respect to a distribution function as follows. For a real-valued random variable with the distribution function, F and define the function $dF : \mathbb{R} \to \mathbb{R}$, as

$$dF(x) := \lim_{y \uparrow x} F(x) - F(y)$$

where $x, y \in \mathbb{R}$ with $y < x$.

Then by the definition of a distribution function,

$$\begin{aligned} dF(x) &= \lim_{y \uparrow x} F(x) - F(y) \\ &= \lim_{y \uparrow x} \mathbb{P}(X \leq x) - \mathbb{P}(X \leq y) \\ &= \lim_{y \uparrow x} \mathbb{P}(y \leq X \leq x) \\ &= \mathbb{P}(X \in dx) \\ &= \mathbb{P}_X(dx). \end{aligned}$$

This suggests that for some nice function f

$$\begin{aligned} \int_\Omega f(X(\omega)) d\mathbb{P}(\omega) &= \int_{\mathbb{R}} f(x) \mathbb{P}(dx). \\ &= \int_{-\infty}^{\infty} f(x) dF(x) \end{aligned}$$

where the the right-hand is a *Riemann-Stieltjes* integral (*cf.* [**66**, Ch. 3.4]). This is captured in the following corollary. □

COROLLARY 1.24. *Let X be a real-valued random variable with the distribution function F. Then for any measurable function $g \in L^1(\mathbb{R})$*

$$\int_\Omega g(X(\omega)) d\mathbb{P}(\omega) = \int_{-\infty}^{\infty} g(x) dF(x). \tag{1.4.2}$$

Moreover, if the distribution is absolutely continuous with density f, then

$$\int_\Omega g(X(\omega)) d\mathbb{P}(\omega) = \int_{-\infty}^{\infty} g(x) f(x)\, dx.$$

In particular, if $g \equiv 1$, then

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

NOTATION. $dF(x)$ is sometimes written as $F(dx)$. □

The following result permits the explicit calculation of the probability of the occurrence of a particular event when the random variable admits a density function. The proof follows directly from the definitions and the fundamental theorem of calculus (*cf.* [**8**, pg. 237-238]).

COROLLARY 1.25. *If X is a real-valued random variable with density f, then for any $B \in \mathcal{B}$,*

$$\mathbb{P}(X \in B) = \int_B f(x)\, dx.$$

The next result also follows from the fundamental theorem of calculus (*cf.* [**8**, pg. 237-238]).

COROLLARY 1.26. *If X is a real-valued random variable with density f, then*

$$\mathbb{P}(dx) = f(x)\, dx.$$

Integration can be easily extended to vector-valued functions as follows.

DEFINITION. If $X = (X^1, \ldots, X^n)$ is a random vector, its ***integral with respect to*** $\mathbb{P}$ is defined as

$$\int_\Omega X d\mathbb{P} = \left(\int_\Omega X^1 d\mathbb{P}, \ldots, \int_\Omega X^n d\mathbb{P} \right).$$

The next result follows from the definitions of product spaces and product measures above (*cf.* Section 1.1.4). A proof for the case a two-dimensional product space is given in Durrett [**23**, pgs. 37-38]. The n-dimensional case follows directly by induction.

THEOREM 1.27 (Fubini's theorem in two dimensions). *Let* $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega_1, \mathcal{F}_1, \mathbb{P}_1) \times (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ *be a product measure space. If* X *is a* $\mathbb{R}^n$*-valued random vector and* $f \in L^1(\mathbb{R}^n)$*, then*

$$\begin{aligned} \int_\Omega f(X(\omega)) d\mathbb{P}(\omega) &= \int_{\Omega_1} \int_{\Omega_2} f\left(X^1(\omega_1), X^2(\omega_2)\right) d\mathbb{P}_2(\omega_2) d\mathbb{P}_1(\omega_1) \\ &= \int_{\Omega_2} \int_{\Omega_1} f\left(X^1(\omega_1), X^2(\omega_2)\right) d\mathbb{P}_1(\omega_1) d\mathbb{P}_2(\omega_2). \end{aligned}$$

NOTATION. $d\mathbb{P}(\omega) = \mathbb{P}(d\omega_1 \times d\omega_2) = d\mathbb{P}(\omega_1) d\mathbb{P}(\omega_2)$ where $\omega = (\omega_1, \omega_2)$. □

The next two results, which rely on Fubini's theorem, help illustrate the importance of convolution and kernels in probability theory that will be used extensively in the sequel.

THEOREM 1.28. *Let* X *and* Y *be independent, real-valued random variables with distribution functions* F *and* G *respectively. Then for any* $z \in \mathbb{R}$*,*

$$\mathbb{P}(X + Y \leq z) = \int_{-\infty}^{\infty} F(z - y) dG(y).$$

Moreover, if X *and* Y *are absolutely continuous with densities* f *and* g *respectively, then the density of the random variable* $Z = X + Y$ *exists and is given by*

$$h(x) = \int_{-\infty}^{\infty} f(x - y) g(y) dx.$$

PROOF. For any $x, y \in \mathbb{R}$, let $h(x, y) = 1_{\{x+y \leq z\}}(x, y)$. Then for a fixed y

$$\begin{aligned} \int_{\mathbb{R}} h(x, y) \mathbb{P}_X(dx) &= \int_{\mathbb{R}} 1_{\{x+y \leq z\}}(x, y) \mathbb{P}_X(dx) \\ &= \int_{\mathbb{R}} 1_{\{x \leq z-y\}}(x, y) \mathbb{P}_X(dx) \\ &= \int_{\mathbb{R}} 1_{\{-\infty \leq z-y\}}(x) \mathbb{P}_X(dx) \\ &= \mathbb{P}(X \leq z - y) \\ &= F(z - y). \end{aligned}$$

By Fubini's theorem

$$\begin{aligned}\mathbb{P}(X+Y\leq z) &= \int_{\mathbb{R}^2} h(x,y)\,\mathbb{P}_{X,Y}(dx\times dy)\\ &\quad \int_{\mathbb{R}}\left[\int_{\mathbb{R}} h(x,y)\,\mathbb{P}_X(dx)\right]\mathbb{P}_Y(dy)\\ &= \int_{\mathbb{R}} F(z-y)\,\mathbb{P}_Y(dy)\\ &= \int_{-\infty}^{\infty} F(z-y)\,dG(y).\end{aligned}$$

The second result follows immediately from this and the definition of a density function. □

1.5. Expectation and Expected Value

Given a random variable, it is natural to ask what should we expect its value to be after conducting a given experiment? This is captured in the concept of *expectation*.

Assume in this section that $(\Omega,\mathcal{F},\mathbb{P})$ is a given probability space and that all random variables are real-valued and $\mathcal{F}$-measurable unless otherwise noted.

DEFINITION. The ***expectation*** (*a.k.a.* ***expected value***) of a real-valued random variable X is an operator into the extended reals (*i.e.* $\mathbb{R}\cup\{-\infty,\infty\}$) given by

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

If the random variable is vector-valued, then, for any $X=(X^1,\ldots,X^n)$,

$$\mathbb{E}_{\mathbb{P}}[X] := \left(\mathbb{E}_{\mathbb{P}}\left[X^1\right],\ldots,\mathbb{E}_{\mathbb{P}}\left[X^n\right]\right)$$

where

$$\mathbb{E}_{\mathbb{P}}\left[X^k\right] := \int_{\Omega} X^k d\mathbb{P} \quad (\text{for } k=1,\ldots,n).$$

The ***variance*** of a random variable, X is defined as

$$\mathrm{Var}_{\mathbb{P}}[X] := \mathbb{E}_{\mathbb{P}}\left[(X-\mathbb{E}_{\mathbb{P}}[X])^2\right] = \int_{\Omega} |X-\mathbb{E}_{\mathbb{P}}[X]|^2\, d\mathbb{P}.$$

If the random variable is vector-valued, the variance is a positive, semi-definite matrix known as the ***covariance matrix***, $\mathrm{Var}_{\mathbb{P}}[\mathbf{X}]$, where the ij^{th} entry is

$$\mathrm{Var}_{\mathbb{P}}[X]^{ij} = \int_{\Omega}\left(X^i-\mathbb{E}_{\mathbb{P}}\left[X^i\right]\right)\left(X^j-\mathbb{E}_{\mathbb{P}}\left[X^j\right]\right)d\mathbb{P}.$$

NOTATION. Generally, the dependency on the specific probability measure is dropped unless it is needed to avoid confusion, or

$$\mathbb{E}[\,\cdot\,] := \mathbb{E}_{\mathbb{P}}[\,\cdot\,] \quad \text{and} \quad \mathrm{Var}[\,\cdot\,] := \mathrm{Var}_{\mathbb{P}}[\,\cdot\,].$$

□

THEOREM 1.29. *Given any random variable X,*

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

PROOF. By definition,

$$\begin{aligned}\mathrm{Var}[X] &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right] \\ &= \mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\ &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.\end{aligned}$$

□

DEFINITION. If X and Y are random variables, then their ***covariance*** is defined as

$$\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

The distribution of a random variable can be characterized in terms of its *moments*.

DEFINITION. If X is real-valued random variable, then the expression

$$\mathbb{E}[X^n] = \int_{\mathbb{R}} X^n \mathbb{P}(dx)$$

is called the n^{th}***-moment***. Similarly, the expression

$$\mathbb{E}[(X - \mathbb{E}[X])^n] = \int_{\mathbb{R}} (X - \mathbb{E}[X])^n \mathbb{P}(dx)$$

is called the n^{th}***-centered moment***.

EXAMPLE. The expected value of a random variable is its first moment. Its variance is its second centered moment. □

The next result, which follows directly from the definitions, shows how expectation can be expressed in terms of the distribution function.

THEOREM 1.30. *If X is a real-valued random variable with the distribution function F, then for any measurable function $g \in L^1(\mathbb{R})$,*

$$\begin{aligned}\mathbb{E}[g(X)] &= \int_{\mathbb{R}} g(x)\mathbb{P}(dx) \\ &= \int_{-\infty}^{\infty} g(x)\,dF(x).\end{aligned}$$

Moreover, if a density function f exists, then

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx. \tag{1.5.1}$$

The distribution function of a real-valued random variable provides a natural way to compute its moments.

COROLLARY 1.31. *If X is a real-valued random variable with the distribution function F, then*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, dF(x)$$

and

$$\operatorname{Var}[X] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2\, dF(x)$$

Moreover, if a density function f exists, then

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

and

$$\operatorname{Var}[X] = \int_{-\infty}^{\infty} (x - \mathbb{E}[x])^2 f(x)\, dx.$$

PROOF. In the first case, let $g(x) = x$ and, in the second, let $g(x) = (x - \bar{x})^2$ where $\bar{x} := \mathbb{E}[X]$. □

The following results hold for independent random variables.

THEOREM 1.32. *Let X and Y be independent, real-valued random variables and let $g \in L^1(\mathbb{R}^2)$ be a measurable function, then*

$$\mathbb{E}[g(X,Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y)\, \mathbb{P}_X(dx)\, \mathbb{P}_Y(dy).$$

In particular, if $g(x,y) = xy$, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y].$$

PROOF. By Theorem 1.18 and Fubini's theorem

$$\begin{aligned} \mathbb{E}[g(X,Y)] &= \int_{\mathbb{R}^2} g(x,y)\, \mathbb{P}_{X,Y}(dx \times dy) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y)\, \mathbb{P}_X(dx)\, d\mathbb{P}_Y(dy). \end{aligned}$$

Now, if $g(x,y) = xy$, then

$$\begin{aligned} \mathbb{E}[XY] &= \int_{\mathbb{R}} \int_{\mathbb{R}} xy\, \mathbb{P}_X(dx)\, \mathbb{P}_Y(dy) \\ &= \int_{\mathbb{R}} x\, \mathbb{P}_X(dx) \int_{\mathbb{R}} y\, \mathbb{P}_Y(dy) \\ &= \mathbb{E}[X]\,\mathbb{E}[Y]. \end{aligned}$$

□

The next two results follow directly from the above theorem.

COROLLARY 1.33. *Let X and Y be independent, real-valued random variables. Then* $\mathrm{Cov}[X,Y] = 0$.

COROLLARY 1.34. *If $\{X_i\}$ is a finite collection of independent, real-valued random variables, then*

$$\mathbb{E}\left[\prod_{i=1}^{k} X_i\right] = \prod_{i=1}^{k} \mathbb{E}[X_i].$$

COROLLARY 1.35. *If $\{X_i\}$ is a finite collection of independent, real-valued random variables, then*

$$\mathrm{Var}\left[\sum_{i=1}^{k} X_i\right] = \sum_{i=1}^{k} \mathrm{Var}[X_i].$$

PROOF. This is a proof by induction. Let $k = 2$. Then for two independent, real-valued random variables, X_1 and X_2,

$$\begin{aligned}
\mathrm{Var}[X_1 + X_2] &= \mathbb{E}\left[((X_1 + X_2) - \mathbb{E}[X_1 + X_2])^2\right] \\
&= \mathbb{E}\left[((X_1 - \mathbb{E}[X_1]) + (X_2 - \mathbb{E}[X_2]))^2\right] \\
&= \mathbb{E}\left[(X_1 - \mathbb{E}[X_1])^2 + (X_2 - \mathbb{E}[X_2])^2\right. \\
&\quad \left. + 2(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])\right] \\
&= \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + 2\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])].
\end{aligned}$$

Now, X_1 and X_2 are independent and $\mathbb{E}[X_1]$ and $\mathbb{E}[X_2]$ are numbers. Hence, $(X_1 - \mathbb{E}[X_1])$ and $(X_2 - \mathbb{E}[X_2])$ are independent. Thus, by Corollary 1.33,

$$\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] = \mathbb{E}[X_1 - \mathbb{E}[X_1]]\,\mathbb{E}[X_2 - \mathbb{E}[X_2]] = 0.$$

Hence,

$$\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2].$$

Now, assume that the result holds for k. Then letting $Y = \sum_{i=1}^{k} X_i$ and applying the above result to $\mathrm{Var}[Y + X_{k+1}]$, the result holds. □

I shall end this section with some important inequalities.

THEOREM 1.36 (Markov's inequality). *Let X be a non-negative real-valued, random variable. Then for any $\alpha > 0$*

$$\mathbb{P}(X \geq \alpha) \leq \frac{1}{\alpha}\mathbb{E}[X].$$

PROOF. Let X be a non-negative, real-valued, random variable and $\alpha > 0$. Then

$$\begin{aligned}\mathbb{E}[X] &= \int_{\mathbb{R}} x\,\mathbb{P}(dx)\\ &\geq \int_{x\geq\alpha} x\,\mathbb{P}(dx)\\ &\geq \alpha\int_{x\geq\alpha}\mathbb{P}(dx)\\ &\geq \alpha\mathbb{P}(X\geq\alpha).\end{aligned}$$

□

COROLLARY 1.37 (Chebyshev's inequality.). *Let X be real-valued random variable. Then for any $\alpha > 0$*

$$\mathbb{P}(|X|\geq\alpha)\leq\frac{1}{\alpha^k}\mathbb{E}\left[|X|^k\right]\quad(\textit{for } 1\leq k<\infty).$$

PROOF. Let $Y = |X|^k$ for $1 \leq k < \infty$ and apply Markov's theorem above. □

COROLLARY 1.38. *Let X be real-valued random variable. Then for any $\alpha > 0$*

$$\mathbb{P}(|X-\mathbb{E}[X]|\geq\alpha)\leq\frac{1}{\alpha^2}\mathrm{Var}[X]\quad(\textit{for } 1\leq k<\infty).$$

PROOF. Let $Y = X - \mathbb{E}[X]^2$ and apply Markov's theorem above. □

1.6. Spaces of Random Variables and Uniform Integrability

Viewed as functions, it seems reasonable to ask what properties, if any, a collection of random variables may have in a functional analysis sense. In fact, under the proper norm, the collection of real-valued, measurable random variables forms a Banach space. This is simply an application of standard L^p spaces to probability spaces. See Çinlar[**14**, ch. 2, sec. 3] for a rigorous treatment of the subject.

Throughout this section, assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space and all random variables are real-valued and $\mathcal{F}$-measurable unless otherwise noted.

My first task is to show that the collection of all real-valued, measurable random variables forms a linear space. For this, I will need the following lemma.

LEMMA 1.39. *Assume that $(E, \mathcal{E})$ and $(G, \mathcal{G})$ are two measurable spaces. Further, let $X : (\Omega; \mathcal{F}) \to (E; \mathcal{E})$ and $f : (E; \mathcal{E}) \to (G, \mathcal{G})$ be $\mathcal{F}-$ and $\mathcal{E}$-measurable, respectively. Then $f(X)$ is $\mathcal{F}$-measurable.*

PROOF. Let $(E, \mathcal{E})$ and $(G; \mathcal{G})$ be two measurable spaces. Further, let $X : (\Omega; \mathcal{F}) \to (E; \mathcal{E})$ and $f : (E; \mathcal{E}) \to (G, \mathcal{G})$ be $\mathcal{F}-$ and $\mathcal{E}$-measurable, respectively. Define the mapping $g : (\Omega; \mathcal{F}) \to (G; \mathcal{G})$ as

$$g(\omega) = f(X(\omega))\quad(\text{for } \omega\in\Omega).$$

Then for $A \in \mathcal{G}$

$$\begin{aligned} g^{-1}(A) &= \{\omega \in \Omega \mid g(\omega) \in A\} \\ &= \{\omega \in \Omega \mid f(X(\omega)) \in A\} \\ &= \left\{\omega \in \Omega \mid X(\omega) \in f^{-1}(A)\right\} \\ &= X^{-1}\left(f^{-1}(A)\right). \end{aligned}$$

Since, f is $\mathcal{E}$-measurable, $f^{-1}(A) \in \mathcal{E}$. Consequently, since X is $\mathcal{F}$-measurable, $g^{-1}(A) \in \mathcal{F}$. Hence, $f(X)$ is $\mathcal{F}$-measurable. □

THEOREM 1.40. *Let $\mathcal{V}$ be the collection of all real-valued, measurable random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then $\mathcal{V}$ with the operations $(+, \cdot)$ is a (possibly infinite) linear space over $\mathbb{R}$.*

PROOF. Let $\mathcal{V}$ be the collection of all real-valued, measurable random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $X, Y \in \mathcal{V}$. Further define Z as

$$Z(\omega) := \alpha X(\omega) + \beta Y(\omega) \quad (\text{for } \omega \in \Omega)$$

where α and β real constants. Clearly, Z is a real-valued random variable. Applying the lemma above, Z is also measurable. Hence, $Z \in \mathcal{V}$. □

Importantly, the space of real-valued, measurable random variables is a Banach space under the following norm.

DEFINITION. For any $1 \leq p < \infty$, define the space $L^p(\Omega, \mathcal{F}, \mathbb{P})$ to be collection of real-valued, measurable random variables such that

$$\|X\|_{L^p(\Omega,\mathcal{F},\mathbb{P})} < \infty$$

where

$$\|X\|_{L^p(\Omega,\mathcal{F},\mathbb{P})} := \text{ess sup}_\Omega |X| := \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p} = \left(\int_\Omega |X|^p \, d\mathbb{P}\right)^{1/p}.$$

Further, define the space $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ to be collection of real-valued, measurable random variables such that

$$\|X\|_{L^\infty(\Omega,\mathcal{F},\mathbb{P})} < \infty$$

$$\|X\|_{L^\infty(\Omega,\mathcal{F},\mathbb{P})} := \inf\left(C \in \mathbb{R}_+ \times \{\infty\} \mid |X| \leq C \text{ a.s.}\right).$$

REMARK. While the full description of the probability space is necessary to define the function space $L^p(\Omega, \mathcal{F}, \mathbb{P})$, for this paper, I will adopt the common abuse of notation and refer to it simply as $L^p(\Omega)$. □

The nest result shows that the collection of measurable random variables form a Banach space. The proof, which relies on the *Riesz-Fischer theorem*, is given in Billingsley [**8**, sec. 19].

THEOREM 1.41. *For any $1 \leq p \leq \infty$, $L^p(\Omega)$ is a Banach space under the norm, $\|\cdot\|_{L^p(\Omega)}$.*

The next result follows directly from the definitions of expectation and the $L^2(\Omega)$ norm.

COROLLARY 1.42. *$L^2(\Omega)$ is a Hilbert space under the inner product*

$$(\cdot,\cdot)_{L^2(\Omega)} : L^2(\Omega) \times L^2(\Omega) \to \mathbb{R}$$

defined as

$$(X,Y)_{L^2(\Omega)} : = \int_\Omega XY d\mathbb{P} = \mathbb{E}[XY]$$

for $X, Y \in L^2(\Omega)$.

NOTATION. Unless there is some confusion, I will refer to $(\cdot,\cdot)_{L^2(\Omega)}$ simply as $(\cdot,\cdot)$ herein. □

The next result is well-known inequality from analysis.

THEOREM 1.43 (Lyapunov's inequality). *Let $X \in L^p \cap L^q$ for $1 \leq p < \infty$. Then for $1 \leq q < p$,*

$$\|X\|_{L^q(\Omega)} \leq \|X\|_{L^p(\Omega).}$$

PROOF. Let $X \in L^p \cap L^q$ for $1 \leq p < \infty$ and let $1 \leq q < p$. Define the set S as

$$S := \{x \in \mathbb{R} \mid x \geq 0\}$$

and the function $\varphi : \mathbb{R} \to \mathbb{R}$

$$\varphi(x) := x^{p/q}.$$

Then both S and φ are convex. Therefore, applying Jensen's inequality for some summable random variable Y

$$\varphi(\mathbb{E}[Y]) \leq \mathbb{E}[\varphi(Y)].$$

Hence,

$$\mathbb{E}[Y]^{\frac{p}{q}} \leq \mathbb{E}\left[Y^{\frac{p}{q}}\right].$$

Setting $Y := |X|^q$, then

$$\mathbb{E}[|X|^q]^{p/q} \leq \mathbb{E}[|X|^p].$$

Hence,

$$\mathbb{E}[|X|^q]^{1/q} \leq \mathbb{E}[|X|^p]^{1/p}$$

and

$$\|X\|_{L^q(\Omega)} \leq \|X\|_{L^p(\Omega).}$$

□

As in traditional analysis, L^p spaces of random variables admit a natural dual space. Consequently, a number important results under the *weak topology* from functional analysis can then be stated in terms of random variables. I shall begin with a number of definitions.

DEFINITION. Assume $1 \leq p < \infty$. Then its ***conjugate*** q is the number such that

$$\frac{1}{p} + \frac{1}{q} = 1$$

or

$$q = \frac{p}{p-1}.$$

If $p = \infty$, the its conjugate $q = 1$ and vice versa.

DEFINITION. Let $1 \leq p \leq \infty$. A ***linear functional*** on $L^p(\Omega)$ is a real-valued mapping φ from $L^p(\Omega)$ to $\mathbb{R}$ such that

$$\varphi(\alpha X + \beta Y) = \alpha\varphi(X) + \beta\varphi(Y)$$

for $X, Y \in L^p(\Omega)$ and $\alpha, \beta \in \mathbb{R}$.

DEFINITION. Let $1 \leq p \leq \infty$. A linear functional φ on $L^p(\Omega)$ is said to be ***bounded*** if there exists a finite $C \in \mathbb{R}$ such that

$$|\varphi(X)| < C \|X\|_{L^p(\Omega)} \quad (\text{for all } X \in L^p(\Omega)).$$

DEFINITION. Let $1 \leq p \leq \infty$. The collection of all bounded linear functionals on $L^p(\Omega)$ is said to be the ***dual space of*** $L^p(\Omega)$ and denoted as $L^{p*}(\Omega)$.

The next result is a classical result from analysis and is proved in Billingsley [**8**, pgs. 259-260]

THEOREM 1.44. *Assume $1 \leq p \leq \infty$. The dual space $L^{p*}(\Omega)$ on $L^p(\Omega)$is a Banach space with the norm $\|\cdot\|_{L^{p*}(\Omega)}$ where*

$$\begin{aligned} \|\varphi\|_{L^{p*}(\Omega)} &= \sup_{X \in L^p(\Omega)} \left\{ \frac{|\varphi(X)|}{\|X\|_{L^p(\Omega)}} \mid X \neq 0 \right\} \\ &= \sup \left\{ |\varphi(X)| \mid \|X\|_{L^p(\Omega)} \leq 1 \right\}. \end{aligned}$$

Moreover, $L^{p*}(\Omega) = L^q(\Omega)$ where q is the conjugate of (i.e. $\frac{1}{p} + \frac{1}{q} = 1$).

The next result provides some motivation behind the concept of *uniform integrability.*

LEMMA 1.45. *X is summable if and only if*

$$\lim_{\varepsilon \to \infty} \mathbb{E}\left[|X| \, 1_{\{|X| > \varepsilon\}} \right] = 0.$$

PROOF. Assume that X is summable. Then it is clear that

$$|X| \, 1_{\{|X| > \varepsilon\}} \to 0 \quad \text{as } \varepsilon \to \infty.$$

Also, $|X|\,1_{\{|X|>\varepsilon\}} < |X| < \infty$ for all ε. Then since X is summable, by the dominated convergence theorem (*cf.* Theorem 1.55), the result holds.

Conversely, if the result holds, choose ε large enough so that $\mathbb{E}\left[|X|\,1_{\{|X|>\varepsilon\}}\right] < C$ for some finite constant. Also, for any $\varepsilon < \infty$

$$\begin{aligned} |X| &= |X|\,1_{\{|X|>\varepsilon\}} + |X|\,1_{\{|X|\leq\varepsilon\}} \\ &\leq |X|\,1_{\{|X|>\varepsilon\}} + \varepsilon. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}\left[|X|\right] &\leq \mathbb{E}\left[|X|\,1_{\{|X|>\varepsilon\}} + \varepsilon\right] \\ &\leq C + \varepsilon < \infty. \end{aligned}$$

□

DEFINITION. A collection of real-valued random variables $\mathcal{C}$ is said to be ***uniformly integrable*** if

$$\lim_{\varepsilon\to\infty} \sup_{X\in\mathcal{C}} \mathbb{E}\left[|X|\,1_{\{|X|>\varepsilon\}}\right] = 0.$$

By the comparison property of integration with respect to a probability measure (*cf.* Theorem 1.22) and the above lemma, the following result holds.

THEOREM 1.46. *Let $\mathcal{C}$ be a uniformly integrable collection of real-valued, measurable random variables, then $\mathcal{C} \subset L^1(\Omega)$.*

REMARK. The converse is false (*cf.* [**14**, pg. 72]).

1.7. Properties of the Integral under a Probability Measure

As constructed, integration with respect to a probability measure is simply a special case of integration with respect to a finite measure. Hence, such integrals possess of number of important measure-theoretic and analytic properties that I will state without proof. See Durrett [**23**, ch. 1.4-1.5] for a rigorous treatment of the subject including proofs for the following results.

Assume in this section that $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space and that all random variables are real-valued and $\mathcal{F}$-measurable unless otherwise noted.

DEFINITION 1.47. We say that X is ***integrable*** if $\int_\Omega |X|^p\, d\mathbb{P} < \infty$ for some $1 \leq p < \infty$. In particular, we say that X is ***summable*** if $\int_\Omega |X|\, d\mathbb{P} < \infty$.

THEOREM 1.48 (Jensen's inequality). *Suppose that $\varphi \in L^1(\mathbb{R})$ is convex and X is a summable random variable. Then*

$$\varphi\left(\int_\Omega X d\mathbb{P}\right) \leq \int_\Omega \varphi(X)\, d\mathbb{P}$$

or

$$\varphi\left(\mathbb{E}\left[X\right]\right) \leq \mathbb{E}\left[\varphi(X)\right].$$

THEOREM 1.49 (Hölder's inequality). *Let $1 < p, q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then given $X \in L^p(\Omega)$ and $Y \in L^q(\Omega)$*

$$\int_\Omega |XY| \, d\mathbb{P}(\omega) \leq \|X\|_{L^p(\Omega)} \|Y\|_{L^q(\Omega)}$$

or

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{1/p} (\mathbb{E}[|Y|^q])^{1/q}.$$

Setting $p = q = 1/2$ gives the next result

COROLLARY 1.50 (Cauchy-Schwarz inequality). *Given $X, Y \in L^2(\Omega)$*

$$\int_\Omega |XY| \, d\mathbb{P} \leq \|X\|_{L^2(\Omega)} \|Y\|_{L^2(\Omega)}$$

or

$$\mathbb{E}[|XY|] \leq \left(\mathbb{E}\left[|X|^2\right]\right)^{1/2} \left(\mathbb{E}\left[|Y|^2\right]\right)^{1/2}.$$

THEOREM 1.51 (Minkowski's inequality). *Let $1 < p < \infty$. Then for any $X, Y \in L^p(\Omega)$*

$$\|X + Y\|_{L^p(\Omega)} \leq \|X\|_{L^p(\Omega)} + \|Y\|_{L^p(\Omega)}$$

or

$$(\mathbb{E}[|X + Y|^p])^{1/p} \leq (\mathbb{E}[|X|^p])^{1/p} + (\mathbb{E}[|Y|^q])^{1/q}.$$

The following results are useful in passing to the limits of a convergence sequence.

THEOREM 1.52 (Bounded convergence theorem). *Assume X is a summable random variable and $\{X_k\}$ is a countable sequence of summable random variables such that $X_k \to X$ a.s. If there exists some finite constant C such that $|X_k| \leq C$ for all k, then*

$$\begin{aligned} \lim_{k\to\infty} \int_\Omega X_k d\mathbb{P} &= \int_\Omega \lim_{k\to\infty} X_k d\mathbb{P} \\ &= \int_\Omega X d\mathbb{P} \end{aligned}$$

or

$$\lim_{k\to\infty} \mathbb{E}[X_k] = \mathbb{E}[X].$$

THEOREM 1.53 (Fatou's lemma). *If $\{X_k\}$ is countable sequence of summable random variables such that $X_k \geq 0$ for all k, then*

$$\int_\Omega \liminf_{k\to\infty} X_k d\mathbb{P} \leq \liminf_{k\to\infty} \int_\Omega X_k d\mathbb{P}$$

or

$$\mathbb{E}\left[\liminf_{k\to\infty} X_k\right] \leq \liminf_{k\to\infty} \mathbb{E}[X_k].$$

THEOREM 1.54 (Monotone convergence theorem). *If X is an summable random variable and $\{X_k\}$ is countable sequence of random variables such that $0 \leq X_1 \leq X_2 \leq \ldots$ a.s. where $X_k \uparrow X$ a.s., then*

$$\begin{aligned} \lim_{k\uparrow\infty} \int_\Omega X_k d\mathbb{P} &\leq \int_\Omega \lim_{k\uparrow\infty} X_k d\mathbb{P} \\ &\leq \int_\Omega X d\mathbb{P} \end{aligned}$$

or

$$\lim_{k\uparrow\infty} \mathbb{E}[X_k] \leq \mathbb{E}[X].$$

THEOREM 1.55 (Dominated convergence theorem). *If X and Y are summable random variables and $\{X_k\}$ is countable sequence of summable random variables such that $X_k \to X$ a.s. and $|X_k| \leq Y$ a.s. for all k, then*

$$\begin{aligned} \lim_{k\to\infty} \int_\Omega X_k d\mathbb{P} &\leq \int_\Omega \lim_{k\to\infty} X_k d\mathbb{P} \\ &\leq \int_\Omega X d\mathbb{P} \end{aligned}$$

or

$$\lim_{k\to\infty} \mathbb{E}[X_k] \leq \mathbb{E}[X].$$

1.8. Conditional Expectation

Often the price of a financial instrument is represented by the random variable. If it is independent of a given σ-algebra, then the information contained in that σ-algebra has no bearing on its price. On the other hand, if the random variable is measurable with respect to that σ-algebra, then the σ-algebra contains all the information necessary to determine price. In other words, if the state of the σ-algebra is known, the price of that financial instrument is known. However, it is quite possible that σ-algebra may contain some, but not all of the information required to determine the price completely. Mathematically, this is captured by the concept of *conditional expectation* - the expected value of a random variable predicated on the occurrence (or not) of one or more observable events.

Assume in this section that $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space and that all random variables are real-valued and $\mathcal{F}$-measurable unless otherwise noted.

Motivation. Recall that we wanted to determine the probability of tossing two heads in some experiment (*cf.* motivation on page 8). This could be modeled by tossing two individual coins simultaneously. In this case, the sample space is

$$\Omega = \{\{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}\}$$

and we can select the family of events to be $\mathcal{F} = 2^{\Omega}$. Given this, the probability of tossing two heads is 25% since the results of each coin are assumed to be independent and fair (*i.e.* the probability of tossing a head or a tail is equal).

Now, this experiment can also be modeled as two sequential tosses of the same coin. As I showed earlier (*cf.* Section 1.1.4), this can be modeled by the probability space,

$$(\Lambda, \mathcal{G}, \mathbb{P})^2 = (\Lambda, \mathcal{G}, \mathbb{P}) \times (\Lambda, \mathcal{G}, \mathbb{P})$$

where

$$\Lambda = \{\{H\}, \{T\}\}$$

and the family of events is

$$\mathcal{G} = (\emptyset, \Omega, \{H\}, \{T\}).$$

Then

$$\Lambda \times \Lambda = \Omega \quad \text{and} \quad \mathcal{G} \otimes \mathcal{G} = \mathcal{F}.$$

Hence, this model is probabilistically equivalent to the first and, thus, the probability of tossing two heads is also 25% .

Finally, this experiment can also be modeled as follows. Prior to the first toss, the sample space of possible outcomes of the first toss can be thought of as $\Omega_1 := \{\{H\}, \{T\}\}$ and the family of events as $\mathcal{F}_1 := (\emptyset, \Omega_1, \{H\}, \{T\})$ while the possible outcomes of the second toss can be thought of as $\Omega_2 := \{\{\cdot, H\}, \{\cdot, T\}\}$ and the family of events as $\mathcal{F}_2 := (\emptyset, \Omega_2, \{\cdot, H\}, \{\cdot, T\})$. Here, $\{\cdot, H\}$ indicates that the outcome is ***unresolved*** and is dependent on the result of the first toss.

For instance, assuming that the first toss is a head, prior to the second toss, the sample space of possible outcomes following the second toss is

$$\Omega_2 := \{\{H, H\}, \{H, T\}\}$$

and the family of events is

$$\mathcal{F}_2 := (\emptyset, \Omega_2, \{H, H\}, \{H, T\})$$

Therefore, knowing that the first toss was a head, the probability of tossing two heads is 50%. Obviously, if the first toss was a tail, then

$$\Omega_2 = \{\{T, H\}, \{T, T\}\}$$

and

$$\mathcal{F}_2 = (\emptyset, \Omega_2, \{T, H\}, \{T, T\}).$$

Hence, the probability of tossing two heads is zero.

In other words, the probability of tossing two heads is conditioned on the first outcome and, therefore, is not fixed, but depends on the result of the first toss. As such, the probability of the second toss is a random variable.

This simple example suggests that the occurrence of given event should inform our expectation of some subsequent event. In the above example, knowing the result of the first toss (*i.e.* having more information) changes our expectation regarding the result of second toss absent that information. However, it also seems reasonable that if the two events are independent, then the outcome of one event should not have any impact on our expectation of the other.

Now, recall that the expectation of some summable, real-valued random variable is define as

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

Essentially, the expected value is the "average" value of X over the complete sample space. This suggests that the expected value of X conditioned on the occurrence of some event $B \in \mathcal{B}$ should be the "average" value of X on the subspace generated by B. Recalling Theorem 1.17, this probability measure on this subspace is

$$\begin{aligned} \tilde{\mathbb{P}}(\cdot) &= \mathbb{P}(\cdot \mid B) \\ &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \end{aligned}$$

This suggests defining conditional expectation as

$$\begin{aligned} \mathbb{E}[X \mid B] &= \int_{B} X d\tilde{\mathbb{P}} \\ &= \frac{1}{\mathbb{P}(B)} \int_{B} X d\mathbb{P} \\ &= \frac{1}{\mathbb{P}(B)} \mathbb{E}[X 1_B] \end{aligned}$$

where the symbol $\mathbb{E}[\cdot \mid B]$ indicates the explicit dependence of the expression on the choice of the event, B. In other words, conditional expectation is a function of B. Since the occurrence of B is random, $\mathbb{E}[\cdot \mid B]$ is a random variable. Given this insight, let me first provide a definition of expectation conditioned on the occurrence of a given event. □

DEFINITION. For any event $B \in \mathcal{F}$ and a summable random variable X, the ***conditional expectation of*** X ***given*** B, denoted by

$$\mathbb{E}[X \mid B],$$

is any random variable Y that is $\sigma(B)$-measurable and satisfies

$$\int_{B} Y d\mathbb{P} = \int_{B} X d\mathbb{P}. \tag{1.8.1}$$

This definition can be naturally extended to any sub-family of events as follows.

DEFINITION. Given a proper sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ and a summable random variable X, the ***conditional expectation of*** X ***given*** $\mathcal{G}$, denoted by

$$\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X(\omega) \mid \omega \in G \text{ for all } G \in \mathcal{G}],$$

is any random variable Y that is $\mathcal{G}$-measurable and satisfies

$$\int_G Y d\mathbb{P} = \int_G X d\mathbb{P} \quad (\text{for all } G \in \mathcal{G}). \tag{1.8.2}$$

NOTATION. If the σ-algebra is generated by some other random variable Z, then the ***conditional expectation of*** X ***given*** $\sigma(Z)$ is typically written as $\mathbb{E}[X \mid Z]$.

The following result provides that such a random variable exists and is unique. It's proof is given in Shreve [**?**, pg. 69].

THEOREM 1.56. *Given a proper sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ and a summable random variable X, the conditional expectation of X given $\mathcal{G}$ exists and is unique a.s.*

1.8.1. Conditional Expectations with respect to Hilbert spaces. Conditional expectation can also be viewed in geometric terms using the fact that square-integrable random variables form a Hilbert space (*cf.* Corollary 1.42).

Motivation. Intuitively, if a random variable X is $\mathcal{G}$-measurable for some proper sub-σ-algebra, $\mathcal{G} \subset \mathcal{F}$, then $\mathcal{G}$ contains all the information we need to determine X (*i.e.* if we know $\mathcal{G}$, then we know X). In other words, we should expect $\mathbb{E}[X \mid \mathcal{G}] = X$. On the other hand, if X is independent of $\mathcal{G}$, then knowing $\mathcal{G}$ should provide no insight into X and we should expect $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$. Geometrically, this suggest if X is $\mathcal{G}$-measurable random variable, it must lie in the subspace of $\mathcal{F}$ generated by $\mathcal{G}$. Conversely, if knowing $\mathcal{G}$ provides no information about X whatsoever, it suggests that X lies in a subspace orthogonal to the one generated by $\mathcal{G}$.

Recall that the space of all square-integrable, real-valued measurable random vectors on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $L^2(\Omega, \mathcal{F}, \mathbb{P})$, is a Hilbert space (*cf.* Corollary 1.42). Then for any sub-σ-algebra, $\mathcal{G} \subset \mathcal{F}$, $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is also a Hilbert space such that $L^2(\Omega, \mathcal{G}, \mathbb{P}) \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$. Hence, $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed linear subspace since it is complete. Consequently, for any $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, I can construct a $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ such that Y is the "closest" random variable to X in the following "least-squared" sense

$$Y \in \underset{Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})}{\arg\min} \|X - Z\|^2_{L^2(\Omega, \mathcal{F}, \mathbb{P})}. \tag{1.8.3}$$

In other words, Y is the best approximation of X that is in the subspace, $L^2(\Omega, \mathcal{G}, \mathbb{P})$ (clearly, if $X \in L^2(\Omega, \mathcal{G}, \mathbb{P})$, then the minimizer is X itself).

To prove this claim, I need to rely on the following result which is the generalization of the least square method applied to infinite dimensional spaces. The result is proved in Chorin and Hald [**15**, Ch 1.1]. □

LEMMA 1.57. *Let V be a closed linear subspace of a linear space U. Then for any $x \in U$, there exists a unique $y \in V$ such that*

$$\|x - y\|^2 = \min_{z \in V} \|x - z\|^2.$$

Moreover, for any $z \in V$, $x - y$ is orthogonal to z, or

$$(z, x - y) = 0 \quad (\textit{for all } z \in V).$$

DEFINITION. Given the set up above, the minimizing vector y is called the ***projection of*** x ***onto*** V and represents the best of estimation of x restricted to V**.** The vector, $x - y$, is called the ***residual vector*** and represents the error in the estimation of x by y.

The above lemma can be used to prove the following result.

THEOREM 1.58. *Let $\mathcal{G} \subset \mathcal{F}$ be a proper sub-σ-algebra and $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Then $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ is the almost surely unique minimizer of*

$$\min_{Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})} \|X - Z\|^2$$

if and only $Y = \mathbb{E}[X \,|\mathcal{G}]$.

PROOF. Assume that $\mathcal{G} \subset \mathcal{F}$ is a proper sub-σ-algebra and let $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. First, assume that Y a minimizer. Then by Lemma 1.57, for any random variable $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$,

$$\begin{aligned} 0 &= (Z, X - Y) \\ &= (Z, X) - (Z, Y). \end{aligned}$$

This implies

$$(Z, X) = (Z, Y) \quad \big(\text{for all } Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})\big).$$

In particular, let $Z = 1_A$ for any event $A \in \mathcal{G}$. Then by the definition of the inner product on $L^2(\Omega, \mathcal{G}, \mathbb{P})$

$$\int_\Omega X 1_A d\mathbb{P} = \int_\Omega Y 1_A d\mathbb{P} \quad (\text{for any } A \in \mathcal{G}).$$

Thus,

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

However, by (1.8.1)

$$Y = \mathbb{E}[X \mid A] \quad (\text{for any } A \in \mathcal{G}).$$

Then by (1.8.2)

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

Therefore, the conditional expectation of X on $\mathcal{G}$ is the $\mathcal{G}$-measurable random variable that is the best approximation of X in $\mathcal{G}$ in the least squares sense. Reversing the steps proves the converse. □

1.8.2. Properties of Conditional Expectation. Conditional expectation has a number of useful properties that are captured in the following results which are proved in Williams [**88**, ch. 9.3].

THEOREM 1.59. *Let $\mathcal{G} \subset \mathcal{F}$ be a proper sub-σ-algebra and let $X \in L^1(\Omega)$. Then*

(1) If Y is any version of $\mathbb{E}[X \mid \mathcal{G}]$, then $\mathbb{E}[Y] = \mathbb{E}[X]$;
(2) $\mathbb{E}[X \mid \mathcal{G}] = X$ *a.s. if X is* $\mathcal{G}$-measurable;
(3) $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$;
(4) $\mathbb{E}[k_1 X + k_2 Y \mid \mathcal{G}] = k_1 \mathbb{E}[X \mid \mathcal{G}] + k_2 \mathbb{E}[Y \mid \mathcal{G}]$ *for $Y \in L^1(\Omega)$, and real constants,* k_1 and k_2 (*linearity*);
(5) If $X \geq 0$, then $\mathbb{E}[X \mid \mathcal{G}] \geq 0$ (*positivity*) ;
(6) $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X \mid \mathcal{H}]$ *if $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$* (*tower property*) ;
(7) $\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$ *if Y is $\mathcal{G}$-measurable* ("*taking out what is known*") ;
(8) $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$ *if X is independent of $\mathcal{G}$* (*independence*); and
(9) $\mathbb{E}[\varphi(X) \mid \mathcal{G}] \geq \varphi(\mathbb{E}[X])$ *for any summable, convex function φ* (*Jensen's inequality*).

A number of important results of expectation which are proved in Section 1.7 an be extended to conditional expectation.

THEOREM 1.60. *Let $\{X_k\}$ be a countable sequence of summable, measurable random variables and let $\mathcal{G} \subset \mathcal{F}$ be a proper sub-σ-algebra.*

(1) (Conditional Fatou's lemma) If $X_k \geq 0$ *for all k, then*
$$\mathbb{E}\left[\liminf_{k\to\infty} X_k \mid \mathcal{G}\right] \leq \liminf_{k\to\infty} \mathbb{E}[X_k \mid \mathcal{G}];$$
(2) (Conditional monotone convergence theorem) If $0 \leq X_1 \leq X_2 \leq \ldots$ *a.s. such that $X_k \uparrow X$ a.s. where X is a summable random variable, then*
$$\lim_{k\uparrow\infty} \mathbb{E}[X_k \mid \mathcal{G}] \leq \mathbb{E}[X \mid \mathcal{G}];$$
(3) (Conditional dominated convergence theorem) *If X and Y are summable random variables and $X_k \to X$ a.s. where $|X_k| \leq Y$ a.s. for all k, then*
$$\lim_{k\to\infty} \mathbb{E}[X_k \mid \mathcal{G}] \leq \mathbb{E}[X \mid \mathcal{G}].$$

1.8.3. Conditional Independence. Recall that two events, $A, B \in \mathcal{F}$ are independent if
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$$
where
$$\mathbb{P}(A \cap B) = \mathbb{P}(\omega \in A \text{ and } \omega \in B).$$
Therefore, it seems natural to define independence with respect to conditioning on some event $C \in \mathcal{F}$ as
$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\,\mathbb{P}(B \mid C)$$

where

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}((\omega \in A \text{ and } \omega \in B) \mid C) = \mathbb{P}(\omega \in A \mid C \text{ and } \omega \in B \mid C).$$

This is captured in the following definition.

DEFINITION. Let $A, B_1, \ldots, B_k \in \mathcal{F}$, then the $B_1, \ldots, B_k$ are said to be ***conditionally independent given*** A if

$$\mathbb{P}\left(\bigcap_{j=1}^{k} B_j \mid A\right) = \prod_{j=1}^{k} \mathbb{P}[B_j \mid A].$$

This concept can be extended to σ-algebras as follows.

DEFINITION. Let $\mathcal{F}, \mathcal{G}_1, \ldots, \mathcal{G}_k$ be sub-σ-algebras of $\mathcal{H}$, then the $\mathcal{G}_1, \ldots, \mathcal{G}_k$ are said to be ***conditionally independent given*** $\mathcal{F}$ if

$$\mathbb{P}\left(\bigcap_{j=1}^{k} G_j \mid \mathcal{F}\right) = \prod_{j=1}^{k} \mathbb{P}[G_j \mid \mathcal{F}]$$

for all events $G_1, \ldots, G_k$ such $G_j \subseteq \mathcal{G}_j$ for $j = 1, \ldots, k$.

In particular if $\mathcal{F} = \sigma(X)$ for some random process X, then the $\mathcal{G}_1, \ldots, \mathcal{G}_k$ are said to be ***conditionally independent given*** X if the above condition holds.

Consider two σ-algebras that are conditionally independent given a third one. Then any information contained in the first should be of no benefit in regard to predicting the future state of the other once the state of the third is known and visa versa. This is capture is the following result whose proof is analogous to the non-conditional case (*cf.* Theorem 1.32).

THEOREM 1.61. *Let $\mathcal{F}, \mathcal{G}_1, \ldots, \mathcal{G}_k$ be sub-σ-algebras of $\mathcal{H}$. Then the $\mathcal{G}_1, \ldots, \mathcal{G}_k$ are conditionally independent if*

$$\mathbb{E}\left[\prod_{j=1}^{k} G_j \mid \mathcal{F}\right] = \prod_{j=1}^{k} \mathbb{E}[G_j \mid \mathcal{F}]$$

for all $G_1, \ldots, G_k$ such $G_j \subseteq \mathcal{G}_i$ for $j = 1, \ldots, k$.

PROOF. Let $\mathcal{F}, \mathcal{G}_1, \mathcal{G}_2$ be sub-σ-algebras of $\mathcal{H}$ and assume that

$$\mathbb{E}[G_1 G_2 \mid \mathcal{F}] = \mathbb{E}[G_1 \mid \mathcal{F}]\,\mathbb{E}[G_2 \mid \mathcal{F}]$$

for all $G_1 \subseteq \mathcal{G}_1$ and $G_2 \subseteq \mathcal{G}_2$.

By the definition of conditional expectation

$$\begin{aligned} \mathbb{P}(G_1 \mid \mathcal{F}) &= \mathbb{E}[1_{G_1} \mid \mathcal{F}], \\ \mathbb{P}(G_2 \mid \mathcal{F}) &= \mathbb{E}[1_{G_2} \mid \mathcal{F}], \\ \mathbb{P}(G_1 \cap G_2 \mid \mathcal{F}) &= \mathbb{E}[1_{G_1 \cap G_2} \mid \mathcal{F}]. \end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{P}(G_1 \cap G_2 \mid \mathcal{F}) &= \mathbb{E}[1_{G_1 \cap G_2} \mid \mathcal{F}] \\ &= \mathbb{E}[1_{G_1} 1_{G_2} \mid \mathcal{F}] \\ &= \mathbb{E}[1_{G_1} \mid \mathcal{F}]\,\mathbb{E}[1_{G_2} \mid \mathcal{F}] \quad \text{(by assumption)} \\ &= \mathbb{P}(G_1 \mid \mathcal{F})\,\mathbb{P}(G_2 \mid \mathcal{F}).\end{aligned}$$

Thus, $\mathcal{G}_1$ and $\mathcal{G}_2$ are conditionally independent. The full result follows by induction. □

1.9. Characteristic and Moment Generating Functions

Generally, computations involving random variables that do not admit a density function are difficult. In these cases, integral transforms can be used that completely characterize the random variable's distribution. I will consider two such transforms in this paper: the *characteristic function* and the *moment generating function*. This discussion generally follows Billingsley [**8**, sec. 26] and Bauer [**6**, ch. 4].

There is some confusion in the literature regarding the specific definition of these transforms in the context of probability theory. I have chosen to follow the above texts in this regard. □

Assume in this section that $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space and that all random variables are real-valued and $\mathcal{F}$-measurable unless otherwise noted.

1.9.1. Characteristic Functions. Assume that X has density $f \in L^1(\mathbb{R})$. Then its inverse Fourier transform exists and is given by

$$\check{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\xi x} f(x)\, dx \quad \text{(for any } \xi \in \mathbb{R}).$$

However, since $f(\cdot)$ is a density function

$$\check{f}(\xi) = \mathbb{E}\left[\frac{1}{\sqrt{2\pi}} e^{i\xi X}\right].$$

Moreover, by the Fourier inversion theorem

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \check{f}(\xi)\, e^{-i\xi x} d\xi.$$

In other words, if it exists, the density of a random variable X is the Fourier transform of $\mathbb{E}\left[\frac{1}{\sqrt{2\pi}} e^{i\xi X}\right]$. This is formalized as follows. □

DEFINITION. The ***Fourier transform*** (*a.k.a.* ***Fourier-Stieltjes transform***) of a probability measure μ is defined as

$$\hat{\mu}(x) := \int_{\mathbb{R}^n} e^{ix \cdot y} \mu(dy)$$

where $x \in \mathbb{R}^n$.

REMARK. This is generally considered to be the inverse Fourier transform in classical analysis. □

Motivation. Since, the distribution function of some real-valued random variable X is a probability measure, its Fourier transform is

$$\begin{aligned}\hat{\mathbb{P}}_X(x): &= \int_{\mathbb{R}} e^{ixy}\mathbb{P}_X(dy) \\ &= \mathbb{E}\left[e^{ixX}\right]\end{aligned}$$

where $x \in \mathbb{R}$.

This is formalized below. □

DEFINITION. The ***characteristic function of*** X is defined as

$$\varphi_X(x): = \mathbb{E}\left[e^{ix\cdot X}\right] \quad (\text{for } x \in \mathbb{R}^n).$$

REMARK. This definition is somewhat of an abuse of notation. The characteristic function is actually the Fourier transform of the distribution of X and not X itself. As such, characteristic functions are unique up to random variables that are identically distributed. □

REMARK. Note that since $e^{ix\cdot X}$ is bounded, the characteristic function always exists. □

EXAMPLE (The characteristic function of a Dirac measure). Let X be a random variable on the probability space $(\mathbb{R}^n, \mathcal{B}^n, \delta_{x_0})$ where δ_{x_0} is the Dirac measure centered at $x_0 \in \mathbb{R}^n$. Then

$$\varphi_X(x) = e^{ix\cdot x_0}.$$

□

EXAMPLE (The characteristic function of a Gaussian variable). Let $X \sim N(\mu, \sigma^2)$. Then for any $x \in \mathbb{R}$

$$\varphi_X(x) = e^{i\mu x - \frac{1}{2}\sigma^2 x^2}. \tag{1.9.1}$$

For a derivation, see Bauer [**6**, pg. 187-188]. □

The next theorem follows directly from the definitions.

THEOREM 1.62. *Let X have the distribution function F and the characteristic function φ_X. Then*

$$\begin{aligned}\varphi_X(x) &= \int_\Omega e^{ixX} d\mathbb{P} \\ &= \int_{\mathbb{R}} e^{ixy}\mathbb{P}_X(dy) \\ &= \int_{-\infty}^{\infty} e^{ixy} dF(y).\end{aligned}$$

Further, if X has the density function f, then $\varphi_X(x)$ is the inverse Fourier transform of the density function, or

$$\varphi_X(x) = \int_{-\infty}^{\infty} e^{ixy} f(y)\, dy$$

and

$$f(x) = \int_{-\infty}^{\infty} e^{-ixy} \varphi_X(y)\, dy.$$

In other words, when a density does exist, it is the characteristic function's *dual* in the sense that they are the Fourier transform of each other.

COROLLARY 1.63. *Let X have the characteristic function φ_X. If $\varphi_X \in L^1(\mathbb{R})$. Then X has a density function f defined as*

$$f(x) = \int_{-\infty}^{\infty} e^{-ixy} \varphi_X(y)\, dy.$$

PROOF. For such a f to exist, I merely have to show that it is bounded. Then for some fixed $y \in \mathbb{R}$, $|e^{-ixy}| = e^{-ixy}\widehat{e^{-ixy}} = e^{-ixy}e^{ixy} = 1$ for any $x \in \mathbb{R}$, by definition

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-ixy} \varphi_X(y)\, dy &\leq \int_{-\infty}^{\infty} \left| e^{-ixy} \varphi_X(y) \right| dy \\ &\leq \int_{-\infty}^{\infty} \left| e^{-ixy} \right| |\varphi_X(y)|\, dy \\ &\leq \int_{-\infty}^{\infty} |\varphi_X(y)|\, dy < \infty. \end{aligned}$$

Hence, $f(x) = \int_{-\infty}^{\infty} e^{-ixy} \varphi_X(y)\, dy$ is well defined. □

Characteristic functions have a number of important properties.

THEOREM 1.64. *Let X have the characteristic function φ_X. Then*

(1) $\varphi_X(0) = 1$;
(2) $|\varphi_X(x)| \leq 1$ for all $x \in \mathbb{R}$; *and*
(3) φ_X *is uniformly continuous.*

PROOF. (1) $\varphi_X(0) = 1$ follows from the fact that the mass of the sample space is one under a probability measure.

(2) As shown in the proof of Corollary 1.63, $|e^{ixy}| = 1$. Hence, for any $x \in \mathbb{R}$,

$$\begin{aligned} |\varphi_X(x)| &\leq \int_{\Omega} \left| e^{ixX} \right| d\mathbb{P} \\ &\leq 1. \end{aligned}$$

(3) By definition, φ_X is continuous. To see that it is uniformly continuous, let $h > 0$. Then using the fact that $|e^{ixy}| = 1$ for all $x \in \mathbb{R}$

$$\begin{aligned} |\varphi_X(x+h) - \varphi_X(x)| &= \left| \int_\Omega e^{i(x+h)X} - e^{ixX} d\mathbb{P} \right| \\ &\leq \int_\Omega \left| e^{ixX} \left(e^{ihX} - 1 \right) \right| d\mathbb{P} \\ &\leq \int_\Omega \left| e^{ixX} \right| \left| \left(e^{ihX} - 1 \right) \right| d\mathbb{P} \\ &\leq \int_\Omega \left| e^{ihX} - 1 \right| d\mathbb{P}. \end{aligned}$$

Now

$$e^{ihx} \to 1 \text{ as } h \to 0 \text{ and } \left| e^{ihx} \right| = 1 \quad (\text{for all } x \in \mathbb{R}).$$

Then by the bounded convergence (Theorem 1.52), the right-hand side converges to zero as $h \to \infty$ independently from x. Thus, φ_X is uniformly continuous. □

THEOREM 1.65. *Let $\{X_j\}_{j=1}^m$ be a finitely countable collection of random variables. Then for any $x \in \mathbb{R}$,*

$$\varphi_{X_1+\cdots+X_m}(x) = \prod_{j=1}^m \varphi_{X_j}(x).$$

PROOF. Let X and Y be independent real-valued, measurable random variables. Then

$$\begin{aligned} \varphi_{X+Y}(x) &= \mathbb{E}\left[e^{ix(X+Y)} \right] \\ &= \mathbb{E}\left[e^{ixX} e^{ixY} \right] \\ &= \mathbb{E}\left[e^{ixX} \right] \mathbb{E}\left[e^{ixY} \right] \\ &= \varphi_X(x) \varphi_Y(x). \end{aligned}$$

The result follows by induction by defining $X = \sum_{j=1}^k X_j$ and $Y = X_{k+1}$ for some $k < m$. □

As stated earlier, the advantage of a characteristic function is that it exists for any random variable even when the random variable does not have a density. Moreover, it is unique in the *weak* sense (*i.e.* up to modulo distribution). In other words, given φ_X for a real-valued random variable X, it is possible to determine F_X. This follows from a result known as the *inversion theorem* which, along with the following corollary, are proved in Billingsley[**8**, pg. 369-370].

THEOREM 1.66 (Inversion theorem). *Let X have the distribution function F_X and characteristic function φ_X. Then*

$$F_X(a,b] = \lim_{h \to \infty} \frac{1}{2\pi} \int_{-h}^h \frac{e^{-ixa} - e^{-ixb}}{ix} \varphi_X(x)\, dx.$$

COROLLARY 1.67 (Uniqueness of the characteristic function modulo distribution). *If Y is a real-valued, measurable random variable, then $X \overset{d}{=} Y$ if and only if $\varphi_X(x) = \varphi_Y(x)$ for all $x \in \mathbb{R}$.*

This corollary points to one the most useful aspects of characteristic functions which is captured in the following result that is proved in Billingsley [**8**, pg. 372].

THEOREM 1.68. *Let $\{X_m\}$ be a sequence of real-valued, measurable random variables. Then for some real-valued, measurable random variable X*

$$\lim_{m\to\infty} X_m \to X \text{ if and only if } \lim_{m\to\infty} \varphi_{X_m}(x) \to \varphi_X(x) \quad (\text{for all } x \in \mathbb{R}).$$

THEOREM 1.69. *Let X have the distribution function F_X and characteristic function $\varphi_X \in L^1(\mathbb{R})$. Then F_X has density f of the form*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixy}\varphi(y)\,dy.$$

Moreover, f is continuous.

PROOF. Let X be a real-valued random variable with the distribution function F_X and characteristic function $\varphi_X \in L^1(\mathbb{R})$. The plan is to use the inversion theorem to compute the density. First, I need to show that in the limit, the integral is defined over all of $\mathbb{R}$. In other words, I need to show

$$\lim_{h\to\infty}\int_{-h}^{h}\frac{e^{-ixa}-e^{-ixb}}{ix}\varphi_X(x)\,dx = \int_{-\infty}^{\infty}\frac{e^{-ixa}-e^{-ixb}}{ix}\varphi_X(x)\,dx < \infty.$$

Now,

$$\lim_{h\to\infty}\left|\int_{-h}^{h}\frac{e^{-ixa}-e^{-ixb}}{ix}\varphi_X(x)\right|dx \le \int_{-\infty}^{\infty}\left|\frac{e^{-ixa}-e^{-ixb}}{ix}\right||\varphi_X(x)|\,dx.$$

But,

$$\begin{aligned}\left|\frac{e^{-ixa}-e^{-ixb}}{ix}\right| &= \left|\frac{i^2e^{-ixa}\left(e^{-ix(b-a)}-1\right)}{ix}\right| \\ &= \left|e^{-ixa}\right|\left|\frac{\left(e^{-ix(b-a)}-1\right)}{x}\right| \\ &= \left|\frac{\left(e^{-ix(b-a)}-1\right)}{x}\right|.\end{aligned}$$

Using Taylor's formula,

$$\left|\frac{\left(e^{-ix(b-a)}-1\right)}{x}\right| \le |b-a|.$$

Hence,

$$\begin{aligned}\lim_{h\to\infty}\left|\int_{-h}^{h}\frac{e^{-ixa}-e^{-ixb}}{ix}\varphi_X(x)\right|dx &\leq \int_{-\infty}^{\infty}\left|\frac{e^{-ixa}-e^{-ixb}}{ix}\right||\varphi_X(x)|\,dx\\ &\leq |b-a|\int_{-\infty}^{\infty}|\varphi_X(x)|\,dx<\infty.\end{aligned}$$

Since $\varphi_X \in L^1(\mathbb{R})$, the right-hand side is bounded for a and b finite. Thus,

$$\lim_{h\to\infty}\int_{-h}^{h}\frac{e^{-ixa}-e^{-ixb}}{ix}\varphi_X(x)\,dx=\int_{-\infty}^{\infty}\frac{e^{-ixa}-e^{-ixb}}{ix}\varphi_X(x)\,dx<\infty.$$

Consequently, I can compute the derivative of the distribution function using the fact the integrand is continuous as follows. For some $x\in\mathbb{R}$, by the inversion theorem (*cf.* Theorem 1.66)

$$\begin{aligned}\lim_{h\to 0}\frac{F_X(x+h)-F_X(h)}{h} &= \frac{1}{2\pi}\lim_{h\to 0}\int_{-\infty}^{\infty}\frac{e^{-ixy}-e^{-i(x+h)y}}{ihy}\varphi_X(y)\,dy\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\lim_{h\to 0}\frac{e^{-ixy}-e^{-i(x+h)y}}{ihy}\varphi_X(y)\,dy\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\lim_{h\to 0}\frac{e^{-ixy}-e^{-ixy}e^{-ihy}}{ihy}\varphi_X(y)\,dy\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\lim_{h\to 0}\frac{\left(1-e^{-ihy}\right)}{ihy}e^{-ixy}\varphi_X(y)\,dy\end{aligned}$$

By L'Hôpital's rule (*cf.* [**78**, pg. 204])

$$\lim_{h\to 0}\frac{\left(1-e^{-ihy}\right)}{iyh} = 1.$$

Thus,

$$\lim_{h\to 0}\frac{F_X(x+h)-F_X(h)}{h} = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-ixy}\varphi_X(y)\,dy<\infty \quad (\text{for all } x\in\mathbb{R}).$$

Hence, F_X has a density. Letting,

$$f(x):=\lim_{h\to 0}\frac{F_X(x+h)-F_X(h)}{h}$$

then

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-ixy}\varphi_X(y)\,dy.$$

Moreover, since φ_X is continuous, f is continuous. By the fundamental theorem of calculus (*cf.* [**78**, pg. 285-287]), f is unique and

$$F_X(x) = \int_{-\infty}^{x} f(y)\, dy < \infty \quad (\text{for all } x \in \mathbb{R}).$$

Hence, f is the density of F. □

Finally, the characteristic function of random variable completely determines its moments and, hence, the random variable itself.

THEOREM 1.70. *Let X have the characteristic function φ_X with countably many (possibly infinite) derivatives. Then if the moments $\mathbb{E}\left[X^k\right]$ exist for $k = 1, \ldots$*

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}\left[X^k\right] \quad (k = 1, \ldots)$$

PROOF. Let $k = 1$. Then for any $x \in \mathbb{R}$,

$$\begin{aligned} \varphi_X'(x) &= \left(\mathbb{E}\left[e^{ixX}\right]\right)', \\ &= \int_\Omega iXe^{ixX} d\mathbb{P} \\ &= i\mathbb{E}\left[Xe^{ixX}\right] \end{aligned}$$

where "$'$"$= \dfrac{d}{dx}$.

Hence,

$$\varphi_X'(0) = i\mathbb{E}[X].$$

A similar argument holds for $k > 1$. □

This leads directly to the following result which characterizes the regularity of characteristic functions.

COROLLARY 1.71. *Let X have m moments. Then its characteristic function (near zero) is of the form*

$$\varphi_X(x) = \sum_{k=1}^{m} \frac{(ix)^k}{k!} \mathbb{E}\left[X^k\right] \quad (\textit{for } x \in \mathbb{R}).$$

REMARK. This is essentially a Taylor expansion (*cf.* [**78**, pg. 511]) which shows that random variables are generally infinite dimensional. □

The following theorem is helpful in practice.

THEOREM 1.72. *Let $\{X_j\}_{j=1}^m$ be countable collection of independent, real-valued, measurable random variables with characteristic functions φ_{X_j} for $j = 1, \ldots, m$. Then*

$$\varphi_{X_1+\cdots+X_m}(x) = \prod_{j=1}^{k} \varphi_{X_k}(x).$$

PROOF. Let $S_m = \sum_{j=1}^m X_j$. Then for any $x \in \mathbb{R}$,

$$\begin{aligned}
\varphi_{S_m}(x) &= \mathbb{E}\left[e^{ixS_m}\right] \\
&= \mathbb{E}\left[e^{ix\sum_{j=1}^m X_j}\right] \\
&= \mathbb{E}\left[\prod_{j=1}^m e^{ixX_j}\right] \\
&= \prod_{j=1}^k \mathbb{E}\left[e^{ixX_j}\right] \quad \text{(the } X_j \text{ are independent)} \\
&= \prod_{j=1}^k \varphi_{X_k}(x).
\end{aligned}$$

□

1.9.2. Moment Generating Functions. As discussed above, the characteristic function as the advantage that it always exist even when a density function does not. However, it involves complex numbers. In many cases, it is possible to characterize the distribution of a non-negative random variable using a form of the *Laplace transform* known as the *moment generating function.*

DEFINITION. Let f be a real-valued function on the non-negative reals $\mathbb{R}_+$, the (***one-sided***) ***Laplace transform of*** f is given by

$$L(x) := \int_0^\infty e^{-xy} f(y)\, dy \quad \text{(for any } x \in \mathbb{R}_+)$$

and

$$L(x) = \int_{-\infty}^\infty e^{-xy} f(y)\, dy$$

is known as the ***bilateral Laplace transform***.

Motivation. If f is the density of some real-valued, random variable X that is positive *a.s.,* then for any $x \in \mathbb{R}_+$, then the bilinear Laplace transform is

$$\begin{aligned}
L(x) &= \int_{-\infty}^\infty e^{-xy} f(y)\, dy \\
&= \mathbb{E}\left[e^{-xX}\right].
\end{aligned}$$

Moreover, the derivatives of the Laplace transform give the moments of X, since

$$L^{(k)}(x) = (-1)^k \int_{-\infty}^\infty y^k e^{-xy} f(y)\, dy \quad \text{(for } x \in \mathbb{R}_+).$$

Hence,

$$\begin{aligned} L^{(k)}(0) &= (-1)^k \int_{-\infty}^{\infty} y^k f(y)\, dy \\ &= (-1)^k \mathbb{E}\left[X^k\right]. \end{aligned}$$

This is formalized below. □

DEFINITION. If X is an almost surely non-negative, measurable random variable with the distribution function F_X, then the ***moment generating function*** of X is defined as

$$M_X(x) := \int_0^{\infty} \cdots \int_0^{\infty} e^{x \cdot y} dF_X(y) \quad \left(\text{for } x \in \mathbb{R}_+^n\right).$$

REMARK. Under this definition, the moment generating function is typically said to be the Laplace transform of X. Again, this is somewhat of an abuse of notation since $M_{\mathbf{X}}(-x)$ is actually the Laplace transform of the *distribution function* of X. □

Given a real-valued, measurable random variable X such that $\mathbb{P}(X < 0) = 0$, upon inspection

$$\varphi_X(x) = M_{iX}(x) = M_X(ix). \tag{1.9.2}$$

In other words, the characteristic function is the moment generating function of the complex-valued, measurable random variable iX. Equivalently, the characteristic function is the moment generating function confined to the complex plane.

Given the relationship expressed in 1.9.2, the moments of a real-valued random variable can be expressed in terms of the derivatives of the moment generating function.

THEOREM 1.73. *Let X be an almost surely non-negative, measurable random variable. Then*

$$M_X^{(k)}(0) = (-1)^k \mathbb{E}\left[X^k\right] \quad (k = 1, \ldots)$$

PROOF. For any $k = 1,\ldots$ and $x \in \mathbb{R}_+^n$, by (1.9.2)

$$\begin{aligned} M_X^{(k)}(x) &= i^k \varphi_X^{(k)}(ix) \\ &= i^{2k} \mathbb{E}\left[X^k e^{ixX}\right]. \end{aligned}$$

by Theorem 1.70.

In particular,

$$\begin{aligned} M_X^{(k)}(0) &= i^k \varphi_X^{(k)}(0) \\ &= i^{2k} \mathbb{E}\left[X^k\right] \\ &= (-1)^k \mathbb{E}\left[X^k\right]. \end{aligned}$$

□

COROLLARY 1.74. *Let X be an almost surely non-negative, measurable random variable with m moments. Then its moment generating function (near zero) is of the form*

$$M_X(x) = \sum_{k=1}^{m} \frac{(-x)^k}{k!} \mathbb{E}\left[X^k\right].$$

The following two theorems also follow immediately from (1.9.2).

THEOREM 1.75. *If X and Y are almost surely non-negative, measurable random variables, then $X \stackrel{d}{=} Y$ if and only if $M_X(x) = M_Y(x)$ for all $x \in \mathbb{R}_+$.*

THEOREM 1.76. *Let $\{X_j\}_{j=1}^m$ be independent, almost surely non-negative, measurable random variables with characteristic functions, M_{X_j} for $j = 1, \cdots, m$. Then for any $x \in \mathbb{R}^+$*

$$M_{X_1+\cdots+X_m}(x) = \prod_{j=1}^{k} M_{X_k}(x).$$

1.10. Convergence of Random Variables

Considering random variables as functions, there are number of notions of convergence that are commonly used in continuous-time finance. These are captured in the following definitions.

Assume in this section that $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space and that all random variables are real-valued and $\mathcal{F}$-measurable unless otherwise noted.

DEFINITION. Let X have the distribution function F. A countable sequence of real-valued, measurable random variables $\{X_k\}$ with distribution functions F_k for each $k = 1, \ldots$ is said to ***converge in distribution*** (*a.k.a.* ***weakly***) ***to*** X, denoted as $X_k \xrightarrow{d} X$, if

$$\lim_{k\to\infty} F_k(x) = F(x) \quad (\text{for all } x \in \mathbb{R}).$$

Alternatively, $X_k \xrightarrow{d} X$ if

$$\lim_{k\to\infty} \mathbb{P}(X_k \le x) = \mathbb{P}(X_k \le x) \quad (\text{for all } x \in \mathbb{R}).$$

DEFINITION. A countable sequence of real-valued, measurable random variables $\{X_k\}$ is said to ***converge in probability to*** X, denoted as $X_k \xrightarrow{p} X$, if for every $\varepsilon > 0$,

$$\lim_{k\to\infty} \mathbb{P}(|X_k - X| > \varepsilon) = \lim_{k\to\infty} \mathbb{P}\{\omega \in \Omega \mid |X_k(\omega) - X(\omega)| > \varepsilon\} = 0.$$

The following result, which is proved in Billingsley (*cf.* [**8**, pg. 353]), shows that convergence in probability implies convergence in distribution. However, the converse is false (*cf.* [**14**, pg. 110]).

THEOREM 1.77. *Let $\{X_k\}$ be a countable sequence of real-valued, measurable random variables such that $X_k \xrightarrow{p} X$. Then $X_k \xrightarrow{d} X$.*

Another notion of the convergence of random variables is analogous to almost everywhere convergence in measure theory. Here, I need to use the fact that the liminf and limsup of a sequence of random variables are also random variables. For a rigorous proof, see Durrett [**23**, pg. 15].

THEOREM 1.78. *Let* $\{X_k\}$ *a sequence of real-valued, measurable random variables. Then* $\liminf_{k\to\infty} X_k$ *and* $\limsup_{k\to\infty} X_k$ *are also real-valued, measurable random variables. In fact,*

$$\liminf_{k\to\infty} X_k = \sup_{k\geq 0}\left(\inf_{j\geq k}\left(X_j\right)\right) \quad \textit{and} \quad \liminf_{k\to\infty} X_k = \inf_{k\geq 0}\left(\sup_{j\geq k}\left(X_j\right)\right).$$

This leads to the standard definition of almost surely convergence.

DEFINITION. A countable sequence of real-valued, measurable random variables $\{X_k\}$ is said to ***converge almost surely to*** X, denoted as $X_k \to X$ a.s., if

$$\mathbb{P}\left\{\omega\in\Omega \mid \liminf_{k\to\infty} X_k(\omega) = \limsup_{k\to\infty} X_k(\omega)\right\} = 1.$$

Alternatively, $X_k \to X$ a.s. if

$$\mathbb{P}\left\{\omega\in\Omega \mid \lim_{k\to\infty} X_k(\omega) = X(\omega)\right\} = 1.$$

The following result shows that almost surely convergence implies convergence in probability, hence, in distribution as well. However, the converse is false (*cf.* [**14**, pg. 102]).

THEOREM 1.79. *Let* $\{X_k\}$ *be a countable sequence of real-valued, measurable random variables such that* $X_k \to X$ *a.s.. Then* $X_k \xrightarrow{p} X$ *and* $X_k \xrightarrow{d} X$.

Given the fact that the collection of measurable of random variables $L^p(\Omega)$ forms a Banach space under the L^p-norm, it seems to be reasonable to define convergence under this norm.

DEFINITION. A countable sequence of real-valued, measurable random variables $\{X_k\}$ is said to ***converge in*** L^p ***to*** X for any $1 \leq p \leq \infty$, denoted as $X_k \xrightarrow{L^p} X$, if X and $X_k \in L^p(\Omega)$ for all k and

$$\lim_{k\to\infty} \|X_k - X\|_{L^p(\Omega)} = 0.$$

Alternatively, $X_k \xrightarrow{L^p} X$ if

$$\lim_{k\to\infty} \mathbb{E}\left[|X_k - X|^p\right] = 0.$$

If $p = 1$ and $X_k \xrightarrow{L^1} X$, the $\{X_k\}$ are said to ***converge in mean to*** X.

The following result shows that L^p convergence implies convergence in probability, hence, in distribution as well. Again, the converse is false (*cf.* [**14**, pg. 106-107]).

THEOREM 1.80. *Let $\{X_k\}$ be a countable sequence of real-valued, measurable random variables such that $X_k \xrightarrow{L^p} X$ for $1 \leq p \leq \infty$. Then $X_k \xrightarrow{p} X$ and and $X_k \xrightarrow{d} X$.*

The next results comes directly from the properties of general L^p spaces.

THEOREM 1.81. *For any $1 \leq q \leq p \leq \infty$, if $\{X_k\}$ is a countable sequence of real-valued, measurable random variables such that X and $X_k \in L^p(\Omega) \cap L^q(\Omega)$ and $X_k \xrightarrow{L^p} X$, then $X_k \xrightarrow{L^q} X$.*

PROOF. Let $Y \in L^p(\Omega) \cap L^q(\Omega)$ be some random variable where $1 \leq q \leq p < \infty$. Further, define $\varphi(x) = |x|^{p/q}$ for $x \in \mathbb{R}$. Since, $p \geq q$, φ is convex. Thus, I can apply Jensen's inequality (*cf.* Theorem 1.48).

$$\begin{aligned}(\mathbb{E}[|Y|^q])^{p/q} &\leq \mathbb{E}\left[(|Y|^q)^{p/q}\right] \\ &\leq \mathbb{E}[|Y|^p].\end{aligned}$$

Therefore,

$$(\mathbb{E}[|Y|^q])^{1/q} \leq (\mathbb{E}[|Y|^p])^{1/p}$$

or

$$\|Y\|_{L^q(\Omega)} \leq \|Y\|_{L^p(\Omega)}.$$

Now, if $\{X_k\}$ is a countable sequence of real-valued, measurable random variables such that X and $X_k \in L^p(\Omega) \cap L^q(\Omega)$ and $X_k \xrightarrow{L^p} X$, then

$$\lim_{k\to\infty} \|X_k - X\|_{L^p(\Omega)} = 0.$$

But, for each k, by the above calculation,

$$\|X_k - X\|_{L^p(\Omega)} \geq \|X_k - X\|_{L^q(\Omega)} \geq 0.$$

Therefore,

$$\lim_{k\to\infty} \|X_k - X\|_{L^q(\Omega)} = 0.$$

If $p = \infty$,

$$\begin{aligned}\|Y\|_{L^q(\Omega)} &= \left(\int |Y|^q \, d\mathbb{P}(\omega)\right)^{1/q} \\ &\leq \operatorname{ess\,sup}_\Omega |Y| \left(\int d\mathbb{P}(\omega)\right)^{1/q} \\ &\leq \|Y\|_{L^\infty(\Omega)}\end{aligned}$$

and the result follows as above. □

The next result will be helpful in the sequel and it is proved in Çinlar [**14**, pg. 106-107]).

THEOREM 1.82. *Let $\{X_k\}$ be a uniformly integrable, countable sequence of real-valued, measurable random variables such that $X_k \xrightarrow{p} X$. Then X is summable and $X_k \xrightarrow{L^1} X$.*

1.11. Change of Probability Measure

As I show in the sequel, a core tenet of continuous-time finance is valuing a financial instrument under a risk-neutral probability measure (*cf.* Section 4.6). The construction of such a measure requires a number of results, some of which are stated below. However, many are very technical and beyond the scope of this paper. See Shreve [**76**, ch. 1.6] for a rigorous treatment of the subject with a good discussion in the context of continuous-time finance.

Assume in this section that $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space and that all random variables are real-valued and $\mathcal{F}$-measurable unless otherwise noted.

Motivation. In standard n-dimensional integration, a change of variables for some nice real-valued functions f and ψ on some open set $U \subseteq \mathbb{R}^n$ can generally be expressed as

$$\int_{\psi(U)} f(y)\, dy = \int_U f(\psi(x)) \left|\det D\psi(x)\right| dx.$$

In terms of differentials,

$$dy = Z(\psi(x))\, dx$$

where Z is the Jacobian ($Z(\psi(x)) = \left|\det D\psi(x)\right|$).

Thus, it seems reasonable to define a change in probability measure as:

$$d\mathbb{Q} = Z d\mathbb{P} \tag{1.11.1}$$

where $\mathbb{Q}$ and $\mathbb{P}$ are probability measures on the same measurable space and Z is some appropriate random variable playing the role of the Jacobian. This is expressed in the following theorem whose proof can be found in Shreve [**76**, pgs. 33-34]. □

THEOREM 1.83. *Let Z be a real-valued, measurable random variable such that $Z \geq 0$ a.s. with $\mathbb{E}_{\mathbb{P}}[Z] = 1$. Further, for all $A \in \mathcal{F}$, define*

$$\mathbb{Q}(A) := \int_A Z d\mathbb{P}.$$

Then $\mathbb{Q}$ is a probability measure equivalent to $\mathbb{P}$. Additionally, if X is summable, then

$$\mathbb{E}_{\mathbb{Q}}[X] = \mathbb{E}_{\mathbb{P}}[XZ]. \tag{1.11.2}$$

Moreover, if $Z > 0$ a.s., then

$$\mathbb{E}_{\mathbb{P}}[X] = \mathbb{E}_{\mathbb{Q}}\left[\frac{X}{Z}\right]. \tag{1.11.3}$$

The next result follows immediately from the above theorem.

COROLLARY 1.84. *If Z and X are independent, then $\mathbb{E}_{\mathbb{Q}}[X] = \mathbb{E}_{\mathbb{P}}[X]$.*

DEFINITION. Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on the same measurable space. $\mathbb{Q}$ is said to be ***absolute continuous with respect to*** $\mathbb{P}$ if $\mathbb{P}(A) = 0$ implies $\mathbb{Q}(A) = 0$ for all $A \in \mathcal{F}$. The measures are said to be ***equivalent*** if they have the same null-sets (*i.e.* $\mathbb{P}(A) = 0$ if and only if $\mathbb{Q}(A) = 0$ for $A \in \Omega$).

REMARK. Null-sets and certain events are invariant under equivalent measures. □

THEOREM 1.85. *If $\mathbb{P}$ and $\mathbb{Q}$ are equivalent probability measures, then*

$$\mathbb{P}(A) = 1 \text{ if and only if } \mathbb{Q}(A) = 1 \quad (\text{for } A \in \mathcal{F}).$$

PROOF. The proof follows from the fact that if an event has probability zero, its complement has probability one. Hence, if $\mathbb{P}$ and $\mathbb{Q}$ agree on any event with probability zero, they will also agree on any event with probability one. □

REMARK. If two measures are equivalent, they agree on what events are certain and on what events are impossible. □

Motivation. Theorem 1.83 provides a clue into the random variable needed to play the part of the Jacobian in the change of measure formula (*cf.* (1.11.1)). Rewriting (1.11.2) as

$$\int_{\Omega} X d\mathbb{Q} = \int_{\Omega} XZ d\mathbb{P},$$

it seems natural to define Z as

$$Z = \frac{d\mathbb{Q}}{d\mathbb{P}}$$

whatever that may mean. This is formalized in the following definition. □

DEFINITION. Let $\mathbb{P}$ and $\mathbb{Q}$ be two equivalent probability measures and let Z be the random variable implicitly defined as

$$\mathbb{Q}(A) := \int_A Z d\mathbb{P} \quad (\text{for all } A \in \mathcal{F}).$$

Then Z is called the ***Radon-Nikodým derivative of*** $\mathbb{Q}$ ***with respect to*** $\mathbb{P}$ and is denoted as

$$Z = \frac{d\mathbb{Q}}{d\mathbb{P}}. \tag{1.11.4}$$

The following result proves the existence and uniqueness of such a random variable. See Çinlar [**14**, pgs. 208-210] for a rigorous proof.

THEOREM 1.86 (Radon-Nikodým theorem). *Let $\mathbb{P}$ and $\mathbb{Q}$ be two equivalent probability measures. Then there exists a unique random variable Z such that*

(1) *$Z > 0$ a.s.*;

(2) $\mathbb{E}_{\mathbb{P}}[Z] = 1$; *and*
(3) $\mathbb{Q}(A) = \int_A Z d\mathbb{P}$ (*for all* $A \in \mathcal{F}$).

Given sufficient regularity, the Radon-Nikodým theorem can be expressed in terms of densities for a given real-valued random variable.

COROLLARY 1.87. *Let* $\mathbb{P}$ *and* $\mathbb{Q}$ *be two equivalent probability measures. Further, let* X *have densities* f *and* g *under* $\mathbb{P}$ *and* $\mathbb{Q}$ *respectively. Then there exists a unique random variable such that* $Z > 0$ *a.s.,* $\mathbb{E}_{\mathbb{P}}[Z] = 1$*, and*

$$Z = \frac{f(X)}{g(X)}.$$

CHAPTER 2

Stochastic Processes

Stochastic processes provide the means to describe time-dependent random variables such the price of a financial instrument. As such, they form the core of the mathematical foundations of continuous-time finance. First, however, it is necessary to formalize how to mathematically express the evolution of information with time.

2.1. Information Evolution

As mentioned earlier, the concept of information is a central character in continuous-time finance. In the context of probability theory, we saw that information is expressed in terms of a family of observable events captured in the form of a σ-algebra. Intuitively, however, information is not static, but becomes richer over time as investors learn. In probability theory, such an evolution of information is described by a growing family of σ-algebras known as a *filtration*. It is reasonable to assume that investors' expectations regarding future markets will change when presented with such evolving information.

Before I jump into information evolution, I need to formalize what is meant by time in a probabilistic context. Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space in this section.

2.1.1. Time Indexes.

DEFINITION. A ***time index*** is a member of the ordered set of unique elements known as the (***time) index set*** that may be any abstract set. In the case that the index set is (possibly infinitely) countable, it is called a ***discrete-time set*** (*e.g.* $\mathbb{T} = \{t_0, t_1, \ldots, t_m = T\}$) where $t_k < t_{k+1}$ for $k = 0, \ldots, m$. In the case that the index set is an (possible unbounded) interval, it is called ***continuous-time set*** (*e.g.* $\mathbb{T} = [t_0, T]$). The upper bound of the index set T is called the ***time horizon*** which may be infinite.

NOTATION. Unless I need to distinguish finite from infinite time horizons, I will simply use the symbol, $\mathbb{T}$, to indicate the use of a time index throughout this paper. Further, I will assume, without loss of generality, that $t_0 = 0$ and that index set is closed. Lastly, I will generally use the symbol "k" to indicate a discrete-time and the symbol "t" to indicate continuous-time. □

DEFINITION. Given a time index set $\mathbb{T}$, a ***partition*** Π is ordered finite collection of distinct points in $\mathbb{T}$ including both endpoints $\{0 = t_0 < t_1 < \cdots < t_k = T\}$ where $t_j \in \mathbb{T}$ for $0 \leq j \leq k$. Further, the ***mesh size*** of a partition Π is defined as

$$|\Pi| = \max_{0 \leq j \leq k-1} |t_{j+1} - t_j| .$$

Lastly, if Π and $\tilde{\Pi}$ are two partitions of some index set $\mathbb{T}$, Π is said to be ***finer*** than $\tilde{\Pi}$ if $|\Pi| < \left|\tilde{\Pi}\right|$.

2.1.2. Filtrations. With the introduction of time, it is natural to ask how best to describe the evolution of information over time. Recall that for a given probability space, the observable information is captured by its σ-algebra which is a static family of sets in the sample space. Therefore, it seems reasonable that evolving information should is captured by a family of σ-algebras *increasing* with time.

DEFINITION. The family of σ-algebras $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is said to be ***increasing*** if $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s, t \in \mathbb{T}$ such that $s \leq t$.

In other words, if a family of σ-algebras is increasing, the observable events (hence, information) increase with time.

DEFINITION. A ***filtration***, denoted as $F = \{\mathcal{F}_t\}_{t\in\mathbb{T}}$, is an increasing family of σ-algebras on the same measurable space $(\Omega, \mathcal{F})$ such that $\mathcal{F}_t \subseteq \mathcal{F}$ for all $t \in \mathbb{T}$.

The next result follows directly from the fact that σ-algebras are closed under countable unions by definition (*cf.* on page 2).

THEOREM 2.1. *Let $F = \{\mathcal{F}_t\}_{t\in\mathbb{T}}$ be a filtration. Then the union of the σ-algebras $\mathcal{F}_t$, denoted as $\mathcal{F}_\infty := \cup_{t\in\mathbb{T}}\mathcal{F}_t$, is a σ-algebra and is called the* **σ-algebra generated by** F.

REMARK. By this definition, for an event $A \in \mathcal{F}_s$, then $A \in \mathcal{F}_t$ *for all* $s \leq t$ and $A \in \mathcal{F}$. In other words, once an event becomes observable, it is observable for all time. □

EXAMPLE. Assume that the sample space Ω corresponds to the sides of coin $\{\{H\}, \{T\}\}$ and the events are $\{\emptyset, \Omega, \{H\}, \{T\}\}$. We can now construct a (discrete) filtration $F = (\mathcal{F}_k)_{k\in\mathbb{T}}$ as follows. Let $\{a_i\}_{i=0}^k$ be the sequence of the outcome after k tosses. Then $\{a_1 = \{H\}, a_2 = \{H\} \ldots, a_j = \{H\}\}$ is an event in the σ-algebra $\mathcal{F}_j$ for all $j \leq k$. Moreover, it is a event in the σ-algebra $\mathcal{F}_k$ since all earlier events are observable. Thus, $\mathcal{F}_j \subseteq \mathcal{F}_k$ for all $j \leq k$.

On the other hand, the event $\{a_1 = \{H\}, a_2 = \{H\}, \ldots, a_k = \{H\}\}$ is not in the σ-algebra $\mathcal{F}_j$ for any for $j < k$. Thus, $\mathcal{F}_k \nsubseteq \mathcal{F}_j$ for all $j \leq k$. In other words, there is more information $\mathcal{F}_k$ than $\mathcal{F}_j$. □

DEFINITION. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $F = \{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is called a ***filtered probability space*** and is denoted as $(\Omega, \mathcal{F}, F, \mathbb{P})$.

In order to preserve many of the measure-theoretic results, I also need to ensure that $\mathcal{F}_t$ is complete for each $t \in \mathbb{T}$. This is captured in the following definition.

DEFINITION. A filtration $F = \{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is said to be ***augmented*** if $\mathcal{F}_0$ is complete, or if $A \subset \Omega$ and $\mathbb{P}(A) = 0$, then $A \in \mathcal{F}_0$.

Motivation. To be a useful tool in the context of continuous-time finance, we should require any filtration to be *non-anticipating* in the sense that we should not be able to *see* into the future. In other words, for the filtration $F = \{\mathcal{F}_t\}_{t\in\mathbb{T}}$, information about future states should not be available at earlier times, or

$$\mathcal{F}_t \cap \mathcal{F}_s = \mathcal{F}_s \text{ if } s \leq t.$$

This is captured in the following definition. □

DEFINITION. Let $F = \{\mathcal{F}_t\}_{t\in\mathbb{T}}$ be a filtration on the measurable space $(\Omega, \mathcal{F})$. Then for all $s, t \in \mathbb{T}$ with $s \geq t$, set

$$\mathcal{F}_{t^+} := \bigcap_{s>t} \mathcal{F}_s.$$

Then F is said to be ***right-continuous*** if $\mathcal{F}_{t^+} = \mathcal{F}_t$ for all $t \in \mathbb{T}$.

Similarly, set

$$\mathcal{F}_{t^-} = \begin{cases} \sigma\left(\bigcup_{s<t} \mathcal{F}_t\right), & t > 0 \\ \mathcal{F}_0, & t = 0. \end{cases}$$

Then F is said to be ***left continuous*** if $\mathcal{F}_{t^-} = \mathcal{F}_t$ for all $t \in \mathbb{T}$.

Under these definitions, a right-continuous filtration is non-anticipating in the sense that there is no material gain in information for a infinitesimal peek into the future. On the other hand, if the filtration was not right-continuous, there are instantaneous jumps in the amount of observable information at specific points in time. In other words, a small peek into the future would yield material rewards. Additionally, I should require that all impossible and certain outcomes are known at the outset. Given this, I will generally insist that all filtrations are right-continuous and augmented which is captured in the following definition.

DEFINITION. A filtration $F = \{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is said to have the ***usual conditions*** (*a.k.a.* the ***usual hypothesis***) if F is ***right continuous*** and F is ***augmented***.

Filtrations with the usual conditions play a key role in continuous-time finance since they represent the evolution of observable information available to investors with time. Because a filtration is increasing with time (*i.e.* additional information becomes available with time), investors gain from experience. However, there is no point to trying to *see* into the future since no material information can be gained by looking an infinitesimal amount into the future. In short, the return from any investment other than one in the riskless asset is uncertain (*i.e.* it is *non-anticipatory*). Finally, since any filtration with the usual conditions is augmented, all impossible events are known to all investors at the outset. Hence, rational investors will never make investments based on the anticipation of such events occurring.

For the remainder of the paper, I will assume all filtrations have the usual conditions.

2.2. General Properties of Stochastic Processes

A stochastic process is a mathematical structure used to model the occurrence random events over time. As such, they are well suited to describing non-deterministic evolution systems such as the price of a financial instrument. In fact, much of modern finance involves the study of various classes of stochastic processes. As usual, I shall begin with some basic definitions.

Assume that $\mathbb{T}$ is some arbitrary time index set which may be discrete or infinite as well as bounded or unbounded and let $(\Omega, \mathcal{F}, F, \mathbb{P})$ be a given filtered probability space unless stated otherwise.

2.2.1. Basic Definitions.

DEFINITION 2.2. A ***stochastic process*** X is an ordered family of random variables each from the measurable space $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$ indexed by the set $\mathbb{T}$ and is denoted as

$$X = (X(t))_{t \in \mathbb{T}}.$$

If the index set is countable, the process is said to be ***a discrete-time process***. Conversely, if the index set is an interval (possibly infinite), the process is said to be ***continuous in time*** or a ***continuous-time process***.

If the stochastic process is vector-valued, then it is an ordered family of random vectors from the measurable space $(\Omega, \mathcal{F})^n$ to $(E, \mathcal{E})^n$ indexed by the set $\mathbb{T}$ and is denoted as

$$X = (X(t))_{t \in \mathbb{T}} = \left(X^1(t), \ldots, X^n(t)\right)_{t \in \mathbb{T}}.$$

NOTATION. Since $X(t)$ is a random variable for each t, a more correct notation would be $X(\omega(t), t)$ to show the dependency on the sample space and time. However, this is commonly suppressed as in the case of a normal random variable. Additionally, $X_t := X(t)$ is the common notation used to describe a stochastic process's value at time t. However, this short-hand notation can also be interpreted as the partial derivative with respect to time. Given the focus of this paper, I have chosen the notation $X(t)$ for continuous-time processes to remove any confusion. □

EXAMPLE (Symmetric random walk). In a discrete-time setting, a stochastic process can be defined in terms of countable sequence of *independent, identically distributed* (*i.i.d.*) random variables on some given sample space. In effect, such a process can be viewed a sequence of independent experiments carried out at distinct times. Here, the family of events represents the collection of the outcomes of each experiment up to a given time and the distribution of the random variables then can be used to forecast future results.

Recall that I discussed the experiment in which a coin is tossed countably many times (*cf.* on page 38). The probability space was defined as $(\Omega, \mathcal{F}, \mathbb{P})^n$ with $\Omega = \{\{H\}, \{T\}\}$ and $\mathcal{F} = (\emptyset, \Omega, \{H\}, \{T\})$. The probability measure for each toss was given by

$$\mathbb{P}(A) = \begin{cases} 1/2, & A = \{H\} \\ 1/2, & A = \{T\}. \end{cases}$$

After k tosses for any $k \in \mathbb{T}$, an outcome in the product sample space Ω^k is an order collection of outcomes in Ω representing the successive tosses of the coin (*i.e.* a k-dimensional vector). Specifically, any outcome ω in Ω^k is the "path" $\omega = (\omega_1, \ldots, \omega_k)^{\mathrm{T}}$of outcomes up to time $k \in \mathbb{T}$ whose probability is given by

$$\mathbb{P}(\omega) = \prod_{j=1}^{k} \mathbb{P}(\omega_k).$$

REMARK. Note that $\mathbb{P}(\omega)$ represents that the probability that a specific path $(\omega_1, \ldots, \omega_k)$ will occur and not simply the probability of end point ω_k. In other words, $\mathbb{P}(\omega) \neq \mathbb{P}(\omega_k)$. □

I can model such a path by the stochastic process $X = (X_k)_{k\in\mathbb{T}}$ such that X_k is the result of the k^{th} coin toss for all $k \in \mathbb{T}$, or

$$X_k(\omega) = \begin{cases} 1, & \omega_k = \{H\} \\ -1, & \omega_k = \{T\}. \end{cases}$$

Then the vector $X = (X_1, \ldots, X_k)$ represents the path of outcomes in the state space up to and including the k^{th} toss.

Now, define the stochastic process $S = (S_k)_{k\in\mathbb{T}}$ where $S_0 = 0$ and

$$\begin{aligned} S_k &= S_{k-1} + X_k, \quad (\text{for all } k \in \mathbb{T}) \\ &= \sum_{j=1}^{k} X_j. \end{aligned}$$

Essentially, S represents the net result of the coin tosses $(X_1, \ldots, X_k)$ up to and including the k^{th} coin toss for all $k \in \mathbb{T}$. Both S and X are examples of a *discrete-time* stochastic process. In fact, they are very important discrete-time stochastic processes, known as a *symmetric random walk* and the *increments of a symmetric random walk* respectively. □

DEFINITION. Let $X = (X_k)_{k\in\mathbb{T}}$ and $S = (S_k)_{k\in\mathbb{T}}$ be defined as above. Then S is called a ***symmetric random walk*** (***centered at*** 0) ***with increments*** X.

REMARK. By definition, a symmetric random walk has independent increments since $S_k - S_{k-1} = X_k$ and the X_k are assumed to be i.i.d. □

THEOREM 2.3 (Moments of symmetric random walks). *Let $S = (S_k)_{k\in\mathbb{T}}$ be a symmetric random walk generated by $X = (X_k)_{k\in\mathbb{T}}$ such that $X_k \sim N(0,1)$ and are independent. Then*

$$S_j \sim N(0, j) \quad (\textit{for any } j \in \mathbb{T})$$

and

$$\mathrm{Cov}[S_i, S_j] = i \wedge j \quad (\textit{for } i, j \in \mathbb{T} \textit{ with } i \neq j)$$

where "$\wedge$" is the "minimum of" operator.

PROOF. Since the $X_k \sim N(0,1)$ for all $k \in \mathbb{T}$, then given any $j \in \mathbb{T}$,

$$\mathbb{E}[S_j] = \mathbb{E}\left[\sum_{i=1}^{j} X_i\right] = \sum_{i=1}^{j} \mathbb{E}[X_1] = 0.$$

and

$$\begin{aligned}
\mathrm{Var}[S_j] &= \mathbb{E}\left[S_j^2\right] - \mathbb{E}[S_j]^2 \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{j} X_i\right)\left(\sum_{i=1}^{j} X_i\right)\right] \quad (\mathbb{E}[S_j] = 0) \\
&= \mathbb{E}\left[\sum_{i=1}^{j} X_i^2\right] + \mathbb{E}\left[\sum_{\substack{l\neq i \\ i,l=1}}^{j} X_l X_i\right] \\
&= \sum_{i=1}^{j} \mathrm{Var}[X_1] + \sum_{\substack{l\neq i \\ i,l=1}}^{j} \mathbb{E}[X_l X_i] \quad (X_i \text{ are i.i.d.}) \\
&= j + \sum_{\substack{l\neq i \\ i,l=1}}^{j} \mathbb{E}[X_l]\,\mathbb{E}[X_i] \quad (X_k \perp X_i) \\
&= j.
\end{aligned}$$

Hence, $S_j \sim N(0,j)$.

To prove the second statement, assume, without loss of generality, that $i,j \in \mathbb{T}$ and $i<j$, then

$$\begin{aligned}
\mathrm{Cov}[S_i,S_j] &= \mathbb{E}[S_i S_j] - \mathbb{E}[S_i]\,\mathbb{E}[S_j] \\
&= \mathbb{E}\left[\left(\sum_{k=1}^{i} X_k\right)\left(\sum_{l=1}^{j} X_l\right)\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{i} X_k\left(\sum_{l=1}^{i} X_l + \sum_{i=1+1}^{j} X_l\right)\right] \\
&= \mathbb{E}\left[\left(\sum_{k=1}^{i} X_k\right)\left(\sum_{k=1}^{i} X_k\right)\right] + \mathbb{E}\left[\left(\sum_{k=1}^{i} X_k\right)\left(\sum_{i=1+1}^{j} X_l\right)\right].
\end{aligned}$$

Noting that the X_k are i.i.d. and using the proof above

$$\mathrm{Cov}[S_i,S_j] = i + \mathbb{E}\left[\left(\sum_{k=1}^{i} X_k\right)\right]\mathbb{E}\left[\left(\sum_{i=1+1}^{j} X_l\right)\right] = i.$$

Other the other hand, I could have just as easily assumed that $j < i$. In this case, $\text{Cov}[S_i, S_j] = j$. Hence,

$$\text{Cov}[S_i, S_j] = i \wedge j.$$

□

Random walks are often used to represent price dynamics, especially in regard to numerical estimation and they will be used extensively in the sequel. In fact, one of the first important pricing models to consider general options is known as the *binomial option pricing model* and is based on a discrete random walk. See Shreve **[76]** for an exhaustive discussion of the model. □

Independence can also defined in terms of the distribution of a stochastic process as follows.

DEFINITION. Let $X = (X(t))_{t\in\mathbb{T}}$ and $Y = (Y(t))_{t\in\mathbb{T}}$ be stochastic processes. X and Y are said to be ***independent*** if, for any finite partitions of $\mathbb{T}$

$$\{0 \le t_1 < \cdots < t_m \le T\} \text{ and } \{0 \le s_1 < \cdots < s_m \le T\}$$

the vectors

$$(X(t_1), \ldots, X(t_m)) \text{ and } (Y(s_1), \ldots, Y(s_m))$$

are independent.

2.2.2. The Distribution of a Stochastic Process. The concept of a distribution can be extended to a stochastic process $X = (X(t))_{t\in\mathbb{T}}$ using the fact that for each fixed time t, $X(t)$ is random variable. This is formalized in the following definitions.

DEFINITION. Assume $\{t_1, \ldots, t_m\}$ is any finite collection of points in of $\mathbb{T}$. Then the (***finite-dimensional***) ***distribution*** of a real-valued stochastic process $X = (X(t))_{t\in\mathbb{T}}$ is the joint distribution,

$$\mathbb{P}_{X(t_1),\ldots,X(t_m)}(B_1, \ldots, B_m) := \mathbb{P}(X(t_1) \in B_1, \ldots, X(t_m) \in B_m)$$

for $B_j \in \mathcal{B}$ where $1 \le j \le m$.

Equivalently,

$$\mathbb{P}_{X(t_1),\ldots,X(t_m)}(B) := \mathbb{P}(X(t_1), \ldots, X(t_m) \in B)$$

where $B \in \mathcal{B}^k$.

Stochastic processes with joint normal distributions are especially important in continuous-time finance. They also help establish the connection between price processes and PDE which will be heavily exploited in the sequel (*cf.* Chapter 4).

DEFINITION (Gaussian process). A real-valued stochastic process $X = (X(t))_{t\in\mathbb{T}}$ is said to be a ***Gaussian*** (*a.k.a.* ***Normal***) ***process*** if, for all finite partitions of $\mathbb{T}$ its finite-dimensional distribution is a joint normal distribution. □

To better understand the importance of normal stochastic processes in continuous-time finance, note that for a normal stochastic process $X = (X(t))_{t\in\mathbb{T}}$, the random variable $X(t)$ at any time $t > 0$ is normal (*i.e.* $X(t) \sim N(0,t)$) with density

$$u(x) = \frac{1}{\sqrt{2\pi t}} e^{\frac{-|x|^2}{2t}}.$$

It easy to show that this is the solution to the *heat equation*

$$u_t - \frac{1}{2}\triangle u = 0 \quad \text{in } \mathbb{R}\times(0,\infty).$$

In other words, if X represents the dynamics of the price of a financial instrument, then its distribution evolves over time according to the heat equation. This is discussed in much greater depth below (*cf.* Section 4.1).

An important class of stochastic process are those whose finite-dimensional distribution is independent time.

DEFINITION. A stochastic process $X = (X(t))_{t\in\mathbb{T}}$ is said to be ***stationary*** if its finite-dimensional distribution is invariant under shifts in time. In other words, for any finite collection of points $\{t_1,\ldots,t_m\}$ in $\mathbb{T}$ and any reasonable fixed $\tau > 0$,

$$\mathbb{P}(X(t_1+\tau)\in B_1,\ldots,X(t_m+\tau)\in B_m) \;=\; \mathbb{P}(X(t_1)\in B_1,\ldots,X(t_m)\in B_m)$$

for all $B_j \in \mathcal{B}$ with $1 \le j \le m$.

An interesting questions is: *does a stochastic process exist for a given finite-dimensional distribution*? The answer is *yes* provided that the two following consistency conditions hold. First, let me begin with a definition.

THEOREM 2.4 (First consistency condition). *Assume that $X = (X(t))_{t\in\mathbb{T}}$ is a real-valued stochastic process and $\{t_1,\ldots,t_m\}$ is a finite collection of points in $\mathbb{T}$. Further, assume that $\pi(\cdot)$ is some permutation of $(1,\ldots m)$. Then for all $B_j \in \mathcal{B}$ with $1 \le j \le m$,*

$$\mathbb{P}_{X(t_1),\ldots,X(t_m)}(B_1,\ldots,B_m) \;=\; \mathbb{P}_{X\left(t_{\pi(1)}\right),\ldots,X\left(t_{\pi(m)}\right)}\left(B_{\pi(1)},\ldots,B_{\pi(m)}\right). \tag{2.2.1}$$

PROOF. Given the setup above, the result follows directly from the fact that

$$\mathbb{P}(X(t_1)\in B_1,\ldots,X(t_m)\in B_m) = \mathbb{P}\left(X\left(t_{\pi(1)}\right)\in B_{\pi(1)},\ldots,X\left(t_{\pi(m)}\right)\in B_{\pi(m)}\right)$$

for any permutation π. □

Consequently, for any m-tuple $(s_1,\ldots,s_m)$ of distinct elements of $\mathbb{T}$ such that $s_j = t_i$ for some $1 \le i,j \le m$.

$$\mathbb{P}_{X(t_1),\ldots,X(t_m)}(B_1,\ldots,B_m) = \mathbb{P}_{X(s_1),\ldots,X(s_m)}(B_{s_1},\ldots,B_{s_m}).$$

In other words, the finite distribution of a stochastic process does not depend on the order of the time index set (*i.e.* it is invariant under coordinate permutations).

The second consistency result provides the means to extend finite products spaces to infinite ones.

THEOREM 2.5 (Second consistency condition). *Assume that $X = (X(t))_{t\in\mathbb{T}}$ is a real-valued stochastic process and $\{t_1, \ldots, t_m\}$ is a finite collection of points in $\mathbb{T}$. Then for all $B_j \in \mathcal{B}$ with $1 \leq j \leq m$*

$$\mathbb{P}_{X(t_1),\ldots,X(t_{m-1})}(B_1, \ldots, B_{m-1}) = \mathbb{P}_{X(t_1),\ldots,X(t_{m-1}),X(t_m)}(B_1, \ldots, B_{m-1}, \mathbb{R}). \tag{2.2.2}$$

PROOF. Assume that $X = (X(t))_{t\in\mathbb{T}}$ is a real-valued stochastic process and $\{t_1, \ldots, t_m\}$ is a finite collection of points in $\mathbb{T}$. Since $\mathbb{P}(X(t_m) \in \mathbb{R}) = 1$ and $t_i < t_j$ for $1 \leq i < j \leq m$

$$\begin{aligned}\mathbb{P}_{X(t_1),\ldots,X(t_{m-1}),X(t_m)}(B_1, \ldots, B_{m-1}, \mathbb{R}) &= \mathbb{P}(X(t_1) \in B_1, \ldots, X(t_{m-1}) \in B_{m-1} \mid X(t_m) \in \mathbb{R}) \\ &= \mathbb{P}(X(t_1) \in B_1, \ldots, X(t_{m-1}) \in B_{m-1}).\end{aligned}$$

□

The following result, which combines these two consistency conditions, will be helpful in the sequel. Its proof follows Billingsley [**8**, pg. 514-515].

THEOREM 2.6 (Consistency condition). *Assume that $X = (X(t))_{t\in\mathbb{T}}$ is a real-valued stochastic process and let $(t_1, \ldots, t_m)$ be distinct points in $\mathbb{T}$. Further, assume that $\pi(\cdot)$ is some permutation of $(1, \ldots m)$. Then for all $B_j \in \mathcal{B}$ with $1 \leq j \leq m$ and any $k \leq m$*

$$\mathbb{P}_{X(t_1),\ldots,X(t_k)}(B_1, \ldots, B_k) = \mathbb{P}_{X\left(t_{\pi(1)}\right),\ldots,X\left(t_{\pi(k)}\right)}\left(B_{\pi(1)}, \ldots, B_{\pi(k)}, \mathbb{R}^{m-k}\right). \tag{2.2.3}$$

PROOF. Let's begin by showing that the above statement incorporates the first consistency condition. Assume the finite k-tuple $(t_1, \ldots, t_k)$ has distinct elements of $\mathbb{T}$ such that $k \leq m$. Further, let π_k be some permutation of $(1, \ldots k)$ and let $\varphi : \mathbb{R}^k \to \mathbb{R}^k$ be a coordinate permutation defined as

$$\varphi_\pi(x_1, \ldots, x_k) = \left(x_{\pi^{-1}(1)}, \ldots, x_{\pi^{-1}(k)}\right).$$

Then its inverse is also a coordinate permutation, or

$$\varphi_\pi^{-1}(x_1, \ldots, x_k) = \left(x_{\pi(1)}, \ldots, x_{\pi(k)}\right).$$

Then, ,

$$\varphi_\pi^{-1}(X(t_1), \ldots, X(t_k)) = X\left(t_{\pi(1)}\right), \ldots, X\left(t_{\pi(k)}\right).$$

Then by (2.2.1)

$$\mathbb{P}_{X(t_1),\ldots,X(t_k)}(B_1, \ldots, B_k) = \mathbb{P}_{X\left(t_{\pi(1)}\right),\ldots,X\left(t_{\pi(k)}\right)} \circ \varphi_\pi^{-1}(B_1, \ldots, B_k)$$

for all $B_j \in \mathcal{B}$ with $1 \leq j \leq k$ where

$$\varphi_\pi^{-1}(B_1, \ldots, B_k) := \left(B_{\pi(1)}, \ldots, B_{\pi(k)}\right).$$

Hence,

$$\mathbb{P}_{X(t_1),\ldots,X(t_k)}(B_1, \ldots, B_k) = \mathbb{P}_{X(t_1),\ldots,X(t_k)\varphi_\pi^{-1}}(B_1, \ldots, B_k)$$

for any permutation π and for all $B_j \in \mathcal{B}$ with $1 \leq j \leq k$.

Now, let's show that the second consistency condition is included in the above expression. Define the projection $\phi : \mathbb{R}^k \to \mathbb{R}^{k-1}$ as

$$\phi(x_1, \ldots, x_k) = (x_1, \ldots, x_{k-1}).$$

Then following the above argument and applying (2.2.2)

$$\mathbb{P}_{X(t_1),\ldots,X(t_{k-1})}(B_1, \ldots, B_{k-1}) = \mathbb{P}_{X(t_1),\ldots,X(t_k)}\phi^{-1}(B_1, \ldots, B_{k-1})$$

for all $B_i \in \mathcal{B}$ with $1 \leq i \leq k$ where

$$\phi^{-1}(B_1, \ldots, B_{k-1}) := (B_1, \ldots, B_{k-1}, \mathbb{R}).$$

Further, I can combine both of these results as follows. Let $(\tau_1, \ldots, \tau_k, \tau_{k+1}, \ldots \tau_m)$ be any m-tuple of distinct elements of $\mathbb{T}$ such that $t_j = \tau_i$ for some $1 \leq j \leq k$ and $1 \leq i \leq m$. Then there exists some permutation $\hat{\pi}$ such that

$$\begin{aligned} \hat{\pi}(\tau_1, \ldots, \tau_k, \tau_{k+1}, \ldots \tau_m) &= \left(\tau_{\hat{\pi}(1)}, \ldots, \tau_{\hat{\pi}(k)}, \tau_{\hat{\pi}(k+1)}, \ldots \tau_{\hat{\pi}(m)}\right) \\ &= (t_1, \ldots, t_k, \tau_{k+1}, \ldots \tau_m). \end{aligned}$$

Further, define $\psi : \mathbb{R}^m \to \mathbb{R}^k$ where $k \leq m$ as

$$\begin{aligned} \psi(x_1, \ldots, x_m) &= \varphi_{\hat{\pi}} \circ \phi_k(x_1, \ldots, x_m) \\ &= x_{\hat{\pi}^{-1}(1)}, \ldots, x_{\hat{\pi}^{-1}(k)} \end{aligned}$$

where $\phi_k(\cdot) := \prod_{i=1}^{k} \phi(\cdot)$. Therefore,

$$\mathbb{P}_{X(t_1),\ldots,X(t_k)}(B_1, \ldots, B_k) = \mathbb{P}_{X\left(t_{\hat{\pi}(1)}\right),\ldots,X\left(t_{\hat{\pi}(k)}\right)}\psi^{-1}(B_1, \ldots, B_k)$$

where

$$\psi^{-1}(B_1, \ldots, B_k) := \left(B_{\pi(1)}, \ldots, B_{\pi(k)}, \mathbb{R}^{m-k}\right).$$

□

REMARK. The finite distribution of a stochastic process does not completely characterize the process since it is not path-wise property. In other words, it is possible for two stochastic processes to have the same finite distribution and, yet, have different paths *a.s.* See Billingsley [**8**, pages 523-525] for discussion on this topic. □

2.2.3. Infinite Product Spaces. Above, I defined a real-valued stochastic process $X = (X(t))_{t\in\mathbb{T}}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as a collection of random variables $\{X(t)\}_{t\in\mathbb{T}}$ where $X(t) : (\Omega; \mathcal{F}) \to (\mathbb{R}; \mathcal{B})$ for each $t \in \mathbb{T}$ in some arbitrary time index (*cf.* definition on page 63). However, it is not clear that such a process exists, especially if the time index is infinite.

Recall that earlier, I discussed how to construct such a product space from finitely many probability spaces. (*cf.* Section 1.1.4). If the index set is finite, say $\mathbb{T} = (t_1, \ldots, t_n)$, then process X can be

seen as mapping $(\Omega : \mathcal{F})^n \to (\mathbb{R} : \mathcal{B})^n$. In other words, a stochastic process can viewed as X is a random variable on the product space $(\Omega, \mathcal{F}, \mathbb{P})^n$. Hence, such a stochastic process exists.

However, in the case of continuous-time stochastic process, I will need something equivalent to infinite product spaces since any time interval will be infinite. The issue here is confirming that a σ-algebra and probability measure can be constructed for such spaces. Unfortunately, this is non-trivial and requires a number of highly technical results from measure theory, the key result being the *Kolmogorov's extension theorem.*

REMARK. In the finite case, I was able to construct a product space from a collection of probability spaces. In the case of an infinite product space, this is limited to a single probability one (*cf.* Theorem 2.7). However, this is not a real limitation since this is sufficient for all stochastic process discussed herein (*cf.* [TBD]). □

While the task at hand is to construct a infinite product space given some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, since a random variable maps one measurable space to another, let me begin by constructing a infinite product of measurable spaces.

Motivation. Assume that $X = (X_k)_{k \in \mathbb{T}}$ is a discrete-time, real-valued stochastic process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then it maps a point in Ω to a point in $\mathbb{R}$ for each fixed time n (*i.e.* $\omega_n \mapsto x_n$). However, the process can also be thought it terms of its *history* of states up to time n. In other words, it can be viewed as a function that, for each $n \in \mathbb{T}$, maps a sequence of n-outcomes to a vector in $\mathbb{R}^n$, or

$$(\omega_1, \ldots, \omega_n) \mapsto (x_1, \ldots, x_n) \in \mathbb{R}^n$$

where $x_i := X(\omega_i)$ for $1 \leq i \leq n$.

From this perspective, for each time $n \in \mathbb{T}$, X_n is a random vector from taking a vector of outcomes $\omega \in \Omega^n$ to points in $\mathbb{R}^n$. Essentially, this can also be thought of a mapping from $\Omega \times \mathbb{T} \to \mathbb{R} \times \mathbb{T}$ when $\mathbb{T} \subseteq \mathbb{N}$.

This suggests that a continuous-time, real-valued stochastic process might be thought of as a mapping from $\Omega \times \mathbb{T} \to \mathbb{R} \times \mathbb{T}$ where $\mathbb{T} = [0, T]$. Consequently, I will need to define an appropriate σ-algebra each space which preserves the concept of measurability for them to be of any use.

The following discussion generally follows Billingsley (*cf.* [**8**, Section 36]) and Çinlar (*cf.* [**14**, pg. 164-166]). Let me begin with a couple of definitions. □

DEFINITION. Let $R^{\mathbb{T}}$ be the collection of real-valued, measurable functions on $\mathbb{T}$. Then an element $v \in R^{\mathbb{T}}$ is a function from $\mathbb{T} \to \mathbb{R}$ and its ***history*** up to time $t \in \mathbb{T}$ is the set

$$\{v(s) \mid s \leq t \text{ such that } s \in \mathbb{T}\}.$$

REMARK. Note that I have not made any assumptions regarding the regularity of the elements of $R^{\mathbb{T}}$. □

Following the discussion above, if $\mathbb{T}$ is a finite discrete set with n elements, then the history of $v(\cdot) \in R^{\mathbb{T}}$ is the real-valued vector $(v(t_1), \ldots, v(t_n))$. Since any element of $R^{\mathbb{T}}$ is uniquely identified

by its history, $R^{\mathbb{T}}$ can be thought of as product space $\mathsf{X}_{j=1}^{n} R^1$. Additionally, changing notation so that $x_i := v(t_i)$ for $1 \leq i \leq n$, then the history of any $v(\cdot) \in R^{\mathbb{T}}$ is vector in $\mathbb{R}^n$, or

$$(v(t_1), \ldots, v(t_n)) = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

As such, for a discrete index set $\mathbb{T}$, it is easy to see that $R^{\mathbb{T}}$ can be identified with $\mathbb{R}^n$ and, in particular, R^1 is equivalent to $\mathbb{R}$.Consequently, in order construct a σ-algebra on $R^{\mathbb{T}}$, lets begin by first looking at R^1.

Remark. Two stochastic processes that have the same history on some time index set are indistinguishable. Hence, the path of a stochastic process creates an equivalence class among members of $R^{\mathbb{T}}$. It also establishes an isomorphism between $R^{\mathbb{T}}$ and $\mathbb{R}^{\mathbb{T}}$ (modulo this equivalence class) when $\mathbb{T}$ is discrete.

As seen above, R^1 is equivalent to $\mathbb{R}$, so we should expect any σ-algebra on R^1 to be equivalent to $\mathcal{B}(\mathbb{R})$. Now, consider the collection of sets of the form

$$R_A := \left\{ v(\cdot) \in R^1 \mid v(t) \in A \right\}$$

where $A \in \mathcal{B}(\mathbb{R})$. It is easy to show that $\mathcal{R}^1 := \{R_A \mid A \in \mathcal{B}(\mathbb{R})\}$ is a σ-algebra since $\mathcal{B}(\mathbb{R})$ is one.

Also, recall that we constructed a σ-algebra on a product space of two measurable spaces using measurable rectangles. Specifically, for two measurable space $(E, \mathcal{E})$ and $(F, \mathcal{F})$, the collection $\mathcal{G}$ of all measurable rectangles of the form

$$\{(x, y) \mid x \in A, y \in B\}$$

where $A \in \mathcal{E}$ and $B \in \mathcal{B}$ is a σ-algebra on the product space $E \times F$ and $(E \times F, \mathcal{G})$ is a measurable space (*cf.* Theorem 1.6).

Therefore, consider R^2 as the product space $R^1 \times R^1$ along with the collection of sets of the form

$$R_{A \times B} = \{(v(\cdot), w(\cdot)) \mid v(\cdot) \in R_A \text{ and } w(\cdot) \in R_B\}$$

where $R_A, R_B \in \mathcal{R}^1$. From the definition of R_A and R_B above

$$R_{A \times B} = \{(v(\cdot), w(\cdot)) \mid v(t) \in A \text{ and } w(t) \in B\}$$

Then, it is clear that collection of such sets $\mathcal{R}^2 := \{R_{A \times B} \mid A, B \in \mathcal{B}(\mathbb{R})\}$ is equivalent to $\mathcal{B}^2(\mathbb{R})$. Hence, there is a σ-algebra on the product space $R^2 = R^1 \times R^1$ and $(R^2, \mathcal{R}^2)$ is a measurable space.

On the other hand, setting $\mathbb{T} = (t_1, t_2)$, we can also consider R^2 in terms of the histories of its elements, or

$$\left\{ (v(t_1), v(t_2)) \mid v(\cdot) \in R^2 \right\}$$

In this case, $\mathcal{R}^2$ is the collection of sets of the form

$$\{(v(t_1), v(t_2)) \mid v(t_1) \in A \text{ and } v(t_2) \in B\}$$

where $A, B \in \mathcal{B}(\mathbb{R})$.

In other words, this σ-algebra is constructed using the projection of each of the two components of $(v(t_1), v(t_2))$ onto R^1. This suggests a general method to construct a σ-algebra on $R^{\mathbb{T}}$ for any arbitrary index set $\mathbb{T}$.

For instance, assume that $\mathbb{T} = [0, T]$, it is possible to define a *projection* from $R^{\mathbb{T}} \to R^1$ which represents selecting one point in the history of a given function in $R^{\mathbb{T}}$. Then, for some $v(\cdot) \in R^{\mathbb{T}}$ and any $t \in [0, T]$, the projection of the t^{th} component of $v(\cdot)$ is simply $v(t)$. This leads to the following definition which will be used to generate a σ-algebra on the set $R^{\mathbb{T}}$ for any arbitrary index set $\mathbb{T}$.

DEFINITION. Assume that $\mathbb{T}$ is an arbitrary (possibly infinite) index set. Then for any $t \in \mathbb{T}$, the ***projection*** $Z_t : R^{\mathbb{T}} \to R^1$ is defined as

$$Z_t(v(\cdot)) := v(t) \in \mathbb{R}$$

where $v(t)$ is the value of function $v(\cdot) \in R^{\mathbb{T}}$ at time t.

REMARK. In particular, if $\mathbb{T} = (t_1, \ldots, t_n)$, then $(v(t_1), \ldots, v(t_n)) = (x_1, \ldots, x_n) \in \mathbb{R}^n$. □

REMARK. In the event that the function in question is a stochastic process, then for some $t \in \mathbb{T}$, its history up to time t is its path up to time t and this projection is the value of the stochastic process at time t. □

Following the motivation above, let $\mathcal{R}^{\mathbb{T}}$ be the σ-algebra generated by the collection of all projections Z_t on $R^{\mathbb{T}}$, or

$$\mathcal{R}^{\mathbb{T}} = \sigma\left(Z_t(v(\cdot)) \mid t \in \mathbb{T} \text{ and } v(\cdot) \in R^{\mathbb{T}}\right).$$

Then $\mathcal{R}^{\mathbb{T}}$ consists of sets of the form

$$\left\{v(\cdot) \in R^{\mathbb{T}} \mid Z_t(v(\cdot)) \in A\right\} = \left\{v(\cdot) \in R^{\mathbb{T}} \mid v(t) \in A\right\}$$

for $t \in \mathbb{T}$ and some $A \in \mathcal{B}(\mathbb{R})$. In other words, $\mathcal{R}^{\mathbb{T}}$ is generated by functions that are in the set A at time t.

REMARK. If $\mathbb{T} = (t_1, \ldots, t_n)$, then by the motivation above, $\mathcal{R}^{\mathbb{T}}$ is equivalent to $\mathcal{B}^n(\mathbb{R})$. □

This projection can be used to construct the analog of measurable rectangles for infinite product spaces which, in turn, can be used to show that $\mathcal{R}^{\mathbb{T}}$ is a σ-algebra for any arbitrary (possibly infinite) index set $\mathbb{T}$. Let me begin by defining a generalization of measurable rectangles.

DEFINITION. Assume that $\mathbb{T}$ is an arbitrary (possibly infinite) index set and Z is the projection of $R^{\mathbb{T}}$ onto R^1 defined above. Then for the finite k-tuple $(t_1, \ldots, t_k)$ of distinct points in $\mathbb{T}$, define a ***k-cylinder*** as a set of the form

$$\left\{v(\cdot) \in R^{\mathbb{T}} \mid Z_{t_1}(v(\cdot)), \ldots, Z_{t_k}(v(\cdot)) \in A\right\} = \left\{v(\cdot) \in R^{\mathbb{T}} \mid v(t_j) \in A_j \text{ for } 1 \leq j \leq k\right\}$$

where $A = (A_1, \ldots, A_k) \in \mathcal{B}^k(\mathbb{R})$. Further, define $\mathcal{R}_0^{\mathbb{T}}$ to be the collection of all such cylinders.

REMARK. A k-cylinder consists of real-valued functions on $\mathbb{T}$ which are in the set A_j at time t_j. □

To see the similarity to measurable rectangles, set $x_j = v(t_j)$ for $1 \leq j \leq k$, then a k-cylinder is of the form

$$\left\{v(\cdot) \in R^{\mathbb{T}} \mid x_{t_j} \in A_j \text{ for } 1 \leq j \leq k\right\}$$

which is equivalent to the measurable rectangle

$$\left\{(x_{t_1}, \ldots, x_{t_k}) \mid x_{t_j} \in A_j \text{ for } 1 \leq j \leq k\right\}.$$

However, unlike the collection of measurable rectangles, it can be shown that $\mathcal{R}_0^{\mathbb{T}}$ is not a σ-algebra; however, it is a field and it can be shown that $\mathcal{R}^{\mathbb{T}} = \sigma\left(\mathcal{R}_0^{\mathbb{T}}\right)$ (*cf.* [**8**, Section 36]). Hence, $\left(R^{\mathbb{T}}, \mathcal{R}^{\mathbb{T}}\right)$ is a measurable space. Consequently, all that is left to show is that it is possible to construct an probability measure on the measurable space $\left(R^{\mathbb{T}}, \mathcal{R}^{\mathbb{T}}\right)$.

Let me begin by expressing the projections dependency on time explicitly, or

$$Z((\cdot), t) := Z_t((\cdot)).$$

Then for some fixed $v(\cdot) \in R^{\mathbb{T}}$, the projection $Z(v(\cdot), \cdot)$ is a real-valued function on $\mathbb{T}$. On the other hand, for some fixed $t \in \mathbb{T}$, $Z(\cdot, t)$ is a real-valued function on $R^{\mathbb{T}}$. Moreover, if a probability measure μ on $\left(R^{\mathbb{T}}, \mathcal{R}^{\mathbb{T}}\right)$ exists, then by construction, $Z(\cdot, t)$ is a real-valued, measurable function for each $t \in \mathbb{T}$.

Motivation. Assume that $X = ((X(t))_{t \in \mathbb{T}}$ is a real-valued stochastic process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then $X \in R^{\mathbb{T}}$. Hence, $Z(t) = (Z(X(t), t))_{t \in \mathbb{T}}$ can be viewed as stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, Z will have the finite-dimensional distribution $\mathbb{P}_{X(t_1), \ldots, X(t_k)}$ for some finite k where $t_1, \ldots, t_k \in \mathbb{T}$. This leads to the following definitions. □

Definition. Assume that μ is a probability measure on the measurable space $\left(R^{\mathbb{T}}, \mathcal{R}^{\mathbb{T}}\right)$, then the real-valued stochastic process $Z = (Z(t))_{t \in \mathbb{T}}$ on the probability space $\left(R^{\mathbb{T}}, \mathcal{R}^{\mathbb{T}}, \mu\right)$ defined as

$$Z(t) = Z(v(\cdot), t)$$

for some $v(\cdot) \in R^{\mathbb{T}}$ and $t \in \mathbb{T}$ is called the ***projection process***.

Remark. When $v(\cdot) \in R^{\mathbb{T}}$ is fixed, then $Z(v(\cdot), \cdot)$ is the history of $v(\cdot)$ over $t \in \mathbb{T}$. When $t \in \mathbb{T}$ is fixed, then $Z((\cdot), t)$ is the value of some real-valued, random function at time t. □

Definition. Let μ be a probability measure on some measurable space $(E, \mathcal{E})$, then for some real-valued measurable function $f : E \to \mathbb{R}$ for $1 \leq j \leq n$, the product measure on $(\mathbb{R}^n, \mathcal{B}^n)$ defined by

$$\mu_{t_1, \ldots, t_n}(f, B) = \mu(f(t_1), \ldots f(t_n) \in B) \quad (B \in \mathcal{B}^n).$$

is called a ***(n-) system of distributions***. Further, if the system meets the consistency condition (2.2.3), then the system is said to be ***consistent***.

Notation. Often, the measurable function is not important. In the case, a system of distributions is denoted as $\mu_{t_1, \ldots, t_n}(\cdot)$. □

I am now ready to prove to the Kolmogorov's extension theorem. This result can be stated in two ways. The first shows that under certain conditions, a probability measurable exists on an infinite product of measurable spaces isomorphic to$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such as $(R^1, \mathcal{R}^1)$. The second shows that given the product of measurable spaces of this form, it is possible to construct a stochastic process with a given distribution on that product space. The second essentially demonstrates the existence of stochastic processes as defined herein. Both of these are proved in Billingsley [**8**, page 520-523].

THEOREM 2.7 (Kolmogorov's extension theorem version 1). . *Assume that $\mu_{t_1,\ldots,t_k}(\cdot)$ is a consistent system distributions for some finite k. Then there is a probability measure $\mathbb{P}$ on $\left(R^{\mathbb{T}}, \mathcal{R}^{\mathbb{T}}\right)$ such that the projection process $Z = (Z(t))_{t\in\mathbb{T}}$ on $\left(R^{\mathbb{T}}, \mathcal{R}^{\mathbb{T}}, \mathbb{P}\right)$ has finite-dimensional distribution $\mu_{t_1,\ldots,t_k}(\cdot)$ (i.e. $\mathbb{P}(Z(t_j) \in (\cdot)) = \mu_{t_j}(\cdot)$ for all $1 \leq j \leq k$).*

THEOREM 2.8 (Kolmogorov's extension theorem version 2). . *Assume that $\mu_{t_1,\ldots,t_k}(\cdot)$ is a consistent system distributions for some finite k. Then there is exists a stochastic process $X = (X(t))_{t\in\mathbb{T}}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that X has finite-dimensional distribution $\mu_{t_1,\ldots,t_k}(\cdot)$.*

EXAMPLE. Assume that $\mathbb{T} = (t_1, \ldots, t_n)$ and that $\mu_{1,\ldots,n}(\cdot)$ is a system of distributions on $(\mathbb{R}^n, \mathcal{B}^n(\mathbb{R}))$. Then $\mathbb{P}$ defined as

$$\mathbb{P}(A) = \mu_{1,\ldots,n}(B)$$

where $B \in \mathcal{B}^n(\mathbb{R})$ and $A \in \mathcal{R}^{\mathbb{T}}$ is the form

$$A = \left\{v(\cdot) \in R^{\mathbb{T}} \mid Z_{t_1}(v(\cdot)), \ldots, Z_{t_k}(v(\cdot)) \in B\right\}.$$

EXAMPLE. Assume that $\mathbb{T}$ is some arbitrary bounded interval $[0, T]$. Further, assume that $X = (X(t))_{t\in\mathbb{T}}$ is a real-valued, measurable stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the projection process $Z = (Z(X(t), t))_{t\in\mathbb{T}}$ on X is a stochastic process $Z = (Z(X(\omega, t), t))_{t\in\mathbb{T}}$ from $(\Omega \times \mathbb{T}; \mathcal{F} \times \mathcal{B}(\mathbb{R})) \to (\mathbb{R} \times \mathbb{T}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{T}))$ with the finite-dimensional distribution

$$\mathbb{P}_{X(t_1),\ldots,X(t_k)}$$

for some finite k. □

2.2.4. The Characterization of Stochastic Processes. At the outset of this chapter, I characterized a stochastic process as a family of (possibly uncountably many) random variables on the same sample space (*cf.* 2.2). In other words, given a real-valued stochastic process $X = (X(t))_{t\in\mathbb{T}}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, for each fixed$t \in \mathbb{T}$, $X(t)$ is some real-valued random variable $Y_t : (\Omega; \mathcal{F}) \to (\mathbb{R}; \mathcal{B})$ where

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

Essentially, this is the point-wise perspective of a stochastic process.

In the previous section, I showed that a stochastic process can be viewed in terms of its paths. For example, given a finite index set $\mathbb{T} = (t_1, \ldots t_n)$, a real-valued stochastic process $X = (X_k)_{k\in\mathbb{T}}$ can be thought of as a function that maps n-outcomes to a vector in $\mathbb{R}^n$ for each $n \in \mathbb{T}$ (*i.e.* $(\omega_1, \ldots \omega_n) \mapsto (x_1, \ldots x_n)$). From this perspective, X is a random vector from taking a vector of

outcomes $\omega = (\omega_1, \ldots \omega_n) \in \Omega^n$ to points in $\mathbb{R}^n$ for each time $n \in \mathbb{T}$. This suggested that such a random variable can be thought of a mapping from $\Omega \times \mathbb{T} \to \mathbb{R} \times \mathbb{T}$ when $\mathbb{T} \subseteq \mathbb{N}$.

On the other hand, if the index set is an interval, say $\mathbb{T} = [0, T]$ for some finite T, then a continuous-time stochastic process can be thought of as a mapping from $\Omega \times \mathbb{T} \to \mathbb{R} \times \mathbb{T}$. Here, an outcome can be thought of as path starting $\omega(0) \in \Omega$ and ending at $\omega(t) \in \Omega$ for each $t \in \mathbb{T}$. In other words, the sample space of a stochastic process can be viewed as the product space $\Omega \times \mathbb{T}$ and outcomes are of the form,

$$\omega = \{(\omega(s)) \mid 0 \leq s \leq t\}$$

where $\omega : \mathbb{T} \to \Omega$. As such, it also seems reasonable to view a real-valued stochastic process as a function taking a path in $\Omega \times \mathbb{T}$ to a path in $\mathbb{R} \times \mathbb{T}$. In this context, a stochastic process is a *random function* or $\omega(\cdot) \mapsto X(\omega(\cdot), \cdot)$ where the graph of $\omega(\cdot)$ is a random path in $\Omega \times \mathbb{T}$ starting at time zero up and ending at time t and the graph of $X(\omega(\cdot), \cdot)$ is a random path in $\mathbb{R} \times \mathbb{T}$ also starting at time zero up and ending at time t. In other words, the stochastic process takes Ω-valued paths to real-valued paths, or

$$X : (\Omega \times \mathbb{T}; \mathcal{F} \otimes \mathcal{B}(\mathbb{T})) \to (\mathbb{R} \times \mathbb{T}; \mathcal{B} \otimes \mathcal{B}(\mathbb{T})).$$

As such, we can consider stochastic processes to be random variables on *path-spaces*. This is captured in the following definition.

DEFINITION. Assume $\mathbb{T} = [0, T]$ where $T < \infty$ and let $X = (X(t))_{t \in \mathbb{T}}$ be a a stochastic process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then for a fixed $\omega(\cdot) \in \Omega$, then for any fixed $t \in \mathbb{T}$, the set

$$\{X(s)\}_{0 \leq s \leq t} := \{X(\omega(s), s)\}_{0 \leq s \leq t} := \{X(\omega(s), s) \mid 0 \leq s \leq t\}$$

is called a ***sample path*** up to time t and represents the continuum of outcomes from time zero up to time t.

In other words, a sample path represents the continuous history of outcomes up to the time t. Since the outcome is random at each time, starting the process again will result in a different sample path almost surely. For the remainder of the discussion, I will preliminary view a stochastic process is this context.

EXAMPLE. In the context of continuous-time finance, a stochastic process often represents the price of a given financial instrument (or portfolio of such instruments) as it evolves over time. As such, for each fixed time t, $X(t)$ is random variable which represents the price of a financial instrument at any given time t. On the other hand, for a fixed $\omega(\cdot)$, $\{X(\omega(s), s)\}_{0 \leq s \leq t}$ represents the price history of a given financial instrument up to time t. □

NOTATION. It is customary to denote a path of outcomes $\omega(\cdot)$ simply as ω. □

Given that a stochastic process is a random function, it seems reasonable to consider the regularity of such functions.

DEFINITION. A stochastic process $X = (X(t))_{t\in\mathbb{T}}$ is said to have a ***right limit*** if, for each sample path ω, $\lim_{s\downarrow t} X(\omega, s)$ exists and is bounded a.s. Likewise, it is said to have a ***left limit*** if, for each sample path ω, $\lim_{s\uparrow t} X(\omega, s)$ exists and is bounded.

Similarly, X is said to be ***right continuous*** if, for each sample path ω, it has a right limit and $\lim_{s\downarrow t} X(\omega, s) = X(\omega, t)$ a.s. Likewise, it is said to be ***left continuous*** if, for each sample path ω, it has a left limit and $\lim_{s\uparrow t} X(\omega, s) = X(\omega, t)$ a.s. It is said to be ***continuous*** if right and left continuous.

If X is right continuous with left limits, it is said to be ***RCLL*** or ***càdlàg*** which represents the equivalent French phrase *continue à gauche, limite à droite*.

2.2.5. Equality of Stochastic Processes. Like random variables, there a several means of equating stochastic process. In the first case, two stochastic processes $X = (X(t))_{t\in\mathbb{T}}$ and $Y = (Y(t))_{t\in\mathbb{T}}$ are said to ***have the same law*** (*a.k.a.* are ***equal in distribution***), if, for every finite collection points in $\mathbb{T}$, $\{t_1, \ldots, t_m\}$,

$$X(t_j) \stackrel{d}{=} Y(t_j) \quad (\text{for all } j = 1, \ldots, m).$$

This is clearly equivalent to the two processes having the same finite-dimensional distribution. However, as mentioned above, finite-dimensional distribution is point-wise property and it is possible to have the same finite-dimensional distribution, yet have different paths. The next definitions consider path-wise equivalence.

DEFINITION. Given two stochastic processes $X = (X(t))_{t\in\mathbb{T}}$ and $Y = (Y(t))_{t\in\mathbb{T}}$, X is said to be a ***modification*** of Y if, for any fixed $t \in \mathbb{T}$,

$$X(t) = Y(t) \text{ a.s.}$$

REMARK. If one process is the modification of another, then at each fixed time, the two processes agree except on a set of outcomes of probability zero. In other words, their sample paths over the interval, $[0, t]$, have the same end points almost surely at each time t. □

Viewed as integrable functions, the definition of a stochastic process admits an equivalent class on stochastic processes (*i.e.* they agree everywhere except on sets of probability zero). This is expressed in the next definitions which are the analogs of equivalency results for random variables.

DEFINITION. If the process $X = (X(t))_{t\in\mathbb{T}}$ is a modification of $Y = (Y(t))_{t\in\mathbb{T}}$ and the set $A := \{\omega \in \Omega \,|X(\omega, t) = Y(\omega, t)\}$ is measurable, then X is said to a ***version*** of Y (*a.k.a.* they are ***indistinguishable***) if

$$\mathbb{P}(A) = 1.$$

If X is a version of Y, they are said to be ***stochastically equivalent***.

REMARK. In other words, two processes have the same sample paths almost surely if they are stochastically equivalent. □

THEOREM 2.9. *Any stochastic process is stochastically equivalent to a càdlàg process.*

Given this result, which is proved in Stroock and Varadhan [**81**, page 51], I will assume that all stochastic processes are càdlàg in the sequel.

2.2.6. Admissible Processes. As defined above, a stochastic process is rather abstract and has virtually no properties other than it is family of random variables on some arbitrary probability space. In this paper, I will generally be concerned with stochastic processes which have "nice" properties, chief among them is that they are measurable in some meaningful sense.

Motivation. Recall that a real-valued stochastic process $X = (X(t))_{t\in\mathbb{T}}$ can be characterized as family of a random variables mapping the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the reals. In particular, for each fixed $t \in \mathbb{T}$, $Y_t(\omega) = X(\omega, t)$ is a random variable not only from $(\Omega, \mathcal{F}, P)$ to $\mathbb{R}$, but also from the probability space $(\Omega, \mathcal{F}_t, P)$ to $\mathbb{R}$. As such, there should be form of measurability with respect to the σ-algebra $\mathcal{F}_t$ for each t. This notion is captured in the following definition. □

DEFINITION. A stochastic process $X = (X(t))_{t\in\mathbb{T}}$ on the filtered probability space $(\Omega, \mathcal{F}, F, \mathbb{P})$ is said to be F-***adapted*** (*a.k.a.* ***non-anticipating***) if $X(t)$ is $\mathcal{F}_t$-measurable for all $t \in \mathbb{T}$. When the filtration is understood, the process is simply called ***adapted***.

It turns out that simply being adapted (which is a point-wise in time property) is not sufficient to ensure the measurability of a sample path unless the paths are sufficiently regular. As such, other technical conditions will often be necessary. Generally, these conditions view a stochastic process as a random function mapping paths in $\Omega \times \mathbb{T}$ to paths $\mathbb{R} \times \mathbb{T}$. For the purpose of this paper, I will only be concerned with one such a technical condition known as *progressive measurability* which ensures that the path of a process stopped at any $t \in \mathbb{T}$ is observable.

DEFINITION. Given a filtered probability space $(\Omega, \mathcal{F}, F, \mathbb{P})$, a real-valued stochastic process $X = (X(t))_{t\in\mathbb{T}}$ on the probability space of the form $X : (\Omega \times \mathbb{T}; \mathcal{F} \otimes \mathcal{B}(\mathbb{T})) \to (\mathbb{R} \times \mathbb{T}; \mathcal{B} \otimes \mathcal{B}(\mathbb{T}))$ is called ***progressively measurable*** (***with respect to the filtration*** F) or simply ***progressive***, if, for each fixed time $t \in \mathbb{T}$, $X(\omega(t), t)$ is $\mathcal{F}_t \otimes \mathcal{B}(\mathbb{T})$-measurable where $\omega(\cdot) \in \Omega \times \mathbb{T}$.

REMARK. In the case of an adapted process, the end-point of each possible path is observable while in the case of a progressively measurable process, each possible path is fully observable. □

The following result captures the relationship between progressive and adapted processes which is proved in Mörters and Peres [**61**, pg. 191].

THEOREM 2.10. *Let $X = (X(t))_{t\in\mathbb{T}}$ be a stochastic process on the filtered probability space $(\Omega, \mathcal{F}, F, \mathbb{P})$. If X is progressive, then it is adapted. Conversely, if an adapted process is right or left continuous, then it is progressive.*

The following corollary follows directly from the fact that all stochastic process are stochastically equivalent to a càdlàg process (*cf.* Theorem 2.9).

COROLLARY 2.11. *Let $X = (X(t))_{t\in\mathbb{T}}$ be stochastic processes. Then it has a progressively measurable modification.*

REMARK. Given this corollary, I will assume that every process is progressive herein. □

Recall that in the case of a random variable, there exists a unique σ-algebra that is the smallest σ-algebra with respect to which that random variable is measurable (*cf.* Theorem 1.2). Thus, it seems seems reasonable to ask is there a "smallest" filtration with respect to which a stochastic process is adapted? Any such construct, however, must include the fact that once an event is observable, it is observable for all time.

Motivation. Let $X = (X_k)_{k\in\mathbb{T}}$ be a discrete-time stochastic process with a finite, countable index set $\mathbb{T} = \{0, \ldots, T\}$. Then for each $k \in \mathbb{T}$, the σ-algebra $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$ contains all the information about the process up to and including time k and is the smallest σ-algebra for which X_k is measurable (*cf.* Corollary 1.11). This suggests that in the continuous-time case, there should be some smallest filtration which is generated by each $X(t)$ such that $\sigma(X(s)) \subseteq \sigma(X(t))$ for each $s, t \in \mathbb{T}$ with $s \leq t$. In other words, $\sigma(X(t))$ contains all the information in $\sigma(X(s))$ for $s \leq t$ and $X(t)$ is measurable with respect to $\sigma(X(t))$. Importantly, since $\sigma(X(t))$ must contain all prior information, it must be generated by the sample paths of X up to time t and not simply the value of X at time t. This is captured in the following definition. □

DEFINITION. Let $X = (X(t))_{t\in\mathbb{T}}$ be a stochastic process. Then its ***natural filtration*** is the smallest filtration with respect to which X is adapted and denoted as $\tilde{F}^X = \left(\tilde{\mathcal{F}}_t^X\right)_{t\in\mathbb{T}}$ where

$$\tilde{\mathcal{F}}_t^X = \sigma(X(s) \mid t, s \in \mathbb{T} \text{ and } s \leq t)$$

REMARK. If $A \in \tilde{\mathcal{F}}_t^X$, then by time t, an observer knows if the event A occurred or not since $\tilde{\mathcal{F}}_s^X \subseteq \tilde{\mathcal{F}}_t^X$ for $s \leq t$. □

Motivation. Earlier, I defined a filtration to have the *usual conditions* if it was right-continuous, complete, and augmented. The first property ensures that, if for any fixed $t \in \mathbb{T}$, a process is $\mathcal{F}_s$-measurable for all $s > t$, then it is $\mathcal{F}_t$-measurable. The latter two properties prevent the situation in which two processes $X = (X(t))_{t\in\mathbb{T}}$ and $Y = (Y(t))_{t\in\mathbb{T}}$ are equal almost surely with X adapted, but not Y. Unfortunately, there is no guarantee that the natural filtration of a arbitrary continuous-time process will satisfy the usual conditions, especially right-continuity which is necessary to guarantee progressivity. This is accounted for in the following definition. □

DEFINITION. Let $X = (X(t))_{t\in\mathbb{T}}$ be a stochastic process. Then its ***standard filtration***, denoted as $F^X = \left\{\mathcal{F}_t^X\right\}_{t\in\mathbb{T}}$, is defined as

$$\mathcal{F}_t^X := \cap_{\varepsilon>0}\sigma\left(\tilde{\mathcal{F}}_{t+\varepsilon}^X \cup \mathcal{N}\right) \quad \text{for all } t \in \mathbb{T}.$$

where $\mathcal{N}$ is the collection of all null-sets of Ω.

The next result follows directly from the definitions.

THEOREM 2.12. *Given a given stochastic process* $X = (X(t))_{t\in\mathbb{T}}$*, its standard filtration* $F^X = \left(\mathcal{F}_t^X\right)_{t\in\mathbb{T}}$ *satisfies the usual conditions.*

Given the above results, I can now define admissible stochastic processes as follows.

DEFINITION. A real-valued stochastic process $X = (X(t))_{t\in\mathbb{T}}$ on the filtered probability space $(\Omega, \mathcal{F}, F, \mathbb{P})$ where F is its standard filtration is said to be ***admissible*** if it is adapted, càdlàg (hence, progressive), and summable for all time $t \in \mathbb{T}$.

Going forward, I will assume that all processes are admissible unless noted otherwise which is sufficient for most price processes. However, it is worth mentioning that many of the results in the sequel apply to much larger classes of stochastic process with few, if any, additional technical conditions.

2.3. Martingales

Much of modern mathematical finance involves a very rich class of stochastic processes known as *martingales*. These were first used to model fair games of chance and, as such, they are well-suited to model price dynamics in efficient markets. In fact, martingales play *the* central role in the risk-neutral approach to dynamic asset pricing.

For the most part, I have chosen to present the material in terms of proper martingales rather than a more general setting of local martingales or semi-martingales. While the more general setting allows for a expanded statement of stochastic calculus than that presented herein, the analysis becomes quite technical and is not necessary for this paper. See Jeanblanc, *et al.* [**47**] for a discussion of this material in the more general setting.

For this section, assume that $(\Omega, \mathcal{F}, F, \mathbb{P})$ is some filtered probability space and that all stochastic processes on $(\Omega, \mathcal{F}, F, \mathbb{P})$ are admissible. Further, assume that $\mathbb{T}$ is some bounded time index.

I shall begin by examining discrete-time martingales.

2.3.1. Discrete-Time Martingales. Motivation (Random walks as martingales)**.** Let me construct the following game that will be referred to throughout the discussion of martingales. Assume that I bet \$1 on the outcome of a coin toss. If the result is a head, I receive an additional dollar. If the result is tails, I lose the dollar. As seen in 2.2.1 above, this can be modeled by a symmetric random walk $S = (S_k)_{k\in\mathbb{T}}$ where $S_0 = C$ (*i.e.* my initial wealth). Then

$$S_k = S_{k-1} + X_k \quad (\text{for all } k > 0 \in \mathbb{T}),$$

and the $X_k \sim N(0,1)$ are i.i.d. random variables such that $X_k = \pm 1$ for all k. In other words, S_k represents my total wealth after k tosses starting with the initial wealth S_0. Now, the expectation

of S_k conditioned on the previous tosses is given by

$$\begin{aligned}\mathbb{E}\left[S_k \mid S_0, \ldots, S_{k-1}\right] &= \mathbb{E}\left[S_{k-1} + X_k \mid X_0, \ldots, X_{k-1}\right] \\ &= \mathbb{E}\left[S_{k-1} \mid X_0, \ldots, X_{k-1}\right] + \mathbb{E}\left[X_k \mid X_0, \ldots, X_{k-1}\right] \\ &= S_{k-1} + \mathbb{E}\left[X_k\right] \quad (\text{the } X_k \text{ are i.i.d.}) \\ &= S_{k-1} \quad (X_k \sim N(0,1)).\end{aligned}$$

Hence, the best guess of my wealth after the next coin toss based on what has happened so far is simply my current wealth. This is formalized in the following definition. □

DEFINITION. Let $X = (X_k)_{k\in\mathbb{T}}$ and $Y = (Y_k)_{k\in\mathbb{T}}$ be summable, discrete-time stochastic processes. Then for any $k \in \mathbb{T}$, we say that

(1) X is a (***discrete***) ***submartingale*** (***with respect to*** Y) if
$$\mathbb{E}\left[X_{k+1} \mid Y_k\right] \geq X_k \text{ a.s.;}$$
(2) X is (***discrete***) ***supermartingale*** (***with respect to*** Y) if
$$\mathbb{E}\left[X_{k+1} \mid Y_k\right] \leq X_k \text{ a.s.; and}$$
(3) X is (***discrete***) ***martingale*** (***with respect to*** Y) if it is submartingale and supermartingale, or
$$\mathbb{E}\left[X_{k+1} \mid Y_k\right] = X_k \text{ a.s.}$$

REMARK. An important feature of this definition of a martingale is that Y can be any stochastic process on the same filtered probability space. In other words, it does not have to be related to X. □

By definition, a martingale is *non-anticipatory* in the sense that the best guess about the future state of the process is simply its current state. This is captured in the following result.

THEOREM 2.13. *Let $X = (X_k)_{k\in\mathbb{T}}$ be a martingale with respect to $Y = (Y_k)_{k\in\mathbb{T}}$. Then for any $j, k \in \mathbb{T}$ with $j \leq k$,*
$$\mathbb{E}\left[X_k \mid Y_j\right] = X_j \text{ a.s.}$$
Moreover, if X is submartingale, for $j, k \in \mathbb{T}$ with $j \leq k$,
$$\mathbb{E}\left[X_k \mid Y_j\right] \geq X_j \text{ a.s.}$$
and if X is submartingale, then for any $j, k \in \mathbb{T}$ with $j \leq k$,
$$\mathbb{E}\left[X_k \mid Y_j\right] \leq X_j \text{ a.s.}$$

PROOF. If X is a martingale with respect to Y, then for any $j, k \in \mathbb{T}$ with $j \leq k$,
$$\mathbb{E}\left[X_k \mid Y_{k-1}\right] = X_{k-1} \text{ a.s.}$$

By the tower principle (*cf.* Theorem 1.59), the expected value of the left-hand side of the above expression conditioned on Y_{k-2} becomes

$$\mathbb{E}\left[\mathbb{E}\left[X_k \mid Y_{k-1}\right] \mid Y_{k-2}\right] \;=\; \mathbb{E}\left[X_k \mid Y_{k-2}\right] \text{ a.s.}$$

Hence, by repeatedly applying the tower principle, the left-hand side becomes $\mathbb{E}\left[X_k \mid Y_j\right]$ for any $j \in \mathbb{T}$ such that $j \leq k$.

On the other hand, since X is a martingale with respect to Y, the expected value of the right-hand side of the above expression conditioned on Y_{k-2} is

$$\mathbb{E}\left[X_{k-1} \mid Y_{k-2}\right] \;=\; X_{k-2}.$$

Hence, by repeatedly applying the tower principle, the right-hand side becomes X_j. Therefore, combining the two results

$$\mathbb{E}\left[X_k \mid Y_j\right] = X_j.$$

A similar argument proves the other statements. □

The above theorem is true for all choices of $j \leq k$, in particular $j = 0$. The gives the following corollary.

COROLLARY 2.14. *Let $X = (X_k)_{t\in\mathbb{T}}$ be a martingale with respect to $Y = (Y_k)_{k\in\mathbb{T}}$. Then for any $k \in \mathbb{T}$,*

$$\mathbb{E}\left[X_k \mid Y_0\right] = X_0 \text{ a.s.}$$

Moreover, for any $k \in \mathbb{T}$, if X is submartingale,

$$\mathbb{E}\left[X_k \mid Y_0\right] \geq X_0 \text{ a.s.}$$

and if X is supermartingale,

$$\mathbb{E}\left[X_k \mid Y_0\right] \leq X_0 \text{ a.s.}$$

Since these results hold for any arbitrary summable, discrete-time stochastic process $Y = (Y_k)_{k\in\mathbb{T}}$, they hold when $Y = X$. This is captured in the following result.

COROLLARY 2.15. *Let $X = (X_k)_{t\in\mathbb{T}}$ be a martingale. Then for any $k \in \mathbb{T}$,*

$$\mathbb{E}\left[X_k \mid X_0\right] = X_0 \text{ a.s.}$$

Moreover, for any $k \in \mathbb{T}$, if X is submartingale,

$$\mathbb{E}\left[X_k \mid X_0\right] \geq X_0 \text{ a.s.}$$

and if X is supermartingale,

$$\mathbb{E}\left[X_k \mid X_0\right] \leq X_0 \text{ a.s.}$$

We saw at the beginning of this section that the coin-tossing game was a martingale. Given the above result, after k tosses, I would expect that $\mathbb{E}[S_k] = S_0$. In other words, my winnings and losses should net to be zero over the long haul. As such, the game is considered to be *fair* since it does not favor either me or the house. On the other hand, if the game had been modeled by a submartingale, then $\mathbb{E}[S_k] \geq S_0$ and my wealth would tend to increase over time. Hence, the game is biased in my favor. Conversely, if the game had been modeled by a supermartingale, then $\mathbb{E}[S_k] \leq S_0$ and my wealth would tend to decrease over time. Thus, the game is biased in favor of the house.

In terms of continuous-time finance, martingales represent fair markets in which agents acting in those markets cannot know future outcomes with certainty based solely on past information. A submartingale favors an investor since his wealth is expected to grow while an investor should expect to lose wealth in the case of a supermartingale.

2.3.2. Continuous-Time Martingales. Continuous-time stochastic processes can also be martingales. Here, it is common to model the evolution of information using some filtration rather than the history of a random variable.

DEFINITION. Let $X = (X(t))_{t\in\mathbb{T}}$ be an admissible processes adapted to the filtration $F = \{\mathcal{F}_t\}_{t\in\mathbb{T}}$. Then for any $s, t \in \mathbb{T}$ such that $s < t$, we say that

(1) X is a ***submartingale*** (***with respect to*** F) if
$$\mathbb{E}[X_t \mid \mathcal{F}_s] \geq X_s \text{ a.s.;}$$
(2) X is ***supermartingale*** (***with respect to*** F) if
$$\mathbb{E}[X_t \mid \mathcal{F}_s] \leq X_s \text{ a.s.; and}$$
(3) X is ***martingale*** (***with respect to*** F) if it is submartingale and supermartingale, or
$$\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s \text{ a.s.}$$

The next result follows directly from the definitions.

THEOREM 2.16. *Let $X = (X(t))_{t\in\mathbb{T}}$ be an admissible process.*

(1) *If X is a martingale, then $(X(t) - X(0))_{t\in\mathbb{T}}$ is also a martingale; and*
(2) *If X is a submartingale, $-X$ is a supermartingale and vice versa.*

Given this result, I can assume that $X(0) = 0$ for all martingales without loss of generality. Also, any results pertaining to supermartingales apply equally to submartingales.

Motivation. Assume that I adopt a strategy to stop playing the above coin tossing game once I have won a sufficient amount (or, equivalently, when I have lost a certain amount). In this case, is it possible to predict when that might occur? Clearly, such a time is a random event. Therefore, to answer such a question, such a time must be observable. The following class of random variables, known as *stopping times*, will help answer this question. □

DEFINITION. A ***stopping time*** (***with respect to the filtration*** F) (*a.k.a.* a ***Markov time***) is a random variable $\tau : \Omega \to \mathbb{T}$ such that for every $t \in \mathbb{T}$ and $\mathcal{F}_t \in F$

$$\{\tau \leq t\} = \{\omega \in \Omega \mid \tau(\omega) \leq t\} \in \mathcal{F}_t.$$

In other words, all possible stopping times up to a given time t are observable.

THEOREM 2.17. *Let $F = (\mathcal{F}_t)_{t\in\mathbb{T}}$ be a right continuous filtration and suppose a given random variable $\tau : \Omega \to \mathbb{T}$ satisfies*

$$\{\tau < t\} \in \mathcal{F}_t \quad (\textit{for every } t \in \mathbb{T}).$$

Then τ is a stopping time with respect to F.

PROOF. Let τ satisfy the theorem's hypothesis. Then for any fixed $t \in \mathbb{T}$, pick ε small enough such that $t + 1/\varepsilon \in \mathbb{T}$. Hence,

$$\{\tau < t + 1/\varepsilon\} \in \mathcal{F}_{t+1/\varepsilon}.$$

Thus,

$$\cap_{\epsilon\to\infty} \{\tau < t + 1/\varepsilon\} \in \cap_{\epsilon\to\infty} \mathcal{F}_{t+1/\varepsilon}.$$

Now, the left-hand side is

$$\cap_{\epsilon\to\infty} \{\tau < t + 1/\varepsilon\} = \{\tau < t\}.$$

On the other hand, since F is right-continuous, the right-hand side is

$$\cap_{\epsilon\to\infty} \mathcal{F}_{t+1/\varepsilon} = \mathcal{F}_t.$$

Hence, $\{\tau < t\} = \mathcal{F}_t$. □

DEFINITION. Let $X = (X(t))_{t\in\mathbb{T}}$ be a stochastic process. Then the process $X^\tau = (X^\tau(t))_{t\in\mathbb{T}}$ defined as

$$X^\tau(t) := X(\tau \wedge t) \quad (\text{for all } t \in \mathbb{T})$$

is called a ***stopped process*** (***at*** τ) where τ is a stopping time.

The following theorem, which is known by a variety of names and is stated in various forms, characterizes martingales in terms of stopped processes.

THEOREM 2.18 (Doob's optional sampling theorem). *Let $X = (X(t))_{t\in\mathbb{T}}$ be a stochastic process. Then X is a martingale if and only if for all finite stopping times τ, X^τ is a martingale. In particular, X is a martingale if and only if*

$$\mathbb{E}[X(\tau)] = X(0) \quad (\textit{for all stopping times } \tau).$$

PROOF. See Williams [**88**, A14.3 and A14.4] □

2.4. Markov Processes

Markov processes are another exceptionally rich class of stochastic processes that are used to represent the random evolution of systems that are *memoryless* (*i.e.* future outcomes do not depend the full history of past outcomes). Essentially, such processes can be thought of as starting anew at any point in time. Mathematically, this is captured by what is known as the *Markov property.*

In terms of continuous-time finance, the Markov property is important since it basically represents the belief that all the information necessary to estimate future prices is observable in the market today. Knowing the history of price movements provides no additional value than simply knowing the current price. In other words, the current price level is all that is important, not the path it took to get there (this is a big blow for adherents of a school of investments known as *technical analysis*).

For this section, assume that $(\Omega, \mathcal{F}, F, \mathbb{P})$ is some filtered probability space and that all stochastic processes on $(\Omega, \mathcal{F}, F, \mathbb{P})$ are admissible. Further, assume that $\mathbb{T}$ is some bounded time index.

Motivation (Random walks as Markov processes)**.** Recall that for some discrete-time random process $X = (X_k)_{k\in\mathbb{T}}$where X_k are independent, standard normal random variables for all $k \in \mathbb{T}$, a symmetric random walk $S = (S_k)_{k\in\mathbb{T}}$ is a discrete-time, real-valued stochastic process such that

$$S_{k+1} = S_k + X_{k+1} \quad (\text{for any } k \in \mathbb{T}).$$

As such, the value at the next state only depends on the current state (*i.e.* S_k) and not the path it took to get there (*i.e.* $S_1, \ldots, S_{k-1}$).

Mathematically, I can state this lack of dependence on past history as

$$\mathbb{P}(S_{k+1} = x \mid S_1 = x_1, \ldots, S_k = x_k) = P(S_{k+1} = x \mid S_k = x_k)$$

where $x, x_1, \ldots x_k \in \mathbb{R}$ for all $k \in \mathbb{T}$.

To see this, by the definition of conditional probability (*cf.* on page 19)

$$\mathbb{P}(S_{k+1} = x \mid S_1 = x_1, \ldots, S_k = x_k) = \frac{\mathbb{P}(S_1 = x_1, \ldots, S_k = x_k, S_{k+1} = x)}{\mathbb{P}(S_1 = x_1, S_2 = x_2, \ldots, S_k = x_k)}.$$

Using the fact that $S_j = x_j$ for $1 \leq j \leq k$,

$$\begin{aligned}\mathbb{P}(S_{k+1} = x \mid S_1 = x_1, \ldots, S_k = x_k) &= \frac{\mathbb{P}(S_1 = x_1, \ldots, S_k = x_k, S_{k+1} = x)}{\mathbb{P}(S_1 = x_1, S_2 - S_1 = x_2 - x_1, \ldots, S_k = x_k)} \\ &= \frac{\mathbb{P}(S_1 = x_1, S_2 - S_1 = x_2 - x_1, \ldots, S_{k+1} - S_k = x - x_k)}{\mathbb{P}(S_1 = x_1, S_2 - S_1 = x_2 - x_1, \ldots, S_k - S_{k-1} = x_k - x_{k-1})}.\end{aligned}$$

Recall that for any $k \in \mathbb{T}$, S_k is given by

$$S_k = \sum_{j=1}^{k} X_j.$$

Hence, for any $k \in \mathbb{T}$,

$$x_k - x_{k-1} = X_k.$$

Therefore,

$$\mathbb{P}\left(S_{k+1} = x \mid S_1 = x_1, \ldots, S_k = x_k\right) = \frac{\mathbb{P}\left(X_1 = x_1, X_2 = x_2 - x_1, \ldots, X_{k+1} = x - x_k\right)}{\mathbb{P}\left(X_1 = x_1, X_2 = x_2 - x_1, \ldots, X_k = x_k - x_{k-1}\right)}.$$

But, the X_k are i.i.d., thus

$$\begin{aligned}
&= \frac{\mathbb{P}\left(X_{k+1} = x - x_k\right)\mathbb{P}\left(X_1 = x_1, X_2 = x_2 - x_1, \ldots, X_k = x_k - x_{k-1}\right)}{\mathbb{P}\left(X_1 = x_1, X_2 = x_2 - x_1, \ldots, X_k = x_k - x_{k-1}\right)} \\
&= \mathbb{P}\left(X_{k+1} = x - x_k\right) \\
&= \frac{\mathbb{P}\left(X_{k+1} = x - x_k\right)\mathbb{P}\left(S_k = x_k\right)}{\mathbb{P}\left(S_k = x_k\right)} \\
&= \frac{\mathbb{P}\left(X_{k+1} = x - x_k, S_k = x_k\right)}{\mathbb{P}\left(S_k = x_k\right)} \quad \left(X_{k+1} \perp S_k\right) \\
&= \mathbb{P}\left(X_{k+1} + S_k = x \mid S_k = x_k\right) \\
&= \mathbb{P}\left(S_{k+1} = x \mid S_k = x_k\right) \quad \left(S_{k+1} = S_k + X_{k+1}\right).
\end{aligned}$$

This property is captured in the following definition. □

DEFINITION 2.19. An admissible stochastic process $X = \left(X\left(t\right)\right)_{t\in\mathbb{T}}$ is said to be ***Markov*** if for each $t, s \in \mathbb{T}$ with $s \leq t$ and any $B \in \mathcal{B}$

$$\mathbb{P}\left(X\left(t\right) \in B \mid \mathcal{F}_s\right) = \mathbb{P}\left(X\left(t\right) \in B \mid X\left(s\right)\right). \tag{2.4.1}$$

Essentially, the Markov property implies that probability of the future of a process depends entirely on its current state, not the path it took to get there. We can impose a stronger condition that if a process is stopped at any time and then allowed to proceed, it can be viewed as beginning anew. In other words, the sample paths from t_0 to t are "indistinguishable" from those from $t_0 + s$ to $t + s$ for some reasonable s. This is captured in the following definition.

DEFINITION. Let $X = \left(X\left(t\right)\right)_{t\in\mathbb{T}}$ be a continuous-time Markov process. For some positive s, define the process $Y = \left(Y\left(t\right)\right)_{t\in\mathbb{T}}$ as

$$Y\left(t\right) := X\left(t + s\right)$$

for all $t \in \mathbb{T}$ such that $t + s \in \mathbb{T}$. Then X is said to be ***time-homogeneous*** if $Y\left(t\right) \overset{d}{=} X\left(t\right)$ for some fixed s such that $t + s \in \mathbb{T}$.

Recall that the σ-algebra $\mathcal{F}_s$ contains all observable information about any adapted process up to and including time s. Therefore, its seems reasonable that the Markov property can be defined in terms conditional expectation. This is captured in the following two results which are proved in Shreve [**76**, ch. 2.4].

THEOREM 2.20. *Let $X = (X(t))_{t\in\mathbb{T}}$ be an admissible stochastic process. Then for each $t, s \in \mathbb{T}$ with $s \leq t$*

$$\mathbb{E}[X(t) \mid \mathcal{F}_s] = \mathbb{E}[X(t) \mid X(s)].$$

COROLLARY 2.21. *If the admissible process $X = (X(t))_{t\in\mathbb{T}}$ is Markov, then for all $s, t \in \mathbb{T}$ with $0 \leq s \leq t$ and for every non-negative, measurable function f, then there is another measurable function g such that*

$$\mathbb{E}[f(X(t)) \mid \mathcal{F}_s] = g(X(s)).$$

PROOF. The result follows from the fact that the conditional expectation on $X(s)$ is both a measurable random variable and a function of $X(s)$. □

REMARK. As discussed in the prequel, a filtration can be thought of as containing all information about the markets up to the current time. However, since most price processes discussed in this paper are Markov, all the information that an investor needs to predict the future price of a particular instrument is its current price. In other words, Markov price processes are *path-independent*. □

2.4.1. Markov Chains. To understand the Markov property better, let me first consider a discrete-time, discrete-state stochastic process with the Markov property process that is commonly called a *Markov chain*. For this discussion, which generally follows Lawler [**57**, ch. 1,2, and 3], assume that $X = (X_k)_{k\in\mathbb{T}}$ is a countable sequence of real-valued random variables indexed by $\mathbb{T} = \{0, 1, \ldots, T\}$ taking discrete values in the state space $\mathbb{R}$ indexed by an ordered countable set $S \subset \mathbb{N}$.

DEFINITION. The discrete stochastic process $X = (X_k)_{k\in\mathbb{T}}$ is said to be a (***discrete***) ***Markov chain*** if

$$\mathbb{P}(X_{k+1} = x \mid X_0 = x_0, \ldots, X_k = x_k) = \mathbb{P}(X_{k+1} = x \mid X_k = x_k)$$

where $k \in \mathbb{T}$ and $x, x_k \in \mathbb{R}$ for all k. In other words, given the sample path up to step k, $(X_0 = x_0, \ldots, X_k = x_k)$, the next state in the chain X_{k+1} depends only on X_k and not the path it took to get there.

REMARK. By definition, the future state of a Markov chain only depends on its current and not its past states. In other words, the future state is independent of its past states. □

EXAMPLE. Based on the calculations above, a symmetric random walk is an example of a Markov chain (*cf.* on page 63). □

Given the above definition, it is possible to define a function that expresses the probability of transitioning from one discrete state to another at any given time step as follows.

DEFINITION. Let $X = (X_k)_{k\in\mathbb{T}}$ be a discrete Markov chain. Then for any $k \in \mathbb{T}$, define the function $\rho_k{:}S \times S \to [0, 1]$ as

$$\rho_k(i, j) := \mathbb{P}(X_{k+1} = x_j \mid X_k = x_i).$$

where $i, j \in S$ and $x_i, x_j \in E$ for all i, j. Then $\rho_k(i,j)$ is said to be a ***transition probability*** from state x_i to x_j at time step $k+1$.

REMARK. In other words, $\rho_k(i,j)$ is the probability of being in the state j at step $k+1$ if the chain is at state i at step k. □

The following class of Markov chains will be very useful in the following discussion.

DEFINITION. Let $X = (X_k)_{k \in \mathbb{T}}$ be a discrete Markov chain. If the transition probabilities are independent of time for all $i, j \in S$, or

$$\rho_k(i,j) = \rho_m(i,j) \quad (\text{for any } k, m \in \mathbb{T}),$$

it is said to be ***time-homogeneous***.

REMARK. In other words, if a Markov chain is time-homogeneous, the probability of transition for one particular state to another is the same for all time. qedsymbol

REMARK. I will assume for the remainder of the discussion that all Markov chains are time-homogeneous unless otherwise noted. □

DEFINITION. Let $X = (X_k)_{k \in \mathbb{T}}$ be a discrete, real-valued Markov chain. Then the (possibly infinite) matrix P_k composed of all transition probabilities at step $k \in \mathbb{T}$,

$$\mathrm{P}_k^{ij} := \rho_k(i,j) = \mathbb{P}(X_{k+1} = x_j \mid X_k = x_i)$$

where $i, j \in S$ and $x_i, x_j \in \mathbb{R}$ for all i, j is called the ***transition*** (***probability***) ***matrix*** at step k. Consequently, X is called the ***Markov chain with transition matrix*** P.

NOTATION. Since I am assuming that all Markov chains are time-homogeneous, I will drop the time index subscript and simply refer to the transition matrix as P. □

The next result follows directly from the definitions above.

THEOREM 2.22. *Let $X = (X_k)_{k \in \mathbb{T}}$ be a time-homogeneous Markov chain with transition matrix* P. *Then*

(1) P *is unique,*
(2) $0 \leq \mathrm{P}^{ij} \leq 1 \quad i, j \in S$, *and*
(3) *For any fixed $i \in S$, $\sum_{j \in S} \mathrm{P}^{ij} = 1$.*

Now, assume for the moment that state space is countably finite (hence, P is finite).

Motivation. Using its transition matrix, it is possible to determine the probability of a time-homogeneous Markov chain moving from the state x_i to x_j in two time steps as follows. For some fixed $k \in \mathbb{T}$,

$$\mathbb{P}(X_{k+2} = x_j \mid X_k = x_i) = \sum_{l \in S} \mathbb{P}(X_{k+2} = x_j, X_{k+1} = x_l \mid X_k = x_i)$$

Since the process is time-homogeneous, by the Markov property

$$\begin{aligned}\mathbb{P}\left(X_{k+2}=x_j \mid X_k=x_i\right) &= \sum_{l\in S}\mathbb{P}\left(X_{k+2}=x_j, X_{k+1}=x_l \mid X_n=x_i\right)\\ &= \sum_{l\in S}\mathbb{P}\left(X_{k+2}=x_j \mid X_{k+1}=x_l\right)\mathbb{P}\left(X_{k+1}=x_l \mid X_k=x_i\right)\\ &= \sum_{l\in S}\mathrm{P}^{il}\mathrm{P}^{lj}\\ &= \left(\mathrm{P}^2\right)^{ij}.\end{aligned}$$

This can be generalized using induction as follows. □

THEOREM 2.23. *Let $X=(X_k)_{k\in\mathbb{T}}$ be a discrete Markov chain with the transition matrix* P. *Then for any $i,j\in S$ and $k,m\in\mathbb{T}$*

$$\mathbb{P}\left(X_{k+m}=x_j \mid X_k=x_i\right)=\left(\mathrm{P}^m\right)^{ij} \quad (x_i,x_j\in\mathbb{R}).$$

Since this is true for any starting time k, in particular, it is true for $k=0$.

COROLLARY 2.24. *Let $X=(X_k)_{k\in\mathbb{T}}$ be a discrete Markov chain with the transition matrix* P. *Then for any $i,j\in S$ and $m\in\mathbb{T}$*

$$\mathbb{P}\left(X_m=x_j \mid X_0=x_i\right)=\left(\mathrm{P}^m\right)^{ij} \quad (x_i,x_j\in\mathbb{R}).$$

Motivation. Up until now, I have assumed that P is countably finite. Then for finite matrices,

$$\mathrm{P}^{k+m}=\mathrm{P}^k\mathrm{P}^m. \tag{2.4.2}$$

Hence, for $i,j\in S$

$$\left(\mathrm{P}^{k+m}\right)^{ij}=\sum_{l\in S}\left(\mathrm{P}^k\right)^{il}\left(\mathrm{P}^m\right)^{lj}. \tag{2.4.3}$$

By Theorem 2.22, this is well-defined for an infinite transition matrix as well. Hence, I can relax the requirement that the state space be countably finite and consider countable Markov chains generally. □

Through the next result, which is known as the *Chapman-Kolmogorov equation*, equation (2.4.2) shows the family of transition matrices process the *semigroup property* which will be discussed at length in the sequel (*cf.* [TBD]).

THEOREM 2.25 (Chapman-Kolmogorov equation for Markov chains). *Let $X=(X_k)_{k\in\mathbb{T}}$ be a discrete Markov chain with transition matrix* P. *Then for any $i,j\in S$ and $k,m\in\mathbb{T}$*

$$\left(\mathrm{P}^{k+m}\right)^{ij}=\sum_{l\in S}\left(\mathrm{P}^k\right)^{il}\left(\mathrm{P}^m\right)^{lj}$$

or more generally

$$\mathrm{P}^{k+m}=\mathrm{P}^k\mathrm{P}^m.$$

Given these results, it is clear that the future states of any Markov chain are determined entirely from its transition matrix and the probability distribution of its initial state. This is captured in the following result.

THEOREM 2.26. *Let $X = (X_k)_{k\in\mathbb{T}}$ be a discrete Markov chain with transition matrix* P*. If X's initial probability distribution is ϕ_i for all $i \in S$, then the probability of being in the j^{th} state after m time steps is given*

$$\mathbb{P}(X_m = x_j) = \sum_{i\in S} \phi_i \left(\mathrm{P}^m\right)^{ij}.$$

PROOF. Let $X = (X_k)_{k\in\mathbb{T}}$ be a discrete Markov chain with transition matrix P and initial distribution ϕ_i for all $i \in S$. Then

$$\phi_i = \mathbb{P}(X_0 = x_i).$$

Thus, by the definition of conditional probability (*cf.* on page 19) and Theorem 2.23,

$$\begin{aligned}\mathbb{P}(X_m = x_j) &= \sum_{i\in S} \mathbb{P}(X_0 = x_i)\,\mathbb{P}(X_m = x_j \mid X_0 = x_i)\\ &= \sum_{i\in S} \phi_i \left(\mathrm{P}^m\right)^{ij}.\end{aligned}$$

□

NOTATION. Recall that vectors are in column major form (*i.e.* $x = (x^1, \ldots, x^n, \ldots)^{\mathrm{T}}$). Hence, the above expression can be written in vector notion as

$$\mathbb{P}(X_m = x_j) = (\mathrm{P}^m\boldsymbol{\phi})\cdot\mathbf{e}^j = (\mathrm{P}^m\boldsymbol{\phi})^j \tag{2.4.4}$$

where $\boldsymbol{\phi} = (\phi^1, \ldots, \phi^n, \ldots)^{\mathrm{T}}$ are the probabilities of the initial states and $\mathbf{e}^j$ is the j^{th} unit vector.□

Importantly, transition probabilities can be used to compute the expectation of any arbitrary function acting on a Markov chain.

THEOREM 2.27. *Let $X = (X_k)_{k\in\mathbb{T}}$ be a discrete, real-valued Markov chain with transition matrix* P *and initial distribution $\boldsymbol{\phi}$. Then for any summable function f,*

$$\mathbb{E}[f(X_m)] = (\mathrm{P}^m\boldsymbol{\phi})\,\mathbf{f}$$

where $\mathbf{f}(x) = (f(x_1), \ldots, f(x_m), \ldots)^{\mathrm{T}}$ for any $x = (x_1, \ldots, x_m, \ldots) \in \mathbb{R}^\infty$.

PROOF. Given a summable function f acting on a random variable $X_0 : (\Omega; \mathcal{F}) \to (E; \mathcal{E})$, its expected value is

$$\mathbb{E}[f(X_0)] = \int_\Omega f(X_0)\, d\mathbb{P}.$$

In particular, when the sample space is discrete the expected value is simply the probability weighted average of the observable outcomes in the state space, or

$$\mathbb{E}[f(X_0)] = \sum_{i\in S} f(x_i)\,\phi^i$$

where $\phi^i := \mathbb{P}(X_0 = x_i)$ with $x_i \in E$ for all $i \in S$. Then the conditional expectation of f given X_0 at the m^{th} time step is

$$\begin{aligned}\mathbb{E}[f(X_m) \mid X_0 = x_i] &= \sum_{j\in S} f(x_i)\,\mathbb{P}(X_m = x_j \mid X_0 = x_i)\\ &= \sum_{j\in S} (\mathrm{P}^m)^{ij} f(x_j)\\ &= (\mathrm{P}^m\mathbf{f})^i.\end{aligned}$$

Then using the properties of conditional expectation (*cf.* Theorem 1.59)

$$\begin{aligned}\mathbb{E}[f(X_m)] &= \mathbb{E}[\mathbb{E}[f(X_m) \mid X_0 = x_i]]\\ &= \sum_{i\in S} (\mathrm{P}^m\mathbf{f})^i \phi^i\\ &= (\mathrm{P}^m\mathbf{f})^{\mathrm{T}} \boldsymbol{\phi}.\end{aligned}$$

□

2.4.2. Transition Probability Functions. As was seen above, a Markov chain is completely characterized by its initial distribution and its transition matrix. This can be extended to continuous-time processes as follows.

For the following discussing, all Markov processes are assumed to be continuous-time processes on some filtered probability space $(\Omega, \mathcal{F}, F, \mathbb{P})$.

Motivation. Assume $X = (X(t))_{t\in\mathbb{T}}$ is a real-valued Markov process. Then for any $s, t \in \mathbb{T}$ such that $s \le t$, by the Markov property

$$\mathbb{P}(X(t) \in B \mid \mathcal{F}_s) = \mathbb{P}(X(t) \in B \mid X(s) = x) \quad (\text{for any } x \in \mathbb{R}, B \in \mathcal{B}).$$

This can be interpreted as the probability that the process will be in the set B at time t after starting at x at time s. This suggest defining a transition probability for a continuous-time Markov process as

$$p_{s,t}(x, B) := \mathbb{P}(X(t) \in B \mid X(s) = x).$$

NOTATION. The transition probability is commonly expressed as $p(x, s, B, t)$ to emphasis that it is function of the starting and ending times s and t respectively. □

Now, consider the probability of the processing ending up in the state $B \in \mathcal{B}$ at time $t + r$ after starting at $x \in \mathbb{R}$ at time $s \in \mathbb{T}$ for some reasonable time r, or

$$\mathbb{P}(X(t + r) \in B \mid X(s) = x).$$

Given what we learned from discrete-time Markov chains, it seems reasonable that this can be computed using the Chapman-Kolmogorov equation. In other words, we first consider all paths

stopping at time t which started at x at some initial time s and then consider all the paths that start at time t and end up in B at time $t+r$, or

$$\begin{aligned}\mathbb{P}\left(X\left(t+r\right)\in B\mid X\left(s\right)=x\right) &= \int_{\mathbb{R}}\mathbb{P}\left(X\left(r\right)\in B\mid X\left(t\right)=y\right)\mathbb{P}\left(X\left(t\right)=dy\mid X\left(s\right)=x\right)\\ &= \int_{\mathbb{R}}p_{t,r}\left(y,B\right)p_{s,t}\left(x,dy\right).\end{aligned}$$

Recalling that $p_{s,t+r}\left(x,B\right)=\mathbb{P}\left(X\left(t+r\right)\in B\,|X\left(s\right)=x\right)$, this gives the continuous-time Chapman-Kolmogorov equation

$$p_{s,t+r}\left(x,B\right)=\int_{\mathbb{R}}p_{t,r}\left(y,B\right)p_{s,t}\left(x,dy\right).$$

Further, it follows that for some "nice" real-valued function f on

$$\begin{aligned}\mathbb{E}\left[f\left(X\left(t\right)\right)\mid X\left(s\right)=x\right] &= \int_{1_{\{X(s)=x\}}}f\left(y\right)\mathbb{P}\left(dy\right)\\ &= \int_{\mathbb{R}}f\left(y\right)\mathbb{P}\left(X\in dy\mid X\left(s\right)=x\right)\\ &= \int_{\mathbb{R}}f\left(y\right)p_{s,t}\left(x,dy\right).\end{aligned}\tag{2.4.5}$$

In particular, letting $p_t := p_{0,t}$

$$p_t\left(x,B\right)=\mathbb{P}\left(X\left(t\right)\in B\mid X\left(0\right)=0\right)$$

and

$$\mathbb{E}\left[f\left(X\left(t\right)\right)\mid X\left(0\right)=x\right]=\int_{\mathbb{R}}f\left(y\right)p_t\left(x,dy\right).\tag{2.4.6}$$

NOTATION. I will typically use the following notation for such conditional expectation throughout the sequel

$$\mathbb{E}^{x,s}\left[f\left(X\left(t\right)\right)\right]:=\mathbb{E}\left[f\left(X\left(t\right)\right)\mid X\left(s\right)=x\right]$$

and, in particular,

$$\mathbb{E}^{x}\left[f\left(X\left(t\right)\right)\right]:=\mathbb{E}\left[f\left(X\left(t\right)\right)\mid X\left(0\right)=x\right].$$

This is formalized as follows. □

DEFINITION. Given some fixed $t\in\mathbb{T}$ and filtered probability space $(\Omega,\mathcal{F},F,\mathbb{P})$, the function p_t on $\mathbb{R}\times\mathcal{B}$ is called a ***transition*** (***probability***) ***function*** if

(1) $p_t\left(x,\cdot\right)$ is a non-negative measure on $\mathcal{B}$ such that $p_t\left(x,B\right)\leq 1$ for all $x\in\mathbb{R}$ and $B\in\mathcal{B}$ (*i.e.* $p_t:\mathbb{R}\times\mathcal{B}\to[0,1]$);
(2) $p_t\left(\cdot,B\right)$ is a measurable function for all $B\in\mathcal{B}$;

(3) $p_0(x, \cdot) = \delta_x(\cdot)$ for all $x \in \mathbb{R}$; and
(4) For any $t, s \in \mathbb{T}, B \in \mathcal{B}$, and $x \in \mathbb{R}$

$$p_{t+s}(x, B) = \int_{\mathbb{R}} p_t(y, B)\, p_s(x, dy).$$

Further, if $p_t(\cdot, \mathbb{R}) = 1$, then $p_t(\cdot, \cdot)$ is called a ***Markov transition probability function***.

REMARK. A transition probability function is also commonly called a ***transition kernel***. The reason will become clear shortly (*cf.* remark on the following page). □

REMARK. All transition probability functions discussed in this paper will be Markov. Hence, they will be referred to simply as *transition probability functions.* □

DEFINITION. If a Markov process admits a family of transition probability functions $\{p_t(\cdot, \cdot)\}_{t\in\mathbb{T}}$, it is said to be a ***Markov process with transition probability functions*** $\{p_t\}_{t\in\mathbb{T}}$.

REMARK. All of the Markov processes of interest in this paper admit a family of transition probability functions. For an example of a Markov process without a family transition probability functions, see Çinlar [**14**, pg. 416]. □

Given these definitions, it seems reasonable to ask: (a) *under what conditions do Markov processes admit transition probability functions*? (b) *is it possible to construct a Markov process from a family of transition probability functions*? and (c) *does a family of transition probability functions uniquely characterize a Markov process*? The following results help answer these questions the first of which follows directly from the definitions and the motivation above.

THEOREM 2.28. *Let $X = (X(t))_{t\in\mathbb{T}}$ be a real-valued Markov process. Then for a fixed $t \in \mathbb{T}$, the function p_t defined as*

$$p_t(x, B) = \mathbb{P}(X(t) \in B \mid X(0) = x) \quad (\textit{for any } x \in \mathbb{R}, B \in \mathcal{B}, \textit{ and } t \in \mathbb{T}),$$

is a transition probability function.

In turns out that a Markov process is completely determined by its transition probability functions which is proved in Stroock (*cf.* [**80**, pg. 84]).

THEOREM 2.29. *The distribution of any Markov process is uniquely determined by its initial distribution and its family of transition probability functions.*

As in the case of Markov chains, I will generally be concerned with time-homogeneous transition probability functions in the sequel.

DEFINITION. Let $X = (X_k)_{k\in\mathbb{T}}$ be a real-valued Markov process with transition probability functions $\{p_t\}_{t\in\mathbb{T}}$. If the transition probabilities are independent of time for all $x \in \mathbb{R}, B \in \mathcal{B}$, or, for any suitable r

$$p_{s,t}(x, B) = p_{s+r,t+r}(x, B) \quad (\text{for any } s, t \in \mathbb{T}),$$

it is said to be ***time-homogeneous***.

The next result follows directly from (2.4.5).

THEOREM 2.30. *Let $X = (X(t))_{t\in\mathbb{T}}$ be a real-valued Markov process. Then for any summable function f and any $s,t \in \mathbb{T}$ such that $s \leq t$ and $t+s \in \mathbb{T}$*

$$\mathbb{E}^{x,s}\left[f\left(X\left(t+s\right)\right)\right] = \int_{\mathbb{R}} f\left(y\right) p_{s,t}\left(x, dy\right).$$

In particular, when $s = 0$

$$\mathbb{E}^{x}\left[f\left(X\left(t\right)\right)\right] = \int_{\mathbb{R}} f\left(y\right) p_{t}\left(x, dy\right).$$

If the process is time-homogeneous, then

$$\mathbb{E}^{x,s}\left[f\left(X\left(t+s\right)\right)\right] = \int_{\mathbb{R}} f\left(y\right) p_{t}\left(x, dy\right).$$

EXAMPLE. Let $X = (X(t))_{t\in\mathbb{T}}$ be a real-valued Gaussian process with mean x and variance t. Then for each $0 < t \in \mathbb{T}$ and $B \in \mathcal{B}$, $X(t)$ has the conditional distribution

$$\mathbb{P}\left\{X\left(t\right) \in B \mid X\left(0\right) = x\right\} = \frac{1}{\sqrt{2\pi t}} \int_{B} e^{-\frac{(y-x)^2}{2t}} dy$$

Therefore, I can define the transition probability function for each time $t \in \mathbb{T}$ as

$$\begin{cases} p_t\left(x, B\right) & = \frac{1}{\sqrt{2\pi t}} \int_B e^{-\frac{(y-x)^2}{2t}} dy \\ p_0\left(x, B\right) & = \delta_x\left(B\right). \end{cases} \tag{2.4.7}$$

Hence, for f sufficiently smooth,

$$\mathbb{E}^{x}\left[f\left(X\left(t\right)\right)\right] = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} f\left(y\right) e^{-\frac{(y-x)^2}{2t}} dy \quad \left(0 < t \in \mathbb{T}\right). \tag{2.4.8}$$

In particular,

$$\mathbb{E}^{x}\left[X\left(t\right)\right] = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} y e^{-\frac{(y-x)^2}{2t}} dy \quad \left(0 < t \in \mathbb{T}\right).$$

□

REMARK. The function, $\Phi(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$ for $0 < t < \infty$, is known as the ***Gaussian kernel*** or ***heat kernel*** and is the *fundamental solution* of the heat equation (*cf.* Section 4.1). As such, (2.4.8) can be written is terms of convolution, or, for $0 < t \in \mathbb{T}$

$$\begin{aligned} \mathbb{E}^{x}\left[f\left(X\left(t\right)\right)\right] &= \Phi\left(x,t\right) * f\left(x\right) \\ &= \int_{-\infty}^{\infty} f\left(y\right) \Phi\left(y-x,t\right) dy \\ &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} f\left(y\right) e^{-\frac{(y-x)^2}{2t}} dy. \end{aligned}$$

□

In the case of real-valued processes, it is natural to define the *transition distribution function* as follows.

DEFINITION. Let $X = (X(t))_{t\in\mathbb{T}}$ be a real-valued Markov process. Then the function p_t on $\mathbb{R}\times\mathcal{B}$ for all $t \in \mathbb{T}$ defined as

$$p_t(x, y) = \mathbb{P}(X(t) \leq y \mid X(0) = x) \quad (\text{for } x, y \in \mathbb{R})$$

is called a ***transition distribution function***.

In the case that the distribution admits a density as above, we can define probability of transitioning from one state to another in terms these densities as follows.

DEFINITION. Assume that $X = (X(t))_{t\in\mathbb{T}}$ is a real-valued Markov process with transition probability functions $\{p_t\}$. Then, if it exists, the function ρ_t defined for each $t \in \mathbb{T}$ as

$$\rho_t(x, y) = \frac{d}{dy}p_t(x, y) \quad (x, y \in \mathbb{R})$$

where

$$p_t(x, y) = \mathbb{P}(X(t) \leq y \mid X(0) = x)$$

is called the ***transition density functions of*** X and is denoted as $\{\rho_t\}_{t\in\mathbb{T}}$.

REMARK. It is easy to show the transition distributions and transition densities have the semigroup properties using the previous results for transition probability functions. □

EXAMPLE. Let $X = (X(t))_{t\in\mathbb{T}}$ be a Gaussian process with mean x and variance t. Then for any $0 < t \in \mathbb{T}$, it has transition distributions

$$p_t(x, y) = \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} e^{-\frac{(y-x)^2}{2t}}dy \tag{2.4.9}$$

and transition densities

$$\rho_t(x, y) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{(y-x)^2}{2t}}.$$

□

REMARK. While transition probability functions completely determine a Markov process (*cf.* Theorem 2.29), as we shall see in the sequel, in most cases, explicit formulas for transition probability functions8 to do not exist since they are solutions to general PDE. As such, they are generally not a useful computational tool. □

2.5. Brownian Motion

Brownian motion is the "nicest" stochastic process in many ways and it is the bedrock of modern continuous-time finance. On the other hand, while continuous, it is extremely erratic. In fact, Brownian motion is so erratic that it is nowhere differentiable almost surely (*cf.* Theorem 2.40). I will only present a somewhat superficial treatment of a very rich subject here. See Karatzas and Shreve [**51**] and Mörters and Peres [**61**] for a much more complete discussion of Brownian motion.

For this section, assume that $(\Omega, \mathcal{F}, F, \mathbb{P})$ is some filtered probability space with usual conditions and that all stochastic processes on $(\Omega, \mathcal{F}, F, \mathbb{P})$ are admissible. Further, assume that $\mathbb{T}$ is some bounded time index.

Motivation (Symmetric random walk as a proxy for Brownian motion)**.** So far, we have seen that symmetric random walks have a number of attractive properties that are well suited to pricing financial instruments in a discrete setting. In particular, for a symmetric random walk $S = (S_k)_{k\in\mathbb{T}}$ generated by a family of i.i.d. random variables $\{X_k\}_{k\in\mathbb{T}}$ where $X_k \sim N(0,1)$ for all $k \in \mathbb{T}$, we know from the Discrete-Time Martingales motivation starting on page 79

(1) $S_k \sim N(0,k)$ for all $k \in \mathbb{T}$;
(2) S is a martingale; and
(3) S is Markov.

Additionally, the increments of a symmetric random walk defined as

$$S_k - S_l \text{ for any } k, l \in \mathbb{T} \text{ with } l \leq k$$

possess a number of important properties as well. For instance, we also know the increments of a random walk are random variables with mean zero since, for any $k, l \in \mathbb{T}$ with $l \leq k$,

$$\mathbb{E}[S_k - S_l] = \mathbb{E}\left[\sum_{j=l}^{k} X_j\right] = \sum_{j=l}^{k} \mathbb{E}[X_j] = 0.$$

Additionally, they have variance $k - l$ since

$$\text{Var}[S_k - S_l] = \text{Var}\left[\sum_{j=l}^{k} X_j\right] = \sum_{j=l}^{k} \text{Var}[X_j] = k - l.$$

Hence, the increments are normally distributed with mean zero and variance $k - l$ (*i.e.* $S_k - S_l \sim N(0, k-l)$). Also, the increments are independent since, for any $m, r \in \mathbb{T}$ with $r \leq m$,

$$\begin{aligned}
\mathbb{P}(S_k - S_l \,|S_m - S_r) &= \mathbb{P}\left(\sum_{j=l}^{k} X_j \mid \sum_{j=r}^{m} X_j\right) \\
&= \sum_{j=l}^{k} \mathbb{P}\left(X_j \mid \sum_{j=r}^{m} X_j\right) \\
&= \sum_{j=l}^{k} \mathbb{P}(X_j)\, \mathbb{P}\left(\sum_{j=r}^{m} X_j\right) \quad (X_i \text{ are i.i.d.}) \\
&= \mathbb{P}\left(\sum_{j=l}^{k} X_j\right) \mathbb{P}\left(\sum_{j=r}^{m} X_j\right) \\
&= \mathbb{P}(S_k - S_l)\, \mathbb{P}(S_m - S_r).
\end{aligned}$$

Lastly, it is clear that $\mathbb{E}[S_{k-l}] = \mathbb{E}[S_k - S_l]$ and $\text{Var}[S_{k-l}] = \text{Var}[S_k - S_l]$. Therefore, the increments are also stationary since the distribution of normal variable is completely characterized by its first two moments.

Given this, it seems reasonable to ask if a continuous-time process with these properties exists. I will begin with a formal definition of such a process. □

DEFINITION (One dimensional Brownian motion). The F-adapted, real-valued stochastic process $W = (W(t))_{t\in\mathbb{T}}$ is called a (***one-dimensional***) ***Brownian motion*** (***centered at*** $x_0 \in \mathbb{R}$) (*a.k.a.* a ***Wiener process***), if

(1) $W(0) = x_0$;
(2) The map $t \mapsto W(t)$ is continuous a.s. (*continuous sample paths*);
(3) For all times $0 = t_0 < t_1 < t_2 \cdots < t_k$ such that $t_k \in \mathbb{T}$ for all k, the increments $W(t_0), W(t_1) - W(t_0), \ldots, W(t_k) - W(t_{k-1})$ are independent (*independent increments*); and
(4) For each $s, t \in \mathbb{T}$ such that $0 \leq s \leq t$, the increment $W(t) - W(s)$ is a normal random variable with $W(t) - W(s) \sim N(0, t-s)$ (*Gaussian stationary increments*).

If $x_0 = 0$, the process is said to ***standard Brownian motion***.

Brownian motion can be extended to n-dimensions in the natural way.

DEFINITION. The m-dimensional, real-valued stochastic process

$$W = \left(W^1(t), \ldots, W^m(t)\right)_{t\in\mathbb{T}}$$

is called a (***m-dimensional***) ***Brownian motion*** (***centered at*** $x_0 \in \mathbb{R}^n$) if W^k is a one-dimensional Brownian motion started at $x_0 \cdot \mathbf{e}_k$ where $\mathbf{e}_k$ is the k^{th} unit vector.

NOTATION. A one-dimensional Brownian motion is often said to be ***linear***. A two-dimensional Brownian motion is often said to be ***planar***. □

Using the above definition, it is easy to determine first two moments of Brownian motion.

THEOREM 2.31. *Let $W = (W(t))_{t\in\mathbb{T}}$ be standard Brownian motion. Then $W(t)$ is a Gaussian random variable for all $t \in \mathbb{T}$ such that*

$$W(t) \sim N(0, t).$$

PROOF. By definition, for any $s, t \in \mathbb{T}$ with $s \leq t$,

$$W(t) - W(s) \sim N(0, t-s).$$

In particular, this true for $s = 0$. Since $W(0) = 0$,

$$W(t) \sim N(0, t).$$

□

COROLLARY 2.32. *Let $W = (W(t))_{t\in\mathbb{T}}$ be standard Brownian motion. Then for any $s, t\in \mathbb{T}$,*

$$\mathrm{Cov}\left[W(t), W(s)\right] = t \wedge s.$$

PROOF. For any $s, t\in \mathbb{T}$ such that $s \leq t$,

$$\begin{aligned}
\mathrm{Cov}\left[W(t), W(s)\right] &= \mathbb{E}\left[\left(W(t) - \mathbb{E}\left[W(t)\right]\right)\left(W(s) - \mathbb{E}\left[W(s)\right]\right)\right] \\
&= \mathbb{E}\left[W(t) W(s)\right] \quad (W(\cdot) \sim N(0,1)) \\
&= \mathbb{E}\left[W(s)\left(W(t) - W(s)\right) + W^2(s)\right] \\
&= \mathbb{E}\left[W(s)\left(W(t) - W(s)\right)\right] + \mathbb{E}\left[W^2(s)\right] \\
&= \mathbb{E}\left[W^2(s)\right] \quad \text{(independent increments)} \\
&= \mathrm{Var}\left[W(s)\right] = s.
\end{aligned}$$

□

As will be shown shortly, the first two moments completely characterize standard Brownian motion since it is a Gaussian process (*cf.* Corollary 2.34).

Motivation. The fact that the increments of Brownian motion are Gaussian provides a clue regarding an important property of its "derivative". Assume $s, t \in \mathbb{T}$ with $s \leq t$. Then

$$\begin{aligned}
\mathbb{E}\left[\left(W(t) - W(s)\right)^2\right] &= \mathbb{E}\left[W^2(t) - 2W(t) W(s) + W^2(s)\right] \\
&= \mathbb{E}\left[W^2(t)\right] - 2\mathbb{E}\left[W(t) W(s)\right] + \mathbb{E}\left[W^2(s)\right] \\
&= t - 2s + s = t - s.
\end{aligned}$$

Now, it can be shown that if X is Gaussian, $\mathbb{E}\left[X^4\right] = 3\mathrm{Var}\left[X\right]^2$ by direct calculation using characteristic functions (*cf.* (1.9.1) and Theorem 1.70). Hence,

$$\begin{aligned}
\mathrm{Var}\left[\left(W(t) - W(s)\right)^2\right] &= \mathbb{E}\left[\left(W(t) - W(s)\right)^4\right] - \mathbb{E}\left[\left(W(t) - W(s)\right)^2\right]^2 \\
&= 3\mathrm{Var}\left[\left(W(t) - W(s)\right)\right]^2 - (t-s)^2 \\
&= 2(t-s)^2
\end{aligned}$$

In other words, as the increments get smaller, their variance is very small. This suggests as $t - s \to 0$

$$\left(W(t) - W(s)\right)^2 = \mathbb{E}\left[\left(W(t) - W(s)\right)^2\right] = t - s$$

or

$$dW(t)\, dW(t) = dt.$$

While intuitive, this argument is, in fact, incorrect. It will be shown that Brownian motion is simply too erratic for such a result to hold. However, it does provide some motivation as to an important result proved below (*cf.* Theorem 2.40). □

2.5.1. Existence. While such a process clearly has a number of very attractive properties, proving the existence of Brownian motion is non-trivial. In fact, there are several ways that give deep insight into the nature of Brownian motion and its generalization (for a good discussion of one such method due to Lévy, see Mörters and Peres [**61**, ch 1.1]. Given the importance of random walks in the numerical calculation of price processes, I have chosen an existence proof in which Brownian motion is seen to be the limiting process of symmetric random walks. The proof generally follows Shreve[**76**, Ch. 3.2] and Cox [**17**, Ch. 5.2].

Motivation. Let be $S = (S_k)_{k\in\mathbb{T}}$ a symmetric random walk driven by with the coin flipping process $X = (X_k)_{k\in\mathbb{T}}$ and assume that I will flip the coin over some finite, discrete time horizon, $\mathbb{T} = \{0, 1, \ldots, T\}$. Further, assume the time step is constant (say equal to one) and the wealth step size is also constant and equal to ± 1 (*i.e.* $X_k = \pm 1$ for all $k \in \mathbb{T}$).

Now, I can construct a new random walk $\hat{S} = \left(\hat{S}_k\right)_{k\in\mathbb{T}}$ by rescaling the time step (*i.e.* flipping the coin more or less often in the same time span). If I want to ensure that the new random walk has the same variance as the original, I will also have to rescale the wealth step size to account for the change in the time step. Specifically, for some fixed $m \in \mathbb{N}$,

$$\hat{S}_k = {}^1/_{\sqrt{m}} S_{mk}.$$

To see this, recall that since the $X_k \sim N(0, 1)$ and i.i.d., then

$$\text{Var}(S_k) \quad = \quad \mathbb{E}\left[X_k^2\right] = \sum_{i=1}^{k} X_1^2 = kX_1^2$$

(*cf.* 2.3).

However, in the case of scaled random walk, I have to toss the coin m more often to reach time k. Hence, its variance is given by

$$\text{Var}\left(\hat{S}_{mk}\right) = \mathbb{E}\left[\hat{X}_k^2\right] = \sum_{i=1}^{mk} \hat{X}_1^2 = mk\hat{X}_1^2.$$

Therefore, in order to have the same variance

$$mk\hat{X}_1^2 = kX_1^2 \;\Rightarrow \hat{X}_1 = \frac{1}{\sqrt{m}} X_1.$$

Consequently, if the coin is tossed m times as often, in order to have the same variance, we will have to reduce the wealth step size by factor of ${}^1/_{\sqrt{m}}$. In other words, for scaled random walk $\hat{S} = \left(\hat{S}_k\right)_{k\in\mathbb{T}}$ with the time step ${}^1/_m$, then the rescaled random variables must be of the form, $\hat{X}_k \sim N\left(0, \frac{1}{\sqrt{m}}\right)$.

However, such a process is not a good candidate for Brownian motion since it is still discrete and Brownian motion is continuous (albeit still highly erratic). Therefore, I can use linear interpolation between the times at which I flip the coin to create piece-wise continuous increments. Letting

$W^{(m)} = \left(W_k^{(m)}\right)_{k\in\mathbb{T}}$ be such a rescaled random walk with continuous increments, the hope is that $W^{(m)} \to W$ as $m \to \infty$. This is formalized as follows. □

DEFINITION. Let be $S = (S_k)_{k\in\mathbb{T}}$ a symmetric random walk driven by the process $X = (X_k)_{k\in\mathbb{T}}$ where $X_k \sim N(0,1)$ for all $k \in \mathbb{T}$. Then for any $t \in [0,T]$ and some positive integer $k \in \mathbb{T}$, define the process $W^m = (W^m(t))_{t\in\mathbb{T}}$

$$W^m(t) = \begin{cases} \dfrac{1}{\sqrt{m}} S_{mt}, & (mt = [mt]) \\ S_{[mt]}\left(1 + ([mt] - mt)\right) + S_{[mt]+1}\left([mt] - mt\right), & (mt \neq [mt]) \end{cases}$$

where "$[\,\cdot\,]$" is the "largest integer" operator.

Such a process is said to be a ***scaled symmetric random walk***.

To see that a such a process is a good approximation for Brownian motion, note that it is continuous by construction. Additionally, at each time that $mt = [mt]$, the process is a symmetric random walk so it is a martingale and Markov. Moreover, its is Gaussian with stationary, independent increments (*cf.* 2.5).

To see this, fix $m \in \mathbb{N}$ and let $t \in [0,T]$ such that $mt = [mt]$. Then

$$\mathbb{E}\left[W^{(m)}(t)\right] = \frac{1}{\sqrt{m}}\mathbb{E}(S_{mt}) = 0$$

and

$$\mathrm{Var}\left(W^{(m)}(t)\right) = \mathrm{Var}\left(\frac{1}{\sqrt{m}}S_{mt}\right) = \frac{1}{m}\mathrm{Var}(S_{mt}) = t.$$

As for increments, fix $m \in \mathbb{N}$ and let $\{0 = t_0 < t_1 < \cdots < t_k = T\}$ be a partition such that $mt_j = [mt_j]$ for $0 \le j \le k$. Then for any $0 \le j \le i \le k$,

$$\begin{aligned} &\mathbb{P}\left(W^{(m)}(t_{i+1}) - W^{(m)}(t_i) \mid W^{(m)}(t_{j+1}) - W^{(m)}(t_j)\right) \\ &= \frac{1}{m}\mathbb{P}\left(S_{mt_{i+1}} - S_{mt_i} \mid S_{mt_{j+1}} - S_{mt_j}\right) \\ &= \frac{1}{m}\mathbb{P}\left(S_{mt_{i+1}} - S_{mt_i}\right)\mathbb{P}\left(S_{mt_{j+1}} - S_{mt_j}\right) \\ &= \mathbb{P}\left(W^{(m)}(t_{i+1}) - W^{(m)}(t_i)\right)\mathbb{P}\left(W^{(m)}(t_{j+1}) - W^{(m)}(t_j)\right). \end{aligned}$$

Hence, the increments are independent. Additionally, they are stationary since, for any $0 \le s \le t \le T$ such $mt = [mt]$ and $ms = [ms]$

$$\mathbb{E}\left(W^{(m)}(t) - W^{(m)}(s)\right) = \frac{1}{\sqrt{m}}\mathbb{E}(S_{mt} - S_{ms}) = 0.$$

and

$$\begin{aligned}\operatorname{Var}\left(W^{(m)}(t)-W^{(m)}(s)\right) &= \operatorname{Var}\left(\frac{1}{\sqrt{m}}\left(S_{mt}-S_{ms}\right)\right)\\ &= \frac{1}{m}\operatorname{Var}\left(\sum_{k=ms}^{mt}X_k\right)\\ &= \frac{1}{m}\sum_{k=ms}^{mt}\operatorname{Var}\left(X_1\right)\\ &= t-s.\end{aligned}$$

In fact, scaled symmetric random walk has all the properties of Brownian except that it is not Gaussian for any finite m. While it is Gaussian at each time $mt=[mt]$, due to the piece-wise linear increments, scaled symmetric random walks are still too smooth compared to Brownian motion. Fortunately, it is possible to show that the limiting process of a scaled random walk is, in fact, Brownian motion. Here, I will rely on the following results which are proved in Durrett (*cf.* [**23**, pg. 355]).

LEMMA 2.33. *Let $W=(W(t))_{t\in\mathbb{T}}$ be some stochastic process such that for any $\omega\in\Omega$, the mapping $t\mapsto W(\omega,t)$ is continuous for $t\in\mathbb{T}$. Then for any partition $\{0=t_0<t_1\cdots<t_m=T\}$, the following statements are equivalent.*

(1) *The increments $W(t_{k+1})-W(t_k)\sim N(0,t_{k+1}-t_k)$ and are independent for $0\leq k<m$;*
(2) *The random variables $W(t_0),\ldots,W(t_m)$ are jointly normally distributed with mean zero and covariance matrix Σ where*

$$\Sigma^{i,j}=t_i\wedge t_j.$$

The next result follows immediately.

LEMMA 2.34. *A continuous Gaussian process with the covariance matrix $\Sigma^{ij}=\Sigma^{ij}=t_i\wedge t_j$ is Brownian motion.*

Therefore, in order to prove that scaled random walk converges to Brownian motion, I only to show that limiting process is Gaussian. For this, I can turn to the following well-known result which is proved in Shreve [**76**, pg. 89-91].

THEOREM 2.35 (Central limit). *Let $W^{(m)}=\left(W_t^{(m)}\right)_{t\in\mathbb{T}}$ be a scaled random walk centered at zero. Then for any fixed $t\in\mathbb{T}$*

$$\lim_{m\to\infty}W_t^{(m)}\sim N(0,t).$$

Using the above the results, I have shown that scaled random walk, indeed, converges to Brownian motion.

THEOREM 2.36. *Let* $W^{(m)} = \left(W^{(m)}(t)\right)_{t\in\mathbb{T}}$ *be scaled symmetric random walk. Then the process* $W = (W(t))_{t\in\mathbb{T}}$ *defined as*

$$W(t) := \lim_{m\to\infty} W^{(m)}(t) \quad (\textit{for each } t \in \mathbb{T})$$

is standard Brownian motion.

2.5.2. Simple Invariance Properties of Brownian Motion. The following results are analogous to properties of scaled symmetric random walk derived above (*cf.* definition on page 99).

THEOREM 2.37 (Scaling invariance). *Let* $W = (W(t))_{t\in\mathbb{T}}$ *be standard Brownian motion. Then for any reasonable* $\alpha > 0$*, the process* $B = (B(t))_{t\in\mathbb{T}}$ *defined as*

$$B(t) = \frac{1}{\sqrt{\alpha}} B(\alpha t)$$

is also a standard Brownian motion.

The following result is another form of Brownian motion scaling which follows directly from the above result.

COROLLARY 2.38 (Time inversion). *Let* $W = (W(t))_{t\in\mathbb{T}}$ *be standard Brownian motion. Then the process* $B = (B(t))_{t\in\mathbb{T}}$ *defined as*

$$B(t) = \begin{cases} tW\left(\frac{1}{t}\right), & t > 0 \\ 0, & t = 0 \end{cases}$$

is also standard Brownian motion.

2.5.3. Sample Path Properties of Brownian Motion. The next results give some insight into the regularity of Brownian motion which are proved in Mörters and Peres [**61**, pg. 18 and 21] respectively.

THEOREM 2.39. *Let* $W = (W(t))_{t\in\mathbb{T}}$ *be standard Brownian motion. Then* W *is (locally) Hölder continuous for any modulus of continuity* $0 < \gamma < 1/2$*, or, for some constant* C*,*

$$|W(t) - W(s)| \leq C|t - s|^{\gamma} \quad (\textit{for } s, t \in \mathbb{T}).$$

THEOREM 2.40 (Paley, Wiener, and Zygmund theorem). *Let* $W = (W(t))_{t\in\mathbb{T}}$ *be standard Brownian motion. Then* W *is nowhere differentiable almost surely.*

2.5.4. Quadratic Variation. Given the two results above, Brownian motion, although continuous, is extremely erratic indeed. However, my hope is to develop a calculus for price processes driven by Brownian motion and it would appear the standard notions of integration and differentiation do not apply given the extreme volatility of Brownian motion.

To better understand the problem, it is worth looking into the regularity of Brownian motion a little more deeply. I will begin by examining Brownian motion in the context of its *variation*.

Motivation. Consider a scaled symmetric random walk as a proxy for Brownian motion since it has many of the same properties. A natural question is: *how regular is the typical path of such a process*? One way to measure this is to look at the length of a path over some fixed time: the longer the path over a given time span, the more erratic the process. Since each increment of the path is piecewise linear, it seems reasonable to measure volatility in terms of the sum the Euclidean length of the increments of the path. Such a measure of volatility is known as the *variation*.

Remark. Variance is another a measure of the volatility of a stochastic process. However, unlike variation which depends on the entire path of the process up to some time t, variance depends only the value of the process at time t and is not suited for our purpose. □

In order to define variation, assume that f is some differentiable function on a bounded interval $[0,T]$ with some partition $(0 = t_0 < t_1 < \cdots < t_k = T)$. Then by the mean value theorem (*cf.* [**78**, pg. 193]), for all $0 \leq j \leq k-1$, there exist some point $t_j \leq \tau_j \leq t_{j+1}$ such that

$$f'(\tau_j) = \frac{f(t_{j+1}) - f(t_j)}{t_{j+1} - t_j} < \infty$$

since f is differentiable on $[0,T]$.

Now, $t_{j+1} > t_j$ for all $0 \leq j \leq k-1$. Hence,

$$|f'(\tau_j)| = \frac{|f(t_{j+1}) - f(t_j)|}{t_{j+1} - t_j} < \infty. \tag{2.5.1}$$

Then after rearranging terms

$$\sum_{j=0}^{k-1} |f(t_{j+1}) - f(t_j)| = \sum_{j=0}^{k-1} |f'(\tau_j)| (t_{j+1} - t_j). \tag{2.5.2}$$

Taking the limit as the mesh size goes to zero, the right-hand side is a bounded Riemann-Stieltjes integral (*cf.* [**66**, Ch. 3.4]). Hence,

$$\lim_{k\to\infty} \sum_{j=0}^{k-1} |f(t_{j+1}) - f(t_j)| \quad = \quad \int_0^T |f'(t)|\, dt < \infty.$$

Upon inspection, the left-side is a measure of the length of the graph of f which is captured in the following definition.

Definition. Assume that f is a real-valued function on some bounded interval $[0,T]$. Then for some partition $\Pi = (0 = t_0 < t_1 \cdots < t_k = T)$, the (***first***) ***variation of*** f ***on*** $[0,T]$ ***relative to*** Π, denoted as $\mathrm{V}_T(f,\Pi)$, is defined as

$$\mathrm{V}_T(f,\Pi) = \sum_{j=0}^{k-1} |f(t_{j+1}) - f(t_j)|.$$

Now, let $\{\Pi\}$ be the collection of all finite partitions of $[0,T]$. Then f is said to have ***bounded variation*** if

$$\mathrm{V}_T(f) := \sup_{\{\Pi\}} \mathrm{V}_T(f,\Pi) < \infty.$$

$\mathrm{V}_T(f)$ is called (***first***) ***variation of*** f ***on*** $[0,T]$ (*a.k.a.* ***total variation***) and the collection of all such functions is denoted $\mathrm{BV}(0,T)$ (*i.e.* $\mathrm{BV}(0,T) := \{f : [0,T] \to \mathbb{R} \mid \mathrm{V}_T(f) < \infty\}$.)

The following result, which is proved in Pascucci (*cf.* [**66**, pg. 129]), provides a convenient way to compute the first variation of a function.

THEOREM 2.41. *Assume that* $f \in \mathrm{BV}(0,T) \cap C(0,T)$ *for some bounded interval* $[0,T]$ *with the partition* $\Pi = (0 = t_0 < t_1 \cdots < t_k = T)$. *Then*

$$\mathrm{V}_T(f) = \lim_{|\Pi|\to 0} \mathrm{V}_T(f,\Pi).$$

While the first variation may be good measure of the volatility of a differentiable function, it is not suitable for Brownian motion since it nowhere differentiable by the Paley, Wiener, Zygmund theorem (*cf.* Theorem 2.40). This gives rise to the second measure of volatility, *quadratic variation*.

DEFINITION. Assume that f is a real-valued function on some bounded interval $[0,T]$. Then for some partition $\Pi = (0 = t_0 < t_1 \cdots < t_k = T)$, the ***quadratic*** (***second***) ***variation of*** f ***on*** $[0,T]$ ***relative to*** Π, denoted as $\mathrm{V}_T^2(f,\Pi)$, is defined as

$$\mathrm{V}_T^2(f,\Pi) = \sum_{j=0}^{k-1} |f(t_{j+1}) - f(t_j)|^2.$$

Now, let $\{\Pi\}$ be the collection of all finite partitions of $[0,T]$. Then the ***quadratic variation of*** f ***on*** $[0,T]$, denoted as $\mathrm{V}_T^2(f)$, is defined as

$$\mathrm{V}_T^2(f) = \sup_{\{\Pi\}} \mathrm{V}_T^2(f,\Pi).$$

We can use this definitions of these definitions of volatility to characterize continuous-time deterministic functions. First, I need the following lemma.

LEMMA 2.42. *Assume that* $f \in \mathrm{BV}(0,T) \cap C(0,T)$. *Then* f *is uniformly continuous on* $[0,T]$.

PROOF. Assume that $f \in \mathrm{BV}(0,T) \cap C(0,T)$ and choose j such that $0 \leq j \leq k-1$. Then for any $\varepsilon > 0$, there exists some δ_j such that for $t_{j+1} - t_j < \delta_j$,

$$|f(t_{j+1}) - f(t_j)| < \epsilon.$$

Taking $\delta = |\Pi| = \min_{0\leq j\leq k-1} \delta_j$, then f is seen to be uniformly continuous. □

Using this lemma, it is possible to show that continuous functions with bounded variation have zero quadratic variation.

THEOREM 2.43. *Assume that* $f \in \mathrm{BV}(0,T) \cap C(0,T)$. *Then* $\mathrm{V}_T^2(f) = 0$.

PROOF. Assume that $f \in \mathrm{BV}(0,T) \cap C(0,T)$ on some bounded interval $[0,T]$ with the partition $\Pi = (0 = t_0 < t_1 < \cdots < t_k = T)$. By the above lemma, f is uniformly continuous and for any finite $\varepsilon > 0$, there exists $\delta > 0$ such that, for $0 \leq j \leq k-1$

$$|f(t_{j+1}) - f(t_j)| < \varepsilon$$

Therefore,

$$\begin{aligned} \sum_{j=0}^{k-1} |f(t_{j+1}) - f(t_j)|^2 &\leq \varepsilon \sum_{j=0}^{k-1} |f(t_{j+1}) - f(t_j)| \\ &\leq \varepsilon \mathrm{V}_T(f) < \infty \end{aligned}$$

since f has bounded variation. Therefore,

$$\lim_{\varepsilon \to 0} \sum_{j=0}^{k-1} |f(t_{j+1}) - f(t_j)|^2 = 0.$$

Since this true for any partition, $\mathrm{V}_T^2(f) = 0$. □

The concept of quadratic variation can be extended to stochastic processes as follows.

DEFINITION. Assume that $X = (X(t))_{t \in [0,T]}$ is an admissible process on some filtered probability space $(\Omega, \mathcal{F}, F, \mathbb{P})$ with the usual conditions. Then for some partition $(0 = t_0 < t_1 < \cdots < t_k = t)$ where $t \in [0,T]$, the ***quadratic variation of*** X ***on*** $[0,t]$, denoted as$[X](t)$, is defined as

$$[X](t) := \lim_{k \to \infty} \sum_{j=0}^{k-1} |X(\omega, t_{j+1}) - X(\omega, t_j)|^2$$

where $\omega \in \Omega \times [0,t]$ is some fixed path.

REMARK. Although the notation ignores the dependency of quadratic variation on the path, the quadratic variation is *path-wise* property which measures the volatility along a specific path. □

Motivation. Given the above discussion, a scaled random walk should be in $\mathrm{BV}(0,T) \cap (0,T)$ if its first variation is finite and its second variation is zero. Thus, for some fixed $k < \infty$, select a partition $\{0 = t_0 < t_1 < \cdots < t_k = T\}$ such that $mt_j = [mt_j]$ for all $j = 0, \ldots, k$. Then

$$\begin{aligned} \sum_{j=1}^{k} \left| \left(W^{(m)}(t_j) - W^{(m)}(t_{j-1}) \right) \right| &= \frac{1}{\sqrt{m}} \sum_{j=1}^{k} \left| \left(S_{mt_j} - S_{mt_{j-1}} \right) \right| \\ &= \frac{1}{\sqrt{m}} \sum_{j=mt_1}^{mt_k} |(S_j - S_{j-1})| . \end{aligned}$$

Since $S_j - S_{j-1} = X_j = \pm 1$ for all j,

$$\begin{aligned}\sum_{j=1}^{k} \left|\left(W^{(m)}(t_j) - W^{(m)}(t_{j-1})\right)\right| &= \frac{1}{\sqrt{m}} \sum_{j=mt_1}^{mt_k} |X_j| \\ &= \sqrt{m}t_k = \sqrt{m}t.\end{aligned}$$

Therefore, $\mathrm{V}_t\left(W^{(m)}\right) \to \infty$ as $m \to \infty$.

As for its second variation, by definition,

$$\begin{aligned}\sum_{j=1}^{k} \left|\left(W^{(m)}(t_j) - W^{(m)}(t_{j-1})\right)\right|^2 &= \frac{1}{m} \sum_{j=1}^{k} \left|S_{mt_J} - S_{mt_{J-1}}\right|^2 \\ &= \frac{1}{m} \sum_{j=mt_1}^{mt_k} |S_j - S_{j-1}|^2 .\end{aligned}$$

Again, $S_j - S_{j-1} = X_j = \pm 1$ for all j,

$$\sum_{j=1}^{k} \left|\left(W^{(m)}(t_j) - W^{(m)}(t_{j-1})\right)\right|^2 = \frac{1}{m} \sum_{j=mt_1}^{mt_k} |X_j|^2 = t_k = t.$$

Therefore, $\left[W^{(m)}\right](t) = t$ for all m.

□

The following result confirms that, in fact, the quadratic variation of Brownian motion is non-zero almost surely. It is proved in Shreve [**?**, pgs. 102-103].

THEOREM 2.44. *Let $W = (W(t))_{t\in[0,T]}$ be standard Brownian motion. Then $[W](t) = t$ for all $t \in [0,T]$ a.s.*

COROLLARY 2.45. *Let $W = (W(t))_{t\in[0,T]}$ be standard Brownian motion. Then W on any interval $[0,T]$ has infinite total variation a.s.*

PROOF. Assume that$\{0 = t_0 < t_1 < \cdots < t_k = t\}$ is a partition on $[0,t]$ where $t \in [0,T]$ and that $W = (W(t))_{t\in[0,T]}$ is standard Brownian motion. Now, assume that $\mathrm{V}_t(W) < \infty$. Since $[W](t) = t$, then for any path a.s.

$$\begin{aligned}\sum_{j=0}^{k-1} (W(\omega, t_{j+1}) - W(\omega, t_j))^2 &\leq \sup_{0\leq j\leq k-1} |W(\omega, t_{j+1}) - W(\omega, t_j)| \sum_{j=0}^{k-1} |W(\omega, t_{j+1}) - W(\omega, t_j)| \\ &\leq \sup_{0\leq j\leq k-1} |W(\omega, t_{j+1}) - W(\omega, t_j)| \, \mathrm{V}_t(W) .\end{aligned}$$

Now,

$$\lim_{k\to\infty}\sum_{j=0}^{k-1}\left(W\left(\omega,t_{k+1}\right)-W\left(\omega,t_k\right)\right)^2=[W]\left(t\right)=T.$$

On the other hand, since W is continuous,

$$\lim_{k\to\infty}\left|W\left(\omega,t_{j+1}\right)-W\left(\omega,t_j\right)\right|=0\quad(0\leq j\leq k-1)$$

which is a contraction and $\mathrm{V}_t\left(W\right)<\infty$. □

Above, I showed that the variance of standard Brownian motion is also equal to t (*cf.* Theorem 2.31). However, quadratic variation is a path-wise property while variance is a statistic averaged over all paths up to time t. Hence, they capture different properties of Brownian motion. The following corollary is trivial, but helpful since variance is often easier to compute than the quadratic variation.

COROLLARY 2.46. *Let $W=\left(W\left(t\right)\right)_{t\in\mathbb{T}}$ be standard Brownian motion. Then*

$$\mathrm{Var}\left[W\left(t\right)\right]=[W]\left(t\right). \tag{2.5.3}$$

This suggests the following definition.

DEFINITION. Assume that $X=\left(X\left(t\right)\right)_{t\in[0,T]}$ and and $Y=\left(Y\left(t\right)\right)_{t\in\mathbb{T}}$ are two real-valued stochastic processes on some filtered probability space $(\Omega,\mathcal{F},F,\mathbb{P})$ with the usual conditions. Then for some partition $(0=t_0<t_1\cdots<t_k=t)$ where $t\in[0,T]$, the ***covariation*** of X ***and*** Y ***on*** $[0,t]$ is defined as

$$[X,Y]\left(t\right)=\lim_{k\to\infty}\sum_{j=0}^{k-1}\left(X\left(\omega,t_{j+1}\right)-X_{t_j^m}\right)\left(Y\left(\omega,t_{j+1}\right)-Y\left(\omega,t_j\right)\right)$$

where $\omega\in\Omega\times[0,t]$ is some fixed path.

NOTATION. The symbol $[\cdot,\cdot]$ indicates that covariation is a bilinear form. □

2.5.5. Brownian Filtration. Earlier, I defined the standard filtration of a stochastic process to ensure that it satisfied the usual conditions (*cf.* on page 62). The following result, which is proved in Pascucci [**66**, pg. 118-120], shows that in the case of Brownian motion, the standard filtration is simply the augmented, natural filtration.

THEOREM 2.47. *Let $W=\left(W\left(t\right)\right)_{t\in\mathbb{T}}$ be a standard Brownian motion with respect to some filtration. Then the filtration $F^W=\left(\mathcal{F}_t^W\right)_{t\in\mathbb{T}}$ is equivalent to its standard filtration if*

$$\mathcal{F}_t^W=\sigma\left(\tilde{\mathcal{F}}_t^W\cup\mathcal{N}\right) \tag{2.5.4}$$

where $\mathcal{N}$ is the collection of negligible sets and $\tilde{\mathcal{F}}_t^W$ is the natural filtration of W satisfying the usual conditions.

DEFINITION. Given a standard Brownian motion $W=\left(W\left(t\right)\right)_{t\in\mathbb{T}}$, the filtration defined in (2.5.4) is called ***Brownian filtration***.

2.5.6. Other Properties of Brownian Motion. Given that fact a symmetric random walk is a martingale and Markov. It makes sense to expect that Brownian motion is as well. This is captured in the two following results.

THEOREM 2.48. *Let $W = (W(t))_{t\in\mathbb{T}}$ be a standard Brownian motion. Then W is a martingale.*

PROOF. Let $W = (W(t))_{t\in\mathbb{T}}$ be standard Brownian motion and let $F = \{\mathcal{F}_t\}$ be its standard filtration. Then recalling the properties of conditional expectation (Theorem 1.59) and the fact that $W(t) - W(s) \sim N(0, t-s)$, for $s, t \in \mathbb{T}$ with $s \leq t$

$$\begin{aligned}
\mathbb{E}[W(t) \mid \mathcal{F}_s] &= \mathbb{E}[W(t) - W(s) + W(s) \mid \mathcal{F}_s] \\
&= \mathbb{E}[W(t) - W(s) \mid \mathcal{F}_s] + \mathbb{E}[W(s) \mid \mathcal{F}_s] \\
&= \mathbb{E}[W(t) - W(s)] + W(s) \\
&= W(s) \qquad \text{(Gaussian increments)}
\end{aligned}$$

□

COROLLARY 2.49. *Let $W = (W(t))_{t\in\mathbb{T}}$ be standard Brownian motion. Then W is Markov.*

PROOF. Let $W = (W(t))_{t\in\mathbb{T}}$ be standard Brownian motion and let $F = \{\mathcal{F}_t\}$ be its standard filtration. For $s, t \in \mathbb{T}$ with $s \leq t$,

$$\mathbb{E}[W(t) \mid \mathcal{F}_s] = W(s)$$

since Brownian motion is a martingale. Additionally, by assumption $\mathcal{F}_s = \sigma(W(r) \mid 0 \leq r \leq s)$ and

$$\mathbb{E}[W(t) \mid W(s)] = W(s), \quad (\text{for } s \leq t).$$

Hence, $\mathbb{E}[W(t) \mid \mathcal{F}_s] = \mathbb{E}[W(t) \mid W(s)]$ and Brownian motion is Markov by Theorem 2.20. □

In fact, the following stronger result holds which is proved in Mörters and Peres [**61**, pgs. 43-44].

2.5.7. Correlated Brownian Motion. Typically, the sources of risk in a financial market are expressed in terms of m-dimensional Brownian motion. In other words, the randomness of the price of any financial instrument trading in the market is described by m independent sources of uncertainty (*i.e.* risk), each driven by an independent Brownian motion. Mathematically, this is captured in the concept of *correlated Brownian Motion*.

DEFINITION. Two standard Brownian motions W and B are said to ***correlated*** with the ***correlation*** (***factor***) ρ such that $|\rho| \leq 1$ if, for all $t \in \mathbb{T}$

$$[W, B](t) = \rho t \quad (\text{for all } t \in \mathbb{T}).$$

If $\rho = 0$, the Brownian motions are said to be ***independent***.

Motivation. Let $W^1 = (W^1(t))_{t\in\mathbb{T}}$ and $W^2 = (W^2(t))_{t\in\mathbb{T}}$ be independent, standard Brownian motions. It is possible to construct a new Brownian motion $W^3 = (W^3(t))_{t\in\mathbb{T}}$ defined as

$$W^3(t) = \rho W^1(t) + \sqrt{(1-\rho^2)}W^2(t) \quad (\text{for } t \in \mathbb{T})$$

such that W^1 and W^3 are correlated Brownian motions with

$$\left[W^1, W^3\right](t) = \rho t.$$

To show that W^3 is a standard Brownian motion, recall that a continuous Gaussian process with covariance matrix Σ of the form

$$\Sigma^{ij} = t_i \wedge t_j \quad (\text{for any } 0 \leq i, j \leq k)$$

is a standard Brownian motion (*cf.* Corollary 2.34). Now, W^3 is continuous and $W^3(0) = 0$ since W^1 and W^2 are standard Brownian motions. Additionally, for any $s \in [0, t]$

$$\begin{aligned}
\mathbb{E}\left[W^3(s)\right] &= \mathbb{E}\left[\rho W^1(s) + \sqrt{(1-\rho^2)}W^2(s)\right] \\
&= \rho\mathbb{E}\left[W^1(s)\right] + \sqrt{(1-\rho^2)}\mathbb{E}\left[W^2(s)\right] \\
&= 0.
\end{aligned}$$

Further,

$$\begin{aligned}
\text{Var}\left[W^3(s)\right] &= \text{Var}\left[\rho W^1(s) + \sqrt{(1-\rho^2)}W^2(s)\right] \\
&= \rho^2\text{Var}\left[W^1(s)\right] + \left(1-\rho^2\right)Var\left[W^2(s)\right] \\
&= \rho^2 s + \left(1-\rho^2\right)s \\
&= s.
\end{aligned}$$

Hence, $W^3(t) \sim N(0, t)$ for all $t \in \mathbb{T}$.

Also, by Lemma 2.33, W^3 has Gaussian, independent increments. Given this, I can compute the covariance of W^3 for any $s, t \in \mathbb{T}$ with $s \leq t$ since

$$\begin{aligned}
\text{Cov}\left[W^3(t), W^3(s)\right] &= \mathbb{E}\left[\left(W^3(t) - \mathbb{E}\left[W^3(s)\right]\right)\left(W^3(t) - \mathbb{E}\left[W^3(s)\right]\right)\right] \\
&= \mathbb{E}\left[W^3(t)\,W^3(s)\right] \\
&= \mathbb{E}\left[W^3(s)\left(W^3(t) - W^3(t)\right) + W^3(s)^2\right] \\
&= \mathbb{E}\left[W^3(s)\left(W^3(t) - W^3(s)\right)\right] + \mathbb{E}\left[W^3(s)^2\right] \\
&= \mathbb{E}\left[W^3(s)\left(W^3(t) - W^3(s)\right)\right] + s \\
&= \mathbb{E}\left[W^3(s)\right]\mathbb{E}\left[\left(W^3(t) - W^3(s)\right)\right] + s \\
&= s = s \wedge t.
\end{aligned}$$

Hence, W^3 is standard Brownian motion.

Moreover, for any $t \in \mathbb{T}$

$$\begin{aligned}
\operatorname{Cov}\left[W^1(t), W^3(t)\right] &= \mathbb{E}\left[W^1(t)\left(\rho W^1(t) + \sqrt{(1-\rho^2)}W^2(t)\right)\right] \\
&= \rho\mathbb{E}\left[W^1(t)^2\right] + \sqrt{(1-\rho^2)}\mathbb{E}\left[W^1(t)W^2(t)\right] \\
&= \rho t + \sqrt{(1-\rho^2)}\mathbb{E}\left[W^1(t)\right]\mathbb{E}\left[W^2(t)\right] \\
&= \rho t.
\end{aligned}$$

This gives rise to the following result which can be extended to finite number of Brownian motions. Importantly, this results shows that collection of m independent Brownian motions form a m-dimensional vector space. □

THEOREM 2.50. *Let* $\left\{W^1 = (W^1(t))_{t\in\mathbb{T}}, \ldots, W^m = (W^m(t))_{t\in\mathbb{T}}\right\}$ *be a collection* m *independent standard Brownian motions on the same filtered probability space. Then the process* $W = (W(t))_{t\in\mathbb{T}}$ *as*

$$W(t) := \sum_{j=1}^{m} \rho_j W^j(t) \quad (\textit{for all } t \in \mathbb{T})$$

where $\sum_{j=1}^{m} \rho^2 = 1$ *is standard Brownian motion.*

PROOF. The proof follows from the above motivation and induction. □

The next result can be used to show that the price of a financial instrument can be uniquely expressed in terms of the m independent sources of risk in the market.

THEOREM 2.51. *Let* $\left\{W^1 = (W^1(t))_{t\in\mathbb{T}}, \ldots, W^m = (W^m(t))_{t\in\mathbb{T}}\right\}$ *be a collection of* m *independent standard Brownian motions on the same filtered probability space. Further, assume that* $W = (W(t))_{t\in\mathbb{T}}$ *is standard Brownian motion. Then there exists a unique* $\rho = (\rho_1, \ldots, \rho_m)^{\mathrm{T}}$ *where* $\sum_{j=1}^{m} \rho^2 = 1$ *such that*

$$W(t) := \sum_{j=1}^{m} \rho_j W^j(t) \quad (\textit{for all } t \in \mathbb{T}).$$

is standard Brownian motion.

THEOREM 2.52. *Let* $W = (W^1(t))_{t\in\mathbb{T}}$ *and* $W^2 = (W^2(t))_{t\in\mathbb{T}}$ *be correlated standard Brownian motions such that*

$$\left[W^1, W^2\right](t) = \rho t \quad (\textit{for all } t \in \mathbb{T})$$

where $|\rho| < 1$. *Then there exists* $W^3 = (W^3(t))_{t\in\mathbb{T}}$ *defined as*

$$W^3(t) = \frac{W^2(t) - \rho W^1(t)}{\sqrt{(1-\rho^2)}} \quad (\textit{for } t \in \mathbb{T}).$$

such that W^1 *and* W^3*are independent Brownian motions.*

PROOF. Let $W^1 = (W^1(t))_{t\in\mathbb{T}}$ and $W^2 = (W^2(t))_{t\in\mathbb{T}}$ be correlated, standard Brownian motions such that

$$\left[W^1, W^2\right](t) = \rho t \quad (\text{for all } t \in \mathbb{T})$$

with $|\rho| < 1$ and define $W^3 = (W^3(t))_{t\in\mathbb{T}}$ as

$$W^3(t) = \frac{W^2(t) - \rho W^1(t)}{\sqrt{(1-\rho^2)}} \quad (\text{for } t \in \mathbb{T}).$$

Clearly, W^3 is continuous and $W^3(0) = 0$ since W^1 and W^2 are standard Brownian motions. Additionally, for any $s \in [0,t]$,

$$\begin{aligned}\mathbb{E}\left[W^3(s)\right] &= \frac{1}{\sqrt{(1-\rho^2)}}\mathbb{E}\left[W^2(s) - \rho W^1(s)\right] \\ &= \frac{1}{\sqrt{(1-\rho^2)}}\left(\mathbb{E}\left[W^2(s)\right] - \rho\mathbb{E}\left[W^1(s)\right]\right) = 0.\end{aligned}$$

Further,

$$\begin{aligned}\mathrm{Var}\left[W^3(s)\right] &= \frac{1}{(1-\rho^2)}\mathrm{Var}\left[W^2(s) - \rho W^1(s)\right] \\ &= \frac{1}{(1-\rho^2)}\left(\mathrm{Var}\left[W^2(s)\right] - \rho^2\mathbb{V}\mathrm{ar}\left[W^1(s)\right]\right) \\ &= s\left(\frac{1-\rho^2}{1-\rho^2}\right) = s.\end{aligned}$$

Hence, $W^3(t) \sim N(0,t)$ for all $t \in \mathbb{T}$. Moreover,

$$\begin{aligned}\mathrm{Cov}\left[W^1(t), W^3(t)\right] &= \mathbb{E}\left[W^1(t) W^3(t)\right] \\ &= \frac{1}{\sqrt{(1-\rho^2)}}\mathbb{E}\left[W^1(t)\left(W^2(t) - \rho W^1(t)\right)\right] \\ &= \frac{1}{\sqrt{(1-\rho^2)}}\left(\mathbb{E}\left[W^1(t) W^2(t)\right] - \rho\mathbb{E}\left[W^1(t) W^1(t)\right]\right) \\ &= \frac{1}{\sqrt{(1-\rho^2)}}(\rho t - \rho t) = 0.\end{aligned}$$

Then by Lemma 2.33, W^1 and W^3 are independent Brownian motions. □

CHAPTER 3

Itô Calculus and Stochastic Differential Equations

In the prequel, I suggested that the price a financial instrument can be modeled by a stochastic process whose graph is a path in space and time. Frequently, the paths of deterministic dynamical systems are expressed as solutions to an ordinary differential equation (ODE). This begs the question: *can a stochastic process be expressed as the solution to an ODE is some meaningful sense*? Fortunately, the answer is yes. However, it requires a different type of calculus to make sense of both the integration and differentiation.

The following discussion largely follows Evans [**30**, ch 1.1] which places stochastic differential equations (SDE) in the context of classical dynamical systems. For this chapter, assume that $(\Omega, \mathcal{F}, F, \mathbb{P})$ is some filtered probability space with the usual conditions and that all stochastic processes on $(\Omega, \mathcal{F}, F, \mathbb{P})$ are real-valued and admissible unless otherwise noted. Also, assume that the time index set is a bounded interval $\mathbb{T} = [0, T]$ with $T < \infty$.

3.1. Motivation

Let x_0 be some point in $\mathbb{R}^n$ and assume that $\mathbf{x} : [0, T] \to \mathbb{R}^n$ describes the position of a particle at time $t \in [0, T]$ where $T < \infty$. Defining the *state of the system* to be the particle's position and velocity at any point in time, assume that it is governed by the following *dynamics*

$$\begin{cases} \dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t)) \quad (\text{for } 0 < t \leq T) \\ \mathbf{x}(0) &= x_0 \end{cases} \tag{3.1.1}$$

where "$\dot{}$"$= \dfrac{d}{dt}$ and $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is a given smooth vector field representing the instantaneous velocity of the particle. Then the solution to (3.1.1), if it exists, describes the evolution of the system over time. This is a purely deterministic system whose graph is a path in $\mathbb{R}^n \times [0, \infty)$. Absence any external forces, by this equation, the system will evolve at the (instantaneous) velocity $\mathbf{f}$ for all time.

It seems reasonable to ask if there is a similar expression to (3.1.1) when randomness is added to the system. In other words, if state of the system is stochastic, I should expect the position and velocity to now be $\mathbb{R}^n$-valued stochastic processes, $\mathbf{X} = (\mathbf{X}(t))_{t\in[0,T]}$ and $\dot{\mathbf{X}} = \left(\dot{\mathbf{X}}(t)\right)_{t\in[0,T]}$ respectively, that are governed by the above dynamics perturbed by some random process, or

$$\begin{cases} \dot{\mathbf{X}}(t) &= \mathbf{f}(\mathbf{X}(t)) + \boldsymbol{\xi}(t) \quad (\text{for } 0 < t \leq T) \\ \mathbf{X}(0) &= x_0 \end{cases} \tag{3.1.2}$$

where the last term accounts for the randomness that has been added to the system.

While intuitive, this equation is purely symbology unless the following mathematical questions can be answered:

(1) *What does "$\dot{\mathbf{X}}$" mean in terms of a stochastic process* $\mathbf{X}$ *and under what conditions is it valid*?
(2) *What type of process is* $\boldsymbol{\xi}$? *Does it exist*? *And, what properties does it have*?
(3) *What does it mean for* $\mathbf{X}$ *to be a solution to* (3.1.2) *and is it, in fact, an admissible process*?
(4) *How does the solution of* (3.1.2)*, if it exists, depend on the parameters* x_0 *and* $\mathbf{f}$ *in terms of existence, uniqueness, and regularity*?

To begin to answer these questions, (3.1.2) can be written as (again, symbologically)

$$\frac{d}{dt}\mathbf{X}(t) = \mathbf{f}(\mathbf{X}(t)) + \boldsymbol{\xi}(t).$$

Abusing the notation further, the stochastic version of (3.1.1) can be written as

$$\begin{cases} d\mathbf{X}(t) &= \mathbf{f}(\mathbf{X}(t))\,dt + \boldsymbol{\xi}(t)\,dt \\ \mathbf{X}(0) &= x_0. \end{cases} \tag{3.1.3}$$

where "$d\mathbf{X}$" is called a ***stochastic differential***.

Given this, it seems reasonable to say that the stochastic process $\mathbf{X}$ "solves" (3.1.3) if

$$\mathbf{X}(t) = x_0 + \int_0^t \mathbf{f}(\mathbf{X}(s))\,ds + \int_0^t \boldsymbol{\xi}(s)\,ds \quad (\text{for any } t \in [0,T]). \tag{3.1.4}$$

Here, the first integral represents the deterministic change in position and presents no problem since the integral is well-defined for all "nice" $\mathbf{f}$ despite the fact that $\mathbf{X}$ is a stochastic process. However, the second integral represents integration of a stochastic process with respect to time, whatever that may mean. Therefore, I will begin by exploring the properties of such an integral and its integrand. This will involve a calculus for stochastic processes generally referred to as *Itô calculus*.

NOTATION. For the remainder of this chapter, I will denote to $\mathbb{R}^n$-valued stochastic processes simply as $X = (X(t))_{t\in[0,T]}$. □

3.2. White Noise

Based on the heuristics above and given the numerous attractive properties of Brownian motion, ideally, I would like to express the dynamics of a price process as the solution of an SDE of the form (3.1.3) in which its stochastic behavior is modeled by Brownian motion. However, even though Brownian motion has continuous paths, recall that it is nowhere differentiable by the Paley, Wiener, Zygmund theorem (*cf.* Theorem 2.40). Hence, a true derivative in the traditional sense cannot exist.

In the following derivation, which follows Evans [**30**, ch. 3.2.3], I will show that the "derivative" of Brownian motion is a "generalized" function in some negative Sobolev space (*i.e.* a *distribution* in

the Schwartz sense). For simplicity's sake, I will only consider the one-dimensional case. However, the result can easily be generalized to higher dimensions by creating a white noise *vector* whose components are one-dimensional white noise processes.

Motivation. Let me begin by naively applying the standard definition of a derivative to Brownian motion, or

$$\xi(t) := \frac{d}{dt} W(t) \quad (t \in [0,T])$$

where $T < \infty$ and $W = (W(t))_{t\in[0,T]}$ is standard Brownian motion. Approximating this by a finite difference for some fixed $h > 0$ gives

$$\xi^h(t) := \frac{W(t+h) - W(t)}{h} \quad (\text{for all } t \in [0,T]).$$

Now, since W is standard Brownian motion, $W(t+h) - W(t)$ is an increment for all $h > 0$. Hence, for all $h > 0$ and $t \in [0,T]$

$$\mathbb{E}\left[\xi^h(t)\right] = \mathbb{E}\left[\frac{W(t+h) - W(t)}{h}\right] = 0$$

and

$$\begin{aligned}
\text{Var}\left[\xi^h(t)\right] &= \mathbb{E}\left[\xi^h(t)^2\right] \\
&= \frac{1}{h^2}\mathbb{E}\left[(W(t+h) - W(t))^2\right] \\
&= \frac{1}{h} \quad (\text{standard normal increments}).
\end{aligned}$$

Therefore,

$$\lim_{h\to 0} \text{Var}\left[\xi^h(t)\right] = \infty \quad (\text{for all } t \in [0,T])$$

and, thus, $\xi \sim N(0,\infty)$.

Moreover, for $s, t \in [0,T]$

$$\begin{aligned}
\text{Cov}\left[\xi^h(t), \xi^h(s)\right] &= \mathbb{E}\left[\xi^h(t)\,\xi^h(s)\right] \\
&= \frac{1}{h^2}\mathbb{E}\left[(W(t+h) - W(t))(W(s+h) - W(s))\right] \\
&= \frac{1}{h^2}\left(\mathbb{E}\left[W(t+h)W(s+h)\right] - \mathbb{E}\left[W(t+h)W(s)\right]\right| \\
&\quad \left. -\mathbb{E}\left[(W(s+h)W(t))\right] + \mathbb{E}\left[W(t)W(s)\right]\right) \\
&= \frac{1}{h^2}\left((t+h)\wedge(s+h) - (t+h)\wedge s - t\wedge(s+h) + t\wedge s\right).
\end{aligned}$$

Hence, for $s < t - h$

$$\text{Cov}\left[\xi^h(t), \xi^h(s)\right] = \frac{1}{h^2}\left[(s+h) - s - (s+h) + s\right] = 0.$$

Likewise, for $s > t + h$,

$$\text{Cov}\left[\xi^h(t), \xi^h(s)\right] = \frac{1}{h^2}\left[(t+h) - (t+h) - t + t\right] = 0.$$

Now, for $t - h < s < t$,

$$\text{Cov}\left[\xi^h(t), \xi^h(s)\right] = \frac{1}{h^2}\left[(s+h) - t\right].$$

Finally, for $t < s < t + h$

$$\text{Cov}\left[\xi^h(t), \xi^h(s)\right] = \frac{1}{h^2}\left[(t+h) - s\right].$$

Therefore, the support of $\text{Cov}\left[\xi^h(t), \xi^h(s)\right]$ is the interval $(t-h, t+h)$. Now, for fixed $t \in [0,T]$ and any $s \in [0,T]$ such that $s \neq t$ and any $h > 0$, $\text{Cov}\left[\xi^h(t), \xi^h(s)\right] \geq 0$. Moreover,

$$\begin{aligned}
\int_0^T \text{Cov}\left[\xi^h(t), \xi^h(s)\right] ds &= \int_{t-h}^{t} \text{Cov}\left[\xi^h(t), \xi^h(s)\right] ds + \int_t^{t+h} \text{Cov}\left[\xi^h(t), \xi^h(s)\right] ds \\
&= \frac{1}{h^2}\left(\int_{t-h}^{t} ((s+h) - t)\, ds + \int_t^{t+h} ((t+h) - s)\, ds\right) \\
&= 1.
\end{aligned}$$

Now, for any fixed $s, t \in [0,T]$, $\text{Cov}\left[\xi^h(t), \xi^h(s)\right] \to \infty$ as $h \to \infty$. Hence, for fixed $t \in [0,T]$ and any $s \in [0,T]$

$$\lim_{h\to 0} \text{Cov}\left[\xi^h(t), \xi^h(s)\right] = \delta_t(s).$$

In other words, the "process" $\xi = (\xi(t))_{t\in[0,T]}$ defined as

$$\xi(t) := \lim_{h\to 0} \xi^h(t) \quad \text{for all } t \in [0,T]$$

is a stationary, "generalized" Gaussian process with independent point mass distributions at each time t. This suggest the following definition. □

DEFINITION. A stochastic "generalized" process, $\xi = (\xi(t))_{t\in[0,T]}$, is called ***white noise*** if

(1) $\xi(t)$ is Gaussian with zero mean and infinite variance for all $t \in [0,T]$; and
(2) $\mathbb{E}[\xi(t)\,\xi(s)] = \delta_t(s)$ for all $t, s \in [0,T]$.

NOTATION. Generally, white noise is denoted as $dW = (dW(t))_{t\in[0,T]}$. □

3.3. The Itô Integral

As discussed above, modeling price changes in terms of an SDE seems like a natural extension to a purely deterministic dynamical system adjusted to account for Gaussian randomness. However, it is clear that by the Paley, Wiener, Zygmund theorem that standard calculus no longer applies to stochastic processes driven by Brownian motion (*cf.* Theorem 2.40). The standard approach

in mathematical finance is to employ *Itô calculus* to provide the necessary rigorous mathematical foundation.

Itô calculus is a rich subject with many important results, many of which are outside the scope of this paper. Fortunately, for my purposes, I will only need to present three key results from Itô calculus: the *Itô integral*, *Itô's formula*, and *Itô's product rule*. See Karatzas and Shreve [**51**] and Øksendal [**65**] for a more complete treatment of Itô calculus and stochastic calculus generally.

In constructing the Itô integral, my plan is to generally follow Evans [**30**, Section 4.2] which adapts the construction of general integration under a probability measure as I did in the prequel (*cf.* Section 1.4). Recall this started by first describing integration in terms of simple random variables and then this was extended to admit random variables of greater and greater complexity .

As usual, I will begin with a number of definitions.

DEFINITION. Let $\mathbb{L}^2(0,T)$ denote the space of all real-valued, progressively measurable, admissible processes $X = (X(t))_{t\in[0,T]}$ such that

$$\|X\|_{\mathbb{L}^2(0,T)} := \left(\mathbb{E}\left[\int_0^T X^2(t)\,ds\right]\right)^{1/2} < \infty$$

where $T < \infty$. Further, let $\mathbb{L}^1(0,T)$ denote the space of all real-valued, progressively measurable processes $X = (X(t))_{t\in[0,T]}$ such that

$$\|X\|_{\mathbb{L}^1(0,T)} := \mathbb{E}\left[\int_0^T |X(t)|\,ds\right] < \infty.$$

DEFINITION. A stochastic process $X = (X(t))_{t\in[0,T]}$ in $\mathbb{L}^2(0,T)$ is said to be a ***step process*** if there exists a partition $\{0 = t_0 < t_1 < \cdots < t_m = T\}$ such that

$$X(t) \equiv X_j \quad \text{for } t_j \leq t < t_{j+1} \text{ where } j = 0,\ldots,m.$$

DEFINITION. Assume that $\{0 = t_0 < t_1 < \cdots < t_k = T\}$ be a partition and let $X = (X(t))_{t\in[0,T]}$ be a step process in $\mathbb{L}^2(0,T)$ on Π. Then the ***Itô integral*** is the stochastic process $I[\cdot] = (I[\cdot](t))_{t\in\mathbb{T}}$ defined as

$$I[X](t) := \sum_{j=0}^{m-1} X_j\left(W(t_{j+1}) - W(t_j)\right)$$

where $X = (X(t))_{t\in[0,T]}$ and $t \in [0,T]$.

NOTATION. Typically, the explicitly dependence on the stochastic process is ignored and the integral is simply denoted as

$$I(t) := I[X](t) \quad (\text{for all } t \in [0,T]).$$

However, going forward, I will use the more familiar notation

$$\int_0^t X(s)\,dW(s) := I[X](t) \quad (\text{for all } t \in [0,T]).$$

□

As expected, this integral is independent of the partition and, therefore, is unique. The proof of this result follows the same logic as the proof of Theorem 1.21.

THEOREM 3.1. *Let* $\{0 = t_0 < t_1 < \cdots < t_k = T\}$ *and* $\{0 = s_0 < t_1 < \cdots < s_l = T\}$ *be finite partitions of* $[0,T]$ *and let* $X = (X(t))_{t\in[0,T]}$ *be a step process in* $\mathbb{L}^2(0,T)$ *with respect to both partitions. Then*

$$\sum_{i=0}^{k-1} a_i \left(W(t_{i+1}) - W(t_i)\right) = \sum_{j=0}^{l-1} b_j \left(W(t_{j+1}) - W(t_j)\right)$$

where

$$X(t) \equiv a_i \quad (\textit{for } t_i \le t < t_{i+1} \textit{ with } i = 0,\ldots,k)$$

and

$$X(t) \equiv b_j \quad (\textit{for } t_j \le t < t_{j+1} \textit{ with } j = 0,\ldots,l).$$

Hence, $\int_0^t X(s)\,dW(s)$ *as defined above is independent of the partition and is therefore unique.*

As defined above, such integrals have a number of important properties which are proved in Evans [**30**, pg. 66-67]

THEOREM 3.2 (Properties of Itô integrals with step process integrands). *Let* $X = (X(t))_{t\in[0,T]}$ *and* $Y = (Y(t))_{t\in[0,T]}$ *be step processes in* $\mathbb{L}^2(0,T)$*. Then for any* $t \in [0,T]$

(1) $\int_0^t (aX(s) + bY(s))\,dW(s) = a\int_0^t X(s)\,dW(s) + b\int_0^t Y(s)\,dW(s)$ (for any $a,b \in \mathbb{R}$);
(2) $\mathbb{E}\left[\int_0^t X(s)\,dW(s)\right] = 0$;
(3) $\mathbb{E}\left[\left(\int_0^t X(s)\,dW(s)\right)^2\right] = \mathbb{E}\left[\int_0^t X^2(s)\,ds\right]$; *and*
(4) $\mathbb{E}\left[\int_0^t X(s)\,dW(s)\int_0^t Y(s)\,dW(s)\right] = \mathbb{E}\left[\int_0^t X(s)\,Y(s)\,ds\right]$.

In order to extend this definition to general stochastic processes, the following result shows that any process in $\mathbb{L}^2(0,T)$ can be approximated by step processes whose proof is outlined in Evans [**30**, pg. 68].

THEOREM 3.3. *Let* $X = (X(t))_{t\in[0,T]}$ *be a step process in* $\mathbb{L}^2(0,T)$*. Then there exists a sequence of bounded step processes* $X^m = (X^m(t))_{t\in[0,T]}$*, such that*

$$\mathbb{E}\left[\int_0^t |X^m(s) - X(s)|^2\,ds\right] \to 0 \quad (\textit{for any } t \in [0,T]).$$

Using the above theorem, it is now possible to define the Itô integral for general stochastic processes.

DEFINITION. Let $X = (X(t))_{t\in[0,T]}$ be a stochastic process in $\mathbb{L}^2(0,T)$ approximated by a sequence of bounded step processes $X^m = (X^m(t))_{t\in[0,T]}$ such that $\mathbb{E}\left[\int_0^t |X^m(s) - X(s)|^2 ds\right] \to 0$. Then the stochastic process $I[X] = (I[X](t))_{t\in[0,T]}$ defined as

$$I[X](t) := \lim_{m\to\infty} \int_0^t X^m(s)\, dW(s) \quad \text{(for all } t \in [0,T])$$

is called the ***indefinite Itô integral of*** X and is commonly denoted simply as $I(\cdot)$ where the integrand, X, is understood.

NOTATION. Typically, the integral is denoted as

$$\int_0^t X(s)\, dW(s) := I(t) \quad \text{(for any } t \in [0,T]).$$

□

REMARK. From a finance perspective, the Itô integral represents the profit (or loss) of holding a position of X up to time t in a particular financial instrument due solely to price changes driven by the Brownian motion. Moreover, it is non-anticipating since it is $\mathcal{F}_t$-adapted for $t \in [0,T]$. □

EXAMPLE. Let $W = (W(t))_{t\in[0,T]}$ be a standard Brownian motion. Then for all $t \in [0,T]$

$$\int_0^t W(s)\, dW(s) = \frac{1}{2}W^2(t) - \frac{t}{2}.$$

To see this, let $\left\{0 = t_0^k < t_1^k < \cdots < t_{m_k}^k = t\right\}$ be a partition of $[0,t]$ for any $t \in [0,T]$ and define the step processes $W^k = \left(W^k(t)\right)_{t\in[0,T]}$ as

$$W^k(t) = \sum_{j=0}^{m_k-1} W\left(t_j^k\right) 1_{\left[t_j^k, t_{j+1}^k\right)}(t).$$

Then the W^k approximate Brownian motion in $\mathbb{L}^2(0,T)$ since, for any $t \in [0,T]$

$$\begin{aligned}
\mathbb{E}\left[\int_0^t \left|W^k(s) - W(s)\right|^2 ds\right] &= \mathbb{E}\left[\sum_{j=0}^{m_k-1} \int_{t_j^k}^{t_{j+1}^k} \left(W^k(s) - W(s)\right)^2 ds\right] \\
&= \mathbb{E}\left[\sum_{j=0}^{m_k-1} \int_{t_j^k}^{t_{j+1}^k} \left(W\left(t_j^k\right) - W(s)\right)^2 ds\right] \\
&= \sum_{j=0}^{m_k-1} \int_{t_j^k}^{t_{j+1}^k} \mathbb{E}\left[\left(W\left(t_j^k\right) - W(s)\right)^2\right] ds \\
&\quad \sum_{j=0}^{m_k-1} \int_{t_j^k}^{t_{j+1}^k} \mathrm{Var}\left[W\left(t_j^k\right) - W(s)\right] ds \\
&= \sum_{j=0}^{m_k-1} \int_{t_j^k}^{t_{j+1}^k} \left(s - t_j^k\right) ds \\
&= \sum_{j=0}^{m_k-1} \left(\frac{1}{2}\left(\left(t_{j+1}^k\right)^2 - \left(t_{j+1}^k\right)^2\right) - t_j^k\left(t_{j+1}^k - t_j^k\right)\right) \\
&= \sum_{j=0}^{m_k-1} \frac{1}{2}\left(\left(t_{j+1}^k\right)^2 + \left(t_{j+1}^k\right)^2\right) - t_j^k t_{j+1}^k \\
&= \frac{1}{2}\sum_{j=0}^{m_k-1} \left(t_{j+1}^k - t_j^k\right)^2 \to 0 \text{ as } k \to \infty.
\end{aligned}$$

Therefore, using the fact that W^k is step function for each k, I can express the stochastic integral of Brownian motion for any $t \in [0,T]$ as

$$\begin{aligned}
\int_0^t W(s)\, dW(s) &= \lim_{k\to\infty} \int_0^t W^k(s)\, dW(s) \\
&= \lim_{k\to\infty} \sum_{j=0}^{l_k-1} W\left(t_j^k\right)\left(W\left(t_{j+1}^k\right) - W\left(t_j^k\right)\right). \qquad (3.3.1)
\end{aligned}$$

The following calculation will be helpful in what follows.

$$\begin{aligned}
\frac{1}{2}\sum_{j=0}^{l_k-1}\left(W\left(t_{j+1}^k\right)-W\left(t_j^k\right)\right)^2 &= \frac{1}{2}\sum_{j=0}^{l_k-1}W^2\left(t_{j+1}^k\right)+\frac{1}{2}\sum_{j=0}^{l_k-1}W^2\left(t_j^k\right)-\sum_{j=0}^{l_k-1}W\left(t_j^k\right)W\left(t_{j+1}^k\right) \\
&= \frac{1}{2}\sum_{j=1}^{l_k}W^2\left(t_j^k\right)+\frac{1}{2}\sum_{j=0}^{l_k-1}W^2\left(t_j^k\right)-\sum_{j=0}^{l_k-1}W\left(t_j^k\right)W\left(t_{j+1}^k\right) \\
&= \frac{1}{2}W^2\left(t_{l_k}^k\right)+\frac{1}{2}W^2\left(t_0^k\right)+\sum_{j=1}^{l_k-1}W^2\left(t_j^k\right)-\sum_{j=0}^{l_k-1}W\left(t_j^k\right)W\left(t_{j+1}^k\right). \\
&= \frac{1}{2}W^2\left(t_{l_k}^k\right)-\frac{1}{2}W^2\left(t_0^k\right)+\sum_{j=0}^{l_k-1}W^2\left(t_j^k\right)-\sum_{j=0}^{l_k-1}W\left(t_j^k\right)W\left(t_{j+1}^k\right).
\end{aligned}$$

But, $t_{l_k}^k = t$ and $t_0^k = 0$ for all k. Moreover, $W(0) = 0$. Hence,

$$\begin{aligned}
\frac{1}{2}\sum_{j=0}^{l_k-1}\left(W\left(t_{j+1}^k\right)-W\left(t_j^k\right)\right)^2 &= \sum_{j=0}^{l_k-1}W\left(t_j^k\right)\left(W\left(t_j^k\right)-W\left(t_{j+1}^k\right)\right)-\frac{1}{2}W^2(t) \\
&= -\sum_{j=0}^{l_k-1}W\left(t_j^k\right)\left(W\left(t_{j+1}^k\right)-W\left(t_j^k\right)\right)-\frac{1}{2}W^2(t)
\end{aligned}$$

Therefore, rearranging terms

$$\sum_{j=0}^{l_k-1}W(t_j)\left(W(t_{j+1})-W(t_j)\right)=\frac{1}{2}W^2(t)-\frac{1}{2}\sum_{j=0}^{l_k-1}\left(W(t_{j+1})-W(t_j)\right)^2.$$

Using this expression, I can rewrite (3.3.1) as

$$\begin{aligned}
\int_0^t W(s)\,dW(s) &= \lim_{k\to\infty}\sum_{j=0}^{l_k-1}W\left(t_j^k\right)\left(W\left(t_{j+1}^k\right)-W\left(t_j^k\right)\right) \\
&= \frac{1}{2}W^2(t)-\frac{1}{2}\lim_{k\to\infty}\sum_{j=0}^{l_k-1}\left(W\left(t_{j+1}^k\right)-W\left(t_j^k\right)\right).^2
\end{aligned}$$

Recalling the definition of quadratic variation (*cf.* 2.5.4)

$$\int_0^t W(s)\,dW(s)=\frac{1}{2}W^2(t)-\frac{1}{2}[W](t).$$

Since, $[W](t) = t$ for standard Brownian motion (*cf.* 2.44)

$$\int_0^t W(s)\,dW(s)=\frac{1}{2}W^2(t)-\frac{t}{2}. \tag{3.3.2}$$

□

REMARK. Recall that for a real-valued, integrable, deterministic function f such that $f(0) = 0$, by the chain rule,

$$\begin{aligned}\int_0^t f(s)\, d(f(s)) &= \int_0^t f(s) f'(s)\, ds \\ &= \frac{1}{2} f^2(s) \Big|_0^t = \frac{1}{2} f^2(t).\end{aligned}$$

Hence, the extra term in (3.3.2) accounts for the fact that Brownian motion has non-zero quadratic variation. □

Under the definition above, an Itô integral possesses a number of very useful properties which are proved in Pascucci [**66**, pg. 145-149]

THEOREM 3.4. *Let $X = (X(t))_{t\in[0,T]}$ be a stochastic process in $\mathbb{L}^2(0,T)$. Then $\int_0^t X(s)\, dW(s)$ is a martingale and has a version that has continuous paths.*

THEOREM 3.5 (Properties of Itô integrals). *Let $X = (X(t))_{t\in[0,T]}$ and $Y = (Y(t))_{t\in[0,T]}$ be stochastic processes in $\mathbb{L}^2(0,T)$. Then for any $t \in [0,T]$*

(1) $\int_0^t (aX(s) + bY(s))\, dW(s) = a\int_0^t X(s)\, dW(s) + b\int_0^t Y(s)\, dW(s) \quad a, b \in \mathbb{R}$ *(linearity)*;
(2) $\int_s^t X(\tau)\, dW(\tau) = \int_0^t X(\tau)\, dW(\tau) - \int_0^s X(\tau)\, dW(\tau)$ for $0 \le s \le t$;
(3) $\mathbb{E}\left[\int_0^t X(s)\, dW(s)\right] = 0$;
(4) $\mathbb{E}\left[\left(\int_0^t X(s)\, dW(s)\right)^2\right] = \mathbb{E}\left[\int_0^t X^2(s)\, ds\right]$ *(Itô's isometry); and*
(5) $\mathbb{E}\left[\int_0^t X(s)\, dW(s) \int_0^t Y(s)\, dW(s)\right] = \mathbb{E}\left[\int_0^t X(s) Y(s)\, ds\right]$.

4 is known as ***Itô's isometry*** since it is equivalent to

$$\|I[X]\|_{L^2(\Omega)} = \|X\|_{\mathbb{L}^2(0,T)}.$$

The following useful corollary is proved in Shreve [**76**, Page 134]

COROLLARY 3.6. *Let $X = (X(t))_{t\in[0,T]}$ be a stochastic process in $\mathbb{L}^2(0,T)$. Then for any $t \in [0,T]$*

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

REMARK. Itô's isometry and the above corollary describe the relationship between the quadratic variation of the Itô integral and its variance. Since the expected value of an Itô integral is zero, then for all $t \in [0,T]$

$$\begin{aligned}\operatorname{Var}\left[\int_0^t X(s)\, dW(s)\right] &= \mathbb{E}\left[\int_0^t X^2(s)\, ds\right] \\ &= \mathbb{E}\left[[I[X]](t)\right]\end{aligned}$$

Here, the quadratic variation is computed over a single path and, although it is bounded a.s. for each path, it may vary path to path. Conversely, the variance is the average quadratic variation over all paths. □

3.4. The Stochastic Differential

Recall above (*cf.* (3.1.4)) that if a $\mathbb{R}^n$-valued stochastic process $\mathbf{X} = (\mathbf{X}(t))_{t\in[0,T]}$ could be written as

$$\mathbf{X}(t) = \mathbf{X}(0) + \int_0^t \mathbf{f}(\mathbf{X}(s))\,ds + \int_0^t \boldsymbol{\xi}(s)\,ds \quad (\text{for any } t \in [0,T])$$

where $\mathbf{f}$ is a smooth vector field and $\boldsymbol{\xi}$ is n-dimensional white noise, I "defined" its *stochastic differential* "$d\mathbf{X}$" as

$$d\mathbf{X}(t) = \mathbf{f}(\mathbf{X}(t))\,dt + \boldsymbol{\xi}(t)\,dt.$$

Given the definition of the Itô integral and white noise, I can now provide a formal definition.

Here, I will restrict the discussion to a real-valued stochastic processes which can easily be extended to higher dimensions. Further, assume that $W = (W(t))_{t\in[0,T]}$ is adapted standard Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, F, \mathbb{P})$ and the time index is a bounded interval $[0,T]$ where $T < \infty$.

DEFINITION. Let $X = (X(t))_{t\in[0,T]}$ be a real-valued, admissible stochastic process. If X can be expressed as

$$X(t) = X(s) + \int_s^t \mu(X(\tau),\tau)\,d\tau + \int_s^t \sigma(X(\tau),\tau)\,dW(\tau) \quad (s,t \in [0,T] \text{ with } s \le t).$$

for some real-valued (deterministic) functions $\mu \in \mathbb{L}^1(0,T)$ and $\sigma \in \mathbb{L}^2(0,T)$, then it is called an ***Itô process*** and we say that X has the ***stochastic differential***

$$dX(t) = \mu(X(t),t)\,dt + \sigma(X(t),t)\,dW(t) \quad (t \in [0,T]). \tag{3.4.1}$$

NOTATION. The stochastic differential is only symbolically defined since "dX" has no meaning on its own. □

The next result follows immediately from the definition.

THEOREM 3.7. *Let $X = (X(t))_{t\in[0,T]}$ be an Itô process. Then X is a Gaussian process.*

The next result, which is proved in Shreve [**76**, pg. 143-144], shows the computation of the quadratic variation of an Itô process is rather straightforward.

THEOREM 3.8. *Let $X = (X(t))_{t\in[0,T]}$ be an Itô process. Then for any $t \in [0,T]$,*

$$[X](t) = \int_0^t \sigma^2(X(s),s)\,ds. \tag{3.4.2}$$

EXAMPLE. In the case of Brownian motion, $\sigma(X(s),s) \equiv 1$. Hence, $[W](t) = t$ which agrees with the result from Theorem 3.16 above. □

EXAMPLE. Given that the Itô integral is an Itô process of the form

$$I[X](t) := \int_0^t X(s)\, dW(s)$$

then its differential is

$$dI[X](t) = X(t)\, dW(t).$$

Additionally, by 3.6, the quadratic variation of the Itô integral is

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

Thus, the integral's quadratic variation is an Itô process and its differential is

$$d[I[X]](t) = X^2(t)\, dt.$$

□

Motivation. Recall above (*cf.* Theorem 2.44), I showed the quadratic variation of Brownian motion increases linearly with time, or

$$[W](t) = t$$

This suggests that the rate at which the quadratic variation of Brownian motion grows should be constant, or, in terms of its differential

$$d[W](t) = dt. \tag{3.4.3}$$

However, if $\{0 = t_0 < t_1 \cdots < t_k = t\}$ is some partition for any $t \in [0, T]$, then by definition (*cf.* 2.5.4)

$$[W](t) = \lim_{k\to\infty} \sum_{j=0}^{k-1} \left(W(t_{j+1}) - W(t_{j+1})\right)^2.$$

which suggest

$$d[W](t) = dW(t)\, dW(t) \;\Rightarrow\; dW(t)\, dW(t) = dt.$$

This can be formalized as follows. □

THEOREM 3.9. *If $W = (W(t))_{t\in[0,T]}$ is Brownian motion, then for all $t \in [0, T]$*

(1) $dW(t)\, dW(t) = dt$*;*
(2) $dW(t)\, dt = 0$*; and*
(3) $dtdt = 0$.

PROOF. Let $W = (W(t))_{t\in[0,T]}$ be standard Brownian motion and assume for for some $t \in [0, T]$ that the partition

$$\Pi = \{0 = t_0 < t_1 \cdots < t_k = t\}$$

is of the form $t_j = \frac{j}{k}t$ for $j = 0, \ldots, k$. Then

$$t_{j+1} - t_j = \frac{t}{k} \quad (0 \leq j \leq k-1).$$

(1). Define the discrete-time stochastic processes $(X_j)_{t_j \in \Pi}$ such that, for $j = 0, \ldots, k-1$,

$$X_j = \frac{W(t_{j+1}) - W(t_j)}{\sqrt{t_{j+1} - t_j}}.$$

Obviously, the X_j are i.i.d. such that $X_j = N(0,1)$. Squaring both sides and rearranging terms gives

$$(W(t_{j+1}) - W(t_j))^2 = \frac{tX_{j+1}^2}{k}$$

By the *law of large numbers* (*cf.* [**8**, pg. 8-11])

$$\begin{aligned} \lim_{k\to\infty} \sum_{j=0}^{k-1} \frac{X_{j+1}^2}{k} &= \mathbb{E}\left[X_{j+1}^2\right]. \\ &= \mathrm{Var}\left[X_{j+1}\right] \\ &= 1. \end{aligned}$$

Therefore,

$$\lim_{k\to\infty} \sum_{j=0}^{k-1} (W(t_{j+1}) - W(t_j))^2 = t.$$

Since, $t = \sum_{j=0}^{k-1} t_{j+1} - t_j$, then

$$\lim_{k\to\infty} \sum_{j=0}^{k-1} (W(t_{j+1}) - W(t_j))^2 - (t_{j+1} - t_j) = 0.$$

This limit can be written as

$$dW(t)\, dW(t) = dt.$$

(2). By definition, the covariation of Brownian motion and t is

$$\begin{aligned} [W,t](t) &= \lim_{k\to\infty} \sum_{j=0}^{k-1} (W(t_{j+1}) - W(t_j))(t_{j+1} - t_j) \\ &\leq \lim_{k\to\infty} \sup_{0\leq i\leq k-1} |W(t_{j+1}) - W(t_j)| \sum_{j=0}^{k-1} (t_{j+1} - t_j) \\ &\leq \lim_{k\to\infty} \sup_{0\leq i\leq k-1} |W(t_{i+1}) - W(t_i)|\, t_{k-1} = 0 \end{aligned}$$

Since Brownian motion has continuous paths,$[W,t](t) \leq 0$. On the other hand,

$$\begin{aligned}
[W,t](t) &= \lim_{k\to\infty} \sum_{j=0}^{k-1} (W(t_{j+1}) - W(t_j))(t_{j+1} - t_j) \\
&\geq \lim_{k\to\infty} \inf_{0\leq i\leq k-1} -|W(t_{j+1}) - W(t_j)| \sum_{j=0}^{k-1} (t_{j+1} - t_j) \\
&\geq \lim_{k\to\infty} \inf_{0\leq i\leq k-1} -|W(t_{i+1}) - W(t_i)| \, t_{k+1} = 0
\end{aligned}$$

Hence, $[W,t](t) = 0$ and $dW(t)\,dt = 0$.

(3). Likewise, the mapping $t \mapsto t$ is continuous. Hence, its quadratic variation is zero and

$$dtdt = 0.$$

□

This result has a number of useful corollaries. The first follows immediately from the definition correlated Brownian motions (*cf.* 2.5.7).

COROLLARY 3.10. *If $W = (W(t))_{t\in[0,T]}$ and $B = (B(t))_{t\in[0,T]}$ are standard Brownian motions, then for all $t \in \mathbb{T}$*

(1) *$dW(t)\,dB(t) = 0$ if W* and *B are independent*; *and*
(2) *$dW(t)\,dB(t) = \rho dt$ if they are correlated where ρ is the correlation coefficient.*

COROLLARY 3.11. *Let $X = (X(t))_{t\in[0,T]}$ be an Itô process. Then for any $t \in [0,T]$,*

$$d[X](t) = dX(t)\,dX(t).$$

PROOF. Let $X = (X(t))_{t\in[0,T]}$ be an Itô process of the form

$$dX(t) = \mu(X(t),t)\,dt + \sigma(X(t),t)\,dW(t).$$

Then

$$\begin{aligned}
dX(t)\,dX(t) &= (\mu(X(t),t)\,dt + \sigma(X(t),t)\,dW(t))^2 \\
&= \sigma^2(X(t),t)\,dW(t)\,dW(t) \\
&= \sigma^2(X(t),t)\,dt \\
&= d[X](t).
\end{aligned}$$

□

COROLLARY 3.12. *Let $X = (X(t))_{t\in[0,T]}$ and $Y = (Y(t))_{t\in[0,T]}$ be Itô processes. Then for any $t \in [0,T]$,*

$$d[X,Y](t) = dX(t)\,dY(t).$$

PROOF. Let $X = (X(t))_{t\in[0,T]}$ and $Y = (Y(t))_{t\in[0,T]}$ be Itô processes of the form

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

and

$$dY(t) = \hat{\mu}(X(t), t)\,dt + \hat{\sigma}(X(t), t)\,dB(t)$$

with

$$[W(t), B(t)] = \rho t$$

for $|\rho| \leq 1$ and $t \in [0, T]$.

Then by 2.5.7

$$\begin{aligned} dX(t)\,dY(t) &= (\mu(X(t), t)\,dt + \sigma(X(t), t)\,dW(t))(\hat{\mu}(X(t), t)\,dt + \hat{\sigma}(X(t), t)\,dB(t)) \\ &= \sigma(X(t), t)\,\hat{\sigma}(X(t), t)\,dW(t)\,dB(t) \\ &= \sigma(X(t), t)\,\hat{\sigma}(X(t), t)\,\rho dt \\ &= d[X, Y](t). \end{aligned}$$

The next two results, both of which are due to Lévy and proved in Shreve [**76**, Section 4.6.3], characterize Brownian motion in terms of its quadratic variation. □

THEOREM 3.13. *Let $M = (M(t))_{t\in[0,T]}$ be a martingale such that $M(0) = 0$, M has continuous paths, and $[M](t) = t$ for all $t \in [0, T]$. Then M is standard Brownian motion.*

THEOREM 3.14. *Let $M_1 = (M_1(t))_{t\in[0,T]}$ and $M_2 = (M_2(t))_{t\in[0,T]}$ be martingales such that for $j = 1, 2$, $M_j(0) = 0$, M_j has continuous paths, and $[M_j](t) = t$ for all $t \in [0, T]$. Then M is standard Brownian motion. Moreover, if $[M_1, M_2](t) = 0$, they are independent.*

3.5. Itô's Chain and Product Rules

The fact that Brownian motion has non-zero quadratic variation also requires changes to the chain and product rules of standard calculus.

3.5.1. Itô's Chain Rule. The following result, which is one of the most famous in mathematical finance, is the stochastic version of the chain rule known as *Itô's formula*. It is proved in Øksendal [**65**, pg. 46-48.]

THEOREM 3.15 (Itô's formula in one-dimension). *Let $X = (X(t))_{t\in[0,T]}$ be an Itô process with the stochastic differential*

$$dX(t) = \mu(X(t), t)\,dt + \sigma(X(t), t)\,dW(t) \quad \text{for any } t \in [0, T].$$

Further, assume that $f \in C_1^2(\mathbb{R} \times [0,T])$. *Then for any* $t \in [0,T]$, *the* ***stochastic differential of*** f ***acting on*** X *is*

$$\begin{aligned} df(X(t),t) &= f_t(X(t),t)\,dt + f_x(X(t),t)\,dX(t) + \frac{1}{2}f_{xx}(X(t),t)\,dX(t)\,dX(t) \\ &= \left(f_t(X(t),t) + f_x(X(t),t)\,f(X(t),t) + \frac{1}{2}f_{xx}(X(t),t)\right)dt \\ &\quad + f_x(X(t),t)\,g(X(t),t)\,dW(t). \end{aligned}$$

Further, if f *has no explicit dependency on time* (*i.e.* $f(x,t) = f(x)$), *then*

$$\begin{aligned} df(X(t)) &= f_x(X(t))\,dX(t) + \frac{1}{2}f_{xx}(X(t))\,dX(t)\,dX(t) \\ &= \left(f_x(X(t))\,f(X(t)) + \frac{1}{2}f_{xx}(X(t))\right)dt \\ &\quad + f_x(X(t))\,g(X(t))\,dW(t). \end{aligned}$$

3.5.2. Itô's Product Rule. Another useful result from stochastic calculus is known as *Itô's product rule* and, as the name implies, is the analog of the deterministic product rule. It follows directly from Itô's formula and is proved in Evans [**30**, pg. 74-75.]

THEOREM 3.16 (Itô's product rule in one-dimension). *Let* $X = (X(t))_{t\in[0,T]}$ *and* $Y = (Y(t))_{t\in[0,T]}$ *be Itô processes such that, for each* $t \in [0,T]$,

$$\begin{cases} dX(t) &= \mu(X(t),t)\,dt + \sigma(X(t),t)\,dW(t) \\ dY(t) &= \hat{\mu}(Y(t),t)\,dt + \hat{\sigma}(Y(t),t)\,dW(t) \end{cases}$$

with $\mu, \hat{\mu} \in L^1(\mathbb{R}\times[0,T];\mathbb{R})$, $\sigma, \hat{\sigma} \in L^2(\mathbb{R}\times[0,T];\mathbb{R})$. *Then*

$$d(X(t)\,Y(t)) = Y(t)\,dX(t) + X(t)\,dY(t) + \sigma\hat{\sigma}dt.$$

REMARK. Again, the extra term is due to the fact that Brownian motion has non-zero quadratic variation. □

3.6. Itô Calculus in Higher Dimensions

In this section, assume all stochastic processes are $\mathbb{R}^n$-valued and admissible unless otherwise noted. In particular, $W = ((W^1(t),\ldots,W^n(t)))_{t\in[0,T]}$ is standard n-dimensional Brownian motion where W^k is standard one-dimensional Brownian motion for $k = 1,\ldots,n$.

3.6.1. The Itô Integral in Higher Dimensions.

DEFINITION. Let $X = ((X^{ij}(t)))_{t\in[0,T]}$ be an $\mathbb{M}^{n\times m}$-valued stochastic process. X is said belong to $\mathbb{L}^2_{n\times m}(0,T)$ if

$$X^{ij} \in \mathbb{L}^2(0,T) \quad (i = 1,\ldots,n; j = 1,\ldots,m).$$

In particular, an $\mathbb{R}^n$-valued stochastic process $Y = ((Y^1(t), \ldots, Y^n(t)))_{t\in[0,T]}$ is in $\mathbb{L}_n^2(0,T)$ if

$$X^i \in \mathbb{L}^2(0,T) \quad (i = 1, \ldots, n).$$

DEFINITION. Let $X \in \mathbb{L}_n^2(0,T)$. Then for any $t \in [0,T]$, the ***n*-dimensional Itô's integral of** X is the n-dimensional stochastic process given by

$$\int_0^t X dW = \left(\sum_{j=1}^m \int_0^t X^{1j}(s)\, dW^j(s), \ldots, \sum_{j=1}^m \int_0^t X^{nj}(s)\, dW^j(s) \right).$$

Again, under the definition above, an n-dimensional Itô integral possesses a number of very useful properties.

THEOREM 3.17 (Properties of Itô integrals in n-dimensions). *Let* $X = ((X^{ij}(t)))_{t\in[0,T]}$ *and* $X = ((Y^{ij}(t)))_{t\in\mathbb{T}}$ *be in* $\mathbb{L}_{n\times m}^2(0,T)$. *Then*

(1) $\int_0^t (aX(s) + bY(s))\, dW(s) = a\int_0^t X(s)\, dW(s) + b\int_0^t Y(s)\, dW(s) \quad (a, b \in \mathbb{R})$;
(2) $\mathbb{E}\left[\int_0^t X(s)\, dW(s)\right] = 0$;
(3) $\mathbb{E}\left[\left|\int_0^t X(s)\, dW(s)\right|^2\right] = \mathbb{E}\left[\int_0^t |X(s)|^2\, ds\right]$; *and*
(4) $\mathbb{E}\left[\int_0^t X(s)\, dW_s \int_0^t X(s)\, dW_s\right] = \mathbb{E}\left[\int_0^t X(s) X(s)\, ds\right]$

where

$$|X|^2 = \sum_{\substack{1\le i\le n \\ 1\le j\le m}} \left|X^{ij}\right|^2.$$

3.6.2. Itô's Chain and Product Rule in Higher Dimensions. Itô calculus can be easily extended to higher dimensions in the natural way. Here, I begin with the Itô integral.

DEFINITION. Let $X = ((X^1(t), \ldots, X^n(t)))_{t\in[0,T]}$ be an $\mathbb{R}^n$-valued stochastic process in $\mathbb{L}_n^2(0,T)$. If, for any $0 \le s \le t \le T$, X can be expressed as

$$X(t) = X(s) + \int_s^t \mu(\tau)\, d\tau + \int_s^t \sigma(\tau)\, dW(s)$$

for some $\mu \in \mathbb{L}_n^1(0,T), \sigma \in \mathbb{L}_{n\times m}^2(0,T)$, then, we say that X has the ***stochastic differential***

$$dX(t) = \mu(t)\, dt + \sigma(t)\, dW(t) \quad (t \in [0,T])$$

where for each $t \in [0,T]$.

$$dX^i(t) = \mu^i(t)\, dt + \sum_{j=1}^m \sigma^{ij}(t)\, dW^j(t) \quad (i = 1, \ldots, n).$$

THEOREM 3.18 (Itô's formula in n-dimensions). *Assume* $X = ((X^1(t), \ldots, X^n(t)))_{t\in[0,T]}$ *be an* $\mathbb{R}^n$*-valued stochastic process in* $\mathbb{L}_n^2(0,T)$ *such that it has the stochastic differential*

$$dX(t) = \mu(t)\,dt + \sigma(t)\,dW(t) \quad (t \in [0,T]).$$

Further, let $f \in C_1^2(\mathbb{R}\times[0,T];\mathbb{R})$. *Then for any* $t \in [0,T]$, *the* ***stochastic differential of*** f ***acting on*** X *is*

$$df(X(t),t) = f_t(X(t),t)\,dt + \sum_{i=1}^{n} f_{x_i}(X(t),t)\,dX^i(t) + \frac{1}{2}\sum_{i,j=1}^{n} f_{x_ix_j}(X(t),t)\sum_{k=1}^{m}\Lambda^{ik}\Lambda^{jk}dt.$$

Further, if f *has no explicit dependency on time* (*i.e.* $f(x,t) = f(x)$), *then*

$$df(X(t)) = \sum_{i=1}^{n} f_{x_i}(X(t))\,dX^i(t) + \frac{1}{2}\sum_{i,j=1}^{n} f_{x_ix_j}(X(t))\sum_{k=1}^{m}\Lambda^{ik}\Lambda^{jk}dt.$$

PROOF. See Evans [**30**, pg. 79-80] for an outline of the proof. □

THEOREM 3.19 (Itô's product rule with m Brownian motions). *Assume that for each* $t \in [0,T]$,

$$\begin{cases} dX(t) &= \mu(X(t),t)\,dt + \sum_{j=1}^{m}\sigma^j(X(t),t)\,dW^j(t) \\ dY(t) &= \hat{\mu}(Y(t),t)\,dt + \sum_{j=1}^{m}\hat{\sigma}^j(Y(t),t)\,dW^j(t) \end{cases}$$

with $\mu, \hat{\mu} \in \mathbb{L}^1(0,T)$ *and* $\sigma^k, \hat{\sigma}^k \in \mathbb{L}^2(0,T)$ *for* $k = 1,\ldots,m$. *Then*

$$d(X(t)Y(t)) = Y(t)\,dX(t) + X(t)\,dY(t) + \sum_{j=1}^{m}\sigma^j\hat{\sigma}^j dt.$$

PROOF. The proof follows directly from Theorem 3.16. □

3.7. Important Applications

In addition to being the foundation of SDE, Itô calculus provides the tools to establish a number of important results in continuous-time finance. For the purpose of this paper, I will only present three such results. The first is useful since the coefficients in many of the price models commonly used are deterministic.

THEOREM 3.20. *Assume that* $X = (X(t))_{t\in[0,T]}$ *is some stochastic process in* $\mathbb{L}^2(0,T)$. *Then for any* $t \in [0,T]$, $I[X]$ *is normally distributed with mean zero and variance* $\int_0^t X^2(s)\,ds$ *where*

$$I[X](t) = \int_0^t X(s)\,dW(s).$$

PROOF. See Shreve [**76**, pg. 149-150]. □

The next result is known as the *martingale representation theorem* which shows that when a price process is driven by a standard Brownian motion under its standard filtration, then the only source of uncertainty in the future price is due to that Brownian motion. It is proved in Øksendal [**65**, pg. 53-54].

THEOREM 3.21 (The martingale representation theorem). *Let $W = (W(t))_{t\in[0,T]}$ be standard Brownian motion. Then for any adapted martingale $M = (M(t))_{t\in\mathbb{T}}$, there is a unique stochastic process $X = (X(t))_{t\in[0,T]}$ in $\mathbb{L}^2(0,T)$ such that, for all $t \in [0,T]$.*

$$M_t = M_0 + \int_0^t X(s)\, dW(s).$$

This final result from Itô's calculus is known Girsanov's theorem and is used in continuous-time finance to transform any given (discounted) price process into a martingale using the risk-neutral probability measure using the Radon-Nikodým derivative (*cf.* (1.11.4)).

Motivation. Consider standard Brownian motion $W = (W(t))_{t\in\mathbb{T}}$ and some process $\Theta = (\Theta(t))_{t\in\mathbb{T}}$ in $\mathbb{L}^2(0,T)$ on a filtered probability space $(\Omega, \mathcal{F}, F, \mathbb{P})$. Now, define the process $Z_\Theta = (Z_\Theta(t))_{t\in\mathbb{T}}$ as

$$Z_\Theta(t) = \exp\left(-\frac{1}{2}\int_0^t |\Theta(s)|^2\, ds - \int_0^t \Theta(s)\, dW(s)\right) \quad (\text{for } t \in [0,T]).$$

Clearly, $Z_\Theta > 0$ and by Itô's formula

$$\begin{aligned} dZ_\Theta(t) &= -Z_\Theta(t)\left(\left(\frac{1}{2}\Theta^2(t)\, dt + \Theta(t)\, dW(t)\right) - \frac{1}{2}\Theta^2(t)\, dt\right) \\ &= -Z_\Theta(t)\,\Theta(t)\, dW(t). \end{aligned}$$

Hence, Z_Θ is a martingale. Then, for any $t \in \mathbb{T}$

$$\mathbb{E}[Z_\Theta(t)] = \mathbb{E}[Z_\Theta(0)] = 1.$$

Therefore, for $t \in \mathbb{T}$, I can the apply the change of measure theorem (*cf.* Theorem 1.83) to construct an equivalent probability $\mathbb{Q}_t$ to $\mathbb{P}$ on $(\Omega, \mathcal{F})$ where

$$\mathbb{Q}_t(A) = \int_A Z_\Theta(t)\, d\mathbb{P} \quad (\text{for any } A \in \mathcal{F})$$

or in terms of the Radon-Nikodým derivative

$$\frac{d\mathbb{Q}_t}{d\mathbb{P}} = Z_\Theta(t).$$

This is captured in the following results. □

DEFINITION. Assume $W = (W(t))_{t\in\mathbb{T}}$ is standard Brownian motion and $\Theta = (\Theta(t))_{t\in\mathbb{T}}$ is some process in $\mathbb{L}^2(0,T)$ both on a filtered probability space with the usual conditions $(\Omega, \mathcal{F}, F, \mathbb{P})$. Then

the process $Z_\Theta = (Z_\Theta(t))_{t\in\mathbb{T}}$ defined as

$$Z_\Theta(t) = \exp\left(-\frac{1}{2}\int_0^t |\Theta(s)|^2\, ds - \int_0^t \Theta(s)\, dW(s)\right) \quad (\text{for } t \in [0,T])$$

is called the ***exponential martingale associated to*** Θ.

THEOREM 3.22 (Girsanov's theorem). *Assume $W = (W(t))_{t\in\mathbb{T}}$ is standard Brownian motion and $\Theta = (\Theta(t))_{t\in\mathbb{T}}$ is some process in $\mathbb{L}^2(0,T)$ both on a filtered probability space with the usual conditions $(\Omega, \mathcal{F}, F, \mathbb{P})$. Further, consider the exponential martingale $Z_\Theta = (Z_\Theta(t))_{t\in\mathbb{T}}$ defined as*

$$Z_\Theta(t) = \exp\left(-\frac{1}{2}\int_0^t |\Theta(s)|^2\, ds - \int_0^t \Theta(s)\, dW(s)\right) \quad (\textit{for } t \in [0,T]). \tag{3.7.1}$$

Then $\mathbb{Q}_t$ defined implicitly as

$$\frac{d\mathbb{Q}_t}{d\mathbb{P}} = Z_\Theta(t) \quad \textit{for each } t \in \mathbb{T}$$

is an equivalent probability measure on $(\Omega, \mathcal{F})$.

Additionally, the process $\hat{W} = \left(\hat{W}(t)\right)_{t\in\mathbb{T}}$ defined as

$$\hat{W}(t) = \int_0^t \Theta(s)\, ds + W(t) \quad (\textit{for } t \in [0,T])$$

is standard Brownian motion under the probability measure $\mathbb{Q} := \mathbb{Q}_T$ provided

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^t |\Theta(s)|^2 Z^2(s)\, ds\right] < \infty.$$

PROOF. See Shreve [**76**, pg. 212-214]. □

3.8. Stochastic Differential Equations

Given the definitions of an Itô integral and a stochastic differential, it is possible to formally define SDE. As before, $W = (W(t))_{t\in[0,T]}$ is an adapted, standard Brownian motion on on a filtered probably space $(\Omega, \mathcal{F}, F, \mathbb{P})$ with the usual conditions and the time index set is the bounded interval $\mathbb{T} = [0,T]$ where $T < \infty$.

I shall begin with a one-dimensional SDE.

3.8.1. One-Dimensional SDE.

DEFINITION. Let $X = (X(t))_{t\in[0,T]}$ be an admissible stochastic process. X is said to be a ***solution of the one-dimensional (Itô) stochastic differential equation***

$$\begin{cases} dX(t) &= \mu(X(t), t)\, dt + \sigma(X(t), t)\, dW(t) \\ X(0) &= x_0 \end{cases} \tag{3.8.1}$$

provided that

(1) X is progressively measurable;
(2) $\mu \in L^1(\mathbb{R} \times [0,T];\mathbb{R}), g \in L^2(\mathbb{R} \times [0,T];\mathbb{R})$; and
(3) $X(t) = X(0) + \int_0^t \mu(X(s),s)\,ds + \int_0^t \sigma(X(s),s)\,dW(s)$ a.s. $(t \in \mathbb{T})$.

REMARK. This is form of SDE is only dependent on the state of process at each time t and not on the path up to that time. While this is sufficient for this paper, a number of over-the-counter derivatives such as caps, floors, and knock-outs are, in fact, path-dependent. □

DEFINITION. An admissible stochastic process $X = (X(t))_{t\in[0,T]}$ is said to be an (***Itô***) ***diffusion*** if it is the solution to a SDE of the form 3.8.1. The drift and diffusion coefficients (μ, σ) are called the ***characteristics*** of the diffusion.

As with any other class of differential equations, the existence and uniqueness of solutions to problems of the form 3.8.1 are not assured. However, with some minimal assumptions concerning the regularity of the coefficients and some minor conditions on the initial data, unique solutions do, in fact, exist as set forth in the following result which is proved in Evans [**30**, pg. 92-96].

THEOREM 3.23. *A SDE of the form of 3.8.1 has a unique solution, if, for any $t \in [0,T]$ and $x, y \in \mathbb{R}$, $X = (X(t))_{t\in[0,T]}$ satisfies the following conditions known as the* ***Itô conditions***

(1) $|f(x,t) - f(y,t)| + |g(x,t) - g(y,t)| \leq C|x-y|$ *for some constant C;* (*uniform Lipschitz continuity in the state space*)
(2) $|f(x,t)| + |g(x,t)| \leq C(1+|x|)$ *for some constant C;* (*linear growth*).

REMARK. By uniqueness, I mean that if $X = (X(t))_{t\in[0,T]}$ and $Y = (Y(t))_{t\in[0,T]}$ are in $\mathbb{L}^2(0,T)$ and are solutions to an SDE of the form of 3.8.1, then

$$\mathbb{P}(X(t) = Y(t) \text{ for all } t \in [0,T]) = 1.$$

□

EXAMPLE (Log-normal process). One objection to modeling price movements using standard Brownian motion is that, as a pure Gaussian process, it admits negative prices which do not generally occur in practice. To deal with this possibility, a ***log-normal*** process $X = (X(t))_{t\in[0,T]}$ is often used to ensure positive prices. Here, the diffusion coefficients are separable in time and space and are of the form:

$$\mu(X(t),t) = \mu(t)X(t) \text{ and } \sigma(X(t),t) = \sigma(t)X(t)$$

where μ and σ are positive constants and it satisfies the following SDE

$$\begin{cases} dX(t) &= \mu(t)X(t)\,dt + \sigma(t)X(t)\,dW(t) \\ X(0) &= x_0. \end{cases} \tag{3.8.2}$$

To see that geometric Brownian motion is positive, ignoring the fact this is a stochastic process for the moment and assuming that the drift and diffusion are constants, the left-hand side of (3.8.2) suggests

$$\frac{dX(t)}{X(t)} = d(\ln X(t)).$$

However, since X is a stochastic process, the right-hand side can be computed by applying Itô's formula

$$d\left(\ln X\left(t\right)\right)=\frac{dX\left(t\right)}{X\left(t\right)}-\frac{1}{2X^{2}\left(t\right)}dX\left(t\right)dX\left(t\right).$$

Since $dX\left(t\right)dX\left(t\right)=\sigma^{2}X^{2}\left(t\right)dt$

$$d\left(\ln X\left(t\right)\right)=\frac{dX\left(t\right)}{X\left(t\right)}-\frac{\sigma^{2}}{2}dt.$$

Then by (3.8.2)

$$\begin{aligned} d\left(\ln X\left(t\right)\right) &= \mu dt+\sigma dW\left(t\right)-\frac{\sigma^{2}}{2}dt \\ &= \left(\mu-\frac{\sigma^{2}}{2}\right)dt+\sigma dW\left(t\right). \end{aligned}$$

Expressing this in integral form

$$\begin{aligned} \ln X\left(t\right) &= \ln X\left(0\right)+\int_{0}^{t}\left(\mu-\frac{\sigma^{2}}{2}\right)dt+\int_{0}^{t}\sigma dW\left(t\right), \text{ or} \\ X\left(t\right) &= x_{0}\exp\left(\left(\mu-\frac{\sigma^{2}}{2}\right)t+\sigma W\left(t\right)\right). \end{aligned}$$

Hence, $X\left(t\right)$ is positive for all $t\in\left[0,T\right]$. □

Definition. A stochastic process $X=\left(X\left(t\right)\right)_{t\in[0,T]}$ is said to be ***geometric Brownian motion*** if it satisfies the following SDE

$$\begin{cases} dX\left(t\right) &= \mu\left(t\right)X\left(t\right)dt+\sigma\left(t\right)X\left(t\right)dW\left(t\right) \\ X\left(0\right) &= x_{0} \end{cases}$$

and is of the form

$$X\left(t\right)=x_{0}\exp\left(\left(\mu-\frac{\sigma^{2}}{2}\right)t+\sigma W\left(t\right)\right).$$

Example (Langevin's equation). Langevin's equation describes Brownian motion with constant friction. It is of the form:

$$\begin{cases} dX\left(t\right) &= -\mu X\left(t\right)dt+\sigma dW\left(t\right) \\ X\left(0\right) &= x_{0}. \end{cases}$$

where $\mu>0$ is the coefficient of friction and $\sigma>0$ the constant diffusion coefficient. Here, the process $X=\left(X\left(t\right)\right)_{t\in[0,T]}$ is interpreted as the velocity of a particle driven by the Brownian motion. □

Example (Ornstein-Uhlenbeck process). Ornstein-Uhlenbeck processes are one of the most common forms of Itô diffusions used in continuous-time finance. Such a process has coefficients of the form:

$$\mu\left(X\left(t\right),t\right)=\mu\left(X\left(t\right)\right)\text{ and }\sigma\left(X\left(t\right),t\right)=\sigma$$

where μ is a non-negative, Lipschitz continuous function and σ is a positive constant. Hence, it satisfies the Itô conditions (*cf.* on page 131) is the unique solution the following SDE

$$\begin{cases} dX(t) & = \mu(X(t))\,dt + \sigma dW(t) \\ X(0) & = x_0. \end{cases}$$

In many cases, the drift term is *mean-reverting* of the form,

$$\mu(X(t)) = \alpha(m - X(t)),$$

where $m > 0$ represents the long-term, invariant mean of the process and $\alpha > 0$ is a real-valued constant known as the ***rate of mean reversion***. □

Example (Hull-White process). A Hull-White process is a special case of an Ornstein-Uhlenbeck process that is commonly used in continuous-time finance to model interest rates. It is mean-reverting with coefficients of the form

$$\mu(X(t), t) = (a(t) - b(t)X(t)) \text{ and } \sigma(X(t), t) = \sigma(t)$$

where $a(\cdot), b(\cdot),$ and $\sigma(\cdot)$are deterministic, positive, real-valued functions. Again, it satisfies the Itô conditions and is the unique solution to the following SDE

$$\begin{cases} dX(t) & = (a(t) - b(t)X(t))\,dt + \sigma dW(t) \\ X(0) & = x_0. \end{cases}$$

□

Example (Constant elasticity of variance process). Constant elasticity of variance (CEV) processes are another common form of dynamics used in continuous-time finance. In its most general form, it has coefficients of the form

$$\mu(X(t), t) = \mu(t)X(t) \text{ and } \sigma(X(t), t) = \sigma X^{\gamma+1}(t)$$

where $-1 < \gamma < 0$, $\mu(\cdot)$ is deterministic, positive, real-valued function, and σ is some positive constant. It also satisfies the Itô conditions and is the unique solution to the following SDE

$$\begin{cases} dX(t) & = \mu X(t)\,dt + \sigma X^{\gamma+1}(t)\,dW(t) \\ X(0) & = x_0. \end{cases}$$

Remark. Note that in some of the literature, there are no bounds on the value of γ. Here, I have chosen to follow the convention of the original authors. □

Example (Cox-Ingersoll-Ross process). A Cox-Ingersoll-Ross (CIR) process is a special case of a CEV process that is commonly used in dynamic asset pricing models. It is a mean-reverting process with coefficients of the form

$$\mu(X(t), t) = \alpha(m - X(t)) \text{ and } \sigma(X(t), t) = X^{1/2}(t)$$

where $m > 0$ is the long-term, invariant mean of the process, $\alpha > 0$ is the rate of mean reversion, and σ is a positive constant. It satisfies the following SDE

$$\begin{cases} dX(t) & = \alpha(m - X(t))\,dt + \sigma X^{1/2}(t)\,dW(t) \\ X(0) & = x_0. \end{cases}$$

□

3.8.2. Higher Dimensional SDE. SDE can also be stated in n-dimensions as follows. Assume $W = (W^1(t), \ldots, W^m(t))_{t\in[0,T]}$ is standard m-dimensional Brownian motion and that all stochastic process are $\mathbb{R}^n$-valued and admissible.

DEFINITION. Let $X = (X^1(t), \ldots, X^n(t))_{t\in\mathbb{T}}$ be an admissible, $\mathbb{R}^n$-valued stochastic process. X is said to be a ***solution of the -dimensional Itô stochastic differential equation***

$$\begin{cases} dX(t) & = \mu(X(t), t)\,dt + \sigma(X(t), t)\,dX(t) \\ X(0) & = x_0 \in \mathbb{R}^n. \end{cases} \tag{3.8.3}$$

provided that

(1) $\mu \in \mathbb{L}_n^1(0,T), \sigma \in \mathbb{L}_{n\times m}^2(0,T)$; and
(2) $X(t) = X(0) + \int_0^t \mu(X(s), s)\,ds + \int_0^t \Lambda(X(s), s)\,dW(s)$ a.s. $(t \in \mathbb{T})$.

The next result is an extension of the one-dimensional case and is proved in Evans [**30**, pg. 92-96].

THEOREM 3.24. *For A SDE of the form of 3.8.1has a unique solution, if, for any $t \in [0,T]$ and $x, y \in \mathbb{R}$, X satisfies the following conditions known as the* ***Itô conditions***

(1) $|\mu(x,t) - \mu(y,t)| + |\sigma(x,t) - \sigma(y,t)| \leq C|x - y|$ *for some e constant C;* (*uniform Lipschitz continuity in state space*)
(2) $|\mu(x,t)| + |\sigma(x,t)| \leq C(1 + |x|)$ *for some constant C;* (*linear growth*).

CHAPTER 4

Connections with Partial Differential Equations

There are many deep connections between probability theory and PDE. From the point of view of this paper, I will be concerned primarily with the connection between Itô diffusions and linear parabolic equations. The principal result of this chapter is the *Feynman-Kac formula* which provides the basis of linking PDE and martingale pricing methods that are commonly used to price financial instruments. For example, it will turn out that under the proper conditions, the Feynman-Kac formula produces the *Black-Scholes PDE* which describes the time-evolution of the price of a specific financial instrument known as an *option*.

I will motivate this discussion using the well-known connection between the Brownian motion and the heat equation. For this chapter, assume that $(\Omega, \mathcal{F}, F, \mathbb{P})$ is some filtered probability space with the usual conditions and that all stochastic processes on $(\Omega, \mathcal{F}, F, \mathbb{P})$ are $\mathbb{R}^n$-valued and admissible. Also, assume that the time index set is a bounded interval $\mathbb{T} = [0, T]$ with $T < \infty$.

NOTATION. As before, I will abuse the notation by not differentiating between scalar- and vector-valued stochastic processes other than in the definition of the process itself. In other words, the notation $X = (X(t))_{t \in \mathbb{T}}$ refers both to a real-valued process (*i.e.* $X : \Omega \times \mathbb{T} \to \mathbb{R}$) and a $\mathbb{R}^n$-valued process (*i.e.* $X : \Omega^n \times \mathbb{T} \to \mathbb{R}^n$). □

4.1. Brownian Motion and the Heat Equation

First, let me show that the solution to the heat equation describes the evolution of the density of Brownian motion over time. Recall that the initial-valued, homogeneous heat equation is a linear, second-order parabolic PDE of the form

$$\begin{cases} u_t - \frac{1}{2}\triangle u = 0 & \text{in } \mathbb{R}^n \times (0, \infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{4.1.1}$$

where $f, g : \mathbb{R}^n \to \mathbb{R}$ are specified functions and $u : \mathbb{R}^n \to \mathbb{R}$ is the unknown function (*cf.* [**29**, Ch. 2.3]). This equation describes the diffusion (*i.e.* the evolution of the density of heat in some region over time given some initial heat density. The existence of solutions to such problems is well-known and is discussed at length in Evans [**29**, ch. 2.3].

For simplicity's sake, in the following discussion I will assume that $n = 1$. Then (4.1.1) becomes

$$\begin{cases} u_t - \frac{1}{2}u_{xx} = 0 & \text{in } \mathbb{R} \times (0, \infty) \\ u = g & \text{on } \mathbb{R} \times \{t = 0\}. \end{cases} \tag{4.1.2}$$

It is well known that if $g \in C(\mathbb{R}) \cap L^\infty(\mathbb{R})$, the solution of (4.1.2) is given by

$$u(x,t) = \int_{-\infty}^{\infty} \Phi(y - x, t)\, g(y)\, dy$$

where for any $x \in \mathbb{R}$, the function

$$\Phi(x,t) = \begin{cases} \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}, & t > 0 \\ 0 & t < 0 \end{cases} \tag{4.1.3}$$

is called the ***fundamental solution*** (*a.k.a.* the ***heat kernel***) of the heat equation (*cf.* [**28**, ch. 2.3]). In other words, the solution to the one dimensional heat equation is

$$u(x,t) = \begin{cases} \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} g(y)\, e^{-\frac{(x-y)^2}{2t}} dy, & x \in \mathbb{R}, t > 0 \\ g(x), & x \in \mathbb{R}, t = 0. \end{cases} \tag{4.1.4}$$

In particular, the fundamental solution is the solution to the heat with initial condition that all of the heat is concentrated at a single point or

$$\begin{cases} \Phi_t - \frac{1}{2}\Phi_{xx} = 0 & \text{in } \mathbb{R} \times (0, \infty) \\ \Phi = \delta_0 & \text{on } \mathbb{R} \times \{t = 0\}. \end{cases} \tag{4.1.5}$$

Remark. If $g \in C(\mathbb{R}) \cap L^\infty(\mathbb{R})$, non-negative, and not identically zero, then

$$u(x,t) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} g(y)\, e^{-\frac{(x-y)^2}{2t}} dy > 0 \quad \text{for all } x \in \mathbb{R} \text{ and all } t > 0.$$

Importantly, even if all the heat in a given system is initially concentrated at a single point at time zero (*i.e.* a Dirac distribution), it is distributed along the entire domain at all other times no matter how small. This suggests that the heat diffuses with an *infinite propagation speed.* □

Now, Recall (*cf.* example on page 94) that the transition probability density functions of Brownian motion centered at $x \in \mathbb{R}$ are of the form

$$\rho(x,t) = \begin{cases} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}, & t > 0 \\ \delta_y(x), & t = 0 \end{cases} \tag{4.1.6}$$

Comparing equations (4.1.3) and (4.1.6) for any $t \in (0, T]$ are the identical when the Brownian motion is centered at zero. Hence, the fundamental solution of the heat equation determines the

family of transition probability density functions for standard Brownian motion. Moreover, for any $x \in \mathbb{R}$

$$\begin{aligned}\mathbb{E}^x\left[g\left(W\left(t\right)\right)\right] &= \begin{cases}\frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} g\left(y\right)e^{-\frac{(x-y)^2}{2t}}dy, & t>0\\ g\left(x\right), & t=0.\end{cases}\\ &= u\left(x,t\right)\end{aligned}$$

In other words, the solution to the heat equation at the point (x,t) is the "average" of all the Brownian paths starting on the x-axis (*i.e.* $(y,0)$ for some $y \in \mathbb{R}$) that ended up at the point (x,t).

Conversely, Brownian motion can be used to construct the solution to the heat equation. Here, I will generally follow Chorin and Hald (*cf.* [**15**, ch. 3.3]).

Specifically, suppose that I want to know $u\left(\hat{x},\hat{t}\right)$ for a specific $\hat{x}$ and $\hat{t}$ given the fact that u is the solution to the heat equation (4.1.2). The first step is to approximate the heat equation (4.1.2) by a *finite difference equation*.

Therefore, create a stencil $S = \{(ih,nk)\,|i,j \in \mathbb{Z}\}$ on $\mathbb{R}^2$ where h and k are positive constants representing the step size for the space and time variables, respectively. Further, discretize the space and time variables $x = ih$ and $t = nk$ and define the discrete function v such that v agrees with u on all points in the stencil S, or

$$v_i^n := u\left(ih,nk\right).$$

In particular, the initial conditions (*i.e.* $t = nk = 0$) are given by

$$v_i^0 = g\left(ih\right) \quad (\text{for any } i \in \mathbb{Z}).$$

Using v, I can approximate the partial derivatives of u by finite differences using the following *finite difference operators*

$$\begin{aligned}u_t \approx D_t^+ v_i^n &:= \frac{v_i^{n+1}-v_i^n}{k};\\ u_x \approx D_x^+ v_i^n &:= \frac{v_{i+1}^n - v_i^n}{h};\\ u_x \approx D_x^- v_i^n &:= \frac{v_i^n - v_{i-1}^n}{h}; \quad \text{and}\\ u_{xx} \approx D_x^+ D_x^- v_i^n &= \frac{v_{i+1}^j - 2v_i^j + v_{i-1}^j}{h^2}.\end{aligned}$$

Substituting these expressions into the heat equation (4.1.2) results in the following *finite difference equation*,

$$\frac{v_i^{n+1}-v_i^n}{k} = \frac{v_{i+1}^n - 2v_i^n + v_{i-1}^n}{2h^2}.$$

Rearranging terms,

$$v_i^{n+1} = (1-2\lambda)\, v_i^n + \lambda \left(v_{i+1}^n + v_{i-1}^n \right) \tag{4.1.7}$$

where

$$\lambda = \frac{1}{2}\frac{k}{h^2}.$$

Choosing $\lambda = 1/2$ to remove the first term on the right-hand side of (4.1.7), then

$$v_i^{n+1} = \frac{v_{i+1}^n + v_{i-1}^n}{2}.$$

In other words, the value of v at the next time step $n+1$ is the average of values of the adjacent space steps at the previous time step n. This is a recursive relationship which can be made explicit on the initial conditions as follows

$$\begin{aligned}
v_i^{n+1} &= \frac{1}{2}\left(\frac{v_{i+2}^{n-1} + v_i^{n-1}}{2} + \frac{v_i^{n-1} + v_{i-2}^{n-1}}{2} \right) \\
&= \frac{1}{4}\left(v_{i+2}^{n-1} + 2v_i^{n-1} + \frac{1}{4} v_{i-2}^{n-1} \right) \\
&\vdots \\
&= \sum_{j=0}^{n} C_{j,n} v_{i+2j-n}^0 \\
&= \sum_{j=0}^{n} C_{j,n} g\left((i+2j-n)\, h \right) \\
&= \sum_{j=0}^{n} C_{j,n} g\left(ih - (n-2j)\, h \right) \qquad (4.1.8)
\end{aligned}$$

where

$$C_{j,n} = \frac{1}{2^n} \left(\begin{array}{c} n \\ j \end{array} \right).$$

Note that as j ranges from zero to n, $i - n + 2j$ ranges from $i - n$ to $i + n$. Here, the interval $(i-n, i+n)$ on the represents the *numerical domain of dependence* which consists of all the values on the y-axis that impact the calculation of the v at the point (ih, nk), or

$$v_j^0 = g\left(jh\right) \quad \left(\text{for } i - n \le j \le i + n\right).$$

In other words, only the initial values on this interval affects the value of v at the point (ih, nk) and the amount of that affect is weighted by the factor

$$C_{j,n} = \frac{1}{2^n} \left(\begin{array}{c} n \\ j \end{array} \right).$$

Upon inspection, $C_{j,n}$ is a binomial distribution with equal probability of success or failure (*cf.* example on page 14). This leads to the probabilistic interpretation that $C_{j,n}$ is the probability of a *backwards* symmetric random walk starting at the point $(0, nk)$ and ending up at the point $(jh, 0)$.

To see this, recall that a symmetric random walk $S = (S_m)_{m\in\mathbb{N}}$ is generated by a family of independent normal random variables $\{X_m\}_{m\in\mathbb{N}}$ such that

$$S_m = S_{m-1} + X_m \quad (\text{for } 1 \leq m \leq n)$$

(*cf.* example on page 63). In this case, we are considering a backwards walk starting at the grid point $(0, nk)$ (*i.e.* $S_0 = 0$) with

$$X_m = \begin{cases} h, & \text{with probability } \frac{1}{2} \\ -h, & \text{with probability } \frac{1}{2}. \end{cases}$$

Therefore, if the process starts at the grid point $(0, nk)$, then after n moves, $S_n \in [-nh, nh]$. Moreover, I can compute that probability the random walk will have a specific value in the interval $[-nh, nh]$, say $(-n + 2j)\, h$ for some $j \in [0, n]$, by dividing the number of paths from $(0, nk)$ to $((-n + 2j)\, h, 0)$ by the total number of paths from $(0, nk)$ that intersect $[-nh, nh]$. In other words, for $0 \leq j \leq n$

$$\mathbb{P}\left(S_n = (-n + 2j)\, h\right) = \frac{1}{2^n} \begin{pmatrix} n \\ j \end{pmatrix} = C_{j,n} \tag{4.1.9}$$

where $\begin{pmatrix} n \\ j \end{pmatrix}$ is the number of paths from $(0, nk)$ to $((-n + 2j)\, h, 0)$ and $1/2^n$ is the total number of paths that intersect $[-nh, nh]$.

Hence, by (4.1.8)

$$\begin{aligned} v_i^n &= \sum_{j=0}^{n} C_{j,n} g\left(ih - (n - 2j)\, h\right) \\ &= \sum_{j=0}^{n} g\left(ih - (n - 2j)\, h\right) \mathbb{P}\left(S_n = (n - 2j)\, h\right) \\ &= \mathbb{E}\left[g\left(ih - S_n\right)\right]. \end{aligned}$$

In order words, v_i^n is the average of all paths ending up at the point (ih, nk) weighted by their value at time zero (*i.e.* $g\left(jh\right)$ for $j \in [-n, n]$).

Now, using the central limit theorem (*cf.* [**23**, pg. 124])

$$\lim_{n\to\infty} S_n \sim N\left(0, nk\right)$$

Also recall, that I set

$$\lambda = \frac{k}{2h^2} = \frac{1}{2}.$$

Hence, $h^2 = k$ and consequently $nk = nh^2$. Therefore, for a specific point on the stencil $\tilde{x} := ih$ and $\tilde{t} := nk = nh^2$,

$$\begin{aligned} u\left(\tilde{x}, \tilde{t}\right) &= v_i^n \\ &= \mathbb{E}\left[g\left(\tilde{x} - S_n\right)\right]. \end{aligned}$$

Now, in order to stay at the point (ih, nk) as $n \to \infty$, h must go to zero. Thus, for large n and small h, S_n is a Gaussian random variable with mean 0 and variance $\hat{t}$ and

$$\mathbb{E}\left[g\left(\tilde{x} - S_n\right)\right] \sim \frac{1}{\sqrt{2\pi\hat{t}}} \int_{-\infty}^{\infty} g\left(\tilde{x} - y\right) e^{-\frac{y^2}{2\hat{t}}} dy.$$

Since this true for any point on the stencil, then with a change of variables and letting $n \to \infty$

$$u\left(\tilde{x}, \tilde{t}\right) = \mathbb{E}\left[g\left(\tilde{x}\right)\right] = \frac{1}{\sqrt{2\pi\tilde{t}}} \int_{-\infty}^{\infty} g\left(y\right) e^{-\frac{(y-\tilde{x})^2}{2\tilde{t}}} dy.$$

Since I assumed that $g \in C\left(\mathbb{R}\right) \cap L^{\infty}\left(\mathbb{R}\right)$, the right-hand side is well-defined and continuous for all $x, t \in \mathbb{R}$. Hence,

$$u\left(x, t\right) = \begin{cases} \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} g\left(y\right) e^{-\frac{(y-x)^2}{2t}} dy & t > 0 \\ g\left(x\right) & t = 0. \end{cases}$$

u is the solution to the initial-valued heat-equation.

4.2. Semigroup Theory of Transition Probability Operators

Semigroup theory of linear operators provides an elegant framework that establishes a formal relationship between Markov processes and certain classes of PDE. Specifically, a family of bounded operators based on the transition probability functions of a Markov process form a linear semigroup whose *infinitesimal generator* is a linear, second order differential operator. Moreover, this differential operator is unique allowing us to uniquely characterize a Markov process by a particular PDE.

Generally, I follow Evans [**29**, ch. 7.4] where any missing proofs may be found.

Motivation. Recall that I introduced transition probability functions in the context of Markov processes (*cf.* Section 2.4.2). In particular, for a given stationary, $\mathbb{R}^n$-valued admissible Markov process $X = \left(X\left(t\right)\right)_{t \in [0,T]}$ with transition probability functions $\left\{p_t\right\}_{t \in [0,T]}$, the probability that the process will be in the target set $B \in \mathcal{B}$ at time t after starting at $x \in \mathbb{R}^n$ at time zero is given by

$$p_t\left(x, B\right) = \mathbb{P}\left(X\left(t\right) \in B \mid X\left(0\right) = x\right).$$

Further, one of the properties of such functions is that they satisfied the Chapman-Kolmogorov equation (*cf.* Theorem 2.25), or, for $t, s \in [0, T]$ such that $t + s \in [0, T]$

$$p_{t+s}(x, B) = \int_{\mathbb{R}^n} p_t\left(y, B\right) p_s\left(x, dy\right) \quad \left(\text{for any } B \in \mathcal{B}\right).$$

Lastly, for $f \in C(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$, I can express the expected value of $f(X(t))$ at any time $t \in [0,T]$ conditioned on $X(0) = x \in \mathbb{R}^n$ in terms of the transition probability functions

$$\mathbb{E}^x [f(X(t))] = \int_{\mathbb{R}} f(y)\, p_t(x, dy)$$

(*cf.* (2.4.6)).

This suggests that for a given Markov process $X = (X(t))_{t\in[0,T]}$ with the transition probability functions $\{p_t\}_{t\in[0,T]}$, I can define a family of operators on the space $C(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ in terms of transition probability functions such that for any $t \in [0,T]$ and any fixed $x \in \mathbb{R}^n$

$$\begin{aligned} (\mathrm{P}_t f)(x) &:= \int_{\mathbb{R}} f(y)\, p_t(x, dy) \\ &= \mathbb{E}^x [f(X(t))]. \end{aligned} \tag{4.2.1}$$

where $f \in C(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$.

Remark. Although somewhat confusing, it is common to write the left-hand side of (4.2.1) as $\mathrm{P}_t f(x)$. However, it is important to note that the operators $\{\mathrm{P}_t\}_{t\in[0,T]}$ are acting on functions in $C(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ and not on points in $\mathbb{R}^n$. □

Remark. Also note that $(\mathrm{P}_t f)(\cdot)$ is a random variable. □

Given this definition, the transition probability operators also satisfy the Chapman-Kolmogorov equation, since

$$\begin{aligned} \mathrm{P}_{t+s} f(x) &= \mathbb{E}^x [f(X(t+s))]. \\ &= \int_{\mathbb{R}} f(y)\, p_{t+s}(x, dy) \\ &= \int_{\mathbb{R}} f(y) \int_{\mathbb{R}} p_t(z, dy)\, p_s(x, dz) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(y)\, p_t(z, dy) \right) p_s(x, dz) \\ &= \int_{\mathbb{R}} \mathbb{E}^z [f(X(t))]\, p_s(x, dz). \end{aligned}$$

Letting

$$g(z) = \mathbb{E}^z [f(X(t))] = (\mathrm{P}_t f)(z).$$

Then

$$\begin{aligned} \mathrm{P_{t+s}} f(x) &= \int_{\mathbb{R}} g(z)\, p_s(x, dz) \\ &= \mathbb{E}\left[g(X(s)) \mid X(0) = x\right] \\ &= \mathrm{P}_s g(x) \\ &= (\mathrm{P_s} \circ \mathrm{P_t})\, f(x). \end{aligned}$$

Additionally,

$$\mathrm{P}_0 f(x) = \mathbb{E}^x\left[f(X(0))\right] = f(x) \Rightarrow \mathrm{P}_0 = \mathrm{I}.$$

Lastly, the mapping $t \mapsto \mathrm{P_t} f$ is obviously continuous since $f \in C(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$. This suggests that the family of transition probability operators of a Markov process form a semigroup under the composition operator "$\circ$" with the following properties:

(1) $\mathrm{P_{t+s}} = \mathrm{P_t P_s} \quad t, s \in [0, T]$ with $t + s \in [0, T]$;
(2) $\mathrm{P}_0 = \mathrm{I}$; and
(3) $t \mapsto \mathrm{P_t} f$.

This is formalized below. □

DEFINITION. Let S be an arbitrary Banach space with norm, $\|\cdot\|$. A (one parameter) family of bounded (in the operator norm) linear operators $\{\mathrm{T_t} : S \to S\}_{t \in \mathbb{T}}$ is called a ***semigroup*** if, for any $f \in S$, the mapping $t \to \mathrm{T}_t f$ is continuous from $[0, T]$ into S, and for any $s, t \in [0, T]$ such that $s + t \in [0, T]$,

$$\mathrm{T}_{s+t}(\,\cdot\,) := \mathrm{T}_s \mathrm{T}_t(\,\cdot\,) := \mathrm{T}_s \circ \mathrm{T}_t(\,\cdot\,)$$

with $\mathrm{T}_0 = \mathrm{I}$. A semigroup is said to be ***strongly continuous*** if, for a fixed $t_0 \in [0, T]$ and any $f \in S$

$$\lim_{t \downarrow t_0} \|\mathrm{T}_t f - \mathrm{T}_{t_0} f\| = 0. \tag{4.2.2}$$

In particular,

$$\lim_{t \downarrow 0} \|\mathrm{T}_t f - f\| = 0.$$

DEFINITION. A semigroup $\{\mathrm{T_t}\}_{t \in [0,T]}$ on some Banach space S is called a ***contraction semigroup*** if, for any $f \in S$,

$$\|\mathrm{T}_t\|_{op} \leq 1 \quad (t \in [0, T])$$

where

$$\|\mathrm{T}_t\|_{op} = \sup\left\{\|\mathrm{T}_t f\|_S \mid \|f\|_S \leq 1\right\}$$

The next result follows directly from the definitions.

THEOREM 4.1. *If* $\{\mathrm{T_t}\}_{t\in[0,T]}$ *is a contraction semigroup on some Banach* S*, then for any* $f \in S$*,*

$$\|\mathrm{T}_t f\| \leq \|f\| \quad (t \in [0,T]).$$

EXAMPLE. The powers of the transition matrix of a time-homogeneous Markov chain form a contraction semigroup. To see this, recall that given a finite-state Markov chain with the transition matrix P and initial distribution vector $\boldsymbol{\phi}$,

$$\mathbb{P}(X_m = x_j) = (\mathrm{P}^m \boldsymbol{\phi})^j$$

(*cf.* Theorem 2.23). Further, I showed that the powers of the transition matrix $\{\mathrm{P}^m\}$ possess the semigroup property and $\sum_{j\in S} (\mathrm{P}^m)^{i,j} = 1$ for each $i \in S$ (Theorem 2.25 and Theorem 2.22, respectively). Then, for any finite m

$$\|\mathrm{P}^m\|_{op} = \sup\{|\mathrm{P}^m \boldsymbol{\phi}| \mid |\boldsymbol{\phi}| \leq 1\}.$$

However,

$$\begin{aligned} \|\mathrm{P}^m \boldsymbol{\phi}\|_{\mathbb{R}^\infty} &= (\mathrm{P}^m \boldsymbol{\phi})^{\mathrm{T}} (\mathrm{P}^m \boldsymbol{\phi}). \\ &= \boldsymbol{\phi}^{\mathrm{T}} \left((\mathrm{P}^m)^{\mathrm{T}} \mathrm{P}^m \right) \boldsymbol{\phi} \\ &\leq \boldsymbol{\phi}^{\mathrm{T}} \cdot \boldsymbol{\phi} \\ &\leq 1. \end{aligned}$$

Hence, $\{\mathrm{P}^m\}$ form a contraction semigroup. □

DEFINITION. A strongly continuous, contraction semigroup of operators on $\mathbb{R}$ is called a ***Feller semigroup***.

THEOREM 4.2. *Given a real-valued Markov process* $X = (X(t))_{t\in[0,T]}$ *with transition probability functions* $\{p_t\}_{t\in[0,T]}$*, the family of transition probability operators* $\{\mathrm{P}_t\}_{t\in[0,T]}$ *form a Feller semigroup where, for any* $t \in [0,T]$ *and* $f \in C(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$

$$\begin{aligned} \mathrm{P}_t f(x) &= \int_{\mathbb{R}} f(y)\, p_t(x, dy) \\ &= \mathbb{E}^x [f(X(t))] \end{aligned}$$

and

$$\mathrm{P}_0 f(x) = f(x).$$

PROOF. Let $X = (X(t))_{t\in[0,T]}$ be a real-valued Markov process with transition probability functions $\{p_t\}_{t\in[0,T]}$ and let $f \in C(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$. Earlier, I showed that the $\{\mathrm{P}_t\}_{t\in[0,T]}$ form a semigroup (*cf.* pg. 142). As such, I need only show that the collection is strongly continuous and contracting.

Now, since f is continuous and bounded

$$\begin{aligned}\lim_{t\downarrow t_0}\left\|\mathrm{P}_t f-\mathrm{P}_{t_0} f\right\|_{L^\infty(\mathbb{R}^n)} &= \lim_{t\downarrow t_0}\left\|\mathbb{E}^x\left[f\left(X\left(t\right)\right)\right]-\mathbb{E}^x\left[f\left(X\left(t_0\right)\right)\right]\right\|_{L^\infty(\mathbb{R}^n)}\\ &= \lim_{t\downarrow t_0}\left\|\mathbb{E}^x\left[f\left(X\left(t\right)\right)-f\left(X\left(t_0\right)\right)\right]\right\|_{L^\infty(\mathbb{R}^n)}=0.\end{aligned}$$

Hence, $\left\{\mathrm{P}_t\right\}_{t\in[0,T]}$ is strongly continuous.

Additionally, for any $t\in[0,T]$,

$$\left\|\mathrm{P}_t\right\|_{op} = \sup\left\{\left\|\mathrm{P}_t f\right\|_{L^\infty(\mathbb{R}^n)} \mid \|f\|_{L^\infty(\mathbb{R}^n)}\leq 1\right\}.$$

Now,

$$\begin{aligned}\left\|\mathrm{P}_t f\right\|_{L^\infty(\mathbb{R})} &= \left\|\mathbb{E}^x\left[f\left(X\left(t\right)\right)\right]\right\|_{L^\infty(\mathbb{R}^n)}\\ &\leq \int_{\mathbb{R}}\left\|f\left(y\right)p_t\left(x,dy\right)\right\|_{L^\infty(\mathbb{R}^n)}\\ &\leq \|f\|_{L^\infty(\mathbb{R}^n)}\int_{\mathbb{R}}p_t\left(x,dy\right)\\ &\leq \|f\|_{L^\infty(\mathbb{R}^n)}\,p_t\left(x,\mathbb{R}^n\right).\end{aligned}$$

By definition, $p_t\left(x,\mathbb{R}^n\right)\leq 1$ (*cf.* definition on page 91). Thus

$$\left\|\mathrm{P}_t f\right\|_{L^\infty(\mathbb{R}^n)}\leq\|f\|_{L^\infty(\mathbb{R}^n)} \;\Rightarrow\; \left\|\mathrm{P}_t\right\|_{op}\leq 1$$

and $\left\{\mathrm{P}_t\right\}_{t\in[0,T]}$ is contracting. □

Under certain conditions, it is possible to find a linear operator which essentially acts as the time "derivative" of a transition probability operator.

Motivation. Recall that a particle moving through n-dimensional space can be model by first-order system of ODE of the form,

$$\begin{cases}\dot{\mathbf{x}}\left(t\right) &= \mathrm{A}\mathbf{x}\left(t\right)\quad(t>0)\\ \mathbf{x}\left(0\right) &= x_0\quad(x_0\in\mathbb{R}^n)\end{cases}\tag{4.2.3}$$

where "$\dot{}$"$=\dfrac{d}{dt}$ and $\mathbf{x}:[0,T]\to\mathbb{R}^n$ represents the location of the particle at time t, A is an $n\times n$ matrix representing the instantaneous velocity of the particle at time t, and the initial starting point is $x_0\in\mathbb{R}^n$. In other words, $\mathbf{x}$ represents the state of the system at any time t. Then by the *fundamental theorem for linear systems* (*cf.* [**67**, pg. 17-18]), the solution is

$$\mathbf{x}\left(t\right)=x_0e^{\mathrm{A}t}$$

where

$$e^{\mathrm{A}t}:=\sum_{k=0}^{\infty}\frac{\mathrm{A}^k t^k}{k!}.\tag{4.2.4}$$

Now, consider the function $\mathrm{T} : [0,T] \to \mathbb{R}^n$ where

$$\mathrm{T}(t) := x_0 e^{\mathrm{A}t} \quad \text{for any } t \in [0,T].$$

Clearly, T satisfies the Chapman-Kolmogorov equation since, for $s,t \in [0,T]$ such that $s+t \in [0,T]$

$$\begin{aligned} \mathrm{T}(t+s) &= e^{(t+s)\mathrm{A}} \\ &= e^{t\mathrm{A}} e^{s\mathrm{A}} \\ &= \mathrm{T}(t)\,\mathrm{T}(s). \end{aligned}$$

Moreover, by (4.2.3), T describes the evolution of the state of the particle, since

$$\begin{cases} \frac{d}{dt}\mathrm{T}(t) &= \mathrm{AT}(t) \quad (t \in \mathbb{T}) \\ \mathrm{T}(0) &= \mathrm{I}. \end{cases} \tag{4.2.5}$$

This can be extended to random variables as follows.

Let $X = (X(t))_{t\in[0,T]}$ be a Markov process with transition probability functions $\{p_t\}_{t\in[0,T]}$ which represents the state of a particle at each time $t \in [0,T]$ where $X(0) = x \in \mathbb{R}^n$. Then by Theorem 4.2, the collection of operators $\{\mathrm{P}_t\}_{t\in[0,T]}$ form a semigroup and

$$\mathrm{P}_t f(x) = \mathbb{E}^x [f(X(t))]$$

for some $f \in S$ where S is an arbitrary Banach space. In other words, $\mathrm{P}_t f(x)$ is the expected state of the particle at time $t \in [0,T]$ having started at $f(x)$. Then (3.1.1) becomes

$$\begin{cases} \frac{d}{dt}\mathrm{P}_t f &= A\mathrm{P}_t f \quad (t \in [0,T]) \\ \mathrm{P}_0 f &= f \end{cases} \tag{4.2.6}$$

since P_0 is assumed to be the identity. Here, A is some operator on S. Hence, if such an equation makes sense and a solution exists (for instance, if A is bounded),

$$\mathrm{P}_t f = e^{At} f = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k f \quad (t \in [0,T]). \tag{4.2.7}$$

In other words, the operator A "generates" the family of transition probability operators. This is formalized below. □

NOTATION. If (4.2.4) makes sense mathematically, e^{At} is the "symbol" representing the the family of operators of the form of (4.2.7). Such symbology is known as ***semigroup notation***. For more on this, see Tao [**84**, pg. 40-49]. □

DEFINITION. Let $\{\mathrm{T_t}\}_{t\in[0,T]}$ be a Feller semigroup on a Banach space S and define the operator $A : S \to S$ as:

$$Af = \lim_{t\downarrow 0} \frac{\mathrm{T}_t f - f}{t}$$

for some arbitrary $f \in S$.

Then A, if it exists, is called the (***infinitesimal***) ***generator*** of the semigroup $\{\mathrm{T}_t\}_{t\in[0,T]}$ where the ***domain of*** A is given by

$$\mathcal{D}(A) = \left\{ f \in S \mid \lim_{t\downarrow 0} \frac{\mathrm{T}_t f - f}{t} \text{ exists in } S \right\}.$$

The following theorem, which is proved in Evans [**29**, pg. 435-436], shows that infinitesimal generators have many desirable properties.

THEOREM 4.3 (Differential properties of semigroups). *Let $\{\mathrm{T}_t\}_{t\in[0,T]}$ be a Feller semigroup on a Banach space S. Then for any $f \in \mathcal{D}(A)$ and $t \in [0,T]$,*

(1) $\mathrm{T}_t f \in \mathcal{D}(A)$,
(2) $A\mathrm{T}_t f = \mathrm{T}_t A f$,
(3) *The mapping $t \mapsto \mathrm{T}_t f$ is differentiable for $t > 0$, and*
(4) $\frac{d}{dt}\mathrm{T}_t f = A\mathrm{T}_t f$.

However, for an infinitesimal generator to be of any use, its domain must be non-empty. Fortunately, the next result, which is proved in Evans [**29**, pg. 436-437] shows, in fact, the domain is dense.

THEOREM 4.4 (General properties of generators). *Let A be the generator of the Feller semigroup $\{\mathrm{T}_t\}_{t\in[0,T]}$ on a Banach space S. Then*

(1) *$\mathcal{D}(A)$ is a linear subspace of S;*
(2) *$\mathcal{D}(A)$ is dense in S;*
(3) *A is a **closed operator** in the sense that for $\{f_k\} \in \mathcal{D}(A)$ with $k = 1, \ldots$ where $f_k \to f$* and *$Af_k \to g \in S$ as $k \to \infty$, then*

$$f \in \mathcal{D}(A) \quad \text{and} \quad g = Af.$$

REMARK. While the generator of semigroup has many desirable properties, an important question is: *under what conditions do such generators exist*? Fortunately, the *Hille-Yosida theorem*, which is outside the scope of this paper, provides the necessary and sufficient conditions for a closed, linear operator to be the generator of a Feller semigroup (*cf.* [**29**, pg. 439]). Importantly, most price processes commonly used in practice admit generators and I will simply assume this fact in the sequel. □

Then assuming their existence, the following theorem describes the form of the generator of a general Itô diffusion. A proof of the result in the more general setting of a Lévy process is given in Kallenberg [**50**, pg. 374]

THEOREM 4.5. *Let $\{p_t\}_{t\in[0,T]}$ be the transition probability functions of a $\mathbb{R}^n$-valued Itô diffusion with characteristics (μ, σ) that is the solution to the following SDE*

$$dX(t) = \mu(t)\,dt + \sigma dW(t)$$

where σ is a positive definite $n \times m$ matrix and $W = (W(t))_{t\in[0,T]}$ is m-dimensional standard Brownian motion.

Then the generator of the process has the form

$$A(\cdot) = \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}\frac{\partial}{\partial x_i \partial x_j}(\cdot) + \sum_{i=1}^{n}\mu^i \frac{\partial}{\partial x_i}(\cdot). \tag{4.2.8}$$

where $a^{ij} = \left(\sigma^{\mathrm{T}}\sigma\right)^{ij}$.

4.3. Kolmogorov's Equations

From Theorem 4.3, we know that given a n-dimensional, time-homogeneous admissible Markov process $X = (X(t))_{t\in[0,T]}$ with characteristics (μ, σ), transition probability functions $\{p_t\}_{t\in[0,T]}$, and generator A, then

$$\frac{d}{dt}\mathrm{P}_t g = A\mathrm{P}_t g \tag{4.3.1}$$

for any $g \in C(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ where

$$\mathrm{P}_t g(\cdot) = \int_{\mathbb{R}} g(y)\, p_t(\cdot, dy).$$

Also, recall from (4.2.1)

$$\mathrm{P}_t g(x) = \mathbb{E}^x[g(X(t))] \quad (\text{for any } x \in \mathbb{R}^n).$$

Hence, letting $u(\cdot, t) := \mathrm{P}_t g(\cdot)$ for all $t \in [0,T]$

$$u(x,t) = \mathbb{E}^x[g(X(t))] \quad (\text{for any } x \in \mathbb{R}^n) \tag{4.3.2}$$

and, in particular,

$$u(x,0) \;=\; \mathbb{E}^x[g(X(0))] = g(x).$$

Consequently, by (4.3.1), u is the solution to the following PDE,

$$\begin{cases} u_t - Au = 0 & \text{in } \mathbb{R}^n \times (0,T] \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{4.3.3}$$

Additionally, from Theorem 4.5, we also know that the generator A is a differential operator of the form

$$A(\cdot) = \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}(x,t)\frac{\partial}{\partial x_i \partial x_j}(\cdot) + \sum_{i=1}^{n}\mu^i(x,t)\frac{\partial}{\partial x_i}(\cdot)$$

where $a^{ij} = \left(\sigma^{\mathrm{T}}\sigma\right)^{ij}$.

Therefore I can rewrite (4.3.3) as the following initial valued, linear, homogeneous, second-order PDE

$$\begin{cases} u_t - \frac{1}{2}\sum_{i,j=1}^{n} a^{ij} u_{x_i x_j} - \sum_{i=1}^{n} \mu^i u_{x_i} = 0 & \text{in } \mathbb{R}^n \times (0,T] \\ u = g & \text{on } \mathbb{R}^n \times \{t=0\}. \end{cases} \tag{4.3.4}$$

Now, assume that the transition probability functions admit densities such that (with a change of notation that is explicit on time)

$$\rho(x,y,t)\,dy := p_t(x,dy).$$

Again letting $u(\cdot,t) := \mathrm{P}_t g(\cdot)$, then for any $x \in \mathbb{R}^n$ and $t \in [0,T]$

$$\begin{aligned} u(x,t) &= \mathbb{E}^x\left[g(X(t))\right] \\ &= \int_{-\infty}^{\infty} g(y)\,\rho(x,y,t)\,dy. \end{aligned} \tag{4.3.5}$$

Thus

$$\begin{aligned} u_t &= \int_{-\infty}^{\infty} g(y)\,\rho_t(x,y,t)\,dy, \\ u_{x_i} &= \int_{-\infty}^{\infty} g(y)\,\rho_{x_i}(x,y,t)\,dy, \quad \text{and} \\ u_{x_i x_j} &= \int_{-\infty}^{\infty} g(y)\,\rho_{x_i x_j}(x,y,t)\,dy. \end{aligned}$$

Therefore by (4.3.4)

$$u_t - \frac{1}{2}\sum_{i,j=1}^{n} a^{ij} u_{x_i x_j} - \sum_{i=1}^{n} \mu^i u_{x_i} = 0 \quad (\text{in } \mathbb{R}^n \times (0,T])$$

or

$$\int_{\mathbb{R}} g(y)\left(\rho_t - \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}\rho_{x_i x_j} - \sum_{i=1}^{n}\mu^i \rho_{x_i}\right) dy = 0.$$

Since this is true for any $g \in C(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$, the integrand must be zero. Hence,

$$\rho_t - \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}\rho_{x_i x_j} - \sum_{i=1}^{n}\mu^i \rho_{x_i} = 0 \tag{4.3.6}$$

Additionally,

$$\begin{aligned}\int_{-\infty}^{\infty} g(y)\,\rho(x,y,0)\,dy &= u(x,0)\\ &= g(x).\end{aligned}$$

Hence $\rho(x,y,0)=\delta_y(x)$ for any fixed $y\in\mathbb{R}^n$ and any $x\in\mathbb{R}^n$ and, thus,

$$\begin{cases}\rho_t - A\rho &= 0 \quad \text{on } \mathbb{R}^n\times(0,T]\\ \rho(x,y,0) &= \delta_y(x) \quad \text{on } \mathbb{R}^n\times\{t=0\}\end{cases} \tag{4.3.7}$$

(4.3.7) represents the probability that the system will end up in some fixed terminal state $y\in\mathbb{R}^n$ at some time $t\in[0,T]$ after initially starting at some state x at time zero. Here, the starting state and the terminal time vary while the terminal state and the starting time are fixed. This is easier to see if the transition densities dependency on the initial time is explicit, or

$$\rho(x,y,t)=\rho(x,0,y,t).$$

Again, the starting time and ending state are fixed while the ending time and the starting state are allowed to vary. As such, (4.3.7) is often referred to as the ***Kolmogorov backwards equation*** because once the process has traveled from $X(0)\to X(t)=y$, an observer can look "backward" along the path to discovery where he started.

On the other hand, it seems reasonable to ask what happens if the ending time and starting state are fixed and starting time and ending state are allowed to vary. In other words, what is the probability that the system ends up in some state y at some fixed time $\tau\in[0,T]$ after starting at time $t\in[0,T]$ in some fixed initial state $x\in\mathbb{R}^n$. Such a system is governed by transition probability densities of the form

$$\rho(x,y,t)=\rho(x,t,y,\tau)$$

where $x\in\mathbb{R}^n$ and $\tau\in[0,T]$ are fixed. In this case, I have the following relationship for any $y\in\mathbb{R}^n$ and $t\in[0,T]$

$$\begin{aligned}u(y,t) &= \int_{-\infty}^{\infty} g(x)\,\rho(x,t,y,\tau)\,dx\\ &= \mathbb{E}\left[g(X(t)) \mid X(\tau)=y\right].\end{aligned}$$

While this is similar to (4.3.5), the expectation is conditioned on the terminal state rather than the initial state of the system. However, since we assumed that the process is non-anticipating, the terminal state is not observable at any earlier time. This is similar to the "backwards" heat equation (*i.e.* $u_t+\triangle u=0$ which is ill-posed. As is the case of the backwards heat equation, the way around the problem is to consider the *adjoint* problem (*cf.* [**66**, A.3.5]).

DEFINITION. Let A be some operator on the Hilbert space H. Then the operator A is called the ***adjoint of*** A and denoted as A^*, if it satisfies the following relationship

$$\langle A^*\varphi, \psi\rangle = \langle \varphi, A\psi\rangle \quad \text{(for any } \varphi, \psi \in H)$$

where $\langle \cdot, \cdot \rangle$ denotes the paring with of H with its dual . In our case, assume that the Hilbert space is $L^2(\mathbb{R}^n)$. Then

$$\langle \cdot, \cdot \rangle = (\cdot, \cdot)_{L^2(\mathbb{R}^n)}.$$

Thus, for any test functions $\varphi, \psi \in C_c^\infty(\mathbb{R}^n)$, the adjoint A^* of the generator A, satisfies

$$(A^*\varphi, \psi)_{L^2(\mathbb{R}^n)} = (\varphi, A\psi)_{L^2(\mathbb{R}^n)}. \tag{4.3.8}$$

Then by the definition of the L^2-inner product.

$$\begin{aligned}\int_{-\infty}^{\infty} \psi A^*\varphi dx &= \int_{-\infty}^{\infty} \varphi A\psi dx \\ &= \int_{-\infty}^{\infty} \varphi \left(\frac{1}{2}\sum_{i,j=1}^{n} a^{ij}(x,t)\,\psi_{x_i x_j} + \sum_{i=1}^{n} \mu^i(x,t)\,\psi_{x_i}\right) dx.\end{aligned}$$

Since both test functions have compact support, integrating by parts gives,

$$\int_{-\infty}^{\infty} \psi A^*\varphi dx = -\sum_{i,j=1}^{n} \int_{-\infty}^{\infty} \frac{1}{2}\left(\varphi a^{ij}\right)_{x_i} \psi_{x_j} dy - \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left(\varphi \mu^i\right)_{x_i} \psi dx.$$

Integrating the first term on the right-hand side by parts once more,

$$\int_{-\infty}^{\infty} \psi A^*\varphi dx = \int_{-\infty}^{\infty} \frac{1}{2}\sum_{i,j=1}^{n} \left(\varphi a^{ij}\right)_{x_i x_j} \psi - \sum_{i=1}^{n} \left(\varphi \mu^i\right)_{x_i} \psi dx.$$

Rearranging terms,

$$\int_{-\infty}^{\infty} \psi \left(A^*\varphi - \left(\frac{1}{2}\sum_{i,j=1}^{n} \left(\varphi a^{ij}\right)_{x_i x_j} - \sum_{i=1}^{n} \left(\varphi \mu^i\right)_{x_i}\right)\right) dx = 0.$$

Since this is true for any test functions, the integrand must be identically zero, or

$$A^*\varphi = \frac{1}{2}\sum_{i,j=1}^{n} \left(\varphi a^{ij}\right)_{x_i x_j} - \sum_{i=1}^{n} \left(\varphi \mu^i\right)_{x_i}.$$

Hence, A^* is a second-order (possibly unbounded) differential operator (in *divergence* form) with respect to y, or

$$A^*(\cdot) = \frac{1}{2}\sum_{i,j=1}^{n} \left((\cdot)\, a^{ij}\right)_{y_i y_j} - \sum_{i=1}^{n} \left((\cdot)\, \mu^i\right)_{y_i}.$$

Now, *assume* that $\tilde{\rho}$ is a solution to the terminal-valued PDE

$$\begin{cases} \tilde{\rho}_t(x,t,y,\tau) - A^*\tilde{\rho}(x,t,y,\tau) & = 0 \quad \text{on } \mathbb{R}^n \times [0,\tau) \\ \tilde{\rho}(x,\tau,y,\tau) & = \delta_x(y) \quad \text{on } \mathbb{R}^n \times \{t=\tau\} \end{cases} \tag{4.3.9}$$

where $x \in \mathbb{R}^n$ and $\tau \in [0,T]$ are fixed.

Recall (*cf.* (4.3.5)) that

$$u(x,t) = \mathbb{E}^x[g(X(t))] \quad (\text{for any } x \in \mathbb{R}^n, t \in [0,T])$$

where u is the solution to (4.3.3). Hence

$$u_t(y,t) = Au(y,t) \quad (\text{for any } y \in \mathbb{R}^n, t \in [0,T]).$$

Thus, for $x \in \mathbb{R}^n$ and $t, \tau \in [0,T]$, taking the derivative with respect to time gives

$$\begin{aligned} \frac{d}{dt}\int_{-\infty}^{\infty} u(y,t)\,\tilde{\rho}(x,t,y,\tau)\,dy &= \int_{-\infty}^{\infty} u(y,t)\,\tilde{\rho}_t(x,t,y,\tau) - \tilde{\rho}(x,t,y,\tau)\,u_t(y,t)\,dy \\ &= \int_{-\infty}^{\infty} u(y,t)\,A^*\tilde{\rho}(x,t,y,\tau) - \tilde{\rho}(x,t,y,\tau)\,Au(y,t)\,dy \\ &= (A^*\tilde{\rho}, u)_{L^2(\mathbb{R})} - (\tilde{\rho}, Au)_{L^2(\mathbb{R})} = 0. \end{aligned}$$

by the fact that A^* is the adjoint of A.

Therefore, $\int_{-\infty}^{\infty} u(y,t)\,\tilde{\rho}(x,t,y,\tau)\,dy$ does not explicitly depend on time and for any $x \in \mathbb{R}^n$ and $t, \tau \in [0,T]$

$$\begin{aligned} \int_{-\infty}^{\infty} \tilde{\rho}(x,t,y,\tau)\,u(y,t)\,dy &= \int_{-\infty}^{\infty} \tilde{\rho}(x,\tau,y,\tau)\,u(y,\tau)\,dy \\ &= \int_{-\infty}^{\infty} \delta_x(y)\,u(y,\tau)\,dy \\ &= u(x,T). \end{aligned}$$

In particular, this is true when $t = 0$. Hence, by (4.3.4)

$$\begin{aligned} u(x,\tau) &= \int_{-\infty}^{\infty} \tilde{\rho}(x,0,y,\tau)\,u(y,0)\,dy \\ &= \int_{-\infty}^{\infty} \tilde{\rho}(x,0,y,\tau)\,g(y)\,dy. \end{aligned}$$

However, by (4.3.5)

$$u(x,\tau) = \int_{-\infty}^{\infty} \rho(x,0,y,\tau)\,g(y)\,dy$$

Thus,

$$\int_{-\infty}^{\infty} (\tilde{\rho}(x,0,y,\tau) - \rho(x,0,y,\tau))\,g(y)\,dy = 0.$$

Since this true for any $g \in C(\mathbb{R}^n) \cap L^\infty(\mathbb{R})$,

$$\tilde{\rho}(x, 0, y, \tau) = \rho(x, 0, y, \tau) \quad (x, y \in \mathbb{R}^n \text{ and any } \tau \in \mathbb{T}).$$

Consequently, $\{\rho_t\}_{t\in[0,T]}$ are the solutions to the adjoint problem

$$\begin{cases} \rho_\tau - A^*\rho & = 0 \\ \rho(x, \tau, y, \tau) & = \delta_y(x). \end{cases} \tag{4.3.10}$$

(4.3.10) is known as ***Kolmogorov forward equation*** (*a.k.a.* ***Fokker-Planck equation***).

REMARK. The forward equation is more restrictive than the backward equation in the sense the drift and diffusion terms must be differentiable (twice differentiable in the case of the diffusions).□

4.4. Feynman-Kac Formula

Above, I showed that the evolution of distribution of any $\mathbb{R}^n$-valued admissible Itô diffusion with characteristics (μ, σ) and transition probability functions $\{p_t\}_{t\in[0,T]}$ can be described explicitly in terms of a PDE using its generator A. Specifically, if $u(x,t) := \mathrm{P}_t g(x)$ for some $g \in C(\mathbb{R}^n)$ and $x \in \mathbb{R}^n, t \in [0,T]$ with $T < \infty$ where

$$\mathrm{P}_t g(\cdot) = \int_{\mathbb{R}^n} g(y)\, p_t(\cdot, dy)$$

then u is the solution to the Kolmogorov backward equation

$$\begin{cases} u_t - Au = 0 & \text{in } \mathbb{R}^n \times (0,T] \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\} \end{cases}$$

Given this, it seems natural to ask if Kolmogorov's backward equation can be generalized by adding a zeroth order "potential term" or

$$\begin{cases} u_t - Au - V(x,t)\, u = 0 & \text{in } \mathbb{R}^n \times (0,T] \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

In other words, is there a connection between some class of stochastic processes whose generator is a general linear second-order differential operator with possibly non-constant coefficients and a zeroth-order term, or

$$A(\cdot) = \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}(x,t)\, \frac{\partial^2}{\partial x_i \partial x_j}(\cdot) + \sum_{i=1}^{n} \mu^i(x,t)\, \frac{\partial}{\partial x_i}(\cdot) + c(x,t)(\cdot).$$

NOTATION. While I have tried to use capital letters to denote stochastic variables, here I have chosen to adopt the standard notation of V for the potential term. □

In fact, there is such a generalization and it is the central result showing the duality between the probabilistic statement of the stochastic dynamical system and its PDE representation. This is

formalized in the next result known as the *Feynman-Kac formula* which is proved in Pascucci (*cf.* [**66**, pg. 311-313]).

THEOREM 4.6 (Feynman-Kac formula). *Let $X = (X(t))_{t\in[0,T]}$ be a $\mathbb{R}^n$-valued, time-homogeneous Itô diffusion with characteristics (μ, σ) and generator A of the form,*

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

where σ is an $n \times m$ matrix and $W = (W(t))_{t\in[0,T]}$ is m-dimensional standard Brownian motion. Further, let $V : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ be bounded and smooth. If u is the solution to the following PDE,

$$\begin{cases} u_t - Au - V(x,t)\,u = 0 & \text{in } \mathbb{R}^n \times (0,T] \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{4.4.1}$$

where $a^{ij} = \left(\sigma^{\mathrm{T}}\sigma\right)^{ij}$, then for any $x \in \mathbb{R}^n$ and $t \in [0,T]$

$$u(x,t) = \mathbb{E}^x\left[g(X(t))\,e^{-\int_0^T V(X(s))ds}\right]. \tag{4.4.2}$$

Moreover, if $g \in \mathcal{D}(A)$, then (4.4.2) *is a solution to the PDE expressed in* (4.4.1).

Now, assume that $g(X(t))$ represents the future payoff of some financial instrument A given the price $X(t)$ of some other instrument B at time t. Then by (4.4.2), $u(x,t)$ is the expected discounted payoff given that the price of B is currently x (*i.e.* $X(0) = x$). In other words, u is the price of A conditioned on the current price of B. Then by the Feynman-Kac formula, the price can be seen to evolve purely deterministically. From PDE theory, it is clear when such solutions exist and are unique (*cf.* [**29**, Ch. 7.1]). Hence, the price of fairly complex financial instrument can be stated equivalently as a conditional expectation of a discounted price process or as the solution of linear second-order PDE.

4.5. Derivation of the Black-Scholes PDE by Replication

The following derivation of the *Black-Scholes PDE* is based on the proposition that price of a financial instrument known as an *option* can determined using a *replicating portfolio* whose aggregate price is observable in the market at any time. Then, by an important result from finance theory known as the *no arbitrage condition*, the price of the option must equal that of the portfolio up to and including *expiry*.

As usual, let me begin with some basic definitions and results.

DEFINITION. A ***call*** (***put***) ***option*** gives the buyer the right, but not the obligation, to purchase (sell) (*a.k.a.* ***exercise***) the *underlying instrument* (*a.k.a.* ***underlying***) at some point in the future on or prior to the option's *expiration* (*a.k.a.* ***expiry***) at a price set on trade date (*a.k.a.* the ***strike price*** or ***exercise price***).

For the purpose of this paper, I will assume all options are call options and that any option is uniquely identified by three parameters: its underlying, its expiry, and its strike price. For this derivation, I will also assume that the option can only be exercised at expiry. Such an option is known as an ***European*** option.

DEFINITION. The ***payoff*** of a European call option with strike price $K > 0$ and expiry $T > 0$ is given by

$$g(T) := (X(T) - K)^+.$$

where $X = (X(t))_{t\in[0,T]}$ is the price process of the underlying

In general, investors are interested in the *present value* of a payoff and not simply its magnitude. This essentially *discounts* the payoff to account of the *time value of money*. This phenomenon is captured in the following results.

DEFINITION. A financial instrument trading in some market with a fully deterministic price process of the form

$$\begin{cases} dX(t) &= r(t) X(t)\, dt \quad (t \in (0,T]) \\ X(0) &= r_0. \end{cases}$$

where $r : [0,T] \to \mathbb{R}_+$ is a smooth, non-negative function is called a ***riskless instrument*** and r is known as the ***risk-free rate***.

THEOREM 4.7. *Assume that $X = (X(t))_{t\in[0,T]}$ is the price process of a riskless instrument. Then*

$$X(t) = r_0 \exp\left(\int_0^t r(s)\, ds\right)$$

where $r \in C^\infty(0,T;\mathbb{R}_+)$ is the risk-free rate and $r_0 = r(0)$.

PROOF. Let $X = (X(t))_{t\in[0,T]}$ be the price process of a riskless instrument with the risk-free $r \in C^\infty(0,T;\mathbb{R}_+)$ where $r_0 = r(0)$. Since, X is a deterministic price process, its dynamics are of the form

$$\begin{cases} dX(t) &= r(t) X(t)\, dt \quad (t \in (0,T]) \\ X(0) &= r_0. \end{cases}$$

Hence,

$$\frac{dX(t)}{X(t)} = r(t)\, dt.$$

Integrating both sides with respect to time,

$$\ln X(t) = \int_0^t r(s)\, ds + C \;\Rightarrow\; X(t) = r_0 \exp\left(\int_0^t r(s)\, ds\right).$$

Importantly, by the no arbitrage condition, if two financial are riskless, they must both yield the risk-free rate. In other words, they are indistinguishable (*a.k.a. fungible*). As such, I will assume that there is only one riskless instrument in the sequel. □

NOTATION. Since all riskless instruments are indistinguishable, I will denote the risk-free rate process as

$$X_0 = (X_0(t))_{t\in[0,T]}$$

REMARK. For simplicity's sake, going forward, I will assume that the risk-free rate is constant and that $r_0 = 1$. Hence, the risk-rate process of any riskless instrument $X_0 = (X_0(t))_{t\in[0,T]}$ is given by

$$X_0(t) = e^{rt} \quad (t \in [0,T]). \tag{4.5.1}$$

□

DEFINITION. Let $X = (X(t))_{t\in[0,T]}$ be the price process of a general financial instrument. Then, its ***discounted price process*** is given by

$$\tilde{X}(t) = \frac{X(t)}{X_0(t)} = \frac{X(t)}{e^{rt}} = e^{-rt}X(t).$$

REMARK. Effectively, the discounted price process expresses the future price of a given financial instrument in terms of the value of a dollar today. As such, the denominator is a process which converts one financial instrument (a future dollar) into units of another (today's dollar) . Many applications in continuous-time finance involve such conversions. For example, expressing dollars in terms of Euros. Such a conversion generally involves applying a suitable denominator, known as the ***numeraire***, in the above expression. □

DEFINITION. The ***price*** of a European call option with strike price $K > 0$ and expiry $T > 0$ is its expected discounted payoff, or

$$\begin{aligned} u(x,t) &= \mathbb{E}\left[e^{-r(T-t)}g(T) \mid X(t) = x\right] \\ &= \mathbb{E}^{x,t}\left[e^{-r(T-t)}g(T)\right]. \end{aligned} \tag{4.5.2}$$

where $X = (X(t))_{t\in[0,T]}$ is the price process of the underlying and $g(T) = (X(T) - K)^+$.

REMARK. For an European option, the payoff is a *terminal payoff* of a single payment at expiry if $X(T) > K$. In the event that a financial instrument makes payments $h(\cdot)$ prior to some finite investment horizon $T < \infty$ (*i.e.* a *running payoff*), then its price is given by

$$u(x,t) = \mathbb{E}^{x,t}\left[\int_0^t e^{-r(T-s)}h(s)\,ds + e^{-r(T-t)}g(T)\right].$$

□

DEFINITION. A ***market*** is a finitely countable collection of financial instruments. A ***portfolio*** is a weighted subset of the market denoted as $\{w_j\}_{j=1}^n$ where n is the number of financial instruments in the portfolio and w_j is the weight of the j^{th} constituent. We say that a market is ***ideal*** if the payoff

of any financial instrument can be replicated by a weighted portfolio whose price is observable in the market at all times. The ***value*** or ***price*** of a portfolio at time $t \in [0,T]$ is given by

$$v(t) = \sum_{j=1}^{n} w_j(t)\, p_j(t)$$

where p_j is the price of the j^{th} constituent. Further, a portfolio is said to ***self-financing*** if all changes in value to portfolio are due to changes to in the market value of the constituents (*i.e. capital gains* only) or for any $t \in [0,T]$

$$\frac{d}{dt}v(t) = \sum_{j=1}^{n} w_j(t)\, \frac{d}{dt}p_j(t).$$

REMARK. Importantly, any rebalancing of the portfolio (*i.e.* changing of the weights) does not change the value of the portfolio. In other words, the proceeds of any sale of a constituent are used to by additional units of other constituents and purchasing additional units without some offsetting sale is prohibited. Self-financing is a necessary condition ensuring the existence of a replicating portfolio (*cf.* **[66**, Ch. 7.1]). □

DEFINITION. We say that a price process of a financial instrument $X = (X(t))_{t\in[0,T]}$ is governed by ***Black-Scholes dynamics*** if X is the solution the following SDE

$$dX(t) = \mu X(t)\, dt + \sigma X(t)\, dW(t) \quad (t \in [0,T]) \tag{4.5.3}$$

where

μ and σ are non-negative constants; and

$W(\cdot)$ is standard Brownian motion.

REMARK. In other words, such a price process is an Itô diffusion with constant coefficients. Hence, a unique solution to (4.5.3) exists by Theorem 3.23 and X is well-defined. □

The next result describes the price of a European option as the solution to a second-order, parabolic PDE.

THEOREM 4.8 (Black-Scholes PDE). *Assume that some European option has expiry $T > 0$, exercise price $K>0$, and is trading in an ideal market. Further, assume that the price of its underlying instrument is governed by Black-Scholes dynamics. Then, the price of the option u is the unique solution to the following terminal-valued PDE*

$$\begin{cases} u_t + \frac{1}{2}\sigma^2 x^2 u_{xx} + rxu_x - ru = 0 & \mathbb{R}\times[0,T) \\ u = g & \mathbb{R}\times\{t=T\} \end{cases}$$

where r is the constant risk-free rate and g is the payoff at expiry given by

$$g(x) = (x-K)^+$$

PROOF. Select some European option with expiry $T > 0$, strike price $K > 0$, and payoff $g \in C^{\infty}(\mathbb{R})$ that is trading in an ideal market. Further, assume the price of its underlying is governed by Black-Scholes dynamics with constant characteristics (μ, σ). Since the market is ideal, there is some replicating portfolio with the same payoff $g(\cdot)$ at each time $t \in [0, T]$.

Now, by the no arbitrage condition, all replicating portfolios have must have the same price at all times. Therefore, I need to consider only one and assume the *ansatz* that this replicating portfolio consists two instruments: (a) the underlying and (b) a riskless zero-coupon bond (*i.e.* make no income payments until maturity) with a constant interest rate r equal to the risk-free rate and that matures on the expiry date of the option. Further assume that prices of these two instruments are observable at all times. Without loss of generality, this portfolio can assumed to be self-financing since any income generated by the constituents can be removed from the portfolio. Then, the task at hand is to determine the weighting of the underlying and the bond at each point in time.

Therefore, let $X = (X(t))_{t\in[0,T]}$ and $Y = (Y(t))_{t\in[0,T]}$ be the price processes of the underlying and the bond, respectively. Then for any $t \in [0, T]$, the price of the portfolio is given by

$$V(t) := w(t) X(t) + v(t) Y(t) \tag{4.5.4}$$

where $w(\cdot)$ and $v(\cdot)$ are smooth, real-valued functions representing the amount invested in the underlying and the bond respectively.

Since the replicating portfolio is assumed to be self-financing, by (4.5.4), any change in its price is due solely to changes in the price of the underlying and/or the bond. Hence,

$$dV(t) = w(t) dX(t) + v(t) dY(t). \tag{4.5.5}$$

Further, by assumption, the price of the underlying is governed by Black-Scholes dynamics. Thus, for any $t \in [0, T]$

$$dX(t) = \mu X(t) dt + \sigma X(t) dW(t)$$

where

μ and σ are non-negative constants; and

$W(\cdot)$ is standard Brownian motion.

On the other hand, the bond is assumed to be riskless yielding the risk-free rate r. Hence, by (4.5.1), the price process of the bond is

$$Y(t) = e^{rt} \quad (\text{for } t \in [0, T]).$$

Substituting these expressions into (4.5.5), the change in portfolio value is given by

$$\begin{aligned} dV(t) &= w(t) dX(t) + v(t) dY(t) \\ &= w(t) (\mu X(t) dt + \sigma X(t) dW(t)) + v(t) re^{rt} dt \\ &= \left(w(t) \mu X(t) + v(t) re^{rt}\right) dt + w(t) \sigma X(t) dW(t). \end{aligned} \tag{4.5.6}$$

Now, I want to select $w(\cdot)$ and $v(\cdot)$ such that replicating portfolio has same payoff as the option at all times or

$$\begin{cases} V(t) & = u(X(t),t) \quad (t \in [0,T]) \\ V(T) & = g(X(T)) \end{cases} \tag{4.5.7}$$

where $u(X(t),t)$ is the price of the option at time t with the payoff at expiry $g(X(T))$. Then for $t \in [0,T]$, by Itô's formula,

$$\begin{aligned} dV(t) &= du(X(t),t) \\ &= u_t(X(t),t)\,dt + u_x(X(t),t)\,dX(t) + \frac{1}{2}u_{xx}(X(t),t)\,dX(t)\,dX(t). \end{aligned}$$

Noting that

$$\begin{aligned} dX(t)\,dX(t) &= \mu^2 X^2(t)\,dt^2 + 2\mu\sigma X^2 dt dW(t) + \sigma^2 X^2(t)\,dW(t)\,dW(t) \\ &= \sigma^2 X^2 dt, \end{aligned}$$

the value of the portfolio is

$$\begin{aligned} dV(t) &= u_t dt + u_x dX(t) + \frac{1}{2}\sigma^2 X^2(t)\,u_{xx} dt \\ &= \left(u_t + \frac{1}{2}\sigma^2 X^2(t)\,u_{xx}\right)dt + u_x(\mu X(t)\,dt + \sigma X(t)\,dW(t)) \\ &= \left(u_t + \mu X(t)\,u_x + \frac{1}{2}\sigma^2 X^2(t)\,u_{xx}\right)dt + \sigma X(t)\,u_x dW(t). \end{aligned} \tag{4.5.8}$$

Equating the dW terms in (4.5.6) and (4.5.8), the weight of the underlying is given by

$$w(t)\,\sigma X(t) = \sigma X(t)\,u_x \quad \Rightarrow \quad w(t) = u_x. \tag{4.5.9}$$

Now, equating the dt terms and using the above result,

$$\mu X(t)\,u_x + v(t)\,re^{rt} = u_t + \mu X(t)\,u_x + \frac{1}{2}\sigma^2 X^2(t)\,u_{xx}.$$

Hence,

$$v(t)\,re^{rt} = u_t + \frac{1}{2}\sigma^2 X^2(t)\,u_{xx}.$$

Rearranging terms, the weight of the bond is given by

$$v(t) = \frac{1}{r}e^{-rt}\left(u_t + \frac{1}{2}\sigma^2 X^2(t)\,u_{xx}\right).$$

Substituting these weights into (4.5.4), the price of the portfolio is given by

$$V(t) = u_x X(t) + \frac{1}{r}e^{-rt}\left(u_t + \frac{1}{2}\sigma^2 X^2 u_{xx}\right)e^{rt}.$$

Recalling that the price of the option is equal to the price of the portfolio at all times by (4.5.7) and that $X(t) = x$, the price of the option is the solution to the following (deterministic) PDE known as the ***Black-Scholes PDE***

$$\begin{cases} u_t + \frac{1}{2}\sigma^2 x^2 u_{xx} + rxu_x - ru = 0 & \mathbb{R} \times [0, T) \\ u = g & \mathbb{R} \times \{t = T\}. \end{cases} \tag{4.5.10}$$

□

Comparing the Black-Scholes PDE with the Feynman-Kac PDE (4.4.1), it is clear that the first order terms representing the price appreciation (*i.e.* drift terms) of the underlying are different. This suggests that the price of the option as expressed in (4.5.2) needs to be modified. The solution to this problem is to express the expected payoff under a different probability measure known as the *risk-neutral probability measure* (*a.k.a. the equivalent martingale measure*) as well be shown in the next section.

4.6. Derivation of the Black-Scholes PDE using Martingale Methods

Again, select some European call option with expiry $T > 0$, strike price $K>0$, and payoff $g(\cdot) = (\cdot - K)^+$. Further, assume that the price of the underlying $X = (X(t))_{t\in[0,T]}$ is governed by Black-Scholes dynamics with constant characteristics (μ, σ) (4.5.3). Then by the Feynman-Kac formula (*cf.* Theorem 4.6), the price of the option u is the unique solution to the following PDE

$$\begin{cases} u_t + \frac{1}{2}\sigma^2 x^2 u_{xx} + \mu x u_x - ru = 0 & \mathbb{R} \times [0, T) \\ u = g & \mathbb{R} \times \{t = T\}. \end{cases} \tag{4.6.1}$$

.

However, upon inspection, the Black-Scholes PDE (4.5.10) and this PDE are only the same when $\mu = r$ (*i.e.* when the appreciation rate of the underlying is equal to the free rate).

To understand under what conditions this may happen, consider the discounted price process $\tilde{X} = \left(\tilde{X}(t)\right)_{t\in[0,T]}$ where

$$\tilde{X}(t) = e^{-rt} X(t) \quad (\text{for } t \in [0, T]).$$

Then for $t \in [0, T]$, by Itô's formula,

$$\begin{aligned} d\tilde{X}(t) &= d\left(e^{-rt} X(t)\right) \\ &= e^{-rt}(-rX(t)\,dt + dX(t)) \\ &= e^{-rt}(-rX(t)\,dt + \mu X(t)\,dt + \sigma X(t)\,dW(t)) \\ &= e^{-rt}((\mu - r)X(t)\,dt + \sigma X(t)\,dW(t)) \\ &= (\mu - r)\tilde{X}(t)\,dt + \sigma\tilde{X}(t)\,dW(t). \end{aligned} \tag{4.6.2}$$

Setting $\mu = r$, then

$$d\tilde{X}(t) = \sigma \tilde{X}(t)\, dW(t)$$

or

$$\tilde{X}(t) = \tilde{X}(0)\, e^{\sigma W(t)}.$$

Hence, when $\mu = r$, the discounted price process$\tilde{X}(t)$ is pure geometric Brownian motion and, thus, is a martingale. This sheds some light on the way out of the dilemma of reconciling the Black-Scholes PDE and the Feynman-Kac formula. We need to express the expected payoff of the option in terms of a probability measure, known as the *risk-neutral probability measure* (*a.k.a. equivalent martingale measure* or *EMM*) that, after applying the Feynman-Kac theorem, will result in the Black-Scholes PDE. Fortunately, deriving this risk-neutral probability measure is a straightforward application of Girsanov's theorem (*cf.* Theorem 3.22).

Stating the discounted price process (4.6.2) in the more general setting where the drift and diffusion coefficients are deterministic functions of time,

$$d\tilde{X}(t) = \sigma \tilde{X}(t) \left(\frac{(\mu(t) - r(t))}{\sigma(t)} dt + dW(t) \right). \tag{4.6.3}$$

Now, define the ***market price of risk*** as

$$\Theta(t) = \frac{(\mu(t) - r(t))}{\sigma(t)}.$$

Then (4.6.3) becomes

$$d\tilde{X}(t) = \sigma \tilde{X}(t) \left(\Theta(t)\, dt + dW(t) \right). \tag{4.6.4}$$

Further, define the process $\tilde{W} = \left(\tilde{W}(t) \right)_{t \in \mathbb{T}}$ as

$$\tilde{W}(t) = \int_0^t \Theta(s)\, ds + W(t)$$

or, in differential form

$$d\tilde{W}(t) = \Theta(t)\, dt + dW(t).$$

Hence, $\tilde{W}$ is Brownian motion with drift $\Theta(t)$ and by (4.6.4)

$$d\tilde{X}(t) = \sigma \tilde{X}(t)\, d\tilde{W}(t).$$

Now, by Girsanov's theorem (*cf.* Theorem 3.22), $\tilde{W}$ can be transformed into a martingale under the probably measure $\mathbb{Q}$ defined as

$$\mathbb{Q}(A) = \int_A Z(T)\, d\mathbb{P} \quad (A \in \mathcal{B})$$

where $Z = (Z(t))_{t\in[0,T]}$

$$Z(t) = \exp\left(-\frac{1}{2}\int_0^t \Theta^2(s)\,ds - \int_0^t \Theta(s)\,dW(s)\right) \quad (\text{for } t \in [0,T]).$$

Consequently, under the probability measure $\mathbb{Q}$, the discount price process $\tilde{X}$ is a martingale. Further, under this new probability measure $\mathbb{Q}$, the price of a European option with a strike price K and expiry T by definition is the expected discounted payoff given by

$$u(x,t) = \mathbb{E}_{\mathbb{Q}}^{x,t}\left[e^{-r(T-t)}g(X(T))\right].$$

where r is the risk-free rate and $g(\cdot)$ is the payoff function. Further, the dynamics of the underlying can be expressed in terms of the risk-neutral measure as follows

$$\begin{aligned} dX(t) &= \mu X(t)\,dt + \sigma X(t)\,dW(t) \\ &= \mu X(t)\,dt + \sigma X(t)\left(d\tilde{W}(t) - \Theta dt\right) \\ &= (\mu - \sigma\Theta)X(t)\,dt + \sigma X(t)\,d\tilde{W}(t) \\ &= rX(t)\,dt + \sigma X(t)\,d\tilde{W}(t). \end{aligned}$$

Then by the Feynman-Kac formula under the EMM

$$\begin{cases} u_t + \frac{1}{2}\sigma^2 x^2 u_{xx} + rxu_x - ru = 0 & \mathbb{R}\times[0,T) \\ u = g & \mathbb{R}\times\{t=T\}. \end{cases}$$

which is the Black-Scholes PDE (*cf.* (4.5.10)).

REMARK. Given this, the implication is that all pricing, whether by replication or martingale methods, should be performed under the risk-neutral measure. □

REMARK. An interesting question is: *under what conditions does a risk-neutral probability measure exist and is it unique*? This is captured in two important results in continuous-time finance known as the *fundamental theorems of asset pricing* that are outside the scope of this paper. The first theorem states that if a market has a risk neutral measure, then it is arbitrage free. The second theorem states that if a market has a risk neutral measure, then it is complete if and only if the risk-neutral probability measure is unique. For a discussion of these theorems, see Shreve [**76**, pg 231-234].

Commonly Used Notation

Abbreviations and Acronyms

a.s.:	almost surely.
cdf:	the cumulative distribution function.
càdlàg:	right continuous stochastic process with left limits.
EMM:	the equivalent martingale measure.
i.i.d.:	independent, identicially distributed random variables
PDE:	partial differential equation(s).
pdf:	the probability density function of a random variable.
RCLL:	right continuous stochastic process with left limits.
SDE:	stochastic differential equation(s).

Functions and Operators

$[X, Y](\cdot)$:	the covariation of the processes X and Y.
$[X](\cdot)$:	the quadratic variation of the process X.
$\mathbb{P}_X(\cdot)$:	the distribution of the random variable X.
$1_A(\cdot)$:	the indicator function of the set A.
$\delta_\omega(\cdot)$:	the Dirac measure centered at ω.
$u_{x_i x_j}$:	$\frac{\partial^2}{\partial x_i \partial x_j} u$ where $x \in \mathbb{R}^n$.
$(\cdot)^+$:	$\max(\cdot, 0)$.
$\|\cdot\|_{\mathbb{L}^p(0,T)}$:	the $\mathbb{L}^p$-norm $\left(i.e.\ \|\cdot\|_{\mathbb{L}^p(0,T)} = \left(\mathbb{E}\left[\int_0^T X^p(t)\, ds \right] \right)^{1/p} \right)$.
$\|\cdot\|_{L^p(\Omega)}$:	the L^p-norm $\left(i.e.\ \|\cdot\|_{L^p(\Omega)} = \left(\int_\Omega (\cdot)^p\, d\mathbb{P} \right)^{1/p} \right)$.
$\|\cdot\|_{op}$:	the operator norm on some Banach space.
$\mathbb{E}^{x,s}[X(t)]$:	the conditional expectation of X given $X(s) = x$.
$\mathbb{E}^x[X(t)]$:	the conditional expectation of X given $X(0) = x$.
$\mathbb{E}_{\mathbb{P}}[\cdot]$:	the expectation with respect to the probability measure $\mathbb{P}$.
$\otimes$:	the set product.
$\triangle(\cdot)$:	the Laplacian $\left(i.e. \triangle(\cdot) = \sum_{j=1}^n \frac{\partial^2}{\partial x_j \partial x_j}(\cdot) \right)$.
u_t:	$\frac{\partial}{\partial t} u(x,t)$.
u_{x_i}:	$\frac{\partial}{\partial x_i} u$ where $x \in \mathbb{R}^n$.
u_{xx}:	$\frac{\partial^2}{\partial x^2} u(x,t)$ where $x \in \mathbb{R}$.

Sets and Spaces

$\mathbb{N}$:	the non-negative integers $\{0, 1, 2, \ldots\}$.

$\mathbb{T}$: the time index set.

2^{Ω}: the discrete σ-algebra on the set Ω.

$\mathbb{R}_+$: the non-negative reals, $(i.e.\ \mathbb{R}_+ := [0,\infty))$.

$\sigma(A)$: the smallest σ-algebra generated by A.

$(\Omega;\mathcal{F})$: indicates that the domain Ω of a function is equipped with a σ-algebra $\mathcal{F}$.

$\mathbb{L}^p(0,T)$: the collection of all admissible processes bounded in the $\mathbb{L}^p$-norm.

$\mathbb{Z}$: the integers.

$\mathcal{B}(\mathrm{S})$: the σ-algebra of Borel sets on the space S.

$\mathcal{N}$: the collection of $\mathbb{P}$-null sets.

$\mathrm{BV}(0,T)$: the collection of all real-valued functions with bounded variation on $[0,T]$.

A^c: the complement of the set A.

$C_1^2(U\times[0,T])$: $\{f : U\times[0,T]\to\mathbb{R} \mid f, D_x f, D_x^2 f, f_t \in C(U\times[0,T];V)\}$.

$L^{p*}(\Omega)$: the dual space of L^p.

$L^p(\Omega)$: the collection of all L^p-bounded random variables on $(\Omega,\mathcal{F},\mathbb{P})$.

$\arg\min(\cdot)$: a minimizer.

Symbols

A^{ij}: the $i^{th}j^{th}$ component of the matrix A.

$\overset{d}{=}$: equal in distribution.

$\overset{d}{\to}$: convergence in distribution.

$\overset{p}{\to}$: convergence in probability.

$\downarrow$: the right limit or $\lim_{x_k\downarrow x}$.

$\overset{L^p}{\to}$: convergence in the L^p norm.

$(\cdot)^{\mathrm{T}}$: the transpose.

$\perp$: independent of.

$\uparrow$: the left limit or $\lim_{A_k\uparrow A}$.

$N(\mu,\sigma^2)$: a normal distribution with mean μ and variance σ^2.

Bibliography

1. C.D. Aliprantis and K.C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer, 2007.
2. Marco Avellaneda, Arnon Levy, and Antonio Paras, *Pricing and Hedging Derivative Securities in Markets with Uncertain Volatilities*, Applied Mathematical Finance **2** (1995), 73–88.
3. L. Bachelier, M. Davis, A. Etheridge, and P.A. Samuelson, *Louis Bachelier's Theory of Speculation: The Origins of Modern Finance*, Princeton University Press, 2011.
4. Kerry Back, *A Course in Derivative Securities: Introduction to Theory and Computation*, Springer, 2005.
5. H. Bauer, *Measure and Integration Theory*, De Gruyter studies in mathematics, W. de Gruyter, 2001.
6. Heinz Bauer, *Probability Theory*, De Gruyter Studies in Mathematics, Walter de Gruyter & Co., 1996.
7. Bruno Biais and Wolfgang J Runggaldier (eds.), *Financial Mathematics: Lectures Given at the 3rd Session of the Centro Internazionale Matematico Estivo (CIME) Held in Bressanone, Italy, July 8-13, 1996*, no. 1656, Springer, 1997.
8. Patrick Billingsley, *Probability and Measure*, Wiley, 2006.
9. Nicholas H. Bingham and Rüdiger Kiesel, *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*, second ed., Springer, 2010.
10. Fischer Black, *The Pricing of Commodity Contracts*, Journal of Financial Economics **3** (1976), 167–179.
11. Fischer Black and Myron S. Scholes, *The Pricing of Options and Corporate Liabilities*, Journal of Political Economy **81** (1973), no. 3, 637–654.
12. R. Buff, *Uncertain Volatility Models: Theory and Application*, Springer Finance, Springer Verlag, 2002.
13. Erhan Çinlar, *Introduction to Stochastic Processes*, Dover, 2010.
14. ———, *Probability and Stochastics*, Springer, 2011.
15. Alexandre Chorin and Ole H. Hald, *Stochastic Tools in Mathematics and Science*, second ed., Springer, 2009.
16. Jean-Michel Courtault, Yuri Kabanov, Bernard Bru, Pierre Crepel, Isabelle Lebon, and Arnaud Le Marchand, *Louis Bachelier on the Centenary of Theorie de la Speculation*, Mathematical Finance **10** (2000), no. 3, 341–350.
17. D.R. Cox and H.D. Miller, *The Theory of Stochastic Processes*, Chapman & Hall/CRC, 1965.
18. John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross, *A Theory of the Term Structure of Interest Rates*, Econometrica **53** (1985), no. 2, 385–408.
19. John C. Cox and Stephen A. Ross, *The Valuation of Options for Alternative Stochastic Processes*, Journal of Finance **3** (1976), no. 20292, 145–166.
20. John C. Cox, Stephen A. Ross, and Mark B. Rubinstein, *Option Pricing: A Simplified Approach*, Journal of Financial Economics **7** (1979), 229–263.
21. John C. Cox and Mark Rubinstein, *Options Markets*, Prentice-Hall, 1985.
22. Darrell Duffie, *Dynamic Asset Pricing Theory*, third ed., Princeton University Press, 2001.
23. Rick Durrett, *Probability: Theory and Examples*, fourth ed., Cambridge University Press, 2011.
24. E.B. Dynkin, D.E. Brown, and T. Kovary, *Theory of Markov Processes*, Dover Books on Mathematics, Dover Publications, 2012.
25. Edwin J. Elton, Martin J. Gruber, Stephen J. Brown, and William N. Goetzmann, *Modern Portfolio Theory and Investment Analysis*, eighth ed., Wiley, 2009.
26. T. Wake Epps, *Quantitative Finance: Its Development, Mathematical Foundations, and Current Scope*, Wiley, 2009.

27. S.N. Ethier and T.G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley Series in Probability and Statistics, Wiley, 2009.
28. Lawrence C. Evans, *Measure Theory and Fine Properties of Functions*, CRC Press, 1992.
29. ———, *Partial Differential Equations*, second ed., American Mathematical Society, 2010.
30. ———, *An Introduction to Stochastic Differential Equations*, American Mathematical Society, 2014.
31. William Feller, *On the Theory of Stochastic Processes, with Particular Reference to Applications*, University of California Press, Berkeley, Calif., 1949.
32. ———, *An Introduction to Probability Theory and Its Applications*, second ed., vol. 2, Wiley, 1971.
33. Jean-Pierre Fouque, George Papanicolaou, and K. Ronnie Sircar, *Derivatives in Financial Markets with Stochastic Volatility*, Cambridge University Press, 2000.
34. Christian Fries, *Mathematical Finance: Theory, Modeling, Implementation*, Wiley, 2007.
35. Marco Frittelli and Wolfgang Runggaldier (eds.), *Stochastic Methods in Finance*, Springer, 2003.
36. Rajna. Gibson, Francois-Serge Lhabitant, and Denis Talay, *Modeling the Term Structure of Interest Rates: A Review of the Literature*, Foundations and Trends in Finance, Now, 2010.
37. Paul Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer, 2003.
38. J. Michael Harrison and David Kreps, *Martingales and Arbitrage in Multiperiod Securities Markets*, Journal of Economic Theory **20** (1979), no. 3, 381–408.
39. J. Michael Harrison and Stanley R. Pliska, *Martingales and Stochastic Integrals in the Theory of Continuous Trading*, Stochastic processes and their applications **11** (1981), 215–260.
40. ———, *A Stochastic Calculus Model of Continuous Trading: Complete markets*, Stochastic Processes and their Applications **15** (1983), no. 3, 313–316.
41. John Heaney and Geoffrey Poitras, *Semigroup Properties of Arbitrage Free Pricing Operators*, 1994.
42. David Heath, Robert Jarrow, and Andrew Morton, *Bond Pricing and the Term Structure of Interest Rates*, Econometrica **60** (1992), no. 1, 29.
43. J. Hull, *Pricing Interest-Rate-Derivative Securities*, Review of Financial Studies **3** (1990), no. 4, 573–592.
44. Jonathan E. Ingersoll, *Theory of Financial Decision Making*, vol. 43, Studies in Financial Economics, no. 1, Rowman & Littlefield, 1987.
45. Michael Jacobs Jr., *A Comparison of Fixed Income Valuation Models: Pricing and Econometric Analysis of Interest Rate Derivatives*, Ph.d. disseration, The City University of New York, 2001.
46. Jessica James and Nick Webber, *Interest Rate Modelling: Financial Engineering*, Wiley, 2000.
47. Monique Jeanblanc, Marc Yor, and Marc Chesney, *Mathematical Methods for Financial Markets*, Springer, 2009.
48. Frank Jones, *Lebesgue Integration on Euclidean Space*, Jones & Bartlett, 2001.
49. E. Jouini, J. Cvitanic, and M. Musiela, *Handbooks in Mathematical Finance: Option Pricing, Interest Rates and Risk Management*, Handbooks in Mathematical Finance, Cambridge University Press, 2001.
50. Olav Kallenberg, *Foundations of Modern Probability*, second ed., Springer, 2002.
51. Ioannis Karatzas and Steven E. Shreve, *Brownian Motion and Stochastic Calculus*, second ed., Springer, 1991.
52. Samuel Karlin and Howard M. Taylor, *A First Course in Stochastic Processes*, second ed., Academic Press, 1975.
53. ———, *A Second Course in Stochastic Processes*, Academic Press, 1981.
54. Ralf Korn and Elke Korn, *Option Pricing and Portfolio Optimization: Modern Methods of Financial Mathematics*, American Mathematical Society, 2001.
55. John W. Lamperti, *Stochastic Processes: A Survey of the Mathematical Theory* , Springer, 1977.
56. ———, *Probability: A Survey of the Mathematical Theory*, Wiley Series in Probability and Statistics, Wiley, 2011.
57. Gregory F. Lawler, *Introduction to Stochastic Processes*, second ed., Chapman & Hall/CRC, 2006.
58. G.M. Lieberman, *Second Order Parabolic Differential Equations*, World Scientific, 1996.
59. Francis A. Longstaff, *A Nonlinear General Equilibrim Model of the Term Structure of Interest Rates*, Journal of Financial Economics **23** (1989), 195–224.
60. Robert C. Merton, *Theory of Rational Option Pricing*, (1900).

61. P. Mörters and Y. Peres, *Brownian Motion*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2010.
62. Ravi Myneni, *The Pricing of the American Option*, The Annals of Applied Probability **2** (1992), no. 1, 1–23.
63. Salih N. Neftci, *An Introduction to the Mathematics of Financial Derivatives*, second ed., Academic Press, 2000.
64. ———, *Principles of Financial Engineering*, second ed., Academic Press, 2008.
65. Bernt Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, 2010.
66. Andrea Pascucci, *PDE and Martingale Methods in Option Pricing*, Springer, 2011.
67. L. Perko, *Differential Equations and Dynamical Systems*, Texts in Applied Mathematics, Springer New York, 2013.
68. Nicolas Privault, *Understanding Markov Chains: Examples and Applications*, Springer, 2013.
69. Sidney I. Resnick, *Adventures in Stochastic Processes*, Birkhäuser Boston, 1992.
70. Walter Rudin, *Principles of Mathematical Analysis*, third ed., Springer, 1976.
71. Paul A. Samuelson, *Rational Theory of Warrant Pricing*, Industrial Management Review **6** (1965), no. 2, 13–39.
72. ———, *Foundations of Economic Analysis*, enlarged ed., Harvard University Press, 1983.
73. Stephen M. Schaefer and Eduardo S. Schwartz, *A Two-Factor Model of the Term Structure: An Approximate Analytical Solution*, Journal of Financial and Quantitative Analysis (1984).
74. William Sharpe, Gordon J. Alexander, and Jeffrey W. Bailey, *Investments*, sixth ed., Prentice Hall, 1998.
75. A.N. Shiryayev, *On Analytical Methods In Probability Theory*, Selected Works of A. N. Kolmogorov (A.N. Shiryayev, ed.), Mathematics and Its Applications (Soviet Series), vol. 26, Springer Netherlands, 1992, pp. 62–108 (English).
76. Steven E. Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*, Springer, 2004.
77. ———, *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*, Springer, 2005.
78. M. Spivak, *Calculus*, Publish or Perish, 2008.
79. J. Michael Steele, *Stochastic Calculus and Financial Applications*, Springer, 2010.
80. Daniel W. Stroock, *An Introduction to Markov Processes*, Graduate Texts in Mathematics, Springer Berlin Heidelberg, 2013.
81. D.W. Stroock and S.R.S. Varadhan, Springer, 1979.
82. Suresh M. Sundaresan, *Continuous-Time Methods in Finance: A Review and an Assessment*, The Journal of Finance **LV** (2000), no. 4, 1569–1622.
83. K. Taira, *Semigroups, Boundary Value Problems and Markov Processes*, Springer Monographs in Mathematics, Springer, 2004.
84. Terence Tao, *Nonlinear Dispersive Equations: Local and Global Analysis*, Conference Board of the Mathematical Sciences. Regional Conference Series in Mathematics, no. 106, American Mathematical Society.
85. ———, *An Introduction to Measure Theory*, Graduate Studies in Mathematics, American Mathematical Society, 2011.
86. S.R.S. Varadhan, *Lectures on Diffusion Problems and Partial Differential Equations*, Tata Institute of Fundamental Research Lectures on Mathematics and Physics: Mathematics, Springer-Verlag, 1980.
87. ———, *Probability Theory*, Courant Lecture Notes, Courant Institute of Mathematical Sciences, 2001.
88. David Williams, *Probability with Martingales*, third ed., Cambridge University Press, 1976.
89. P. Wilmott, S. Howison, and J. Dewynne, *The Mathematics of Financial Derivatives: A Student Introduction*, Cambridge University Press, 1995.
90. Paul Wilmott, *Paul Wilmott on Quantitative Finance*, second ed., vol. 1, 2, &3, Wiley, 2006.
91. Kosaku Yosida, *Functional Analysis*, Classics in Mathematics, Cambridge University Press, 1995.